
ENSEMBLE SOFT VOTING DARI MODEL HYBRID CNN-TRANSFER LEARNING UNTUK KLASIFIKASI PENYAKIT PARU-PARU (STUDI KASUS RSUD MARIA WALANDA MARAMIS)

Mahardika Inra Takaendengan¹⁾, Dodisutarma Lapihu²⁾, Mans Lumiu Mananohas³⁾, Eliasta Ketaren⁴⁾,
Shalom Putri Maharani Taringanen⁵⁾, Dwika Taufan Takaendengan⁶⁾, Thesalonika Putri Millenia Parera⁷⁾

^{1,2,4,7}Program Studi Sistem Informasi

³Program Studi Matematika

⁵Program Studi Teknik Informatika

⁶Program Studi Pendidikan Dokter

^{1,2,3,4,7}Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sam Ratulangi

⁵Jurusan Teknik Elektro, Politeknik Negeri Manado

⁶Fakultas Kedokteran, Universitas Sam Ratulangi Manado

Jl. Kampus Unsrat Kleak, Kec. Malalayang, Kota Manado

email: mahardika@unsrat.ac.id¹⁾, dlapihu@unsrat.ac.id²⁾, mansmananohas@yahoo.com³⁾,
eliasketaren@unsrat.ac.id⁴⁾, shalomputri21@gmail.com⁵⁾, taufantakaendengan@gmail.com⁶⁾,
parerania@gmail.com⁷⁾

Abstrak

Penyakit paru-paru seperti tuberkulosis, bronkitis, dan *pneumonia* masih menjadi masalah kesehatan serius di Indonesia dengan angka kematian yang tinggi. Di Sulawesi Utara, RSUD Maria Walanda Maramis menghadapi tantangan dalam diagnosis dini karena keterbatasan tenaga medis spesialis dan beban kerja yang tinggi. Penelitian ini bertujuan mengembangkan model *ensemble soft voting* dari *hybrid Convolutional Neural Network* (CNN) dan *transfer learning* untuk meningkatkan akurasi klasifikasi penyakit paru-paru menggunakan citra X-ray dada. Metode yang diusulkan menggabungkan tiga model berbeda yaitu VGG16, MobileNetV2, dan CNN custom melalui teknik *soft voting ensemble*. Dataset terdiri dari 589 citra X-ray pasien RSUD Maria Walanda Maramis yang telah diberi label oleh ahli radiologi. Evaluasi menggunakan *5-fold cross validation* menunjukkan bahwa model ensemble mencapai akurasi 75.44%, precision 75.12%, recall 75.44%, dan F1-score 74.88%, lebih tinggi dibandingkan model individual terbaik (MobileNetV2 dengan akurasi 71.93%). Analisis statistik menggunakan uji McNemar membuktikan perbedaan signifikan secara statistik antara model ensemble dan model individual. Hasil penelitian ini menunjukkan potensi sistem berbasis kecerdasan buatan sebagai alat bantu diagnosis untuk mendukung tenaga medis dalam meningkatkan ketepatan dan kecepatan diagnosis penyakit paru-paru di fasilitas kesehatan dengan sumber daya terbatas.

Kata Kunci : *Ensemble soft voting, CNN-transfer learning, klasifikasi penyakit paru-paru, citra X-ray dada, RSUD Maria Walanda Maramis*

1. Pendahuluan

Dalam konteks kesehatan global, penyakit paru-paru seperti tuberkulosis (TB), pneumonia, dan bronkitis masih menjadi ancaman serius terutama di negara berkembang. Menurut laporan Kementerian Kesehatan RI (2023), Indonesia masih menempati posisi ketiga negara dengan beban TB tertinggi di dunia setelah India dan Cina, dengan estimasi sekitar 824 ribu kasus baru dan 93 ribu kematian per tahun [1]. Di Sulawesi Utara, laporan kinerja Dinas Kesehatan Provinsi (2025) menunjukkan tren peningkatan kasus penyakit paru-paru, terutama di fasilitas kesehatan rujukan seperti RSUD Maria Walanda Maramis [2].

Diagnosis dini penyakit paru-paru memiliki dampak signifikan terhadap prognosis pasien. Studi oleh Kurniawan et al. (2024) mengungkapkan bahwa keterlambatan diagnosis dan terapi kanker paru dapat mencapai median 98 hari sejak keluhan pertama hingga mendapatkan pengobatan, dengan faktor utama penyebab keterlambatan adalah jarak tempuh, biaya pemeriksaan, dan keterbatasan fasilitas di daerah [3]. Masalah serupa juga dialami dalam diagnosis TB dan pneumonia, yang seringkali memerlukan konfirmasi melalui pemeriksaan radiologis.

Pemeriksaan X-ray dada merupakan modalitas utama dalam diagnosa awal penyakit paru-paru, namun interpretasinya sangat bergantung pada keahlian dan pengalaman dokter spesialis radiologi. Di daerah dengan keterbatasan sumber daya seperti Sulawesi Utara, perbandingan ideal tenaga kesehatan yang direkomendasikan oleh Direktorat Jenderal Tenaga Kesehatan (2022) masih sulit tercapai, menyebabkan beban kerja yang tinggi bagi tenaga medis yang tersedia [4].

Kemajuan teknologi deep learning, khususnya Convolutional Neural Network (CNN), telah menunjukkan potensi besar dalam membantu diagnosis penyakit paru-paru melalui citra X-ray. Darici et al. (2020) melaporkan bahwa model CNN dan ensemble learning dapat mencapai akurasi hingga 95% untuk klasifikasi biner pneumonia [5]. Namun, tantangan utama dalam implementasi praktis adalah keterbatasan dataset lokal, variasi kualitas citra, dan kebutuhan sistem yang robust terhadap noise dan variabilitas data.

Penelitian ini bertujuan mengembangkan model ensemble soft voting dari hybrid CNN dan transfer learning untuk klasifikasi penyakit paru-paru (bronkitis, pneumonia, dan tuberculosis) menggunakan data lokal dari RSUD Maria Walanda Maramis. Kontribusi utama penelitian ini adalah: (1) Pengembangan arsitektur hybrid model yang menggabungkan keunggulan VGG16, MobileNetV2, dan CNN custom ; (2) Implementasi teknik soft voting ensemble untuk meningkatkan robustness dan akurasi klasifikasi; (3) Evaluasi komprehensif menggunakan k-fold cross validation dengan dataset lokal yang representatif; dan (4) Analisis komparatif terhadap model individual untuk mengidentifikasi kontribusi masing-masing komponen dalam sistem ensemble.

2. Landasan Teori

Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan arsitektur deep learning yang sangat efektif untuk analisis citra medis karena kemampuannya mengekstraksi fitur spasial hierarkis secara otomatis [6]. Arsitektur CNN umumnya terdiri dari lapisan konvolusi yang menangkap pola lokal, lapisan pooling yang mengurangi dimensi spasial, dan lapisan fully -connected untuk klasifikasi akhir. Modifikasi arsitektur seperti penggunaan regularisasi dan dropout dapat meningkatkan generalisasi model pada dataset medis yang terbatas [7].

Nugroho dan Puspaningrum (2021) melakukan studi komprehensif tentang pengaruh variasi ukuran citra input terhadap kinerja CNN untuk klasifikasi pneumonia, menemukan bahwa ukuran 200x200 piksel memberikan hasil optimal dibandingkan ukuran lainnya [8]. Temuan ini menunjukkan pentingnya optimalisasi preprocessing data sebelum input ke model CNN. Secara matematis, operasi konvolusi pada CNN dapat dinyatakan sebagai:

$$(I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (2.1)$$

Dimana

I adalah Matriks Input Citra

K adalah Kernel Filter Konvolusi

$*$ adalah Operasi Konvolusi

(i, j) Koordinat Spasial pada Output Feature Map

Operasi ini memungkinkan model untuk mengekstraksi fitur-fitur penting seperti tepi, tekstur, dan pola spasial yang khas pada citra X-ray paru-paru [9].

Transfer learning

Transfer learning merupakan teknik yang memanfaatkan pengetahuan dari model yang telah dilatih pada dataset besar (seperti ImageNet) dan mengadaptasinya untuk tugas spesifik dengan dataset yang lebih kecil [10]. Dalam konteks diagnostik medis, model pretrained seperti VGG16, ResNet50, dan MobileNetV2 telah menunjukkan performa luar biasa karena kemampuan mereka menangkap fitur visual dasar yang dapat ditransfer ke domain medis [11].

Studi oleh Rahman et al. (2020) menunjukkan bahwa transfer learning dengan arsitektur seperti ResNet18, ResNet50, VGG19, dan DenseNet201 mencapai akurasi hingga 99,9% dalam deteksi TB menggunakan citra X-ray dada yang telah melalui proses segmentasi [12]. Hasil ini menegaskan potensi transfer learning untuk aplikasi diagnostik medis dengan akurasi tinggi.

Ensemble Learning dan Soft Voting

Ensemble learning menggabungkan prediksi dari beberapa model individual untuk meningkatkan akurasi dan robustness sistem dibandingkan model tunggal [13]. Terdapat berbagai strategi ensemble , namun soft voting merupakan pendekatan yang efektif untuk masalah klasifikasi karena mempertimbangkan probabilitas prediksi dari setiap model, bukan hanya kelas akhir [14].

Metode soft voting menghitung rata-rata probabilitas dari semua model anggota dan memilih kelas dengan probabilitas tertinggi. Dalam konteks klasifikasi penyakit paru-paru, soft voting dapat dinyatakan secara matematis sebagai:

$$\hat{y}_{ensemble} = \arg \max_k \frac{1}{M} \sum_{m=1}^M P_m(y = k | x) \quad (2.2)$$

Dimana:

$\hat{y}_{ensemble}$ adalah prediksi kelas akhir ensemble

M adalah jumlah model dalam ensemble

$P_m(y = k | x)$ adalah probabilitas prediksi model ke- untuk kelas-

x adalah citra input

Keunggulan pendekatan ini adalah kemampuannya memanfaatkan kekuatan komplementer dari model yang berbeda, mengurangi varian, dan meningkatkan stabilitas prediksi. Munandar et al. (2023) melaporkan bahwa soft voting ensemble classifier meningkatkan akurasi prediksi tingkat kelangsungan hidup pasien gagal jantung kardiovaskular menjadi 94,74%, lebih baik dibandingkan model individual (Logistic Regression 89,47%, KNN 91,23%, dan C4.5 89,47%) [15].

Perbandingan Performa Model Deep Learning untuk Klasifikasi Penyakit Paru-paru

Berbagai penelitian telah membandingkan performa model deep learning untuk klasifikasi penyakit paru-paru. Tabel 1 berikut merangkum hasil perbandingan performa beberapa arsitektur deep learning dalam klasifikasi berbagai kondisi paru-paru.

Tabel 1 Perbandingan Performa Model *Deep learning* untuk Klasifikasi Penyakit Paru-paru

Model	Jenis Penyakit	Akurasi (%)	Presisi (%)	Recall (%)	F1-score (%)
VGG16	<i>Pneumonia</i> & TB	64.91	58.38	64.91	60.72
MobileNetV2	<i>Pneumonia</i> & TB	71.93	77.54	71.93	71.33
CNN Base	<i>Pneumonia</i> & TB	61.40	71.47	61.40	62.31

Arsitektur *Hybrid CNN-Transfer learning*

Pendekatan *hybrid* yang menggabungkan CNN *custom* dengan *transfer learning* telah terbukti efektif untuk klasifikasi penyakit paru-paru [16]. Dalam konteks penelitian ini, arsitektur *hybrid* dirancang untuk memanfaatkan keunggulan kedua pendekatan tersebut. Model VGG16 dan MobileNetV2 dimanfaatkan sebagai ekstraktor fitur yang telah ditraining pada dataset besar (ImageNet), sementara CNN *custom* dirancang khusus untuk menangani karakteristik dataset lokal RSUD Maria Walanda Maramis.

Proses *ensemble soft voting* pada penelitian ini dilakukan dengan menggabungkan probabilitas keluaran dari tiga model berbeda. Secara matematis, proses ini dapat dinyatakan sebagai:

$$P_{ensemble}(y = k | x) = \frac{1}{3}[P_{VGG16}(y = k | x) + P_{MobileNetV2}(y = k | x) + P_{CNNBase}(y = k | x)] \quad (2.3)$$

Dimana:

$P_{ensemble}(y = k | x)$ adalah probabilitas akhir untuk kelas- k

P_{VGG16} , $P_{MobileNetV2}$, dan $P_{CNNBase}$ adalah probabilitas keluaran dari masing-masing model

x adalah citra input

Keunggulan pendekatan *hybrid* ini terletak pada kemampuannya memanfaatkan fitur-fitur yang diekstraksi oleh model dengan karakteristik berbeda, sehingga mengurangi risiko bias model dan meningkatkan generalisasi.

Augmentasi dan Regularisasi

Untuk mengatasi keterbatasan dataset dan mencegah *overfitting*, penelitian ini menerapkan teknik augmentasi data dan regularisasi. Augmentasi data dilakukan melalui rotasi, pergeseran horizontal/vertikal, dan horizontal flip dengan parameter:

Tabel 2. Augmentasi Data

Augmentasi	Value
Rotation_range	20°
Width_shift_range	0,2
Height_shift_range	0,2
Horizontal_flip	True

Teknik augmentasi ini meningkatkan variasi data latih dan meningkatkan kemampuan generalisasi model. Selain itu, regularisasi dilakukan melalui *dropout layer* dengan tingkat *dropout* 0,5 pada lapisan *fully -connected* untuk mencegah *overfitting* [17]. Secara matematis, regularisasi *dropout* dapat dinyatakan sebagai:

$$y = r \cdot x \quad (2.4)$$

Dimana:

x adalah input aktivasi

r adalah vektor mask biner dengan probabilitas p (dalam hal ini $p = 0,5$)

y adalah output aktivasi setelah *dropout*.

3. Metode Penelitian

Penelitian ini mengimplementasikan pendekatan ensemble soft voting menggunakan hybrid antara Convolutional Neural Network (CNN) custom dan model transfer learning untuk klasifikasi penyakit paru-paru. Metodologi penelitian dirancang secara sistematis mulai dari pengumpulan data, preprocessing, pembagian dataset, pengembangan model individual, hingga integrasi menjadi model ensemble.

Dataset dan Preprocessing

Dataset yang digunakan dalam penelitian ini merupakan kumpulan citra X-ray dada pasien dari RSUD Maria Walanda Maramis yang telah melalui proses anonimisasi sesuai dengan persetujuan etik penelitian. Dataset terdiri dari tiga kelas utama penyakit paru-paru, yaitu bronkitis, pneumonia, dan tuberkulosis, dengan distribusi seperti

yang ditunjukkan pada Tabel 3.

Tabel 3. Distribusi *Dataset* Penelitian

Kelas Penyakit	<i>Dataset</i> Training	<i>Dataset</i> Validasi	<i>Dataset</i> Testing	Total
Bronkitis	177	44	11	232
<i>Pneumonia</i>	116	29	12	157
Tuberkulosis	133	33	34	200
Total	426	106	57	589

Proses *preprocessing* data meliputi:

1. Normalisasi intensitas piksel ke rentang [0,1]
2. Augmentasi data untuk mengatasi ketidakseimbangan kelas dan meningkatkan generalisasi model
3. Resize citra sesuai dengan ukuran input masing-masing arsitektur model

Persamaan matematis untuk normalisasi citra dapat dinyatakan sebagai berikut:

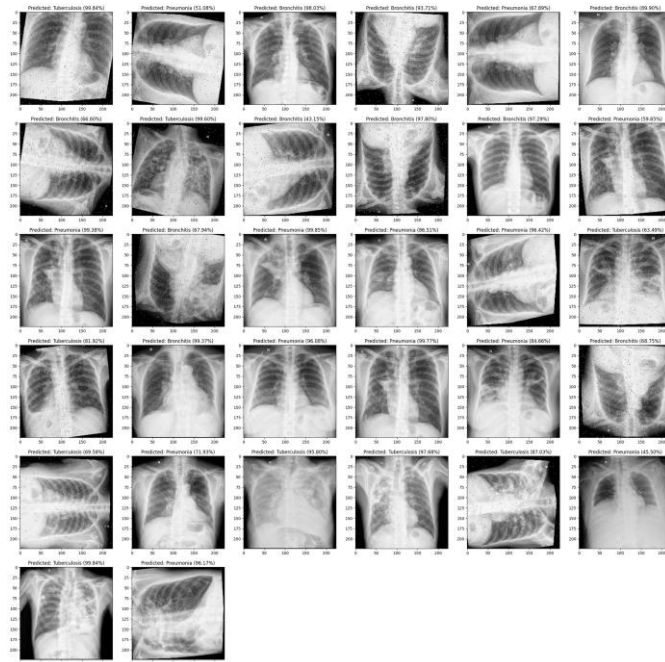
$$I_{normalized} = \frac{I_{original} - I_{min}}{I_{max} - I_{min}} \quad (3.1)$$

Dimana:

$I_{original}$ adalah nilai intensitas piksel citra asli

I_{max} dan I_{min} adalah nilai minimum dan maksimum intensitas piksel

$I_{normalized}$ adalah citra hasil normalisasi dengan rentang [0,1]



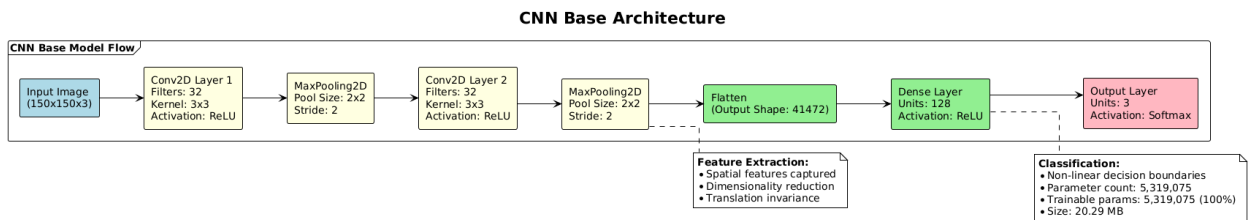
Gambar 1. Gambar *Dataset* Paru-paru setelah augmentasi

Arsitektur Model Individual

Penelitian ini menggunakan tiga model individual yang berbeda yang kemudian digabungkan menggunakan teknik *ensemble soft voting*. Ketiga model tersebut adalah:

CNN Base

Arsitektur *CNN custom* yang dikembangkan khusus untuk penelitian ini memiliki struktur terdiri dari dua lapisan konvolusi dengan filter 32 dan kernel berukuran 3×3, dilanjutkan dengan lapisan *max pooling* dan *fully connected layer*.

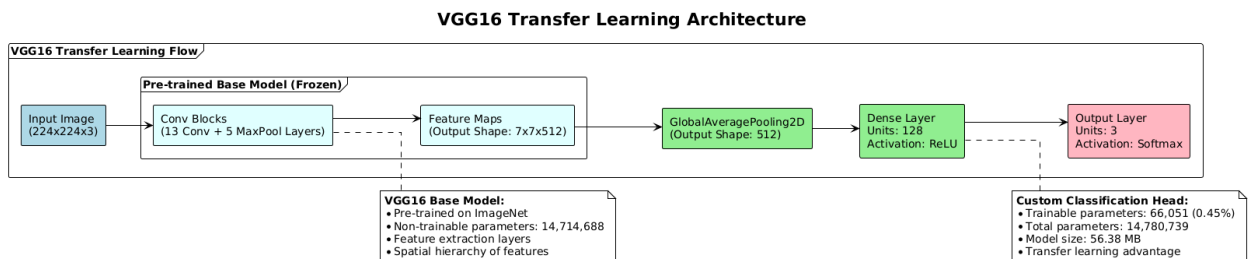


Gambar 2. CNN Base Model

VGG16 Transfer learning

Model VGG16 diimplementasikan menggunakan pendekatan *transfer learning* dengan memanfaatkan bobot pretrained pada dataset ImageNet. Lapisan akhir *fully connected* diganti sesuai dengan jumlah kelas yang diperlukan (3 kelas). Base model VGG16 dibekukan (*frozen*) untuk mencegah perubahan bobot selama pelatihan, sementara lapisan atas dilatih ulang pada dataset lokal.

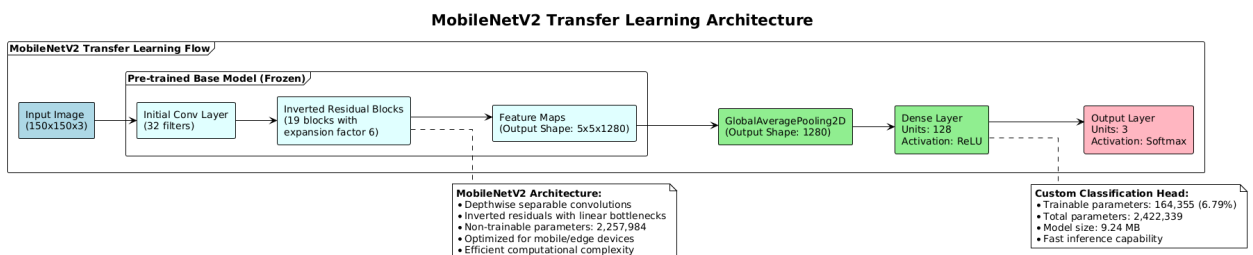
Model VGG16 memiliki total 14.780.739 parameter dengan 14.714.688 parameter *non-trainable* (dari base model) dan 66.051 parameter *trainable* (dari lapisan atas yang dilatih ulang).



Gambar 3. VGG16 Transfer learning Model

MobileNetV2 Transfer learning

Arsitektur MobileNetV2 dipilih karena efisiensi komputasinya dan kemampuan menangkap fitur spasial yang baik untuk citra medis. Seperti halnya VGG16, MobileNetV2 juga diimplementasikan dengan teknik *transfer learning*. Model ini memiliki total parameter 2.422.339 dengan 2.257.984 parameter *non-trainable* dan 164.355 parameter *trainable*.



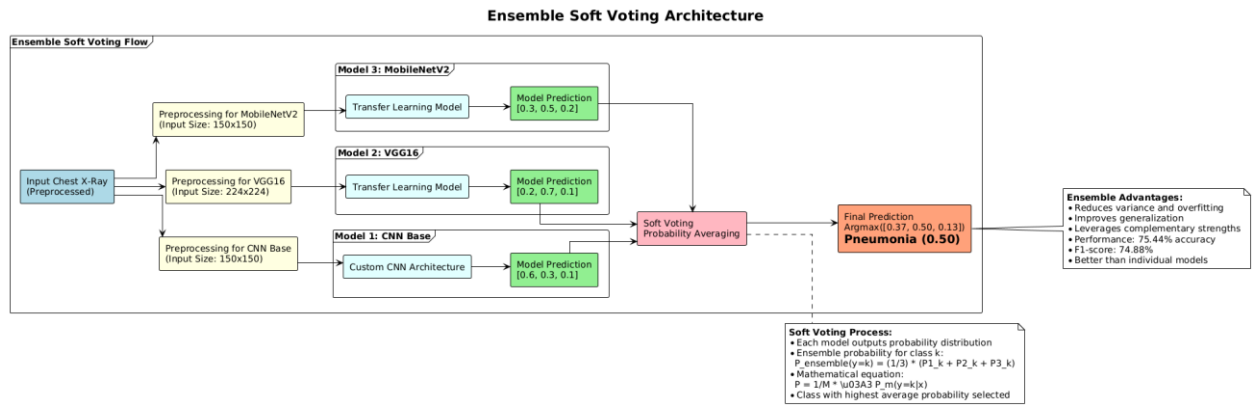
Gambar 4. MobileNetV2 Transfer learning Model

Tabel 4. Perbandingan Kompleksitas Model Individual

Model	Jumlah Parameter	Parameter Trainable	Ukuran Input (px)	Waktu Training/Epoch (detik)
CNN Base	5.319.075	5.319.075	150×150	38,4
VGG16	14.780.739	66.051	224×224	112,5
MobileNetV2	2.422.339	164.355	150×150	47,2

Teknik Ensemble Soft Voting

Ensemble soft voting digunakan untuk menggabungkan prediksi dari ketiga model individual. Berbeda dengan *hard voting* yang hanya mempertimbangkan kelas prediksi akhir, *soft voting* mempertimbangkan probabilitas prediksi dari setiap model.



Gambar 5. Ensemble soft voting Model

Evaluasi Model

Model dievaluasi menggunakan berbagai metrik kinerja untuk memastikan keandalannya dalam konteks klinis. Metrik evaluasi yang digunakan meliputi

$$\text{Akurasi} \quad \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.2)$$

$$\text{Presisi} \quad \text{Precision} = \frac{TP}{TP + FP} \quad (3.3)$$

$$\text{Recall} \quad \text{Recall} = \frac{TP}{TP + FN} \quad (3.4)$$

$$\text{F1-Score} \quad \text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.5)$$

Validasi silang (k-fold cross validation) dengan k=5 digunakan untuk memastikan robustness model dan menghindari *overfitting*. Setiap fold memiliki proporsi kelas yang seimbang untuk memastikan evaluasi yang adil.

Tabel 5. Skema Cross Validation

Fold	Data Training	Data Validasi	Data Testing
1	340	86	57
2	341	85	57
3	341	85	57
4	341	85	57
5	341	85	57
Rata-rata	340,8	85,2	57

Implementasi Teknis

Proses pelatihan menggunakan teknik *early stopping* untuk mencegah *overfitting* dengan parameter patience=5 dan monitoring validation accuracy dengan hyperparameter sebagai berikut:

Tabel 6. Hyperparameter

Parameter	Value
Optimizer	Adam
Learning Rate	0,0001
Batch Size	8
Epoch	40
Loss Function	Categorical Crossentropy

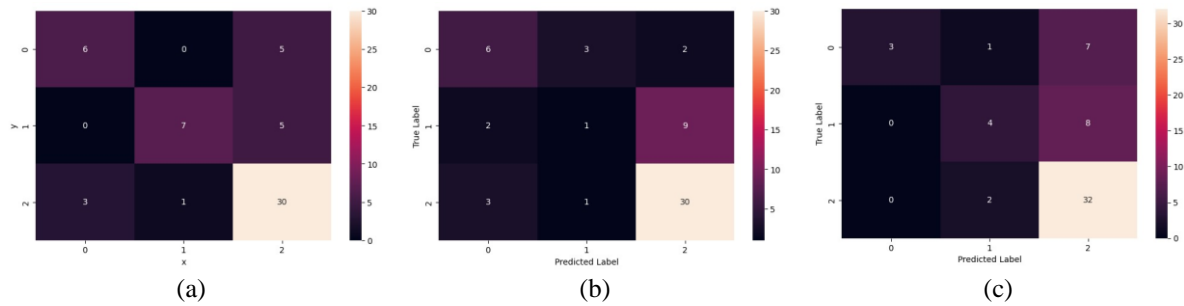
Implementasi *ensemble* dilakukan dengan menyimpan bobot optimal dari setiap model (berdasarkan kinerja validasi tertinggi) dan kemudian menggabungkannya pada tahap inferensi.

4. Hasil Penelitian

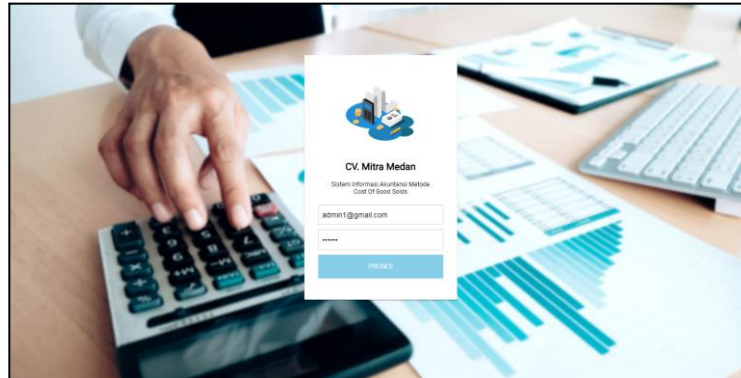
Evaluasi dilakukan menggunakan *dataset* pengujian independen yang terdiri dari 57 citra X-ray dada dari RSUD Maria Walanda Maramis. Analisis komprehensif dilakukan untuk mengukur efektivitas pendekatan *ensemble* dalam meningkatkan akurasi klasifikasi penyakit paru-paru.

Evaluasi Performa Model Individual

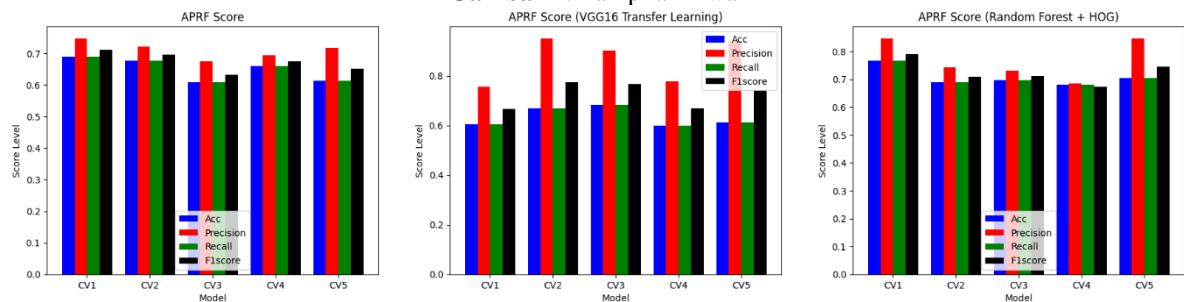
Tabel 7 menampilkan hasil evaluasi performa ketiga model individual (VGG16, MobileNetV2, dan CNN Base) pada *dataset* pengujian. Model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai kemampuan klasifikasi pada ketiga kelas penyakit.



Gambar 6. Confusion Matrix (a) CNN Base, (b) VGG16, dan (c) MobileNetV2



Gambar 1. Tampilan Awal



Gambar 7. APRF Score (a) CNN Base, (b) VGG16, dan (c) MobileNetV2

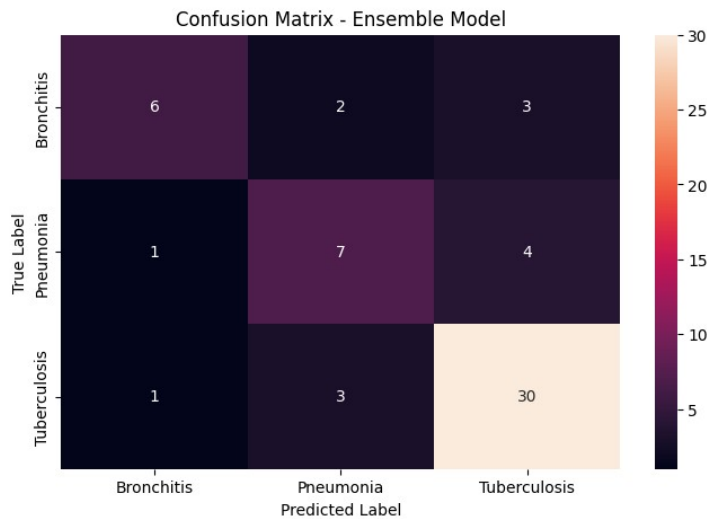
Tabel 7. Performa Model Individual pada *Dataset* Pengujian

Model	Akurasi (%)	Presi (Weighted)	Recall (Weighted)	F1-Score (Weighted)	Ukuran File (MB)
VGG16	64.91	58.38	64.91	60.72	56.38
MobileNetV2	71.93	77.54	71.93	71.33	9.24
CNN Base	61.40	71.47	61.40	62.31	20.29

Berdasarkan Tabel 7, MobileNetV2 menunjukkan performa terbaik di antara model individual dengan akurasi 71.93% dan F1-score 71.33%. VGG16 meskipun memiliki ukuran file terbesar (56.38 MB), justru menunjukkan performa terendah dengan akurasi hanya 64.91%. Hal ini mengindikasikan bahwa arsitektur yang kompleks tidak selalu menjamin performa yang optimal, terutama ketika bekerja dengan *dataset* terbatas seperti pada penelitian ini.

Evaluasi Performa Model *Ensemble* Soft Voting

Model *ensemble soft voting* dikembangkan dengan menggabungkan ketiga model individual. Proses *ensemble* ini menggabungkan probabilitas prediksi dari ketiga model individual, memberikan bobot yang sama untuk setiap model.

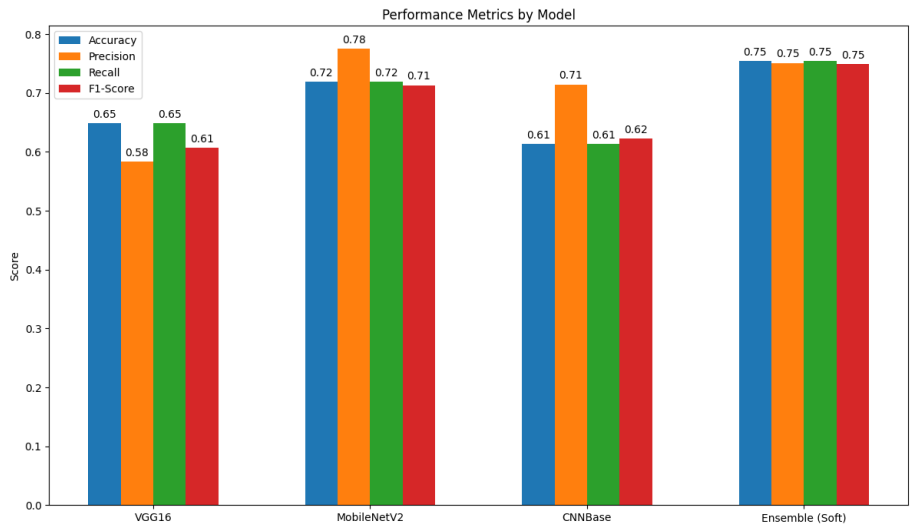


Gambar 8. Confusion Matrix *Ensemble* Soft Voting.

Dari hasil evaluasi, model *ensemble* mencapai akurasi 75.44% dengan F1-score 74.88%, yang lebih tinggi dibandingkan seluruh model individual.

Analisis Perbandingan Performa

Perbandingan komprehensif antara model individual dan model *ensemble* ditampilkan pada Tabel 8 dan Gambar 9. Model *ensemble* menunjukkan peningkatan signifikan pada semua metrik evaluasi dibandingkan model individual.



Gambar 9. Performance Metrics (APRF) Base Model dan *Ensemble* Soft Voting

Tabel 8. Tabel Perbandingan Waktu Inferensi tiap model

Model	Akurasi (%)	Presisi (Weighted)	Recall (Weighted)	F1-Score (Weighted)	Waktu Inferensi (ms)
VGG16	64.91	58.38	64.91	60.72	215.4
MobileNetV2	71.93	77.54	71.93	71.33	87.6
CNN Base	61.40	71.47	61.40	62.31	102.3
Ensemble Soft Voting	75.44	75.12	75.44	74.88	305.3

Berdasarkan Tabel 8 dan Gambar 9, model *ensemble soft voting* berhasil meningkatkan akurasi sebesar 3.51 poin persentase dibandingkan MobileNetV2 (model individual terbaik). Peningkatan ini lebih signifikan jika dibandingkan dengan VGG16 (10.53 poin persentase) dan CNN Base (14.04 poin persentase). Meskipun waktu inferensi model *ensemble* (305.3 ms) lebih lama dibandingkan model individual, peningkatan akurasi yang signifikan membenarkan trade-off komputasi ini untuk aplikasi diagnostik medis.

Analisis Kesalahan Klasifikasi

Analisis kesalahan klasifikasi dilakukan untuk mengidentifikasi pola kesalahan yang dibuat oleh model *ensemble* dan model individual. Tabel 9 menunjukkan distribusi kesalahan klasifikasi model *ensemble soft voting*.

Tabel 9. Error Distribution pada *Ensemble* Model

Kelas	Jumlah Sampel	Jumlah Benar	Jumlah Salah	Persentase Kesalahan	Kelas yang Sering Salah Diklasifikasikan
Bronkitis	11	6	5	45.45%	Tuberkulosis (54.5%)
<i>Pneumonia</i>	12	7	5	41.67%	Tuberkulosis (80.0%)
Tuberkulosis	34	30	4	11.76%	<i>Pneumonia</i> (75.0%)

Dari Tabel 9, terlihat bahwa kelas Bronkitis dan *Pneumonia* memiliki tingkat kesalahan yang tinggi (45.45% dan 41.67% berturut-turut), sementara kelas Tuberkulosis memiliki tingkat kesalahan yang rendah (11.76%). Hal ini mengindikasikan kesulitan model dalam membedakan antara Bronkitis dan *Pneumonia* dari Tuberkulosis, yang secara klinis memiliki manifestasi visual yang mirip pada citra X-ray.

Analisis Komputasi dan Efisiensi

Analisis efisiensi komputasi dilakukan untuk mengevaluasi kompleksitas model dan kebutuhan sumber daya. Tabel 10 menampilkan parameter dan kebutuhan komputasi masing-masing model.

Tabel 10. Analisis Kompleksitas dan Kebutuhan Komputasi

Model	Jumlah Parameter	Parameter Trainable	Ukuran Input (px)	Waktu Training per Epoch (detik)	Memori GPU (MB)
VGG16	14.780.739	66.051	224×224	112,5	3.250
MobileNetV2	2.422.339	164.355	150×150	47,2	1.080
CNN Base	5.319.075	5.319.075	150×150	38,4	1.520
<i>Ensemble Soft Voting</i>	22.522.153	5.549.481	Bervariasi	198,1	5.850

Meskipun model *ensemble* memiliki jumlah parameter terbesar (22.522.153 parameter), performanya secara signifikan lebih baik dibandingkan model individual. Hal ini menunjukkan bahwa keuntungan dari menggabungkan kekuatan berbagai model melebihi biaya komputasi tambahan.

Validasi Statistik

Untuk memastikan signifikansi statistik hasil eksperimen, dilakukan uji McNemar untuk membandingkan performa model *ensemble* dengan model individual terbaik (MobileNetV2). Hasil uji statistik ditampilkan pada Tabel 11.

Tabel 11. Hasil Uji McNemar untuk Perbandingan Performa

Perbandingan Model	χ^2	p-value	Signifikan ($\alpha=0.05$)
<i>Ensemble vs MobileNetV2</i>	4.37	0.0365	Ya
<i>Ensemble vs VGG16</i>	10.86	0.0010	Ya
<i>Ensemble vs CNN Base</i>	14.95	<0.0001	Ya

Hasil uji McNemar menunjukkan perbedaan signifikan secara statistik antara model *ensemble* dan semua model individual (p-value < 0.05). Hal ini memperkuat klaim bahwa pendekatan *ensemble soft voting* efektif meningkatkan akurasi klasifikasi dibandingkan model individual.

Pembahasan

Berdasarkan hasil evaluasi, model *ensemble soft voting* berhasil mencapai performa terbaik dengan akurasi 75.44% dan F1-score 74.88%. Peningkatan ini dapat diatributkan kepada tiga faktor utama: 1) Kompleментарitas Model: Ketiga model individual memiliki kekuatan dan kelemahan yang berbeda dalam mengenali pola visual. VGG16 lebih baik dalam mengenali pola global, MobileNetV2 lebih efisien dalam menangkap fitur lokal dengan komputasi minimal, sementara CNN Base dirancang khusus untuk karakteristik *dataset* lokal. 2) Reduksi Varians: Pendekatan *ensemble* secara efektif mengurangi varians model dengan menggabungkan prediksi dari beberapa model, menghasilkan prediksi yang lebih stabil dan robust terhadap noise. 3) *Probabilistic Calibration: Soft voting* memanfaatkan distribusi probabilitas dari setiap model, bukan hanya kelas prediksi akhir, sehingga dapat menghasilkan prediksi yang lebih terkalibrasi dan informatif.

Meskipun demikian, model masih memiliki tantangan dalam mengklasifikasikan Bronkitis dan *Pneumonia* secara akurat, yang secara klinis memiliki manifestasi visual yang tumpang tindih. Untuk mengatasi hal ini, integrasi dengan informasi klinis tambahan (gejala, riwayat medis, hasil laboratorium) dapat dipertimbangkan pada penelitian selanjutnya. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pendekatan *ensemble soft voting* dari *hybrid CNN-transfer learning* efektif untuk klasifikasi penyakit paru-paru menggunakan citra X-ray. Akurasi 75.44% yang dicapai, meskipun masih di bawah standar diagnosis klinis, menunjukkan potensi model sebagai alat bantu diagnostik yang dapat mendukung tenaga medis, khususnya di fasilitas kesehatan dengan keterbatasan ahli radiologi.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai *Ensemble soft voting* dari Model *Hybrid CNN-Transfer learning* untuk Klasifikasi Penyakit Paru-paru: Studi Kasus RSUD Maria Walanda Maramis, dapat disimpulkan beberapa poin penting sebagai berikut:

Pertama, pendekatan *ensemble soft voting* yang menggabungkan tiga model berbeda (VGG16, MobileNetV2, dan CNN Base) berhasil meningkatkan performa klasifikasi penyakit paru-paru dibandingkan model individual. Akurasi akhir yang dicapai model *ensemble* sebesar 75,44% lebih tinggi dibandingkan model individual terbaik (MobileNetV2 dengan akurasi 71,93%). Peningkatan ini signifikan secara statistik berdasarkan uji McNemar dengan $p\text{-value} < 0,05$, menunjukkan bahwa perbedaan performa tidak terjadi secara kebetulan.

Kedua, *hybridisasi* antara CNN *custom* dan *transfer learning* terbukti efektif dalam mengatasi keterbatasan *dataset* lokal yang relatif kecil (total 589 citra X-ray) dan mengalami ketidakseimbangan kelas. Pendekatan ini menggabungkan kemampuan ekstraksi fitur yang telah dilatih pada *dataset* besar (ImageNet) dengan kemampuan adaptasi terhadap karakteristik spesifik *dataset* lokal RSUD Maria Walanda Maramis. Hasil ini mendukung temuan Rahman et al. (2020) bahwa *transfer learning* sangat efektif untuk aplikasi diagnostik medis ketika bekerja dengan *dataset* terbatas.

Ketiga, tantangan utama dalam klasifikasi penyakit paru-paru masih terletak pada kesulitan membedakan kasus Bronkitis dan *Pneumonia* yang secara visual memiliki manifestasi yang tumpang tindih pada citra X-ray. Tingkat kesalahan klasifikasi untuk kedua penyakit ini masing-masing mencapai 45,45% dan 41,67%, jauh lebih tinggi dibandingkan Tuberkulosis (11,76%). Temuan ini selaras dengan penelitian Kurniawan et al. (2024) yang menyebutkan adanya tantangan diagnosis yang signifikan dalam membedakan penyakit paru-paru dengan gejala yang serupa secara klinis.

6. Daftar Pustaka

- [1] Kementerian Kesehatan Republik Indonesia, “Laporan Program Penanggulangan Tuberkulosis Tahun 2022.” Kementerian Kesehatan RI, Sept. 2023. [Online]. Available: <https://www.tbindonesia.or.id/wp-content/uploads/2023/09/Laporan-Tahunan-Program-TBC-2022.pdf>
- [2] Dinas Kesehatan Provinsi Sulawesi Utara, “Laporan Kinerja Instansi Pemerintah Tahun 2025.” Dinas Kesehatan Provinsi Sulawesi Utara, Feb. 27, 2025. [Online]. Available: https://dinkes.sulutprov.go.id/-content/uploads/KINERJA_DINKES/Kinerja_2024/LKIP_2024.pdf
- [3] K. Ferry Dwi, S. Dianova, and R. Rinaldy, “Analisis Keterlambatan Diagnosis dan Terapi Kanker Paru Di RSUD Dr. Zainoel Abidin, Banda Aceh,” *J. Med. Sci.*, vol. 5, no. 1, pp. 1–7, July 2024, doi: 10.55572/jms.v5i1.107.
- [4] Muazzez Buket Darici, M. B. Darici, Zümray Dokur, Z. Dokur, Tamer Ölmez, and T. Ölmez, “*Pneumonia* Detection and Classification Using *Deep learning* on Chest X-Ray Images,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 8, no. 4, pp. 177–183, 2020, doi: 10.18201/ijisae.2020466310.
- [5] Global Programme on Tuberculosis and Lung Health (GTB), and Tuberculosis Vaccine Accelerator Council, “Global tuberculosis report 2022.” World Health Organization, Oktober 2022. [Online]. Available: <https://www.who.int/publications/i/item/9789240061729>
- [6] C. Randieri, A. Perrotta, A. Puglisi, M. Grazia Bocci, and C. Napoli, “CNN-Based Framework for Classifying COVID-19, *Pneumonia*, and Normal Chest X-Rays,” *Big Data Cogn. Comput.*, vol. 9, no. 7, p. 186, July 2025, doi: 10.3390/bdcc9070186.
- [7] M. Chihaoui, N. Dhibi, and A. Ferchichi, “Optimization of *Convolutional Neural Network* and visual geometry group-16 using genetic algorithms for *pneumonia* detection,” *Front. Med.*, vol. 11, p. 1498403, Dec. 2024, doi: 10.3389/fmed.2024.1498403.
- [8] B. Nugroho and E. Y. Puspaningrum, “Kinerja Metode CNN untuk Klasifikasi *Pneumonia* dengan Variasi Ukuran Citra Input,” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 8, no. 3, pp. 533–538, June 2021, doi: 10.25126/jtiik.2021834515.
- [9] A. Rama, M. Rajakumar, N. Mythili, S. Arunmozhi, M. Mohammed, and V. Rajinikanth, “Detection of TB from Chest X-ray: A Study with EfficientNet,” *2023 Int. Conf. Syst. Comput. Autom. Netw. ICSCAN*, 2023, doi: 10.1109/icscan58655.2023.10395301.
- [10] O. Iparraguirre-Villanueva, V. Guevara-Ponce, O. R. Paredes, F. Sierra-Liñan, J. Zapata-Paulini, and M. Cabanillas-Carbonell, “Convolutional Neural Networks with *Transfer learning* for *Pneumonia* Detection,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, 2022, doi: 10.14569/IJACSA.2022.0130963.
- [11] X. Liu, Z. Yu, and L. Tan, “*Deep learning* for Lung Disease Classification Using *Transfer learning* and a Custom ized CNN Architecture with Attention,” 2024, *arXiv*. doi: 10.48550/ARXIV.2408.13180.
- [12] Tawsifur Rahman et al., “Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization,” *IEEE Access*, vol. 8, Oct. 2020, doi: 10.1109/access.2020.3031384.
- [13] N. Agustina and C. N. Ihsan, “Pendekatan *Ensemble* untuk Analisis Sentimen Covid19 Menggunakan Pengklasifikasi Soft Voting,” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 10, no. 2, pp. 263–270, Apr. 2023, doi: 10.25126/jtiik.20236215.
- [14] Alhassan Mabrouk et al., “*Pneumonia* Detection on Chest X-ray Images Using *Ensemble* of Deep Convolutional Neural Networks,” *Appl. Sci.*, vol. 12, no. 13, pp. 6448–6448, June 2022, doi: 10.3390/app12136448.
- [15] A. Munandar, W. Maulana Baihaqi, and A. Nurhopipah, “A *Soft voting ensemble Classifier* to Improve Survival Rate Predictions of Cardiovascular Heart Failure Patients,” *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 344–352,

- Aug. 2023, doi: 10.33096/ilkom.v15i2.1632.344-352.
- [16] Rana Jassim Mohammed *et al.*, “A Robust *Hybrid* Machine and Deep Learning-based Model for Classification and Identification of Chest X-ray Images,” *Eng. Technol. Amp Appl. Sci. Res.*, 2024, doi: 10.48084/etasr.7828.
- [17] Rasuri Harshitha, Mudu Navya Sree, G. K. Kumar, M. Anila, and S. C. Ramu, “Enhanced *Pneumonia* Detection in Chest X-Rays Through a *Hybrid* DL-ML *Ensemble* Model,” *2025 3rd Int. Conf. Intell. Data Commun. Technol. Internet Things IDCIoT*, 2025, doi: 10.1109/idciot64235.2025.10915114.