

Comparative Analysis of Convolutional Neural Network Architectures in Pneumonia Detection

Indah Putianik^{1*}, Wise Herowati^{2**}

* PJJ Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

** Kelompok Penelitian Kuantum Komputer dan Informatika Material, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
118202300052@mhs.dinus.ac.id¹, wise@dsn.dinus.ac.id²

Article Info

Article history:

Received 2026-04-08

Revised 2026-05-15

Accepted 2026-05-25

Keyword:

CNN,
MobileNetV2,
ResNet-50.

ABSTRACT

The lungs are one of the important organs for humans. One of the lung diseases that can potentially cause death is pneumonia. This study compares the performance of the Convolutional Neural Network (CNN), ResNet-50, and MobileNetV2 models in detecting pneumonia using chest X-ray images. The dataset used was obtained from Kaggle with a total of 2,000 chest X-ray images, which were divided into 70% training data, 15% validation data, and 15% test data. All images were resized to 256×256 pixels. The experiment was carried out 10 times with 20 training epochs and a batch size of 32. The results showed that the CNN model obtained an average accuracy of 95.76%, precision of 95.39%, recall of 96.20%, and F1-score of 95.77%. The ResNet-50 model produced an average accuracy of 96.99%, precision of 99.79%, recall of 94.20%, and F1-score of 96.90%. Meanwhile, MobileNetV2 achieved an average accuracy of 97.33%, precision of 97.23%, recall of 98.13%, and F1-score of 97.66%. ResNet-50 was better at minimizing False Positives, while MobileNetV2 was better at minimizing False Negatives. Overall, all three models performed well in detecting pneumonia, although there were variations in the confusion matrix results across experiments.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Paru-paru merupakan salah satu organ pernapasan dirongga dada yang sangat penting bagi manusia dalam pertukaran gas atau respirasi, yakni menghirup oksigen dan membuang karbon dioksida [1]. Masalah yang kerap terjadi dalam sistem pernapasan adalah kualitas udara yang terkontaminasi oleh bakteri yang dapat menyerang sistem pernapasan [2][3][4]. Saat ini, gangguan pada paru-paru menjadi salah satu kondisi yang bisa berujung pada kematian jika tidak ditangani dengan benar. Penyakit paru-paru yang diakibatkan oleh bakteri atau virus akan mengakibatkan masalah dalam bernafas dan melakukan kegiatan sehari-hari. Salah satu jenis penyakit yang menyerang paru-paru adalah pneumonia.[5].

Pada pasien penderita pneumonia, kumpulan kantong udara kecil di ujung saluran pernapasan di paru-paru (alveoli) akan meradang dan diisi dengan cairan atau nanah. Oleh karena itu, penderita mengalami gejala batuk, demam, kedinginan, nyeri dada dan sesak napas [6]. Pneumonia

disebabkan oleh gangguan yang dilakukan mikroba parasit terhadap sistem kekebalan tubuh. Hal ini mengakibatkan peradangan pada jaringan parenkim paru-paru dan menimbulkan pneumonia. Ini adalah jenis peradangan yang paling umum disebabkan oleh virus atau kuman yang ada di udara dan dihirup oleh manusia [7].

Pneumonia menyebabkan 81% kematian dalam dua tahun pertama kehidupan. Hampir semua kematian terkait pneumonia terjadi di negara berpendapatan rendah dan menengah, dengan Afrika Sub-Sahara menghadapi proporsi besar dari kasus pneumonia (43% kematian akibat pneumonia global) [8]. Menurut data WHO pada tahun 2019, pneumonia merupakan penyebab 14% kematian pada anak-anak usia dibawah 5 tahun dengan total kematian 740.180 orang [9].

Citra rontgen thorax merupakan salah satu yang digunakan dalam mengetahui gangguan pada paru-paru [3]. Untuk mendeteksi hasil rontgen thorax, maka tenaga medis melakukan diagnosa secara langsung melalui hasil citra

rontgen thorax karena dapat menyajikan visual keadaan organ tubuh dengan jelas [10].

Terdapat beberapa penelitian yang dilakukan menggunakan CNN dalam mendeteksi dan klasifikasi pada gambar. Pada penelitian yang dilakukan oleh Kurniawati mengenai identifikasi pneumonia pada balita melalui citra x-ray dengan metode CNN menunjukkan bahwa proses klasifikasi CNN dengan 15 epoch, 5 lapisan tersembunyi, dan 5700 data dapat mencapai akurasi pelatihan sebesar 92,48% dan validasi sebesar 91% [11]. Penelitian lain terkait CNN dilakukan oleh Khairunnisak untuk mendeteksi pneumonia dari citra rontgen dada menunjukkan akurasi 91,4%, presisi 91,7%, recall 97,3%, dan F1-score 94,4% [12].

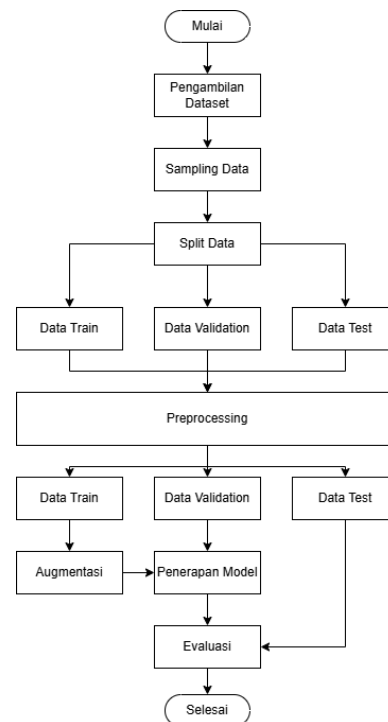
Kemudian terdapat beberapa penelitian mengenai ResNet-50. Penelitian yang dilakukan oleh Adzkiya mengenai ResNet-50 dan ResNet-101 pada citra pneumonia untuk mendeteksi pneumonia menggunakan ResNet-50 menunjukkan hasil presisi teraugmentasi sebesar 61% dan non augmentasi sebesar 49%. Nilai akurasi pada citra teraugmentasi sebesar 74% dan non augmentasi sebesar 65%. Pada ResNet-101 menunjukkan hasil presisi teraugmentasi sebesar 46% dan nonaugmentasi sebesar 46%. Nilai akurasi pada citra teraugmentasi sebesar 61% dan non augmentasi sebesar 61% [13]. Penelitian lain mengenai ResNet-50 dilakukan oleh Prasetyo & Ichwan mengenai deteksi penyakit pneumonia menggunakan ResNet-50 dan ResNet-152. Pada penelitian ini, ResNet-50 menunjukkan nilai rata-rata akurasi sebesar 88,7%, precision 88%, recall 89% dan f-measure 89%. Untuk ResNet-152 menunjukkan rata-rata hasil akurasi 84,1%, precision 83%, recall 85%, dan f-measure 84% [14].

Penelitian terkait MobileNetV2 dilakukan oleh Evita Setyaningrum dan Bambang Wirawan mengenai klasifikasi pneumonia menggunakan citra x-ray paru-paru menunjukkan hasil pengujian model mendapatkan akurasi 97,47% dengan recall kelas pneumonia adalah 99% [15]. Penelitian lain terkait MobileNetV2 dilakukan oleh M Choirul Amri tentang klasifikasi citra x-ray menunjukkan hasil evaluasi pada akurasi adalah 95,40%, presisi 95,62%, recall 95,40%, dan F1-score 95,46% [16].

Penerapan model deep learning pada penelitian ini yaitu CNN, ResNet-50, dan MobileNetV2, pada citra rontgen thorax bertujuan untuk mengklasifikasikan kondisi pneumonia dan normal sekaligus membandingkan kinerja masing-masing model serta menentukan model yang paling optimal dalam klasifikasi citra rontgen thorax untuk mendeteksi penyakit pneumonia.

II. METODE

Penelitian ini terdapat 7 tahapan yaitu pengambilan dataset, sampling data, split data (data training, validasi, data uji), preprocessing, augmentasi, penerapan model dan mengevaluasi model. Tahapan analisis ditampilkan pada Gambar 1 sebagai berikut.



Gambar 1. Tahapan Penelitian

A. Pengambilan dataset

Pada Gambar 1 yang menunjukkan tahapan penelitian pertama yaitu pengambilan dataset. Dataset diperoleh dari situs Kaggle berupa citra rontgen thorax (<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data>). Dataset yang terdapat pada situs ini berjumlah 5856 citra rontgen thorax.

B. Sampling data

Dari total 5856 citra rontgen thorax dilakukan proses sampling untuk mengambil 2000 citra sebagai dataset penelitian. Teknik yang digunakan adalah Random Under Sampling. Metode ini dapat digunakan untuk menghasilkan jumlah data yang sama untuk kelompok besar dan kelompok kecil, atau dapat digunakan untuk memperkecil kelompok besar hingga ukuran tertentu [17]. Pemilihan data dilakukan secara acak untuk mendapatkan 1000 citra Normal dan 1000 citra Pneumonia dari folder train untuk memastikan ketersediaan data cukup. Hal ini dilakukan karena distribusi data asli yang tidak seimbang dan data pada folder validasi sangat terbatas.

C. Split data menjadi data training, data uji dan data validasi

Pada Gambar 1 yang menunjukkan tahapan penelitian pada proses ketiga yang dilakukan adalah membagi data. Untuk memastikan evaluasi yang lebih stabil, dataset dibuat ulang dengan membagi data secara acak menjadi data training, data testing, dan data validasi. Data yang diambil berjumlah 2000 citra rontgen thorax dibagi dengan menggunakan random split dengan rasio 70:15:15. Dengan

70% digunakan sebagai data training, 15% digunakan sebagai data validasi, dan 15% digunakan untuk data uji. Maka 1400 data citra rontgen thorax yang terdiri dari 700 citra normal dan 700 citra rontgen thorax terinfeksi pneumonia digunakan sebagai data training. Kemudian 300 citra rontgen thorax menjadi data uji dan 300 data validasi, dengan distribusi kelas seimbang yaitu 150 normal dan 150 pneumonia pada setiap subset. Untuk mencegah bias pada salah satu kelas, pembagian data dilakukan secara acak dengan tetap mempertahankan proporsi kelas [18]. Distribusi data ditunjukkan pada table I.

TABEL I.
DISTRIBUSI DATA

Kelas	Data Training	Data Testing	Data Validasi	Total
Normal	700	150	150	1000
Pneumonia	700	150	150	1000
Total	1400	300	300	2000

D. Preprocessing

Pada Gambar 1 yang menunjukkan tahapan penelitian pada proses ke empat yang dilakukan adalah preprocessing. Citra rontgen thorax tersebut di-resize dan segmentasi agar lebih efisien saat melakukan proses training [19]. Semua citra rontgen thorax di resize ke ukuran 256x256 dan dinormalisasi.

E. Augmentasi

Pada Gambar 1 yang menunjukkan tahapan penelitian pada proses kelima yaitu Augmentasi. Proses augmentasi dilakukan pada data training untuk meningkatkan variasi citra dari data yang sudah tersedia [20]. Teknik augmentasi yang digunakan pada penelitian ini yaitu rotasi (rotation range = 10), pergeseran horizontal (width shift range=0.05), pergeseran vertical (height shift range = 0.05), zoom (zoom range = 0.05) dan pembalikan horizontal (horizontal flip=true). Pada area kosong yang dihasilkan dari transformasi diatasi dengan metode pengisian piksel fill mode='nearest'.

F. Penerapan metode

Pada Gambar 1 yang menunjukkan tahapan penelitian Dimana proses ketiga yang dilakukan adalah penerapan metode CNN, ResNet-50, MobileNetV2.

1) CNN

CNN akan melakukan klasifikasi dari data uji. Pada tahap ini, citra rontgen thorax akan di input. Ukuran yang digunakan adalah 256x256 pixel. Kemudian dilakukan *convolutional layer* untuk melihat pola-pola penting dan menghasilkan feature map. Hasil pada *feature map* akan diaktivasi menggunakan ReLu. Kemudian dilakukan *pooling layer* yang bertujuan untuk memperkecil ukuran pada gambar. Hasil dari *pooling layer* akan di flatten menjadi vector untuk dimasukan ke tahap akhir yaitu klasifikasi [11].

2) ResNet-50

ResNet-50 merupakan salah satu arsitektur CNN yang memiliki 50 layer. Citra rontgen thorax diinput dan di resize dengan ukuran 256x256. Citra thorax yang digunakan adalah RGB yang kemudian di augmentasi. Lapisan pertama dilakukan proses *convolutional* menggunakan filter ukuran 7x7 dengan 64 channel dengan nilai *stride* 2. Hasil konvolusi diproses dengan *batch normalization* dan menambahkan fungsi aktivasi *ReLU*. Kemudian diakhiri dengan operasi *max pooling* dengan filter 3x3 dan *stride* bernilai 2. Tahapan selanjutnya adalah *identity block*. Operasi ini menggunakan filter 1x1, 3x3, dan 1x1 dengan channel 64,64, dan 256. Proses ini dilakukan dua kali. Kemudian dilakukan kembali *convolutional block* dan tiga *identity block* dengan filter yang sama dan jumlah channel yang meningkat. Ditahap ke empat dilakukan *convolutional block* dengan lima *identity block*. Pada tahap kelima, dilakukan satu *convolutional block* dan dua *identity block*. Setelah semua tahapan selesai, maka dilakukan proses *average pooling* lalu proses *flatten* sehingga mendapatkan output array satu dimensi. Output diproses menggunakan lapisan *fully connected* untuk mendapatkan hasil akhir [21].

3) MobileNetV2

Arsitektur MobileNetV2 adalah jaringan syaraf konvolusional yang memanfaatkan struktur residual terbalik serta penggabungan tingkat bottleneck residual yang disesuaikan untuk digunakan pada perangkat mobile. Untuk menyaring fitur, digunakan lapisan konvolusi kedalaman yang ringan yang melalui lapisan ekspansi menengah. MobileNetV2 memiliki rancangan lapisan yang terhubung sepenuhnya dengan 32 filter dan 19 lapisan bottleneck [22]. Tahap pertama struktur ini adalah konvolusi dua dimensi menggunakan 32 filter. Filter tersebut beroperasi di atas gambar untuk menciptakan peta fitur yang mampu mendeteksi berbagai elemen dasar dalam gambar. Setelah konvolusi awal, arsitektur berlanjut dengan serangkaian lapisan bottleneck yang terdiri dari tujuh blok konvolusi. Setiap lapisan bottleneck menerapkan teknik konvolusi separable depthwise untuk meminimalkan jumlah parameter dan beban kalkulasi tanpa menurunkan performa. Metode ini terbagi menjadi dua tahap: konvolusi depthwise, yang melakukan konvolusi pada setiap channel secara individual, serta konvolusi pointwise, yang menggabungkan hasil dari konvolusi depthwise. Setelah melewati tahap bottleneck, konvolusi 1x1 dengan 1280 filter diterapkan untuk mengubah ukuran fitur agar lebih sesuai untuk langkah selanjutnya. Kemudian, pooling rata-rata dilakukan untuk mengurangi dimensi spasial dari peta fitur dengan menghitung nilai rata-rata di setiap area. Tahap terakhir terdiri dari konvolusi 1x1, yang mengintegrasikan

karakteristik yang diperoleh dari pooling rata-rata dan menyiapkannya untuk keputusan klasifikasi akhir. Hasil dari konvolusi ini akan diteruskan ke lapisan fully connected dan softmax untuk menghasilkan kemungkinan klasifikasi. [23].

G. *Evaluasi model*

Pada Gambar 1 yang menunjukkan tahapan penelitian dimana proses keempat yang dilakukan adalah mengevaluasi model. Setelah dilakukan penerapan model, diperoleh dua jenis klasifikasi yaitu normal dan pneumonia. Pada tahapan evaluasi, digunakan *Confusion Matrix* yang didapatkan dari penerapan model. *Confusion Matrix* digunakan untuk menilai kinerja model. *Confusion Matrix* adalah tabel evaluasi yang mencakup empat nilai yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [24]. Gambaran struktur umum *Confusion Matrix* ditunjukkan pada tabel II:

TABEL II.
CONFUSION MATRIX

	Prediksi	Normal	Pneumonia
Aktual	Normal	TN	FP
	Pneumonia	FN	TP

Hasil penerapan CNN, ResNet-50, dan MobileNetV2 akan dievaluasi. Evaluasi yang dilakukan adalah mengecek *F1-score*, *accuracy*, *precision*, *recall*, untuk mengetahui performa CNN, ResNet-50, MobileNetV2 pada data uji [2].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Accuracy adalah indikator yang mencerminkan seberapa tepat sistem dapat mengenali objek dari semua yang diproses selama proses pendeteksian. *Precision* adalah nilai yang dipakai untuk mengukur kemampuan model dalam memberikan prediksi yang tepat pada kategori positif dari semua prediksi positif yang dihasilkan. *Recall* adalah ukuran yang dipakai untuk mengevaluasi sejauh mana model klasifikasi mampu mendeteksi semua data positif yang sebenarnya. *F1-Score* adalah gabungan antara nilai presisi dan recall yang bertujuan untuk mengevaluasi kinerja model klasifikasi secara komprehensif [25].

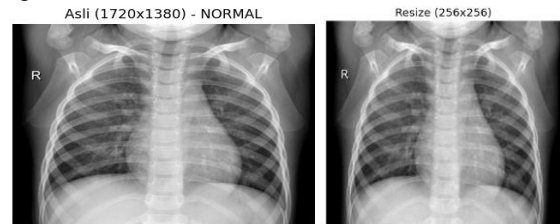
III. HASIL DAN PEMBAHASAN

Data citra rontgen thorax yang digunakan merupakan data dari Kaggle. Dataset yang digunakan adalah folder train yang terdiri dari 2 kelompok yaitu citra normal dan citra

pneumonia. Data diambil secara acak sebanyak 2000 citra rontgen thorax.

Pada penelitian ini digunakan Google Colaboratory Python dengan GPU T4 untuk mengimplementasikan CNN. Framework yang digunakan adalah TensorFlow.

Data train pada citra rontgen thorax di augmentasi dengan rotation range = 10, width shift range = 0.05, height shift range = 0.05, zoom range = 0.05, horizontal flip = True, fill mode = 'nearest' dan seluruh data citra rontgen thorax di-resize dengan ukuran 256x256 pixel. Berikut adalah gambar yang telah di-resize:



Gambar 2. Resize citra thorax ke ukuran 256x256

Pada Gambar 2 merupakan visualisasi gambar setelah dilakukan resize, data dibagi menjadi tiga bagian yaitu data training, data uji, dan validasi. Data training yang digunakan adalah 1400 citra rontgen thorax. Data validasi yang digunakan adalah 300 citra rontgen thorax (150 citra rontgen normal dan 150 citra rontgen pneumonia) dan data uji yang digunakan sebanyak 300 citra rontgen thorax (150 citra rontgen normal dan 150 citra rontgen pneumonia).

A. *Implementasi CNN*

Setelah pembagian dataset, maka dilakukan arsitektur CNN menggunakan Google Colaboratory Python.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_1 (Conv2D)	(None, 125, 125, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d_2 (Conv2D)	(None, 60, 60, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 128)	0
flatten (Flatten)	(None, 115200)	0
dense (Dense)	(None, 128)	14,745,728
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Gambar 3. Layer pada CNN

Gambar 3 menunjukkan proses dalam pembangunan layer pada CNN memanfaatkan library keras_model_sequential. Tahapan pertama yaitu mendeteksi fitur pada citra dengan Convolutional 2D. Pada tahapan ini dilakukan ekstraksi dan menghasilkan 32 feature map dengan ukuran kernel 3x3. Fungsi aktivasi ReLu pada Convolutional 2D adalah untuk mengenali pola-pola pada citra.

Kemudian dilakukan proses max pooling untuk mereduksi feature map dan menjaga hasil dari ekstraksi. selanjutnya dilakukan Convolutional 2D tahap 2 dengan menggunakan 64 feature map dengan ukuran kernel 3x3. Hal ini berfungsi untuk mengurangi kompleksitas. tahapan

berikutnya dilakukan flatten layer yang berfungsi untuk transisi dari Convolutional 2D ke connected layer (lapisan terhubung).

Kemudian dilakukan konvolusi tahap 3 dengan 128 feature map agar model dapat mengenali fitur lebih lanjut pada citra rontgen thorax.

Hasil feature map, diubah menjadi bentuk vector melalui tahapan flatten layer. Kemudian tahapan yang dilakukan selanjutnya yaitu dense layer dengan 128 unit. Dense layer berfungsi untuk memperkuat klasifikasi dari ekstraksi sebelumnya. Kemudian digunakan dropout layer untuk menghindari overfitting. Neuron yang digunakan yaitu sebesar 0,5 atau 50%.

Tahapan terakhir yaitu output layer dengan fungsi sigmoid, yaitu mengubah data citra menjadi output 0 dan 1 sesuai dengan hasil klasifikasi dari model. Pada klasifikasi ini, nilai yang mendekati 0 akan dikategorikan sebagai normal dan yang mendekati 1 dikategorikan sebagai pneumonia. Model ini menggunakan optimizer Adam agar training lebih cepat dan dapat menyesuaikan learning rate secara otomatis.

Berdasarkan penelitian yang dilakukan oleh Maxwell, disebutkan bahwa reproducibility dapat ditingkatkan dengan mengatur random seed [26]. Reproducibility adalah hasil konsisten saat menggunakan input dataset yang sama,

tahapan komputasi, metode, dan kondisi analisis [27]. Maxwell meninjau 100 penelitian pada Deep Learning mengenai penginderaan jarak jauh yang tidak memberikan informasi lengkap dalam menilai akurasi pada model, terutama pada confusion matrix dan pembagian dataset. Menurut Maxwell, ada beberapa rekomendasi yang dapat dilakukan, yaitu image pre-processing seperti transformasi koordinat, metode resampling, meningkatkan kontras, menajamkan gambar, pengaturan GPU, mengatur random seed, data terdokumentasi dengan jelas, melakukan augmentasi atau transformasi. Model perlu dirincikan dan evaluasi untuk menentukan seberapa baik suatu model bekerja [26]. Karena adanya krisis reproduksibilitas dalam penelitian ilmiah, maka penting untuk mengikuti pedoman agar dapat mengurangi dan mengukur faktor yang menyebabkan hasil penelitian berubah-ubah [27].

Pada proses training data, digunakan 20 epoch dengan batch size 32 dan random seed 42. Tiap epoch terdapat 44 batch. Eksperimen dilakukan sepuluh 10 kali dengan kondisi yang sama untuk melihat konsistensi kinerja model. Berikut adalah hasil evaluasi model untuk 10 kali percobaan yang dijalankan dalam waktu yang berbeda-beda yang ditunjukkan pada table III.

TABEL III.
HASIL EVALUASI MODEL CNN TIAP PERCOBAAN

Percobaan	True Negative	True Positive	False Negative	False Positive	Durasi
1	146	142	8	4	7,72 menit
2	146	146	4	4	7,67 menit
3	145	142	8	5	7,68 menit
4	146	144	6	4	7,68 menit
5	146	145	5	5	7,78 menit
6	143	144	6	7	7,78 menit
7	143	145	5	7	7,82 menit
8	132	148	2	18	7,90 menit
9	144	144	6	6	7,62 menit
10	139	143	7	11	7,63 menit

Pada Tabel III menunjukan hasil evaluasi model dilakukan melalui 10 kali percobaan untuk memastikan konsistensi performa. Setiap percobaan mendapatkan nilai akurasi dan confusion matrix yang berbeda.

Pada 10 kali percobaan dengan random seed 42 yang telah ditetapkan, sesuai penelitian Maxwell, ditemukan variabilitas pada confusion matrix dan nilai akurasi pada setiap percobaan. namun ditemukan bahwa variasi waktu pelatihan pada model CNN ini menunjukkan kestabilan yang berada di kisaran waktu 7 menit. hal ini menunjukkan bahwa sifat non-deterministic pada GPU sesuai pada penelitian Renard [27]. Dari 10 kali percobaan yang dilakukan,

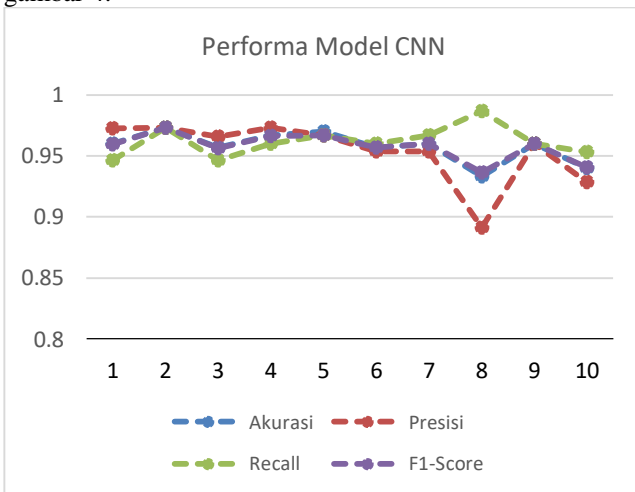
percobaan ke-2 menghasilkan nilai false negative berjumlah 4 gambar dan nilai false positive berjumlah 4 gambar. Hal ini menunjukkan keseimbangan yang sangat baik pada sensitivitas. Pada percobaan ke 8, nilai false positive sebanyak 18 gambar menunjukkan kenaikan nilai false positive dari percobaan sebelumnya (*false alarm*). Pada percobaan ke 10, nilai false positif berjumlah 11 gambar. Ini menunjukkan adanya ketidakstabilan pada iterasi tertentu.

Berikut adalah hasil perhitungan akurasi, presisi, recall dan F1-score pada tiap percobaan yang ditunjukkan pada tabel IV.

TABEL IV.
HASIL AKURASI, PRESISI, RECALL, dan F1-SCORE TIAP
PERCOBAAN PADA MODEL CNN

Percobaan	Akurasi	Presisi	Recall	F1-Score
1	0,96	0,9726	0,9467	0,9595
2	0,9733	0,9733	0,9733	0,9733
3	0,9567	0,9660	0,9467	0,9562
4	0,9667	0,9730	0,9600	0,9664
5	0,97	0,9667	0,9667	0,9667
6	0,9567	0,9536	0,9600	0,9568
7	0,9600	0,9539	0,9667	0,9603
8	0,9333	0,8916	0,9867	0,9367
9	0,96	0,9600	0,9600	0,9600
10	0,94	0,9286	0,9533	0,9408

Tabel IV menunjukkan hasil Akurasi yang didapat pada model CNN dengan ukuran citra 256x256 adalah 97,33% pada percobaan kedua. Hal ini menunjukkan model dapat memberikan prediksi benar sebanyak 97,33% dari data set yang digunakan. Recall tertinggi yang didapatkan yaitu 97,33% pada percobaan kedua yang menunjukkan model mampu mendeteksi dengan benar dan memiliki sensitivitas tinggi. Pada presisi, nilai yang didapatkan yaitu 98,67% pada percobaan kedelapan. Hal ini menunjukkan bahwa model masih ada prediksi true negative pada data set. Pada F1-score tertinggi, nilai yang didapatkan yaitu 97,33% pada percobaan kedua. Performa model CNN ditunjukkan pada gambar 4.

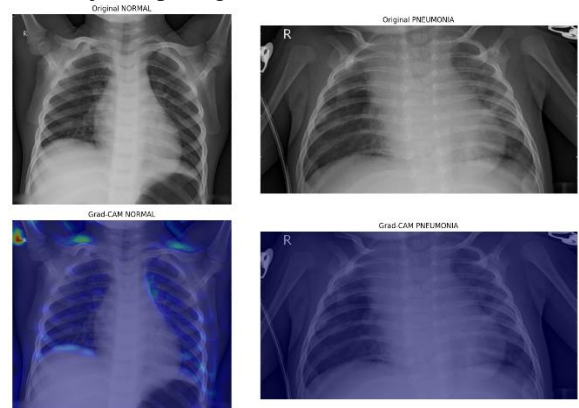


Gambar 4. Performa Model CNN Pada Tiap Percobaan

Gambar 4 menunjukkan grafik performa model CNN, Dimana model CNN menunjukkan nilai akurasi, presisi, recall, dan F1-score yang cukup stabil pada 10 kali percobaan yang dilakukan. Hal ini menunjukkan bahwa model CNN memiliki performa yang baik dalam mengklasifikasi citra pneumonia dan normal. Terdapat sedikit fluktuasi pada

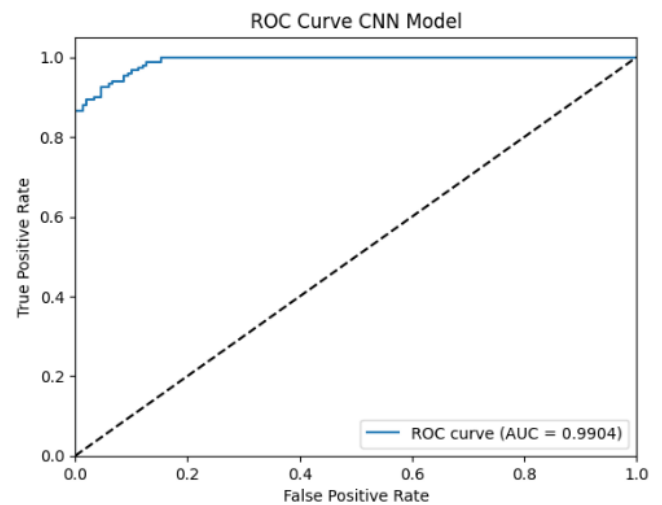
percobaan ke-8, namun secara keseluruhan model tetap menunjukkan hasil yang baik.

Selain percobaan 10 kali tersebut, dilakukan satu percobaan pengujian tambahan untuk menampilkan Grad-CAM, ROC-AUC dan confusion matrix. Grad-CAM model CNN ditunjukkan pada gambar 5.



Gambar 5. Grad-CAM Pada Model CNN

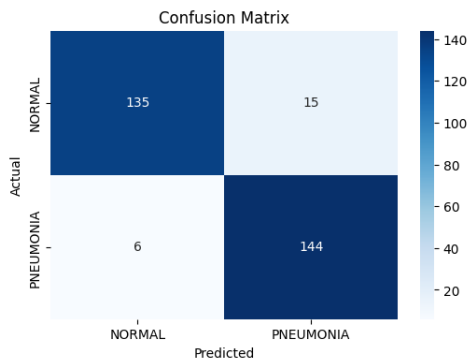
Gambar 5 menunjukkan bahwa Grad-Cam pada model CNN belum bisa fokus secara konsisten pada bagian paru-paru yang terinfeksi. Model CNN masih kurang kuat dalam mendeteksi area yang terinfeksi dan heatmap yang dihasilkan menyebar di area yang bukan paru-paru. Ini menunjukkan bahwa CNN sederhana memiliki keterbatasan dalam mengekstrak fitur kompleks dari citra thoraks. ROC-AUC pada Model CNN ditunjukkan pada gambar 6.



Gambar 6. ROC-AUC Model CNN

Gambar 6 menunjukkan ROC-AUC Model CNN. Berdasarkan hasil evaluasi dengan menggunakan ROC Curve, model CNN mendapatkan nilai AUC sebesar 0,9904. Nilai itu menunjukkan model memiliki kemampuan dalam mengklasifikasikan gambar dengan baik dan bisa membedakan antara kelas normal dan pneumonia secara

optimal. Confusion Matrix pada model CNN ditunjukkan pada gambar 7.



Gambar 7. Confusion Matrix Model CNN

Gambar 7 menunjukkan hasil confusion matrix pada model CNN. Hasil Confusion pada model CNN menunjukkan performa yang baik dalam memprediksi kelas pneumonia dan normal. Nilai 6 pada false negatif menunjukkan model masih terdapat beberapa citra yang diprediksi normal oleh model CNN. Nilai 15 pada false positive menunjukkan masih ada citra normal yang terprediksi pneumonia.

B. Implementasi ResNet-50

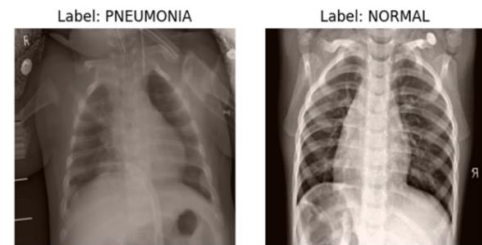
Pada implementasi ResNet-50 digunakan Google Colaboratory Python dengan GPU T4. Framework yang digunakan adalah TensorFlow.

ResNet merupakan salah satu arsitektur dari CNN dengan menggunakan 50 lapisan. Ditahap inisialisasi, bobot jaringan diambil dari pelatihan dataset ImageNet. Citra yang digunakan pada arsitektur ResNet-50 adalah 256x256 piksel dengan 3 channel (RGB). Lapisan-lapisan CNN dari ResNet-50 digunakan sebagai feature extractor dan dihubungkan kembali dengan fully connected layer yang disesuaikan.

Pada arsitektur ResNet-50 digunakan 32 batch size dengan 20 epoch. Data yang digunakan sebanyak 1000 citra normal dan 1000 citra pneumonia yang diambil secara acak. Data tersebut dibagi menjadi 70% data training, 15% data

uji, dan 15% data validasi. Sehingga didapatkan 1400 data training, 300 data validasi dan 300 data uji. Data training tersebut di augmentasi dengan rotation range = 10, width shift range = 0.05, height shift range = 0.05, zoom range = 0.05, horizontal flip = True, fill mode = 'nearest'.

Berikut adalah hasil augmentasi pada data training yang ditunjukkan pada gambar 8.



Gambar 8. Citra thorax setelah proses augmentasi

Gambar 8 merupakan citra yang telah di augmentasi pada data training. Kemudian dilakukan implementasi ResNet-50 dengan transfer learning. Implementasi ResNet-50 menggunakan bobot awal dari ImageNet, Adam Optimizer dengan learning rate 0,00001 dan random seed 42.

Pada tahap awal, seluruh layer dibekukan (*freeze*) untuk mempertahankan fitur dasar yang telah dipelajari. Proses feature extraction dilakukan sebanyak 10 epoch menggunakan optimizer Adam. Kemudian dilakukan proses fine-tuning dengan membuka 30 layer terakhir. Pada proses ini menggunakan learning rate 0,00001 dengan optimizer Adam agar proses berjalan lebih stabil. Pada proses ini juga dilakukan proses reduce learning rate secara otomatis apabila nilai validation loss tidak membaik selama 3 epoch. Tahapan fine tuning dilakukan sebanyak 20 epoch.. Eksperimen dilakukan sebanyak 10 kali dengan kondisi yang sama untuk melihat konsistensi kinerja model.

Berikut adalah hasil tiap percobaan pada implementasi ResNet-50 yang ditunjukkan pada tabel V.

TABEL V.
HASIL EVALUASI MODEL RESNET-50 TIAP PERCOBAAN

Percobaan	True Negative	True Positive	False Negative	False Positive	Durasi
1	149	146	4	1	13,63 menit
2	149	146	4	1	14,88 menit
3	150	144	6	0	14,60 menit
4	150	143	7	0	15,33 menit
5	149	144	6	1	14,38 menit
6	150	136	14	0	13,98 menit
7	150	135	15	0	14,22 menit
8	150	139	11	0	16,07 menit
9	150	139	11	0	16 menit
10	150	141	9	0	13,8 menit

Pada tabel V menunjukkan Hasil evaluasi model ResNet-50 tiap percobaan. Berdasarkan sepuluh percobaan yang dilakukan, model ResNet-50 mampu memprediksi dengan

baik dan kesalahan yang dilakukan sangat sedikit pada false positive (normal terprediksi pneumonia). Nilai False Negative tertinggi terdapat di percobaan ke-7. Nilai False

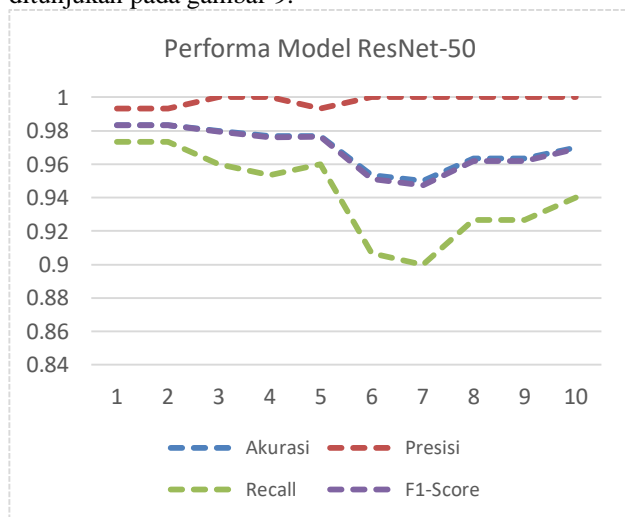
Positive hanya muncul dipercobaan ke-1, ke-2, dan ke-5. Secara keseluruhan, performa klasifikasi pada model ini baik meskipun ada fluktuasi pada tiap percobaan.

Berikut adalah hasil perhitungan akurasi, presisi, recall dan F1-score pada tiap percobaan yang ditunjukkan pada tabel VI.

TABEL VI.
HASIL AKURASI, PRESISI, RECALL, DAN F1-SCORE TIAP PERCOBAAN PADA MODEL RESNET-50

Percobaan	Akurasi	Presisi	Recall	F1-Score
1	0,9833	0,9932	0,9733	0,9832
2	0,9833	0,9932	0,9733	0,9832
3	0,9800	1,0000	0,9600	0,9796
4	0,9767	1,0000	0,9533	0,9761
5	0,9767	0,9931	0,9600	0,9763
6	0,9533	1,0000	0,9067	0,9510
7	0,9500	1,0000	0,9000	0,9474
8	0,9633	1,0000	0,9267	0,9619
9	0,9633	1,0000	0,9267	0,9619
10	0,9700	1,0000	0,9400	0,9691

Berdasarkan tabel VI yang menunjukkan hasil akurasi, presisi, recall dan f1-score pada model ResNet-50. Model menunjukkan hasil terbaik dengan nilai akurasi 98,33%. Ini menandakan ResNet-50 dapat mengklasifikasikan data test dengan benar. Recall tertinggi yang didapatkan yaitu 97,33% yang menunjukkan model mampu mendeteksi dengan benar dan memiliki sensitivitas tinggi. Pada presisi, nilai tertinggi yang didapatkan yaitu 100%. Pada F1-score, nilai tertinggi yang didapatkan yaitu 98,32%. Performa model ResNet-50 ditunjukkan pada gambar 9.

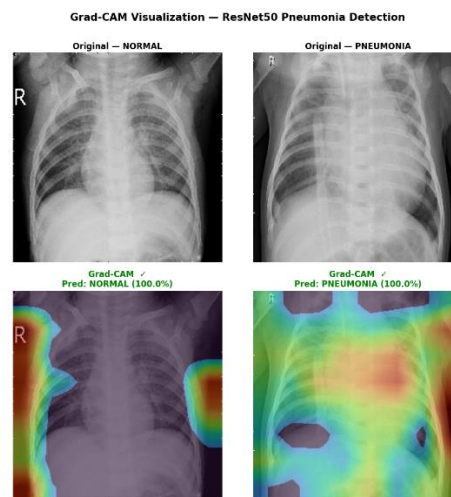


Gambar 9. Performa model ResNet-50 pada tiap percobaan

Gambar 9 menunjukkan performa model ResNet-50 pada 10 kali percobaan. Grafik menunjukkan tingkat presisi yang sangat tinggi serta stabil. Namun, nilai recall mengalami fluktuasi yang cukup signifikan, terutama pada percobaan

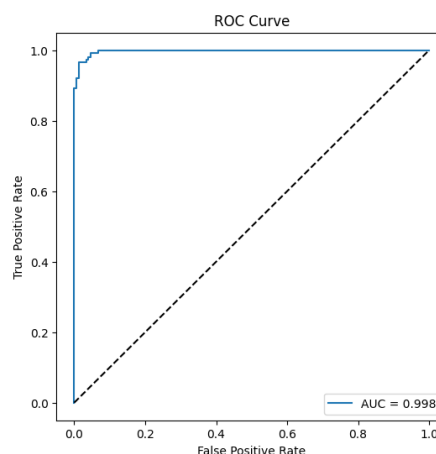
ke-5 dan ke-6 yang mengalami penurunan. Secara keseluruhan, model ini masih memerlukan penstabilan, terutama pada nilai recall agar hasil kerjanya lebih konsisten antar pengujian.

Selain percobaan 10 kali tersebut, dilakukan satu percobaan pengujian tambahan untuk menampilkan Grad-CAM, ROC-AUC dan confusion matrix. Grad-CAM model CNN ditunjukkan pada gambar 10.



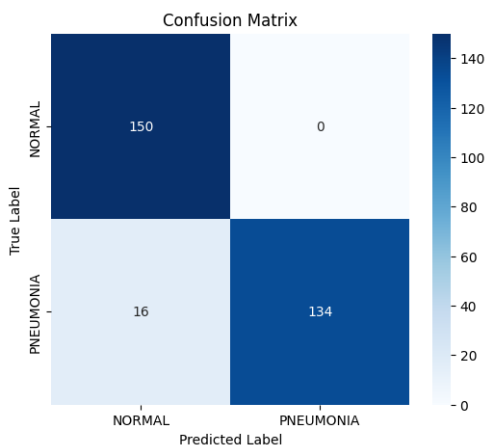
Gambar 10. Grad-CAM pada model ResNet-50

Gambar 10 menunjukkan Grad-CAM pada model ResNet-50. Pada citra normal, model ResNet-50 berhasil mendeteksi kelas normal dengan benar. Pada citra pneumonia, model ResNet-50 berhasil mendeteksi kelas pneumonia dengan benar. Model ResNet-50 menunjukkan fokus pada area paru-paru yang terinfeksi. Namun masih ada area berwarna merah dan kuning diluar area paru-paru pada citra kelas normal dan pneumonia. Hal ini menandakan model masih belum sepenuhnya fokus pada area paru-paru dan perlu ditingkatkan untuk lebih fokus mendeteksi area yang paru-paru. ROC-AUC pada Model ResNet-50 ditunjukkan pada gambar 11.



Gambar 11. ROC-AUC Model ResNet-50

Gambar 10 menunjukkan ROC-AUC model ResNet-50. Kurva ROC menunjukkan bahwa model ResNet-50 memiliki performa sangat baik dalam membedakan kelas normal dan kelas pneumonia. Nilai AUC yaitu 0,998 menunjukkan kemampuan klasifikasi yang baik pada model ResNet-50. Confusion Matrix pada model ResNet-50 ditunjukkan pada gambar 12.



Gambar 12. Confusion Matrix model ResNet-50

Gambar 12 menunjukkan confusion matrix model ResNet-50. Berdasarkan hasil confusion matrix. Nilai false positive bernilai 0 menunjukkan tidak ada citra normal yang terprediksi pneumonia. Nilai False Negative sebanyak 16 menunjukkan model masih belum mendeteksi secara sempurna. Secara keseluruhan model menunjukkan performa yang baik dalam klasifikasi.

C. Implementasi MobileNetV2

Citra yang digunakan pada arsitektur MobileNetV2 adalah 256x256 piksel dengan 3 channel (RGB). Jumlah batch yang digunakan adalah 32 dengan jumlah epoch 20. Data yang digunakan sebanyak 1000 citra normal dan 1000 citra pneumonia yang dibagi menjadi 70% data training, 15% data uji, dan 15% data validasi. Sehingga diperoleh 1400 data training, 300 data validasi, dan 300 data uji. Data train tersebut di augmentasi dengan rotation range = 10, width shift range = 0.05, height shift range = 0.05, zoom range = 0.05, horizontal flip = True, fill mode = 'nearest'.

Kemudian dilakukan implementasi MobileNetV2. Implementasi MobileNetV2 menggunakan transfer learning dengan pembobotan ImageNet, Adam Optimizer dengan learning rate 0,00001 dan random seed 42. Pada tahap awal implementasi MobileNetV2, model terlebih dahulu membekukan (*freeze*) seluruh layer pada *base model* untuk mempertahankan fitur yang telah dipelajari. Selanjutnya dilakukan proses training selama 10 epoch dengan hanya melatih layer klasifikasi yang ditambahkan. Setelah tahap pelatihan selesai, dilakukan proses *fine-tuning* dengan membuka kembali 30 layer terakhir dari arsitektur MobileNetV2. Pada tahap ini model dilatih kembali selama 20 epoch. Eksperimen dilakukan sebanyak 10 kali dengan kondisi dan seed yang sama yaitu 42 untuk melihat konsistensi kinerja model.

Berikut adalah hasil tiap percobaan pada implementasi MobileNetV2 yang ditunjukkan pada tabel VII.

TABEL VII. HASIL EVALUASI MODEL MOBILENETV2 TIAP PERCOBAAN

Percobaan	True Negative	True Positive	False Negative	False Positive	Durasi
1	145	148	2	5	12,8 menit
2	138	148	2	12	12,6 menit
3	140	149	1	10	12,7 menit
4	148	148	2	2	13,3 menit
5	148	148	2	2	13,2 menit
6	147	147	3	3	12,8 menit
7	146	146	4	4	12,9 menit
8	150	146	4	0	13,3 menit
9	150	145	5	0	13,3 menit
10	145	147	3	5	13,1 menit

Pada tabel VII menunjukkan hasil evaluasi model MobileNetV2 tiap percobaan. Nilai False Negative pada sepuluh kali percobaan menunjukkan bahwa model mampu mendeteksi kasus pneumonia dengan baik. Hal ini terlihat dari jumlah kesalahan klasifikasi yang relatif kecil pada setiap percobaan. Pada proses pelatihan, model membutuhkan waktu yang relatif konsisten, yaitu berkisar antara 12 hingga 13 menit pada setiap percobaan. Hal ini menunjukkan bahwa proses pelatihan model berlangsung

secara stabil dan tidak mengalami perbedaan waktu komputasi yang signifikan pada setiap percobaan.

Dari percobaan pertama ke percobaan kedua, nilai False Positive meningkat signifikan. Pada percobaan pertama model MobileNetV2 mendeteksi 5 gambar False Positive, kemudian meningkat menjadi 12 gambar False Positive pada percobaan kedua. Hal ini mengindikasikan bahwa model terlalu agresif dalam memprediksi kelas Pneumonia, sehingga terdapat beberapa citra Normal yang salah terklasifikasi. Pada percobaan ke-4 hingga ke-10, nilai False

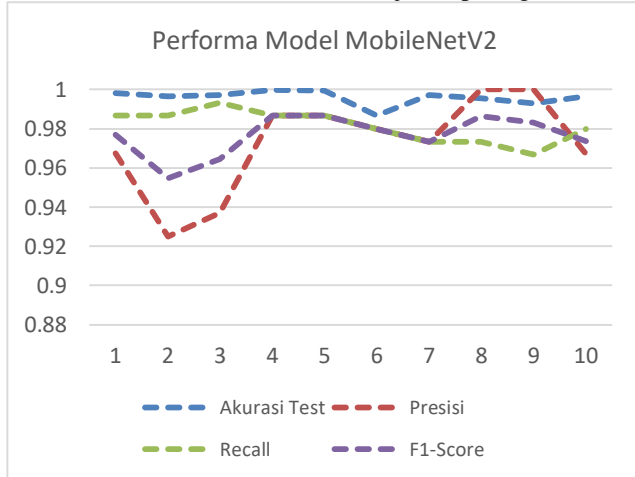
Negative dan False Positive menjadi lebih stabil. Model MobileNetV2 menunjukkan kemampuan dalam mengklasifikasi kedua kelas dan mulai konsisten.

Berikut adalah hasil perhitungan akurasi, presisi, recall dan F1-score pada tiap percobaan yang ditunjukkan pada tabel VIII.

TABEL VIII.
HASIL AKURASI, PRESISI, RECALL, dan F1-SCORE TIAP PERCOBAAN PADA MODEL MOBILENETV2

Percobaan	Akurasi	Presisi	Recall	F1-Score
1	0,9767	0,9673	0,9867	0,9769
2	0,9533	0,9250	0,9867	0,9548
3	0,9633	0,9371	0,9933	0,9644
4	0,9867	0,9867	0,9867	0,9867
5	0,9867	0,9867	0,9867	0,9867
6	0,9800	0,9800	0,9800	0,9800
7	0,9733	0,9733	0,9733	0,9733
8	0,9867	1,0000	0,9733	0,9865
9	0,9833	1,0000	0,9667	0,9831
10	0,9733	0,9671	0,9800	0,9735

Berdasarkan tabel VIII yang menunjukkan hasil akurasi, presisi, recall dan f1-score pada model MobileNetV2. Nilai Akurasi terbesar adalah 98,67%. Ini menandakan MobileNetV2 dapat mengklasifikasikan data test dengan benar. Recall yang didapatkan tertinggi pada percobaan ke tiga yaitu 99,33% yang menunjukkan model mampu mendeteksi dengan benar dan memiliki sensitivitas tinggi. Pada presisi, nilai tertinggi yang didapatkan yaitu 100%. Pada F1-score, nilai tertinggi yang didapatkan yaitu 98,67%. Performa model MobileNetV2 ditunjukkan pada gambar 13.

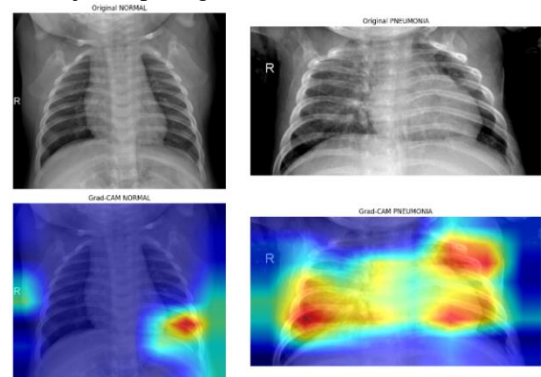


Gambar 13. Performa model MobileNetV2 pada tiap percobaan

Gambar 13 menunjukkan grafik performa model MobileNetV2, di mana model MobileNetV2 menunjukkan nilai Akurasi, Presisi, Recall, dan F1-Score pada 10 kali percobaan yang dilakukan. Hal ini menunjukkan bahwa model MobileNetV2 memiliki performa yang baik dalam mengklasifikasi citra Pneumonia dan Normal. Terdapat

sedikit fluktuasi pada percobaan ke-1 hingga ke-4, namun perlahan membaik dan menjadi lebih stabil.

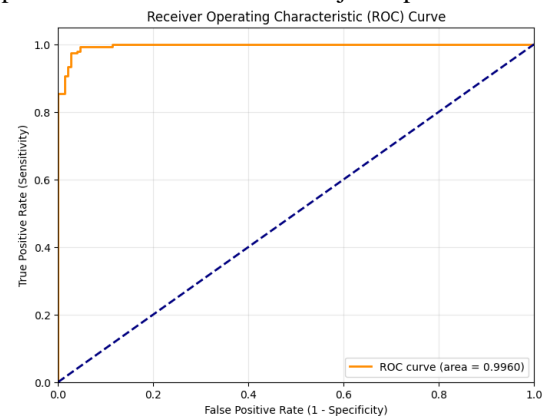
Selain percobaan 10 kali tersebut, dilakukan satu percobaan pengujian tambahan untuk menampilkan Grad-CAM, ROC-AUC dan confusion matrix. Grad-Cam model CNN ditunjukkan pada gambar 14.



Gambar 14. Grad-CAM pada model MobileNetV2

Gambar 14 menunjukkan visualisasi Grad-CAM pada model MobileNetV2. Pada citra normal, Grad-CAM menunjukkan dominasi warna biru dengan sedikit area berwarna kuning dan merah pada bagian bawah paru-paru. Hal ini menunjukkan bahwa model tidak mendeteksi adanya indikasi pneumonia pada area paru-paru. Namun, model masih memberikan heatmap pada area di luar paru sehingga model masih perlu ditingkatkan agar dapat lebih fokus pada area paru-paru.

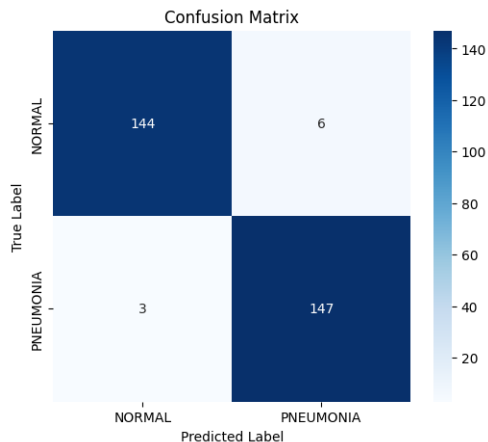
Pada citra pneumonia, Grad-CAM menunjukkan aktivasi tinggi pada area paru-paru kanan dan kiri yang ditandai dengan warna kuning hingga merah. Model memfokuskan perhatian pada area paru yang terdapat indikasi infeksi. Hal ini menunjukkan bahwa model mampu mempelajari fitur penting yang relevan dalam mendeteksi pneumonia. ROC-AUC pada model MobileNetV2 ditunjukkan pada Gambar 15.



Gambar 15. ROC-AUC pada model MobileNetV2

Gambar 15 menunjukkan kurva Receiver Operating Characteristic (ROC) pada model MobileNetV2. Hasil pengujian menunjukkan nilai Area Under Curve (AUC) sebesar 0,9960. Nilai tersebut menunjukkan bahwa model MobileNetV2 memiliki kemampuan klasifikasi yang sangat

baik dalam membedakan kelas pneumonia dan normal. Confusion Matrix pada model MobileNetV2 ditunjukkan pada gambar 16.



Gambar 16. Confusion Matrix pada model MobileNetV2

Gambar 16 menunjukkan hasil Confusion Matrix pada model MobileNetV2. Berdasarkan confusion matrix, model MobileNetV2 memperoleh nilai True Negative sebanyak 144 citra dan True Positive sebanyak 147 citra Nilai False Positive sebanyak 6 citra menunjukkan masih terdapat beberapa citra normal yang diprediksi sebagai pneumonia. Sementara itu, nilai False Negative sebanyak 3 citra menunjukkan masih terdapat beberapa citra pneumonia yang diprediksi sebagai normal.

Secara keseluruhan, model MobileNetV2 memiliki performa klasifikasi yang baik dalam membedakan kelas normal dan pneumonia dengan jumlah kesalahan prediksi yang relatif rendah.

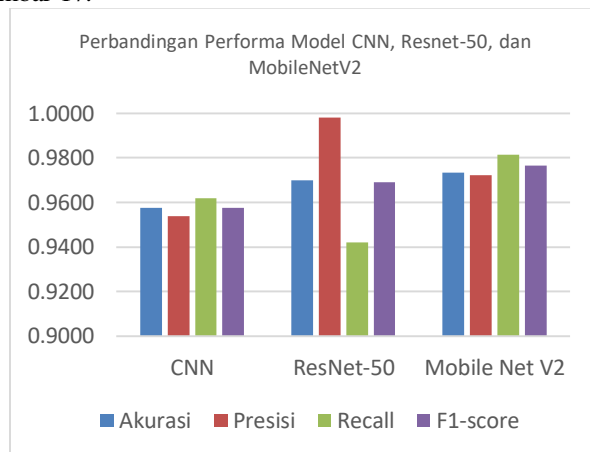
Berdasarkan percobaan yang telah dilakukan sebanyak 10 kali dari model CNN, ResNet-50, dan MobileNetV2, maka rata-rata dari hasil percobaan model CNN, ResNet-50, MobileNetV2 di tunjukan pada Tabel IX.

TABEL IX.
RATA-RATA HASIL PADA MODEL CNN, RESNET-50, MOBILENETV2

Metrik	CNN	ResNet-50	Mobile Net V2
Akurasi	0,9577	0,9699	0,9733
Presisi	0,9539	0,9979	0,9723
Recall	0,9620	0,9420	0,9813
F1-Score	0,9577	0,9690	0,9766
Rata-rata Durasi	7,7 menit	14,6 menit	13 menit

Berdasarkan tabel IX yang menunjukkan rata-rata hasil pada model CNN, ResNet-50 dan MobileNetV2, diperoleh perbedaan pada metrik evaluasi. Pada model CNN menunjukkan bahwa proses pelatihan lebih cepat dengan rata-rata durasi 7,7 menit. Namu performa model CNN pada membedakan citra kelas normal dan pneumonia masih dibawah ResNet-50 dan MobileNetV2. Model ResNet-50 menunjukkan performa yang baik dalam memprediksi False

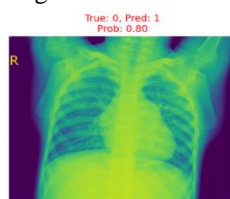
Positive, sehingga prediksi pneumonia yang dihasilkan lebih akurat. Model ResNet-50 memiliki rata-rata durasi training paling lama dibandingkan model CNN dan model MobileNetV2. Hal ini dikarenakan arsitektur jaringan pada ResNet-50 yang dalam dan kompleks. Model MobileNetV2 menunjukkan performa terbaik secara keseluruhan. Model MobileNetV2 mampu memprediksi False Negative dengan baik. Durasi training model MobileNetV2 lebih cepat dibandingkan ResNet-50. Perbandingan performa model CNN, ResNet-50, dan MobileNetV2 ditunjukkan pada gambar 17.



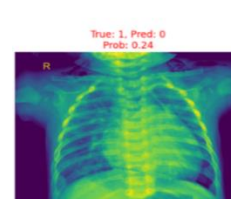
Gambar 17. Perbandingan performa model CNN, ResNet-50, dan MobileNetV2

Gambar 17 menunjukkan perbandingan performa model CNN, ResNet-50 dan MobileNetV2. Grafik menunjukkan performa transfer learning yaitu ResNet-50 dan MobileNetV2 lebih baik dibandingkan model CNN. Grafik MobileNetV2 menunjukkan performa seimbang dibandingkan model lainnya. Model ResNet-50 memiliki nilai rata-rata presisi tertinggi daripada model lainnya. Model CNN memiliki performa yang lebih rendah dibandingkan kedua model transfer learninh, namun menunjukkan performa yang cukup baik.

Adapun contoh gambar prediksi dalam percobaan yang dilakukan. Gambar tersebut ditunjukkan pada gambar 18.a dan gambar 18.b.



Gambar 18.a. Normal diprediksi pneumonia



Gambar 18.b. Pneumonia diprediksi normal

D. Perbandingan Penelitian

Perbandingan Penelitian dilakukan untuk mengetahui performa pada beberapa penelitian sebelumnya yang menggunakan model CNN, ResNet-50 dan MobileNetV2. Tabel perbandingan ditunjukkan pada tabel X.

TABEL X
PERBANDINGAN HASIL PENELITIAN DAN PENELITIAN
SEBELUMNYA

Peneliti	Metode	Dataset	Jumlah Data	Akurasi
Alghozali et al.[28]	CNN	Citra X-ray	771	91%
Fahd et al.[29]	ResNet-50	Skin Lesion	944	76%
Angel[30]	MobileNet V2	Citra X-ray	317	84%
Penelitian ini	CNN	Citra X-ray	2000	95,77%
Penelitian ini	ResNet-50	Citra X-ray	2000	96,99%
Penelitian ini	MobileNet V2	Citra X-ray	2000	97,33%

Tabel X menunjukkan perbandingan penelitian sebelumnya dengan penelitian ini. Penelitian yang dilakukan oleh Alghozali dengan model CNN menggunakan 771 citra x-ray mendapatkan akurasi sebesar 91%. Kemudian Penelitian yang dilakukan oleh Fahd dengan model ResNet-50 menggunakan 944 dataset skin lesion mendapatkan akurasi sebesar 76%. Penelitian yang dilakukan oleh Angel dengan model MobileNetV2 dengan 317 citra x-ray mendapatkan akurasi sebesar 84%. Pada penelitian ini, model CNN memperoleh akurasi sebesar 95,77%, ResNet-50 sebesar 96,99% dan MobileNetV2 sebesar 97,33%. Penggunaan jumlah dataset pada penelitian ini mempengaruhi model dalam mempelajari dan mendeteksi fitur pada citra x-ray dan menghasilkan performa yang lebih baik dibandingkan penelitian sebelumnya.

IV. KESIMPULAN

Penelitian ini berhasil melakukan perbandingan performa model CNN, ResNet-50 dan MobileNetV2 untuk klasifikasi pada citra thorax dalam mendeteksi penyakit pneumonia. Pada setiap percobaan, hasil confusion matrix, akurasi, recall, presisi, dan f1-score mengalami perubahan atau fluktuasi yang menunjukkan adanya pengaruh pada proses training pada setiap model. Pada CNN menghasilkan rata-rata akurasi 95,76%, rata-rata presisi 95,39%, rata-rata recall 96,20%, dan rata-rata f1-score 95,77%. Pada ResNet-50 menunjukkan rata-rata akurasi 96,99%, rata-rata presisi 99,79%, rata-rata recall 94,20%, dan rata-rata f1-score 96,90%. Pada MobileNetV2 menunjukkan rata-rata akurasi 97,33%, rata-rata presisi 97,23%, rata-rata recall 98,13%, dan rata-rata f1-score 97,66%. MobileNetV2 unggul dalam meminimalkan False Negative. Sementara itu, ResNet-50 lebih baik dalam meminimalkan False Positive. Pada penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan beragam, seperti dataset dari rumah sakit untuk meningkatkan generalisasi model.

DAFTAR PUSTAKA

- [1] A. A. Kusmareni, A. Desiani, S. Yahdin, M. Saviera, A. I. Putri, and D. A. Zayanti, "Segmentasi Paru-Paru Pada Citra Thorax Dada Dengan Menggunakan Metode Cnn U-Net," *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 14, no. 2, pp. 2666–2676, 2022, doi: 10.18495/jsi.v14i2.16771.
- [2] D. Kusuma, D. B. Rarasati, U. B. Mulia, and A. Sutera, "Klasifikasi Penyakit Paru-Paru Dengan Citra X-Ray," vol. 9, no. 2, pp. 152–165, 2025.
- [3] A. A. Adiwijaya, Y. S. Hariyani, and S. Hadiyoso, "Peningkatan Citra Untuk Klasifikasi Gangguan Paru-Paru Menggunakan Deep Learning," vol. 10, no. 3, pp. 830–837, 2024.
- [4] F. Ramadhani, H. Syahputra, R. L. Simanjuntak, T. R. Siagian*, U. Nisa, and V. Anggraini, "Klasifikasi Suara Paru Normal Dan Abnormal Berbasis Algoritma CNN (Convolutional Neural Network)," *Jurnal Teknologi Informasi dan Terapan*, vol. 11, no. 1, pp. 15–20, 2024, doi: 10.25047/jtit.v11i1.370.
- [5] A. Milionita, B. Mayor, C. E. Mongi, W. Christo, and D. Weku, "IJIDS Sistem Deteksi Penyakit Pneumonia Menggunakan Algoritma Klasifikasi Convolutional Neural Network dan Support Vector Machine Dengan Citra Chest X-Ray," vol. 01, no. 01, pp. 23–29, 2022, [Online]. Available: <https://ejournal.unsrat.ac.id/v3/index.php/IJIDS/index>
- [6] B. Nugroho and E. Y. Puspaningrum, "Kinerja Metode CNN untuk Klasifikasi Pneumonia dengan Variasi Ukuran Citra Input," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 3, pp. 533–538, 2021, doi: 10.25126/jtiik.2021834515.
- [7] J. Yopento, E. Ernawati, and F. F. Coastera, "Identifikasi Pneumonia Pada Citra X-Ray Paru-Paru Menggunakan Metode Convolutional Neural Network (CNN) Berdasarkan Ekstraksi Fitur Sobel," *Rekursif: Jurnal Informatika*, vol. 10, no. 1, pp. 40–47, 2022, doi: 10.33369/rekursif.v10i1.17247.
- [8] D. N. Rohmah, "Management Kasus Gagal Nafas Pada Penyakit Pneumonia," *Jurnal Berita Ilmu Keperawatan*, vol. 13, no. 1, pp. 22–30, 2020.
- [9] H. N. Ifayatin and R. A. Saputra, "Sistem Deteksi Penyakit Pneumonia Menggunakan Algoritma Faster R-CNN Berbasis Citra Digital Rontgen Dada Detection System of Pneumonia Disease Using Faster R-CNN Algorithm Based on Digital Chest X-ray Images," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 12, no. 4, pp. 645–652, 2024, doi: 10.26418/justin.v12i4.81304.
- [10] F. Hasan, Y. W. Sani, S. Nicholas, and F. T. Anggraeny, "Evaluasi Kualitas Citra Rontgen Dada Menggunakan PSNR dan MSE Setelah Penerapan Teknik Peningkatan Citra," pp. 203–214.
- [11] I. Kurniawati, R. Akbar, Y. Ardina Kusuma, and I. Fahma Kusumawati, "Identifikasi Pneumonia pada Balita melalui Citra X-ray Menggunakan Metode Convolutional Neural Network (CNN)," *Journal of Manufacturing in Industrial Engineering & Technology*, vol. 3, no. 1, pp. 10–23, 2024, doi: 10.30651/mine-tech.v3i1.23391.
- [12] S. Khairunnisak, A. Habibi, S. St, M. J. Afroni, and S. St, "Penerapan Machine Learning Dalam Sistem Klasifikasi Paru-paru Yang Terkena Pneumonia Menggunakan Model Convolutional Neural Network (CNN) Dengan Tensorflow," vol. 19, no. 1, pp. 25–30, 2025.
- [13] M. Adzkia, F. Arland, and A. W. Setiawan, "Deteksi Pneumonia Menggunakan Citra Sinar-X Paru berbasis Residual Network," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 2, pp. 373–380, 2022, doi: 10.25126/jtiik.2022925626.
- [14] R. R. E. Prasetyo and M. Ichwan, "Perbandingan Metode Deep Residual Network 50 dan Deep Residual Network 152 untuk Deteksi Penyakit Pneumonia pada Manusia," *MIND Journal*, vol. 6, no. 2, pp. 168–182, 2021, doi: 10.26760/mindjournal.v6i2.168-182.
- [15] E. Setyaningrum and B. Wirawan, "Implementasi Convolutional Neural Network Berbasis Mobilenetv2 Untuk Klasifikasi

- Pneumonia Menggunakan Citra X-Ray Paru-Paru,” vol. 14, no. 1, pp. 1604–1610.
- [16] M. C. Amri, L. S. Harahap, and A. Rasyid, “Analisis Performa Transfer Learning Menggunakan MobileNetV2 untuk Klasifikasi Citra X-Ray Paru-Paru,” *Jurnal Sains Student Research*, vol. 4, no. 1, pp. 622–630, 2026.
- [17] P. Budi Utomo, M. Faruqziddan, E. Herdika Septa Aulia, and S. Dini Azzahra, “Perbandingan Skenario Balancing Oversampling dan Undersampling dalam Klasifikasi Resiko Kambuh Kanker Tiroid menggunakan Algoritma SVM Linear,” *JAMI: Jurnal Ahli Muda Indonesia*, vol. 5, no. 2, pp. 172–182, 2024, doi: 10.46510/jami.v5i2.320.
- [18] A. H. Nasrullah, “Convolutional Neural Network Untuk Deteksi Penyakit Paru-Paru Menggunakan Citra Radiografi,” vol. 5, no. 2, pp. 37–42, 2021.
- [19] M. B. Sastramandala, R. N. Sari, and N. Novrina, “Pemanfaatan Deep Learning menggunakan Convolutional Neural Network Terhadap Penyakit Tuberkulosis melalui Citra Rontgen,” *Jurnal Minfo Polgan*, vol. 14, no. 1, pp. 631–642, 2025, doi: 10.33395/jmp.v14i1.14818.
- [20] S. Mulyana and I. Muthmainnah, “Analisis Perbandingan Tingkat Akurasi Algoritma Cnn Dan Svm Dalam Klasifikasi Pada Daun Gedi, Daun Pepaya Dan Daun Ubi,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 6157–6162, 2024, doi: 10.36040/jati.v8i4.10144.
- [21] B. M. Sujatmiko, E. Yudaningtyas, and P. M. Raharjo, “Convolution Neural Network Dengan Desain Jaringan Resnet Sebagai Metode Klasifikasi Tumor Kulit Convolution Neural Network Using Resnet Network Design As Skin Tumor,” vol. 11, no. 1, pp. 53–64, 2022.
- [22] A. R. Hermanto, A. Aziz, and S. Sudioanto, “Perbandingan Arsitektur MobileNetV2 dan RestNet50 untuk Klasifikasi Jenis Buah Kurma Comparison of MobileNetV2 and RestNet50 Architectures for Date Fruit Classification by Type,” *Jurnal Sistem dan Teknologi Informasi*, vol. 12, no. 4, pp. 630–637, 2024, doi: 10.26418/justin.v12i4.80358.
- [23] I. S. Lestari, J. Jumadi, and N. Lukman, “Implementasi Convolutional Neural Network Dengan Pre-Trained Model Mobilenetv2 Untuk Deteksi Kolesterol,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 2, pp. 173–183, 2024, doi: 10.36341/rabit.v9i2.4732.
- [24] M. H. Rizky Pratama, M. Akrom, A. P. Santosa, M. R. Rosyid, and L. Mawaddah, “Klasifikasi Otomatis Korosi Menggunakan Convolutional Neural Network dan Transfer Learning dengan Model MobileNetV2,” *Jurnal Algoritma*, vol. 22, no. 1, pp. 138–148, 2025, doi: 10.33364/algoritma/v.22-1.2182.
- [25] C. Nastitie and I. Handayani2, “Perbandingan Metode Mobilenetv2 Dan Efficientnetb0 Untuk Klasifikasi Jenis Pisang,” *INTECOMS: Journal of Information Technology and Computer Science*, vol. 8, no. 6, pp. 1990–1999, 2025, doi: 10.31539/0240cx09.
- [26] A. E. Maxwell, M. S. Bester, and C. A. Ramezan, “Enhancing Reproducibility and Replicability in Remote Sensing Deep Learning Research and Practice,” *Remote Sens. (Basel)*, vol. 14, no. 22, pp. 1–12, 2022, doi: 10.3390/rs14225760.
- [27] F. Renard, S. Guedria, N. De Palma, and N. Vuillerme, “Variability and reproducibility in deep learning for medical image segmentation,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–16, 2020, doi: 10.1038/s41598-020-69920-0.
- [28] M. A. Alghozali, J. R. Triosaputra, and A. Kayan, “Klasifikasi Penyakit Pneumonia Citra Digital X-Ray Menggunakan Metode Convolutional Neural Network dan RGB Equalization,” *Seminar Nasional Teknologi & Sains*, vol. 3, no. 1, pp. 229–236, 2024.
- [29] Isvine Zahroya Jazmine Marzuki Fahd, Abdul Rabi, and Elta Sonalitha, “Perbandingan Akurasi Model ResNet50 dan VGG16 dalam Mengklasifikasi Penyakit Cacar Menggunakan Metode Convolutional Neural Network,” *Uranus : Jurnal Ilmiah Teknik Elektro, Sains dan Informatika*, vol. 3, no. 1, pp. 170–180, 2025, doi: 10.61132/uranus.v3i1.679.
- [30] Angel, “Klasifikasi Citra Paru-Paru Menggunakan Efficient Netb0 Dan Mobilenetv2,” *Jurnal Komputer dan Informatika*, vol. 18, no. 2, pp. 108–115, 2023.