

# Development of MobileNetV2 for CT-Scan Lung Classification Using Transfer Learning

Daffa Fadhil Rajendra <sup>1\*</sup>, Ajie Kusuma Wardhana <sup>2\*</sup>,

\* Teknik Informatika, Universitas AMIKOM Yogyakarta

[dafa.fadil.411@students.amikom.ac.id](mailto:dafa.fadil.411@students.amikom.ac.id) <sup>1</sup>, [ajiekusuma@amikom.ac.id](mailto:ajiekusuma@amikom.ac.id) <sup>2</sup>

## Article Info

### Article history:

Received 2025-07-17

Revised 2025-08-01

Accepted 2025-08-10

### Keyword:

*Lung Cancer,  
CT Scan,  
MobileNetV2,  
Transfer Learning,  
Deep Learning.*

## ABSTRACT

Lung cancer is one of the leading causes of cancer-related deaths worldwide, making early detection crucial for improving patient survival rates. This study proposes an automated classification approach based on deep learning using the MobileNetV2 architecture to identify three categories of lung CT scan images: normal, benign, and malignant. The dataset used is the augmented IQ-OTH/NCCD Lung Cancer Dataset, consisting of 3,609 images with a resolution of 224×224 pixels. All images underwent preprocessing steps including RGB conversion, pixel rescaling, and normalization. The MobileNetV2 model was modified by adding a GlobalAveragePooling2D layer, a dense layer, and dropout to reduce overfitting risk. Training was conducted for 28 epochs using the optimizer Adam, followed by evaluation using accuracy, precision, recall, and F1-score metrics. The model was tested on unseen data and validated using Stratified 5-Fold Cross Validation. The testing results showed an overall accuracy of 97%, with a perfect recall score (1.00) for the malignant class. The cross-validation yielded an average accuracy of 97.26% with a standard deviation of ±0.66%, indicating consistent model performance. Given its lightweight architecture and high accuracy, MobileNetV2 has the potential to be implemented as a decision support system in medical image analysis.



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

## I. PENDAHULUAN

Kanker paru-paru merupakan penyebab utama kematian akibat kanker di seluruh dunia, dengan 1,8 juta kematian pada tahun 2020 atau sekitar 18% dari total kematian akibat kanker [1]. Di Indonesia, kanker paru menempati urutan kedua terbanyak dengan kurang lebih 34.783 kasus baru pada tahun 2020 [2]. Deteksi sejak dini sangat penting untuk meningkatkan peluang hidup pasien, meskipun dalam praktiknya masih menghadapi banyak hambatan [3]. CT Scan banyak dimanfaatkan karena mampu menangkap detail jaringan paru dengan cukup baik, sehingga efektif digunakan dalam deteksi kanker paru [4]. Namun, sebelum diterapkannya teknologi kecerdasan buatan, proses interpretasi citra CT sepenuhnya mengandalkan keahlian radiolog, yang dalam beberapa kasus dapat menimbulkan ketidakakuratan hasil, risiko salah diagnosis, serta keterlambatan dalam penanganan akibat subjektivitas dan kelelahan dalam analisis manual. Kondisi ini mendorong

pemanfaatan metode komputasional seperti deep learning untuk mendukung deteksi kanker secara lebih konsisten dan efisien. Dalam beberapa tahun terakhir, metode deep learning seperti CNN mulai digunakan untuk mengotomatisasi proses klasifikasi citra CT paru [5].

Meskipun CNN menunjukkan performa tinggi dalam pengenalan citra medis, sebagian besar penelitian terdahulu masih terbatas pada dua klasifikasi saja yaitu kanker dan normal, sehingga kurang mencerminkan kompleksitas kasus di dunia nyata [6]. Selain itu, implementasi CNN seringkali membutuhkan sumber daya komputasi besar dan waktu pelatihan yang lama, sehingga menjadi kendala di banyak fasilitas kesehatan dengan keterbatasan infrastruktur [7].

Beberapa studi masih terus mengembangkan dan mengeksplorasi arsitektur ringan seperti MobileNetV2 yang dapat berjalan efisien di perangkat dengan spesifikasi terbatas [8]. Sebaliknya, klasifikasi multi-kelas (normal, benign, dan malignant) dalam konteks citra medis masih menjadi tantangan tersendiri karena kompleksitas visual antar kelas,

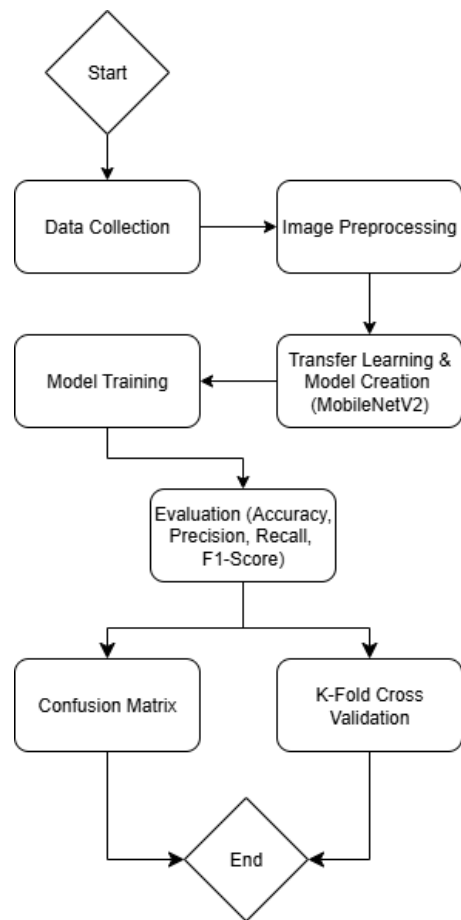
sehingga memerlukan pendekatan yang akurat dan efisien untuk mendukung proses diagnosis [9]. Sejumlah studi menunjukkan bahwa MobileNetV2 memiliki performa kompetitif dibandingkan model dengan arsitektur yang lebih kompleks. Penelitian sebelumnya [10] melakukan perbandingan antara MobileNetV2 dan beberapa model lain seperti VGG16, ResNet50, serta InceptionV3 dalam klasifikasi kanker paru berdasarkan citra CT. MobileNetV2 mendapat akurasi 94,5% dengan keunggulan pada efisiensi waktu inferensi dan kebutuhan komputasi yang lebih rendah. Sementara itu, penelitian lain [11] menunjukkan bahwa MobileNetV2 memperoleh akurasi sebesar 92,3% pada klasifikasi tiga kelas, dan tetap memiliki performa yang lebih unggul terhadap model yang lebih kompleks seperti ResNet152V2 maupun InceptionResNetV2. Dengan demikian, arsitektur MobileNetV2 terbukti mampu bersaing dengan model-model arsitektur lain yang lebih kompleks. Penelitian ini difokuskan pada pengembangan model berbasis MobileNetV2 yang tidak hanya mempertahankan tingkat akurasi yang tinggi, tetapi juga tetap konsisten dengan tujuan awalnya, yaitu menghasilkan arsitektur yang ringan dan efisien. Keunggulan ini menjadikan MobileNetV2 sebagai solusi yang potensial untuk diterapkan di fasilitas kesehatan dengan keterbatasan infrastruktur komputasi.

Di Indonesia sendiri, studi yang menerapkan pendekatan transfer learning untuk klasifikasi kanker paru masih terus dikembangkan dalam berbagai konteks, terutama yang berkaitan dengan efisiensi pemodelan dan keterbatasan sumber daya. Hal ini menunjukkan bahwa topik ini tetap terbuka untuk dieksplorasi lebih lanjut guna memberikan kontribusi yang relevan terhadap kebutuhan dan kondisi [12].

Penelitian ini bertujuan untuk mengatasi permasalahan tersebut dengan menerapkan transfer learning menggunakan arsitektur MobileNetV2 pada citra CT-Scan untuk klasifikasi multi-kelas [13]. Dengan transfer learning, model pelatihan yang telah diuji pada kumpulan data besar dapat digunakan dengan efektif dan tetap akurat [14]. MobileNetV2 dipilih karena dikenal ringan, cepat, dan sesuai untuk lingkungan medis dengan keterbatasan perangkat keras [15]. Diharapkan metode ini mampu menghasilkan sistem klasifikasi kanker paru yang dapat membantu pengambilan keputusan medis, serupa dengan sistem Computer-Aided Diagnosis (CAD) berbasis deep learning yang sudah digunakan dalam mendeteksi nodul paru [16] [17]. Hasil penelitian ini juga diharapkan dapat berkontribusi dalam pengembangan model Deep Learning untuk mendeteksi kanker paru-paru dengan lebih akurat dan lebih efisien.

## II. METODE

Penelitian ini dilakukan melalui beberapa tahapan yang dapat dilihat pada Gambar 1. Metode yang digunakan terdiri dari proses pengumpulan data, preprocessing gambar, pembuatan model klasifikasi menggunakan arsitektur MobileNetV2, pelatihan model, dan evaluasi performa menggunakan metrik klasifikasi multi-kelas.



Gambar 1. Alur Metodologi

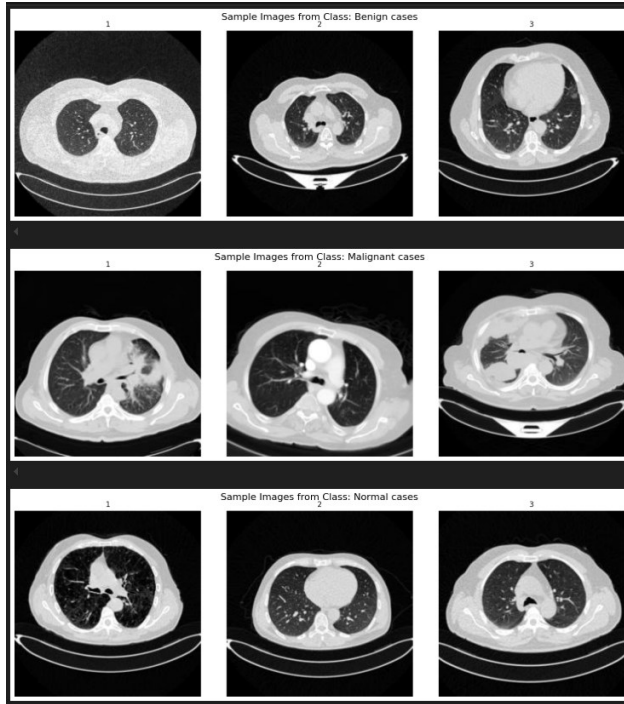
### A. Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini adalah IQ-OTH/NCCD Lung Cancer Dataset (Augmented) yang tersedia dalam platform Kaggle. Dataset ini terdiri dari 3.609 citra CT-Scan yang telah melalui proses augmentasi dan terbagi menjadi tiga kelas, yaitu normal, benign, dan malignant. Setiap kelas memiliki jumlah yang hampir seimbang, sebagaimana ditunjukkan pada Tabel I. Dataset ini dipilih karena representatif terhadap berbagai kondisi jaringan paru-paru yang umum dijumpai dalam diagnosis kanker.

TABEL I  
DISTRIBUSI DATASET

Kelas	Jumlah
Normal	1203
Benign	1201
Malignant	1205
Total	3609

Untuk memberikan gambaran visual terhadap data yang digunakan, Gambar 2 menyajikan contoh citra dari masing-masing kelas dalam dataset.



Gambar 2. Contoh Citra Dataset

### B. Preprocessing Gambar

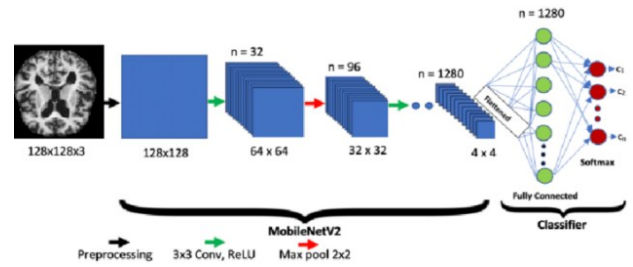
Setiap citra dikonversi ke format RGB dan di-resize menjadi  $224 \times 224$  piksel agar sesuai dengan input model MobileNetV2. Nilai piksel dinormalisasi ke dalam rentang 0 hingga 1 dengan membagi setiap piksel terhadap 255. Label kelas kemudian dikodekan ke dalam format one-hot encoding untuk mendukung klasifikasi multi-kelas. Proses ini bertujuan untuk memastikan keseragaman dimensi dan intensitas piksel dalam seluruh dataset. Dataset yang telah diproses dibagi menjadi tiga bagian: data pelatihan (70%), data validasi (15%), dan data pengujian (15%). Pembagian dilakukan menggunakan metode stratified split agar distribusi kelas tetap proporsional pada masing-masing subset data. Hal ini penting untuk menjaga keseimbangan representasi kelas, khususnya dalam evaluasi performa model terhadap data yang belum pernah dilihat sebelumnya.

### C. Pembangunan Model

Model klasifikasi citra dalam penelitian ini dibangun dengan pendekatan transfer learning, yaitu dengan memanfaatkan arsitektur MobileNetV2 yang sebelumnya telah dilatih pada dataset berskala besar, ImageNet. Proses ini bertujuan untuk memanfaatkan representasi fitur umum yang telah dipelajari oleh model dari dataset lain, sehingga dapat meningkatkan performa dan efisiensi pelatihan saat diaplikasikan pada dataset baru dengan ukuran terbatas, seperti dalam penelitian ini.

Pada tahap awal, bagian atas (top layer) dari MobileNetV2 dihapus agar arsitektur dapat disesuaikan dengan kebutuhan klasifikasi tiga kelas (benign, malignant, normal). Kemudian, dilakukan fine-tuning dengan menambahkan beberapa lapisan

kustom untuk menghasilkan output prediksi sesuai dengan kelas target. Gambar 3 menunjukkan struktur arsitektur model yang digunakan dalam referensi untuk penelitian ini.



Gambar 3. Arsitektur Model [18]

Pada pseudocode Tabel II merupakan arsitektur dan konfigurasi model yang digunakan pada penelitian ini. Lapisan tersebut meliputi GlobalAveragePooling2D yang berfungsi untuk meratakan hasil ekstraksi fitur, kemudian diikuti oleh Dense layer sebanyak 128 unit dengan fungsi aktivasi ReLU untuk menambahkan kompleksitas non-linearitas. Untuk mengurangi risiko overfitting, diterapkan Dropout layer sebesar 0,5, dan pada bagian akhir, digunakan Dense layer dengan 3 unit dan fungsi aktivasi softmax sebagai layer klasifikasi untuk menghasilkan probabilitas dari masing-masing kelas. Model kemudian dikompilasi menggunakan fungsi loss categorical crossentropy karena tugas klasifikasi bersifat multi-kelas, dan dioptimasi menggunakan algoritma Adam dengan learning rate sebesar 0,001.

TABEL II  
PSEUDOCODE

Step	Description
Input:	Preprocessed image data $X$ with shape (IMG_SIZE, IMG_SIZE, 3) One-hot encoded label data $y$ with shape (N, C), where C is number of classes
Output:	Compiled deep learning model using MobileNetV2 as feature extractor
Load pretrained MobileNetV2 base model :	Set include_top = False to remove original classification head Set input_shape = (IMG_SIZE, IMG_SIZE, 3) Set weights = 'imagenet' to use pretrained weights
Freeze all layers in base model	base_mobilenet.trainable ← False
Define number of output classes	num_classes ← number of columns in $y$
Construct new classification head:	$x \leftarrow$ Output of base_mobilenet $x \leftarrow$ GlobalAveragePooling2D( $x$ ) $x \leftarrow$ Dense(128 units, activation = ReLU)( $x$ )

	$x \leftarrow \text{Dropout}(\text{rate} = 0.5)(x)$ $\text{predictions} \leftarrow \text{Dense}(\text{num\_classes}, \text{activation} = \text{softmax})(x)$
Assemble final model:	$\text{model\_mobilenet} \leftarrow \text{Model}(\text{inputs} = \text{base\_mobilenet.input}, \text{outputs} = \text{predictions})$
Compile the model:	$\text{optimizer} \leftarrow \text{Adam}$ $\text{loss} \leftarrow \text{categorical\_crossentropy}$ $\text{metrics} \leftarrow ['\text{accuracy}']$ $\text{model\_mobilenet.compile}(\text{optimizer}, \text{loss}, \text{metrics})$

Fungsi aktivasi ReLU yang digunakan pada lapisan Dense didefinisikan dalam persamaan :

$$f(x) = \max(0, x)$$

Keterangan :

$f(x)$  = nilai output dari fungsi aktivasi ReLU

$x$  = input dari neuron (hasil perkalian bobot dan input ditambah bias)

$\max(0, x)$  = memilih nilai maksimum antara 0 dan nilai input  $x$

Pendekatan transfer learning ini memungkinkan proses pelatihan yang lebih cepat, ringan, dan tetap akurat, terutama ketika diterapkan pada lingkungan dengan keterbatasan sumber daya komputasi seperti sistem pendukung keputusan medis berbasis CT-Scan.

#### D. Pelatihan Model

Model dilatih dengan ukuran batch sebesar 32 selama maksimum 50 epoch. Mekanisme early stopping dengan nilai patience sebesar 5 diterapkan untuk menghentikan proses pelatihan apabila tidak terjadi peningkatan akurasi pada data validasi selama lima epoch berturut-turut. Selama pelatihan, data validasi digunakan untuk memantau kemampuan generalisasi model terhadap data yang tidak terlihat. Proses pelatihan dilakukan dalam lingkungan komputasi GPU guna meningkatkan efisiensi waktu dan performa pemrosesan. Arsitektur model terdiri atas seluruh layer MobileNetV2 tanpa lapisan atas (top layer), yang kemudian ditambahkan tiga lapisan kustom berupa GlobalAveragePooling2D, Dense dengan 128 unit dan fungsi aktivasi ReLU, Dropout sebesar 0,5, serta Dense akhir dengan tiga unit beraktivasi softmax untuk klasifikasi multi-kelas. Model dioptimasi menggunakan algoritma Adam dengan learning rate sebesar 0,001 dan dievaluasi menggunakan metrik akurasi.

#### E. Evaluasi Model

Evaluasi performa dilakukan terhadap data uji dengan menggunakan beberapa metrik umum dalam klasifikasi, yaitu akurasi, precision, recall, dan F1-score. Akurasi menunjukkan proporsi prediksi yang benar terhadap seluruh data, dihitung dengan rumus:

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

Precision mengukur ketepatan prediksi positif model dan dirumuskan sebagai:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall menunjukkan kemampuan model dalam menemukan seluruh data positif yang benar, dengan rumus:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score merupakan rata-rata harmonik dari precision dan recall, berguna saat diperlukan keseimbangan antara keduanya:

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Selain matrik tersebut, evaluasi juga dilengkapi dengan visualisasi confusion matrix untuk melihat secara langsung jumlah klasifikasi benar dan salah pada masing-masing kelas.

#### F. Cross Validation

Evaluasi kinerja model dalam penelitian ini dilakukan menggunakan metode 5-Fold Stratified Cross Validation, yang bertujuan untuk menghasilkan estimasi performa model yang stabil dan bebas dari bias akibat pembagian data yang tidak merata. Teknik ini dipilih karena mampu mempertahankan proporsi kelas yang seimbang dalam setiap fold, terutama penting dalam kasus klasifikasi multi-kelas.

Metode ini bekerja dengan membagi dataset secara acak ke dalam lima subset berukuran kurang lebih sama. Dalam setiap iterasi ke- $i$ , satu subset digunakan sebagai data uji, sedangkan empat subset sisanya digunakan untuk melatih model. Proses ini diulang sebanyak lima kali hingga seluruh data bergilir sebagai data uji.

Evaluasi dilakukan pada setiap iterasi dengan mencatat metrik performa berupa akurasi, precision, recall, dan F1-score. Hasil evaluasi dari setiap fold kemudian dirata-ratakan untuk memperoleh estimasi kinerja keseluruhan model. Rata-rata dari metrik tersebut dihitung dengan rumus:

$$\bar{M} = \frac{1}{k} \sum_{i=1}^k M_i$$

Untuk mengetahui sejauh mana hasil antar fold bervariasi, dilakukan juga perhitungan standar deviasi dengan rumus:

$$\sigma = \sqrt{\frac{1}{k} \sum_{i=1}^k (M_i - \bar{M})^2}$$

keterangan:

$\sigma$  = standar deviasi performa

$M_i$  = nilai metrik pada fold ke- $i$

$\bar{M}$  = rata-rata nilai metrik dari seluruh fold

$k$  = jumlah fold

Dengan menggunakan pendekatan ini, model dapat dievaluasi secara lebih objektif dan menyeluruh terhadap keseluruhan dataset. Hasil dari cross-validation juga memberikan gambaran sejauh mana model konsisten dalam mengenali pola dari data medis yang kompleks, serta meminimalkan risiko overfitting pada subset tertentu.

### III. HASIL DAN PEMBAHASAN

#### A. Image Preprocessing

Tahap awal dalam penelitian ini adalah proses pra-pemrosesan citra, yang bertujuan untuk menyiapkan data agar sesuai dengan kebutuhan arsitektur model. Seluruh gambar CT-Scan paru-paru dikonversi ke dalam format RGB dan diubah ukurannya menjadi  $224 \times 224$  piksel agar kompatibel dengan dimensi input MobileNetV2. Selanjutnya, dilakukan normalisasi nilai piksel ke dalam rentang 0 hingga 1 untuk mempercepat proses konvergensi selama pelatihan dan meningkatkan stabilitas optimisasi. Tahapan ini penting untuk mengurangi variabilitas yang tidak relevan antar citra dan menyamakan distribusi nilai intensitas untuk mendukung proses pembelajaran model secara konsisten.

Setelah proses preprocessing, dataset yang terdiri dari 3.609 gambar dibagi ke dalam tiga subset, yaitu 70% untuk pelatihan, 15% untuk validasi, dan 15% untuk pengujian. Strategi pembagian ini dilakukan secara acak dengan tetap mempertahankan proporsi kelas yang seimbang dalam setiap subset. Distribusi yang seimbang pada kelas benign, malignant, dan normal sangat penting untuk mencegah bias prediksi terhadap kelas mayoritas. Ketidakseimbangan data dapat menyebabkan model cenderung mengklasifikasikan ke kelas yang lebih dominan, sehingga menurunkan akurasi pada kelas minoritas. Dengan menjaga keseimbangan ini sejak tahap awal, model memiliki peluang yang lebih baik dalam mempelajari pola dari semua kelas secara proporsional.

#### B. Pembangunan Model

Arsitektur yang digunakan dalam penelitian ini adalah MobileNetV2, yang dikenal sebagai salah satu model deep learning ringan dan efisien dari segi komputasi, terutama untuk aplikasi pada perangkat dengan sumber daya terbatas. Model ini diimplementasikan sebagai feature extractor dengan memanfaatkan bobot pra-latih (pretrained weights) dari dataset ImageNet. Untuk keperluan klasifikasi khusus dalam penelitian ini, bagian klasifikasi atas (top layer) dari MobileNetV2 dihilangkan (`include_top=False`) agar dapat disesuaikan dengan struktur lapisan klasifikasi baru.

Modifikasi dilakukan dengan menambahkan beberapa lapisan tambahan pada bagian akhir model. Pertama, diterapkan lapisan GlobalAveragePooling2D untuk mereduksi dimensi hasil ekstraksi fitur tanpa kehilangan informasi spasial penting. Selanjutnya, ditambahkan lapisan

Dense dengan 128 unit neuron dan fungsi aktivasi ReLU untuk menangkap representasi non-linier dari fitur yang telah diekstraksi. Untuk mengurangi risiko overfitting, diterapkan lapisan Dropout dengan rasio 0,5, yang berfungsi untuk menonaktifkan sebagian neuron secara acak selama pelatihan.

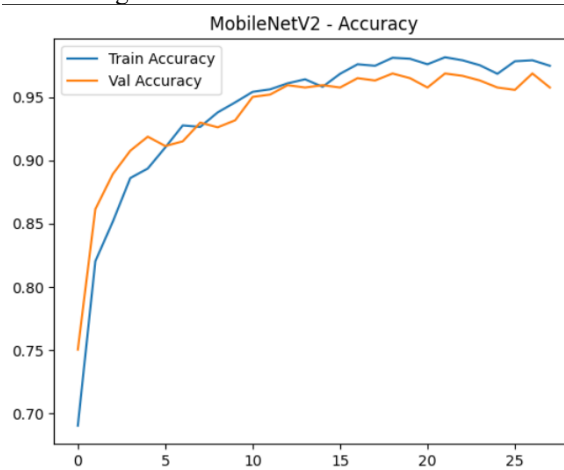
Sebagai lapisan output, digunakan Dense layer dengan tiga unit neuron yang masing-masing merepresentasikan kelas normal, benign, dan malignant, dan menggunakan fungsi aktivasi softmax untuk menghasilkan prediksi probabilistik multi-kelas. Pemilihan MobileNetV2 sebagai arsitektur utama didasarkan pada kombinasi antara efisiensi komputasi dan akurasi yang kompetitif, menjadikannya cocok untuk diterapkan pada sistem pendukung keputusan medis berbasis citra dengan sumber daya komputasi terbatas.

#### C. Pelatihan Model

Proses pelatihan model dilakukan menggunakan algoritma optimisasi Adam, yang dikenal efisien dalam memperbarui bobot jaringan saraf dalam skala besar. Nilai learning rate ditetapkan sebesar 0,001, sedangkan fungsi loss yang digunakan adalah categorical crossentropy, sesuai untuk permasalahan klasifikasi multi-kelas seperti dalam penelitian ini. Model dilatih selama 28 epoch dengan ukuran batch sebesar 32, yang dipilih untuk menyeimbangkan antara efisiensi proses komputasi dan kestabilan pembelajaran.

Untuk mencegah terjadinya overfitting serta menghemat sumber daya pelatihan, diterapkan teknik early stopping dengan patience selama 5 epoch, yang akan menghentikan pelatihan secara otomatis apabila tidak terjadi perbaikan performa pada data validasi dalam lima epoch berturut-turut.

Selama proses pelatihan, model menunjukkan performa yang stabil. Hal ini terlihat dari grafik akurasi pada data pelatihan dan validasi yang meningkat secara konsisten, sebagaimana ditampilkan pada Gambar 4. Tidak ada perbedaan yang signifikan dari grafik training dan validasi sehingga model dapat dinilai tidak mengalami overfitting atau underfitting.

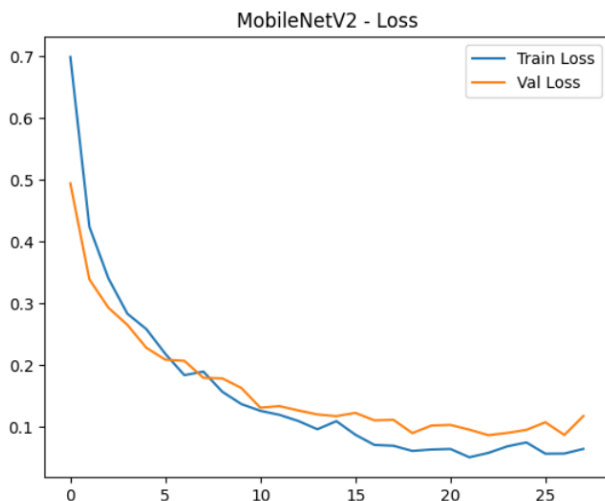


Gambar 4. Train Val Akurasi

Sementara itu, grafik nilai loss pada Gambar 5 menunjukkan penurunan bertahap seiring bertambahnya



epoch, yang mengindikasikan bahwa proses pembelajaran berlangsung secara optimal tanpa gejala overfitting yang signifikan. Pola hasil tersebut menunjukkan bahwa model mampu menyesuaikan bobot secara efektif terhadap data yang tersedia.



Gambar 5. Train Val Loss

#### D. Evaluasi Model

Setelah proses pelatihan selesai, model dievaluasi menggunakan data uji yang belum pernah dilibatkan dalam proses pelatihan maupun validasi. Hasil evaluasi menunjukkan bahwa arsitektur MobileNetV2 mampu mencapai akurasi keseluruhan sebesar 97%, mencerminkan performa klasifikasi yang sangat baik dan stabil. Penilaian dilakukan menggunakan metrik precision, recall, dan F1-score untuk masing-masing kelas guna mengevaluasi kemampuan model dalam mengenali karakteristik visual dari citra CT-Scan.

Rincian hasil pengujian ditampilkan pada Tabel III, yang menunjukkan bahwa model menghasilkan performa yang tinggi dan seimbang di semua kelas. Untuk kelas benign, precision, recall, dan F1-score masing-masing sebesar 0,96, menunjukkan kemampuan model dalam mendeteksi kasus jinak dengan akurasi tinggi. Sementara pada kelas normal, precision mencapai 0,97, recall 0,95, dan F1-score sebesar 0,96, menandakan efektivitas model dalam membedakan citra tanpa indikasi kelainan.

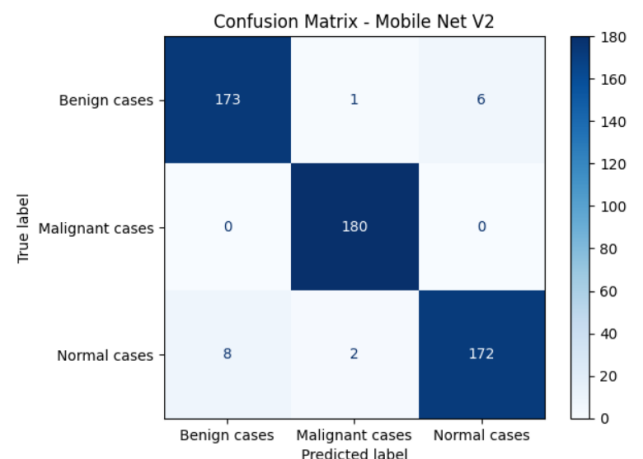
Pencapaian paling menonjol terdapat pada kelas malignant, di mana recall mencapai nilai sempurna yaitu 1,00, artinya seluruh citra dengan indikasi kanker ganas berhasil diklasifikasikan dengan benar. Precision pada kelas ini tercatat sebesar 0,98, dan F1-score sebesar 0,99. Tingginya nilai recall pada kelas malignant sangat penting dalam konteks deteksi penyakit, karena kesalahan dalam mengidentifikasi kategori ini dapat berdampak pada keterlambatan penanganan. Dengan demikian, performa ini memperlihatkan kemampuan model dalam memberikan

prediksi yang akurat pada kondisi yang paling kritis untuk dikenali.

TABEL III  
MOBILENETV2 TESTING RESULT

Class	Precision	Recall	F1-score	Accuracy
Benign	0.96	0.96	0.96	0.97
Malignant	0.98	1.00	0.99	0.97
Normal	0.97	0.95	0.96	0.97

Visualisasi hasil klasifikasi pada Gambar 6 menunjukkan distribusi prediksi model terhadap data uji. Hampir seluruh citra berhasil diklasifikasikan sesuai dengan label aslinya. Pada kelas malignant, seluruh citra terklasifikasi dengan benar, tanpa kesalahan. Kesalahan kecil hanya terjadi pada kelas benign dan normal, dalam jumlah yang tidak signifikan. Temuan ini memperkuat hasil evaluasi sebelumnya, menunjukkan bahwa model tidak hanya akurat, tetapi juga mampu mengenali pola dari citra paru secara konsisten.



Gambar 6. Confusion Matrix

Visualisasi confusion matrix pada Gambar 6 menunjukkan bahwa model MobileNetV2 berhasil mengklasifikasikan sebagian besar citra CT-Scan secara akurat pada masing-masing kelas. Untuk kelas benign, dari total 180 citra, sebanyak 173 citra berhasil diklasifikasikan dengan benar, sementara hanya 1 citra salah diklasifikasikan sebagai malignant dan 6 lainnya sebagai normal. Pada kelas malignant, seluruh 180 citra diklasifikasikan dengan tepat tanpa ada kesalahan, yang menunjukkan performa sempurna dalam mendeteksi jenis kanker ganas. Sementara itu, pada kelas normal, dari 182 citra, sebanyak 172 citra diklasifikasikan dengan benar, dengan 8 citra salah diklasifikasikan sebagai benign dan 2 citra lainnya sebagai malignant.

Distribusi hasil prediksi yang ditampilkan pada confusion matrix tersebut memperkuat metrik evaluasi dalam Tabel II. Kinerja tertinggi terlihat pada kelas malignant, dengan nilai recall sebesar 1,00, menandakan tidak ada kasus yang terlewat oleh model. Hasil ini menunjukkan bahwa model tidak hanya

memberikan skor evaluasi tinggi, tetapi juga memiliki konsistensi dalam mengenali pola citra secara akurat. Selain itu, kemampuan model dalam mengklasifikasikan seluruh citra pada kelas malignant secara tepat juga menunjukkan sensitivitas yang tinggi terhadap fitur visual yang kompleks, termasuk kemungkinan keberadaan lesi berukuran kecil. Lesi kecil sering kali sulit dikenali karena tampilan yang samar atau minim perbedaan kontras dibanding jaringan sehat. Evaluasi ini mengindikasikan bahwa MobileNetV2 mampu menangkap pola-pola penting meskipun ukurannya kecil, yang merupakan aspek penting dalam sistem deteksi otomatis berbasis citra.

#### E. Cross Validation

Untuk memastikan stabilitas dan reliabilitas performa model MobileNetV2 dalam klasifikasi multi-kelas kanker paru, dilakukan evaluasi menggunakan metode Stratified 5-Fold Cross Validation. Teknik ini membagi dataset menjadi lima subset dengan proporsi kelas yang seimbang, di mana pada setiap iterasi, satu subset digunakan sebagai data uji dan empat sisanya sebagai data latih. Hal ini memastikan bahwa seluruh data mendapatkan kesempatan untuk berkontribusi pada proses pelatihan dan pengujian secara merata. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score pada setiap fold. Hasil evaluasi menunjukkan bahwa model memiliki performa yang tinggi dan stabil di seluruh fold, seperti ditampilkan pada Tabel IV berikut:

TABEL IV  
5-FOLD CROSS VALIDATION RESULT

Fold	Accuracy	Precision	Recall	F1-Score
Fold-1	96,95	96,96	96,95	96,95
Fold-2	97,92	97,93	97,93	97,92
Fold-3	98,06	98,08	98,07	98,06
Fold-4	96,26	96,33	96,27	96,27
Fold-5	97,09	97,10	97,09	97,09
Rata-Rata	97,26	97,28	97,26	97,26
Standar Deviasi	± 0.66	± 0.65	± 0.66	± 0.66

Hasil pada Tabel III menunjukkan bahwa model menghasilkan rata-rata akurasi sebesar 97,26%, precision 97,28%, recall 97,26%, dan F1-score 97,26%, dengan standar deviasi di bawah 1% untuk semua metrik. Ini menandakan bahwa model tidak hanya unggul secara akurasi, tetapi juga stabil dalam mengenali pola dari berbagai distribusi data.

Stabilitas antar fold yang tinggi ini membuktikan bahwa arsitektur MobileNetV2 memiliki kemampuan generalisasi yang baik, sehingga dapat diterapkan untuk klasifikasi citra medis tanpa tergantung pada distribusi data tertentu. Validasi silang ini memberikan bukti tambahan bahwa model tidak mengalami overfitting dan dapat bekerja secara konsisten serta mencapai generalisasi.

#### IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan arsitektur MobileNetV2 dalam tugas klasifikasi multi-kelas kanker paru (normal, benign, malignant) berbasis citra CT-Scan. Melalui tahapan preprocessing yang cermat, pembagian data yang proporsional, serta pelatihan model dengan konfigurasi yang optimal, diperoleh model yang tidak hanya akurat tetapi juga stabil. Evaluasi model terhadap data uji menunjukkan akurasi keseluruhan sebesar 97%, dengan nilai precision, recall, dan F1-score yang tinggi dan seimbang di semua kelas. Kinerja tertinggi dicapai pada kelas malignant, dengan recall mencapai 100%, yang menandakan seluruh citra dengan indikasi kanker ganas berhasil diklasifikasikan dengan benar sebuah pencapaian penting dalam konteks deteksi penyakit serius. Visualisasi confusion matrix menunjukkan bahwa sebagian besar prediksi yang dihasilkan sesuai dengan label asli, serta mengindikasikan bahwa model mampu mengenali pola-pola penting meskipun berada pada citra dengan lesi berukuran kecil. Hal ini menandakan sensitivitas model yang baik terhadap fitur visual yang sulit dikenali.

Selanjutnya, validasi dengan metode Stratified 5-Fold Cross Validation menghasilkan akurasi rata-rata sebesar 97,26% dengan standar deviasi  $\pm 0,66$ , memperkuat bukti bahwa model memiliki performa yang konsisten dan generalisasi yang baik pada berbagai subset data. Stabilitas hasil antar fold juga menandakan bahwa model tidak mengalami overfitting dan mampu bekerja secara andal dalam situasi dunia nyata. Dengan demikian, arsitektur MobileNetV2 terbukti efektif dan efisien untuk diterapkan dalam sistem klasifikasi citra medis, khususnya dalam mendeteksi kanker paru secara otomatis, cepat, dan akurat. Penelitian ini memberikan dasar yang kuat bagi pengembangan lebih lanjut ke arah sistem pendukung keputusan berbasis AI di bidang kesehatan, terutama untuk fasilitas medis dengan keterbatasan infrastruktur komputasi.

#### DAFTAR PUSTAKA

- [1] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.
- [2] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview," *Int J Cancer*, vol. 149, no. 4, pp. 778–789, 2021, doi: 10.1002/ijc.33588.
- [3] E. Edelman Saul *et al.*, "The challenges of implementing low-dose computed tomography for lung cancer screening in low- and middle-income countries," *Nat Cancer*, vol. 1, no. 12, pp. 1140–1152, Nov. 2020, doi: 10.1038/s43018-020-00142-z.
- [4] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images," *Diagnostics*, vol. 13, no. 16, p. 2617, Aug. 2023, doi: 10.3390/diagnostics13162617.
- [5] M. Q. Shatnawi, Q. Abuein, and R. Al-Quraan, "Deep learning-based approach to diagnose lung cancer using CT-scan images," *Intell Based Med*, vol. 11, p. 100188, Dec. 2025, doi: 10.1016/j.ibmed.2024.100188.
- [6] W. Wulaningsih, C. Villamaria, A. Akram, J. Benemile, F. Croce, and J. Watkins, "Deep Learning Models for Predicting Malignancy Risk in CT-Detected Pulmonary Nodules: A Systematic Review

- and Meta-analysis,” *Lung*, vol. 202, no. 5, pp. 625–636, 2024, doi: 10.1007/s00408-024-00706-1.
- [7] H. Jiang, S. Tang, W. Liu, and Y. Zhang, “Deep learning for COVID-19 chest CT (computed tomography) image analysis: A lesson from lung cancer,” *Comput Struct Biotechnol J*, vol. 19, pp. 1391–1399, 2021, doi: 10.1016/j.csbj.2021.02.016.
- [8] O. Ukwandu, H. Hindy, and E. Ukwandu, “An evaluation of lightweight deep learning techniques in medical imaging for high precision COVID-19 diagnostics,” *Healthcare Analytics*, vol. 2, p. 100096, Nov. 2022, doi: 10.1016/j.health.2022.100096.
- [9] J. Zheng *et al.*, “Pulmonary nodule risk classification in adenocarcinoma from CT images using deep CNN with scale transfer module,” *IET Image Process*, vol. 14, no. 8, pp. 1481–1489, Jun. 2020, doi: 10.1049/iet-ipr.2019.0248.
- [10] R. Klangbunrueang, P. Pookduang, W. Chansanam, and T. Lunrasri, “AI-Powered Lung Cancer Detection: Assessing VGG16 and CNN Architectures for CT Scan Image Classification,” *Informatics*, vol. 12, no. 1, p. 18, Feb. 2025, doi: 10.3390/informatics12010018.
- [11] M. Kalkan, M. S. Guzel, F. Ekinici, E. Akcapinar Sezer, and T. Asuroglu, “Comparative Analysis of Deep Learning Methods on CT Images for Lung Cancer Specification,” *Cancers (Basel)*, vol. 16, no. 19, p. 3321, Sep. 2024, doi: 10.3390/cancers16193321.
- [12] B. D. S. Isnayni Sugma Rachmatika, Lailil Muflikhah, “Deteksi Mutasi Pada Kanker Paru Melalui Citra CT-Scan Penerapan Model Algoritma Convolutional Neural Networks (CNN) dan Optimizer Adam,” *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. Vol 8 No 9, 2024, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14134/6317>
- [13] Z. Gao, Y. Tian, S.-C. Lin, and J. Lin, “A CT Image Classification Network Framework for Lung Tumors Based on Pre-trained MobileNetV2 Model and Transfer learning, And Its Application and Market Analysis in the Medical field,” Jan. 2025, [Online]. Available: <http://arxiv.org/abs/2501.04996>
- [14] A. Saha, S. M. Ganie, P. K. D. Pramanik, R. K. Yadav, S. Mallik, and Z. Zhao, “VER-Net: a hybrid transfer learning model for lung cancer detection using CT scan images,” *BMC Med Imaging*, vol. 24, no. 1, p. 120, May 2024, doi: 10.1186/s12880-024-01238-z.
- [15] S. Velu, “An efficient, lightweight MobileNetV2-based fine-tuned model for COVID-19 detection using chest X-ray images,” *Mathematical Biosciences and Engineering*, vol. 20, no. 5, pp. 8400–8427, 2023, doi: 10.3934/mbe.2023368.
- [16] A. Gudur, H. Sivaraman, and V. Vimal, “International Journal of Intelligent Systems And Applications In Engineering Deep Learning-Based Detection of Lung Nodules in CT Scans for Cancer Screening,” *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 2023, no. 7s, pp. 20–28, 2023, [Online]. Available: [www.ijisae.org](http://www.ijisae.org)
- [17] M. Aftab *et al.*, “AI in Oncology: Transforming Cancer Detection through Machine Learning and Deep Learning Applications,” Jan. 2025, [Online]. Available: <http://arxiv.org/abs/2501.15489>
- [18] M. Aparna and B. S. Rao, “A novel automated deep learning approach for Alzheimer’s disease classification,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 1, p. 451, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp451-458.