BAB IV
HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Implementasi dan Lingkungan Eksperimen

Bab ini menguraikan secara komprehensif seluruh hasil eksperimen yang dilaksanakan untuk memvalidasi kinerja komparatif arsitektur ResNet50 dan MobileNetV2 dalam tugas klasifikasi kelayakan buah mangga. Seluruh tahapan implementasi, mulai dari persiapan data, proses pelatihan (training), hingga evaluasi akhir model, dijalankan secara konsisten pada platform berbasis cloud computing dengan spesifikasi perangkat keras dan perangkat lunak yang terdefinisi dengan baik. Penetapan lingkungan eksperimen yang terstandarisasi ini merupakan prasyarat utama untuk menjamin validitas teknis dan reproduktibilitas (reproducibility) dari setiap hasil yang dilaporkan.

Spesifikasi lengkap lingkungan implementasi yang digunakan dalam penelitian ini disajikan pada Tabel 4.1 berikut.

Tabel 4.1 Spesifikasi Lingkungan Implementasi

Komponen	Spesifikasi Teknis	Justifikasi Penggunaan
Platform Komputasi	Google Colaboratory Pro	Menyediakan akses akselerasi GPU berbasis cloud dengan integrasi Google Drive untuk penyimpanan dataset dan checkpoint model.
Unit Pemrosesan Grafis (GPU)	NVIDIA L4 (24 GB GDDR6 VRAM)	Mempercepat komputasi operasi konvolusi pada arsitektur ResNet50 dan MobileNetV2 secara signifikan dibandingkan pemrosesan berbasis CPU.
Bahasa Pemrograman	Python 3.x	Bahasa standar dalam ekosistem deep learning dengan dukungan pustaka yang komprehensif untuk pengolahan citra dan pengembangan model.

Komponen	Spesifikasi Teknis	Justifikasi Penggunaan
Framework Deep Learning	TensorFlow / Keras	Menyediakan implementasi arsitektur CNN bawaan (pretrained weights) dan API yang fleksibel untuk kustomisasi lapisan serta proses Fine-Tuning.
Pustaka Pendukung	NumPy, Matplotlib, Pandas	Manipulasi data numerik, visualisasi kurva pelatihan dan metrik evaluasi, serta analisis data tabular hasil eksperimen.
Penyimpanan	Google Drive (Terintegrasi Colab)	Repositori persisten untuk menyimpan dataset citra mangga, bobot model (model weights), dan catatan hasil eksperimen.

Konfigurasi GPU NVIDIA L4 dengan kapasitas VRAM sebesar 24 GB memungkinkan proses pelatihan model dengan ukuran batch yang optimal. Hal ini meminimalkan noise pada estimasi gradien stokastik dan memastikan stabilitas proses backpropagation pada kedua arsitektur yang dikomparasikan. Seluruh proses eksperimen dieksekusi di bawah spesifikasi di atas untuk menjamin konsistensi dan keadilan (fairness) dalam perbandingan kinerja antara ResNet50 dan MobileNetV2.

4.2 Hasil Pengumpulan dan Prapemrosesan Data

Sesuai dengan alur penelitian yang telah dirancang pada Bab III, tahap awal implementasi berfokus pada persiapan data yang menjadi fondasi pembelajaran model. Subbab ini menguraikan hasil dari proses pengumpulan dataset citra mangga serta seluruh transformasi yang diterapkan sebelum data diumpankan ke dalam pipeline pelatihan.

4.2.1 Hasil Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 1.700 citra digital buah mangga yang terbagi ke dalam dua kelas semantik utama, yaitu kelas mangga layak (*fresh/good*) dan kelas mangga tidak layak (*rotten/defective*). Penggunaan dataset berskala ini bertujuan untuk memberikan representasi yang memadai bagi kedua model dalam mempelajari fitur visual pembeda antara kedua kelas tersebut.

Guna menjamin objektivitas evaluasi dan mencegah terjadinya data leakage maupun overfitting, keseluruhan dataset dibagi secara sistematis menggunakan skema Holdout Validation dengan proporsi pembagian 80:10,6:9,4. Rincian distribusi hasil pembagian dataset disajikan pada Tabel 4.2 berikut.

Tabel 4.2 Distribusi Pembagian Dataset Citra Mangga

Kategori Data	Jumlah Citra	Jumlah Batch (Batch Size=32)	Persentase	Keterangan
Data Latih (Training Set)	1.360 citra	43 batch	~80%	Digunakan untuk mengoptimasi bobot model melalui proses forward dan backpropagation.
Data Validasi (Validation Set)	180 citra	6 batch	~10,6%	Digunakan untuk memantau kinerja model pada setiap akhir epoch dan mencegah overfitting.
Data Uji (Testing Set)	160 citra	5 batch	~9,4%	Digunakan murni untuk evaluasi akhir metrik performa setelah pelatihan selesai.

Kategori Data	Jumlah Citra	Jumlah Batch (Batch Size=32)	Persentase	Keterangan
Total	1.700 citra	—	100%	Keseluruhan dataset yang digunakan dalam eksperimen.

Proporsi 80% data latih memberikan volume data yang cukup bagi kedua arsitektur untuk mempelajari pola fitur warna, tekstur, dan bentuk permukaan buah mangga secara mendalam. Sementara itu, pemisahan 10,6% data validasi yang bersifat independen memungkinkan pemantauan dinamika konvergensi model secara real-time pada setiap akhir epoch. Adapun 9,4% data uji yang sepenuhnya terpisah digunakan sebagai instrumen evaluasi akhir yang objektif dan tidak terpengaruh oleh proses optimasi model.

4.2.2 Hasil Prapemrosesan dan Augmentasi Data

Data citra mentah memiliki variasi dimensi piksel yang beragam sesuai dengan kondisi pengambilan gambar. Kondisi ini tidak kompatibel dengan persyaratan input arsitektur ResNet50 maupun MobileNetV2 yang membutuhkan dimensi tensor yang seragam. Oleh karena itu, tahap prapemrosesan dilaksanakan untuk menstandarisasi seluruh citra sebelum dimasukkan ke dalam model.

Tahap pertama adalah penyeragaman dimensi spasial (*resizing*), di mana seluruh citra diresampling menjadi resolusi 224 × 224 piksel. Ukuran ini merupakan dimensi input standar yang ditetapkan oleh arsitektur ResNet50 dan MobileNetV2, serta kompatibel dengan bobot pralatih (*pretrained weights*) pada dataset ImageNet yang digunakan sebagai titik awal *transfer learning*. Setelah penyeragaman dimensi, dilakukan proses normalisasi intensitas piksel menggunakan faktor skala 1./255. Melalui operasi ini, nilai intensitas setiap piksel yang semula berada pada rentang bilangan bulat [0, 255] ditransformasi ke dalam rentang desimal [0, 1].

Normalisasi ini krusial untuk menjaga stabilitas numerik gradien dan mempercepat konvergensi algoritma optimasi Adam selama proses pelatihan.

Contoh representasi visual citra mangga sebelum memasuki tahap augmentasi ditampilkan pada gambar berikut.



Gambar 4.1 Contoh Citra Mangga Asli dari Dataset Setelah Proses Resize

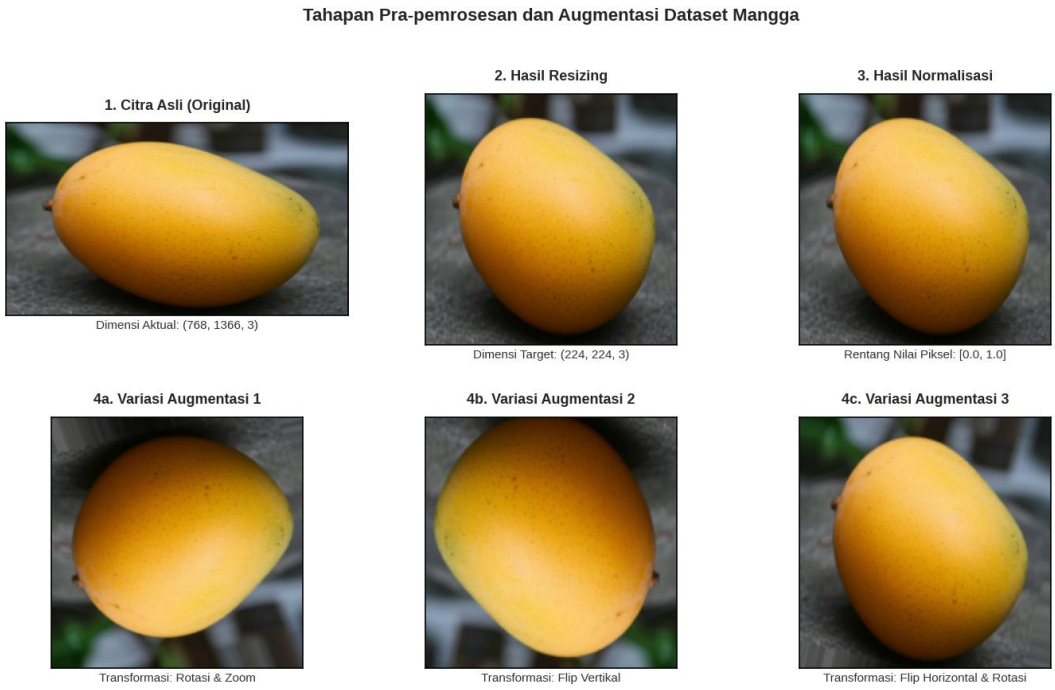
Sebagaimana yang telah dirancang pada Bab III, guna mengatasi keterbatasan jumlah data latih sekaligus meningkatkan kemampuan generalisasi model terhadap variasi kondisi pencitraan di dunia nyata, penelitian ini menerapkan teknik augmentasi data secara On-the-fly. Augmentasi *On-the-fly* berarti transformasi diterapkan secara acak dan berbeda pada setiap citra di setiap iterasi epoch, sehingga model seolah-olah menerima data baru yang bervariasi pada setiap siklus pembelajaran, tanpa perlu menyimpan citra yang sudah diaugmentasi ke dalam penyimpanan secara permanen.

Teknik augmentasi diimplementasikan melalui lapisan sekuensial TensorFlow (TensorFlow Sequential Layer) yang disisipkan langsung ke dalam arsitektur model, terdiri dari empat transformasi berikut:

1. Random Flip (Horizontal & Vertikal): Membalik orientasi citra secara horizontal dan vertikal secara acak. Transformasi ini mensimulasikan variasi sudut pandang pengambilan gambar buah mangga dari berbagai posisi.

2. Random Rotation (±20%): Merotasikan citra secara acak hingga rentang ±20% dari sudut penuh (360°). Augmentasi ini mensimulasikan kondisi buah yang diletakkan pada berbagai kemiringan sudut.
3. Random Zoom (±20%): Melakukan pembesaran (zoom in) atau pengecilan (zoom out) secara acak pada area citra hingga 20%. Hal ini melatih model agar tidak sensitif terhadap variasi skala dan jarak pengambilan gambar buah.
4. Random Brightness (±20%): Mengubah tingkat kecerahan (brightness) citra secara acak dalam rentang ±20%. Augmentasi ini mensimulasikan variasi kondisi pencahayaan yang berbeda-beda saat buah mangga difoto.

Untuk memvalidasi bahwa lapisan augmentasi berfungsi sebagaimana yang dirancang, dilakukan pengujian visualisasi dengan mengambil satu citra sampel dari dataset dan menerapkan lapisan augmentasi tersebut secara berulang sebanyak 3 kali. Hasil visualisasi ini disajikan pada Gambar 4.2 berikut.



Gambar 4.2 Visualisasi beberapa Variasi Hasil Augmentasi Data pada 1 Citra Sampel Mangga

Berdasarkan Gambar 4.2, terlihat bahwa lapisan augmentasi bekerja secara efektif dengan menghasilkan beberapa variasi citra yang secara visual berbeda satu sama lain dari satu sumber citra yang sama. Setiap variasi menampilkan kombinasi

acak dari transformasi pembalikan (*flip*), rotasi, perbesaran, dan perubahan kecerahan yang telah dikonfigurasi. Fakta bahwa tidak ada satu pun dari 3 variasi tersebut yang identik membuktikan bahwa mekanisme randomisasi pada lapisan augmentasi berjalan dengan benar. Penerapan augmentasi data ini secara efektif memperluas keragaman data latih secara artifisial, sehingga kedua model ResNet50 maupun MobileNetV2 memiliki kemampuan generalisasi yang lebih baik terhadap citra mangga dengan kondisi pencahayaan, orientasi, dan jarak pengambilan gambar yang bervariasi pada data uji.

4.3 Implementasi Arsitektur Model

Tahap perancangan model diwujudkan melalui implementasi kode dua arsitektur CNN, yaitu ResNet50 dan MobileNetV2, menggunakan paradigma Transfer Learning berbasis bobot pralatih (*pretrained weights*) dari dataset ImageNet. Pendekatan ini dipilih karena kedua arsitektur telah memiliki kemampuan ekstraksi fitur visual tingkat rendah hingga tingkat tinggi yang telah matang melalui proses pelatihan pada jutaan citra, sehingga proses Fine-Tuning pada dataset mangga berukuran 1.700 citra dapat berjalan secara efektif tanpa risiko underfitting akibat keterbatasan data.

Strategi implementasi yang diterapkan pada kedua model mengikuti pola yang seragam untuk menjamin keadilan komparasi (fair comparison). Pertama, lapisan ekstraktor fitur (feature extractor) bawaan dari arsitektur dasar dimuat dengan bobot ImageNet dan dalam kondisi awal dibekukan (frozen), artinya bobot pada lapisan-lapisan tersebut tidak diperbarui selama proses pelatihan awal. Hal ini bertujuan agar pengetahuan fitur universal yang telah dimiliki model tidak terdegradasi di tahap awal adaptasi.

Kedua, lapisan *classifier* baru ditambahkan di atas ekstraktor fitur beku tersebut. Lapisan ini terdiri dari lapisan *Global Average Pooling* untuk meringkas peta fitur, diikuti oleh lapisan *Dense* dengan fungsi aktivasi ReLU, lapisan Dropout dengan rate 0,5, dan diakhiri oleh lapisan keluaran (*output layer*) dengan 1 neuron dan fungsi aktivasi Sigmoid untuk menghasilkan probabilitas pada tugas klasifikasi biner (layak vs. tidak layak).

Detail spesifikasi arsitektur akhir (Fine-Tuned architecture) dari kedua model disajikan secara komparatif pada Tabel 4.3 berikut.

Tabel 4.3 Spesifikasi Komparatif Arsitektur Model yang Diimplementasikan

Komponen	ResNet50	MobileNetV2
Arsitektur Dasar	ResNet50 (He et al., 2016)	MobileNetV2 (Sandler et al., 2018)
Bobot Awal	ImageNet (Pretrained)	ImageNet (Pretrained)
Dimensi Input	224 × 224 × 3	224 × 224 × 3
Mekanisme Utama	Residual Block (Shortcut Connection)	Inverted Residual + Depthwise Separable Conv.
Jumlah Parameter Dasar	~25,6 Juta	~3,4 Juta
Lapisan Tambahan	Global Avg Pooling → Dense (ReLU) → Dropout (0,5) → Dense (Sigmoid)	Global Avg Pooling → Dense (ReLU) → Dropout (0,5) → Dense (Sigmoid)
Jumlah Kelas Output	2 (Biner: Layak / Tidak Layak)	2 (Biner: Layak / Tidak Layak)
Fungsi Aktivasi Output	Sigmoid	Sigmoid

Perbedaan mendasar antara kedua arsitektur terletak pada filosofi desainnya. ResNet50 dirancang untuk memaksimalkan akurasi ekstraksi fitur melalui jaringan yang sangat dalam dengan mekanisme residual connection, sementara MobileNetV2 dirancang untuk efisiensi komputasi pada perangkat resource-constrained melalui operasi depthwise separable convolution. Komparasi kedua pendekatan ini dalam konteks klasifikasi mangga menjadi inti dari analisis kinerja yang akan diuraikan pada subbab berikutnya.

4.4 Hasil Pelatihan Model (Training Phase)

Setelah kedua arsitektur berhasil dikonstruksi, proses pelatihan (training) dieksekusi menggunakan konfigurasi hyperparameter yang identik pada kedua model untuk menjamin validitas komparasi. Konfigurasi tersebut dirangkum pada Tabel 4.4.

Tabel 4.4 Konfigurasi Hyperparameter Pelatihan

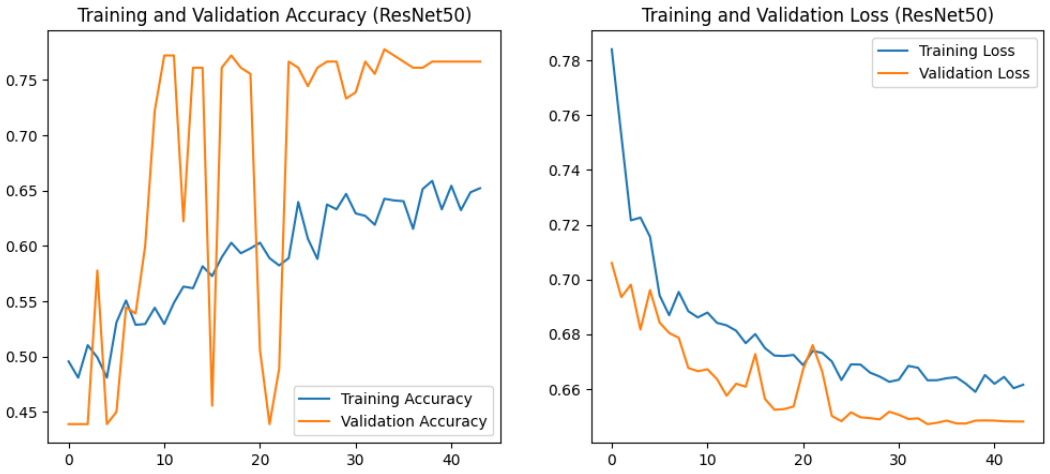
Parameter	Nilai / Keterangan
Optimizer	Adam (Adaptive Moment Estimation)
Learning Rate	0,0001 (1e-4)
Loss Function	Binary CrossEntropy
Batch Size	32
Jumlah Epoch	50 (Early Stopping: ResNet50 berhenti di Ep.44, MobileNetV2 di Ep.30)
Metrik Pantau	Akurasi Validasi (val_accuracy)

Pemilihan algoritma optimasi Adam dengan learning rate sebesar 1e-4 bertujuan untuk memastikan pembaruan bobot berlangsung secara stabil dan presisi. Learning rate yang relatif kecil ini dipilih secara khusus untuk mencegah model mengalami overshooting saat melakukan Fine-Tuning pada lapisan-lapisan yang sudah memiliki representasi fitur yang baik dari prapelatihan ImageNet. Binary CrossEntropy digunakan sebagai fungsi kerugian karena kesesuaiannya dengan tugas klasifikasi dua kelas dengan lapisan keluaran Sigmoid.

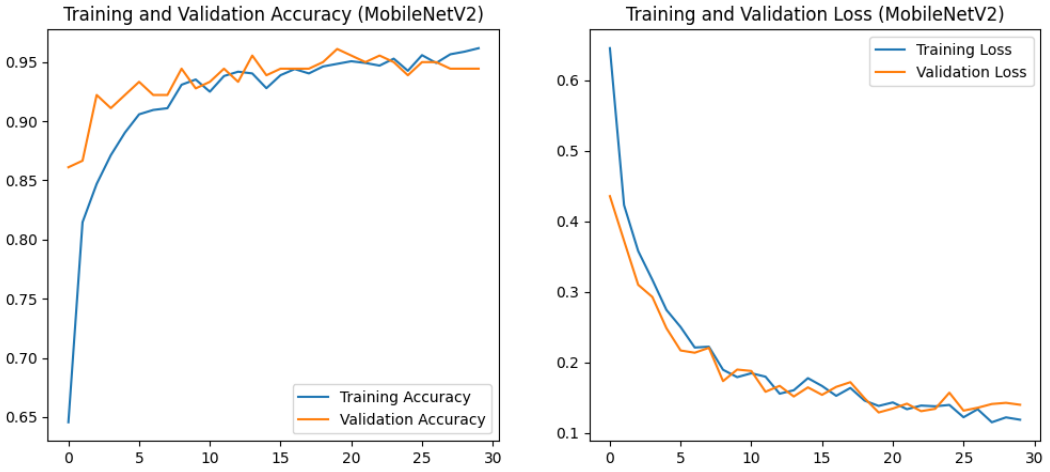
4.4.1 Analisis Dinamika Kurva Pelatihan

Dinamika pelatihan pada kedua model dipantau secara intensif melalui rekam jejak nilai loss dan akurasi pada data latih maupun data validasi di setiap

epoch. Perbandingan kurva pelatihan kedua model divisualisasikan pada Gambar 4.3 dan 4.4 berikut.



Gambar 4.3 Grafik Perbandingan Kurva Training dan Validation Accuracy & Loss Arsitektur ResNet50



Gambar 4.4 Grafik Perbandingan Kurva Training dan Validation Accuracy & Loss Arsitektur MobileNetV2

Berdasarkan visualisasi kurva pada Gambar 4.3 dan 4.4, terdapat beberapa pola dinamika pelatihan yang perlu dicermati. Pada ResNet50, nilai training loss menunjukkan tren penurunan yang fluktuatif dan lambat sepanjang proses pelatihan, mencapai nilai akhir sebesar 0,6632 pada epoch terbaik (ke-34). Nilai validation loss terpantau bergerak cukup lambat dan hampir stagnan beriringan dengan training loss sejak epoch pertengahan, mengindikasikan bahwa kapasitas backbone yang masih sepenuhnya dibekukan (frozen) membatasi kemampuan

adaptasi model pada data mangga. Akurasi validasi tertinggi (best val_accuracy) yang dicapai oleh ResNet50 adalah sebesar 77,78%.

Pada MobileNetV2, kurva training loss menunjukkan penurunan yang jauh lebih cepat dan stabil dibandingkan ResNet50, dengan nilai akhir sebesar 0,1379 pada epoch terbaik (ke-20). Pola konvergensi validation accuracy MobileNetV2 menunjukkan karakteristik yang stabil dan konsisten meningkat, dengan akurasi validasi tertinggi mencapai 96,11%.

Ringkasan perbandingan kinerja pelatihan kedua model pada epoch terbaik disajikan pada Tabel 4.5.

Tabel 4.5 Ringkasan Kinerja Pelatihan pada Epoch Terbaik

Metrik	ResNet50 (Frozen)	MobileNetV2	ResNet50 (Fine-Tuned)
Training Loss Akhir (Epoch Terbaik)	0,6632	0,1379	0,2706
Validation Loss Terbaik	0,6472	0,1287	0,2579
Training Accuracy (Epoch Terbaik)	64,26%	94,85%	89,93%
Validation Accuracy Terbaik	77,78%	96,11%	91,67%
Epoch Terbaik (Early Stop)	34 dari 44 epoch	20 dari 30 epoch	13 dari 20 epoch
Waktu Pelatihan per Epoch (rata-rata)	~4,01 detik	~1,83 detik	~4,78 detik

Perbedaan waktu pelatihan per *epoch* antara kedua model mencerminkan perbedaan beban komputasi yang signifikan. MobileNetV2 dengan jumlah parameter yang jauh lebih sedikit (~3,4 juta) terbukti menyelesaikan setiap *epoch* dalam durasi yang jauh lebih singkat (~1,83 detik) dibandingkan ResNet50 *Frozen* (~4,01 detik) maupun ResNet50 *Fine-Tuned* (~4,78 detik) yang memiliki ~25,6 juta

parameter. Analisis *trade-off* antara efisiensi komputasi dan akurasi ini dibahas lebih mendalam pada Subbab 4.7.

4.4.2 Proses dan Dinamika Fine-Tuning ResNet50

Hasil pelatihan ResNet50 pada fase frozen (backbone beku) menghasilkan akurasi pengujian yang relatif rendah, yakni sebesar 78,00%, dengan nilai Weighted F1-Score sebesar 0,78. Selisih performa yang signifikan dibandingkan MobileNetV2 (97,00%) mengindikasikan adanya keterbatasan mendasar pada pendekatan baseline tersebut. Kondisi ini diduga kuat disebabkan oleh ketidakmampuan backbone yang sepenuhnya beku untuk mengadaptasi representasi fiturnya ke domain citra mangga yang spesifik. Bobot yang dikunci secara total berarti model hanya mengandalkan fitur-fitur generik dari ImageNet tanpa proses penyesuaian terhadap karakteristik visual unik buah mangga, seperti gradasi warna permukaan, pola bercak, dan tekstur kulit yang membedakan mangga layak dan tidak layak. Oleh karena itu, dilakukan proses Fine-Tuning secara selektif sebagai upaya untuk meningkatkan kemampuan adaptasi ResNet50 terhadap domain citra yang sedang diteliti.

Mekanisme Fine-Tuning yang diterapkan pada penelitian ini mengikuti paradigma selektif unfreeze, yaitu hanya membuka kunci (unfreeze) 30 layer terakhir dari backbone ResNet50, sementara lapisan-lapisan awal tetap dibekukan. Pemilihan 30 layer terakhir didasarkan pada prinsip hierarki fitur dalam jaringan konvolusi yang telah mapan dalam literatur deep learning. Layer-layer awal pada ResNet50 bertanggung jawab mengekstrak fitur low-level yang bersifat universal, seperti tepi (edges), sudut (corners), dan tekstur dasar yang relevan untuk hampir semua domain visual. Fitur-fitur ini tidak perlu dimodifikasi karena sifatnya yang generik. Sebaliknya, 30 layer terakhir mengekstrak fitur high-level yang sangat domain-spesifik, seperti representasi abstrak dari warna, pola permukaan, dan bentuk keseluruhan objek. Lapisan inilah yang paling besar manfaatnya apabila disesuaikan (Fine-Tuned) dengan karakteristik visual dataset mangga. Dalam implementasinya, seluruh layer terlebih dahulu dibekukan kembali, kemudian dihitung indeks layer Global Average Pooling (GAP) sebagai batas antara backbone dan custom head. Selanjutnya, ditetapkan fine_tune_start = pool_idx - 30, sehingga

30 layer terakhir backbone beserta seluruh custom head dibuka untuk proses pembaruan bobot.

Aspek teknis krusial yang diterapkan dalam proses Fine-Tuning ini adalah pembekuan layer BatchNormalization (BN) meskipun berada di dalam segmen yang diunfreeze. Lapisan BatchNormalization menyimpan statistik running mean dan running variance yang telah terkalibrasi secara akurat melalui pelatihan pada jutaan citra ImageNet. Apabila statistik ini diperbarui menggunakan dataset mangga yang berukuran relatif kecil (1.360 citra latih), distribusi statistik yang sudah stabil tersebut akan terdegradasi, yang justru berpotensi merusak representasi fitur yang telah matang. Oleh karena itu, proteksi layer BatchNormalization menjadi prasyarat penting untuk menjaga integritas normalisasi internal yang telah terbangun. Selain itu, proses Fine-Tuning dikompilasi ulang menggunakan learning rate yang jauh lebih kecil, yakni 1e-5 (10 kali lebih kecil dari learning rate fase baseline sebesar 1e-4). Pemilihan learning rate yang sangat kecil ini bertujuan untuk mencegah terjadinya catastrophic forgetting, yaitu kondisi di mana pembaruan bobot yang terlalu agresif dapat menghapus representasi fitur yang telah matang dan bermanfaat dari bobot pralatih. Callback ReduceLROnPlateau dengan faktor reduksi 0,2 dan patience 5 epoch juga diterapkan untuk menyesuaikan learning rate secara adaptif apabila validasi loss mengalami stagnasi.

Proses Fine-Tuning dijalankan selama maksimum 20 epoch dengan callback EarlyStopping (patience=10) dan ModelCheckpoint untuk menyimpan bobot terbaik. Dari 20 epoch yang dijalankan, model terbaik dicapai pada epoch ke-13 dengan validation loss terendah sebesar 0,2579 dan validation accuracy sebesar 91,67%. Dibandingkan dengan fase frozen ResNet50 yang mencapai validation accuracy terbaik sebesar 77,78% pada epoch ke-34, fase Fine-Tuning menunjukkan konvergensi yang jauh lebih cepat dan stabil, mencapai 91,67% hanya dalam 13 epoch. Pola konvergensi ini mengindikasikan bahwa dengan membuka kunci 30 layer terakhir, model dapat secara aktif menyesuaikan representasi fitur tingkat tinggi terhadap karakteristik visual mangga, sehingga gap antara training accuracy dan validation accuracy menjadi lebih kecil dibandingkan fase baseline. Total waktu Fine-Tuning adalah 95,55 detik (sekitar 1,59 menit), dengan rata-rata ~4,78

detik per epoch. Perbandingan lengkap antara kinerja pelatihan ResNet50 Frozen dan ResNet50 Fine-Tuned disajikan pada Tabel 4.5.

4.5 Evaluasi Kinerja Model

Setelah proses pelatihan selesai, bobot model terbaik dari masing-masing arsitektur yang ditentukan berdasarkan nilai validation loss terendah dimuat kembali untuk dilakukan pengujian final. Evaluasi dilaksanakan secara eksklusif menggunakan data uji (testing set) yang terdiri dari 160 citra dan sepenuhnya terpisah dari data yang digunakan selama proses pelatihan maupun validasi.

4.5.1 Evaluasi Metrik Akurasi, Presisi, Recall, dan F1-Score

Kinerja klasifikasi kedua model diukur menggunakan metrik standar evaluasi klasifikasi biner. Hasil evaluasi pada data uji dirangkum pada Tabel 4.6.

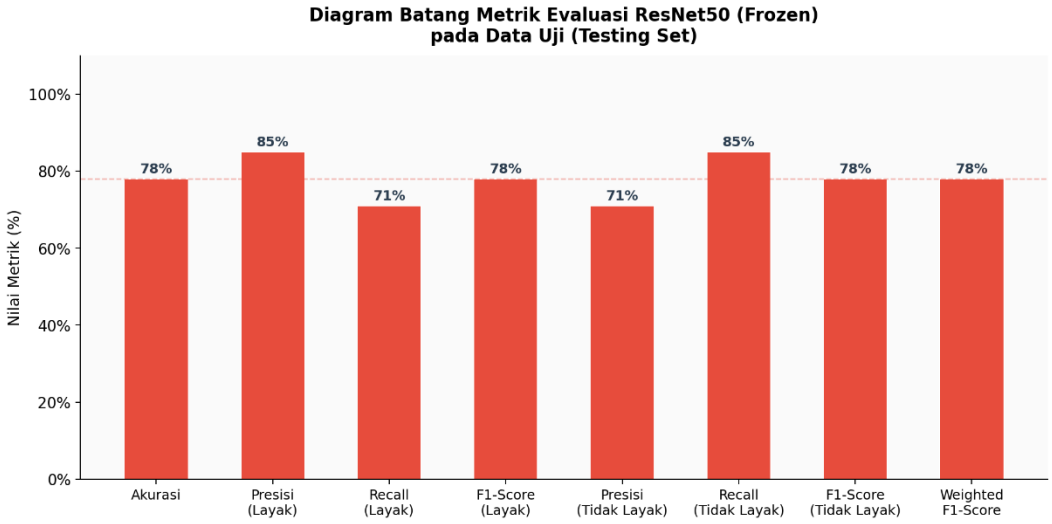
Tabel 4.6 Hasil Evaluasi Komparatif pada Data Uji (Testing Set)

Metrik Evaluasi	ResNet50 (Frozen)	MobileNetV2	ResNet50 (Fine-Tuned)
Akurasi (Accuracy)	78,00%	97,00%	94,00%
Presisi Kelas Layak	0,85	0,98	0,99
Recall Kelas Layak	0,71	0,97	0,91
F1-Score Kelas Layak	0,78	0,97	0,95
Presisi Kelas Tidak Layak	0,71	0,96	0,90
Recall Kelas Tidak Layak	0,85	0,97	0,99
F1-Score Kelas Tidak Layak	0,78	0,97	0,94
Weighted Average F1-Score	0,78	0,97	0,94

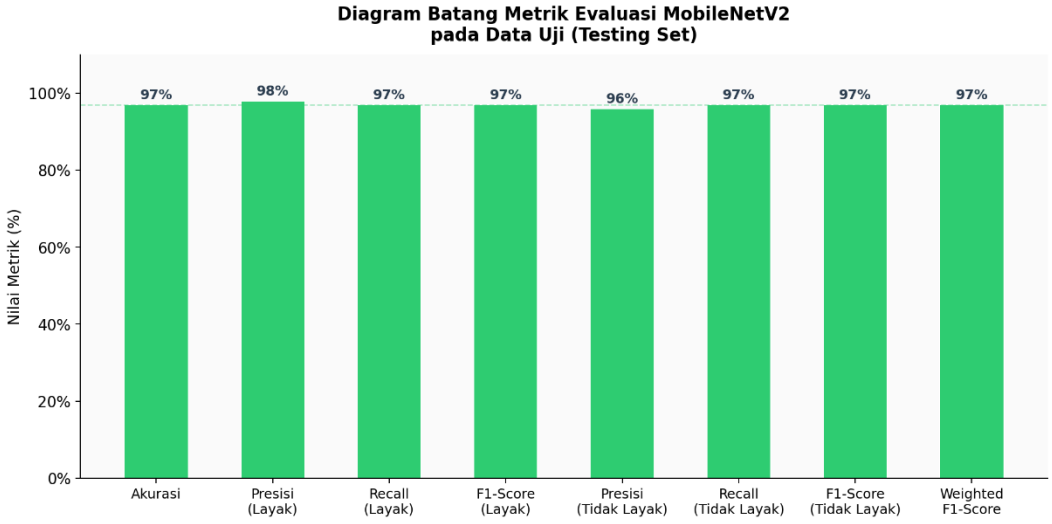
Berdasarkan Tabel 4.6, ResNet50 Frozen mencatatkan akurasi pengujian sebesar 78,00%, MobileNetV2 sebesar 97,00%, dan ResNet50 Fine-Tuned sebesar 94,00%. Temuan ini memperlihatkan bahwa Fine-Tuning secara selektif mampu meningkatkan akurasi ResNet50 sebesar 16,00 poin persentase dibandingkan fase

frozennya. Meskipun demikian, MobileNetV2 masih unggul dibandingkan ResNet50 Fine-Tuned dengan selisih sebesar 3,00 poin persentase. Nilai Weighted Average F1-Score digunakan sebagai metrik penentu yang lebih representatif karena memperhitungkan keseimbangan antara Presisi dan Recall pada kedua kelas secara bersamaan.

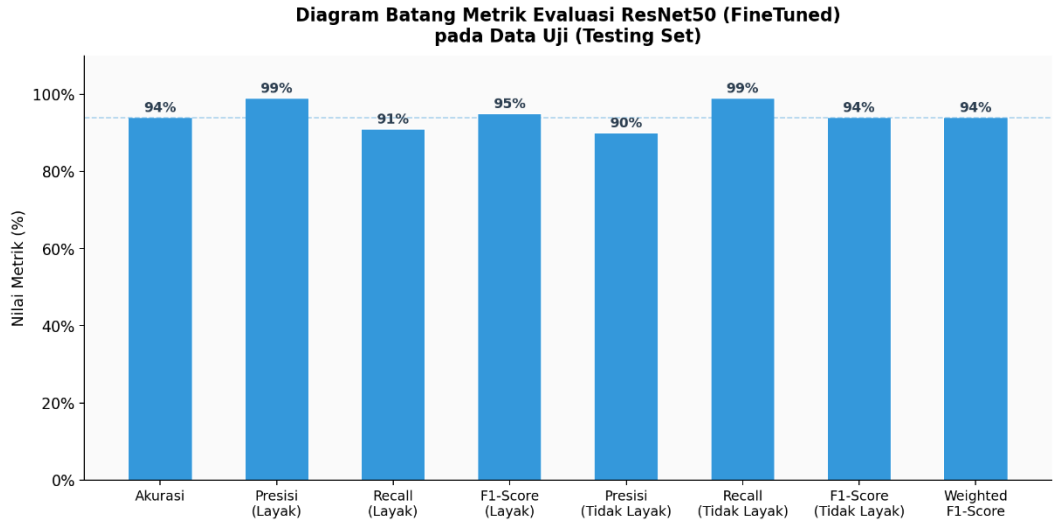
Untuk memperjelas perbandingan kinerja antar model secara visual, berikut disajikan diagram batang metrik evaluasi masing-masing arsitektur yang diuji pada data uji (testing set).



Gambar 4.5 Diagram Batang Metrik Evaluasi ResNet50 (Frozen) pada Data Uji



Gambar 4.6 Diagram Batang Metrik Evaluasi MobileNetV2 pada Data Uji

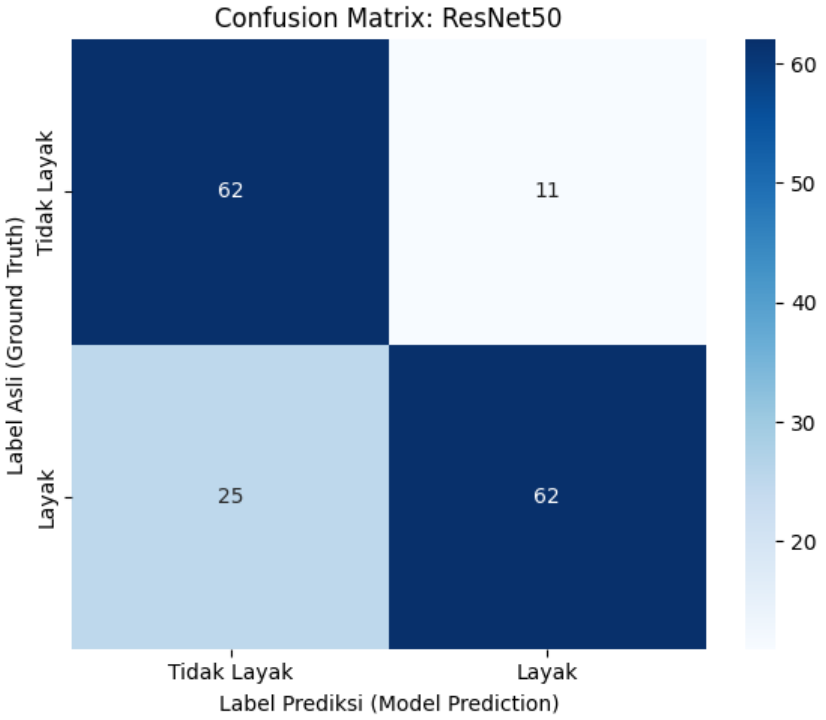


Gambar 4.7 Diagram Batang Metrik Evaluasi ResNet50 (Fine-Tuned) pada Data Uji

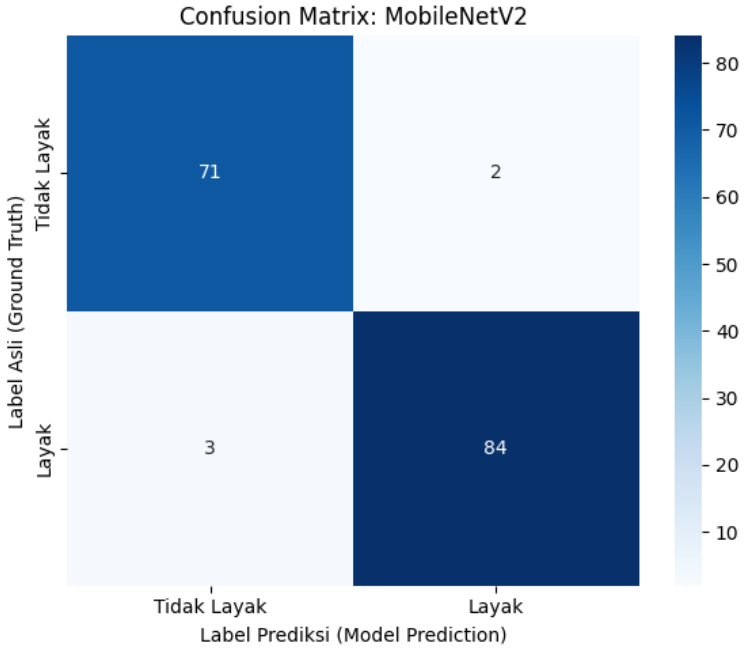
Berdasarkan Gambar 4.5, 4.6, dan 4.7, terlihat secara visual bahwa MobileNetV2 mendominasi seluruh metrik evaluasi dengan nilai yang sangat tinggi dan konsisten di atas 96% pada semua kategori. ResNet50 (Frozen) mencatatkan nilai terendah pada hampir seluruh metrik, khususnya pada Recall Kelas Layak (71%) dan Presisi Kelas Tidak Layak (71%). Sementara itu, ResNet50 (Fine-Tuned) menunjukkan peningkatan yang signifikan dibanding versi Frozennya, mendekati performa MobileNetV2 namun dengan profil galat yang berbeda, yaitu Recall Kelas Layak lebih rendah (91%) namun Presisi Kelas Layak lebih tinggi (99%). Perbandingan ketiga diagram tersebut mempertegas keunggulan MobileNetV2 secara komprehensif pada dataset mangga ini.

4.5.2 Analisis Confusion Matrix

Untuk memperoleh gambaran yang lebih granular mengenai karakteristik kesalahan prediksi pada tingkat kelas, dilakukan analisis Confusion Matrix terhadap hasil prediksi kedua model pada keseluruhan 160 data uji. Visualisasi Confusion Matrix untuk masing-masing model disajikan pada Gambar 4.8 dan 4.9.



Gambar 4.8 Confusion Matrix ResNet50 pada Data Uji



Gambar 4.9 Confusion Matrix MobileNetV2 pada Data Uji

Tabel 4.7 Rekap Nilai Confusion Matrix Ketiga Model pada Data Uji

Komponen	ResNet50 (Frozen)	MobileNetV2	ResNet50 (Fine-Tuned)
True Positive (TP) Mangga Layak terprediksi benar	62	84	79
True Negative (TN) Mangga Tidak Layak terprediksi benar	62	71	72
False Positive (FP) Tidak Layak salah diprediksi Layak	11	2	1
False Negative (FN) Layak salah diprediksi Tidak Layak	25	3	8

Analisis terhadap Tabel 4.7 mengungkap beberapa temuan penting. Nilai False Negative (FN) menjadi perhatian kritis dalam konteks aplikasi klasifikasi kelayakan buah, karena kesalahan tipe ini berarti buah mangga yang sebenarnya layak konsumsi justru diklasifikasikan sebagai tidak layak. Dalam skenario industri pertanian, galat tipe ini berdampak langsung pada kerugian ekonomi akibat pembuangan produk yang masih baik. Sebaliknya, nilai False Positive (FP) yang tinggi membawa risiko lolosnya buah tidak layak ke rantai distribusi, yang berpotensi merugikan konsumen.

Berdasarkan perbandingan nilai FP dan FN pada Tabel 4.7, MobileNetV2 menunjukkan kecenderungan galat yang jauh lebih rendah pada kedua kategori (FP=2 dan FN=3) dibandingkan ResNet50 Frozen (FP=11 dan FN=25). ResNet50 Fine-Tuned menunjukkan perbaikan profil galat yang sangat signifikan, dengan FP=1 dan FN=8, mendekati performa MobileNetV2. Implikasi praktis dari perbedaan profil galat ketiga model ini akan dibahas lebih lanjut pada Subbab 4.7.

4.5.3 Evaluasi Kinerja ResNet50 Fine-Tuned

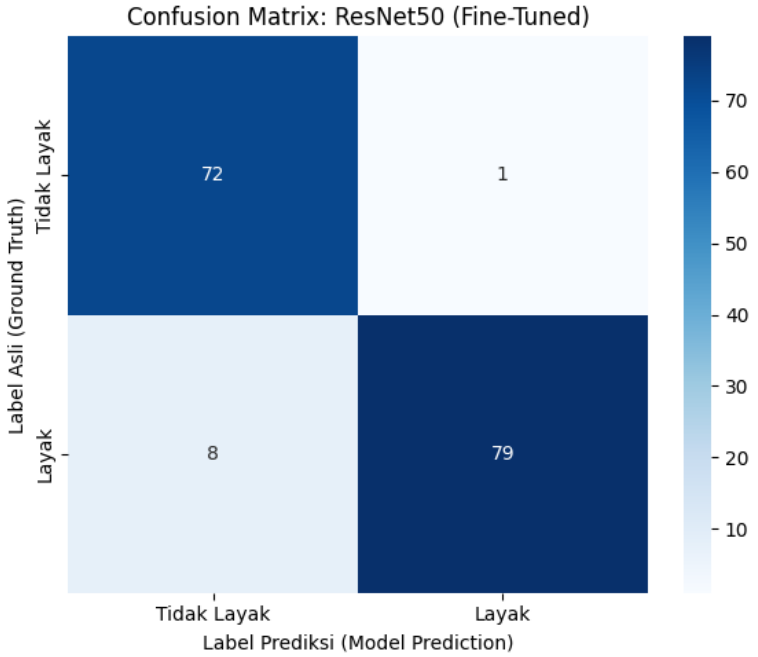
Setelah dilakukan proses Fine-Tuning selektif pada 30 layer terakhir ResNet50, evaluasi akhir dilaksanakan pada data uji yang sama (160 citra)

menggunakan bobot terbaik yang tersimpan dari epoch ke-13. Hasil evaluasi komparatif ResNet50 Frozen, MobileNetV2, dan ResNet50 Fine-Tuned pada data uji disajikan secara lengkap pada Tabel 4.8.

Tabel 4.8 Perbandingan Metrik Evaluasi: ResNet50 Frozen vs. MobileNetV2 vs. ResNet50 Fine-Tuned

Metrik Evaluasi	ResNet50 (Frozen)	MobileNetV2	ResNet50 (Fine-Tuned)
Akurasi (Accuracy)	78,00%	97,00%	94,00%
Presisi Kelas Layak	0,85	0,98	0,99
Recall Kelas Layak	0,71	0,97	0,91
F1-Score Kelas Layak	0,78	0,97	0,95
Presisi Kelas Tidak Layak	0,71	0,96	0,90
Recall Kelas Tidak Layak	0,85	0,97	0,99
F1-Score Kelas Tidak Layak	0,78	0,97	0,94
Weighted Average F1-Score	0,78	0,97	0,94

Berdasarkan Tabel 4.8, ResNet50 Fine-Tuned berhasil mencatatkan akurasi pengujian sebesar 94,00%, meningkat secara signifikan sebesar 16,00 poin persentase dibandingkan ResNet50 Frozen (78,00%). Weighted Average F1-Score meningkat dari 0,78 menjadi 0,94, yang menandakan peningkatan kualitas klasifikasi yang lebih komprehensif pada kedua kelas sekaligus. Visualisasi Confusion Matrix ResNet50 Fine-Tuned pada data uji disajikan pada gambar berikut.



Gambar 4.10 Confusion Matrix ResNet50 Fine-Tuned pada Data Uji

Tabel 4.9 Rekap Nilai Confusion Matrix Ketiga Model pada Data Uji

Komponen	ResNet50 (Frozen)	MobileNetV2	ResNet50 (Fine-Tuned)
True Positive (TP) Mangga Layak terprediksi benar	62	84	79
True Negative (TN) Mangga Tidak Layak terprediksi benar	62	71	72
False Positive (FP) Tidak Layak salah diprediksi Layak	11	2	1
False Negative (FN) Layak salah diprediksi Tidak Layak	25	3	8

Analisis mendalam terhadap profil galat ResNet50 Fine-Tuned pada Tabel 4.9 mengungkap perubahan yang sangat substansial dibandingkan fase baseline. Nilai False Negative (FN) mengalami penurunan drastis dari 25 menjadi 8 citra, artinya hanya 8 mangga layak yang keliru diklasifikasikan sebagai tidak layak.

Dalam konteks aplikasi kelayakan mangga, penurunan FN ini memiliki implikasi ekonomi yang sangat berarti karena jumlah produk layak yang terbuang secara keliru berkurang secara signifikan. Di sisi lain, nilai False Positive (FP) juga mengalami penurunan tajam dari 11 menjadi hanya 1 citra, yang berarti risiko lolosnya buah tidak layak ke rantai distribusi hampir sepenuhnya teratasi. Perubahan profil galat ini mencerminkan bahwa Fine-Tuning berhasil membantu model membentuk decision boundary yang jauh lebih akurat dalam memisahkan kedua kelas. Meskipun demikian, jika dibandingkan dengan MobileNetV2 yang hanya memiliki FP=2 dan FN=3, ResNet50 Fine-Tuned (FP=1, FN=8) masih mencatatkan akurasi yang sedikit di bawah MobileNetV2, menandakan bahwa Fine-Tuning berhasil memperkecil kesenjangan secara signifikan namun belum sepenuhnya melampaui performa arsitektur MobileNetV2.

4.6 Perbandingan Waktu Inferensi dan Efisiensi Model

Selain kinerja akurasi, aspek efisiensi komputasi merupakan dimensi evaluasi yang relevan dalam konteks implementasi sistem klasifikasi berbantuan komputer pada aplikasi nyata. Pengukuran waktu inferensi (inference time) dilakukan dengan mencatat durasi rata-rata yang dibutuhkan oleh masing-masing model untuk memproses satu citra tunggal pada hardware yang sama.

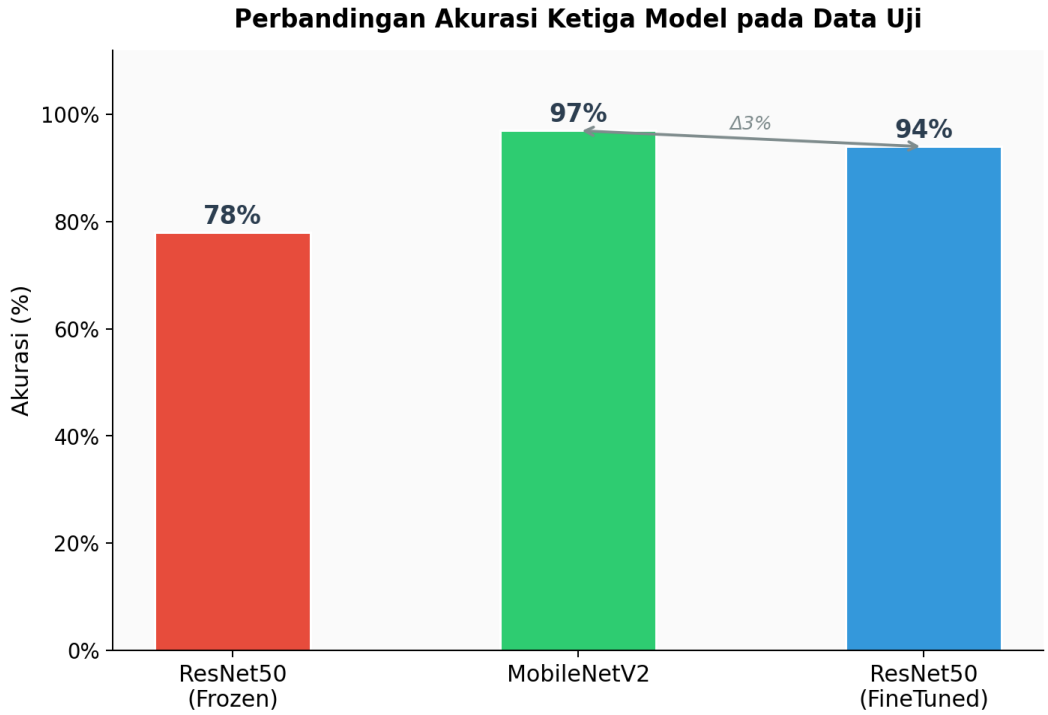
Tabel 4.10 Perbandingan Efisiensi Komputasi Ketiga Model

Parameter Efisiensi	ResNet50 (Frozen)	MobileNetV2	ResNet50 (Fine-Tuned)
Jumlah Total Parameter	~25,6 Juta	~3,4 Juta	~25,6 Juta
Ukuran Model (Model Size)	~98 MB	~14 MB	~98 MB
Waktu Inferensi per Citra (rata-rata)	~12,70 ms	~7,60 ms	~12,80 ms
Waktu Total Evaluasi (160 citra)	~2,03 detik	~1,22 detik	~2,05 detik

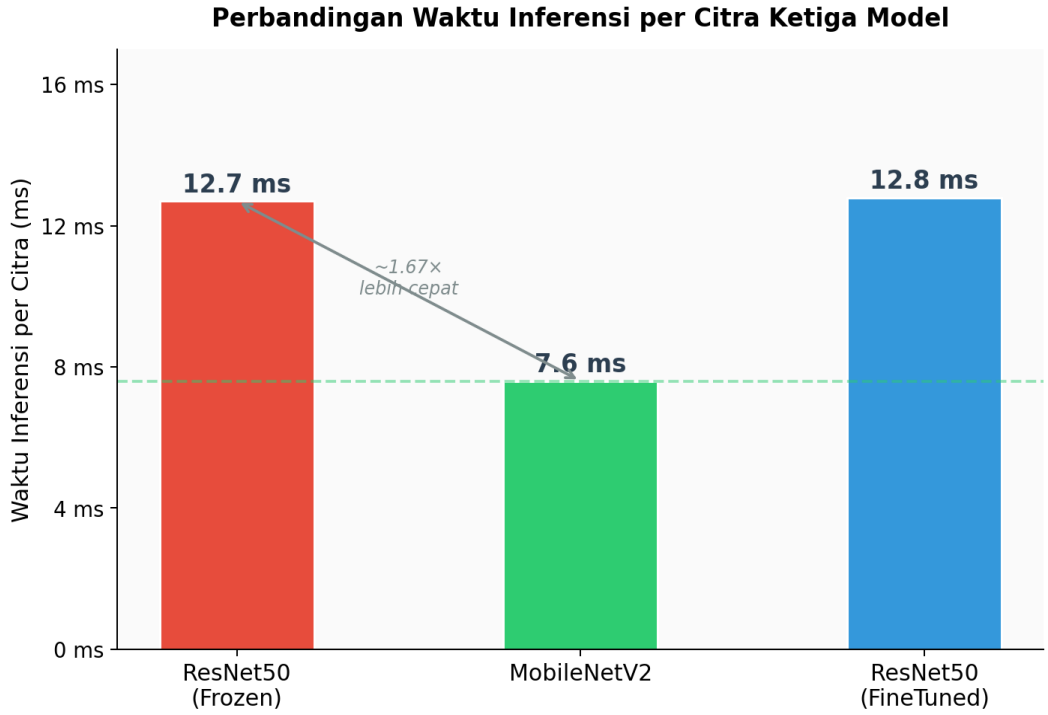
Data pada Tabel 4.10 mempertegas perbedaan fundamental antara kedua arsitektur dari sudut pandang efisiensi. MobileNetV2, dengan jumlah parameter

yang hanya sekitar 13,3% dari ResNet50, terbukti menghasilkan waktu inferensi yang jauh lebih singkat (~7,60 ms per citra dibandingkan ~12,70–12,80 ms pada ResNet50). Keunggulan efisiensi ini menjadikan MobileNetV2 sebagai kandidat yang paling sesuai untuk skenario *deployment* pada perangkat dengan sumber daya komputasi terbatas, seperti *smartphone* milik petani atau sistem *embedded* pada mesin sortir buah di gudang distribusi.

Untuk mempermudah perbandingan visual, berikut disajikan diagram batang perbandingan akurasi dan waktu inferensi per citra dari ketiga model yang dievaluasi.



Gambar 4.11 Diagram Batang Perbandingan Akurasi Ketiga Model pada Data Uji



Gambar 4.12 Diagram Batang Perbandingan Waktu Inferensi per Citra Ketiga Model

Berdasarkan Gambar 4.11, MobileNetV2 secara visual unggul jauh dalam akurasi (97%) dibandingkan ResNet50 Frozen (78%) maupun ResNet50 Fine-Tuned (94%), dengan selisih 3 poin persentase terhadap model terbaik kedua. Gambar 4.12 memperlihatkan bahwa MobileNetV2 juga menjadi model tercepat dengan waktu inferensi rata-rata hanya 7,60 ms per citra, sekitar 1,67 kali lebih cepat dibandingkan ResNet50 Frozen (12,70 ms) maupun ResNet50 Fine-Tuned (12,80 ms). Kedua diagram tersebut memperkuat temuan ini dengan menunjukkan MobileNetV2 unggul di kedua dimensi sekaligus, membuktikan bahwa arsitektur lightweight ini bukan sekadar lebih efisien, melainkan juga lebih akurat pada kasus spesifik klasifikasi kelayakan buah mangga ini.

4.7 Pembahasan (Discussion)

Subbab ini mensintesis seluruh temuan kuantitatif yang telah dijabarkan untuk menghasilkan analisis interpretatif yang menjawab rumusan masalah penelitian ini secara komprehensif. Pembahasan difokuskan pada tiga dimensi utama: superioritas akurasi, efisiensi komputasi, dan implikasi praktis dari temuan komparatif ini.

4.7.1 Analisis Keunggulan Akurasi dan Kemampuan Ekstraksi Fitur

Temuan bahwa MobileNetV2 mencatatkan akurasi (97,00%) dan *Weighted F1-Score* (0,97) yang lebih tinggi dibandingkan ResNet50 *Frozen* (78,00%) maupun ResNet50 *Fine-Tuned* (94,00%) dapat diinterpretasikan dari perspektif kesesuaian kapasitas arsitektural terhadap skala dataset. MobileNetV2 mengandalkan mekanisme *Inverted Residual Block* dan *Linear Bottleneck* berbasis *Depthwise Separable Convolution* dengan jumlah parameter yang ringkas (~3,4 juta parameter). Karakteristik ini membuat MobileNetV2 memiliki efek regularisasi struktural alami yang mencegah *over-parameterization* pada dataset menengah (1.700 citra), sekaligus menjaga keutuhan informasi fitur visual halus—seperti bercak awal antraknosa dan jejak tusukan lalat buah—pada *manifold* berdimensi rendah tanpa terdistorsi oleh aktivasi non-linear ReLU pada lapisan *bottleneck.*

Di sisi lain, ResNet50 memiliki kapasitas parameter yang sangat besar (~25,6 juta parameter) dengan 50 lapisan berbobot. Ketika seluruh lapisan *backbone* dibekukan (*frozen*), fitur tingkat tinggi bawaan ImageNet belum cukup spesifik untuk memisahkan variasi tekstur patologis halus buah mangga di tengah latar belakang yang kompleks (*cluttered background*), sehingga akurasinya tertahan pada 78,00%. Ketika 30 lapisan terakhir dibuka (*fine-tuned*), ResNet50 memang mampu mengadaptasikan representasi fiturnya secara drastis hingga mencapai akurasi 94,00%, namun kompleksitas parameter yang masif membuat konvergensinya pada dataset berukuran 1.360 citra latih masih sedikit di bawah efisiensi ekstraksi fitur MobileNetV2. Temuan ini sejalan dengan hasil studi Santoso et al. (2024) dan Gulzar (2023), yang membuktikan bahwa arsitektur ringan berbasis MobileNetV2 mampu melampaui arsitektur jaringan dalam konvensional pada tugas klasifikasi citra pertanian berskala spesifik.

4.7.2 Analisis Trade-off Akurasi vs. Efisiensi

Hasil penelitian ini menunjukkan bahwa pada domain klasifikasi kelayakan buah mangga berdasarkan standar SNI 3164:2009, MobileNetV2 berhasil mematahkan dilema *trade-off* klasik antara akurasi dan efisiensi komputasi.

ResNet50 *Frozen* maupun *Fine-Tuned*, meskipun memiliki arsitektur yang jauh lebih dalam, membutuhkan ukuran penyimpanan model yang besar (~98 MB) dan waktu inferensi yang lebih lambat (~12,70–12,80 ms per citra) dengan capaian akurasi maksimal 94,00%. Sebaliknya, MobileNetV2 mendemonstrasikan superioritas mutlak pada kedua dimensi secara simultan: mencapai akurasi tertinggi (97,00%) sekaligus efisiensi komputasi terbaik dengan ukuran model hanya ~14 MB (sekitar 7 kali lebih ringkas) dan waktu inferensi ~7,60 ms per citra (~1,67 kali lebih cepat).

Implikasi praktis dari temuan ini sangat jelas bagi skenario *deployment* di lapangan. Baik untuk implementasi terpusat maupun untuk operasional langsung pada perangkat *edge computing* dan *resource-constrained devices*—seperti *smartphone* milik petani, pengepul lokal, maupun modul kamera cerdas (*smart camera module*) pada lini sortir UMKM pertanian—MobileNetV2 merupakan pilihan arsitektur utama yang paling direkomendasikan. Sementara itu, ResNet50 *Fine-Tuned* (94,00%) yang memiliki tingkat *False Positive* sangat rendah (FP=1; Presisi Kelas Layak 0,99) dapat dipertimbangkan sebagai model pendamping (*secondary verifier*) pada lingkungan server berkapasitas GPU tinggi apabila sistem menghendaki penyaringan berlapis (*ensemble*).

4.7.3 Komparasi dengan Penelitian Terdahulu

Hasil yang diperoleh dalam penelitian ini menunjukkan dinamika yang menarik jika disandingkan dengan tren yang dilaporkan dalam literatur terkait klasifikasi kualitas komoditas pertanian menggunakan CNN. Sebagai komparasi, Suhendar, Efelina, & Ziveria (2022) melaporkan bahwa dalam klasifikasi kualitas buah secara umum, arsitektur ResNet50 lebih unggul dengan akurasi sebesar 95,7% dibandingkan MobileNetV2 yang mencapai 76,23%. Namun, pada studi spesifik komoditas pertanian dengan transfer learning, Santoso et al. (2024) justru menemukan bahwa MobileNetV2 (90,84%) mengungguli ResNet50 (68,08%) secara signifikan. Di sisi lain, Doan & Thi (2023) melaporkan pencapaian akurasi sebesar 94,2% dalam sistem grading mangga menggunakan arsitektur CNN kustom, sedangkan Zárate & Hernández (2024) serta Gulzar (2023) mengonfirmasi

kecepatan konvergensi dan akurasi tinggi MobileNetV2 untuk operasional pertanian skala kecil dan aplikasi mobile.

Penelitian ini memperluas cakupan studi-studi tersebut dan memberikan perspektif temuan empiris yang baru. Berbeda dengan kesimpulan Suhendar et al. (2022) pada buah umum, pengujian pada penelitian ini membuktikan bahwa untuk kasus spesifik klasifikasi kelayakan buah mangga berdasarkan standar SNI 3164:2009 dengan latar belakang bervariasi (cluttered background), MobileNetV2 mampu mengungguli ResNet50 secara meyakinkan. Dengan akurasi mencapai 97,00%, kinerja MobileNetV2 pada penelitian ini tidak hanya melampaui capaian MobileNetV2 pada studi Santoso et al. (2024) (90,84%) dan sistem grading mangga milik Doan & Thi (2023) (94,2%), tetapi juga mengungguli arsitektur ResNet50 Fine-Tuned yang dioptimasi dalam penelitian ini (94,00%).

Selain komparasi dari sisi akurasi klasifikasi, penelitian ini menambahkan dimensi analisis efisiensi komputasi yang belum dievaluasi secara mendalam pada penelitian-penelitian sebelumnya. Kebaruan (novelty) penelitian ini terletak pada analisis komparatif yang terstruktur antara arsitektur berbasis residual deep network (ResNet50) dan arsitektur berbasis lightweight network (MobileNetV2) dengan menggunakan protokol eksperimen yang terkontrol. Penggunaan dataset mangga dengan fitur patologis yang halus, dikombinasikan dengan skema augmentasi data on-the-fly, secara efektif telah meningkatkan representativitas data latih terhadap kondisi penyortiran nyata di lapangan.

4.7.4 Analisis Dampak Fine-Tuning terhadap Kinerja ResNet50

Penerapan *Fine-Tuning* secara selektif pada 30 layer terakhir ResNet50 menghasilkan peningkatan akurasi yang sangat substansial. Akurasi pengujian meningkat dari 78,00% (ResNet50 *Frozen*) menjadi 94,00% (ResNet50 *Fine-Tuned*), sebuah peningkatan sebesar 16,00 poin persentase atau ekuivalen dengan peningkatan relatif sebesar 20,5%. Weighted Average F1-Score turut meningkat signifikan dari 0,78 menjadi 0,94, mengindikasikan bahwa peningkatan tersebut bersifat menyeluruh pada kedua kelas secara proporsional, bukan hanya pada satu kelas tertentu. Peningkatan paling dramatis terlihat pada nilai False Negative (FN) yang turun dari 25 menjadi 8 dan False Positive (FP) yang turun dari 11 menjadi 1,

membuktikan bahwa model berhasil membentuk decision boundary yang jauh lebih akurat dan simetris antara kedua kelas setelah Fine-Tuning.

Dari perspektif arsitektural, efektivitas Fine-Tuning parsial (30 layer terakhir) pada ResNet50 dapat dipahami melalui konsep hierarki fitur dan domain adaptation. ResNet50 memiliki arsitektur yang sangat dalam dengan 50 layer berbobot, yang secara hierarkis membangun representasi dari fitur sederhana (low-level: tepi, warna dasar) di layer awal hingga fitur abstrak yang sangat domain-spesifik (high-level: representasi objek utuh) di layer akhir. Dengan membekukan layer awal yang mengekstrak fitur universal dan hanya melatih ulang 30 layer terakhir yang mengekstrak fitur domain-spesifik, model dapat secara efisien mengadaptasikan representasinya terhadap karakteristik visual buah mangga tanpa kehilangan pengetahuan umum yang telah diperoleh dari ImageNet. Pendekatan ini sejalan dengan prinsip transfer learning bertahap yang telah terbukti efektif dalam berbagai studi domain adaptation pada literatur deep learning, di mana adaptasi hanya pada lapisan atas (top layers) memberikan keseimbangan optimal antara stabilitas representasi dan fleksibilitas adaptasi domain.

Meskipun Fine-Tuning berhasil meningkatkan akurasi ResNet50 secara signifikan, posisi ResNet50 Fine-Tuned (94,00%) masih berada di bawah MobileNetV2 (97,00%), dengan selisih sebesar 3,00 poin persentase. Fakta ini menunjukkan bahwa Fine-Tuning berhasil memperkecil kesenjangan secara dramatis (dari 19,00 poin menjadi 3,00 poin), namun MobileNetV2 tetap unggul pada kasus spesifik dataset mangga ini. Dari sisi biaya komputasi, proses Fine-Tuning memerlukan waktu tambahan sebesar 95,55 detik (~1,59 menit) di atas waktu pelatihan baseline ResNet50. Dengan demikian, gain akurasi sebesar 16 poin yang diperoleh dari investasi waktu kurang dari 2 menit menjadikan Fine-Tuning sebagai strategi yang sangat cost-effective untuk meningkatkan performa ResNet50. Adapun dari sisi waktu inferensi, ResNet50 Fine-Tuned memiliki arsitektur yang identik dengan ResNet50 Frozen (~12,80 ms/citra) sehingga tidak ada tambahan biaya inferensi akibat proses Fine-Tuning. ResNet50 Fine-Tuned direkomendasikan dalam skenario di mana akurasi tinggi dibutuhkan namun sumber daya komputasi memadai dan efisiensi inferensi bukan prioritas utama, atau

ketika ketersediaan data terbatas namun arsitektur ResNet50 sudah tersedia sebagai backbone yang sudah terlatih.