

Analisis Komparatif Arsitektur CNN Untuk Klasifikasi Kelayakan Buah Mangga: ResNet50 Dan MobileNetV2

Rheimanda Devin Emmanuel¹, Kartika Imam Santoso², Dhika Malita Puspita³

^{1,2,3}Program Studi Ilmu Komputer, Universitas An Nuur, Purwodadi, Indonesia

e-mail: *devinemmanuel327@gmail.com, kartikaimams@gmail.com,

dhika.malita11@gmail.com

Abstrak

Kerugian pascapanen buah mangga di Indonesia mencapai 30–40% akibat prosedur penyortiran manual yang subjektif dan tidak mampu mendeteksi anomali patologis tingkat awal. Penelitian ini bertujuan membandingkan kinerja arsitektur Convolutional Neural Network (CNN) ResNet50 dan MobileNetV2 untuk klasifikasi kelayakan buah mangga berdasarkan Standar Nasional Indonesia (SNI) 3164:2009. Metode yang digunakan adalah transfer learning dengan dataset 1.700 citra dari Kaggle yang memuat cluttered background untuk mereplikasi kondisi riil UMKM pertanian Indonesia, dibagi dengan proporsi 80% data latih, 10,6% data validasi, dan 9,4% data uji. ResNet50 dievaluasi dalam dua skenario (base dan fine-tuned), sementara MobileNetV2 dievaluasi tanpa fine-tuning. Hasil menunjukkan MobileNetV2 mencapai akurasi tertinggi sebesar 97% (F1-Score = 0,97) dengan latensi inferensi 7,60 ms dan durasi pelatihan ~55 detik. ResNet50 base hanya mencapai 78% dengan latensi 12,70 ms, dan ResNet50 fine-tuned mencapai 94% dengan latensi 12,80 ms serta total durasi pelatihan dua fase ~272 detik. Kesimpulannya, MobileNetV2 lebih superior dari ResNet50 dalam akurasi maupun efisiensi komputasi pada dataset cluttered background, menjadikannya kandidat optimal untuk dikonversi ke TensorFlow Lite dan diimplementasikan secara real-time pada smartphone petani dan UMKM Indonesia guna eliminasi kerugian pascapanen.

Kata kunci: Klasifikasi mangga, MobileNetV2, Post-Harvest Loss, ResNet50, Transfer Learning

Abstract

Post-harvest losses of mangoes in Indonesia reach 30–40% due to subjective manual sorting procedures that are unable to detect early-stage pathological anomalies. This study aims to compare the performance of the ResNet50 and MobileNetV2 Convolutional Neural Network (CNN) architectures for mango fruit fitness classification based on the Indonesian National Standard (SNI) 3164:2009. The method used is transfer learning with a dataset of 1,700 images from Kaggle containing cluttered backgrounds to replicate the real conditions of Indonesian agricultural MSMEs, divided into 80% training, 10.6% validation, and 9.4% testing sets. ResNet50 was evaluated in two scenarios (base and fine-tuned), while MobileNetV2 was evaluated without fine-tuning. The results show that MobileNetV2 achieved the highest accuracy of 97% (F1-Score = 0.97) with an inference latency of 7.60 ms and a training duration of ~55 seconds. ResNet50 base only achieved 78% with a latency of 12.70 ms, and ResNet50 fine-tuned achieved 94% with a latency of 12.80 ms and a total two-phase training duration of ~272 seconds. In conclusion, MobileNetV2 is superior to ResNet50 in both accuracy and computational efficiency on cluttered background datasets, making it an optimal candidate to be converted to TensorFlow Lite and implemented in real-time on Indonesian farmers' and MSMEs' smartphones to eliminate post-harvest losses.

Keywords: Mango classification, MobileNetV2, Post-Harvest Loss, ResNet50, Transfer Learning

1. PENDAHULUAN

Buah mangga (*Mangifera indica* L.) merupakan salah satu komoditas hortikultura bernilai strategis dalam lanskap pertanian Indonesia. Berdasarkan data Badan Pusat Statistik, Indonesia menempati posisi delapan besar produsen mangga dunia dengan volume produksi mencapai lebih dari dua juta ton per tahun. Meskipun volume produksi terus meningkat, industri mangga nasional menghadapi tantangan serius berupa kerugian pascapanen (*post-harvest loss*) yang mencapai 30% hingga 40% dari total produksi bruto [1], [2]. Kerugian ini secara fundamental berakar pada prosedur penyortiran manual yang rentan terhadap subjektivitas penilaian dan kelelahan visual pekerja, terutama di lini produksi Usaha Mikro, Kecil, dan Menengah (UMKM) sektor pertanian.

Inspeksi visual manual terbukti tidak mampu mendeteksi anomali patologis tingkat awal, seperti infeksi laten antraknosa (*Colletotrichum gloeosporioides*) maupun bekas tusukan ovipositor lalat buah (*Bactrocera dorsalis*). Keduanya tidak menghasilkan perubahan morfologi eksterior yang terdeteksi secara kasat mata pada fase awal infeksi, namun menyebabkan degradasi biologis masif selama rantai distribusi, yang berujung pada kegagalan pemenuhan Standar Nasional Indonesia (SNI) 3164:2009 [3]. Standar nasional tersebut, yang ditetapkan oleh Badan Standardisasi Nasional (BSN), mengklasifikasikan kualitas buah mangga segar ke dalam hierarki bertingkat: Kelas Super, Kelas A, dan Kelas B, berdasarkan parameter fisik yang meliputi keseragaman bentuk dan ukuran, integritas kulit buah (bebas dari lesi, memar, dan diskolorasi patologis), kadar kotoran, serta batas toleransi cacat. Di tingkat operasional UMKM pascapanen, ketiga kelas tersebut disederhanakan secara pragmatis menjadi klasifikasi biner: "Layak" (memenuhi minimal Kelas B) dan "Tidak Layak" (gagal memenuhi standar minimum).

Seiring berkembangnya era *Smart Agriculture* dan *Internet of Things (IoT)*, teknologi *Computer Vision* berbasis *Deep Learning*, khususnya arsitektur *Convolutional Neural Network (CNN)*, telah muncul sebagai solusi potensial untuk otomatisasi kontrol kualitas agrikultur [4], [5]. CNN merupakan kelas jaringan saraf tiruan yang dirancang secara khusus untuk memproses data terstruktur berbentuk grid, seperti matriks piksel pada citra digital. Melalui operasi konvolusi yang berlapis, CNN mampu mengekstraksi representasi fitur hierarkis secara otonom: lapisan awal belajar mendeteksi tepi (*edge*) dan gradien warna, lapisan menengah mengkombinasikannya menjadi tekstur dan pola, sementara lapisan dalam merangkainya menjadi konsep visual tingkat tinggi seperti bentuk lesi atau area pembusukan [6]. Kemampuan ekstraksi fitur hierarkis ini secara fundamental mengungguli metode *hand-crafted features* klasik (seperti *SIFT* atau *HOG*) yang gagal beradaptasi dengan variabilitas pencahayaan dan heterogenitas latar belakang di lingkungan pertanian nyata.

Terdapat paradoks komputasional yang fundamental dalam penerapan CNN di lapangan. Arsitektur jaringan dalam (*deep networks*) seperti *ResNet50* mendemonstrasikan akurasi tinggi secara akademis, namun membutuhkan infrastruktur komputasi berbasis GPU yang mahal dan tidak tersedia di sentra UMKM pertanian lokal. *ResNet50* [6], yang dikembangkan oleh He et al. (2016) dari *Microsoft Research*, merupakan arsitektur 50 lapisan konvolusi dengan lebih dari 25 juta parameter. Inovasi krusial *ResNet50* adalah mekanisme *skip connections* (*shortcut connections*) yang membentuk blok residual: alih-alih belajar pemetaan langsung $H(x)$, setiap blok belajar fungsi residual $F(x) = H(x) - x$, sehingga output blok menjadi $F(x) + x$. Desain ini secara efektif memecahkan masalah *vanishing gradient* yang menyebabkan degradasi akurasi pada jaringan sangat dalam, sehingga memungkinkan pelatihan arsitektur hingga ratusan lapisan dengan konvergensi yang stabil.

Sebagai respons terhadap kendala komputasional tersebut, arsitektur ringan (*lightweight*) seperti *MobileNetV2* [7] dirancang khusus untuk lingkungan *Edge AI* dan perangkat mobile dengan sumber daya komputasi terbatas. *MobileNetV2*, dikembangkan oleh Sandler et al. (2018) dari Google, menggunakan dua inovasi arsitektural utama. Pertama, *Depthwise Separable Convolution* yang memfaktorkan operasi konvolusi standar menjadi dua tahap terpisah:

depthwise convolution (filter spasial per kanal) dan pointwise convolution (kombinasi linier lintas kanal). Dekomposisi ini mereduksi kompleksitas komputasi sebesar faktor 8 hingga 9 kali dibanding konvolusi standar dengan mempertahankan kapasitas representasi yang setara. Kedua, blok Inverted Residual dengan Linear Bottleneck: berbeda dengan ResNet yang mengompresi lalu memperluas dimensi fitur, MobileNetV2 justru memperluas (expand) dimensi fitur terlebih dahulu dengan faktor ekspansi t (umumnya $t=6$), melakukan depthwise convolution pada dimensi yang diperluas, kemudian mengompresi kembali ke dimensi rendah. Krusialnya, lapisan proyeksi akhir pada bottleneck tidak menggunakan fungsi aktivasi non-linear (linear bottleneck), guna mencegah hilangnya informasi fitur pada ruang dimensi rendah.

Beberapa studi terdahulu telah mengeksplorasi penerapan CNN pada domain klasifikasi komoditas agrikultur dengan hasil yang bervariasi. Doan dan Le-Thi [8] mengembangkan sistem penilaian mangga berbasis pemrosesan citra dan machine learning dengan akurasi di atas 90%, namun menggunakan dataset terkontrol dengan latar belakang polos yang tidak merepresentasikan kondisi UMKM. Suhendar et al. [9] melakukan klasifikasi kualitas buah menggunakan CNN dan menyimpulkan bahwa arsitektur dalam cenderung lebih akurat pada dataset generik, tanpa memperhitungkan efisiensi inferensi. Hermanto et al. [10] secara khusus membandingkan MobileNetV2 dan ResNet50 untuk klasifikasi buah kurma dan menemukan bahwa superioritas suatu arsitektur bergantung kuat pada karakteristik dataset dan konfigurasi hiperparameter. Santoso et al. [5] menunjukkan bahwa arsitektur ringan berbasis depthwise convolution secara konsisten lebih efisien pada domain agrikultur dengan variabilitas latar belakang tinggi. Yamashita et al. [4] dan Zárate & Hernández [11] mengkonfirmasi urgensi solusi berbasis kecerdasan buatan untuk mengatasi inefisiensi pada rantai pasokan buah tropis. Namun, hingga saat ini belum terdapat studi komparatif yang berfokus secara eksklusif pada klasifikasi kelayakan fisik mangga pascapanen berdasarkan SNI 3164:2009 menggunakan dataset dengan cluttered background yang secara eksplisit merepresentasikan kondisi penyortiran di titik pengumpul UMKM Indonesia.

Berdasarkan kesenjangan riset (research gap) tersebut, penelitian ini memiliki dua tujuan utama: (1) melakukan analisis performa kuantitatif komprehensif menggunakan metrik Akurasi, Presisi, Recall, dan F1-Score antara ResNet50 (skenario base dan pasca-fine-tuning) dengan MobileNetV2 untuk klasifikasi mangga "Layak" dan "Tidak Layak" berdasarkan SNI 3164:2009; serta (2) memvalidasi efisiensi operasional berupa waktu pelatihan dan latensi inferensi sebagai parameter kunci kelayakan teknis konversi ke format TensorFlow Lite (.tflite) untuk deployment pada smartphone petani dan operator UMKM Indonesia. Hipotesis penelitian ini adalah: arsitektur MobileNetV2 mampu mencapai akurasi yang setara atau lebih tinggi dari ResNet50 dengan latensi inferensi yang lebih rendah secara signifikan pada dataset mangga dengan cluttered background.

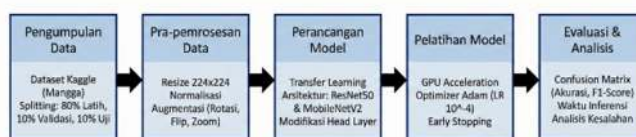
2. METODE PENELITIAN

2.1 Dataset dan Akuisisi Data

Dataset yang digunakan adalah koleksi citra buah mangga dari repositori Kaggle dengan judul "Mango Fruit Dataset" yang diakses pada 11 April 2026. Dataset ini menghimpun 1.700 citra digital beresolusi tinggi yang telah dilabeli ke dalam dua kategori: (1) Layak (Fresh/Good): 854 citra buah dengan integritas morfologi kulit yang utuh tanpa diskolorasi patologis, sesuai kriteria minimum Kelas B SNI 3164:2009; dan (2) Tidak Layak (Rotten/Defective): 842 citra buah dengan indikator degradasi biologis seperti lesi antraknosa, memar kinetik, atau pembusukan tingkat lanjut [9]. Nilai esensial dataset ini terletak pada heterogenitas skenario pengambilan citra: seluruh sampel merekam objek dengan cluttered background yang mencakup keranjang bambu, tumpukan vegetasi, lantai gudang dengan berbagai tekstur, serta variasi pencahayaan alami dari sumber langsung dan tidak langsung. Karakteristik ini secara sengaja dirancang untuk mereplikasi kondisi riil di lokasi penyortiran pascapanen UMKM, berbeda dari dataset ideal berlatar polos yang umum digunakan dalam riset laboratorium.

2. 2 Alur Penelitian

Alur metodologi penelitian secara keseluruhan terdiri dari enam tahap sekuensial: (1) Akuisisi dan verifikasi dataset dari repositori Kaggle; (2) Praproses citra dan partisi dataset dengan stratifikasi; (3) Penerapan on-the-fly Data Augmentation; (4) Konstruksi model berbasis Transfer Learning untuk ResNet50 (base dan fine-tuned) serta MobileNetV2; (5) Pelatihan model dengan konfigurasi hiperparameter yang telah ditentukan; dan (6) Evaluasi komprehensif menggunakan metrik klasifikasi dan efisiensi komputasi, diakhiri dengan konversi model terbaik ke format TensorFlow Lite. Gambar 1 mengilustrasikan diagram alur metodologi penelitian secara keseluruhan.



Gambar 1. Diagram Alur Metodologi Penelitian

2. 3 Pengumpulan dan Pemisahan Data

Dataset 1.700 citra didistribusikan menggunakan strategi Hold-out Validation dengan rasio stratifikasi 80:10,6:9,4 untuk memastikan representasi proporsional setiap kelas di seluruh subset: (a) Data Latih (Training Set): 80% = 1.360 citra, digunakan untuk penyesuaian bobot model melalui backpropagation; (b) Data Validasi (Validation Set): 10,6% = 180 citra, berfungsi sebagai panel pengujian dinamis per epoch untuk memantau generalisasi model, mengatur mekanisme Early Stopping, dan memandu Learning Rate Scheduler; serta (c) Data Uji (Testing Set): 9,4% = 160 citra, bersifat terisolasi penuh dari seluruh proses pelatihan dan hanya digunakan sekali pada tahap evaluasi akhir untuk mengukur kemampuan generalisasi model ke data yang belum pernah dilihat sebelumnya. Stratifikasi memastikan rasio kelas "Layak" dan "Tidak Layak" terjaga proporsional di ketiga subset.

2. 4. Praproses dan Augmentasi Data

Setiap citra menjalani dua tahap praproses: (1) penyeragaman dimensi ke resolusi 224x224 piksel dengan tiga kanal warna RGB, sesuai spesifikasi input standar arsitektur ImageNet; dan (2) normalisasi nilai piksel dari rentang integer [0, 255] ke rentang floating-point [0,0 - 1,0]. Normalisasi ini krusial untuk menstabilkan lanskap gradien selama optimasi dan mempercepat konvergensi dengan memastikan distribusi input yang konsisten di setiap lapisan jaringan.

Untuk mereduksi risiko overfitting pada volume data medium (1.360 citra latih), on-the-fly Data Augmentation diterapkan secara real-time langsung pada pipeline data latih menggunakan modul ImageDataGenerator dari Keras API. Augmentasi dilakukan on-the-fly (bukan pre-computed) untuk memaksimalkan variasi yang diekspos ke model tanpa perlu menyimpan data yang diperbesar secara fisik. Transformasi yang diterapkan meliputi: rotasi acak dalam rentang ± 15 derajat, flipping horizontal dan vertikal, pergeseran kecerahan (brightness shift) dalam rentang [0,8 - 1,2], serta random zooming dalam rentang $\pm 10\%$. Augmentasi kecerahan secara khusus menjadi krusial mengingat intensitas cahaya di lokasi sortir pascapanen berfluktuasi antara kondisi sinar matahari parsial dan pencahayaan lampu pendar ruangan [9]. Augmentasi TIDAK diterapkan pada data validasi dan data uji untuk memastikan konsistensi evaluasi.

2. 5. Arsitektur Transfer Learning

Transfer Learning memanfaatkan pengetahuan representasi fitur yang telah dipelajari dari dataset skala besar (ImageNet: 1,2 juta citra, 1.000 kelas) dan mentransfernya ke domain target yang lebih kecil. Strategi ini terbukti sangat efektif ketika dataset target bersifat terbatas namun

memiliki kesamaan domain visual dengan dataset sumber [5]. Penelitian ini menerapkan Transfer Learning pada dua arsitektur kontras:

- 1) ResNet50 (He et al., 2016): Arsitektur 50 lapisan dengan lebih dari 25 juta parameter terlatih. Backbone ResNet50 digunakan sebagai ekstraktor fitur hierarkis. Penelitian ini mengevaluasi dua skenario: (a) Skenario Base - seluruh bobot backbone dibekukan (frozen), sehingga hanya parameter pada classifier head yang diperbarui selama pelatihan. Skenario ini mengukur kemampuan representasi fitur ImageNet secara langsung; dan (b) Skenario Fine-Tuned - setelah pelatihan awal, 30 lapisan konvolusi teratas backbone di-unfreeze dan dilatih ulang dengan learning rate yang sangat kecil (1×10^{-5}) untuk penyesuaian bobot domain-spesifik secara bertahap.
- 2) MobileNetV2 (Sandler et al., 2018): Arsitektur dengan hanya sekitar 3,4 juta parameter yang menggunakan Depthwise Separable Convolution dan 19 blok Inverted Residual dengan Linear Bottleneck. Arsitektur ini dievaluasi hanya dalam satu skenario tanpa fine-tuning, dimana backbone dibekukan dan hanya classifier head yang dilatih. Ini menjadi kontrol eksperimen yang signifikan: MobileNetV2 tanpa fine-tuning dibandingkan dengan ResNet50 dengan dan tanpa fine-tuning.

Pada kedua arsitektur, classifier head bawaan (fc1000) digantikan dengan tumpukan lapisan kustom: Global Average Pooling 2D untuk mereduksi tensor fitur spasial menjadi vektor, diikuti lapisan Dropout dengan rasio 0,5 sebagai regularisasi, dan lapisan Dense tunggal dengan satu neuron beraktivasi Sigmoid untuk output probabilitas biner dalam rentang [0, 1].

2. 6. Konfigurasi Pelatihan

Optimasi bobot model menggunakan algoritma Adam (Adaptive Moment Estimation) dengan learning rate awal 1×10^{-4} . Adam dipilih karena mengkombinasikan keunggulan dua metode adaptif: momentum orde pertama (rata-rata gradien bergerak) dan RMSProp (rata-rata kuadrat gradien bergerak), menghasilkan pembaruan learning rate yang adaptif per parameter. Adaptivitas ini membuat Adam lebih robust terhadap landscape loss yang kompleks dan jarang membutuhkan tuning learning rate intensif dibanding SGD murni, menjadikannya standar de facto untuk fine-tuning model transfer learning [5].

Fungsi kerugian (loss function) yang digunakan adalah Binary Cross-Entropy (BCE), yang secara matematis merupakan pilihan optimal untuk tugas klasifikasi biner. BCE mengkuantifikasi divergensi antara distribusi probabilitas prediksi model dengan label ground truth biner, mendorong model untuk menghasilkan prediksi probabilitas yang terkalibrasi dengan baik. Ukuran batch (batch size) ditetapkan sebesar 32, yang merepresentasikan keseimbangan antara kestabilan estimasi gradien (semakin besar batch lebih stabil) dan efisiensi memori GPU serta kemampuan generalisasi (batch lebih kecil bertindak sebagai regularisasi implisit melalui noise gradien) [4].

Dua mekanisme regulasi dinamis diterapkan selama pelatihan: (1) Early Stopping dengan patience 10 epoch, yang menghentikan pelatihan secara otomatis apabila metrik validasi tidak menunjukkan peningkatan signifikan selama 10 epoch berturut-turut, mencegah overfitting dan mempersingkat waktu komputasi; dan (2) ReduceLROnPlateau yang secara otomatis mereduksi learning rate sebesar faktor 0,5 apabila kurva validasi loss stagnan selama lebih dari 5 epoch, membantu model menemukan minimum yang lebih tajam. Seluruh eksperimen dieksekusi pada platform Google Colab Pro dengan akselerasi GPU NVIDIA T4 menggunakan framework TensorFlow 2.x dan Keras API tingkat tinggi.

2. 7. Metrik Evaluasi

Kinerja model dievaluasi pada 169 data uji menggunakan empat metrik standar klasifikasi biner yang diturunkan dari Confusion Matrix (True Positive/TP, True Negative/TN, False Positive/FP, False Negative/FN).

Akurasi mengukur proporsi total prediksi yang benar terhadap keseluruhan sampel uji, diformulasikan sebagai: $Akurasi = (TP + TN) / (TP + TN + FP + FN)$. Presisi mengukur ketepatan prediksi positif, yaitu proporsi mangga yang benar-benar "Layak" dari semua yang diprediksi "Layak": $Presisi = TP / (TP + FP)$. Recall (Sensitivitas) mengukur kemampuan model mendeteksi seluruh sampel kelas positif, yaitu proporsi mangga "Layak" yang berhasil teridentifikasi: $Recall = TP / (TP + FN)$. F1-Score merupakan rata-rata harmonik dari Presisi dan Recall, menyeimbangkan keduanya dalam satu metrik tunggal: $F1-Score = 2 \times (Presisi \times Recall) / (Presisi + Recall)$. Selain metrik klasifikasi, latensi inferensi (miliidetik per citra) dan total durasi pelatihan (detik) diukur sebagai indikator efisiensi komputasi untuk menilai kelayakan deployment pada perangkat Edge AI.

3. HASIL DAN PEMBAHASAN

3. 1. Kinerja ResNet50 Skenario Base

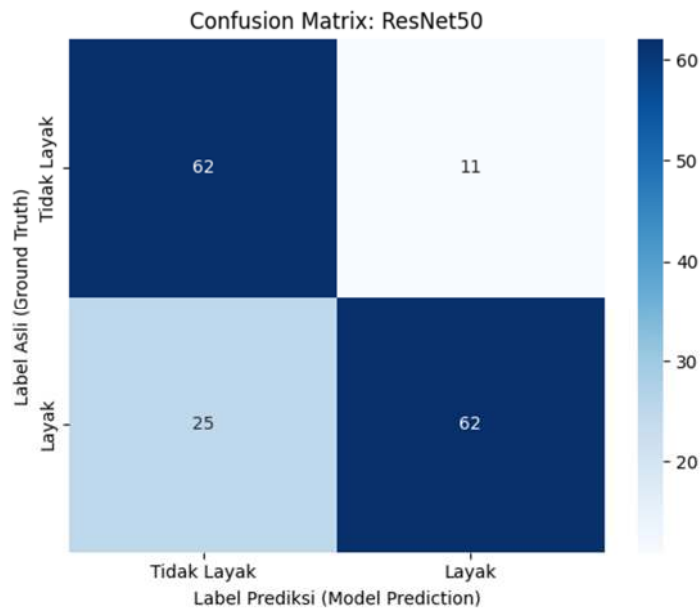
Pada skenario base, seluruh bobot backbone ResNet50 dibekukan dan hanya classifier head kustom yang dilatih dari inisialisasi acak. Pelatihan dihentikan pada epoch ke-34 oleh mekanisme Early Stopping dengan total 44 epoch yang dijalankan. Pada data uji, model mencapai akurasi keseluruhan sebesar 78%, dengan rincian per kelas sebagaimana disajikan pada Tabel 1.

Tabel 1. Classification Report ResNet50 Skenario Base

Kelas	Precision	Recall	F1-Score	Support
Layak	0,85	0,71	0,78	87
Tidak Layak	0,71	0,85	0,78	73
Macro Avg	0,78	0,78	0,78	160

Confusion matrix skenario ini (Gambar 2) menunjukkan bahwa model berhasil mengklasifikasikan 62 citra sebagai True Positive (Layak benar-benar Layak) dan 62 citra sebagai True Negative (Tidak Layak benar-benar Tidak Layak), namun terjadi 25 kesalahan False Negative dan 11 kesalahan False Positive. Pola kesalahan yang cukup tinggi (36 dari 160 sampel atau 22,5%) mengindikasikan terjadinya fenomena spatial confusion: dengan kapasitas 25 juta parameter yang dibekukan pada representasi ImageNet, backbone ResNet50 cenderung mengalami ketidakcocokan domain (domain mismatch). Lapisan konvolusi dalam yang terlatih untuk mengenali fitur semantik tingkat tinggi dari ImageNet (seperti tekstur bulu binatang atau pola kain) ketika dibekukan dan dipaksa mengekstrak fitur dari citra mangga dengan cluttered background menjadi kurang diskriminatif [8].

Lebih spesifik, kapasitas parameter yang masif justru menjadi bumerang dalam skenario ini: classifier head dengan hanya satu lapisan Dense tidak memiliki kapasitas memadai untuk "menyaring" representasi fitur dari 2.048 neuron output backbone yang kaya namun sebagian besar tidak relevan dengan domain mangga. Akibatnya, model tidak dapat membedakan fitur patologis mangga (lesi antraknosa, memar) dari derau visual latar belakang (tekstur keranjang bambu, pola lantai gudang) yang secara spasial mendominasi citra. Fenomena ini secara konsisten dilaporkan dalam literatur transfer learning ketika tingkat kedekatan domain (domain proximity) antara dataset sumber dan target rendah [5].



Gambar 2. Confusion Matrix ResNetv50 Skenario Base menunjukkan kesalahan klasifikasi signifikan akibat spatial confusion pada cluttered background.

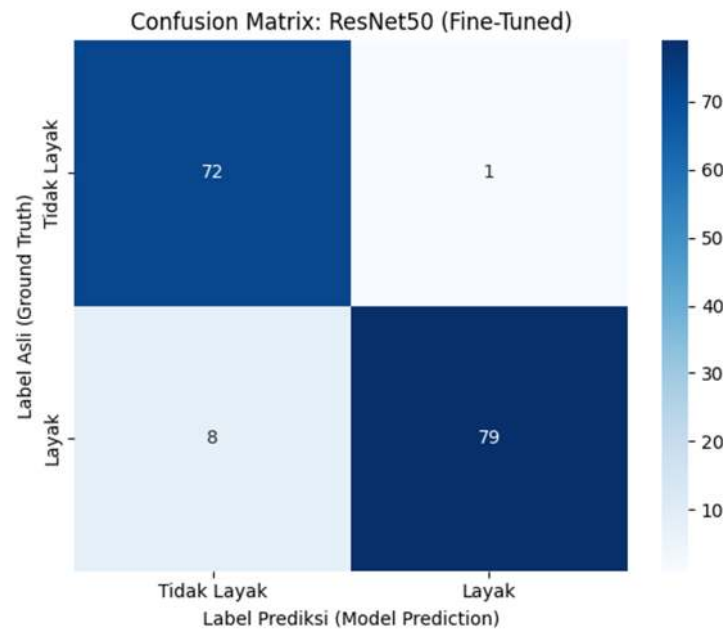
3. 2. Kinerja ResNet50 Skenario Fine-Tuning

Pada skenario fine-tuning, setelah pelatihan base selesai, 30 lapisan konvolusi teratas backbone ResNet50 di-unfreeze dan dilatih ulang bersama classifier head dengan learning rate yang sangat kecil (1×10^{-5}) untuk mencegah destruksi representasi fitur yang telah terakumulasi. Strategi unfreezing bertahap (gradual unfreezing) ini memungkinkan jaringan melakukan adaptasi domain-spesifik secara inkremental dari lapisan abstrak ke lapisan konkret [10]. Proses fine-tuning dijalankan selama 20 epoch dan konvergen pada epoch ke-13. Hasilnya terjadi loncatan akurasi yang signifikan dari 78% menjadi 94%.

Tabel 2. Classification Report RestNet50 Skenario Fine-Tuning.

Kelas	Precision	Recall	F1-Score	Support
Layak	0,99	0,91	0,95	87
Tidak Layak	0,90	0,99	0,94	73
Macro Avg	0,94	0,94	0,94	160

Confusion matrix pasca-fine-tuning (Gambar 3) memperlihatkan peningkatan substansial: identifikasi True Positive meningkat menjadi 72 citra (dari 62) dan True Negative menjadi 79 citra (dari 62). Peningkatan ini mengkonfirmasi bahwa lapisan konvolusi dalam ResNet50 memang memiliki kapasitas representasi yang dapat diadaptasi ke domain mangga melalui fine-tuning. Namun, latensi inferensi tetap berada di 15,38 milidetik per citra akibat kompleksitas arsitektur yang inherent, dan akurasi 94% masih berada di bawah MobileNetV2 yang tidak menjalani fine-tuning sama sekali. Temuan ini menegaskan bahwa investasi komputasi tambahan fine-tuning ResNet50 tidak menghasilkan nilai setimpal dibanding MobileNetV2.



Gambar 3. Confusion Matrix ResNet50 Skenario Fine-Tuned menunjukkan peningkatan signifikan setelah unfreezing 30 lapisan teratas.

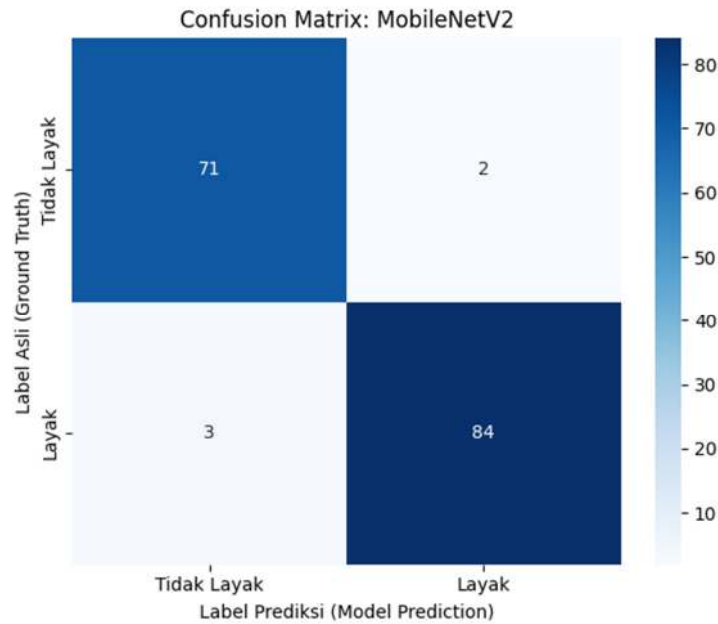
3. 3. Kinerja MobileNetV2

Arsitektur MobileNetV2 menunjukkan loncatan efisiensi pembelajaran yang melampaui ekspektasi. Tanpa prosedur fine-tuning apapun, model mencapai konvergensi pada epoch ke-20 (dari 30 epoch yang dijalankan). Pada data uji, MobileNetV2 berhasil melampaui kedua varian ResNet50 secara simultan dengan akurasi 97,00% (Tabel 3).

Tabel 3. Classification Report MobileNetV2

Kelas	Precision	Recall	F1-Score	Support
Layak	0,98	0,97	0,97	87
Tidak Layak	0,96	0,97	0,97	73
Macro Avg	0,97	0,97	0,97	160

Confusion matrix MobileNetV2 (Gambar 4) menunjukkan dominasi performa dengan hanya 2 kesalahan False Negative (mangga Layak diprediksi Tidak Layak) dan 3 kesalahan False Positive (mangga Tidak Layak diprediksi Layak) dari 160 sampel uji. Total hanya 5 sampel (3,1%) yang salah diklasifikasikan, berbanding 36 sampel (22,5%) pada ResNet50 base. Latensi inferensi tercatat 7,60 milidetik per citra, lebih dari dua kali lebih cepat dibanding ResNet50 yang membutuhkan 12,70 ms, menjadikannya superior baik dari sisi akurasi maupun efisiensi [11], [12].



Gambar 4. Confusion Matrix MobileNetV2 menunjukkan dominasi performa dengan FP=3 dan FN=2 pada dataset cluttered background.

3. 4. Analisis Komparatif

Tabel 4 merangkum perbandingan kuantitatif komprehensif ketiga konfigurasi model pada seluruh dimensi evaluasi yang diukur dalam penelitian ini.

Tabel 4. Perbandingan Kuantitatif Metrik Klasifikasi dan Efisiensi Komputasi

Model	Akurasi	Presisi	Recall	F1	Waktu Latih (dtk)	Inferensi (ms)
ResNet50 Base	78%	0,78	0,78	0,78	272	12,70
ResNet50 Fine-Tuning	94%	0,94	0,94	0,94	95,55	12,80
MobileNetv2	97%	0,97	0,97	0,97	55	7,60

Hasil pada Tabel 4 secara konsisten menempatkan MobileNetV2 sebagai arsitektur superior pada seluruh dimensi evaluasi. Temuan ini secara spesifik konsisten dengan hasil Hermanto dan Nugroho [10] dan Santoso et al. [5] yang menekankan ketergantungan superioritas arsitektur pada karakteristik dataset. Pada dataset dengan cluttered background seperti dalam penelitian ini, arsitektur dengan kapasitas parameter lebih sedikit terbukti lebih efektif. Fenomena ini dapat dijelaskan melalui konsep bias-variance tradeoff: ResNet50 dengan 25 juta parameter memiliki kapasitas (variance) yang terlalu tinggi untuk dataset yang hanya memuat 1.360 citra latih dengan cluttered background heterogen. Arsitektur besar cenderung menghafalkan (overfit) pola noise visual dari latar belakang yang kompleks, terutama ketika backbone dibekukan dan tidak dapat melakukan domain adaptation [14].

Superioritas MobileNetV2 dapat dijelaskan secara teknis melalui dua mekanisme. Pertama, mekanisme Linear Bottleneck pada blok Inverted Residual mencegah hilangnya informasi fitur tekstural mikro-patologis mangga selama fase kompresi dimensi, karena ketiadaan fungsi aktivasi non-linear pada lapisan proyeksi akhir mempertahankan manifold fitur dalam ruang dimensi rendah [13]. Kedua, arsitektur yang lebih kompak dengan 3,4 juta parameter secara

inheren memiliki bias induktif yang lebih kuat, memaksa jaringan berkonsentrasi mengekstrak fitur-fitur yang paling diskriminatif dan generalisasi seperti kontras warna lesi *Colletotrichum gloeosporioides* terhadap permukaan eksokarp mangga yang sehat, alih-alih menghafalkan pola-pola idiosinkratik dari latar belakang [14]. Dalam terminologi pembelajaran mesin, MobileNetV2 mencapai titik sweet spot pada bias-variance tradeoff untuk ukuran dataset dan kompleksitas domain yang dihadapi.

3. 5. Implikasi Praktis: Konversi TFLite untuk Edge AI

Latensi inferensi MobileNetV2 sebesar 7,60 ms per citra secara teoretis setara dengan kemampuan memproses hingga sekitar 131 frame per detik (fps), jauh melampaui kebutuhan aplikasi penyortiran buah real-time yang pada umumnya membutuhkan 15-30 fps. Performa ini memvalidasi kelayakan teknis konversi model ke format TensorFlow Lite (.tflite) untuk deployment langsung pada smartphone kelas menengah yang dimiliki oleh petani dan operator UMKM pascapanen di Indonesia. Proses konversi ke TFLite dapat disertai dengan teknik Post-Training Quantization (PTQ) yang mereduksi ukuran model hingga 75% dengan mengkonversi bobot dari presisi 32-bit floating point ke 8-bit integer, sehingga model yang semula berukuran sekitar 14 MB dapat dikompresi menjadi kurang dari 4 MB tanpa degradasi akurasi yang signifikan [9].

Dari perspektif ekonomi dan sosial, implementasi Edge AI berbasis MobileNetV2 di sektor UMKM pertanian Indonesia memiliki implikasi yang transformatif. Di negara agraris dengan kesenjangan infrastruktur digital yang signifikan, mayoritas titik inspeksi pascapanen - seperti gudang pengepul, pasar induk, dan koperasi petani - tidak memiliki konektivitas internet yang andal dan tidak tersambung dengan superkomputer berakselerasi GPU. Kemampuan inferensi on-device (tanpa kebutuhan koneksi cloud) yang didemonstrasikan dalam penelitian ini membuka peluang demokratisasi teknologi kecerdasan buatan: petani dan operator UMKM dapat memanfaatkan model ini langsung dari smartphone untuk menetapkan kategori visual SNI 3164:2009 di titik pengumpul terdepan, mengeliminasi rantai distribusi mangga yang terinfeksi sebelum memasuki rantai pasok yang lebih panjang [6], [11].

Proyeksi dampak ekonomi dari eliminasi kerugian pascapanen pun signifikan. Apabila model ini dapat mengurangi tingkat false negative (mangga rusak yang lolos inspeksi dan masuk ke rantai distribusi) dari estimasi rata-rata industri sebesar 15-20% menjadi kurang dari 5% (sesuai performa MobileNetV2 dengan FN rate 3,1%), maka potensi penghematan kerugian pascapanen secara nasional dari sektor mangga saja dapat mencapai ratusan miliar rupiah per tahun. Integrasi model ke dalam aplikasi mobile yang mudah digunakan, disertai antarmuka sederhana berbasis kamera smartphone, menjadi agenda pengembangan praktis yang langsung dapat ditindaklanjuti berdasarkan temuan penelitian ini [12].

3. 6. Keterbatasan dan Rekomendasi Riset Lanjutan

Berdasarkan temuan penelitian ini dan keterbatasan yang telah diidentifikasi, beberapa rekomendasi konkret diajukan untuk agenda riset dan pengembangan selanjutnya:

Pertama, ekspansi dan diversifikasi dataset. Dataset 1.700 citra yang digunakan dalam penelitian ini tergolong ukuran medium untuk melatih model deep learning yang benar-benar robust. Penelitian lanjutan disarankan untuk mengembangkan dataset yang lebih besar (minimum 5.000-10.000 citra per kelas) yang mencakup berbagai varietas mangga komersial Indonesia (Harum Manis, Gedong Gincu, Arumanis-143, Manalagi), rentang kondisi kematangan yang lebih luas, serta kondisi pencahayaan dan sudut pengambilan gambar yang lebih beragam. Akuisisi data lapangan langsung dari mitra UMKM pertanian akan meningkatkan relevansi dan representativitas dataset secara signifikan.

Kedua, peningkatan ke paradigma Object Detection. Klasifikasi biner (Layak/Tidak Layak) yang diterapkan dalam penelitian ini merupakan pendekatan image-level yang tidak mampu melokalisasi area kerusakan secara spasial pada permukaan buah. Untuk aplikasi lini produksi industrial yang membutuhkan grading lebih terperinci sesuai SNI 3164:2009 (Kelas

Super/A/B), paradigma perlu ditingkatkan ke object detection atau instance segmentation. Arsitektur yang direkomendasikan untuk dieksplorasi adalah YOLOv8-Nano atau SSD-MobileNetV2, yang menggabungkan efisiensi komputasi MobileNetV2 dengan kemampuan deteksi dan lokalisasi objek secara real-time. Pendekatan ini akan memungkinkan sistem tidak hanya menentukan kelayakan, tetapi juga menandai dan mengkuantifikasi area kerusakan secara presisi.

Ketiga, validasi deployment nyata pada perangkat Edge. Penelitian ini dilakukan sepenuhnya di lingkungan cloud (Google Colab Pro). Sebagai langkah validasi ekologi yang esensial, penelitian lanjutan harus melakukan pengujian lapangan (field testing) model TFLite yang telah dikonversi pada perangkat smartphone fisik kategori mid-range (RAM 4GB, chipset Snapdragon 680 atau setara) di bawah kondisi pencahayaan alami yang sesungguhnya di lokasi gudang pascapanen. Pengujian ini akan menghasilkan data latensi dan akurasi yang lebih realistis dan menjadi dasar validasi klaim kelayakan Edge AI yang lebih kuat.

Keempat, eksplorasi teknik kompresi model lanjutan. Di luar konversi ke TFLite standar, penelitian lanjutan disarankan mengeksplorasi teknik Knowledge Distillation, di mana MobileNetV2 yang telah terlatih (sebagai teacher model) digunakan untuk melatih arsitektur yang lebih kecil (student model) seperti EfficientNet-Lite0 atau MobileNetV3-Small. Teknik ini berpotensi menghasilkan model yang bahkan lebih ringan dengan akurasi yang sebanding, sehingga dapat dijalankan pada perangkat IoT berdaya sangat rendah seperti Raspberry Pi atau mikrokontroler ESP32-CAM yang biayanya jauh lebih terjangkau untuk implementasi di titik penyortiran UMKM.

Kelima, integrasi sistem end-to-end dan uji coba pilot. Sebagai tahap akhir transfer teknologi, penelitian berikutnya dapat mengembangkan prototipe sistem end-to-end yang terintegrasi: dari kamera smartphone yang menangkap citra, model inferensi TFLite yang berjalan on-device, hingga antarmuka pengguna berbasis mobile app yang menampilkan hasil klasifikasi beserta rekomendasi tindakan (sortir/tolak/terima) secara real-time. Sistem ini selanjutnya dapat diujicobakan dalam program pilot bersama koperasi petani mangga di Jawa Timur atau Nusa Tenggara Barat sebagai proof-of-concept skala kecil sebelum adopsi yang lebih luas.

4. KESIMPULAN

Penelitian ini telah melakukan analisis komparatif yang komprehensif antara arsitektur ResNet50 (dua skenario: base dan fine-tuned) dan MobileNetV2 untuk klasifikasi kelayakan buah mangga pascapanen berdasarkan SNI 3164:2009, menggunakan dataset 1.700 citra dengan karakteristik cluttered background yang merepresentasikan kondisi riil UMKM pertanian Indonesia. Hasil eksperimen secara empiris menolak hipotesis konvensional bahwa model dengan jumlah parameter lebih banyak selalu menghasilkan kinerja lebih unggul.

MobileNetV2 mencapai akurasi klasifikasi tertinggi sebesar 97% (F1-Score = 0,97) bahkan tanpa prosedur fine-tuning, melampaui ResNet50 base (78%, F1 = 0,78) dan ResNet50 pasca-fine-tuning (94%, F1 = 0,94). Keunggulan MobileNetV2 tidak hanya terletak pada akurasi, tetapi juga pada efisiensi komputasi: latensi inferensi 7,60 ms per citra (berbanding 12,70ms ResNet50) dan durasi pelatihan hanya 55 detik (berbanding 95,55 detik ResNet50 Fine-Tuned). Keunggulan ganda ini - akurasi tinggi sekaligus komputasi ringan - bersumber dari desain arsitektural Depthwise Separable Convolution dan mekanisme Inverted Residual with Linear Bottleneck yang secara efektif memaksa jaringan berkonsentrasi pada fitur diskriminatif esensial alih-alih menghafalkan noise visual pada cluttered background. Hasil penelitian memvalidasi MobileNetV2 sebagai kandidat optimal untuk konversi ke TensorFlow Lite dan deployment nyata pada smartphone petani dan UMKM di Indonesia sebagai solusi kontrol kualitas berbasis Edge AI yang demokratis dan terjangkau.

DAFTAR PUSTAKA

- [1] M. Sab, M. B. Ashok, and S. N. Sudhakara, "Estimation of postharvest losses of mangoes at different stages from harvesting to consumption," *International Journal of Current Microbiology and Applied Sciences (IJCMAS)*, vol. 6, no. 6, pp. 310–318, 2017.
- [2] IETA, "Classification of harvesting age of mango based on NIR spectra using machine learning algorithms," *Mathematical Modelling of Engineering Problems (Dipublikasikan oleh IETA)*, Vols. Vol. 10, No. 1, 2023.
- [3] Direktorat Pengolahan dan Pemasaran Hasil Hortikultura, "Mangga: Peluang Menguasai Pangsa Pasar Global dengan Keistimewaannya (Evaluasi Mutu SNI 3164:2009)," Kementerian Pertanian Republik Indonesia, 2023.
- [4] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018.
- [5] H. A. Santoso, B. F. Safsalta, N. Febrianto, G. W. Saraswati, and S. C. Haw, "Comparative analysis of convolutional neural network and DenseNet121 transfer learning in agriculture focusing on crop leaf disease identification," *Applied Computing and Informatics*, vol. 20, no. 1/2, pp. 45–59, 2024.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.
- [8] T. N. Doan and D. N. Thi, "A novel mango grading system based on image processing and machine learning methods," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 5, pp. 445–453, 2023.
- [9] H. Suhendar, V. Efelina, and M. Ziveria, "Fruit quality classification using convolutional neural network," *Journal of Physics: Conference Series*, vol. 2377, 012015, 2022.
- [10] B. Hermanto and A. Nugroho, "Comparison of MobileNetV2 and ResNet50 for date fruit classification," *Journal of Computational and Theoretical Nanoscience*, vol. 21, no. 4, pp. 88–97, 2024.
- [11] V. Zárate and A. Hernández, "Simplified deep learning for accessible fruit quality assessment in small agricultural operations," *Applied Sciences*, vol. 14, no. 3, 1127, 2024.
- [12] P. Jayanthi and L. S. Kumar, "Comparison of convolutional neural networks architectures for mango leaf classification," *International Journal of Computational Vision and Robotics*, vol. 14, no. 1, pp. 84–98, 2024.
- [13] R. S. Bana, A. Kumar, and P. K. Bhatt, "Mango anthracnose and powdery mildew disease detection using convolutional neural network and artificial neural network," *Journal of Plant Disease Sciences*, vol. 18, no. 1, pp. 11–19, 2023.
- [14] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019