

Evaluasi Kinerja MobileNetV2 Berbasis Transfer Learning untuk Klasifikasi Multikelas Penyakit Stroberi

Muhammad Doni Saputra✉, Yuwan Jumaryadi*✉

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana,
Indonesia

Abstrak: Identifikasi penyakit stroberi diperlukan untuk mendukung pengendalian penyakit dan mempertahankan produktivitas tanaman. Penelitian ini bertujuan mengevaluasi kinerja MobileNetV2 berbasis *transfer learning* dalam klasifikasi multikelas penyakit stroberi serta menganalisis pengaruh *fine-tuning* terhadap kinerja model. Dataset publik yang digunakan terdiri atas 2.053 citra yang terbagi ke dalam tujuh kelas, yaitu *Angular Leaf Spot*, *Anthrachnose Fruit Rot*, *Blossom Blight*, *Gray Mold*, *Leaf Spot*, *Powdery Mildew Fruit*, dan *Powdery Mildew Leaf*. Dataset dibagi menggunakan *stratified sampling* dengan rasio sekitar 80:10:10 sehingga diperoleh 1.642 citra pelatihan, 205 citra validasi, dan 206 citra pengujian. Seluruh citra diubah ukurannya menjadi 224 × 224 piksel dan dipraproses sesuai dengan kebutuhan masukan MobileNetV2, sedangkan augmentasi hanya diterapkan pada data pelatihan. Dua skenario *transfer learning* dibandingkan, yaitu *feature extraction* dengan membekukan lapisan dasar MobileNetV2 dan *fine-tuning* dengan melatih kembali lapisan dasar tersebut menggunakan *learning rate* yang lebih rendah. Hasil pengujian menunjukkan bahwa *feature extraction* menghasilkan akurasi sebesar 95,63% dan *macro F1-score* 0,85. Penerapan *fine-tuning* meningkatkan akurasi menjadi 98,06% dan *macro F1-score* menjadi 0,96. Peningkatan kinerja juga terlihat pada kelas minoritas *Anthrachnose Fruit Rot* dan *Powdery Mildew Fruit*. Namun, hasil pada kedua kelas tersebut perlu ditafsirkan secara hati-hati karena jumlah sampel pengujiannya terbatas. Secara keseluruhan, *fine-tuning* meningkatkan kinerja MobileNetV2 pada subset pengujian internal. Hasil tersebut menunjukkan bahwa *fine-tuning* meningkatkan akurasi sekaligus menurunkan *test loss* pada data pengujian yang sama.

Kata Kunci: *fine-tuning; klasifikasi penyakit stroberi; MobileNetV2; penyakit stroberi; transfer learning*

Abstract: Identifying strawberry diseases is essential for supporting disease control and maintaining crop productivity. This study aims to evaluate the performance of a transfer learning-based MobileNetV2 model in the multi-class classification of strawberry diseases and to analyze the impact of fine-tuning on model performance. The public dataset used consists of 2,053 images categorized into seven classes: *Angular Leaf Spot*, *Anthrachnose Fruit Rot*, *Blossom Blight*, *Gray Mold*, *Leaf Spot*, *Powdery Mildew Fruit*, and *Powdery Mildew Leaf*. The dataset was split using stratified sampling with an approximate 80:10:10 ratio, resulting in 1,642 training images, 205 validation images, and 206 testing images. All images were resized to 224 × 224 pixels and preprocessed to meet MobileNetV2 input requirements, while data augmentation was applied exclusively to the training set. Two transfer learning scenarios were compared: feature extraction (freezing the

MobileNetV2 base layers) and fine-tuning (retraining the base layers using a lower learning rate). Test results indicate that feature extraction achieved an accuracy of 95.63% and a macro F1-score of 0.85. Applying fine-tuning improved accuracy to 98.06% and the macro F1-score to 0.96. Performance improvements were also observed for the minority classes—Anthracnose Fruit Rot and Powdery Mildew Fruit. However, results for these two classes require cautious interpretation due to the limited number of test samples. Overall, fine-tuning enhanced MobileNetV2's performance on the internal test subset. The results indicate that fine-tuning improves accuracy while simultaneously reducing test loss on the same test data.

Keywords: *fine-tuning; strawberry disease classification; MobileNetV2; strawberry disease; transfer learning;*

1. PENDAHULUAN

Stroberi (*Fragaria × ananassa*) merupakan komoditas hortikultura bernilai ekonomi tinggi yang banyak dibudidayakan karena permintaan pasar dan nilai jualnya [1]. Meskipun memiliki potensi pengembangan yang baik, produktivitas stroberi masih dipengaruhi oleh berbagai faktor, salah satunya serangan penyakit tanaman yang dapat menurunkan kualitas dan hasil produksi [2].

Identifikasi penyakit stroberi secara dini menjadi penting karena keterlambatan deteksi dapat menyebabkan penyebaran penyakit dan meningkatkan risiko kehilangan hasil [3], [4]. Dalam praktiknya, identifikasi masih banyak dilakukan melalui pengamatan visual, sehingga hasilnya bergantung pada pengalaman pengamat dan berpotensi menimbulkan kesalahan ketika beberapa penyakit memiliki karakteristik visual yang serupa. Kondisi tersebut mendorong perlunya metode identifikasi yang lebih objektif, cepat, dan konsisten untuk mendukung proses deteksi serta pengendalian penyakit tanaman.

Perkembangan kecerdasan buatan, khususnya *deep learning*, telah mendorong pengembangan metode otomatis untuk klasifikasi penyakit tanaman berbasis citra [5], [6], [7]. Salah satu pendekatan yang banyak digunakan adalah *Convolutional Neural Network* (CNN), karena mampu mempelajari representasi visual secara langsung dari data, termasuk pola warna, tekstur, tepi, dan bentuk, tanpa bergantung pada proses ekstraksi fitur secara manual [8], [9], [10], [11], [12]. Di antara berbagai arsitektur CNN, MobileNetV2 dirancang sebagai arsitektur ringan yang memanfaatkan *depthwise separable convolution*, *inverted residual*, dan *linear bottleneck* untuk mengurangi kompleksitas komputasi sekaligus mempertahankan kemampuan representasi fitur [13], [14], [15].

Penerapan MobileNetV2 pada klasifikasi citra tanaman telah menunjukkan kinerja yang kompetitif pada berbagai studi. Pada klasifikasi penyakit daun kentang, MobileNetV2 dilaporkan mencapai akurasi sebesar 95,31% [16], sedangkan pada identifikasi daun tanaman obat diperoleh akurasi pengujian sebesar 94% [17]. Temuan tersebut menunjukkan bahwa MobileNetV2 memiliki potensi untuk digunakan pada tugas klasifikasi citra tanaman dengan kebutuhan komputasi yang relatif rendah. Namun, kinerja model tetap dipengaruhi oleh karakteristik dataset, jumlah kelas, tingkat kemiripan visual antarkelas, serta strategi pelatihan yang digunakan.

Meskipun MobileNetV2 telah diterapkan pada berbagai tugas klasifikasi penyakit tanaman, evaluasi terhadap kemampuannya dalam membedakan beberapa penyakit stroberi yang muncul pada daun, buah, dan bunga masih relatif terbatas. Tantangan klasifikasi tidak hanya berasal dari kemiripan karakteristik visual antarkelas, seperti warna, tekstur, dan pola bercak, tetapi juga dari distribusi jumlah citra yang tidak seimbang. Oleh karena itu, diperlukan evaluasi yang tidak hanya menilai akurasi keseluruhan, tetapi juga kinerja per kelas serta pengaruh strategi *transfer learning* terhadap kemampuan model dalam mempelajari karakteristik spesifik penyakit stroberi.

Untuk menjawab kesenjangan tersebut, penelitian ini mengevaluasi MobileNetV2 berbasis *transfer learning* pada 2.053 citra yang terbagi ke dalam tujuh kelas, yaitu Angular Leaf Spot, Anthracnose Fruit Rot, Blossom Blight, Gray Mold, Leaf Spot, Powdery Mildew Fruit, dan Powdery Mildew Leaf. Dua skenario pelatihan dibandingkan, yaitu feature extraction dengan lapisan dasar MobileNetV2 dibekukan dan *fine-tuning* dengan seluruh lapisan dasar diaktifkan kembali. Kontribusi penelitian ini terletak pada evaluasi empiris kedua strategi tersebut pada dataset penyakit stroberi yang tidak seimbang menggunakan accuracy, test loss, *precision*, recall, F1-score, macro average, weighted average, dan support setiap kelas.

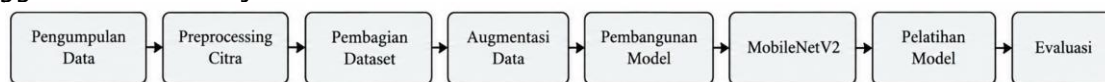
Berdasarkan pendekatan tersebut, penelitian ini bertujuan menganalisis kemampuan MobileNetV2 dalam mengklasifikasikan tujuh kategori penyakit stroberi serta mengukur dampak *fine-tuning* terhadap performa model. Evaluasi dilakukan pada data pengujian yang dipisahkan dari proses pelatihan dan validasi sehingga hasil dapat digunakan untuk menilai kemampuan generalisasi model pada *dataset* penelitian. Temuan yang diperoleh diharapkan menjadi dasar bagi pengembangan sistem identifikasi penyakit stroberi berbasis citra dan penelitian lanjutan mengenai penanganan ketidakseimbangan kelas, validasi menggunakan data lapangan, serta perbandingan dengan arsitektur *deep learning* lainnya.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental kuantitatif untuk membangun dan mengevaluasi model klasifikasi multikelas penyakit stroberi berbasis citra digital. MobileNetV2 dengan bobot pralatih ImageNet digunakan melalui dua skenario *transfer learning*, yaitu *feature extraction* dan *fine-tuning*. Eksperimen dilakukan pada tujuh kelas penyakit stroberi. Tahapan penelitian meliputi pengumpulan dataset, prapemrosesan citra, pembagian dataset, augmentasi data, pembangunan model, pelatihan, *fine-tuning*, dan evaluasi menggunakan accuracy, test loss, *precision*, recall, F1-score, macro average, weighted average, serta support. Data pengujian dipisahkan dari data pelatihan dan validasi untuk mengevaluasi kinerja internal pada citra yang tidak digunakan selama proses pembelajaran.

2.1. Tahapan Penelitian

Tahapan penelitian disusun secara sistematis mulai dari identifikasi permasalahan hingga evaluasi kinerja model.



Gambar 1. Tahapan Penelitian

Pengumpulan Dataset

Dataset penelitian diperoleh dari Kaggle melalui *Strawberry Disease Detection Dataset* dan terdiri atas 2.053 citra yang terbagi ke dalam tujuh kelas, yaitu *Angular Leaf Spot*, *Anthracnose Fruit Rot*, *Blossom Blight*, *Gray Mold*, *Leaf Spot*, *Powdery Mildew Fruit*, dan *Powdery Mildew Leaf*. Label kelas yang tersedia pada dataset digunakan sebagai *ground truth* tanpa pelabelan ulang. Selanjutnya, label kategorikal dikonversi menjadi label numerik 0–6 agar dapat diproses oleh model selama tahap pelatihan dan evaluasi.

Prapemrosesan Citra

Prapemrosesan dilakukan untuk menyesuaikan citra dengan kebutuhan input MobileNetV2. Seluruh citra dibaca dalam format RGB dan diubah ukurannya menjadi 224 × 224 piksel. Nilai piksel kemudian diproses menggunakan fungsi *preprocess_input* MobileNetV2 sehingga ditransformasikan ke rentang sekitar -1 hingga 1. Selanjutnya, label

kategorikal dikonversi menjadi bilangan integer 0–6 untuk merepresentasikan tujuh kelas penyakit stroberi yang digunakan dalam penelitian. Setelah tahap ini, data siap digunakan pada proses pembagian dataset, augmentasi, dan pelatihan model.

Pembagian Dataset

Dataset dibagi menjadi tiga bagian, yaitu data training, validation, dan testing menggunakan fungsi `train_test_split` dengan metode stratified sampling. Pembagian dilakukan dengan rasio sekitar 80:10:10, sehingga dari total 2.053 citra diperoleh 1.642 citra training, 205 citra validation, dan 206 citra testing. Parameter `random_state=42` digunakan agar pembagian data memberikan hasil yang konsisten ketika eksperimen dijalankan kembali.

Metode stratified sampling digunakan untuk mempertahankan proporsi setiap kelas pada masing-masing subset. Dengan demikian, seluruh kelas tetap terwakili pada data training, validation, dan testing, meskipun jumlah citra antar kelas tidak seimbang. Distribusi data hasil pembagian disajikan pada Tabel 1.

Tabel 1. Distribusi data hasil pembagian

Kelas	Training	Validation	Testing	Total
Angular Leaf Spot	348	43	44	435
Anthracoze Fruit Rot	78	10	9	97
Blossom Blight	166	21	21	208
Gray Mold	382	47	48	477
Leaf Spot	453	57	57	567
Powdery Mildew Fruit	44	6	5	55
Powdery Mildew Leaf	171	21	22	214
Total	1.642	205	206	2.053

Berdasarkan Tabel 1, setiap kelas tetap terwakili pada seluruh subset melalui stratified sampling. Teknik ini mempertahankan proporsi kelas pada dataset awal, tetapi tidak menghilangkan ketidakseimbangan jumlah citra. Eksperimen ini tidak menerapkan class weighting, oversampling, atau undersampling; oleh karena itu, pengaruh distribusi kelas dianalisis melalui macro average, weighted average, dan metrik setiap kelas.

Augmentasi Data

Augmentasi data diterapkan hanya pada data *training* untuk meningkatkan keragaman citra dan mengurangi risiko *overfitting* selama proses pelatihan. Data *validation* dan *testing* tidak mengalami augmentasi agar evaluasi kinerja model dilakukan menggunakan citra yang tidak dimodifikasi. Teknik augmentasi yang digunakan meliputi pembalikan citra secara horizontal (*Random Flip*), rotasi dengan faktor 0,2 (*Random Rotation*), perubahan skala dengan faktor 0,2 (*Random Zoom*), serta penyesuaian kontras dengan faktor 0,2 (*Random Contrast*). Seluruh transformasi diimplementasikan menggunakan lapisan augmentasi pada TensorFlow dan diterapkan secara dinamis selama proses pelatihan.

Tabel 2. Teknik Augmentasi Data

Teknik Augmentasi	Parameter
Random Flip	Horizontal
Random Rotation	0,2
Random Zoom	0,2
Random Contrast	0,2

Penerapan augmentasi menghasilkan variasi citra dalam aspek orientasi, sudut, skala, dan kontras tanpa mengubah label kelas. Variasi tersebut digunakan untuk memperluas keragaman representasi data pelatihan sehingga model dapat mempelajari karakteristik visual penyakit secara lebih general dan mengurangi ketergantungan terhadap kondisi citra tertentu.

Pelatihan dan Evaluasi Model

Setelah prapemrosesan dan pembagian *dataset*, MobileNetV2 dilatih dalam dua skenario. Pada skenario *feature extraction*, seluruh lapisan *backbone* dibekukan sehingga pembaruan bobot hanya dilakukan pada *classification head*. Pada skenario *fine-tuning*, seluruh *backbone* diaktifkan kembali agar bobot pralatih dapat disesuaikan dengan karakteristik *dataset* target. Data *validation* digunakan untuk memantau kinerja selama pelatihan, sedangkan data *testing* hanya digunakan setelah proses pelatihan selesai.

Kinerja kedua skenario dievaluasi pada data testing menggunakan test loss, accuracy, *precision*, recall, F1-score, macro average, weighted average, dan support. Macro average memberikan bobot yang sama kepada setiap kelas, sedangkan weighted average mempertimbangkan jumlah sampel per kelas. Kedua ringkasan tersebut digunakan bersama untuk menilai dampak ketidakseimbangan kelas.

2.2. Pemodelan MobileNetV2

MobileNetV2 digunakan sebagai *backbone* dengan konfigurasi *include_top=False* dan bobot pralatih ImageNet. Input model berupa citra RGB berukuran 224×224 piksel. Lapisan klasifikasi bawaan tidak digunakan sehingga keluaran *backbone* diteruskan ke *classification head* yang disesuaikan untuk tujuh kelas penyakit stroberi.

Classification head terdiri atas *GlobalAveragePooling2D*, *Dropout* sebesar 0,5, *Dense* 32 neuron dengan aktivasi ReLU dan regularisasi L2 sebesar 0,001, *Dropout* kedua sebesar 0,5, serta *Dense* tujuh neuron dengan aktivasi *Softmax*. Pada *feature extraction*, *backbone* MobileNetV2 dibekukan, sedangkan pada *fine-tuning* seluruh lapisan *backbone* diaktifkan kembali.

Model dikompilasi menggunakan *optimizer* Adam dan fungsi *loss* *Sparse Categorical Crossentropy* karena label kelas direpresentasikan sebagai bilangan integer. *Learning rate* awal ditetapkan sebesar 0,001 untuk *feature extraction* dan 0,0001 untuk *fine-tuning*. *Batch size* yang digunakan adalah 32 dengan jumlah maksimum 300 *epoch*.

Proses pelatihan menggunakan *ReduceLROnPlateau* dengan parameter *monitor* pada *validation loss*, *factor* sebesar 0,2, *patience* sebesar 3, dan minimum *learning rate* sebesar 0,00001. *EarlyStopping* diterapkan dengan *patience* sebesar 10 dan *restore_best_weights=True* sehingga proses pelatihan dapat dihentikan ketika *validation loss* tidak lagi menunjukkan perbaikan dan bobot terbaik dapat dipulihkan.

Data *training* digunakan untuk memperbarui bobot sesuai skenario pelatihan, sedangkan data *validation* digunakan untuk memantau *accuracy* dan *loss* selama proses pembelajaran. Setelah pelatihan selesai, kedua model dievaluasi menggunakan 206 citra data *testing* yang tidak terlibat dalam proses pelatihan maupun validasi.

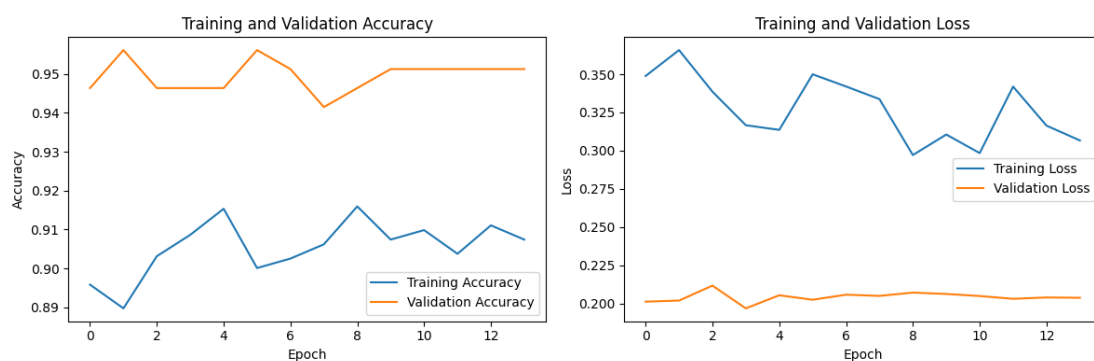
Prediksi akhir ditentukan berdasarkan kelas dengan probabilitas *Softmax* tertinggi. Prediksi kemudian dibandingkan dengan label aktual untuk menghitung test loss, accuracy, *precision*, recall, F1-score, macro average, weighted average, dan support. Seluruh nilai tersebut diperoleh dari model, *test set*, dan prosedur evaluasi yang sama.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Eksperimen dilakukan pada 2.053 citra penyakit stroberi yang dibagi menjadi 1.642 citra *training*, 205 citra *validation*, dan 206 citra *testing*. Dua skenario MobileNetV2 dibandingkan, yaitu *feature extraction* dan *fine-tuning*. Pada skenario *feature extraction*, *backbone* dipertahankan dalam kondisi *frozen*, sedangkan pada skenario *fine-tuning* seluruh *backbone* dibuat *trainable* dengan *learning rate* awal yang lebih rendah.

Perkembangan kinerja model selama proses pelatihan ditunjukkan pada Gambar 2 melalui kurva *training accuracy*, *validation accuracy*, *training loss*, dan *validation loss*. Kurva tersebut digunakan untuk mengamati proses konvergensi model serta perbedaan kinerja antara data pelatihan dan validasi selama proses pembelajaran.



Gambar 2. Grafik akurasi dan loss pada data training dan validation

Selain memantau perubahan nilai *accuracy* dan *loss* sebagaimana ditunjukkan pada Gambar 2, proses pelatihan juga menerapkan *ReduceLROnPlateau* dan *EarlyStopping*. *ReduceLROnPlateau* digunakan untuk menurunkan *learning rate* ketika *validation loss* tidak menunjukkan perbaikan dalam beberapa *epoch*, sedangkan *EarlyStopping* digunakan untuk menghentikan proses pelatihan apabila kinerja validasi tidak lagi mengalami peningkatan. Penerapan kedua *callback* tersebut bertujuan untuk menjaga kestabilan proses optimasi sekaligus mengurangi risiko *overfitting* selama pelatihan model.

Perubahan *learning rate* selama proses pelatihan dirangkum pada Tabel 3. Nilai *learning rate* awal sebesar 0,001000 diturunkan secara bertahap menjadi 0,000200 pada *epoch* ke-8, 0,000040 pada *epoch* ke-11, dan 0,000010 pada *epoch* ke-14. Penurunan tersebut menunjukkan bahwa proses optimasi secara adaptif menggunakan langkah pembaruan bobot yang semakin kecil ketika *validation loss* tidak lagi menunjukkan perbaikan yang memadai.

Tabel 3. Perubahan *Learning Rate* selama Proses Pelatihan

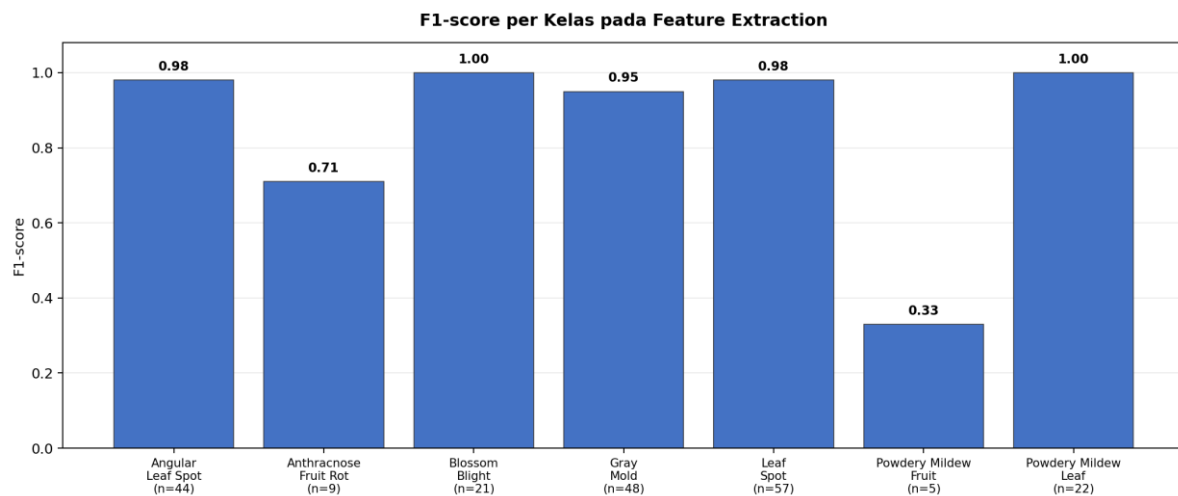
Epoch	Learning Rate Sebelum	Learning Rate Sesudah
8	0,001000	0,000200
11	0,000200	0,000040
14	0,000040	0,000010

Hasil pelatihan menunjukkan bahwa kedua skenario mampu mempelajari pola klasifikasi pada *dataset* yang digunakan. Namun, skenario *fine-tuning* menghasilkan performa validasi dan pengujian yang lebih baik dibandingkan dengan *feature extraction*. Perbedaan kinerja kedua skenario selanjutnya dianalisis secara lebih rinci menggunakan data *testing* serta metrik evaluasi pada masing-masing kelas.

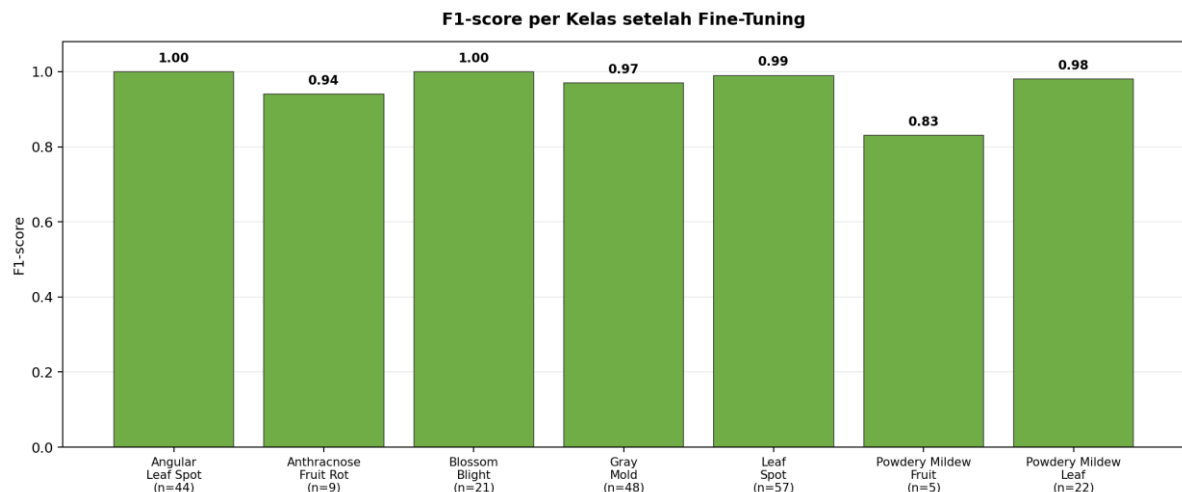
3.2. Hasil Pengujian

Pengujian dilakukan pada 206 citra yang tidak digunakan selama proses pelatihan maupun validasi. Berdasarkan hasil evaluasi pada *test set* yang sama, MobileNetV2 dengan *feature extraction* menghasilkan test loss 0,1748 dan akurasi 95,63%, sedangkan *fine-tuning* menghasilkan test loss 0,1541 dan akurasi 98,06%. Nilai tersebut digunakan secara konsisten pada abstrak, pembahasan, Tabel 6, dan kesimpulan.

Gambar 3 dan Gambar 4 menyajikan F1-score setiap kelas berdasarkan classification report dari keluaran pengujian yang sama. Karena distribusi kelas tidak seimbang, interpretasi tidak hanya didasarkan pada akurasi keseluruhan, tetapi juga pada *precision*, *recall*, *F1-score*, macro average, weighted average, dan support.



Gambar 3. F1-score per kelas pada MobileNetV2 dengan *feature extraction*



Gambar 4. F1-score per kelas pada MobileNetV2 dengan *fine-tuning*

Classification report untuk *feature extraction* disajikan pada Tabel 4, sedangkan hasil *fine-tuning* disajikan pada Tabel 5. Kedua tabel menggunakan prediksi pada 206 citra pengujian yang sama dan melaporkan *precision*, *recall*, *F1-score*, *support*, macro average, serta weighted average.

Tabel 4. Hasil Evaluasi Model MobileNetV2

Kelas	Precision	Recall	F1-Score	Support
Angular Leaf Spot	0,96	1,00	0,98	44
Anthracoese Fruit Rot	0,75	0,67	0,71	9
Blossom Blight	1,00	1,00	1,00	21
Gray Mold	0,91	1,00	0,95	48
Leaf Spot	1,00	0,96	0,98	57
Powdery Mildew Fruit	1,00	0,20	0,33	5
Powdery Mildew Leaf	1,00	1,00	1,00	22
Accuracy	–	–	0,9563	206
Macro Average	0,94	0,83	0,85	206
Weighted Average	0,96	0,96	0,95	206

Berdasarkan Tabel 4, *feature extraction* menghasilkan akurasi 0,9563. Macro *precision*, macro recall, dan macro F1-score masing-masing sebesar 0,94, 0,83, dan 0,85, sedangkan weighted F1-score mencapai 0,95. Perbedaan antara macro F1-score dan weighted F1-score menunjukkan bahwa akurasi keseluruhan yang tinggi masih dipengaruhi oleh performa kelas berukuran besar. Powdery Mildew Fruit merupakan kelas dengan performa terendah, yaitu recall 0,20 dan F1-score 0,33; hanya satu dari lima citra pengujian pada kelas tersebut yang dikenali dengan benar. Anthracnose Fruit Rot juga menunjukkan performa lebih rendah daripada mayoritas kelas, dengan recall 0,67 dan F1-score 0,71.

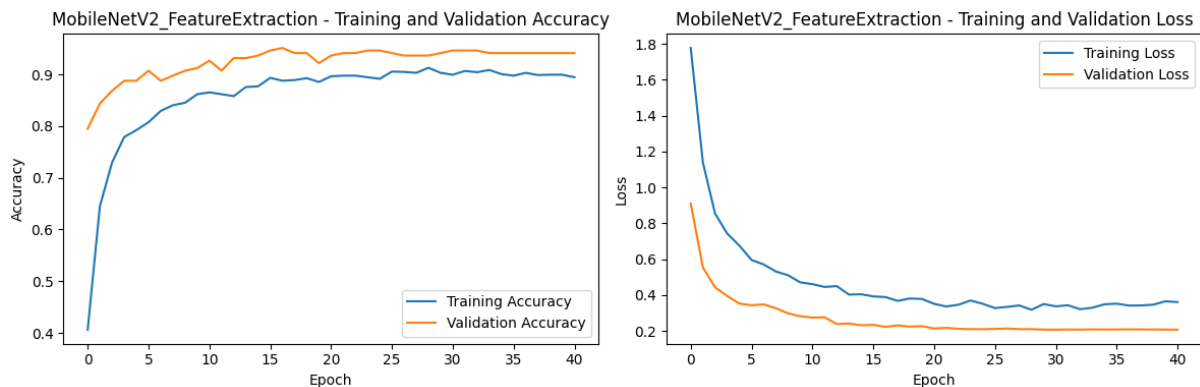
Tabel 5. Hasil Evaluasi Model MobileNetV2+Fine Tuning

Kelas	Precision	Recall	F1-Score	Support
Angular Leaf Spot	1,00	1,00	1,00	44
Anthracoese Fruit Rot	1,00	0,89	0,94	9
Blossom Blight	1,00	1,00	1,00	21
Gray Mold	0,98	0,96	0,97	48
Leaf Spot	1,00	0,98	0,99	57
Powdery Mildew Fruit	0,71	1,00	0,83	5
Powdery Mildew Leaf	0,96	1,00	0,98	22
Accuracy	–	–	0,9806	206
Macro Average	0,95	0,98	0,96	206
Weighted Average	0,98	0,98	0,98	206

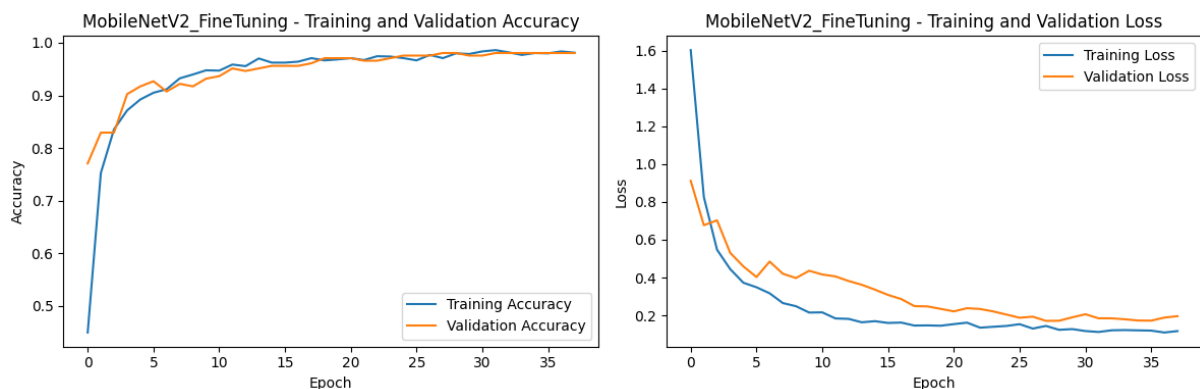
Setelah *fine-tuning*, akurasi meningkat menjadi 0,9806. Macro *precision*, macro recall, dan macro F1-score masing-masing mencapai 0,95, 0,98, dan 0,96, sedangkan weighted F1-score mencapai 0,98. F1-score Anthracnose Fruit Rot meningkat dari 0,71 menjadi 0,94. Pada Powdery Mildew Fruit, recall meningkat dari 0,20 menjadi 1,00 dan F1-score dari 0,33 menjadi 0,83. Namun, *precision* kelas tersebut sebesar 0,71 menunjukkan bahwa sebagian citra dari kelas lain masih diprediksi sebagai Powdery Mildew Fruit.

Dataset memiliki distribusi kelas yang tidak seimbang: Leaf Spot sebagai kelas terbesar terdiri atas 567 citra, sedangkan Powdery Mildew Fruit sebagai kelas terkecil hanya memiliki 55 citra, sehingga rasionya sekitar 10,3:1. Pada data pengujian, Anthracnose Fruit Rot hanya memiliki 9 citra dan Powdery Mildew Fruit hanya 5 citra. Akibatnya, satu perubahan prediksi dapat mengubah recall masing-masing kelas sekitar 11,11 dan 20 poin persentase. Peningkatan kelas minoritas setelah *fine-tuning* merupakan temuan yang positif, tetapi belum cukup untuk menyimpulkan kestabilan performa pada populasi yang lebih luas. Stratified sampling mempertahankan proporsi kelas, tetapi tidak mengatasi keterbatasan jumlah contoh minoritas.

Kurva training dan validation pada Gambar 5 dan Gambar 6 digunakan untuk mengamati kestabilan pembelajaran pada kedua skenario. Kurva tersebut, pemisahan data pengujian, dan penerapan augmentasi hanya pada data pelatihan mendukung validitas internal eksperimen, tetapi tidak dengan sendirinya membuktikan generalisasi terhadap citra dari sumber lain.



Gambar 5. Training dan Validation Akurasi MobilenetV2



Gambar 6. Training dan Validation Akurasi MobilenetV2+Fine Tuning

3.3. Analisis Perbandingan

Perbandingan kinerja kedua skenario MobileNetV2 dirangkum pada Tabel 6. Pada *feature extraction*, bobot *backbone* pralatih ImageNet dipertahankan dan hanya *classification head* yang diperbarui. Sebaliknya, *fine-tuning* memungkinkan seluruh *backbone* menyesuaikan bobot terhadap karakteristik visual *dataset* penyakit stroberi dengan *learning rate* awal yang lebih kecil.

Tabel 6. Perbandingan Kinerja MobileNetV2 *Feature Extraction* dan *Fine-Tuning*

Model	Akurasi Uji	Loss Uji
MobileNetV2+ <i>Feature Extraction</i>	0,9563	0,1748
MobileNetV2+ <i>Fine-Tuning</i>	0,9806	0,1541

Berdasarkan hasil pengujian, *fine-tuning* menunjukkan kinerja yang lebih baik daripada *feature extraction*. Akurasi meningkat dari 95,63% menjadi 98,06%, sedangkan macro F1-score meningkat dari 0,85 menjadi 0,96. Kenaikan macro F1-score menunjukkan

bahwa perbaikan tidak hanya terjadi pada kelas dominan, walaupun estimasi untuk kelas minoritas masih memiliki ketidakpastian karena jumlah sampel pengujiannya terbatas.

Fine-tuning juga menurunkan test loss dari 0,1748 menjadi 0,1541. Nilai loss yang lebih rendah menunjukkan bahwa, pada *test set* internal yang sama, distribusi probabilitas prediksi *fine-tuning* lebih sesuai dengan label aktual dibandingkan dengan *feature extraction*.

Secara keseluruhan, hasil eksperimen mendukung penggunaan *fine-tuning* untuk konfigurasi dan dataset penelitian ini. Pemisahan data pengujian dari proses pelatihan dan validasi serta penerapan augmentasi hanya pada data pelatihan digunakan untuk mengurangi risiko kebocoran data. Namun, seluruh citra berasal dari satu dataset publik dan pembagian dilakukan secara acak pada tingkat citra. Oleh karena itu, hasil pengujian internal belum sepenuhnya membuktikan bahwa model hanya mengenali karakteristik visual penyakit. Model masih mungkin memanfaatkan karakteristik lain yang berkorelasi dengan label, seperti organ tanaman, latar belakang, pencahayaan, perangkat pengambilan, atau pola sumber dataset. Selain itu, eksperimen belum mencakup pemeriksaan citra duplikat, pembagian data berdasarkan tanaman atau sumber pengambilan, analisis lokasi perhatian model, dan pengujian pada dataset eksternal. Dengan demikian, hasil penelitian ini dibatasi pada kemampuan generalisasi internal terhadap subset pengujian dari dataset yang sama.

4. KESIMPULAN

Penelitian ini mengevaluasi kinerja MobileNetV2 berbasis *transfer learning* dalam mengklasifikasikan 2.053 citra yang terbagi ke dalam tujuh kelas penyakit stroberi. Berdasarkan pengujian terhadap 206 citra, skenario *feature extraction* menghasilkan akurasi sebesar 95,63%, *test loss* 0,1748, dan *macro F1-score* 0,85. Penerapan *fine-tuning* meningkatkan akurasi menjadi 98,06% dan *macro F1-score* menjadi 0,96, serta menurunkan *test loss* menjadi 0,1541. Seluruh nilai tersebut diperoleh menggunakan *test set* dan prosedur evaluasi yang sama. Hasil ini menunjukkan bahwa *fine-tuning* meningkatkan kinerja keseluruhan dan keseimbangan klasifikasi antarkelas pada dataset penelitian.

Peningkatan kinerja juga terlihat pada kelas minoritas. F1-score kelas *Anthracoze Fruit Rot* meningkat dari 0,71 menjadi 0,94, sedangkan F1-score kelas *Powdery Mildew Fruit* meningkat dari 0,33 menjadi 0,83. Meskipun demikian, kedua kelas tersebut masing-masing hanya memiliki 9 dan 5 citra pada data pengujian. Dataset juga menunjukkan ketidakseimbangan kelas, dengan jumlah citra pada kelas terbesar dan terkecil masing-masing sebanyak 567 dan 55 citra atau memiliki rasio sekitar 10,3:1. Oleh karena itu, hasil pada kelas minoritas perlu ditafsirkan secara hati-hati karena jumlah sampel pengujian yang terbatas belum cukup untuk menunjukkan kestabilan kinerja model pada distribusi data yang lebih luas.

Secara keseluruhan, MobileNetV2 dengan *fine-tuning* menunjukkan kinerja terbaik pada subset pengujian internal yang digunakan dalam penelitian ini. Namun, hasil tersebut belum membuktikan kemampuan generalisasi model pada kondisi lapangan maupun memastikan bahwa keputusan model sepenuhnya didasarkan pada karakteristik visual penyakit. Model masih mungkin memanfaatkan karakteristik lain yang berkorelasi dengan label, seperti latar belakang, pencahayaan, organ tanaman, perangkat pengambilan, atau karakteristik sumber citra.

Penelitian selanjutnya perlu menggunakan citra lapangan independen, memeriksa citra duplikat atau hampir duplikat, serta menerapkan pembagian data berdasarkan tanaman, kebun, atau sesi pengambilan. Strategi penanganan ketidakseimbangan kelas juga perlu dievaluasi dengan penerapan terbatas pada data pelatihan. Selain itu, analisis Grad-CAM yang diverifikasi oleh pakar dan eksperimen penghilangan latar belakang atau penutupan

area lesi diperlukan untuk menilai apakah keputusan model benar-benar didasarkan pada gejala penyakit, bukan pada karakteristik khusus dataset.

PERNYATAAN KONTRIBUSI PENULIS (CREDIT)

Muhammad Doni Saputra: Konseptualisasi, Metodologi, Perangkat Lunak, Kurasi Data, Investigasi, Analisis Formal, Validasi, Penulisan – Draf Awal, dan Visualisasi. **Yuwan Jumaryadi:** Supervisi, Konseptualisasi, Metodologi, Validasi, Analisis Formal, serta Penulisan – Peninjauan dan Penyuntingan.

PERNYATAAN KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa tidak terdapat kepentingan finansial atau hubungan pribadi yang diketahui yang dapat memengaruhi hasil penelitian yang dilaporkan dalam artikel ini.

PERNYATAAN PENDANAAN

Penulis tidak menerima dukungan finansial untuk penelitian, kepenulisan, dan/atau publikasi artikel ini.

PERNYATAAN PENGGUNAAN AI GENERATIF DAN TEKNOLOGI BERBANTUAN AI DALAM PROSES PENULISAN

Para penulis menggunakan GPT-5.6 Sol untuk meningkatkan kejelasan penulisan artikel ini. Seluruh konten yang dibantu oleh AI telah ditinjau dan disunting oleh penulis, dan para penulis bertanggung jawab sepenuhnya atas isi akhir yang dipublikasikan.

REFERENCES

- [1] H. Naito, Y. Kawasaki, U. Lee, M. Takahashi, F. Hosoi, and T. Ota, "Proof-of-concept of harvest peak control using a strawberry cultivation emulator with artificial weather chambers," *Sci. Rep.*, vol. 16, no. 1, pp. 1–17, 2026, doi: 10.1038/s41598-026-46422-z.
- [2] A. Sitohang, P. Herbert, T. I. Hermanto, and C. D. Lestari, "Klasifikasi Jenis Penyakit Pada Daun Tumbuhan Stroberi Menggunakan Metode Convolutional Neural Network Arsitektur Inceptionv3," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 12, no. 3, pp. 4146–4154, 2024, [Online]. Available: [http://dx.doi.org/10.23960/jitet.v12i3\\$1.5274](http://dx.doi.org/10.23960/jitet.v12i3$1.5274)
- [3] H. Kaur, J. Wesche, J. Gelain, C. Minzheng, C. X. Luo, and G. Schnabel, "Characterization and Fungicide Sensitivity of *Gnomoniopsis fructicola* Causing Gnomonia Leaf Blotch of Strawberry in the Carolinas," *Plant Dis.*, vol. 109, no. 1, pp. 107–114, 2025, doi: 10.1094/PDIS-07-24-1361-RE.
- [4] T. Liao, R. Yang, P. Zhao, W. Zhou, M. He, and L. Li, "MDAM-DRNet: Dual Channel Residual Network With Multi-Directional Attention Mechanism in Strawberry Leaf Diseases Detection," *Front. Plant Sci.*, vol. 13, no. July, pp. 1–20, 2022, doi: 10.3389/fpls.2022.869524.
- [5] M. K. Nath, K. Mookkandi, K. Balakrishnan, M. Mishra, M. Kavitha, and M. Mathumathi, "Bottleneck of AI driven techniques for identifying indian agricultural diseases: A systematic review and future directives," *Results Eng.*, vol. 31, no. February, p. 111212, 2026, doi: 10.1016/j.rineng.2026.111212.
- [6] B. Kaur et al., "A hybrid approach for citrus disease detection using convolutional neural networks and fuzzy inference systems for enhanced accuracy and

- interpretability," *Sci. Rep.*, pp. 1–20, 2026, doi: 10.1038/s41598-026-52135-0.
- [7] I. Astuti, W. W. Ariestya, and B. Solehudin, "Deteksi Objek Daun Semanggi Secara Real Time Menggunakan CNN-Single Shot Multibox Detector (SSD)," *J. Ilm. FIFO*, vol. 14, no. 1, p. 47, 2022, doi: 10.22441/fifo.2022.v14i1.005.
- [8] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
- [9] I. K. Trisiawan and Y. Yuliza, "Penerapan Multi-Label Image Classification Menggunakan Metode Convolutional Neural Network (CNN) Untuk Sortir Botol Minuman," *J. Teknol. Elektro*, vol. 13, no. 1, p. 48, Feb. 2022, doi: 10.22441/jte.2022.v13i1.009.
- [10] W. Suprihatiningsih, D. Romahadi, and A. G. Feleke, "Electroencephalogram-Based Multi-Class Driver Fatigue Detection using Power Spectral Density and Lightweight Convolutional Neural Networks," *J. Eng. Technol. Sci.*, vol. 57, no. 4, pp. 445–457, 2025, doi: 10.5614/j.eng.technol.sci.2025.57.4.2.
- [11] T. M. Kadarina, Z. Iklima, R. Priambodo, Riandini, and R. N. Wardhani, "Dental caries classification using depthwise separable convolutional neural network for teledentistry system," *Bull. Electr. Eng. Informatics*, vol. 12, no. 2, pp. 940–949, 2023, doi: 10.11591/eei.v12i2.4428.
- [12] U. Salamah, N. Ani, Y. Jumaryadi, and Agustiawan, "Multimodal Transfer Learning for Anti-Inflammatory Medicinal Plant Leaf Classification using ResNet50," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 14, no. 1, pp. 131–140, 2026, doi: 10.33558/piksel.v14i1.12279.
- [13] Z. Li *et al.*, "Enhancing Tea Leaf Disease Identification with Lightweight MobileNetV2," *Comput. Mater. Contin.*, vol. 80, no. 1, pp. 679–694, 2024, doi: 10.32604/cmc.2024.051526.
- [14] S. Dwiasnati, Y. Devianto, S. M. Arif, and R. Avrizal, "Pemodelan Wilayah Titik Api Kebakaran Hutan Menggunakan Deep Learning," *J. Ilm. FIFO*, vol. 16, no. 1, p. 1, May 2024, doi: 10.22441/fifo.2024.v16i1.001.
- [15] Y. Jumaryadi, A. Muhammad Ihsan, and B. Priambodo, "Klasifikasi Jenis Buah-Buahan Menggunakan Citra Digital Dengan Metode Convolutional Neural Networks," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1737–1746, 2023, doi: 10.30865/klik.v4i3.1421.
- [16] R. Gunawan, F. Salim, A. I. Wahyudhy, A. Y. Wibowo, G. Yordan, and R. F. Filamori, "Klasifikasi Penyakit Daun Kentang dengan Transfer Learning Menggunakan CNN optimalisasi Arsitektur MobileNetV2 Potato Leaf Disease Classification with Transfer Learning Using CNN Optimization of MobileNetV2 Architecture," *J. Comput. Sci. Inf. Technol.*, vol. 6, no. 2, pp. 254–258, 2025, [Online]. Available: <http://ejurnal.umri.ac.id/index.php/coscitech/indexhttps://doi.org/10.37859/coscitech.v6i2.8599>
- [17] R. J. Hendri Butar-Butar and N. L. Marpaung, "Deep Learning untuk Identifikasi Daun Tanaman Obat Menggunakan Transfer Learning MobileNetV2," *J. Inform. J. Pengemb. IT*, vol. 8, no. 2, pp. 142–148, 2023, doi: 10.30591/jpit.v8i2.5217.