Peningkatan Akurasi Klasifikasi Kematangan Buah Pisang Menggunakan CNN dengan Attention Mechanism

Nazhmi Ahmad Fauzan
Universitas Telkom
Bandung
nazhmiaf@student.telkomuniversity.ac.id

Wikky Fawwaz
Universitas Telkom
Bandung
wikkyfawwaz@telkomuniversity.ac.id

Penyortiran buah otomatis berbasis *deep learning* pada perangkat *edge* menghadapi tantangan besar karena keterbatasan sumber daya komputasi. Arsitektur ringan seperti MobileNetV2 sangat efisien untuk inferensi, namun seringkali kurang sensitif dalam menangkap fitur *fine-grained* seperti gradasi warna dan fitur halus pada tugas klasifikasi tingkat kematangan. Penelitian ini mengusulkan strategi integrasi modul *Squeeze-and-Excitation* (SE) secara selektif pada MobileNetV2 untuk meningkatkan akurasi tanpa membebani parameter model secara signifikan. Metodologi penelitian melibatkan pengujian lima skenario: *Baseline*, *All Block*, serta integrasi selektif pada blok *Early* (1-5), *Middle* (6- 11), dan *Deep* (12-17) guna merekalibrasi fitur pada berbagai tingkatan hierarki. Hasil eksperimen pada *Banana Ripeness Classification Dataset* menunjukkan bahwa skenario Early memberikan performa paling optimal dengan akurasi validasi tertinggi sebesar 96,79% dan nilai *F1-score* sebesar 0,9679. Skenario ini terbukti sangat efisien karena mempertahankan jumlah parameter pada angka 2,88 M, serupa dengan model *baseline*. Sebaliknya, integrasi pada seluruh blok (*All Block*) justru menurunkan akurasi menjadi 95,19% dan meningkatkan parameter menjadi 2,91 M. Analisis *confusion matrix* mengonfirmasi bahwa model memiliki tingkat misklasifikasi yang sangat rendah pada kelas *unripe*. Penelitian ini menyimpulkan bahwa penempatan modul SE pada tahap awal jaringan merupakan strategi terbaik untuk menjaga efisiensi dan presisi klasifikasi pada perangkat *edge*.
Kata kunci: *deep learning*, *edge computing*, klasifikasi kematangan buah, *mobilenetv2*, *squeeze-and- excitation*.

Automated fruit sorting based on deep learning on edge devices faces significant challenges due to limited computational resources. Lightweight architectures such as MobileNetV2 are efficient for inference, yet they often lack sensitivity in capturing fine-grained features like color gradients and subtle features in ripeness classification tasks. This research proposes a selective Squeeze-and- Excitation (SE) module integration strategy on MobileNetV2 to enhance accuracy without significantly increasing the model's parameters. The research methodology involves testing five scenarios: Baseline, All Block, and selective integration in Early (blocks 1-5), Middle (blocks 6-11), and Deep (blocks 12-17) stages to recalibrate features at different hierarchical levels. Experimental results on the Banana Ripeness Classification Dataset demonstrate that the Early scenario yields the most optimal performance, achieving the highest validation accuracy of 96.79% and an F1-score of 0.9679. This scenario proves highly efficient as it maintains the parameter count at 2.88 M, identical to the baseline model. Conversely, integrating across all blocks decreased accuracy to 95.19% while increasing the parameters to 2.91 M. Confusion matrix analysis confirms that the model has a very low misclassification rate for the unripe class. This study concludes that placing SE modules at the early stages of the network is the best strategy to maintain efficiency and classification precision on edge devices.

Keywords: deep learning, edge computing, fruit ripeness classification, mobilenetv2, squeeze-and- excitation.

1. Pendahuluan Latar Belakang

Pisang (*Musa spp.*) merupakan salah satu komoditas hortikultura paling penting secara global, baik dari sisi produksi, nilai ekonomi, maupun kontribusinya terhadap ketahanan pangan. FAO melaporkan bahwa

pisang merupakan buah segar dengan volume ekspor tertinggi di dunia, dengan nilai perdagangan global mencapai lebih dari USD 10 miliar per tahun [1] . Selain bernilai ekonomi tinggi, pisang juga mengalami proses pematangan yang kompleks, ditandai oleh perubahan warna kulit, detil permukaan, kemunculan bintik cokelat, serta perubahan komposisi kimia internal dari pati menjadi gula [2]. Klasifikasi kematangan pisang berperan penting dalam rantai pasok pertanian, mulai dari penentuan waktu panen hingga pengendalian kualitas produk. Namun, proses penyortiran masih banyak dilakukan secara manual sehingga bersifat subjektif, tidak konsisten, dan rentan terhadap kesalahan. Oleh karena itu, diperlukan sistem klasifikasi otomatis berbasis pengolahan citra yang andal dan efisien [3].

Kemajuan teknologi *deep learning*, khususnya *Convolutional Neural Network* (CNN), telah membuka peluang untuk mengotomatisasi proses klasifikasi kematangan buah. CNN terbukti efektif dalam mengekstraksi fitur visual dari citra buah untuk tugas klasifikasi, dengan berbagai studi melaporkan akurasi tinggi pada dataset buah-buahan [4]. Namun, implementasi sistem klasifikasi berbasis *deep learning* pada perangkat *edge* dengan sumber daya komputasi terbatas menghadapi tantangan kritis yang perlu diatasi [5]. Maka dari itu, Arsitektur ringan seperti MobileNetV2 dirancang untuk efisiensi melalui *depthwise separable convolution* dan *inverted residual*. Namun, reduksi kanal yang agresif pada MobileNetV2 berpotensi menyebabkan hilangnya detail halus yang penting pada tugas *fine-grained classification* [6].

Untuk mengatasi permasalahan tersebut, mekanisme atensi seperti *Squeeze-and-Excitation* (SE) banyak diintegrasikan ke dalam CNN untuk meningkatkan kemampuan model dalam menekankan fitur kanal yang informatif [7]. Meskipun integrasi SE-Net secara menyeluruh terbukti meningkatkan akurasi, pendekatan ini berpotensi menambah kompleksitas komputasi [8], [9]. Beberapa penelitian terbaru menunjukkan bahwa integrasi atensi secara selektif pada blok tertentu dapat memberikan keseimbangan yang lebih baik antara akurasi dan efisiensi dengan mempertimbangkan hierarki fitur CNN [10], [11].

Topik dan Batasannya

Rumusan Masalah Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana memodifikasi model ringan dalam menangkap fitur visual halus (*fine-grained*) seperti transisi warna dan dettil pada pisang tanpa harus menggunakan arsitektur berat yang membebani komputasi?
2. Bagaimana merancang strategi integrasi mekanisme atensi yang tetap efektif, namun tidak menambah beban komputasi secara berlebihan pada arsitektur CNN ringan?
3. Bagaimana analisis performa model ringan ketika modul atensi selektif diintegrasikan pada blok residu tertentu?

Batasan Masalah Agar penelitian tetap terarah dan dapat diselesaikan dalam kurun waktu satu semester, ditetapkan batasan-batasan masalah sebagai berikut:

1. Dataset yang digunakan terbatas pada *Banana Ripeness Classification Dataset* dengan latar belakang yang bervariasi, dan tidak mencakup jenis buah lain.
2. Modifikasi arsitektur hanya dilakukan pada MobileNetV2, tanpa membandingkan dengan arsitektur ringan lainnya.
3. Evaluasi dilakukan pada lingkungan PyTorch menggunakan metrik akurasi dan *F1-Score*, tanpa pengujian langsung pada perangkat embedded fisik.
4. Sistem dibatasi pada tugas klasifikasi citra tunggal, bukan deteksi objek jamak.

Tujuan

Berdasarkan topik dan batasan, tujuan dari penelitian ini antara lain :

1. Mengembangkan modifikasi arsitektur CNN ringan yang mampu menangkap fitur *fine-grained* seperti transisi warna pisang tanpa meningkatkan beban komputasi secara signifikan.
2. Menerapkan strategi integrasi atensi selektif untuk memperoleh peningkatan performa tanpa menambah kompleksitas komputasi yang tidak perlu.
3. Menganalisis hasil performa model ringan yang diintegrasikan dengan modul atensi selektif pada blok residu tertentu.

2. Studi Terkait

Pendekatan awal dalam klasifikasi buah umumnya bergantung pada arsitektur Convolutional Neural Network (CNN) berkapasitas besar seperti VGG16, ResNet-50, dan InceptionV3 [4], [12]. Meskipun model- model tersebut menunjukkan akurasi tinggi dalam mengekstraksi fitur semantik, kompleksitas komputasi dan jumlah parameternya yang besar membatasi penerapannya pada perangkat edge dengan sumber daya terbatas

[3]. Hal ini mendorong tren penelitian untuk beralih pada arsitektur ringan (*lightweight*) yang dirancang untuk inferensi waktu nyata, seperti MobileNet dan ShuffleNet [13], [14].

MobileNetV2 berhasil mengurangi latensi melalui penerapan *depthwise separable convolution*[15], Namun, efisiensi ini pada beberapa studi dilaporkan dapat disertai penurunan kapasitas representasi, terutama pada tugas klasifikasi *fine-grained* yang menuntut pembedaan transisi warna dan detail halus [16]. Pada kondisi tersebut, model ringan berpotensi kehilangan informasi selama proses reduksi kanal sehingga sensitivitas terhadap detail menurun [17].

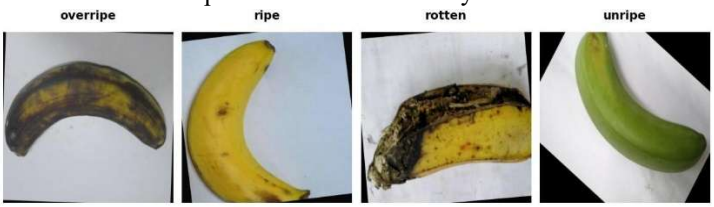
Untuk meningkatkan kualitas representasi, berbagai mekanisme atensi seperti *Squeeze-and-Excitation* (SE) dan CBAM digunakan untuk merekalibrasi bobot kanal secara dinamis [7], [18]. Namun, sebagian besar literatur menerapkan modul atensi secara menyeluruh pada seluruh blok jaringan (*full integration*) [19], [20]. Strategi ini memang meningkatkan kemampuan penekanan fitur relevan, tetapi juga dilaporkan dapat menambah kompleksitas komputasi dan pada beberapa kasus, memengaruhi stabilitas fitur pada lapisan awal yang masih mengekstraksi pola dasar[11]. Berdasarkan celah tersebut, penelitian ini mengusulkan strategi integrasi atensi selektif, yaitu menempatkan modul atensi hanya pada titik *bottleneck* yang dianggap paling berpengaruh terhadap pembentukan fitur. Tujuannya adalah mencapai keseimbangan antara peningkatan akurasi dan efisiensi komputasi pada arsitektur MobileNetV2.

3. Sistem yang Dibangun

3.1. Dataset dan Pra-pemrosesan Data

Penelitian ini menggunakan *Banana Ripeness Classification Dataset* [21]yang terdiri dari 13.478 citra. Berbeda dengan dataset sintetis berlatar belakang seragam, dataset ini dipilih karena merepresentasikan tantangan dunia nyata dengan kondisi pencahayaan variabel, pemandangan yang berantakan, dan latar belakang heterogen seperti terlihat pada Gambar 1. Kompleksitas visual ini krusial untuk menguji kemampuan model dalam memisahkan objek buah dari *noise* lingkungan.

Dataset ini sudah dibagi (pre-split) oleh penyedia dataset di Kaggle ke dalam tiga subset, yaitu Training Set sebanyak 11.793 citra (±87%), Validation Set sebanyak 1.123 citra (±8%), dan Test Set sebanyak 562 citra (±4%). Pembagian ini dipertahankan pada penelitian ini agar hasil yang diperoleh tetap sejalan dengan standar evaluasi yang digunakan pada penelitian sebelumnya serta menghindari potensi *data leakage*. Training set digunakan untuk proses pembelajaran model, validation set dimanfaatkan untuk validasi performa selama pelatihan dan penyesuaian hyperparameter, sedangkan test set hanya digunakan sekali pada tahap akhir untuk mengukur kemampuan generalisasi model terhadap data yang benar-benar belum pernah dilihat sebelumnya.



GAMBAR 1
Contoh gambar di masing-masing kelas

Pada tahap prapemrosesan, seluruh citra terlebih dahulu diubah ke ukuran 224×224 piksel agar sesuai dengan kebutuhan arsitektur MobileNetV2, kemudian setiap citra dinormalisasi menggunakan nilai rata-rata kanal RGB sebesar 0,485; 0,456; dan 0,406 serta standar deviasi 0,229; 0,224; dan 0,225. Nilai-nilai ini merupakan statistik distribusi piksel

pada dataset ImageNet yang secara luas digunakan sebagai standar normalisasi pada model CNN modern. Penerapan normalisasi ini bertujuan untuk menstabilkan distribusi nilai piksel, mengurangi variansi antar fitur, serta membantu proses pelatihan jaringan agar lebih stabil dan konvergen. Pada data latih, diterapkan data augmentation guna mengurangi risiko *overfitting*, meliputi *random resized crop* ke 224×224 piksel, *horizontal flip* acak, penyesuaian acak pada kecerahan, kontras, dan saturasi (*color jitter*), serta rotasi acak hingga ±20°. Sementara itu, data validasi hanya melalui proses *resize* ke 224×224 piksel tanpa augmentasi, dan data uji di-*resize* ke 256 piksel lalu di-*center crop* menjadi 224×224 piksel agar evaluasi tetap konsisten. Seluruh data diproses menggunakan *batch size* sebesar 32.

TABEL 1
Nilai rata-rata RGB sebelum dan setelah di normalisasi

Mode	Red	Green	Blue
Unnormalized	0.6114	0.6259	0.5904
Normalized	0.4830	0.4280	0.3334

Berdasarkan Tabel 1. Perbandingan nilai rata-rata kanal RGB, dapat diamati bahwa sebelum normalisasi, nilai intensitas warna citra pisang berada pada rentang 0–1, dengan dominasi kanal Green (0.6259) dan Red (0.6114). Hal ini wajar karena objek pisang secara visual memang didominasi warna kuning–hijau yang merupakan kombinasi kuat antara kanal merah dan hijau.

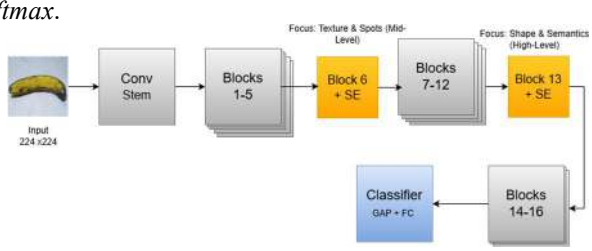
Setelah dilakukan normalisasi menggunakan standar ImageNet, nilai rata-rata kanal RGB mengalami penurunan dan pergeseran distribusi, khususnya pada kanal Blue (0.3334) yang menjadi paling kecil. Perubahan ini terjadi karena setiap kanal RGB dinormalisasi dengan cara dikurangi nilai mean dan dibagi standar deviasi ImageNet, sehingga distribusi data masukan menjadi lebih terpusat dan terstandarisasi. Normalisasi ini tidak bertujuan mengubah informasi visual citra, melainkan untuk menyamakan skala dan distribusi data dengan kondisi data yang digunakan.

3.2. Integrasi Atensi Selektif pada MobileNetV2

Sistem ini dibangun menggunakan *backbone* MobileNetV2 karena efisiensi struktur *inverted residual*-nya yang optimal untuk implementasi pada perangkat *edge*. Guna meningkatkan kemampuan model dalam mengekstraksi fitur *fine-grained* seperti transisi warna dan detil pada kulit buah. Diterapkan strategi modifikasi melalui integrasi modul *Squeeze-and-Excitation* (SE) yang dibagi ke dalam lima skenario pengujian. Selain model *Baseline* (tanpa SE) dan All Block (integrasi pada seluruh blok), penelitian ini memfokuskan analisis pada penempatan modul SE secara selektif pada tingkatan hierarki fitur yang berbeda:

1. Skenario Early (Blok 1-5): Modul SE diintegrasikan pada lapisan awal untuk merekalibrasi fitur dasar seperti tepian, warna dasar, dan kontras sebelum resolusi spasial berkurang secara signifikan.
2. Skenario Middle (Blok 6-11): Modul SE ditempatkan pada lapisan menengah untuk menangkap fitur yang lebih kompleks, seperti pola detil permukaan dan bintik-bintik kematangan.
3. Skenario Deep (Blok 12-17): Modul SE disisipkan pada tahap jaringan yang lebih dalam guna memproses informasi semantik dan menekankan fitur global seperti distribusi warna menyeluruh.

Seluruh fitur hasil ekstraksi dari berbagai skenario tersebut kemudian diproses melalui *Global Average Pooling* (GAP) dan dipetakan oleh *classification head* yang terdiri dari lapisan *Dense* (512 unit), *Batch Normalization*, aktivasi ReLU, serta *Dropout* (p=0.4) untuk menjaga stabilitas pelatihan, sebelum akhirnya mencapai lapisan keluaran *Softmax*.



GAMBAR 2
Arsitektur diagram

3.3 Skenario Pengujian dan Metrik Evaluasi

Pelatihan model dilakukan menggunakan framework PyTorch pada GPU NVIDIA L4 dengan durasi 40 epoch. Optimasi menggunakan Adam dengan learning rate awal 0.0001 yang dikendalikan oleh penjadwal ReduceLROnPlateau (patience=2) untuk memastikan model dapat keluar dari local minima saat konvergensi

melambat. Evaluasi kinerja sistem tidak hanya mengandalkan Akurasi global, tetapi juga menggunakan metrik Presisi, Recall, dan F1-Score. Penggunaan F1-Score sangat ditekankan dalam penelitian ini karena memberikan pandangan holistik keseimbangan antara presisi dan sensitivitas, yang krusial untuk menghindari bias pada kelas mayoritas. Selain itu, analisis Confusion Matrix digunakan untuk mengidentifikasi kesalahan spesifik pada kelas yang berdekatan (misalnya antara *Ripe* dan *Overripe*), memberikan wawasan mendalam mengenai ketahanan model terhadap kemiripan visual antar-kelas.

Accuracy:

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

Precision :

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F1-Score:

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

4. Evaluasi

4.1 Hasil Pengujian

Hasil pengujian pada Tabel 2 menunjukkan bahwa penambahan Squeeze-and-Excitation (SE) Block pada arsitektur MobileNetV2 memberikan dampak yang berbeda terhadap akurasi dan kompleksitas model bergantung pada lokasi penempatannya. Model baseline MobileNetV2 mencapai akurasi validasi sebesar 95,71% dengan jumlah parameter 2,88 juta, yang menunjukkan performa awal yang sudah baik namun belum memanfaatkan mekanisme penimbangan kanal fitur secara adaptif.

Penerapan SE-Block pada seluruh blok (All Block) justru menurunkan akurasi validasi menjadi 95,19%, meskipun jumlah parameter meningkat menjadi 2,91 juta. Kondisi ini

menunjukkan bahwa peningkatan kompleksitas model tidak selalu berbanding lurus dengan peningkatan performa. Penambahan attention secara menyeluruh berpotensi mengganggu efisiensi representasi fitur pada MobileNetV2 dan menyebabkan over-regularization, khususnya pada arsitektur ringan yang dirancang untuk efisiensi komputasi.

Penempatan SE-Block pada lapisan awal (blok 1–5) menghasilkan performa terbaik dengan akurasi validasi tertinggi sebesar 96,79%, sementara jumlah parameter tetap relatif sama dengan baseline (2,88 juta). Hasil ini menunjukkan bahwa peningkatan akurasi diperoleh tanpa menambah kompleksitas model secara signifikan. Lapisan awal CNN berperan dalam mengekstraksi fitur dasar seperti warna yang merupakan indikator utama kematangan buah sehingga penambahan SE-Block pada tahap ini mampu memperkuat kanal fitur yang paling relevan dan meningkatkan kemampuan diskriminasi kelas secara efisien.

Pada lapisan menengah (blok 6–11), akurasi validasi meningkat menjadi 96,17% dengan jumlah parameter yang tetap (2,88 juta). Hal ini menunjukkan bahwa mekanisme attention masih memberikan kontribusi positif, namun pengaruhnya mulai berkurang seiring meningkatnya tingkat abstraksi fitur yang diekstraksi. Sementara itu, penempatan SE-Block pada lapisan dalam (blok 12–17) hanya memberikan peningkatan akurasi yang bersifat marginal (95,99%) dengan jumlah parameter meningkat menjadi 2,90 juta, menunjukkan bahwa penambahan kompleksitas pada tahap ini tidak memberikan keuntungan performa yang signifikan.

Secara keseluruhan, analisis akurasi dan jumlah parameter menunjukkan bahwa penempatan SE-Block secara selektif lebih efektif dibandingkan penerapan pada seluruh lapisan. Skenario lapisan awal memberikan rasio peningkatan akurasi terhadap kompleksitas model yang paling optimal, karena mampu meningkatkan performa tanpa menambah beban parameter secara berarti. Temuan ini selaras dengan prinsip SE-Block yang menekankan *channel-wise feature recalibration* serta karakteristik MobileNetV2 sebagai arsitektur ringan, dan menunjukkan bahwa attention mechanism paling efektif ketika diterapkan pada tahap ekstraksi fitur dasar untuk tugas klasifikasi kematangan buah.

TABEL 2
Hasil pengujian bedasarkan akurasi dan parameter

Skenario	Se-Net (Blocks)	Train Acc (%)	Val Acc (%)	Train Acc (%)	Params
MobileNetV2 (Baseline)	-	93.21 %	95.71 %	95.73 %	2.88 M
All Block	All	91.95 %	95.19 %	94.31 %	2.91 M
Early	1-5	92.87 %	96.79 %	95.55 %	2.88 M
Middle	6-11	92.35 %	96.17 %	95.91 %	2.88 M
Deep	12-17	92.98 %	95.99 %	94.84 %	2.90 M

Selain akurasi dan parameter, pada Tabel 3 terlihat adanya korelasi positif yang signifikan antara penempatan blok *Squeeze-and-Excitation* (SE) dengan peningkatan performa model dibandingkan dengan MobileNetV2 *baseline*. Skenario Early (penggunaan SE pada blok 1-5) muncul sebagai konfigurasi terbaik dengan nilai *Precision*, *Recall*, dan *F1-Score* tertinggi, yakni masing-masing sebesar 0,9684, 0,9679, dan 0,9679. Jika dihubungkan dengan data pada Tabel 2, keunggulan ini juga selaras dengan capaian akurasi validasi tertinggi sebesar 96,79%. Hal ini mengindikasikan bahwa mekanisme *attention* yang diletakkan pada lapisan-lapisan awal jauh lebih efektif dalam mempertajam ekstraksi fitur dasar sebelum informasi tersebut diabstraksi lebih lanjut pada lapisan yang lebih dalam.

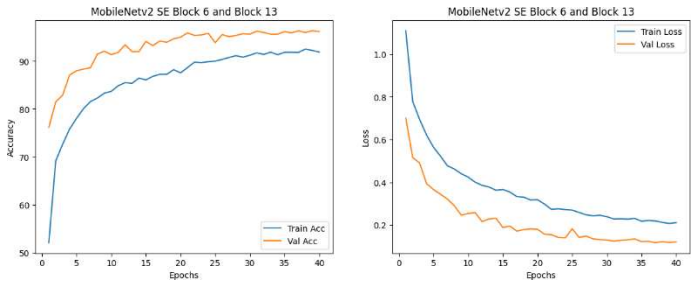
Analisis lebih lanjut mengenai efisiensi menunjukkan bahwa peningkatan performa pada skenario Early dan Middle tidak diikuti oleh peningkatan jumlah parameter yang signifikan, di mana keduanya tetap berada pada angka 2,88 M. Hal ini membuktikan bahwa penempatan blok SE yang strategis dapat meningkatkan kemampuan generalisasi model secara optimal tanpa menambah beban komputasi yang berat. Sebaliknya, skenario Deep dan All Block justru menunjukkan peningkatan jumlah parameter hingga 2,91 M namun dengan hasil metrik yang cenderung menurun. Fenomena ini menunjukkan bahwa penambahan kompleksitas pada lapisan akhir atau pada seluruh blok tidak selalu menjamin hasil yang lebih baik dan justru berisiko menimbulkan redundansi informasi.

Kondisi paling kritis terlihat pada skenario All Block, yang mencatatkan nilai metrik terendah di bawah model *baseline*, dengan *F1-Score* sebesar 0,9518. Sebagai kesimpulan, integrasi SE-Net pada bagian *early stage* arsitektur MobileNetV2 merupakan strategi yang paling ideal bagi sistem ini karena mampu menghasilkan keseimbangan terbaik antara akurasi prediksi dan efisiensi jumlah parameter.

TABEL 3
Hasil metrik dari semua skenario

Skenario	Se-Net (Blocks)	Prec	Rec	F1
MobileNetV2 (Baseline)	-	0.9575	0.9573	0.9572
All Block	All	0.9526	0.9519	0.9518
Early	1-5	0.9684	0.9679	0.9679
Middle	6-11	0.9620	0.9617	0.9617
Deep	12-17	0.9601	0.9599	0.9599

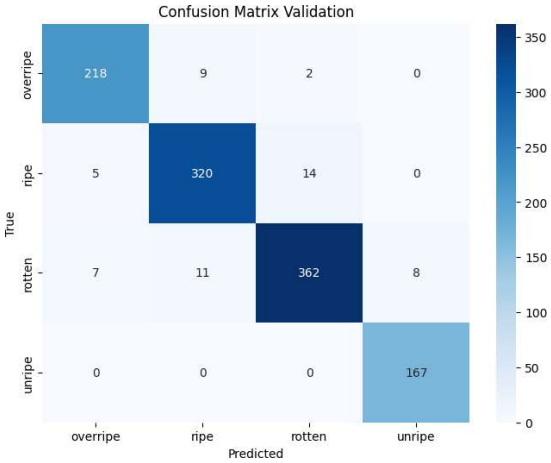
Selain metrik tabel, dinamika pelatihan divisualisasikan pada Gambar 3. Kurva menunjukkan bahwa model usulan memiliki konvergensi yang stabil tanpa divergensi signifikan antara *loss* pelatihan dan validasi, mengindikasikan bahwa model tidak mengalami *overfitting* meskipun dilatih pada data dengan latar belakang yang bervariasi.

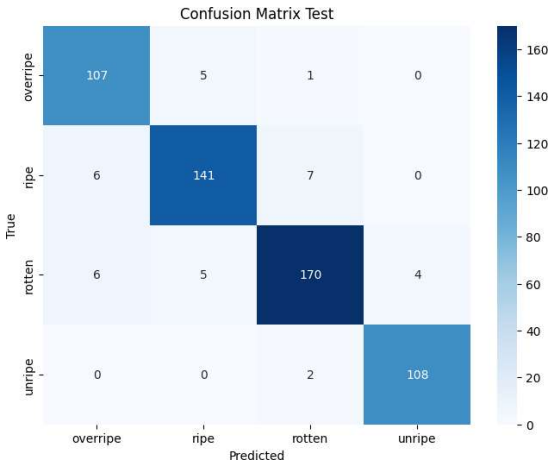


GAMBAR 3
Grafik Pelatihan dan Validasi

4.2 Analisis Hasil Pengujian

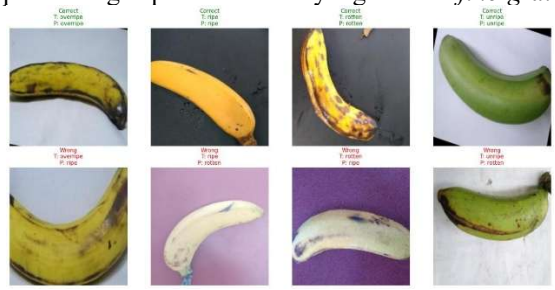
Pada Gambar 4. Memperlihatkan *confusion matrix* pada data *validation* dan *test* menunjukkan bahwa model memiliki performa klasifikasi yang sangat stabil dan konsisten dalam mengenali tingkat kematangan buah. Secara keseluruhan, model mampu mengidentifikasi kelas unripe dengan tingkat akurasi tertinggi bahkan hampir sempurna pada kedua dataset, yang mengindikasikan bahwa fitur visual pada fase mentah memiliki karakteristik yang sangat distingtif dibandingkan fase lainnya. Namun, terdapat sedikit ambiguitas pada klasifikasi kelas ripe, overripe, dan rotten, di mana beberapa sampel saling tertukar, terutama antara kelas *ripe* dan *rotten*. Secara ilmiah, fenomena ini dapat dimaklumi karena pada tahap transisi kematangan menuju pembusukan, terdapat kemiripan detil dan degradasi warna yang cukup tipis sehingga menjadi tantangan tersendiri bagi model dalam proses ekstraksi fitur. Meskipun terdapat sedikit *misclassification* tersebut, dominasi nilai pada diagonal utama di kedua matriks membuktikan bahwa model memiliki kemampuan generalisasi yang baik dan tidak mengalami *overfitting*, sehingga sangat layak untuk diimplementasikan pada data baru yang belum pernah dilihat sebelumnya.





GAMBAR 4
Confusion Matrix data validasi dan tes

Sampel hasil prediksi visual pada Gambar 5 menunjukkan bahwa model klasifikasi berbasis MobileNetV2 ini memiliki performa yang sangat stabil dengan tingkat akurasi tinggi pada seluruh kelas. Kelas unripe secara konsisten menjadi kategori dengan performa terbaik karena fitur warna hijau yang sangat distingtif dibandingkan kelas lainnya, sehingga hampir tidak ditemukan kesalahan prediksi pada kelas ini. Sebaliknya, tantangan klasifikasi muncul pada kelas ripe, overripe, dan rotten, di mana terjadi tumpang tindih prediksi akibat kemiripan fitur dan gradasi warna cokelat pada permukaan kulit buah. Sampel visual menunjukkan bahwa kesalahan prediksi umumnya terjadi pada kondisi transisi, seperti buah *overripe* yang diprediksi sebagai *ripe* atau buah *ripe* yang dianggap *rotten*, yang secara ilmiah disebabkan oleh ambiguitas fitur bintik (*spots*) dan kemiripan persepsi warna pada citra. Secara keseluruhan, konsistensi hasil antara matriks validasi dan pengujian membuktikan bahwa model memiliki kemampuan generalisasi yang handal dan sangat efektif dalam mengenali pola kematangan buah meskipun terdapat tantangan pada fitur-fitur yang bersifat *fine-grained*.



GAMBAR 5
Hasil uji kualitatif

5. Kesimpulan dan Saran

Penelitian ini berhasil meningkatkan performa arsitektur MobileNetV2 dalam mengklasifikasikan tingkat kematangan buah melalui integrasi modul *Squeeze-and-Excitation* (SE) secara selektif. Berdasarkan hasil pengujian, skenario Early (penempatan SE pada blok 1-5) merupakan konfigurasi paling optimal dengan capaian akurasi validasi tertinggi sebesar 96,79% dan nilai *F1-score* 0,9679. Temuan penting dalam riset ini adalah bahwa penambahan mekanisme *attention* pada

lapisan awal cukup efektif dan efisien karena mampu meningkatkan performa tanpa menambah jumlah parameter komputasi (tetap pada 2,88 M), dibandingkan skenario *All Block* yang justru menurunkan performa akibat redundansi fitur. Hasil *confusion matrix* menunjukkan bahwa model memiliki kemampuan generalisasi yang sangat baik, terutama pada kelas *unripe* yang terklasifikasi hampir sempurna. Meskipun terdapat beberapa kesalahan klasifikasi pada kelas *ripe*, *overripe*, dan *rotten* akibat kemiripan detil visual yang sangat halus (*fine-grained*), secara keseluruhan model ini telah memenuhi standar efisiensi dan akurasi untuk kebutuhan klasifikasi otomatis.

Penelitian selanjutnya disarankan untuk menambahkan metode *Explainable AI* seperti Grad-CAM agar proses pengambilan keputusan model dapat divisualisasikan dan dievaluasi secara lebih transparan. Selain itu, implementasi model pada perangkat edge seperti Raspberry Pi atau NVIDIA Jetson perlu dilakukan untuk menguji kinerja dari sisi kecepatan inferensi dan efisiensi komputasi di lingkungan nyata. Pengembangan berikutnya juga dapat diarahkan pada integrasi dengan metode object detection sehingga sistem mampu mendeteksi sekaligus mengklasifikasikan tingkat kematangan beberapa buah dalam satu citra, sehingga lebih relevan untuk aplikasi industri.

REFERENSI

[1] "BANANA Market Review 2023 © iStock Shakeel Sha."

[2] L. Chuquimarca, B. Vintimilla, and S. Velastin, "Banana Ripeness Level Classification using a Simple CNN Model Trained with Real and Synthetic Datasets," Apr. 2025, doi: 10.5220/0011654600003417.

[3] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, and A. Albarelli, "Fruit ripeness classification: A survey," Mar. 01, 2023, *KeAi Communications Co.* doi: 10.1016/j.aiia.2023.02.004.

[4] S. K. Behera, A. K. Rath, and P. K. Sethy, "Maturity status classification of papaya fruits based on machine learning and transfer learning approach," *Information Processing in Agriculture*, vol. 8, no. 2, pp. 244–250, Jun. 2021, doi: 10.1016/j.inpa.2020.05.003.

[5] O. Martínez-Mora *et al.*, "Artificial Vision-Based Dual CNN Classification of Banana Ripeness and Quality Attributes Using RGB Images," *Processes*, vol. 13, no. 7, Jul. 2025, doi: 10.3390/pr13071982.

[6] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Mar. 2019, [Online]. Available: http://arxiv.org/abs/1801.04381

[7] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," May 2019, [Online]. Available: http://arxiv.org/abs/1709.01507

[8] Z. Ullah, M. Hong, T. Mahmood, and J. Kim, "Systematic Integration of Attention Modules into CNNs for Accurate and Generalizable Medical Image Classification," *Mathematics*, vol. 13, no. 22, Nov. 2025, doi: 10.3390/math13223728.

[9] P. Wang, T. Niu, and D. He, "Tomato young fruits detection method under near color background based on improved faster r-cnn with attention mechanism," *Agriculture (Switzerland)*, vol. 11, no. 11, Nov. 2021, doi: 10.3390/agriculture11111059.

[10] Z. Ullah, M. Hong, T. Mahmood, and J. Kim, "Systematic Integration of Attention Modules into CNNs for Accurate and Generalizable Medical Image Diagnosis," Sep. 2025, [Online]. Available: http://arxiv.org/abs/2509.05343

[11] W. Li, Y. Wang, Y. Yu, and J. Liu, "Application of Attention-Enhanced 1D-CNN Algorithm in Hyperspectral Image and Spectral Fusion Detection of Moisture Content in Orah Mandarin (Citrus reticulata Blanco)," *Information (Switzerland)*, vol. 15, no. 7, Jul. 2024, doi: 10.3390/info15070408.

[12] K. Goyal, P. Kumar, and K. Verma, "Tomato ripeness and shelf-life prediction system using machine learning," *Journal of Food Measurement and Characterization*, vol. 18, no. 4, pp. 2715–2730, Apr. 2024, doi: 10.1007/s11694-023-02349-x.

[13] K. Sumathi and V. Vinod, "CLASSIFICATION OF FRUITS RIPENESS USING CNN WITH MULTIVARIATE ANALYSIS BY SGD," *Neural Network World*, vol. 32, no. 6, pp. 319–332, 2022, doi: 10.14311/NNW.2022.32.019.

[14] S. Nuanmeesri, "Enhanced hybrid attention deep learning for avocado ripeness classification on resource constrained devices," *Sci Rep*, vol. 15, no. 1, p. 3719, Dec. 2025, doi: 10.1038/s41598-025- 87173-7.

[15] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability (Switzerland)*, vol. 15, no. 3, Feb. 2023, doi: 10.3390/su15031906.

[16] R. Kurniawan, Samsuryadi, F. S. Mohamad, H. O. L. Wijaya, and B. Santoso, "Classification of palm oil fruit ripeness based on AlexNet deep Convolutional Neural Network," *Sinergi (Indonesia)*, vol. 29, no. 1, pp. 207–220, 2025, doi: 10.22441/sinergi.2025.1.019.

[17] K. K. Yadav and G. Tandan, "FruitNet: Lightweight CNN for High-Throughput Image-Based Fruit Yield Estimation," *SHS Web of Conferences*, p. 2025, 2025, doi: 10.1051/iciaites/2025042701054.

[18] W. Xu, Y. Wan, and D. Zhao, "SFA: Efficient Attention Mechanism for Superior CNN Performance," *Neural Process Lett*, vol. 57, no. 2, Apr. 2025, doi: 10.1007/s11063-025-11748-8.

[19] W. Zhang, "A Fruit Ripeness Detection Method using Adapted Deep Learning-based Approach," 2023. [Online]. Available: www.ijacsa.thesai.org

[20] Z. A. Khan *et al.*, "EA-CNN: Enhanced attention-CNN with explainable AI for fruit and vegetable classification," *Heliyon*, vol. 10, no. 23, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40820.

[21] S.M. Shahriar, "Banana Ripeness Classification Dataset," Kaggle - https://www.kaggle.com/datasets/shahriar26s/banana-ripeness-classification-dataset.