

PENGENALAN TINGKAT KEMATANGAN BUAH PISANG BERBASIS CITRA DENGAN METODE CONVOLUTIONAL NEURAL NETWORK

Moh Raihan Zacky Ramadhan, Anggay Luri Pramana

Informatika, Universitas Nahdlatul Ulama Sidoarjo
Rangkah Kidul, Kec. Sidoarjo, Kabupaten Sidoarjo, Jawa Timur, Indonesia
rehanzacky.rz@gmail.com

ABSTRAK

Pisang merupakan komoditas hortikultura krusial di Indonesia, di mana tingkat kematangannya menjadi indikator utama penentu kualitas pascapanen. Namun, proses penentuan kematangan di lapangan umumnya masih dilakukan secara manual dan subjektif, sehingga sering memicu inkonsistensi klasifikasi yang berdampak pada kerugian ekonomi. Penelitian ini bertujuan untuk mengotomatisasi proses tersebut dengan mengevaluasi dan membandingkan performa tiga arsitektur deep learning, yaitu Convolutional Neural Network (CNN) standar, VGG16, dan ResNet50, dalam mengklasifikasikan citra buah pisang ke dalam empat kategori kematangan. Evaluasi model dilakukan melalui skenario pelatihan 20 dan 50 epoch, dengan kinerja diukur menggunakan confusion matrix berdasarkan metrik Accuracy, Precision, Recall, dan F1-Score. Hasil pengujian pada skenario 50 epoch menunjukkan model CNN standar mencapai akurasi tertinggi sebesar 98,39%, disusul oleh VGG16 dengan 96,62%. Sebaliknya, ResNet50 mengalami underfitting dengan akurasi stagnan di 50,53%. Kesimpulannya, model CNN terbukti lebih efisien dan akurat untuk dataset citra ini dibandingkan arsitektur jaringan yang terlalu kompleks.

Kata kunci : Kematangan buah pisang, Deep Learning, CNN, VGG16, ResNet50, Klasifikasi citra.

1. PENDAHULUAN

Pisang (*Musa spp.*) merupakan salah satu komoditas hortikultura penting di Indonesia dan memiliki tingkat produksi yang tinggi dibandingkan buah lainnya. Produksi pisang secara konsisten menunjukkan kontribusi signifikan terhadap sektor pertanian nasional, khususnya pada varietas Cavendish yang banyak dibudidayakan di berbagai daerah di Indonesia [1].

Selain itu, konsumsi buah dan sayur termasuk pisang masih menjadi bagian penting dalam pola konsumsi masyarakat Indonesia [2]. Tingginya tingkat produksi dan konsumsi ini menuntut adanya sistem pengendalian kualitas yang andal sebelum tahap distribusi ke pasar.

Preferensi konsumen terhadap buah sangat dipengaruhi oleh berbagai faktor visual, seperti warna, tekstur, kesegaran, dan tingkat kematangan [3]. Tingkat kematangan menjadi indikator utama kualitas buah karena berpengaruh langsung terhadap cita rasa, daya simpan, serta nilai jual produk.

Namun demikian, proses penentuan tingkat kematangan buah pisang di lapangan umumnya masih dilakukan secara manual berdasarkan pengamatan visual. Permasalahan utama dari metode konvensional ini adalah sifatnya yang subjektif dan sangat bergantung pada pengalaman individu pengamat, sehingga berpotensi besar menghasilkan inkonsistensi dalam klasifikasi kualitas buah.

Ketidaktepatan dalam proses klasifikasi ini dapat menyebabkan penyortiran dan distribusi buah yang tidak optimal, yang pada akhirnya berdampak pada kerugian ekonomi bagi pelaku industri pertanian. Untuk memecahkan masalah subjektivitas tersebut, rencana penyelesaian yang diajukan dalam penelitian

ini adalah dengan mengimplementasikan teknologi computer vision dan deep learning untuk membangun sistem klasifikasi citra buah yang lebih objektif dan otomatis.

Beberapa penelitian sebelumnya telah menerapkan teknik Convolutional Neural Network (CNN) untuk mendeteksi kualitas dan kesegaran buah berbasis citra digital dengan hasil yang cukup menjanjikan [4]. Selain CNN konvensional, pendekatan transfer learning menggunakan arsitektur yang lebih kompleks seperti VGG dan ResNet juga mulai banyak digunakan dalam berbagai penelitian klasifikasi citra karena diklaim mampu meningkatkan akurasi melalui pemanfaatan bobot pre-trained [5], [6].

Meskipun berbagai penelitian telah mengimplementasikan metode tersebut, studi komparatif yang secara khusus mengevaluasi performa model CNN dasar, VGG16, dan ResNet50 dalam mengenali empat kelas kematangan buah pisang secara spesifik masih sangat terbatas.

Oleh karena itu, penelitian ini bertujuan untuk membandingkan secara komprehensif performa ketiga arsitektur deep learning tersebut dalam mendeteksi tingkat kematangan buah pisang berbasis citra digital. Melalui evaluasi metrik kinerja yang detail pada setiap model, penelitian ini diharapkan dapat memberikan rekomendasi arsitektur mana yang paling efisien, akurat, dan stabil untuk diimplementasikan sebagai solusi sistem penyortiran buah yang otomatis.

2. TINJAUAN PUSTAKA

Penelitian mengenai pengenalan tingkat kematangan buah pisang telah banyak dieksplorasi menggunakan berbagai pendekatan pengolahan citra digital. Salah satu penelitian relevan yang mendeteksi

kematangan pisang Ambon menggunakan metode transformasi ruang warna *Hue, Saturation, Intensity* (HSI) telah dilakukan sebelumnya [7].

Proses pengenalan pada penelitian tersebut dilakukan dengan mengekstraksi fitur warna kulit pisang yang telah dipisahkan dari latar belakangnya (*cropping*). Hasil pengujian menunjukkan tingkat akurasi sebesar 85%, yang membuktikan bahwa perbedaan nilai *Hue* dan *Saturation* dapat membedakan pisang mentah dan matang. Meskipun demikian, metode konvensional ini memiliki keterbatasan yang signifikan karena sangat rentan terhadap variasi pencahayaan dan perubahan latar belakang citra.

Oleh karena itu, penelitian yang diusulkan ini mengembangkan pendekatan yang lebih tangguh dengan memanfaatkan arsitektur *deep learning* (CNN, VGG, dan ResNet) guna meningkatkan akurasi, kecepatan, dan keandalan sistem klasifikasi dibandingkan metode transformasi warna tradisional.

Implementasi *deep learning* untuk klasifikasi tingkat kematangan buah telah dibuktikan efektivitasnya oleh Yanto dkk. [8] pada kasus buah jeruk manis. Menggunakan arsitektur *Convolutional Neural Network* (CNN) dasar, penelitian tersebut membedakan jeruk berkondisi baik dan busuk berdasarkan tekstur dan kecerahan warna.

Melalui penerapan prapemrosesan berupa *resize* dan normalisasi citra, model tersebut berhasil mencapai akurasi 96% pada data latih dan 92% pada data uji. Namun, penelitian tersebut masih terbatas pada penggunaan satu arsitektur tunggal dan belum mengeksplorasi model turunan CNN lainnya.

Mengacu pada celah tersebut, penelitian ini memperluas cakupan evaluasi dengan melakukan analisis komparatif secara langsung terhadap performa arsitektur CNN standar, VGG, dan ResNet untuk mencari model yang paling efisien dalam mengenali tingkat kematangan buah pisang.

Pentingnya melakukan perbandingan antararsitektur *deep learning* juga telah disoroti dalam penelitian Setiawan [9] yang mengevaluasi berbagai model termasuk AlexNet, VGG16, VGG19, ResNet50, hingga Inception untuk klasifikasi citra medis pada retina.

Hasil penelitian tersebut menunjukkan bahwa arsitektur yang lebih dalam seperti VGG16 dan VGG19 mampu memberikan performa terbaik dengan akurasi 92,31%. Temuan ini membuktikan bahwa penambahan lapisan (*layer*) sangat memengaruhi kemampuan model dalam mengekstrak fitur citra secara detail.

Mengadopsi landasan komparatif tersebut, penelitian ini menerapkan konsep perbandingan serupa namun dialihkan pada domain objek alami, yakni buah pisang.

Berbeda dengan penelitian sebelumnya, analisis pada penelitian ini tidak hanya berfokus pada akurasi global, tetapi juga mengevaluasi efisiensi komputasi serta metrik tambahan seperti *Precision*, *Recall*, dan

F1-score untuk memperoleh pemahaman kinerja yang lebih komprehensif.

Terkait penerapan metrik evaluasi yang detail untuk objek pertanian, Palupi dkk. [10] telah mendemonstrasikan evaluasi pendeteksian objek varian jahe menggunakan algoritma YOLOv5.

Model yang dilatih selama 100 *epoch* tersebut berhasil memperoleh nilai *Precision* 0,94 dan *Recall* 0,93, membuktikan efektivitas metode konvolusi dalam mengenali objek pertanian dengan variasi bentuk dan warna yang kompleks.

Penelitian tersebut menegaskan pentingnya validasi menggunakan parameter pengujian yang lengkap, meskipun hasilnya masih dipengaruhi oleh variasi sudut pengambilan gambar. Berangkat dari evaluasi tersebut, penelitian yang dikembangkan saat ini mengadaptasi standarisasi pengujian yang ketat untuk mengukur performa CNN, VGG, dan ResNet.

Pendekatan ini diharapkan tidak hanya mengatasi kendala variasi visual buah di dunia nyata, tetapi juga memberikan rekomendasi arsitektur mana yang secara nyata paling andal untuk diimplementasikan.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental dengan tujuan membandingkan performa beberapa arsitektur *Convolutional Neural Network* (CNN) dalam melakukan klasifikasi tingkat kematangan buah pisang berbasis citra digital. Model yang digunakan dalam penelitian ini meliputi CNN sebagai model dasar (*baseline*), serta dua model *transfer learning* yaitu VGG16 dan ResNet50.

3.1. Dataset

Data citra yang digunakan pada tahapan pelatihan dan pengujian model dalam penelitian ini bersumber dari repositori publik Kaggle.

Dataset sekunder ini memuat kumpulan citra buah pisang yang telah dianotasi dan dikategorikan ke dalam empat kelas tingkat kematangan, yakni mentah, matang, terlalu matang, dan busuk. Penggunaan *dataset* dengan klasifikasi visual seperti ini sangat krusial dalam eksperimen visi komputer.

Hal ini sejalan dengan penelitian yang dilakukan oleh Saranya dkk. [11], yang menegaskan bahwa variasi citra pada setiap fase kematangan sangat dibutuhkan oleh model jaringan konvolusional (seperti CNN dan ResNet) untuk dapat mengekstrak fitur warna dan tekstur secara akurat dari kulit buah.



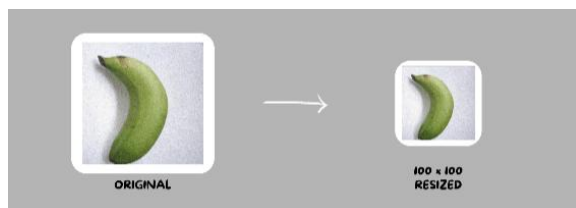
Gambar 1. *Dataset* buah pisang

Seperti pada gambar 1, *Dataset* terdiri dari 13.478 citra buah pisang yang terbagi ke dalam empat kelas, yaitu mentah, matang, terlalu matang, dan busuk. *Dataset* dibagi menjadi tiga bagian, yaitu 11.793 citra sebagai data latih (*training set*), 1.123 citra sebagai data validasi (*validation set*), dan 562 citra sebagai data uji (*testing set*).

Pembagian data dilakukan secara proporsional untuk menjaga keseimbangan jumlah data pada setiap kelas sehingga tidak terjadi bias selama proses pelatihan model.

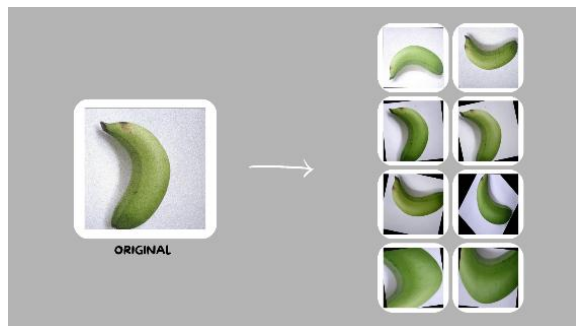
3.2. Preprocessing Data

Sebelum digunakan dalam pelatihan model, seluruh citra melalui tahap *preprocessing* yang meliputi *resize*, normalisasi, dan augmentasi data.



Gambar 2. *Preprocessing resize*

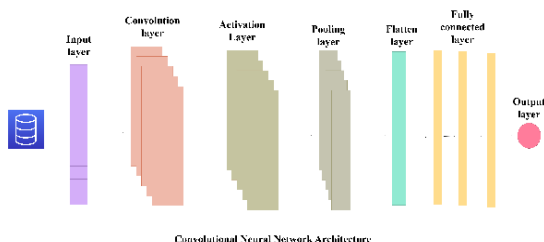
Seperti pada gambar 2, *Resize* dilakukan untuk menyeragamkan ukuran citra menjadi 100×100 piksel. Normalisasi dilakukan dengan mengubah nilai piksel ke dalam rentang 0–1.



Gambar 3. Augmentation data

Selain itu, dilakukan augmentasi data berupa rotasi, flipping, dan zoom seperti pada gambar 3, untuk meningkatkan variasi data dan mengurangi risiko *overfitting* selama proses pelatihan.

3.3. Arsitektur Model



Gambar 4. Arsitektur CNN

Berdasarkan Gambar 4, arsitektur *Convolutional Neural Network* (CNN) yang digunakan dalam penelitian ini direpresentasikan secara sekuensial. Untuk memberikan pemahaman mendalam mengenai mekanisme pengenalan dan klasifikasi citra pada model tersebut, struktur dasar CNN dibangun melalui serangkaian lapisan utama yang saling terhubung [12].

Proses komputasi diawali dari sisi kiri dengan *Input Layer* yang berfungsi menerima representasi data citra mentah. Selanjutnya, data diproses melalui *Convolution Layer* untuk mengekstraksi peta fitur (*feature maps*) secara spasial, yang kemudian diikuti oleh *Activation Layer* (seperti ReLU) guna menyematkan fungsi non-linearitas ke dalam jaringan.

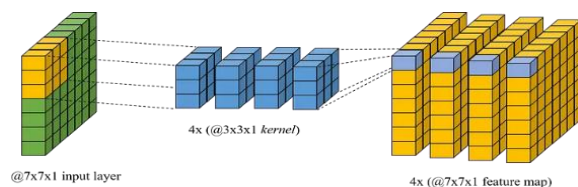
Setelah itu, *Pooling Layer* bertugas untuk melakukan *down-sampling* atau mereduksi dimensi matriks fitur agar proses komputasi lebih efisien. Fitur tingkat tinggi yang telah direduksi tersebut kemudian diratakan melalui *Flatten Layer* menjadi susunan vektor satu dimensi (1D).

Rangkaian vektor ini lalu diteruskan secara menyeluruh ke *Fully connected layer* untuk proses inferensi akhir, yang bermuara pada *Output layer* untuk menghasilkan nilai probabilitas klasifikasi tingkat kematangan buah [12].

3.4. Input layer

Lapisan ini merupakan pintu masuk data citra ke dalam jaringan. Citra buah pisang hasil *Preprocessing* dimasukkan dengan ukuran 100×100 piksel dan tiga kanal warna (RGB). Setiap nilai piksel direpresentasikan dalam bentuk tensor dengan dimensi (100, 100, 3).

3.5. Convolution Layer



Gambar 5. Convolution layer

Seperti pada gambar 5, lapisan konvolusi berfungsi mengekstraksi fitur-fitur penting dari citra, seperti pola warna dan tekstur kulit pisang yang menjadi indikator kematangan. Setiap filter kernel melakukan operasi konvolusi terhadap citra input untuk menghasilkan *feature map*.

Secara matematis, operasi konvolusi dapat dinyatakan sebagai:

$$(I * K)(x, y) = \sum_m \sum_n I(x + m, y + n) K(m, n) \quad (1)$$

di mana I adalah citra input dan K adalah kernel konvolusi. Semakin banyak filter yang digunakan, semakin kompleks fitur yang dapat dipelajari oleh model.

3.6. Activation Layer (ReLU)

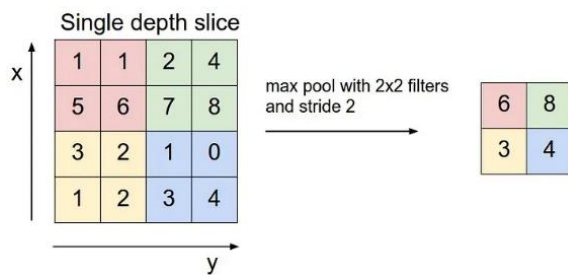
Setelah proses konvolusi, hasilnya diteruskan ke fungsi aktivasi ReLU (Rectified Linear Unit) untuk menambahkan sifat non-linearitas pada model. Fungsi ini mengubah semua nilai negatif menjadi nol dengan rumus:

$$f(x) = \max(0, x) \quad (2)$$

sehingga model dapat mempelajari pola kompleks yang tidak linear.

3.7. Pooling Layer

Pooling Layer berfungsi untuk mengurangi ukuran dimensi (downsampling) dari feature map yang dihasilkan oleh lapisan konvolusi, tanpa menghilangkan informasi penting. Tujuan utama dari tahap ini adalah untuk mengurangi jumlah parameter, mempercepat proses komputasi, serta meningkatkan kemampuan generalisasi model terhadap perubahan posisi atau rotasi objek pada citra.



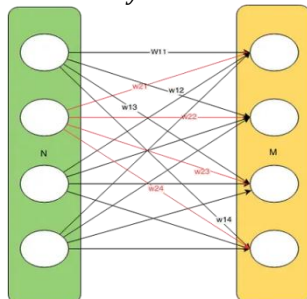
Gambar 6. Pooling layer

Gambar 6 menunjukkan contoh proses Max Pooling dengan filter berukuran 2×2 dan langkah perpindahan (stride) sebesar 2. Feature map pada sisi kiri memiliki ukuran 4×4 , dan setiap kotak warna mewakili area yang akan diproses oleh filter pooling. Filter berukuran 2×2 akan bergerak melintasi citra dengan melompat dua piksel setiap kali (sesuai dengan nilai stride).

3.8. Flattening Layer

Setelah beberapa convolution layer dan pooling, hasil feature map diubah menjadi vektor satu dimensi melalui proses flattening. Langkah ini menghubungkan fitur-fitur yang telah diekstraksi ke dalam lapisan klasifikasi (*fully connected layer*).

3.9. Fully connected layer



Gambar 7. Fully connected layer

Lapisan ini berfungsi menggabungkan seluruh fitur yang telah diekstraksi untuk menentukan kelas keluaran seperti pada gambar 7. Setiap neuron dalam lapisan ini terhubung ke semua neuron pada lapisan sebelumnya dan memproses bobot untuk menghasilkan nilai akhir yang akan diklasifikasikan.

3.10. Output layer

Lapisan terakhir menghasilkan tiga kelas keluaran sesuai dengan tingkat kematangan pisang, yaitu mentah, matang, terlalu matang dan busuk. Fungsi aktivasi yang digunakan adalah Softmax, yang mengubah output menjadi probabilitas dengan rumus:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (3)$$

3.11. Pelatihan Model

Penelitian ini menggunakan tiga arsitektur model, yaitu:

- Model CNN dibangun dari beberapa convolution layer, pooling layer, dan fully connected layer yang dilatih dari awal menggunakan dataset penelitian.
- Model VGG16 menggunakan pendekatan transfer learning dengan memanfaatkan bobot pre-trained dari dataset ImageNet. Lapisan konvolusi dibekukan (freeze), sedangkan lapisan klasifikasi dilatih ulang.
- Model ResNet50 juga menggunakan transfer learning dengan arsitektur residual network yang memiliki skip connection untuk meningkatkan stabilitas pelatihan jaringan dalam.

Ketiga model dilatih menggunakan optimizer Adam dan fungsi loss categorical cross-entropy. Proses pelatihan dilakukan dalam dua skenario, yaitu 20 epoch dan 50 epoch, untuk melihat pengaruh jumlah epoch terhadap performa model. Parameter batch size dan learning rate dibuat sama pada seluruh model agar hasil perbandingan lebih objektif.

3.12. Evaluasi Model

Evaluasi performa model secara objektif dilakukan menggunakan data uji (testing set) yang sama sekali tidak dilibatkan selama proses pelatihan guna mengukur kemampuan generalisasi jaringan. Metrik evaluasi yang digunakan untuk mengukur tingkat keandalan klasifikasi meliputi Accuracy, Precision, Recall, dan F1-score [13]. Kalkulasi nilai dari metrik-metrik tersebut didasarkan pada empat parameter utama, yaitu True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN), yang diekstrak secara langsung dari confusion matrix. Penggunaan confusion matrix dalam mengevaluasi model klasifikasi machine learning sangat esensial karena memberikan rincian performa per kelas, sehingga memungkinkan peneliti menganalisis letak kesalahan klasifikasi (misclassification) secara terperinci [13].

Adapun formulasi matematis standar yang diturunkan dari *confusion matrix* untuk menghitung masing-masing metrik evaluasi tersebut adalah sebagai berikut [14]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

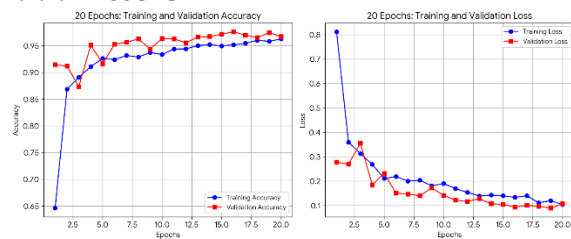
4. HASIL DAN PEMBAHASAN

4.1. Hasil Pelatihan dan Pengujian Model

Pada penelitian ini, dilakukan pengujian terhadap tiga model *deep learning*, yaitu *Convolutional Neural Network* (CNN), VGG16, dan ResNet50 untuk melakukan klasifikasi tingkat kematangan buah. Setiap model dilatih menggunakan dua skenario jumlah *epoch*, yaitu 20 *epoch* dan 50 *epoch*, dengan tujuan untuk melihat pengaruh jumlah *epoch* terhadap performa model.

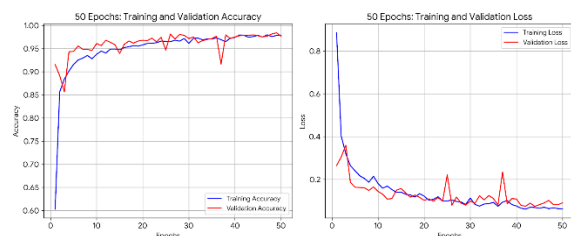
Evaluasi performa dilakukan menggunakan data uji (*testing set*) yang tidak terlibat dalam proses pelatihan. Metrik utama yang digunakan pada tahap ini adalah akurasi (*accuracy*) untuk menggambarkan kemampuan model dalam mengklasifikasikan data secara benar.

4.1.1. Model CNN



Gambar 8. Grafik pelatihan CNN 20 *epoch*

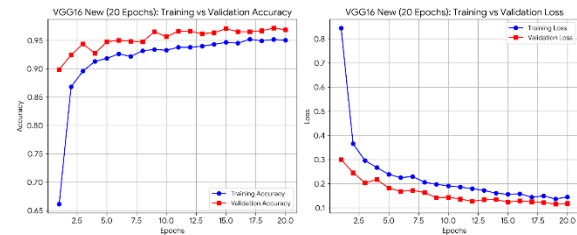
Pada pelatihan 20 *epoch*, model mencapai akurasi pengujian sebesar 97,15%. Setelah dilakukan pelatihan ulang dengan 50 *epoch*, akurasi meningkat menjadi 98,40%.



Gambar 9. Grafik pelatihan CNN 50 *epoch*

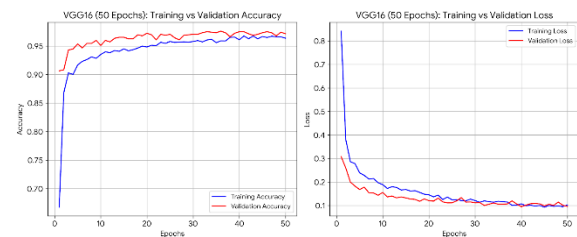
Gambar 9, menunjukkan bahwa penambahan jumlah *epoch* meningkatkan performa, yang mengindikasikan bahwa model CNN masih dapat melakukan pembelajaran lebih optimal hingga *epoch* ke-50 tanpa mengalami *overfitting* yang signifikan.

4.1.2. Model VGG16



Gambar 10. Grafik pelatihan VGG 20 *epoch*

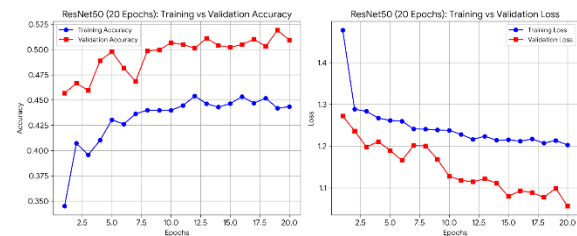
Pada pelatihan 20 *epoch*, model menghasilkan akurasi pengujian sebesar 95,91%. Setelah dilatih sebanyak 50 *epoch*, akurasi meningkat menjadi 96,62%.



Gambar 11. Grafik pelatihan VGG 50 *epoch*

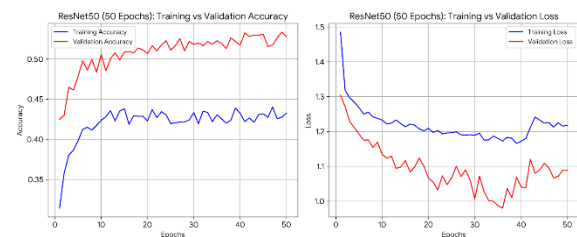
Meskipun peningkatannya tidak sebesar pada model CNN, hasil ini menunjukkan bahwa VGG16 tetap memperoleh manfaat dari penambahan jumlah *epoch*, namun cenderung lebih cepat mencapai kondisi konvergen.

4.1.3. ResNet50



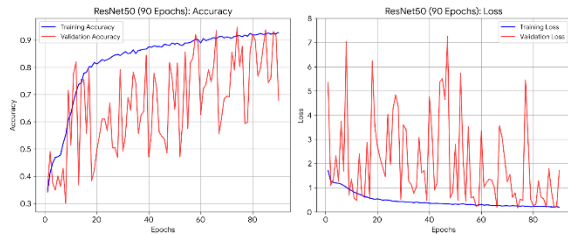
Gambar 12. Grafik pelatihan ResNet 20 *epoch*

Berbeda dengan CNN dan VGG16, model ResNet50 menunjukkan performa yang relatif rendah. Pada pelatihan 20 *epoch*, ResNet50 menghasilkan akurasi sebesar 50,35%, dan setelah ditingkatkan menjadi 50 *epoch* hanya mengalami peningkatan sangat kecil menjadi 50,53%.



Gambar 13. Grafik pelatihan ResNet 50 *epoch*

Gambar 13, menunjukkan bahwa penambahan jumlah *epoch* tidak memberikan peningkatan signifikan terhadap performa ResNet50. Hal ini diduga disebabkan oleh ketidaksesuaian antara kompleksitas arsitektur ResNet50 dengan ukuran *dataset* serta karakteristik citra yang digunakan.



Gambar 14. Grafik pelatihan ResNet 90 epoch

Setelah dilakukan penyesuaian dengan cara menambah lagi jumlah *epoch* ke 90, merubah input dari yang tadinya 100x100 menjadi 224x224, menurunkan *learning rate* menjadi 0.00001, performa pelatihan model menunjukkan peningkatan yang signifikan seperti yang diilustrasikan pada Gambar 14.

Namun, grafik validasi sangat fluktuatif atau tidak stabil. Ketidakstabilan ini mengindikasikan bahwa arsitektur ResNet50 kesulitan untuk melakukan generalisasi dengan baik pada *dataset* ini, sehingga model tersebut kurang dapat diandalkan karena tingkat akurasi yang tidak konsisten.

4.2. Confusion Matrix

Pada penelitian ini, *confusion matrix* dibagi ke dalam dua skenario utama, yaitu pelatihan dengan 20 *epoch* dan 50 *epoch*. Hasil perbandingan metrik dari kedua skenario tersebut selanjutnya dipaparkan secara komprehensif melalui tabel di bawah ini

Tabel 1. *Confusion matrix* CNN

Skenario	Kelas	TP	FP	FN	TN
20 epoch	Busuk	178	11	7	366
	Matang	137	4	17	404
	Mentah	106	1	4	451
	Terlalu Matang	111	14	2	435
50 epoch	Busuk	179	3	6	374
	Matang	153	2	1	406
	Mentah	108	3	2	449
	Terlalu Matang	113	1	0	448

Berdasarkan Tabel 1, model CNN menunjukkan performa yang sangat baik pada seluruh kelas. Nilai TP yang tinggi dan nilai FP serta FN yang rendah menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Kesalahan klasifikasi relatif kecil dan umumnya terjadi pada kelas yang memiliki karakteristik visual yang mirip.

Berdasarkan Tabel 2, Model VGG16 menunjukkan performa yang cukup baik, namun masih terdapat beberapa kesalahan klasifikasi terutama pada kelas terlalu matang dan busuk. Hal ini

mengindikasikan bahwa meskipun VGG16 mampu mengenali pola citra dengan baik, masih terdapat ambiguitas visual pada beberapa kelas yang menyebabkan kesalahan prediksi.

Tabel 2. *Confusion matrix* VGG

Skenario	Kelas	TP	FP	FN	TN
20 epoch	Busuk	175	6	10	371
	Matang	148	9	6	399
	Mentah	107	2	3	450
	Terlalu Matang	105	10	8	439
50 epoch	Busuk	179	5	6	372
	Matang	153	9	1	399
	Mentah	106	1	4	451
	Terlalu Matang	105	4	8	445

Tabel 3. *Confusion matrix* ResNet50

Skenario	Kelas	TP	FP	FN	TN
20 epoch	Busuk	141	179	44	198
	Matang	136	106	18	302
	Mentah	0	0	110	452
	Terlalu Matang	0	0	113	449
50 epoch	Busuk	166	136	19	241
	Matang	118	42	36	366
	Mentah	0	0	110	452
	Terlalu Matang	0	0	113	449

Berdasarkan Tabel 3, dapat dilihat bahwa model ResNet50 memiliki performa yang rendah. Model tidak mampu mengklasifikasikan kelas mentah dan terlalu matang dengan benar, yang ditunjukkan oleh nilai TP sebesar 0 pada kedua kelas tersebut. Hal ini menunjukkan bahwa ResNet50 mengalami kesulitan dalam mempelajari representasi fitur yang relevan untuk seluruh kelas.

Pada klasifikasi multikelas, nilai *True Negative* (TN) cenderung sangat besar dibandingkan TP, FP, dan FN, sehingga pada banyak penelitian hanya ditampilkan nilai TP, FP, dan FN. Namun, pada penelitian ini TN tetap disajikan untuk memberikan gambaran evaluasi yang lebih lengkap.

4.3. Metrik Evaluasi

Untuk mengukur kinerja masing-masing model dalam melakukan klasifikasi tingkat kematangan buah pisang, penelitian ini menggunakan empat metrik evaluasi utama, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

Accuracy digunakan sebagai metrik global untuk mengevaluasi kinerja model secara keseluruhan. Berdasarkan hasil pengujian pada skenario 20 *epoch* dan 50 *epoch*, model CNN mencatatkan peningkatan *accuracy* dari 97.15% menjadi 98.40%. Model VGG16 juga menunjukkan performa yang stabil dengan peningkatan dari 95.91% menjadi 96.62%. Di sisi lain, model ResNet50 memberikan hasil terendah dengan tingkat *accuracy* 50.35% pada 20 *epoch* dan hanya mencapai 50.53% pada 50 *epoch*.

Tabel 4. Metrik Evaluasi CNN

Skenario	Kelas	Precision	Recall	F1-Score
20 epoch	Busuk	0.942	0.962	0.952
	Matang	0.972	0.890	0.929
	Mentah	0.991	0.964	0.977
	Terlalu Matang	0.888	0.982	0.933
	Rata-rata	0.948	0.949	0.948
50 epoch	Busuk	0.983	0.967	0.975
	Matang	0.987	0.993	0.990
	Mentah	0.973	0.982	0.977
	Terlalu Matang	0.991	1.000	0.996
	Rata-rata	0.984	0.986	0.984

Berdasarkan Tabel 4, model CNN menunjukkan performa yang sangat baik. Nilai *precision* dan *recall* yang tinggi pada setiap kelas menunjukkan bahwa model mampu melakukan prediksi dengan tingkat ketepatan yang sangat baik. Nilai *F1-score* yang mendekati 1 pada seluruh kelas menunjukkan bahwa model CNN memiliki keseimbangan yang baik antara *precision* dan *recall*, sehingga dapat dikatakan memiliki performa klasifikasi yang sangat stabil.

Tabel 5. Metrik Evaluasi VGG16

Skenario	Kelas	Precision	Recall	F1-Score
20 epoch	Busuk	0.967	0.946	0.956
	Matang	0.943	0.961	0.952
	Mentah	0.982	0.973	0.977
	Terlalu Matang	0.913	0.929	0.921
	Rata-rata	0.951	0.952	0.952
50 epoch	Busuk	0.973	0.967	0.970
	Matang	0.944	0.993	0.968
	Mentah	0.991	0.964	0.977
	Terlalu Matang	0.963	0.929	0.946
	Rata-rata	0.968	0.963	0.965

Berdasarkan Tabel 5, model VGG16 memberikan performa yang baik, namun masih terdapat penurunan performa pada kelas terlalu matang. Hal ini terlihat dari nilai *recall* dan *F1-score* yang lebih rendah dibandingkan kelas lainnya. Meskipun demikian, secara keseluruhan VGG16 tetap memiliki performa yang cukup stabil dan mampu melakukan klasifikasi dengan tingkat ketepatan yang tinggi.

Berdasarkan Tabel 6, dapat dilihat bahwa model ResNet50 memiliki performa yang rendah. Nilai *precision*, *recall*, dan *F1-score* bernilai nol pada kelas mentah dan terlalu matang, yang menunjukkan bahwa model gagal mengenali kedua kelas tersebut. Hal ini menyebabkan nilai rata-rata metrik evaluasi menjadi sangat rendah, sehingga performa keseluruhan ResNet50 jauh berada di bawah CNN dan VGG16.

Tabel 6. Metrik Evaluasi ResNet50

Skenario	Kelas	Precision	Recall	F1-Score
20 epoch	Busuk	0.441	0.762	0.558
	Matang	0.562	0.883	0.687
	Mentah	0.0000	0.0000	0.0000
	Terlalu Matang	0.0000	0.0000	0.0000
	Rata-rata	0.251	0.411	0.311
50 epoch	Busuk	0.550	0.897	0.682
	Matang	0.737	0.766	0.751
	Mentah	0.000	0.000	0.000
	Terlalu Matang	0.000	0.000	0.000
	Rata-rata	0.322	0.416	0.358

4.4. Perbandingan Model

Pada bagian ini dilakukan perbandingan performa antara tiga model yang digunakan, yaitu CNN murni, VGG16, dan ResNet50, berdasarkan metrik evaluasi yang telah diperoleh pada bagian sebelumnya. Perbandingan difokuskan pada nilai *accuracy*, *precision*, *recall*, dan *F1-score* untuk melihat model mana yang paling optimal dalam mengklasifikasikan tingkat kematangan buah pisang.

Tabel 7. Perbandingan Performa Model

Skenario	Model	Accuracy	Precision	Recall	F1-Score
20 epoch	CNN	97.15%	0.948	0.949	0.948
	VGG16	95.91%	0.951	0.952	0.952
	ResNet50	50.35%	0.251	0.411	0.311
50 epoch	CNN	98.39%	0.984	0.986	0.984
	VGG16	96.62%	0.968	0.963	0.965
	ResNet50	50.53%	0.322	0.416	0.358

Berdasarkan Tabel 7, disajikan evaluasi komprehensif mengenai perbandingan performa tiga arsitektur *deep learning*, yaitu CNN, VGG16, dan ResNet50, yang diuji melalui dua skenario pelatihan: 20 epoch dan 50 epoch. Tabel tersebut memaparkan empat metrik evaluasi utama, yakni *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, guna menilai tingkat keandalan dan keseimbangan prediksi dari masing-masing model secara mendalam.

Pada skenario awal dengan durasi pelatihan 20 epoch, model CNN berhasil memimpin pada metrik akurasi global dengan pencapaian sebesar 97.15%. Meskipun demikian, jika ditinjau lebih rinci, model VGG16 menunjukkan prediksi antar kelas yang sedikit lebih baik, hal ini terbukti dari nilai *Precision* (0.951), *Recall* (0.952), dan *F1-Score* (0.952) yang mengungguli metrik spesifik milik CNN pada fase ini.

Sebaliknya, model ResNet50 memperlihatkan performa yang sangat tertinggal. Tingkat *accuracy* ResNet50 hanya menyentuh angka 50.35%, dengan nilai *F1-Score* sangat rendah di angka 0.311. Temuan awal ini mengindikasikan bahwa pada 20 epoch pertama, arsitektur ResNet50 yang kompleks belum mampu mengekstrak pola fitur *dataset* secara efektif dibandingkan dua arsitektur lainnya.

Ketika durasi pelatihan diperpanjang menjadi 50 *epoch*, mayoritas model menunjukkan respons positif berupa peningkatan performa belajar. Model CNN mendapatkan *accuracy* mencapai 98.39%. Model CNN pada 50 *epoch* juga mendominasi dengan mencatatkan nilai *Precision* 0.984, *Recall* 0.986, dan *F1-Score* 0.984 yang nyaris sempurna. Model VGG16 turut memperlihatkan peningkatan ke angka *accuracy* 96.62% dan *F1-Score* 0.965, membuktikan kapasitasnya sebagai arsitektur alternatif yang sangat kokoh.

Namun, perpanjangan durasi pelatihan tidak memberikan dampak perbaikan yang substansial bagi model ResNet50. Peningkatan dari 20 menjadi 50 *epoch* hanya menaikkan nilai *accuracy* model ini sebesar 0.18% (berubah tipis menjadi 50.53%). Metrik spesifiknya pun masih sangat rendah, dengan *Precision* di angka 0.322 dan *F1-Score* sebesar 0.358.

Kesimpulannya, data pada tabel ini membuktikan bahwa arsitektur CNN dan VGG16 jauh lebih cerdas dan optimal dalam melakukan generalisasi pada studi kasus pengenalan kematangan pisang ini, mengalahkan arsitektur ResNet50 yang dinilai terlalu rumit namun tidak menghasilkan akurasi yang valid secara konsisten.

4.5. Pembahasan

Pada bagian ini, dibahas secara mendalam hasil eksperimen yang telah diperoleh dari pengujian tiga arsitektur *deep learning* yakni CNN, VGG16, dan ResNet50 dalam melakukan klasifikasi tingkat kematangan buah pisang.

Berdasarkan hasil pelatihan dan pengujian, model CNN standar menunjukkan performa terbaik dengan tingkat *accuracy* mencapai 98.39%, diikuti oleh VGG16 dengan 96.62%, dan ResNet50 dengan 50.53%. Tingginya performa CNN ini mengonfirmasi bahwa arsitektur konvensional yang relatif sederhana sudah sangat efektif dalam mengekstraksi fitur visual esensial pada citra objek agrikultur.

Hal ini sejalan dengan penelitian terdahulu [15], yang menyatakan bahwa fitur klasifikasi seperti perubahan warna kulit, tekstur, dan pola degradasi kematangan buah pisang dapat dipelajari secara presisi menggunakan *shallow learning* (pembelajaran dangkal) maupun CNN tanpa memerlukan lapisan jaringan yang terlalu dalam.

Di sisi lain, model VGG16 yang merupakan arsitektur berbasis *transfer learning* juga memberikan performa yang kompetitif. Namun, fakta bahwa hasilnya tidak mampu melampaui CNN dasar menyiratkan sebuah karakteristik penting dalam pemodelan *machine learning*.

Seperti yang dijelaskan dalam literatur [16], penggunaan arsitektur turunan yang lebih dalam tidak selalu berbanding lurus dengan peningkatan akurasi; bahkan varian arsitektur kompleks seperti ResNet sering kali menjadi *overkill* (berlebihan) apabila diaplikasikan pada kasus yang lebih sederhana atau sensitif terhadap sumber daya.

Fenomena tersebut semakin terbukti secara ekstrem pada hasil yang ditunjukkan oleh model ResNet50. Berdasarkan analisis *confusion matrix* dan metrik evaluasi, model ini gagal total dalam mengenali kelas "mentah" dan "terlalu matang" (nilai *True Positive* = 0).

Kondisi ini mengindikasikan bahwa ResNet50 mengalami kesulitan besar dalam mempertahankan representasi fitur yang relevan. Kegagalan ini sejalan dengan temuan riset terbaru [17] yang menegaskan bahwa arsitektur model yang terlalu dalam (*excessively deep models*) justru membebani komputasi secara sia-sia dan memicu terjadinya kegagalan generalisasi (seperti *underfitting* atau *overfitting*) ketika dipaksakan untuk belajar pada *dataset* berskala kecil yang parameter variasinya tidak terlalu masif.

Secara keseluruhan, hasil eksperimen ini membuktikan bahwa pemilihan arsitektur model tidak boleh hanya berpatokan pada tren kedalaman jaringan, melainkan harus disesuaikan secara proporsional dengan tingkat kompleksitas dan karakteristik *dataset* yang digunakan.

Pemodelan yang kompleks tidak menjamin hasil yang superior, dan pada banyak kasus klasifikasi citra, arsitektur shallow CNN yang lebih ringkas justru mampu mengungguli deep neural networks dengan memberikan efisiensi komputasi dan performa generalisasi yang jauh lebih optimal [18].

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa kedalaman suatu arsitektur *deep learning* tidak berbanding lurus dengan tingginya performa pada tugas klasifikasi tingkat kematangan buah pisang.

Berdasarkan pengujian terhadap tiga model yakni CNN standar, VGG16, dan ResNet50 pada skenario 20 dan 50 *epoch*, model CNN terbukti paling superior dan stabil dengan capaian *accuracy* 98.39%, disusul secara kompetitif oleh VGG16 di angka 96.61%. Sebaliknya, arsitektur ResNet50 yang kompleks justru mengalami *underfitting* dengan akurasi yang stagnan di kisaran 50%.

Analisis *confusion matrix* turut menegaskan bahwa model CNN mampu melakukan generalisasi dengan presisi tinggi, meskipun masih terdapat minoritas kesalahan prediksi pada kelas dengan kemiripan fitur visual yang tinggi, seperti fase "matang" dan "terlalu matang".

Temuan ini mengonfirmasi secara empiris bahwa arsitektur yang lebih ringkas namun disesuaikan dengan karakteristik dan ukuran *dataset* mampu mengekstrak fitur secara jauh lebih efisien dibandingkan arsitektur yang terlalu kompleks.

Berdasarkan kesimpulan tersebut, penelitian selanjutnya disarankan untuk menggunakan *dataset* dengan variasi kondisi yang lebih ekstrem, seperti penambahan tingkat pencahayaan yang berbeda, sudut pengambilan gambar, serta latar belakang citra yang

heterogen guna meningkatkan ketahanan (*robustness*) model di kondisi lapangan yang sesungguhnya.

Selain itu, optimalisasi hyperparameter menggunakan teknik *fine-tuning* yang presisi seperti penerapan *learning rate scheduler* sangat direkomendasikan apabila hendak meneliti model *transfer learning* berlapis dalam seperti ResNet50 agar seluruh jaringan dapat beradaptasi mencapai titik konvergensi maksimal.

Terakhir, sebagai tindak lanjut penerapan praktis, integrasi arsitektur terbaik (CNN/VGG16) ke dalam antarmuka aplikasi cerdas berbasis web maupun mobile sangat direkomendasikan. Implementasi ini diharapkan dapat memberikan solusi nyata bagi sektor agrikultur dan distribusi buah untuk mengotomatisasi proses penyortiran kualitas panen secara real-time.

DAFTAR PUSTAKA

- [1] R. Andriani and M. S. Rahayu, "Pertumbuhan dan Produksi Pisang Cavendish Dataran Tinggi di Blitar, Jawa Timur," *Buletin Agrohorti*, vol. 11, no. 3, pp. 323–330, Sep. 2023, doi: 10.29244/agrob.v11i3.49366.
- [2] M. Qibtiyah, C. Rosidati, and M. H. Siregar, "Perilaku Konsumsi Buah dan Sayur Pada Remaja," *Jurnal Gizi Kerja dan Produktivitas*, vol. 2, no. 2, p. 51, Nov. 2021, doi: 10.52742/jgkp.v2i2.12760.
- [3] F. Alham and F. H. Saragih, "ANALISIS PREFERENSI KONSUMEN BUAH PISANG SEBAGAI KOMODITI UNGGULAN DI KOTA LANGSA," *JURNAL AGRICA*, vol. 14, no. 1, pp. 48–57, Apr. 2021, doi: 10.31289/agrica.v14i1.4259.
- [4] S. Darmawan Putra Bahari and U. Latifa, "KLASIFIKASI BUAH SEGAR MENGGUNAKAN TEKNIK COMPUTER VISION UNTUK PENDETEKSIAN KUALITAS DAN KESEGERAN BUAH," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1567–1573, Oct. 2023, doi: 10.36040/jati.v7i3.6871.
- [5] N. Purnama, "Music Genre Recommendations Based on Spectrogram Analysis Using Convolutional Neural Network Algorithm with RESNET-50 and VGG-16 Architecture," *JISA (Jurnal Informatika dan Sains)*, vol. 5, no. 1, pp. 69–74, Jun. 2022, doi: 10.31326/jisa.v5i1.1270.
- [6] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease," *Agronomy*, vol. 13, no. 6, p. 1633, Jun. 2023, doi: 10.3390/agronomy13061633.
- [7] Indarto and Murinto, "Deteksi Kematangan Buah Pisang Berdasarkan Fitur Warna Citra Kulit Pisang Menggunakan Metode Transformasi Ruang Warna HIS (Banana Fruit Detection Based on Banana Skin Image Features Using HSI Color Space Transformation Method)," 2017.
- [8] B. Yanto *et al.*, "Klasifikasi Tekstur Kematangan Buah Jeruk Manis Berdasarkan Tingkat Kecerahan Warna dengan Metode *Deep Learning Convolutional Neural Network*," *Jurnal Invotek POLBENG*, vol. 6, no. 2, p. 2021, 2021.
- [9] W. Setiawan, "Perbandingan Arsitektur Convolutional Neural Network Untuk Klasifikasi Fundus," *Jurnal Simantec*, vol. 7, no. 2, 2019.
- [10] L. Palupi, E. Ihsanto, and F. Nugroho, "Analisis Validasi dan Evaluasi Model Deteksi Objek Varian Jahe Menggunakan Algoritma YOLOv5," *Journal of Information System Research (JOSH)*, vol. 5, no. 1, pp. 234–241, Oct. 2023, doi: 10.47065/josh.v5i1.4380.
- [11] N. Saranya, K. Srinivasan, and S. K. P. Kumar, "Banana ripeness stage identification: a *deep learning* approach," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 8, pp. 4033–4039, Aug. 2022, doi: 10.1007/s12652-021-03267-w.
- [12] L. Alzubaidi *et al.*, "Review of *deep learning*: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
- [13] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.
- [14] Md. M. Islam, S. M. M. Hasan, Md. R. Hossain, Md. P. Uddin, and Md. Al Mamun, "Towards sustainable solutions: Effective waste classification framework via enhanced deep convolutional neural networks," *PLoS One*, vol. 20, no. 6, p. e0324294, Jun. 2025, doi: 10.1371/journal.pone.0324294.
- [15] P. Baglat, A. Hayat, F. Mendonça, A. Gupta, S. S. Mostafa, and F. Morgado-Dias, "Non-Destructive Banana Ripeness Detection Using Shallow and *Deep Learning*: A Systematic Review," *Sensors*, vol. 23, no. 2, p. 738, Jan. 2023, doi: 10.3390/s23020738.
- [16] M. Khatri, M. Sahoo, S. Sayyad, and J. Sayyad, "Adaptive and User-Friendly Framework for Image Classification with *Transfer Learning* Models," *Future Internet*, vol. 17, no. 8, p. 370, Aug. 2025, doi: 10.3390/fi17080370.
- [17] A. Suny, M. S. U. Islam, and Md. I. Hossain, "Reliable *Deep Learning* for Small-Scale Classifications: Experiments on Real-World Image Datasets from Bangladesh," *arXiv preprint arXiv:2601.11911*, Jan. 2026, [Online]. Available: <http://arxiv.org/abs/2601.11911>
- [18] L. Wang *et al.*, "AK-DL: A Shallow Neural Network Model for Diagnosing Actinic Keratosis with Better Performance than Deep Neural Networks," *Diagnostics*, vol. 10, no. 4, p. 217, Apr. 2020, doi: 10.3390/diagnostics10040217