

# Membangun Sistem Identifikasi Kematangan Buah Alpukat menggunakan teknologi Pengolahan Citra Digital

Yosa Adytia Pratama <sup>1\*</sup> 

<sup>1</sup>Universitas Muhammadiyah Ponorogo, Indonesia

## ARTICLE INFO

### Article history:

Received, 2024-07-07

Accepted, 2024-07-14

Published, 2024-07-14

### Kata Kunci:

Era Digital, Komunikasi Sosial, Peluang, Tantangan

### Keywords:

Avocado, Avocado Fruit Ripeness Detection, CNN

### About Article



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2024 by Author. Published by Aksara Shofa.

## ABSTRAK

Penelitian ini bertujuan untuk membangun sistem identifikasi kematangan buah alpukat menggunakan teknologi pengolahan citra digital, dengan fokus pada metode Convolutional Neural Network (CNN). Alpukat, dengan kandungan lemak tinggi dan kandungan gizi yang kaya, merupakan komoditas bernilai tinggi, namun menentukan tingkat kematangan yang tepat menjadi tantangan utama dalam perdagangan buah ini. Metode identifikasi kematangan saat ini sering kali bergantung pada pengamatan visual, yang sulit dilakukan karena alpukat tidak mengalami perubahan warna saat matang. Oleh karena itu, diperlukan sistem otomatis yang dapat mengidentifikasi tingkat kematangan buah alpukat. Penelitian ini menggunakan dataset dari Mendeley dan menerapkan arsitektur CNN untuk mengklasifikasikan tingkat kematangan alpukat menjadi empat kategori: mentah, setengah matang, matang, dan terlalu matang. Namun, hasil pelatihan model menunjukkan tantangan dalam mengatasi overfitting dan ketidakseimbangan kelas. Rekomendasi untuk penelitian selanjutnya adalah menambahkan teknik regularisasi seperti dropout dan memperluas variasi data dengan augmentasi, serta meningkatkan sampel pada kelas minoritas dengan oversampling.

## ABSTRACT

This research aims to build an avocado fruit ripeness identification system using digital image processing technology, focusing on the Convolutional Neural Network (CNN) method. Avocado, with its high-fat content and rich nutritional value, is a high-value commodity, but determining the correct level of ripeness remains a major challenge in avocado trading. Current ripeness identification methods often rely on visual observation, which is difficult to perform because avocados do not undergo color changes when ripe. Therefore, an automated system that can identify the ripeness level of avocados is needed. This study uses a dataset from Mendeley and applies CNN architecture to classify the ripeness levels of avocados into four categories: raw, semi-ripe, ripe, and overripe. However, the results of model training show challenges in overcoming overfitting and class imbalance. Recommendations for further research include adding regularization techniques such as dropout and expanding data variation with augmentation, as well as increasing samples in minority classes through oversampling.

## 1. PENDAHULUAN

Alpukat merupakan buah tinggi lemak dengan kandungan lemak 20 kali lebih banyak dibandingkan buah lainnya. Tanaman alpukat memiliki nama latin *Persea americana* dan diyakini berasal dari Amerika Tengah. Buah ini dibawa ke Indonesia oleh Hindia Belanda pada tahun 1920-1930. Saat itu, Belanda mendatangkan alpukat ke Indonesia untuk memenuhi kebutuhan lemak masyarakat pegunungan (Finaka, 2019).

\*Penanggung Jawab

E-mail : [yosaae24@gmail.com](mailto:yosaae24@gmail.com) (Yosa adytia pratama)



Gambar 1. Tingkat Kematangan Buah Alpukat

Tantangan utama dalam perdagangan buah alpukat yaitu menentukan tingkat kematangan buah yang tepat. Pemilihan buah dengan kematangan yang keliru dapat berimbas pada penjualan produk olahan alpukat dan tentu pada akhirnya akan mengurangi keuntungan, bahkan sampai merugi. Contohnya penjual minuman alpukat kocok, es oyen atau produk olahan alpukat lain sering menghadapi masalah yaitu terpaksa mengurangi penjualan karena alpukat yang telah dibeli belum matang. Atau kebalikannya, akibat alpukat matang terlalu cepat, sedang penjualan sedang sepi maka mereka terpaksa harus membuang stok buah alpukat yang terlanjur busuk. Fenomena di atas sesuai dengan Laporan Kajian *Food Loss and Waste* tahun 2021 yang merupakan hasil riset kolaborasi Kementerian PPN/Bappenas dengan *Waste4Change* dan *World Resource Institut* mencatat sampah makanan Indonesia pada kategori buah-buahan selama periode 2000-2019 terbuang sebanyak 20% dan pada tahun 2017, 30.000 hingga 40.000 ton terbuang sia-sia karena praktik penanganan sebelum dan sesudah panen - dengan kata lain, setengah dari alpukat yang dihasilkan terbuang sia-sia (Wasilwa et al., 2017).

Menanggapi banyaknya buah-buahan yang terbuang percuma, kami berusaha mengurangi makanan terbuang dengan menjadi konsumen yang bijak dalam membeli buah khususnya alpukat. Membeli alpukat dengan kematangan sesuai kehendak membuat kita bisa memperkirakan masa penyimpanan dan mengurangi limbah rumah tangga. Namun terdapat kendala yaitu penentuan kematangan buah alpukat yang saat ini dilakukan secara pengamatan kasat mata sulit dilakukan orang awam. Hal ini disebabkan karena buah tidak mengalami perubahan warna saat matang sehingga tingkat kematangan buah alpukat sulit ditentukan (Azrita et al., 2020). Oleh karena itu seandainya ada sistem yang dapat mengidentifikasi tingkat kematangan buah alpukat dengan tepat tentu akan membantu memecahkan permasalahan ini. Oleh sebab itu dibutuhkan Metode untuk mengatasi masalah dalam membedakan tampilan luar buah alpukat sesuai tingkat kematangannya. Permasalahan identifikasi tampak luar buah alpukat rasanya bisa dilakukan dengan pendekatan teori *deep learning* dan analisis citra digital, yang merupakan suatu teknik mengolah dan menganalisa citra atau gambar dengan menggunakan komputer (Ifmalinda et al., 2022). Contoh dari teori ini salah satunya yaitu metode *Convolutional Neural Network (CNN)*.

*CNN* Bagian dari *deep learning* yang digunakan untuk melatih komputer dalam mencari representasi terbaik adalah *CNN (Convolutional Neural Network)*. Teknologi pengolahan citra digital dengan metode *CNN* dianggap potensial untuk dikembangkan menjadi sistem yang mampu mengidentifikasi tingkat kematangan buah alpukat secara otomatis dan akurat. *CNN* terdiri dari tiga tahapan: Input data, Pembelajaran Fitur, dan Klasifikasi. Dalam kasus ini, setiap citra buah alpukat diproses melalui ketiga tahapan tersebut menggunakan proses penyaringan. (Santoso & Ariyanto, 2018).

Pada penelitian dengan judul "Sistem Deteksi Kematangan Buah Alpukat Menggunakan Metode Pengolahan Citra" dari Universitas Stikubank, Sistem deteksi kematangan buah alpukat dilakukan dengan membobot nilai warna yang terdiri dari komponen Red, Green, dan Blue (RGB) pada sebuah citra. Sistem ini bertujuan untuk mengklasifikasikan tingkat kematangan alpukat menjadi tiga kategori: mentah, masak, dan matang berdasarkan normalisasi nilai rata-rata RGB pada citra tersebut. Keakuratan sistem ini mencapai 90% untuk alpukat matang, 80% untuk yang masak, dan 50% untuk yang mentah, menurut penelitian oleh (Mukhofifah & Nurraharjo, 2019). Namun, analisis citra menggunakan pembobotan memiliki beberapa kelemahan. Kesalahan dapat terjadi karena warna kulit alpukat yang mirip antara tingkat kematangan yang berbeda. Selain itu, bercak-bercak atau goresan kecokelatan dan kehitaman pada tekstur kulit alpukat juga dapat mengurangi akurasi deteksi. Kendala lainnya adalah latar

belakang citra yang lebih luas dibanding objek alpukat itu sendiri, yang juga dapat mempengaruhi hasil deteksi.

Berdasarkan latar belakang yang telah diuraikan diatas, maka penulis mengusulkan penelitian dengan judul “Membangun Sistem Identifikasi Kematangan Buah Alpukat menggunakan teknologi Pengolahan Citra Digital” yang didasari pada teori *Deep Learning* dan analisis citra digital, dengan fokus pada metode *CNN*.

Sistem ini akan memberikan manfaat besar bagi konsumen pedagang dan pelaku bisnis buah alpukat karena dapat mengetahui lama waktu simpan alpukat hingga kapan waktu yang tepat untuk mengolah atau menjual buah alpukat sesuai dengan tingkat kematangannya. Penelitian ini diharapkan dapat menghasilkan sistem komputer yang dapat secara akurat mendeteksi tingkat kematangan buah alpukat sehingga dalam penggunaannya, sistem ini akan membuat mereka bisa mengurangi pemborosan dengan mengurangi limbah makanan dan meningkatkan efisiensi dalam bisnisnya.

## 2. METODE

### 2.1 Pengumpulan Dataset

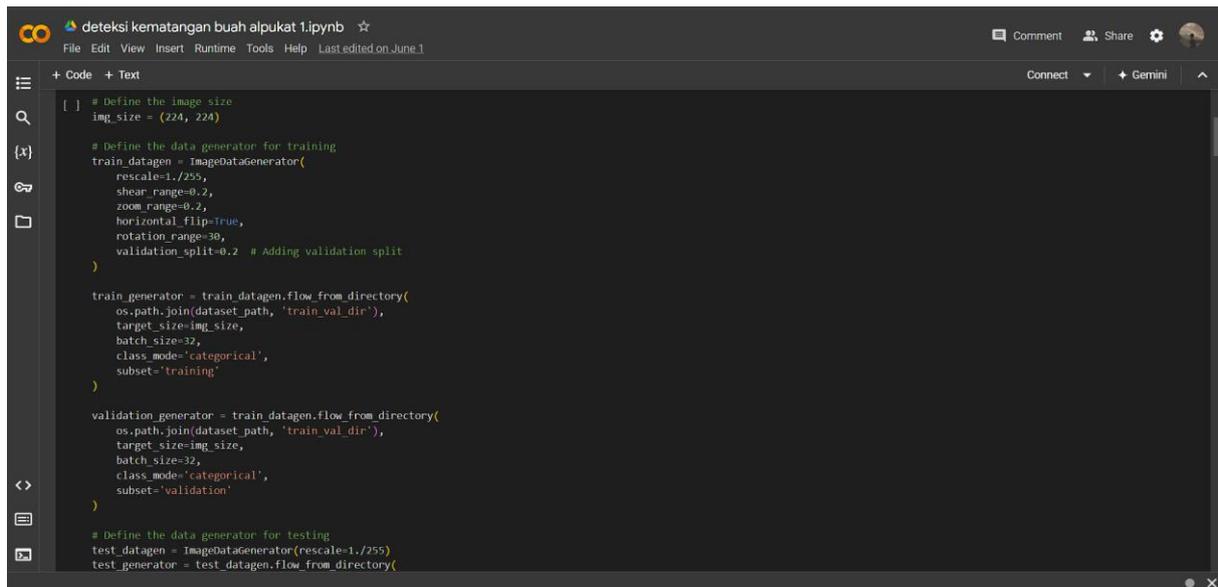
Pada penelitian ini peneliti menggunakan dataset dari dataset Mendeley (Xavier et al., 2024). Di dalam dataset ini terdapat 2, ada folder *train\_val\_dir* yang terdiri 1000 dataset dan terbagi menjadi 4 folder yakni 01 (mentah), 02 (setengah matang), 03 (matang), 04 (terlalu matang) dan folder test yang terdiri 120 dataset dan terbagi menjadi 4 folder yakni 01 (mentah), 02 (setengah matang), 03 (matang), 04 (terlalu matang).



Gambar 2. Citra tingkah kematangan buah alpukat

### 2.2 Preparing Data

Pada tahap awal penelitian ini, persiapan data untuk membandingkan kematangan buah alpukat menggunakan arsitektur pembelajaran transfer sangat penting. Keakuratan dan relevansi hasil penelitian dipengaruhi oleh kualitas dan keragaman data. Foto-foto kematangan buah alpukat yang menunjukkan perubahan warna dan tekstur kulit masih dikumpulkan. Dengan menggunakan sampel yang cukup terdiversifikasi, model yang dihasilkan dapat membedakan kematangan buah alpukat secara efektif (Gede Bintang Arya Budaya & Angga Pradipta, 2023). Pra-pemrosesan data adalah bagian penting dari penelitian. Augmentasi data meningkatkan variasi contoh dalam kumpulan data pelatihan, sementara normalisasi data mengurangi disparitas skala. Membagi kumpulan data menjadi tiga kelompok: pelatihan, validasi, dan pengujian adalah strategi untuk mengurangi overfitting dan memberikan evaluasi kinerja yang tidak memihak. Persiapan data yang cermat sangat penting untuk keandalan dan kredibilitas model CNN dalam mendeteksi kematangan buah alpukat. Ini akan memberikan wawasan tentang efektivitas arsitektur pembelajaran transfer dalam aplikasi ini.



```

# Define the image size
img_size = (224, 224)

# Define the data generator for training
train_datagen = ImageDataGenerator(
    rescale=1./255,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    rotation_range=30,
    validation_split=0.2 # Adding validation split
)

train_generator = train_datagen.flow_from_directory(
    os.path.join(dataset_path, 'train_val_dir'),
    target_size=img_size,
    batch_size=32,
    class_mode='categorical',
    subset='training'
)

validation_generator = train_datagen.flow_from_directory(
    os.path.join(dataset_path, 'train_val_dir'),
    target_size=img_size,
    batch_size=32,
    class_mode='categorical',
    subset='validation'
)

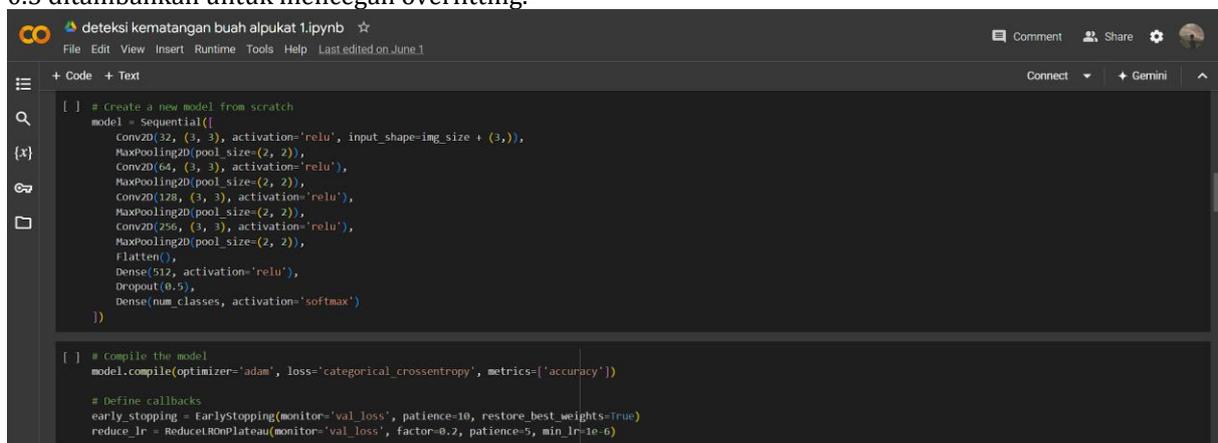
# Define the data generator for testing
test_datagen = ImageDataGenerator(rescale=1./255)
test_generator = test_datagen.flow_from_directory(

```

Gambar 3. Codingan preparing data

### 2.3 Konfigurasi Model CNN

Arsitektur model CNN terdiri dari beberapa lapisan konvolusi dengan ukuran filter yang meningkat secara bertahap (32, 64, 128, dan 256) dan diikuti oleh lapisan pooling. Lapisan konvolusi diaktifkan dengan fungsi aktivasi ReLU. Setelah lapisan konvolusi, data diratakan dan diproses oleh dua lapisan dense, dengan lapisan dense terakhir memiliki unit sebanyak jumlah kelas (kategori kematangan alpukat), diaktifkan dengan fungsi softmax untuk menghasilkan probabilitas kelas. Sebuah lapisan dropout dengan probabilitas 0.5 ditambahkan untuk mencegah overfitting.



```

# Create a new model from scratch
model = Sequential(
    Conv2D(32, (3, 3), activation='relu', input_shape=img_size + (3,)),
    MaxPooling2D(pool_size=(2, 2)),
    Conv2D(64, (3, 3), activation='relu'),
    MaxPooling2D(pool_size=(2, 2)),
    Conv2D(128, (3, 3), activation='relu'),
    MaxPooling2D(pool_size=(2, 2)),
    Conv2D(256, (3, 3), activation='relu'),
    MaxPooling2D(pool_size=(2, 2)),
    Flatten(),
    Dense(512, activation='relu'),
    Dropout(0.5),
    Dense(num_classes, activation='softmax')
)

# Compile the model
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

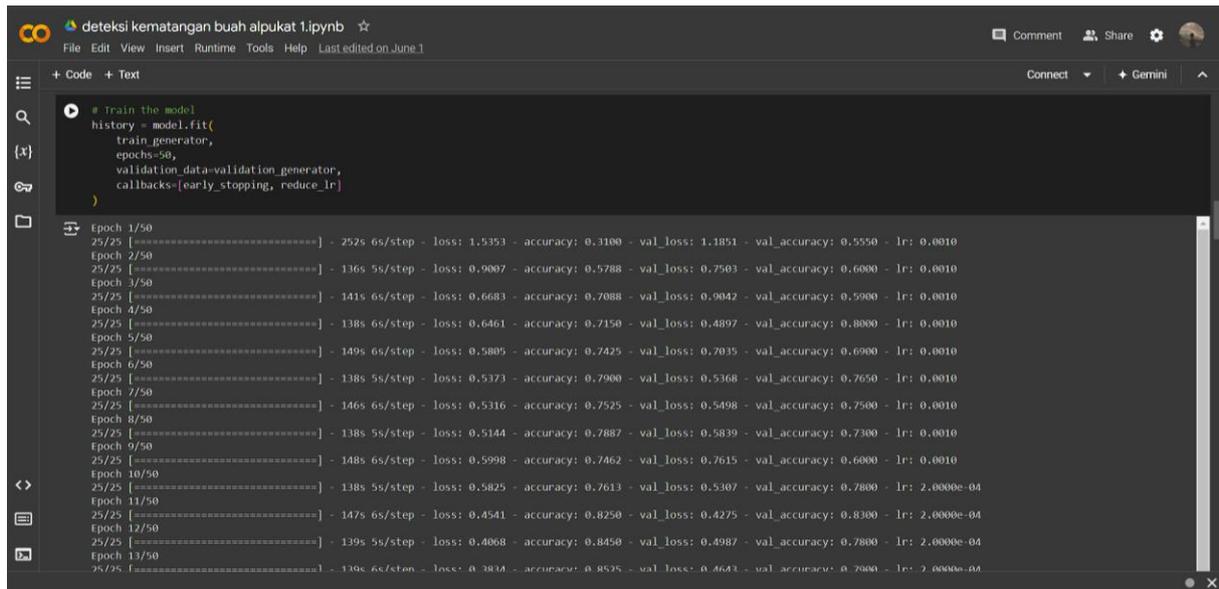
# Define callbacks
early_stopping = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)
reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=5, min_lr=1e-6)

```

Gambar 4. Codingan konfigurasi model

### 2.4 Pelatihan Model

Model ini dikompilasi menggunakan optimizer Adam dan fungsi loss `categorical_crossentropy`, yang cocok untuk masalah klasifikasi multi-kelas. Dua callback digunakan untuk memantau dan mengatur proses pelatihan: `EarlyStopping` untuk menghentikan pelatihan jika tidak ada peningkatan pada loss validasi setelah beberapa epoch, dan `ReduceLROnPlateau` untuk mengurangi laju pembelajaran jika metrik validasi tidak menunjukkan perbaikan. Model dilatih selama maksimal 50 epoch dengan menggunakan generator data untuk pelatihan dan validasi.



Gambar 5. Codingan pelatihan model

### 2.5 Evaluasi dan Analisis

Untuk menganalisis performa model, *Confusion matrix* dan laporan klasifikasi dihasilkan. *Confusion matrix* adalah metode yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menghitung akurasi, loss, dan presisi. Evaluasi ini terdiri dari empat kategori utama seperti yang ditunjukkan berikut (Nabila Asryani Sundari et al., 2022) Sedangkan laporan klasifikasi memberikan metrik seperti precision, recall, dan f1-score untuk setiap kelas.

#### 1. Tabel *Confusion matrix*

		Prediksi	
		TRUE	FALSE
Aktual	TRUE	Tp	Fn
	FALSE	Fp	Tn

Keterangan :

TP: *True Positif* didefinisikan sebagai positif yang diprediksi dengan benar.

TN: *True Negatif* didefinisikan sebagai negatif yang diprediksi dengan benar.

FP: *False Positif* didefinisikan sebagai negatif yang diprediksi dengan data positif.

FN: *False Negatif* didefinisikan sebagai positif yang diprediksi dengan data negatif

#### A. Akurasi

Akurasi sistem adalah ukuran ketepatan suatu sistem dalam mengenali data input sehingga output yang dihasilkan sesuai dengan yang diharapkan. Persamaan untuk menghitung akurasi dapat dilihat pada Persamaan berikut.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

#### B. Loss

Loss adalah ukuran ketidaktepatan suatu sistem dalam mengenali data input, yang menyebabkan output yang dihasilkan tidak sesuai dengan yang diharapkan. Persamaan untuk menghitung loss dapat dilihat pada persamaan berikut.

$$Akurasi = \frac{FP + FN}{TP + TN + FP + FN}$$

#### C. Precision

Precision adalah ukuran seberapa tepat sistem dalam memberikan informasi yang sesuai dengan yang diminta. Persamaan untuk menghitung precision ditunjukkan pada persamaan berikut.

$$Presisi = \frac{TP}{TP + FP}$$

#### D. Recall

Recall adalah ukuran seberapa baik sistem dalam menemukan informasi yang relevan. Persamaan untuk menghitung recall dapat dilihat pada persamaan berikut.

$$Recall = \frac{TP}{TP + Fn}$$

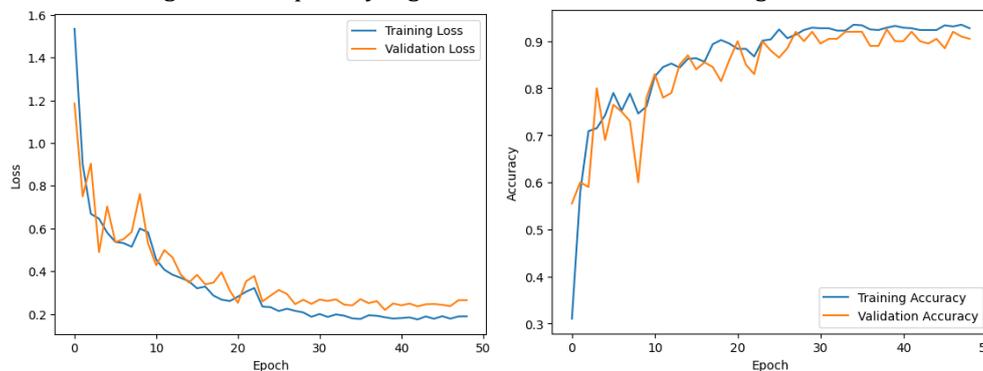
#### E. F1 Score

Recall adalah ukuran seberapa baik sistem dalam menemukan informasi yang relevan. Persamaan untuk menghitung recall dapat dilihat pada persamaan berikut.

$$F1\ Score = \frac{2 \times P \times R}{P + R}$$

### 3. HASIL DAN PEMBAHASAN

Berdasarkan hasil pelatihan yang telah dilakukan dari metode penelitian diatas menghasilkan sistem deteksi kematangan buah alpukat yang memiliki akurasi model sebagai berikut:

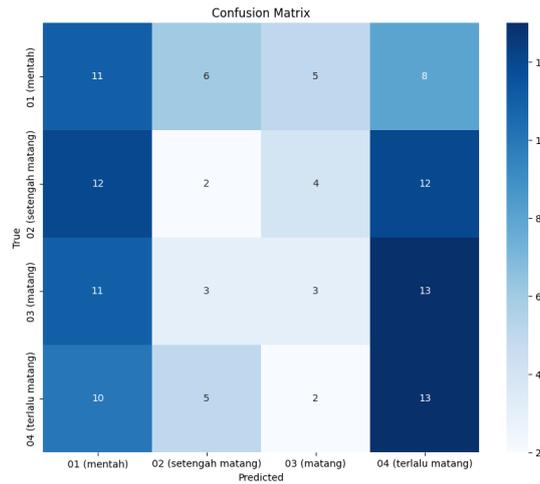


Gambar 6. Grafik perbandingan hasil pelatihan model

Bisa dilihat pada gambar 3 epoch pertama, model menunjukkan nilai loss sebesar 1.5353 dan akurasi sebesar 0.3100, baik pada data pelatihan maupun validasi, yang mencerminkan bahwa model belum belajar banyak dari data tersebut. Namun, seiring bertambahnya epoch, terjadi peningkatan yang signifikan dalam performa pelatihan. Pada epoch ke-50, loss pada data pelatihan menurun hingga 0.1886, sementara akurasi pelatihan meningkat secara bertahap mencapai 0.9275. Ini menunjukkan bahwa model mampu belajar dan menyesuaikan parameter-parameternya dengan data pelatihan.

Pada data validasi, meskipun mengalami beberapa fluktuasi, loss secara umum menunjukkan tren penurunan. Akurasi validasi mencapai nilai tertinggi pada epoch ke-39 dengan angka 0.9250. Namun, fluktuasi yang terjadi setelah epoch tersebut mengindikasikan bahwa model mungkin mulai mengalami overfitting, yaitu situasi di mana model terlalu baik dalam memprediksi data pelatihan namun tidak mampu melakukan generalisasi dengan baik pada data baru.

Learning rate memainkan peran penting dalam proses pelatihan model. Pada epoch ke-10, learning rate awal sebesar 0.0010 diturunkan menjadi 0.0002. Penyesuaian ini membantu mengurangi fluktuasi pada loss validasi dan memberikan kestabilan lebih pada model. Selanjutnya, pada epoch ke-27 dan ke-34, learning rate kembali diturunkan menjadi 0.00008 dan 0.000016. Penurunan bertahap ini membantu mengurangi loss, tetapi tidak memberikan dampak signifikan pada akurasi validasi, yang tetap berfluktuasi.



Gambar 7. Confusion matrix

Evaluasi performa model dengan menggunakan confusion matrix dan classification report pada gambar 4 memberikan gambaran lebih jelas tentang kemampuan model dalam mengklasifikasikan data ke dalam kelas-kelas yang berbeda. Confusion matrix menunjukkan bahwa model sering kali bingung antara kelas-kelas yang ada, dengan tingkat kesalahan yang cukup tinggi di berbagai kelas. Misalnya, pada kelas 0, 3, dan 1, model sering salah mengklasifikasikan sampel ke kelas yang lain.

Tabel 2. Classification report

Classification Report:					
		precision	recall	f1-score	support
	0	0.25	0.37	0.30	30
	1	0.12	0.07	0.09	30
	2	0.21	0.10	0.14	30
	3	0.28	0.43	0.34	30
accu	racy			0.24	120
macro	avg	0.22	0.24	0.22	120
weighted	avg	0.22	0.24	0.22	120

Lebih lanjut, classification report pada tabel 2 mengungkapkan bahwa precision, recall, dan f1-score untuk beberapa kelas sangat rendah. Misalnya, precision dan recall terbaik masing-masing hanya mencapai 0.28 dan 0.43 pada kelas 3, yang menunjukkan bahwa meskipun model lebih baik dalam mengklasifikasikan kelas ini, performanya masih jauh dari memuaskan pada kelas lainnya.

**4. KESIMPULAN**

Berdasarkan hasil dari pembahasan dapat disimpulkan bahwa model CNN yang saya buat kurang berhasil dalam mendeteksi antara kelas-kelas yang ada. Karena *overfitting* menjadi tantangan utama yang terlihat dari perbedaan akurasi antara pelatihan dapat dilihat dari loss validasi yang cenderung berfluktuasi yang mengindikasikan model *overfitting*, Tidak seimbang dalam performa klasifikasi antar kelas menunjukkan adanya class imbalance.

**5. UCAPAN TERIMAKASIH**

Ucapan terima kasih untuk kedua orang tua, dosen pengampu mata kuliah kapita selekta dan yang tidak dapat saya sebutkan satu persatu berkat support kalian semua penelitian ini dapat terselesaikan.

**6. REFERENSI**

- Azrita, M. W., Ahmad, U., & Darmawati, E. (2020). Rancangan Kemasan dengan Indikator Warna untuk Deteksi Tingkat Kematangan Buah Alpukat. *Jurnal Keteknik Pertanian*, 7(2), 155–162. <https://doi.org/10.19028/jtep.07.2.155-162>
- Finaka, A. W. (2019). *Alpukat, Buah Nikmat Kaya Manfaat*. Indonesiabaik. <https://indonesiabaik.id/infografis/alpukat-buah-nikmat-kaya-manfaat>
- Gede Bintang Arya Budaya, I., & Angga Pradipta, G. (2023). Performa Backbone RestNet50 dan MobileNetV2 pada DeeplabV3+ untuk Segmentasi Karakter Komik Lokal. *Seminar Nasional Penelitian Dan Pengabdian Kepada Masyarakat CORISINDO*, 122–127.
- Ifmalinda, I., Andasuryani, A., & Shaufana, L. (2022). IDENTIFIKASI BENTUK BUAH ALPUKAT (Persea americana Mill.) DENGAN ANALISIS CITRA DIGITAL. *Jurnal Teknologi Pertanian*, 23(3), 215–226. <https://doi.org/10.21776/ub.jtp.2022.023.03.5>
- Mukhofifah, M., & Nurraharjo, E. (2019). Sistem Deteksi Kematangan Buah Alpukat Menggunakan Metode Pengolahan Citra. *Jurnal Dinamika Informatika*, 11(1), 12–23. <https://doi.org/10.35315/informatika.v11i1.8144>
- Nabila Asryani Sundari, Rita Magladena, & Sofia Saidah. (2022). Klasifikasi Jenis Kulit Wajah Menggunakan Metode Covolutional Neural Network (CNN) Efficientnet-B0. *E-Proceeding of Engineering*, 8(6), 3180–3187.
- Santoso, A., & Ariyanto, G. (2018). Implementasi Deep Learning Berbasis Keras Untuk. *Jurnal Emitor*, 18(01), 15–21. <http://eprints.ums.ac.id/62956/>
- Wasilwa, L. A., Njuguna, J. K., Okoko, E. N., & Watani, G. W. (2017). Status of Avocado Production in Kenya. *Kenya Agricultural Research Institute (KARI), April 2017*, 1–6.
- Xavier, P., Rodrigues, P., & Silva, C. L. M. (2024). "Hass" Avocado Ripening Photographic Dataset. Mendeley Data, V1. <https://doi.org/10.17632/3xd9n945v8.1>