

**KLASIFIKASI TINGKAT KEMATANGAN BUAH KOPI
ARABICA GAYO MENGGUNAKAN *VISION TRANSFORMER***

TUGAS AKHIR

Diajukan oleh:

RISKI WULANDARI

220705004

Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI AR-RANIRY
BANDA ACEH
2026 M/1447 H**

LEMBAR PERSETUJUAN

KLASIFIKASI TINGKAT KEMATANGAN BUAH KOPI ARABICA GAYO MENGGUNAKAN *VISION TRANSFORMER*

TUGAS AKHIR

Diajukan Kepada Fakultas Sains Dan Teknologi UIN Ar-Raniry Banda Aceh Sebagai
Salah Satu Beban Studi Memperoleh Gelar Sarjana (S1) Dalam Ilmu/Program Studi
Teknologi Informasi

Oleh:


Riski Wulandari
220705004


Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi

Disetujui Untuk Dimunaqasyahkan Oleh:


Pembimbing I,

Pembimbing II,


Malahayati, M.T.
NIP.198301272015032003


Dr. Hendri Ahmadian, M.I.M.
NIP.198301042014031002

Mengetahui,
Ketua Program Studi Teknologi Informasi


Malahayati, M.T.
NIP.198301272015032003

LEMBAR PENGESAHAN

KLASIFIKASI TINGKAT KEMATANGAN BUAH KOPI ARABICA GAYO MENGGUNAKAN *VISION TRANSFORMER*

TUGAS AKHIR

Telah Diuji Oleh Panitia Ujian Munaqasyah Tugas Akhir
Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh Dan Dinyatakan Lulus
Serta Diterima Sebagai Salah Satu Beban Studi Program Sarjana (S1)
Dalam Program Studi Teknologi Informasi

Pada Hari/Tanggal: Rabu, 06 Mei 2026 M
18 Dzulqaidah 1447 H
Di Darussalam Banda Aceh

Panitia Ujian Munaqasyah Tugas Akhir:

Ketua,



Malahayati, M.T.
NIP.198301272015032003

Sekretaris,



Dr. Hendri Ahmadian, M.I.M.
NIP.198301042014031002

Penguji I,



Baihaqi, M.T.
NIP.198802212022031001

Penguji II,



Sarinf Vita Dewi S.T., M.Eng.
NIP.198712222022032001

Mengetahui,
Dekan Fakultas Sains dan Teknologi
UIN Ar-Raniry Banda Aceh



Prof. Dr. Ir. Muhammad Dirhamsyah, M.T., IPU.
NIP.196210021988111001

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini:

Nama : Riski Wulandari
NIM : 220705004
Program studi : Teknologi Informasi
Fakultas : Sains Dan Teknologi
Judul : Klasifikasi Tingkat Kematangan Buah Kopi Arabica Gayo
Menggunkan *Vision Transformer*

Dengan ini menyatakan bahwa dalam penulisan tugas akhir ini, saya:

1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan;
2. Tidak melakukan plagiasi terhadap naskah karya orang lain;
3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa izin pemilik karya
4. Tidak memanipulasi dan memalsukan data
5. Mengerjakan sendiri karya ini dan mampu bertanggung jawab atas karya ini.

Bila dikemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggungjawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenai sanksi berdasarkan aturan yang berlaku di Fakultas Sains Dan Teknologi UIN Ar-raniry Banda Aceh. Demikian pernyataan ini saya buat dengan sesungguhnya dan tanpa paksaan dari pihak manapun

Banda Aceh, 06 Mei 2026
Yang menyatakan




Riski Wulandari

ABSTRAK

Nama : Riski Wulandari
NIM : 220705004
Program studi : Teknologi Informasi
Judul : Klasifikasi Tingkat Kematangan Buah Kopi Arabica Gayo
Menggunakan *Vision Transformer*
Tanggal : 20 April 2026
Jumlah halaman : 67
Pembimbing 1 : Malahayati, M.T.
Pembimbing 2 : Dr. Hendri Ahmadian, M.I.M.

Kualitas kopi arabika Gayo sangat dipengaruhi oleh ketepatan dalam menentukan tingkat kematangan buah saat proses sortasi. Proses manual yang masih digunakan cenderung bersifat subjektif dan kurang konsisten, terutama dalam membedakan buah setengah matang yang memiliki karakteristik visual transisi. Oleh karena itu, diperlukan sistem otomatis berbasis citra untuk meningkatkan akurasi klasifikasi. Penelitian ini bertujuan mengklasifikasikan tingkat kematangan buah kopi menggunakan *Vision Transformer* dengan pendekatan self-supervised learning. Dataset terdiri dari 2.400 citra yang diproses melalui tahap preprocessing, ekstraksi fitur menggunakan DINOv3, dan clustering dengan Fuzzy C-Means menjadi tiga kelas: mentah, setengah matang, dan matang. Hasil evaluasi menunjukkan kualitas cluster yang cukup dengan Silhouette Score 0,2305, Davies-Bouldin Index 1,7560, dan FPC 0,3996. Pada tahap klasifikasi, model mencapai akurasi 92% dengan performa terbaik pada kelas mentah dan matang, sementara kelas setengah matang memiliki recall lebih rendah (0,81) akibat karakteristik visual yang ambigu.

Kata kunci: *Vision Transformer*, self-supervised learning, kopi arabika Gayo, klasifikasi kematangan, deep learning.

ABSTACT

Name : Riski Wulandari
Student ID : 220705004
Study Program : Information Technology
Title : Classification Of Ripeness Levels Of Arabica Gayo Coffee
Cherries Using Vision Transformer
Date : April 20, 2026
Number Of Page : 67
Supervisor 1 : Malahayati, M.T.
Supervisor 2 : Dr. Hendri Ahmadian, M.I.M.

The quality of Arabica Gayo coffee is highly influenced by the accuracy in determining the ripeness level of coffee cherries during the sorting process. Manual sorting methods are still commonly used and tend to be subjective and inconsistent, especially in distinguishing semi-ripe cherries that have transitional visual characteristics. Therefore, an automated image-based system is needed to improve classification accuracy. This study aims to classify the ripeness level of coffee cherries using a Vision Transformer with a self-supervised learning approach. The dataset consists of 2,400 coffee cherry images processed through preprocessing, feature extraction using DINOv3, and clustering using Fuzzy C-Means into three classes: unripe, semi-ripe, and ripe. The evaluation results show moderate clustering quality with a Silhouette Score of 0.2305, Davies-Bouldin Index of 1.7560, and Fuzzy Partition Coefficient (FPC) of 0.3996. In the classification stage, the model achieved an accuracy of 92%, with the best performance in the unripe and ripe classes, while the semi-ripe class showed lower recall (0.81) due to its ambiguous visual characteristics.

Keywords: Vision Transformer, self-supervised learning, Arabica Gayo coffee, ripeness classification, deep learning.

KATA PENGANTAR

Bismillahirrahmanirrahim,

segala puji dan syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan berkah-Nya, sholawat beserta salam tidak lupa kita sanjungkan kepada baginda nabi kita Muhammad SAW. Yang telah membawa kita dari alam jahiliah ke alam yang penuh ilmu pengetahuan. Adapun Tugas Akhir ini berjudul **“Klasifikasi tingkat kematangan buah kopi Arabica Gayo menggunakan *Vision Transformer*”** Tugas Akhir ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S-1) pada Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry Banda Aceh. Pada kesempatan kali ini penulis mengucapkan terimakasih atas bantuan yang telah diberikan kepada penulis baik secara langsung maupun secara tidak langsung dalam rangka penyelesaian Tugas Akhir ini, terutama yang terhormat:

1. Kedua orang tua penulis, bapak Agus Triono dan ibu Rasilawati yang rela memberikan dunianya untuk penulis, segalanya dilakukan demi penulis berjuta-juta terimakasih atas doa dan semangat yang tiada henti atas segala dukungan, kasih sayang, serta perjuangan tanpa kenal lelah dalam memenuhi kebutuhan penulis selama menempuh pendidikan. Semoga segala kebaikan dan pengorbanan yang telah diberikan mendapat balasan yang berlipat ganda dari Allah SWT.
2. Abang penulis Fawira Hidayat, satu-satunya saudara kandung penulis yang selalu menyayangi dan mengasihi penulis dengan sepenuh hati dan senantiasa memberikan semangat kepada penulis.
3. Dekan Fakultas Sains dan Teknologi UIN Ar-raniry Banda Aceh, Prof Dr.Ir.M. Dirhamsyah, M.T., IPU.

4. Ketua dan sekretaris Program Studi Teknologi Informasi, Ibu Malahayati, M.T., dan Bapak Khairan AR, M.Kom., yang memberikan arahan dalam menyelesaikan tugas akhir ini.
5. Ibu Malahayati, M.T., yang telah membimbing dan selalu memberi arahan dalam menyelesaikan tugas akhir ini.
6. Bapak/ibu dosen program Studi Teknologi Informasi yang telah memberikan ilmu dan wawasan selama masa kuliah.
7. Ibu Cut Ida Rahmadiana, S.Si., seorang staf Program Studi Teknologi Informasi, yang telah memberikan bantuan dan dukungan dalam berbagai mata kuliah akademik.
8. Saudara tak sedarahku Ulfa Sasikarani yang setia menemani dari awal memijakkan kaki ke Banda Aceh, serta menjadi saksi suka dan duka selama menjalani kehidupan di perantauan.
9. Sahabatku Riski Mah bengi yang selalu memberi semangat serta dukungan selama penulis melakukan penelitian ini serta selalu ada disaat susah maupun senang.
10. Teman-teman seperjuangan yang selalu ada apabila penulis membutuhkan bantuan selama proses pembelajaran dari semester satu hingga saat ini.

Penulis menyadari masih banyak kekurangan dalam Tugas Akhir ini. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk mencapai kesempurnaan dalam penulisan Tugas Akhir ini.

Banda Aceh, 30 April 2026
penulis,

Riski Wulandari

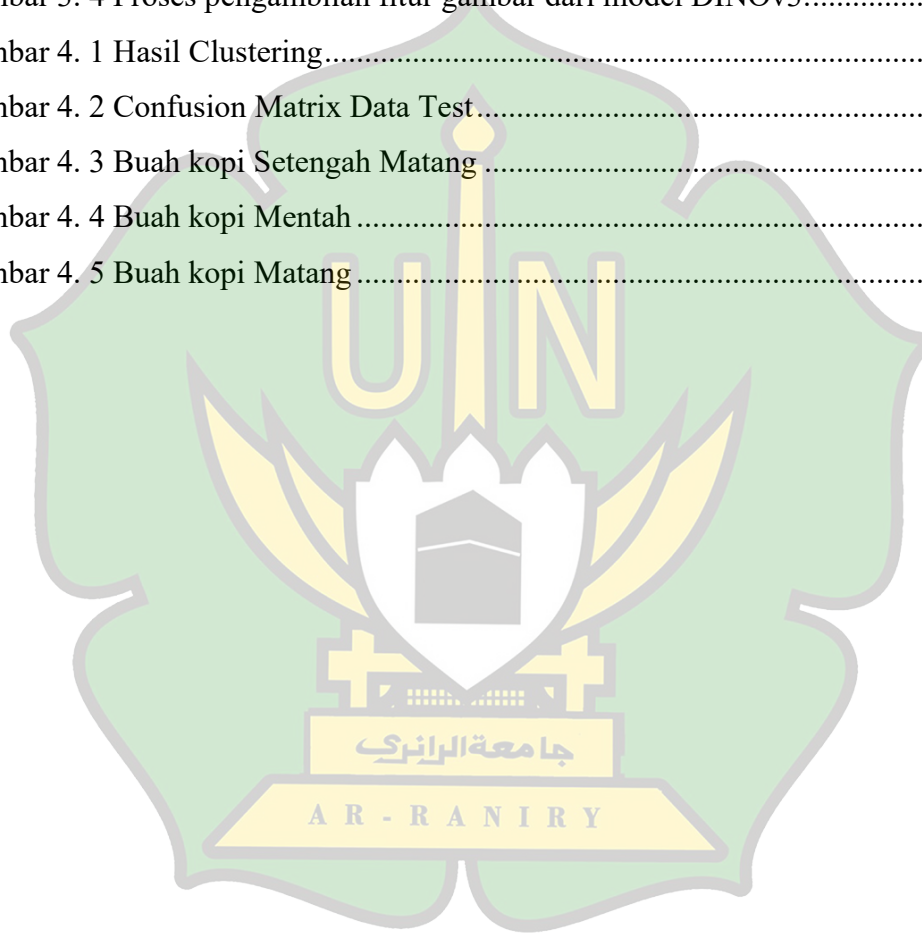
DAFTAR ISI

ABSTRAK	v
ABSTACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB 1 PENDAHULUAN	1
1.1 Latar belakang	1
1.2 Rumusan masalah	3
1.3 Tujuan penelitian	3
1.4 Batasan masalah	3
1.5 Manfaat penelitian	4
BAB II DASAR TEORI	5
2.1 Penelitian terkait	5
2.2 Klasifikasi	13
2.3 Buah kopi Arabica Gayo	14
2.4 Vision Transformer	15
2.5 DINOv3 (Distillation with NO labels)	16
2.6 Self-supervised Learning	16
2.7 Principal Component Analysis (PCA)	18
2.8 Fuzzy C-Means	18
2.7 Deep Learning	19
2.8 Metode Evaluasi	20
BAB III METODE PENELITIAN	22
3.1 Tahapan penelitian	22
3.2 Pengumpulan Data	22

3.3	Preprocessing Data.....	23
3.4	Pretrained Model (DINOv3).....	24
3.1.1.	Input image.....	25
3.1.2.	Normalization.....	26
3.1.3.	Self-Distillation.....	27
3.1.4.	Representation Learning.....	29
3.1.5.	Pretext Task.....	30
3.1.6.	Feature projection.....	31
2.4.1	Fuzzy C-Means (FCM).....	32
3.1.7.	Class Assigment.....	33
3.5	Evaluasi Model.....	34
3.6	Visualisasi hasil.....	36
3.5	Skenario pengujian dan evaluasi kinerja.....	37
3.6	Alat dan bahan.....	39
BAB IV HASIL DAN PEMBAHASAN.....		40
4.1	Hasil Evaluasi Dataset Training dan Validation.....	40
4.1.1.	Hasil Clustering.....	40
4.1.2.	Hasil Evaluasi Pretrained.....	41
4.2	Evaluasi Peforma Model.....	43
4.2.1.	Evaluasi Peforma Model.....	43
4.2.2.	Analisa Confusion Matrix.....	44
4.2.3.	Analisa Visualisasi kopi Setengah Matang.....	46
4.2.4.	Analisa Visualisasi Kopi Mentah.....	47
4.2.5.	Analisa Visualisasi Kopi Matang.....	49
BAB V KESIMPULAN DAN SARAN.....		51
5.1.	Kesimpulan.....	51
5.2.	Saran.....	52
DAFTAR PUSTAKA.....		54

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur Vision Transformer	15
Gambar 3. 1 Tahapan penelitian	22
Gambar 3. 2 Alur sistem ekstraksi fitur DINOv3	25
Gambar 3. 3 Image Resizing.....	26
Gambar 3. 4 Proses pengambilan fitur gambar dari model DINOv3.....	29
Gambar 4. 1 Hasil Clustering.....	40
Gambar 4. 2 Confusion Matrix Data Test.....	45
Gambar 4. 3 Buah kopi Setengah Matang	46
Gambar 4. 4 Buah kopi Mentah	48
Gambar 4. 5 Buah kopi Matang.....	49



DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait.....	5
Tabel 3. 1 Dataset Penelitian.....	23
Tabel 3. 2 Alat dan Bahan.....	39
Tabel 4. 1 Hasil Evaluasi	41
Tabel 4. 2 Hasil Evaluasi Peforma Model.....	43



BAB 1

PENDAHULUAN

1.1 Latar belakang

Mendengar kata kopi mesti sudah tidak asing lagi dikalangan anak muda maupun orang tua, dengan cita rasanya yang khas serta penghambat rasa ngantuk sehingga menjadi minuman yang sangat diminati. Aceh merupakan provinsi yang memiliki luas dan jumlah produksi tertinggi di Indonesia, yang mana jumlah produksinya mencapai 71.084 ton dan luas area 113.968 hektare di tahun 2023. Produksi kopi Indonesia sebagian besar diekspor ke mancanegara dan sisanya dipasarkan di dalam negeri, ekspor dalam negeri Indonesia menjangkau lima benua yaitu Asia, Afrika, Australia, Amerika, dan Eropa dengan pangsa utama di Eropa. Pada tahun 2023, lima besar negara pengimpor Kopi alam Indonesia adalah Amerika Serikat, Egypt, Malaysia, India dan Italy. Arabica, *not roasted, not decaffeinated* dan Robusta, *not roasted, not decaffeinated* merupakan jenis kopi dengan volume dan nilai tertinggi yang di ekspor dari Indonesia (statistik kopi Indonesia, 2023).

Kematangan buah kopi dapat dilihat dari tiga tingkatan yaitu matang, setengah matang dan mentah sehingga warna buah kopi dapat menjadi tolak ukur untuk mengenali tingkat kematangan serta kualitas buah kopi (Nerlys Amelia dkk., 2023). Seiring berkembangnya teknologi kecerdasan buatan, analisis citra digital dapat dimanfaatkan untuk melakukan klasifikasi tingkat kematangan buah kopi secara otomatis. Melalui pendekatan *computer vision*, sistem mampu mempelajari pola visual seperti warna, tekstur, dan bentuk objek untuk membedakan tingkat kematangannya berdasarkan citra digital. Pendekatan *deep learning*, khususnya arsitektur *Vision Transformer*, telah banyak digunakan dalam berbagai tugas klasifikasi citra karena mampu mengekstraksi fitur visual secara lebih efektif melalui mekanisme *self-attention*

(Le dkk., 2024). Dalam proses klasifikasi, analisis performa model menjadi aspek penting untuk menilai kemampuan model dalam mengenali pola pada data citra. Evaluasi performa biasanya dilakukan menggunakan beberapa metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk memberikan gambaran yang lebih komprehensif mengenai kinerja model dalam melakukan prediksi (Rashid et al., 2024). Melalui analisis metrik evaluasi tersebut, kemampuan model dalam memahami karakteristik visual objek dapat diukur secara lebih objektif dan menjadi dasar dalam pengembangan sistem klasifikasi citra yang lebih optimal.

Vision transformer merupakan salah satu algoritma deep learning yang akan penulis gunakan untuk klasifikasi serta indentifikasi objek melalui citra digital, pada penelitian Arya Pangestu dkk (2023) yang berjudul “*Vision Transformer* untuk klasifikasi buah pisang” melakukan penelitian terhadap buah pisang, yang mana buah pisang diberi label mentah, setengah matang, matang, dan terlalu matang. Pada penelitian tersebut, digunakan 5.068 citra pisang yang berbeda dari dataset latih Hasil evaluasi menunjukkan model ViTL/16-in21k memiliki akurasi tertinggi sebesar 91,61%. Berbeda dengan penelitian tersebut, penulis melakukan penelitian menggunakan dataset citra buah kopi yang diambil langsung dari kebun milik Bapak Agus Triono di Kampung Atu Lintang, Kabupaten Aceh Tengah. Dalam penelitian ini penulis menggunakan metode self-supervised learning, yaitu metode pembelajaran yang memungkinkan model belajar langsung dari data gambar tanpa memerlukan label secara eksplisit. Dengan demikian, penelitian ini berbeda dari penelitian sebelumnya karena tidak memerlukan proses pelabelan dataset secara manual, melainkan model akan belajar langsung dari pola dan karakteristik gambar itu sendiri.

Selain itu, penelitian ini juga menganalisis performa model *Vision Transformer* dengan pendekatan Self-Supervised Learning untuk mengetahui efektivitas metode tersebut dalam klasifikasi citra buah kopi dibandingkan

pendekatan supervised learning pada penelitian sebelumnya. Berdasarkan uraian di atas penulis akan melakukan penelitian mengenai “*klasifikasi tingkat kematangan buah kopi Arabica Gayo menggunakan Vision Transformer*”. Penelitian ini bertujuan untuk menerapkan model *Vision Transformer* dalam klasifikasi buah kopi berdasarkan tingkat kematangannya.

1.2 Rumusan masalah

Adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana menerapkan *Vision Transformer* dalam melakukan klasifikasi tingkat kematangan buah kopi Arabica Gayo berbasis citra visual?
2. Bagaimana performa *vision Transformer* dalam mengklasifikasikan tingkat kematangan buah kopi arabica gayo?

1.3 Tujuan penelitian

Adapun tujuan penelitian ini adalah:

1. Menerapkan model *Vision Transformer* untuk melakukan klasifikasi tingkat kematangan buah kopi Arabica Gayo berbasis citra visual.
2. Mengkaji sejauh mana model *Vision Transformer* mampu mengidentifikasi dan membedakan tingkat kematangan buah kopi Arabica Gayo melalui analisis citra digital.

1.4 Batasan masalah

Adapun batasan masalah dari penelitian ini adalah:

1. Sistem akan mendeteksi dan mengidentifikasi buah kopi dengan tiga macam tingkatan yaitu mentah, setengah matang dan matang.
2. Setiap macam buah memiliki 1000 gambar yang akan di klasifikasi menggunakan model *Vision Transformer*.
3. Evaluasi kinerja model dilakukan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.

1.5 Manfaat penelitian

Adapun manfaat dari penelitian ini adalah:

1. Memberikan kontribusi dalam penerapan model *Vision Transformer* untuk klasifikasi tingkat kematangan buah kopi berbasis citra visual, sehingga dapat menunjukkan kemampuan model dalam mengenali pola visual pada objek buah kopi.
2. Menyediakan analisis performa model klasifikasi melalui evaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score untuk mengetahui tingkat efektivitas model dalam mengidentifikasi tingkat kematangan buah kopi secara otomatis.



BAB II DASAR TEORI

2.1 Penelitian terkait

Sebelum melakukan penelitian ini, penulis meninjau beberapa penelitian terdahulu yang berkaitan dengan topik yang diangkat. Tujuan dari pembahasan ini adalah untuk mengetahui metode atau pendekatan yang telah digunakan oleh peneliti sebelumnya, serta menemukan celah penelitian (research gap) yang dapat dijadikan dasar dalam penelitian ini agar hasilnya lebih optimal. Beberapa penelitian terkait dapat dilihat pada Tabel 2.1 berikut.

Tabel 2. 1 Penelitian Terkait

No	Referensi	Metode penelitian	Hasil penelitian
1.	(Rafi Haqul Baasith dkk., 2026)	Vision Transformer (self-supervised)	Penelitian ini menggunakan dataset CheXpert-v1.0-small yang berisi 223.414 citra X-ray dada dengan 14 kelas penyakit paru, dimana label disimpan dalam file CSV dan distribusi data tiap kelas tidak seimbang. Metode yang digunakan adalah self-supervised learning (SSL) sebagai tahap pretraining pada data tanpa label, kemudian dilakukan fine-tuning supervised learning menggunakan model Vision Transformer (ViT) sehingga menghasilkan model SSL-ViT. Hasil eksperimen

No	Referensi	Metode penelitian	Hasil penelitian
			menunjukkan bahwa ViT baseline memperoleh precision 0,66, recall 0,49, F1-score 0,56, dan AUC 0,77, sedangkan SSL-ViT menghasilkan precision 0,28, recall 0,73, F1-score 0,40, dan AUC 0,75, yang menunjukkan peningkatan sensitivitas deteksi. Visualisasi menggunakan Grad-CAM juga menunjukkan bahwa model fokus pada area paru yang relevan, sehingga pendekatan SSL-ViT lebih cocok digunakan untuk screening medis karena mampu mengurangi risiko kesalahan tidak terdeteksinya penyakit (false negative).
2.	(Gregorius Natanael Elwirehardja dkk., 2025)	Vision Transformer (self-supervised)	Penelitian ini menggunakan dataset NIH ChestX-ray14 yang berisi 91.324 citra X-ray dada dengan 15 kelas untuk melakukan klasifikasi penyakit paru menggunakan metode self-supervised learning. Model yang dibandingkan meliputi Vision Transformer (ViT) dengan

No	Referensi	Metode penelitian	Hasil penelitian
			<p>DINOv2, ConvNeXt-V2, serta model baseline seperti ResNet-50, ViT dengan DINO, dan CheXNet. Hasil penelitian menunjukkan bahwa model DINOv2 ViT-S/14 menjadi model terbaik di antara model yang dipretrain pada ImageNet dengan AUC 0,743 dan recall 0,233, namun secara keseluruhan performa model masih rendah karena ketidakseimbangan data dan sekitar 35% label dataset yang tidak akurat. Oleh karena itu, penelitian menyimpulkan bahwa dataset NIH CXR-14 lebih cocok digunakan untuk pretraining self-supervised learning daripada langsung digunakan sebagai dataset klasifikasi, serta pelatihan model menggunakan dataset citra medis khusus berpotensi meningkatkan performa di masa depan.</p>
3.	(Muhammad rifqi.,2024)	Vision Transformer (supervised)	Penerapan CNN dengan arsitektur Vision Transformer menghasilkan klasifikasi

No	Referensi	Metode penelitian	Hasil penelitian
			<p>performa yang cukup baik terhadap seluruh dataset, sehingga memungkinkan klasifikasi penyakit daun kopi dilakukan secara presisi. Dari evaluasi hasil pelatihan, model mencapai tingkat akurasi optimal sebesar 97%, dengan konfigurasi hyperparameter dengan jumlah epoch 20, ukuran batch 16, serta optimizer adam. Selain itu, saat diuji pada data testing (1000 gambar per kelas) model menunjukkan akurasi sebesar 97%, presisi 98% recall 97% dan F-1score 97% untuk masing-masing kelas.</p>
4.	(Dyah Ayu Purbinungtyas., 2024)	Vision Transformers (supervised)	<p>Dari penelitian ini menyimpulkan bahwa pengelompokan tipe kulit wajah melalui pendekatan Vision Transformer sukses meraih performa unggul dalam mengidentifikasi jenis kulit. Model berbasis Vision Transformer yang diimplementasikan berhasil mencapai metrik penilaian yang</p>

No	Referensi	Metode penelitian	Hasil penelitian
			<p>superior, mendapatkan akurasi sebesar 90%, presisi 90,63%, recall 90,63%, dan skor F1 90,62%. Penelitian ini menunjukkan bahwa model yang dikembangkan dapat dipercaya untuk mengenali jenis kulit dengan tingkat presisi yang tinggi, sehingga pendekatan ini memiliki kemungkinan besar untuk diterapkan pada sistem otomatis klasifikasi kulit pada wajah.</p>
5.	(Maharani Nurul Izza1& Gede Putra Kusuma.,2024)	Vision Transformer (supervised)	<p>Pada penelitian ini, langkah-langkah pengolahan data dilakukan dengan pemotongan latar, pembesaran gambar, dan penyesuaian ukuran menjadi 224×224 piksel agar sejalan dengan struktur model. Selain itu, terdapat optimasi hiperparameter yang diterapkan melalui Bayesian Optimization serta teknik augmentasi data seperti rotasi, pembalikan, dan penyesuaian kontras untuk meningkatkan performa model.</p>

No	Referensi	Metode penelitian	Hasil penelitian
			<p>Hasil yang diperoleh menunjukkan bahwa Swin Transformer mencatatkan akurasi tertinggi sebesar 84,75%, diikuti oleh Vision Transformer (ViT) yang mencapai 82,25%, dan Data Efficient Image Transformer (DeiT) dengan akurasi 81,12%, semuanya lebih baik daripada hasil penelitian sebelumnya yang menggunakan arsitektur CNN seperti MobileNetV2 dengan akurasi 81,31%. Sebagai kesimpulan, arsitektur yang berbasis transformer, terutama Swin Transformer, dianggap lebih efisien dalam mengekstraksi fitur gambar dan menghasilkan klasifikasi yang lebih tepat dibandingkan dengan metode CNN tradisional. selain itu, penggunaan Bayesian Optimization terbukti efektif dalam meningkatkan akurasi dan meminimalkan overfitting, sehingga pendekatan deep learning berbasis transformer</p>

No	Referensi	Metode penelitian	Hasil penelitian
			memiliki potensi yang besar untuk pengembangan sistem klasifikasi otomatis guna menilai kualitas biji kopi Arabika hijau di Indonesia.
6.	(Arya Pangestu dkk.,2023)	Vision Transformer (supervised)	Pada penelitian ini, tahapan pisang dikategorikan menjadi empat kelompok, yang mana masing-masing mencerminkan perubahan warna kulit buah pada setiap fase. Temuan dari evaluasi cross-dataset mengindikasikan bahwa model Vision Transformer -L/16-in21k yang telah pre-trained menggunakan dataset ImageNet-21k, mencapai akurasi secara keseluruhan paling unggul di antara varian Vision Transformer dengan nilai 0,9161, ementara model AlexNet yang pre-trained pada dataset ImageNet-1k memperoleh akurasi tertinggi di antara model Convolutional Neural Network (CNN) sebesar 0,7449.Fakta ini menggarisbawahi superioritas model Vision Transformer (ViT)

No	Referensi	Metode penelitian	Hasil penelitian
			<p>dalam hal kemampuan generalisasi dan fleksibilitas penggunaan yang lebih luas, sedangkan CNN unggul dalam hal ukuran model yang lebih ringkas serta durasi pelatihan yang lebih efisien.</p>
7.	(Rayhan Wahyu Taufiqurrahman., 2025)	Vision Transformer (supervised)	<p>Penelitian ini sukses dalam mengimplementasikan serta memverifikasi sebuah alur kerja optimasi yang terorganisir meliputi sparse training, pruning dan fine tuning pada struktur <i>Vision Transformer</i> serta Swin Transformer untuk tugas pengelompokan gambar. Pendekatan ini terbukti sangat efektif, dibuktikan oleh kemampuannya dalam memangkas jumlah parameter model secara substansial yaitu hingga 29,7% pada ViT dan 29,9% pada Swin transformer. Arsitektur Transformer yang telah dioptimasi menampilkan superioritas kompetitif yang tangguh dibandingkan dengan</p>

No	Referensi	Metode penelitian	Hasil penelitian
			metode Convolutional Neural Network (CNN). Model ViT berhasil meraih akurasi maksimal sebesar 72,95%, sementara Swin Transformer mencatatkan nilai 70,52%, yang menempatkan keduanya pada tingkat performa unggul jika dibandingkan dengan hasil tipikal CNN untuk tugas yang sejenis.

2.2 Klasifikasi

Klasifikasi merupakan proses pengelompokan di mana objek atau entitas dengan karakteristik yang serupa ditempatkan dalam satu kelompok, sedangkan entitas yang memiliki perbedaan dipisahkan ke kelompok lain. Proses ini dilakukan dengan mengelompokkan data berdasarkan parameter hasil ekstraksi fitur, yang menekankan pada kemiripan ciri di antara setiap data. Terdapat berbagai pendekatan yang dapat digunakan untuk melakukan klasifikasi, salah satunya adalah dengan menerapkan *Deep Learning*. Sebagai bagian dari *Machine Learning*, *Deep Learning* memiliki kemampuan untuk mempelajari pola dan metode secara otomatis melalui proses komputasi (Mayasari dkk., 2022).

Dalam proses klasifikasi, terdapat dua tahapan utama, yaitu tahap *learning* atau *training* dan tahap *testing*. Pada tahap *training*, dilakukan pembangunan model dengan memanfaatkan data latih agar sistem mampu mengenali pola dari data tersebut. Selanjutnya, pada tahap *testing*, model yang telah terbentuk akan diuji menggunakan data uji untuk menilai sejauh mana performa model dalam melakukan klasifikasi (Siti Raysyah dkk., 2021). Namun, dengan menggunakan

metode *self-supervised*, proses pengklasifikasian dapat dilakukan tanpa perlu memberikan label terlebih dahulu pada data akan tetapi model ini akan belajar sendiri dari datanya sendiri. Pada pendekatan ini, sistem secara otomatis mengelompokkan data berdasarkan kemiripan fitur yang diekstraksi dari citra. Misalnya, melalui penggunaan *Vision Transformer* yang berperan dalam mengekstraksi. Representasi fitur yang dihasilkan dari proses ini menjadi dasar untuk sistem untuk mengklasifikasikan tingkat kematangan buah kopi secara otomatis, tanpa harus melalui pelabelan manual oleh manusia. Dengan pendekatan ini, model bisa belajar secara mandiri dan menemukan pola visual yang bermakna dari data gambar yang tersedia

2.3 Buah kopi Arabica Gayo

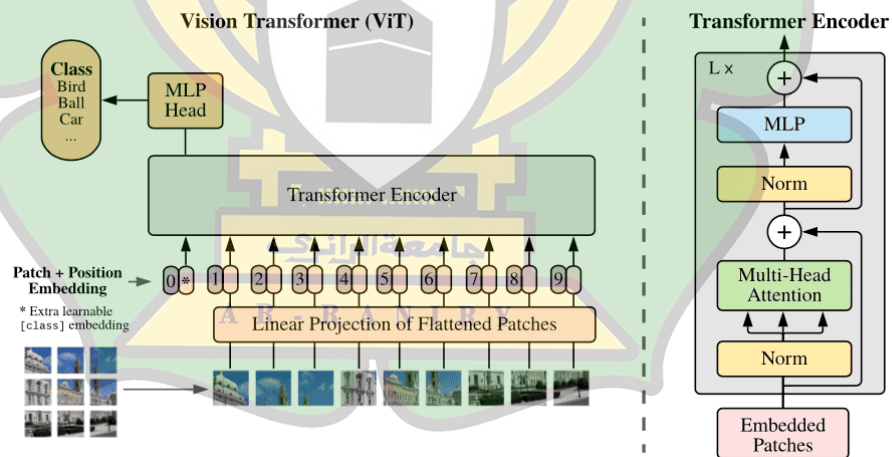
Kopi merupakan salah satu komoditas perkebunan yang memiliki peran penting dan banyak diminati oleh masyarakat di seluruh dunia. Di Indonesia, kopi menjadi salah satu komoditas ekspor unggulan dengan nilai perdagangan yang cukup tinggi. Selain itu, kegiatan perkebunan kopi juga memiliki peran penting dalam meningkatkan kesejahteraan masyarakat sekitar melalui penyediaan lapangan kerja dan sumber pendapatan utama bagi petani. Indonesia sendiri menempati peringkat keempat sebagai negara pengekspor dan produsen kopi terbesar di dunia, setelah Brasil, Kolombia, dan Vietnam. Beberapa jenis kopi dibudidayakan di Indonesia, namun yang paling banyak ditanam oleh masyarakat adalah kopi Arabika dan Robusta (Siti Raysyah dkk., 2021).

Kualitas biji serta cita rasa kopi Arabica Gayo sangat dipengaruhi oleh faktor lingkungan, terutama ketinggian lokasi penanaman dan tingkat kematangan buah saat dipanen. Variasi tingkat kematangan buah kopi yang sering tercampur dalam proses panen dapat berdampak pada penurunan mutu dan cita rasa akhir dari kopi yang dihasilkan (Yusya Abubakar dkk., 2022). Kopi Arabika dan Robusta memiliki perbedaan dalam waktu panen serta proses pengolahan pascapanennya. Kopi Arabika umumnya memiliki cita rasa dan aroma yang lebih disukai oleh konsumen dibandingkan dengan kopi Robusta. Kualitas cita rasa yang lebih

unggul tersebut menjadikan kopi Arabika memiliki nilai jual yang lebih tinggi di pasar (Maya Saputri dkk., 2020).

2.4 Vision Transformer

Vision transformer adalah model deep learning yang dapat digunakan untuk klasifikasi dan analisis citra gambar dengan menggunakan arsitektur transformer. Vision Transformer merupakan perkembangan terkini dalam bidang deep learning yang mengadaptasi arsitektur Transformer, yang sebelumnya banyak digunakan dalam pemrosesan bahasa alami. Dalam pengolahan data visual, Vision Transformer menetapkan mekanisme patch embedding, multi-head attention, serta multi-layer perceptron untuk memproses citra, dan berorientasi pada hubungan antar bagian citra. Gambar input dibagi menjadi beberapa patch berdasarkan ukuran arsitekturnya selama proses penerapan patch. Setelah itu, gambar diberi position embeddings, yang menunjukkan lokasi patch pada gambar. Setelah itu, patch akan dimasukkan ke multi-head attention (Tri Febriyanto & Syuzuki Syofian., 2024).



Gambar 2. 1 Arsitektur Vision Transformer

Pada gambar 2.1 merupakan Arsitektur *Vision Transformer* dapat dilihat proses pemecahan citra menjadi beberapa *patches* (potongan dengan ukuran

seragam). Setiap *patch* kemudian dikonversi menjadi vektor melalui tahap yang disebut *linear embedding*. Setelah itu, ditambahkan *positional embedding* untuk memberikan informasi posisi pada setiap vektor. Selanjutnya, deretan vektor yang telah terbentuk dimasukkan ke dalam *encoder transformer* standar (Dosovitskiy dkk., 2021).

2.5 DINOv3 (Distillation with NO labels)

DINOv3 adalah model vision foundation berbasis Self-Supervised Learning yang dikembangkan untuk memahami dan mengekstraksi fitur dari gambar tanpa membutuhkan label data secara manual. DINOv3 menggunakan arsitektur Vision Transformer dan dilatih menggunakan jumlah data gambar yang sangat besar agar mampu menghasilkan representasi fitur visual yang kuat dan serbaguna. Secara sederhana, DINOv3 bekerja dengan mempelajari pola, bentuk, tekstur, warna, serta hubungan antar objek pada gambar secara otomatis, sehingga menghasilkan feature embedding atau representasi fitur yang dapat digunakan dalam berbagai tugas computer vision seperti klasifikasi gambar, clustering, segmentasi, dan deteksi objek. DINOv3 dirancang untuk memiliki fleksibilitas tinggi melalui peningkatan ukuran model dan jumlah data pelatihan. Model ini mampu mencapai performa yang sangat baik meskipun dalam kondisi *frozen* tanpa *fine-tuning*, sehingga satu forward pass dapat digunakan untuk berbagai tugas computer vision dengan efisiensi komputasi yang tinggi, terutama pada perangkat edge. Selain itu, *pipeline Self-Supervised Learning* yang tidak bergantung pada metadata memungkinkan DINOv3 melakukan generalisasi pada berbagai domain gambar, baik dari web maupun data observasi, serta menghasilkan fitur visual berkualitas tinggi yang mampu membedakan objek dengan jelas (Oriane Siméoni dkk., 2025).

2.6 Self-supervised Learning

Model Transformer berbasis *self-attention* umumnya beroperasi menggunakan mekanisme pelatihan dua tahap. Pada tahap pertama, pre-training dilakukan menggunakan dataset berskala besar (Yen-Chun Chen dkk., 2020).

dengan pendekatan supervised atau self-supervised Selanjutnya, bobot hasil pre-training tersebut diadaptasi untuk tugas lanjutan (*downstream tasks*) menggunakan dataset berskala kecil hingga menengah (Xiujun Li dkk., 2020). Contoh tugas lanjutan meliputi klasifikasi citra, deteksi objek klasifikasi zero-shot, question-answering, dan pengenalan aksi. Efektivitas pre-training pada Transformer berskala besar telah banyak dibuktikan baik pada domain bahasa maupun visi komputer. Sebagai contoh, model *Vision Transformer* (ViT-L) (Alexey Dosovitskiy dkk., 2020) mengalami penurunan akurasi absolut sebesar 13% pada dataset uji ImageNet ketika hanya dilatih menggunakan dataset pelatihan ImageNet saja, dibandingkan dengan model yang terlebih dahulu melalui tahap pre-training menggunakan dataset JFT yang berisi sekitar 300 juta gambar (Yen-Chun Chen dkk., 2020).

Self-supervised learning dimanfaatkan secara luas pada tahap pre-training karena proses pelabelan data secara manual dalam skala besar membutuhkan waktu, tenaga, dan biaya yang tinggi. Pendekatan pre-training berbasis self-supervision memiliki peran penting dalam meningkatkan kemampuan skalabilitas dan generalisasi model, khususnya pada arsitektur Transformer, sehingga memungkinkan pelatihan model dengan jumlah parameter yang sangat besar. Konsep utama dari self-supervised learning adalah mendorong model untuk mempelajari representasi data yang bermakna melalui penyelesaian tugas pendahulu, seperti memprediksi bagian data yang disembunyikan pada citra, memperkirakan informasi temporal pada urutan video, atau menentukan atribut tertentu dari data masukan, termasuk rotasi, susunan potongan citra, dan karakteristik visual lainnya. Selain itu, pendekatan contrastive learning juga banyak digunakan dalam self-supervised learning dengan memanfaatkan berbagai transformasi data untuk menghasilkan pasangan citra yang memiliki kesamaan maupun perbedaan semantik, sehingga model mampu membedakan informasi yang relevan dan tidak relevan dalam proses pembelajaran (Salman Khan dkk., 2022).

2.7 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan metode statistik yang digunakan untuk menyederhanakan data yang memiliki banyak variabel atau dimensi menjadi lebih sedikit variabel tanpa menghilangkan informasi penting yang terdapat di dalam data tersebut. Dataset yang memiliki banyak fitur biasanya lebih sulit dianalisis, divisualisasikan, maupun diproses dalam model machine learning. Oleh karena itu, PCA digunakan untuk mereduksi dimensi data dengan cara mengubah variabel asli menjadi beberapa komponen baru yang disebut principal components atau komponen utama. Komponen ini merupakan kombinasi dari variabel-variabel asli yang mampu mewakili sebagian besar variasi atau informasi yang terdapat dalam dataset.

Proses PCA dimulai dengan melakukan standardisasi data agar semua variabel memiliki skala yang sama sehingga tidak ada variabel yang lebih dominan. Setelah itu dilakukan analisis hubungan antar variabel untuk melihat pola yang terbentuk dalam data. Selanjutnya PCA menentukan arah penyebaran data yang paling besar dan membentuk komponen utama berdasarkan variasi tersebut. Komponen pertama menyimpan informasi paling besar dari data, diikuti oleh komponen kedua dan seterusnya. Namun biasanya hanya beberapa komponen utama saja yang dipilih karena sudah mampu mewakili sebagian besar informasi dari dataset, sedangkan komponen lainnya dapat diabaikan karena kontribusinya kecil atau hanya berupa noise. Dengan cara ini, data yang kompleks dapat disederhanakan sehingga lebih mudah dianalisis dan diproses dalam berbagai metode machine learning.

2.8 Fuzzy C-Means

Fuzzy C-Means (FCM) adalah algoritma clustering yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok (cluster) berdasarkan tingkat kemiripannya. Berbeda dengan metode clustering biasa seperti K-Means, pada FCM setiap data tidak hanya menjadi anggota satu cluster saja, tetapi dapat memiliki tingkat keanggotaan pada beberapa cluster sekaligus. Tingkat

keanggotaan tersebut dinyatakan dalam nilai antara 0 sampai 1, yang menunjukkan seberapa besar suatu data termasuk ke dalam cluster tertentu.

Proses kerja FCM dimulai dengan menentukan jumlah cluster yang diinginkan, kemudian algoritma akan memberikan nilai keanggotaan awal untuk setiap data terhadap setiap cluster. Setelah itu, sistem akan menghitung pusat cluster (centroid) berdasarkan nilai keanggotaan tersebut. Selanjutnya nilai keanggotaan setiap data akan diperbarui berdasarkan jaraknya terhadap pusat cluster. Proses ini dilakukan secara berulang hingga nilai keanggotaan dan pusat cluster menjadi stabil atau tidak banyak berubah lagi. Dengan pendekatan ini, FCM mampu menangani data yang memiliki batas antar kelompok yang tidak terlalu jelas, sehingga sering digunakan dalam berbagai bidang seperti data mining, pengolahan citra, pengenalan pola, dan machine learning.

2.7 Deep Learning

Deep learning merupakan suatu teknik pendekatan berbasis jaringan saraf tiruan, yang telah menjadi salah satu metode utama dalam *Machine Learning* dan semakin populer dalam beberapa tahun terakhir. Banyak penelitian mengatakan bahwa *Deep Learning* bukan hanya terbatas pada bidang tertentu, melainkan dapat dianggap sebagai pendekatan pembelajaran yang bersifat umum dan mampu menangani berbagai jenis masalah di berbagai sektor. Publikasi ilmiah yang membahas penerapan *Deep Learning* dalam bidang Educational Data Mining (EDM) pertama kali muncul pada tahun 2015, dan jumlahnya mengalami peningkatan signifikan dalam setiap tahunnya (Muhammad Haris Diponegoro dkk.,2022).

Pada tahun 2006, *Deep Learning* mulai dikenal dan telah diterapkan di berbagai bidang, seperti pengenalan suara, pengolahan citra, serta pemrosesan bahasa alami. Sebagai bagian dari *Machine Learning*, *Deep Learning* membedakan dirinya melalui struktur yang terdiri dari banyak lapisan yang disusun secara bertumpuk. Kehadiran teknologi ini telah membawa efisiensi yang signifikan dalam waktu pemrosesan dan pembelajaran (Anshori.,2023).

Ada dua kategori utama dalam *Deep Learning* yang dapat diidentifikasi. Yang pertama, *Deep Learning* untuk pembelajaran tanpa pengawasan (*unsupervised Learning*), yang diterapkan ketika label dari target tidak tersedia. Pada pendekatan ini, fokus utama adalah mengidentifikasi korelasi yang lebih tinggi dari nilai-nilai yang diamati untuk mengungkap pola tersembunyi. Dalam bidang pengenalan gambar, *Deep Learning* memungkinkan sistem untuk mendeteksi dan mengenali fitur-fitur kunci seperti huruf atau wajah melalui analisis gambar dan video. Proses ini melibatkan langkah-langkah seperti pemisahan objek dari latar belakang, ekstraksi fitur, konversi data, hingga pengenalan fitur target secara akurat (Vincent.,2023).

2.8 Metode Evaluasi

Evaluasi pada penelitian ini dilakukan untuk mengukur kinerja model dalam mengklasifikasikan tingkat kematangan buah kopi arabika Gayo secara objektif. Proses evaluasi dibagi menjadi dua tahap utama, yaitu evaluasi hasil clustering dan evaluasi performa klasifikasi. Pada tahap clustering, kualitas pengelompokan data dievaluasi menggunakan tiga metrik, yaitu Silhouette Score, Davies-Bouldin Index (DBI), dan Fuzzy Partition Coefficient (FPC). Silhouette Score digunakan untuk mengukur seberapa baik suatu data ditempatkan dalam cluster-nya dibandingkan dengan cluster lain. Davies-Bouldin Index digunakan untuk menilai tingkat pemisahan antar cluster, di mana nilai yang lebih kecil menunjukkan kualitas cluster yang lebih baik. Sedangkan Fuzzy Partition Coefficient digunakan untuk mengukur tingkat kejelasan keanggotaan data pada masing-masing cluster dalam metode Fuzzy C-Means.

Selanjutnya, pada tahap klasifikasi, evaluasi dilakukan menggunakan data uji (testing set) yang tidak digunakan selama proses pelatihan. Hal ini bertujuan untuk mengukur kemampuan generalisasi model terhadap data baru. Kinerja model diukur menggunakan metrik evaluasi berupa accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk mengetahui tingkat ketepatan prediksi secara keseluruhan, sedangkan precision dan recall digunakan untuk mengevaluasi

kemampuan model dalam mengidentifikasi setiap kelas secara spesifik. F1-score merupakan rata-rata harmonis dari precision dan recall yang memberikan gambaran keseimbangan kinerja model.

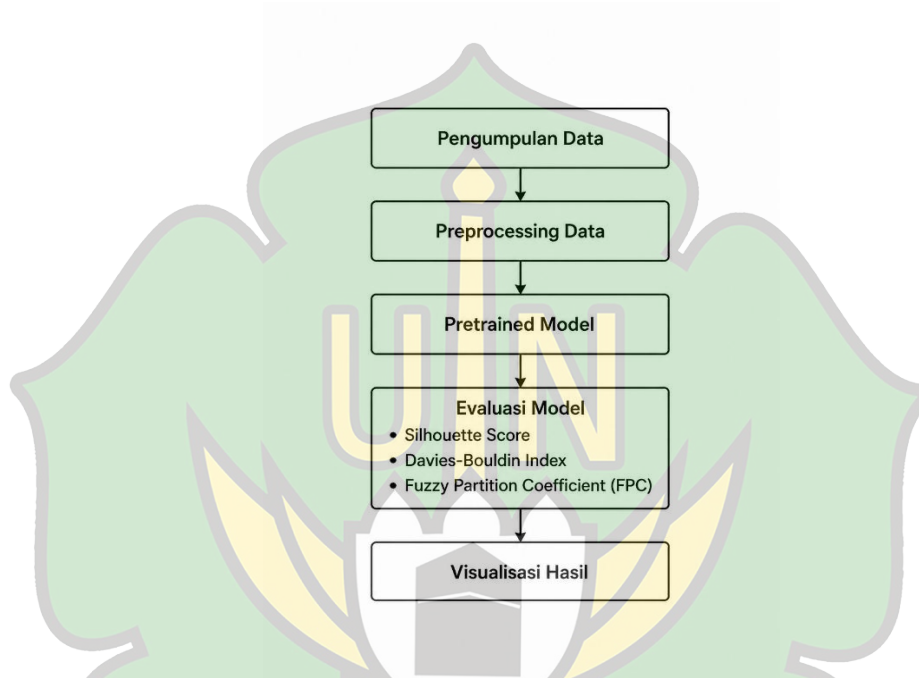
Selain itu, digunakan juga confusion matrix untuk menganalisis distribusi prediksi model terhadap setiap kelas, sehingga dapat diketahui pola kesalahan klasifikasi yang terjadi. Melalui kombinasi metrik evaluasi tersebut, performa model dapat dianalisis secara komprehensif dalam mengklasifikasikan tingkat kematangan buah kopi.



BAB III METODE PENELITIAN

3.1 Tahapan penelitian

Adapun langkah-langkah yang dilakukan pada penelitian ini dapat dilihat pada gambar 3.1.



Gambar 3. 1 Tahapan penelitian

Gambar 3.1 menunjukkan alur proses sistem dalam mengolah gambar mulai dari tahap memasukkan data gambar, melakukan penyesuaian kualitas gambar, mengambil informasi penting, mempelajari pola dari data yang tersedia, hingga menentukan kategori akhir dari gambar tersebut. Setiap tahapan saling terhubung untuk menghasilkan proses klasifikasi yang lebih akurat dan efisien.

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil langsung dari lapangan, yaitu kebun Bapak Agus Triono tepatnya di kampung Atu Lintang, kecamatan Atu

Lintang, kabupaten Aceh Tengah. Data yang digunakan dalam penelitian ini berupa gambar buah kopi Arabica Gayo dengan tingkat kematangan yang berbeda-beda diantaranya yaitu mentah (hijau), setengah matang (merah kekuningan), dan matang (merah). Jumlah gambar yang diambil untuk penelitian ini sebanyak 3000 gambar untuk keseluruhannya. Proses pengambilan gambar langsung dilakukan di kebun menggunakan kamera ponsel Iphone xr dengan resolusi 12 MP, setiap gambar diambil dalam kondisi pencahayaan alami pada siang hari dengan jarak pengambilan sekitar 30-50 cm dari objek. Setiap buah yang difoto dipilih dari beberapa pohon yang berbeda untuk menjaga keberagaman data.

Dataset yang digunakan dalam penelitian ini merupakan citra buah kopi Arabika Gayo yang terdiri dari tiga kategori tingkat kematangan, yaitu mentah, setengah matang, dan matang. Dataset disimpan dalam bentuk folder berdasarkan kelas masing-masing.

Tabel 3. 1 Dataset Penelitian

No	kelas	Jumlah gambar	Jumlah gambar <i>training</i> (80%)	Jumlah gambar <i>testing</i> (20%)
1.	Mentah	1.000	800	200
2.	Setengah matang	1.000	800	200
3.	matang	1.000	800	200
Total		3.000	2.400	600

3.3 Preprocessing Data

Tahap preprocessing data dilakukan untuk meningkatkan kualitas gambar sebelum proses ekstraksi fitur dan clustering. Proses ini bertujuan untuk mengurangi noise, menyeragamkan format data, serta memastikan bahwa informasi yang diproses oleh model berfokus pada objek utama. Adapun tahapan preprocessing yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Penghapusan Background

Pada tahap awal, dilakukan penghapusan latar belakang (background removal) pada citra menggunakan library `rembg`. Proses ini bertujuan untuk memisahkan objek utama dari latar belakang sehingga fitur yang diekstraksi lebih representatif terhadap objek yang dianalisis.

2. Konversi Format dan Penyesuaian Background

Citra hasil penghapusan background dikonversi ke dalam format RGBA, kemudian ditempatkan pada background berwarna hitam (RGB). Langkah ini dilakukan untuk menghindari transparansi pada citra serta menyeragamkan kondisi latar belakang seluruh dataset sehingga meminimalkan variasi yang tidak relevan.

3. Penyesuaian Ukuran dan Normalisasi gambar

Selanjutnya, citra diproses menggunakan `AutoImageProcessor` dari model `DINOv3`. Pada tahap ini dilakukan `resizing` citra ke ukuran yang sesuai dengan input model serta normalisasi nilai piksel. Proses ini bertujuan untuk memastikan bahwa seluruh citra memiliki format yang konsisten dan sesuai dengan kebutuhan model.

4. Normalisasi Fitur

Setelah fitur citra diekstraksi, dilakukan normalisasi menggunakan metode `Normalizer`. Tahap ini bertujuan untuk menyamakan skala antar fitur sehingga tidak ada fitur yang mendominasi dalam proses clustering.

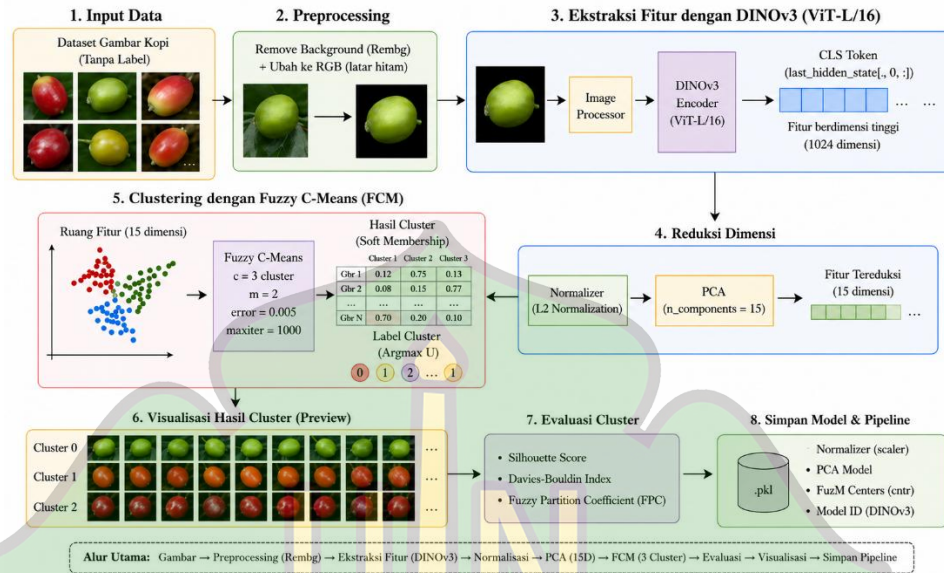
5. Reduksi Dimensi

Tahap terakhir dalam preprocessing adalah reduksi dimensi menggunakan metode `Principal Component Analysis (PCA)`. Reduksi dimensi dilakukan untuk mengurangi kompleksitas data, menghilangkan redundansi fitur, serta meningkatkan efisiensi proses clustering tanpa menghilangkan informasi penting.

3.4 Pretrained Model (DINOv3)

Pada penelitian ini, digunakan model *pretrained* yaitu `DINOv3` dengan arsitektur *Vision Transformer (ViT)* untuk melakukan ekstraksi fitur citra. Model yang digunakan adalah `facebook/dinov3-vitl16-pretrain-lvd1689m` yang telah

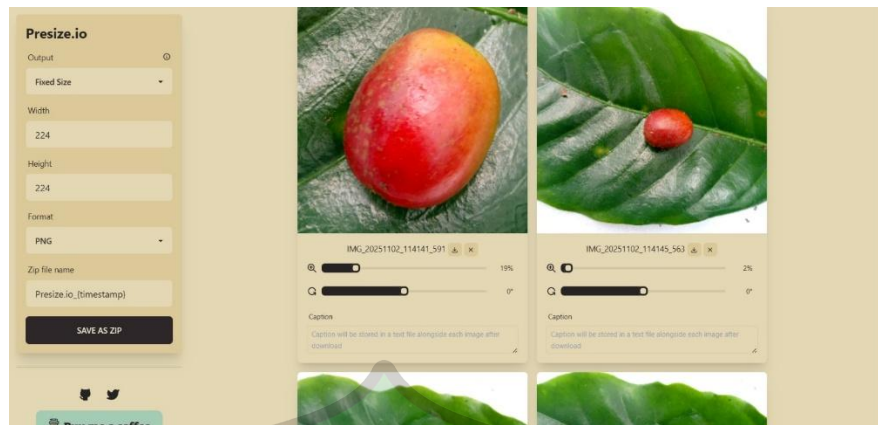
dilatih sebelumnya menggunakan metode *self-supervised learning* pada dataset skala besar. Alur sistem ekstraksi DINOv3 dapat dilihat pada gambar 3.2.



Gambar 3. 2 Alur sistem ekstraksi fitur DINOv3

3.1.1. Input image

Tahap pertama yang dilakukan adalah *input image*, yaitu proses pengumpulan dan persiapan data mentah berupa gambar buah kopi Arabica Gayo yang akan digunakan untuk masukan utama bagi sistem klasifikasi. Pada tahap ini, penulis menggunakan kumpulan gambar buah kopi dari berbagai tingkat kematangan, yaitu mentah, setengah matang, dan matang. Setiap gambar diambil dalam kondisi pencahayaan dan latar belakang yang bermacam-macam agar model dapat mempelajari perbedaan visual agar model dapat mempelajari perbedaan gambar secara menyeluruh. Sebelum digunakan, gambar diperiksa untuk memastikan kualitasnya, seperti kejernihan pada gambar, pencahayaan yang seimbang, serta minim gambar yang dapat memengaruhi hasil analisis.



Gambar 3. 3 Image Resizing

Pada gambar 3.2 merupakan proses standarisasi format dan ukuran gambar agar sesuai dengan kebutuhan model *Vision Transformer*. Seluruh gambar diubah ke ukuran tetap yaitu 224×224 piksel, dan disimpan dalam format seragam seperti PNG. Standarisasi ini diperlukan agar model dapat memproses setiap gambar secara konsisten pada tahap *patch embedding*. Dalam konteks *self-supervised learning*, seluruh data pada tahap ini masih bersifat tidak berlabel, sehingga model akan belajar secara mandiri dari pola-pola gambar yang terdapat dalam citra. Dengan demikian, tahap input image berfungsi sebagai dasar pembelajaran bagi model dalam mengenali karakteristik visual buah kopi dan menjadi fondasi penting bagi keberhasilan proses klasifikasi pada tahap berikutnya.

3.1.2. Normalization R - R A N I R Y

Tahap normalisasi merupakan salah satu proses *preprocessing* yang dilakukan setelah proses ekstraksi fitur dari gambar. Pada penelitian ini, fitur gambar diperoleh menggunakan model *Vision Transformer* berbasis DINOv3 yang menghasilkan representasi data dalam bentuk *embedding vector*. Vektor fitur tersebut memiliki dimensi tinggi yang merepresentasikan karakteristik visual objek pada gambar buah kopi.

Proses normalisasi dilakukan menggunakan metode L2 melalui fungsi *Normalizer* dari pustaka Scikit-learn. Tujuan dari proses ini adalah untuk menyeragamkan skala nilai pada setiap vektor fitur sehingga seluruh data memiliki panjang vektor yang relatif sama tanpa mengubah arah representasi informasi yang terkandung di dalamnya. Dengan demikian, perbedaan magnitudo yang dapat muncul akibat variasi pencahayaan atau kondisi pengambilan gambar tidak akan memengaruhi proses analisis fitur.

Normalisasi menjadi langkah penting karena model transformer seperti DINOv3 menyimpan informasi semantik objek lebih pada pola hubungan antar elemen vektor dibandingkan pada besar kecilnya nilai fitur. Dengan menerapkan normalisasi, proses analisis fitur akan lebih berfokus pada kemiripan pola karakteristik visual objek sehingga representasi data menjadi lebih konsisten.

Selain itu, proses normalisasi juga berpengaruh terhadap tahap pengelompokan menggunakan algoritma Fuzzy C-Means. Algoritma ini menentukan tingkat keanggotaan data terhadap suatu kluster berdasarkan perhitungan jarak antar data. Dengan adanya normalisasi, perhitungan jarak antar vektor fitur menjadi lebih stabil sehingga proses pengelompokan dapat dilakukan secara lebih akurat. Hal ini memungkinkan sistem untuk mengelompokkan buah kopi berdasarkan karakteristik visualnya sehingga klasifikasi tingkat kematangan dapat dilakukan dengan lebih optimal.

3.1.3. Self-Distillation

Self-distillation merupakan salah satu mekanisme pembelajaran yang digunakan dalam pendekatan *self-supervised learning*, model dapat mempelajari representasi data tanpa memerlukan label anotasi dari manusia. Pada penelitian ini, proses ekstraksi fitur dilakukan menggunakan model DINOv3, yang merupakan pengembangan dari arsitektur *Vision Transformer*. Model DINOv3 bekerja menggunakan prinsip *Self-Distillation with No Labels*, yaitu suatu metode pembelajaran yang memungkinkan model untuk berperan sebagai pengajar (*teacher*) sekaligus sebagai pelajar (*student*) dalam proses pelatihannya.

Pada pendekatan ini, dua jaringan dengan arsitektur yang sama digunakan secara bersamaan, yaitu jaringan *teacher* dan jaringan *student*. Kedua jaringan tersebut memproses gambargambar yang sama namun dengan variasi tampilan yang berbeda. Umumnya, citra akan dipotong menjadi beberapa bagian dengan skala yang berbeda, yaitu *global view* yang menampilkan objek secara keseluruhan dan *local view* yang menampilkan sebagian detail objek. Jaringan *student* dilatih menggunakan potongan citra skala kecil (*local crops*) untuk memprediksi representasi fitur dari citra skala besar (*global crops*) yang diproses oleh jaringan *teacher*. Melalui proses ini, jaringan *student* secara bertahap belajar memahami hubungan antara bagian kecil objek dengan bentuk keseluruhan objek.

Selama proses pelatihan berlangsung, bobot pada jaringan *teacher* tidak diperbarui secara langsung melalui proses *backpropagation*. Sebaliknya, bobot jaringan *teacher* diperbarui secara bertahap berdasarkan bobot jaringan *student* menggunakan mekanisme *Exponential Moving Average* (EMA). Pendekatan ini memungkinkan jaringan *teacher* menghasilkan representasi yang lebih stabil sehingga dapat membimbing proses pembelajaran jaringan *student* secara lebih efektif.

Melalui mekanisme *self-distillation* ini, model DINOv3 mampu menghasilkan representasi fitur atau *embedding* yang kaya akan informasi semantik. Representasi tersebut tidak hanya merekam informasi piksel dari citra, tetapi juga mampu menangkap karakteristik visual yang lebih kompleks seperti bentuk objek, tekstur permukaan, serta variasi warna. Dalam konteks penelitian ini, fitur yang dihasilkan oleh model DINOv3 digunakan sebagai dasar untuk proses analisis lebih lanjut menggunakan metode reduksi dimensi dan pengelompokan data. Dengan demikian, mekanisme *self-distillation* berperan penting dalam menghasilkan representasi fitur yang informatif sehingga dapat mendukung proses identifikasi tingkat kematangan buah kopi secara lebih akurat.

3.1.4. Representation Learning

representation learning pada penelitian ini dilakukan dengan memanfaatkan model DINOv3 yang telah melalui proses *pretraining* menggunakan pendekatan *self-supervised learning*. Pada tahap ini, model digunakan sebagai *feature extractor* untuk mengubah data citra menjadi representasi fitur dalam bentuk vektor numerik berdimensi tinggi. Proses *representation learning* dimulai dengan memasukkan citra hasil preprocessing ke dalam model. Sebelumnya, citra terlebih dahulu melalui proses penghapusan latar belakang menggunakan library *rembg*, kemudian objek utama ditempatkan pada latar belakang berwarna hitam. Langkah ini bertujuan untuk mengurangi noise dari latar belakang sehingga model dapat lebih fokus pada objek utama.

Selanjutnya, citra diproses menggunakan *image processor* yang sesuai dengan model DINOv3 untuk menghasilkan tensor input. Tensor tersebut kemudian dimasukkan ke dalam model tanpa dilakukan pelatihan ulang (*fine-tuning*), karena parameter model telah dibekukan (*frozen*). Hal ini menunjukkan bahwa proses pembelajaran representasi sepenuhnya mengandalkan kemampuan model hasil *pretraining*. Representasi fitur diambil dari output model pada bagian *last_hidden_state*, tepatnya pada token pertama ([CLS] token), yang berfungsi sebagai representasi global dari keseluruhan citra. Fitur ini berupa vektor numerik yang mengandung informasi penting seperti bentuk, tekstur, dan pola visual dari objek.

```
outputs = model(**inputs)
feature = outputs.last_hidden_state[:, 0, :]
```

Gambar 3. 4 Proses pengambilan fitur gambar dari model DINOv3.

Vektor fitur yang dihasilkan kemudian dikumpulkan dari seluruh data citra untuk membentuk sekumpulan representasi yang akan digunakan pada tahap selanjutnya. Karena fitur yang dihasilkan memiliki dimensi yang tinggi,

dilakukan proses normalisasi menggunakan *Normalizer* untuk menyesuaikan skala data, serta reduksi dimensi menggunakan *Principal Component Analysis (PCA)* agar fitur menjadi lebih ringkas namun tetap mempertahankan informasi penting.

Dengan demikian, tahap *representation learning* dalam penelitian ini berperan sebagai proses transformasi data citra menjadi representasi numerik yang informatif. Representasi inilah yang menjadi dasar dalam proses clustering menggunakan metode Fuzzy C-Means (FCM), sehingga kualitas fitur yang dihasilkan sangat mempengaruhi hasil pengelompokan data.

3.1.5. Pretext Task

Pretext task merupakan bagian dari pendekatan *self-supervised learning* yang bertujuan untuk membuat model mampu memahami representasi data tanpa menggunakan label. Pada tahap ini, model tidak belajar dari data berlabel, melainkan dari tugas buatan (*pretext*) yang dirancang untuk memaksa model mengenali pola dalam data. Dalam penelitian ini, tahap *pretext task* tidak dilakukan secara langsung oleh penulis, melainkan telah dilakukan sebelumnya pada saat proses *pretraining* model DINOv3. Model tersebut telah dilatih menggunakan strategi *self-distillation*, di mana model belajar menghasilkan representasi fitur yang konsisten dari berbagai variasi (augmentasi) gambar yang sama.

Melalui mekanisme tersebut, model DINOv3 mampu memahami karakteristik visual penting, seperti bentuk, tekstur, dan struktur objek, tanpa memerlukan anotasi label. Hasil dari proses ini adalah representasi fitur yang bersifat umum (*general-purpose*) dan dapat digunakan kembali pada berbagai tugas lanjutan (*downstream tasks*). Pada penelitian ini, representasi fitur yang dihasilkan oleh DINOv3 kemudian dimanfaatkan secara langsung tanpa dilakukan pelatihan ulang (*fine-tuning*). Fitur tersebut selanjutnya diproses pada tahap *feature projection* menggunakan PCA untuk mengurangi dimensi data, sebelum akhirnya digunakan dalam proses pengelompokan menggunakan

metode Fuzzy C-Means (FCM). Dengan demikian, dapat disimpulkan bahwa peran *pretext task* dalam penelitian ini adalah sebagai dasar pembelajaran awal yang memungkinkan model menghasilkan fitur yang informatif, sehingga mendukung keberhasilan proses clustering meskipun tanpa penggunaan label.

3.1.6. Feature projection

Tahap *feature projection* merupakan proses penting yang bertujuan untuk mentransformasikan fitur hasil ekstraksi dari model menjadi representasi yang lebih sederhana, efisien, dan mudah diolah pada tahap selanjutnya tanpa menghilangkan informasi penting yang terkandung di dalamnya. Dalam penelitian ini, fitur awal diperoleh dari model DINOv3 yang menghasilkan vektor dengan dimensi tinggi melalui representasi *embedding*. Vektor fitur tersebut umumnya memiliki ukuran yang besar dan kompleks, sehingga apabila digunakan secara langsung dapat menyebabkan proses clustering menjadi kurang optimal, baik dari segi waktu komputasi maupun kualitas pengelompokan. Oleh karena itu, diperlukan tahap pengolahan lanjutan untuk menyederhanakan struktur data tersebut.

Langkah pertama yang dilakukan adalah normalisasi menggunakan metode *Normalizer*, yang berfungsi untuk menyamakan skala nilai pada setiap fitur. Proses ini penting karena perbedaan rentang nilai antar fitur dapat menyebabkan beberapa fitur memiliki pengaruh yang lebih besar dibandingkan yang lain dalam proses analisis. Dengan normalisasi, setiap fitur akan memiliki kontribusi yang lebih seimbang sehingga hasil pengolahan data menjadi lebih stabil dan tidak bias. Setelah normalisasi, dilakukan reduksi dimensi menggunakan metode *Principal Component Analysis (PCA)* dengan jumlah komponen sebanyak 15. PCA bekerja dengan cara memproyeksikan data ke dalam ruang baru yang dibentuk dari kombinasi linier fitur asli, di mana komponen yang dipilih merupakan komponen yang memiliki variansi terbesar. Dengan demikian, informasi utama dari data tetap dapat dipertahankan meskipun jumlah dimensinya telah dikurangi secara signifikan. Proses ini juga membantu

mengurangi noise serta menghilangkan informasi yang redundan, sehingga data menjadi lebih bersih dan representatif.

Hasil dari tahap ini adalah fitur baru yang memiliki dimensi lebih rendah namun tetap mampu merepresentasikan karakteristik utama dari data citra. Selanjutnya, fitur tersebut ditransformasikan ke dalam bentuk matriks dengan format (fitur \times jumlah sampel) agar sesuai dengan kebutuhan input dari algoritma Fuzzy C-Means (FCM). Dengan adanya tahap *feature projection* ini, data yang awalnya kompleks menjadi lebih terstruktur dan efisien, sehingga dapat meningkatkan performa serta kualitas hasil pengelompokan pada tahap clustering.

2.4.1 Fuzzy C-Means (FCM)

Pada penelitian ini, proses pengelompokan data dilakukan menggunakan algoritma Fuzzy C-Means. Algoritma ini digunakan untuk mengelompokkan vektor fitur yang dihasilkan oleh model DINOv3 ke dalam beberapa kluster yang merepresentasikan tingkat kematangan buah kopi Arabica Gayo. Pendekatan Fuzzy C-Means dipilih karena mampu menangani karakteristik data yang memiliki sifat transisi atau perubahan bertahap, seperti perubahan warna pada buah kopi selama proses pematangan.

Berbeda dengan metode pengelompokan konvensional seperti K-Means yang menggunakan pendekatan *hard clustering*, algoritma Fuzzy C-Means menerapkan konsep *soft clustering*. Pada pendekatan ini, setiap data tidak langsung dimasukkan secara mutlak ke dalam satu kluster tertentu, melainkan diberikan nilai derajat keanggotaan terhadap setiap kluster yang tersedia. Nilai tersebut berada pada rentang antara nol hingga satu yang menunjukkan seberapa besar kemungkinan suatu data termasuk ke dalam kluster tertentu. Dengan pendekatan ini, suatu citra buah kopi dapat memiliki tingkat kedekatan dengan lebih dari satu kategori kematangan.

Proses pengelompokan dimulai dengan menginisialisasi matriks keanggotaan secara acak yang merepresentasikan tingkat keanggotaan setiap data terhadap masing-masing kluster. Selanjutnya, algoritma secara iteratif menghitung posisi pusat kluster berdasarkan nilai keanggotaan tersebut. Setelah pusat kluster diperoleh, sistem akan menghitung jarak antara setiap vektor fitur dengan pusat kluster untuk memperbarui nilai derajat keanggotaannya. Proses ini dilakukan secara berulang hingga perubahan nilai keanggotaan menjadi sangat kecil atau telah mencapai kondisi konvergen.

Dalam penelitian ini, jumlah kluster yang digunakan adalah tiga kluster yang merepresentasikan kategori kematangan buah kopi, yaitu mentah, setengah matang, dan matang. Parameter *fuzzifier* digunakan untuk mengatur tingkat fleksibilitas atau kelembutan batas antar kluster sehingga dapat menangkap sifat transisi pada perubahan warna buah kopi secara lebih baik. Setelah proses iterasi selesai, tahap akhir dilakukan dengan proses *defuzzification*, yaitu menentukan kelas akhir berdasarkan nilai keanggotaan tertinggi dari setiap data citra. Dengan pendekatan ini, algoritma Fuzzy C-Means mampu memberikan hasil pengelompokan yang lebih representatif terhadap karakteristik visual buah kopi yang memiliki perubahan warna secara bertahap.

3.1.7. Class Assignment

Class assignment merupakan proses penentuan label kelas akhir untuk setiap data berdasarkan hasil clustering yang telah dilakukan menggunakan metode Fuzzy C-Means (FCM). Setelah proses pelatihan selesai, algoritma FCM menghasilkan matriks keanggotaan (*membership matrix*) yang menunjukkan derajat keanggotaan setiap data terhadap seluruh cluster yang terbentuk. Nilai pada matriks ini berada pada rentang 0 hingga 1, di mana setiap data dapat memiliki tingkat keanggotaan pada lebih dari satu cluster karena sifat FCM yang termasuk ke dalam *soft clustering*.

Namun, untuk keperluan analisis dan interpretasi hasil, setiap data perlu diberikan satu label kelas yang pasti. Oleh karena itu, pada penelitian ini

dilakukan proses *hard assignment* dengan memilih nilai keanggotaan tertinggi dari masing-masing data. Proses ini dilakukan dengan menggunakan fungsi *argmax*, yang bertujuan untuk menentukan cluster dengan nilai keanggotaan terbesar pada setiap sampel. Hasil dari proses ini berupa label cluster diskrit (misalnya 0, 1, atau 2) yang merepresentasikan kelompok data tertentu.

Label cluster yang telah diperoleh kemudian digunakan dalam tahap visualisasi untuk menampilkan distribusi data pada masing-masing cluster. Selain itu, label tersebut juga dibandingkan dengan label sebenarnya yang diperoleh dari struktur folder dataset guna memberikan gambaran awal terhadap kesesuaian hasil clustering. Dengan demikian, tahap *class assignment* berperan penting dalam mengubah hasil *soft clustering* menjadi bentuk yang lebih mudah dipahami dan dianalisis, sehingga dapat mendukung proses evaluasi dan interpretasi hasil penelitian secara keseluruhan.

3.5 Evaluasi Model

Tahap terakhir dalam penelitian ini adalah evaluasi, yaitu proses untuk menilai seberapa baik model mampu melakukan klasifikasi tingkat kematangan buah kopi Arabika Gayo setelah melewati semua tahapan sebelumnya. Evaluasi ini bertujuan untuk mengukur sejauh mana model dapat mengenali pola visual, membedakan tingkat kematangan buah, dan mengelompokkan citra ke dalam kategori yang sesuai. Proses evaluasi dilakukan menggunakan dua ukuran utama, yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Silhouette Score digunakan untuk melihat seberapa baik setiap gambar cocok dengan kelompoknya. Jika nilainya mendekati 1, artinya hasil pengelompokan sudah baik dan setiap klaster memiliki batas yang jelas satu sama lain. Sementara itu, Davies-Bouldin Index digunakan untuk menilai seberapa mirip antar klaster, semakin kecil nilainya, berarti hasil pengelompokan semakin baik karena tiap klaster makin berbeda satu sama lain. Melalui kedua ukuran ini, penulis dapat menilai seberapa baik model *Vision Transformer* dengan metode *self-supervised learning* mampu mengelompokkan citra buah kopi tanpa harus menggunakan label data

secara manual. Perhitungan metrik dilakukan menggunakan rumus berikut:

$$S = \frac{b - a}{\max(a, b)} \quad (1)$$

Keterangan:

a = rata-rata jarak antara satu data dengan data lain dalam klaster yang sama.

b = rata-rata jarak antara data tersebut dengan data pada klaster terdekat.

Interpretasinya:

S mendekati 1 = pengelompokan sangat baik.

S sekitar 0 = batas antar klaster tidak jelas.

S mendekati -1 = data salah kelompok.

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d_{ij}} \right)$$

Keterangan:

σ_i = rata-rata jarak antara titik-titik di klaster ke- i dengan pusatnya.

d_{ij} = jarak antara pusat klaster ke- i dan ke- j .

Interpretasinya:

Nilai DBI rendah → hasil clustering baik (klaster terpisah dengan jelas).

Nilai DBI tinggi → klaster saling tumpang tindih atau belum jelas batasnya.

Dengan menggunakan kedua metrik tersebut, penulis dapat menilai kualitas

hasil pengelompokan yang dihasilkan oleh model secara objektif. Nilai *Silhouette Score* yang tinggi dan *Davies-Bouldin Index* yang rendah menjadi indikator bahwa model mampu membedakan tingkat kematangan buah kopi dengan baik meskipun tanpa bantuan label. Tahap evaluasi ini menjadi langkah akhir untuk memastikan bahwa metode *self-supervised learning* yang diterapkan pada model *Vision Transformer* benar-benar efektif dalam memahami pola visual dan menghasilkan pengelompokan yang bermakna.

3.6 Visualisasi hasil

Visualisasi hasil dilakukan untuk menampilkan dan menganalisis hasil pengelompokan data yang diperoleh menggunakan metode Fuzzy C-Means. Tahap ini bertujuan untuk memberikan gambaran secara visual terhadap kualitas clustering yang dihasilkan, sehingga memudahkan dalam proses interpretasi dan evaluasi model. Pada penelitian ini, visualisasi dilakukan dengan menampilkan sejumlah citra dari masing-masing cluster dalam bentuk grid. Setiap baris pada grid merepresentasikan satu cluster, sedangkan setiap kolom menunjukkan sampel citra yang termasuk dalam cluster tersebut. Dengan demikian, distribusi data pada setiap cluster dapat diamati secara langsung.

Selain itu, setiap citra juga dilengkapi dengan informasi label asli yang diperoleh dari struktur folder dataset. Hal ini bertujuan untuk membandingkan hasil clustering dengan kategori sebenarnya, sehingga dapat diketahui tingkat kesesuaian antara hasil pengelompokan model dengan kondisi data yang sebenarnya. Visualisasi juga dapat dilengkapi dengan nilai derajat keanggotaan (*membership degree*) yang dihasilkan oleh metode Fuzzy C-Means. Nilai ini menunjukkan tingkat kepercayaan model dalam mengelompokkan suatu data ke dalam cluster tertentu. Dengan adanya informasi tersebut, analisis tidak hanya dilakukan berdasarkan hasil pengelompokan akhir, tetapi juga mempertimbangkan tingkat keyakinan model.

Melalui visualisasi ini, peneliti dapat:

1. Mengidentifikasi pola kemiripan citra dalam setiap cluster

2. Menilai kualitas pemisahan antar cluster
3. Mendeteksi kemungkinan kesalahan pengelompokan data
4. Membandingkan hasil clustering dengan label sebenarnya

Dengan demikian, visualisasi hasil menjadi langkah penting untuk melengkapi evaluasi numerik seperti Silhouette Score, Davies-Bouldin Index, dan Fuzzy Partition Coefficient (FPC), sehingga diperoleh pemahaman yang lebih komprehensif terhadap performa model.

3.5 Skenario pengujian dan evaluasi kinerja

Dalam penelitian ini, evaluasi performa model dilakukan menggunakan beberapa metrik evaluasi yaitu recall, precision, F1-score, dan accuracy. Recall digunakan untuk mengukur kemampuan model dalam mengenali data secara benar pada setiap kelas.

1. Recall digunakan untuk mengukur kemampuan model dalam mengenali dan mengklasifikasikan data secara benar pada setiap kelas. Nilai recall menunjukkan seberapa banyak data aktual pada suatu kelas yang berhasil diprediksi dengan benar oleh model. Semakin tinggi nilai recall, maka semakin baik kemampuan model dalam mendeteksi data pada kelas tersebut.

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

2. Precision digunakan untuk mengukur tingkat ketepatan model dalam melakukan prediksi terhadap suatu kelas.

$$precision = \frac{TP}{TP + FP} \quad (4)$$

3. F1-Score digunakan untuk mengukur keseimbangan antara precision dan recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

4. accuracy digunakan untuk mengukur tingkat keseluruhan prediksi yang benar dibandingkan dengan total data yang diuji

$$Accuracy = \frac{TP + TN}{TP + FN + FP + FN} \quad (6)$$

Keterangan:

TP (*True Positive*) = jumlah data positif yang diprediksi benar

TN (*True Negative*) = jumlah data negatif yang diprediksi benar

FP (*False Positive*) = jumlah data negatif yang diprediksi positif

FN (*False Negative*) = jumlah data positif yang diprediksi negatif

5. Confusion matrix merupakan metode evaluasi yang digunakan untuk mengukur performa model klasifikasi dengan membandingkan hasil prediksi model terhadap kelas sebenarnya (*actual class*). Pada confusion matrix, nilai pada diagonal utama menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar, sedangkan nilai di luar diagonal menunjukkan jumlah kesalahan klasifikasi yang dilakukan oleh model.

$$\left[\frac{TP}{FN} + \frac{FP}{TN} \right] \quad (7)$$

Keterangan:

TP (*True Positive*) merupakan jumlah data yang berhasil diprediksi benar sesuai kelas aktual.

TN (*True Negative*) merupakan jumlah data yang berhasil diprediksi benar sebagai bukan kelas tersebut.

FP (*False Positive*) merupakan jumlah data yang diprediksi sebagai suatu kelas tetapi sebenarnya berasal dari kelas lain.

FN (*False Negative*) merupakan jumlah data yang seharusnya termasuk dalam suatu kelas tetapi diprediksi ke kelas lain.

3.6 Alat dan bahan

penelitian ini memerlukan beberapa alat untuk mendukung proses penelitian dalam klasifikasi tingkat kematangan buah kopi Arabica Gayo menggunakan *Vision Transformer*. Alat yang digunakan dikelompokkan ke dalam dua kategori utama, yaitu perangkat keras dan perangkat lunak. Alat dan bahan dapat dilihat pada tabel 3.2.

Tabel 3. 2 Alat dan Bahan

No	Perangkat keras	Perangkat lunak
1.	Laptop RAM 16 GB	Windows 11
2.	-	Microsoft word 2021
3.	-	Google collab
4.	-	Python 3.10
5.	-	Pytorch

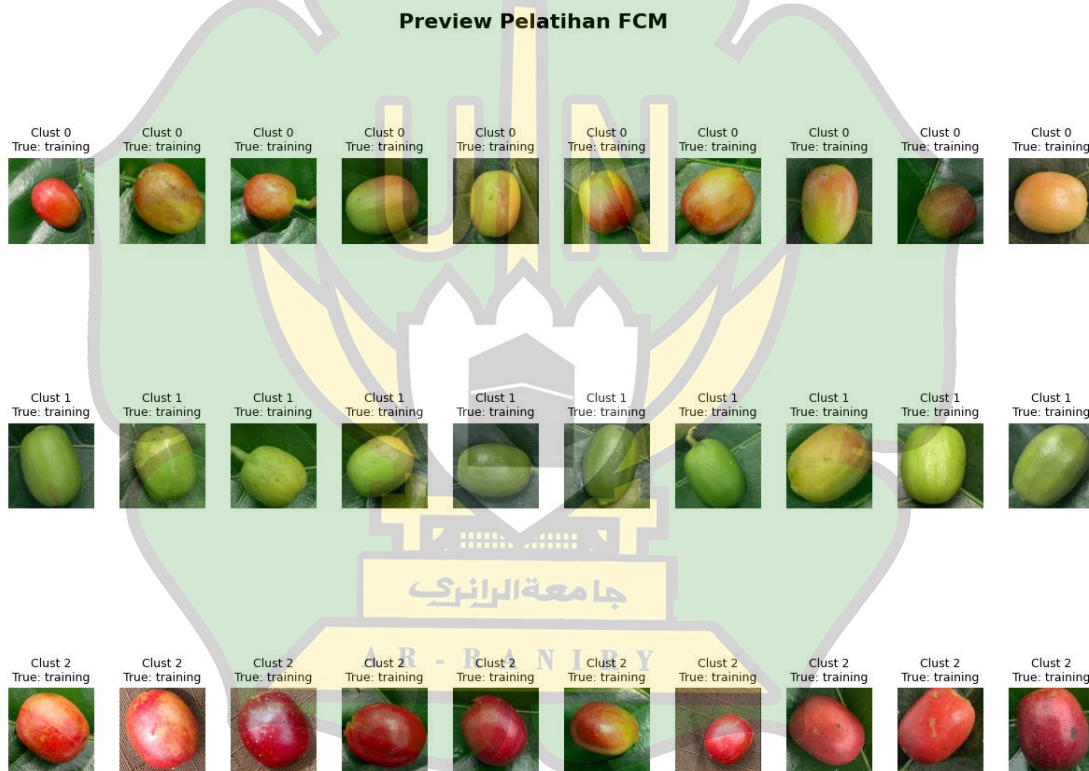


BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Evaluasi Dataset Training dan Validation

4.1.1. Hasil Clustering

Bab ini membahas hasil penelitian yang telah dilakukan serta analisis terhadap metode yang digunakan. Proses penelitian dimulai dari tahap pengolahan data citra kopi, ekstraksi fitur menggunakan model DINOv3, hingga proses pengelompokan citra menggunakan algoritma Fuzzy C-Means. Adapun hasil clustering dapat dilihat pada gambar 4.1



Gambar 4. 1 Hasil Clustering

Gambar 4.1 menampilkan hasil preview pengelompokan citra buah kopi berdasarkan tingkat kematangan menggunakan algoritma Fuzzy C-Means. Pada proses

ini, fitur citra terlebih dahulu diekstraksi menggunakan model DINOv3, kemudian dilakukan proses clustering menjadi tiga kelompok.

Setiap baris pada gambar menunjukkan hasil cluster yang berbeda. Cluster 0 berisi citra buah kopi dengan warna yang cenderung kekuningan hingga kemerahan yang mengindikasikan tingkat kematangan menengah. Cluster 1 didominasi oleh citra buah kopi berwarna hijau yang menunjukkan kondisi buah kopi yang masih mentah. Sedangkan Cluster 2 sebagian besar berisi citra buah kopi dengan warna merah tua yang menunjukkan tingkat kematangan yang lebih tinggi.

Berdasarkan visualisasi tersebut dapat dilihat bahwa algoritma Fuzzy C-Means mampu mengelompokkan citra kopi berdasarkan kemiripan karakteristik warna dan tekstur buah kopi. Meskipun demikian, masih terdapat beberapa citra yang memiliki kemiripan antar cluster karena adanya perbedaan pencahayaan, sudut pengambilan gambar, serta variasi warna pada tingkat kematangan kopi.

4.1.2. Hasil Evaluasi Pretrained

Penelitian ini bertujuan untuk mengelompokkan tingkat kematangan buah kopi menggunakan metode self-supervised learning berbasis *Vision Transformer*. Dataset yang digunakan berjumlah 2.400 gambar buah kopi yang telah melalui tahap preprocessing sebelum dilakukan ekstraksi fitur dan clustering. Berikut merupakan hasil evaluasi pada data validation yang digunakan untuk menilai performa model dalam mengklasifikasikan buah kopi berdasarkan tingkat kematangannya. Adapun hasilnya ditunjukkan pada tabel 4.1.

Tabel 4. 1 Hasil Evaluasi

No	Metode Evaluasi	Nilai	Keterangan
1.	Silhouette Score	0,2305	Menunjukkan kualitas pemisahan cluster yang cukup, dimana sebagian data sudah berada pada cluster yang sesuai namun masih terdapat kemiripan antar cluster

No	Metode Evaluasi	Nilai	Keterangan
2.	Davies-Bouldin Index	1,7560	Nilai yang relatif kecil menunjukkan pemisahan antar cluster cukup baik
3.	Fuzzy Partition Coefficient (FPC)	0,3996	Menunjukkan tingkat kejelasan keanggotaan data pada setiap cluster

Berdasarkan tabel 4.1 nilai Silhouette Score sebesar 0,2305 menunjukkan bahwa struktur cluster yang dihasilkan berada pada kategori cukup. Nilai Silhouette Score berada pada rentang -1 hingga 1, dimana nilai yang mendekati 1 menunjukkan pemisahan cluster yang sangat baik, sedangkan nilai yang mendekati 0 menunjukkan adanya kemiripan antar cluster. Dengan nilai 0,2305 dapat disimpulkan bahwa sebagian data sudah berada pada cluster yang sesuai, namun masih terdapat beberapa data yang memiliki kemiripan dengan cluster lain.

Sementara itu, Davies-Bouldin Index sebesar 1,7560 menunjukkan tingkat pemisahan antar cluster yang cukup baik. Pada metrik ini, nilai yang lebih kecil menunjukkan kualitas cluster yang lebih baik karena jarak antar cluster semakin besar dan variasi data dalam satu cluster semakin kecil. Dengan demikian, nilai 1,7560 menunjukkan bahwa cluster yang dihasilkan sudah memiliki tingkat pemisahan yang cukup namun masih dapat ditingkatkan.

Selain itu, nilai Fuzzy Partition Coefficient (FPC) sebesar 0,3996 menunjukkan tingkat kejelasan pembagian keanggotaan data pada setiap cluster. Nilai FPC berada pada rentang 0 hingga 1, dimana nilai yang mendekati 1 menunjukkan pembagian cluster yang semakin jelas. Nilai 0,3996 menunjukkan bahwa beberapa data masih memiliki derajat keanggotaan yang relatif mirip terhadap lebih dari satu cluster, sehingga batas antar cluster belum sepenuhnya tegas. Berdasarkan ketiga metrik evaluasi tersebut, dapat disimpulkan bahwa metode ekstraksi fitur menggunakan DINOv3 yang dikombinasikan dengan algoritma Fuzzy C-Means mampu

menghasilkan pengelompokan citra kopi berdasarkan tingkat kematangan dengan kualitas cluster yang cukup baik, meskipun masih terdapat beberapa data yang memiliki kemiripan antar cluster.

4.2 Evaluasi Peforma Model

Evaluasi akhir pada penelitian ini dilakukan secara objektif dengan menggunakan data uji (*testing set*), yaitu data baru yang tidak pernah digunakan maupun dilihat oleh model selama proses pelatihan berlangsung. Penggunaan data uji yang terpisah dari data pelatihan bertujuan untuk mengukur kemampuan generalisasi model dalam mengenali pola pada data yang benar-benar baru. Dengan demikian, hasil evaluasi yang diperoleh dapat memberikan gambaran yang lebih akurat mengenai performa model dalam kondisi nyata, serta memastikan bahwa model tidak hanya menghafal data pelatihan, tetapi juga mampu melakukan prediksi secara efektif pada data yang belum pernah ditemui sebelumnya.

4.2.1. Evaluasi Peforma Model

Hasil pengukuran performa model terhadap tiga kelas yang digunakan dalam penelitian ini disajikan secara rinci pada Tabel 4.2. Tabel tersebut menampilkan nilai evaluasi yang digunakan untuk menilai kemampuan model dalam mengklasifikasikan setiap kelas secara akurat.

Tabel 4. 2 Hasil Evaluasi Peforma Model

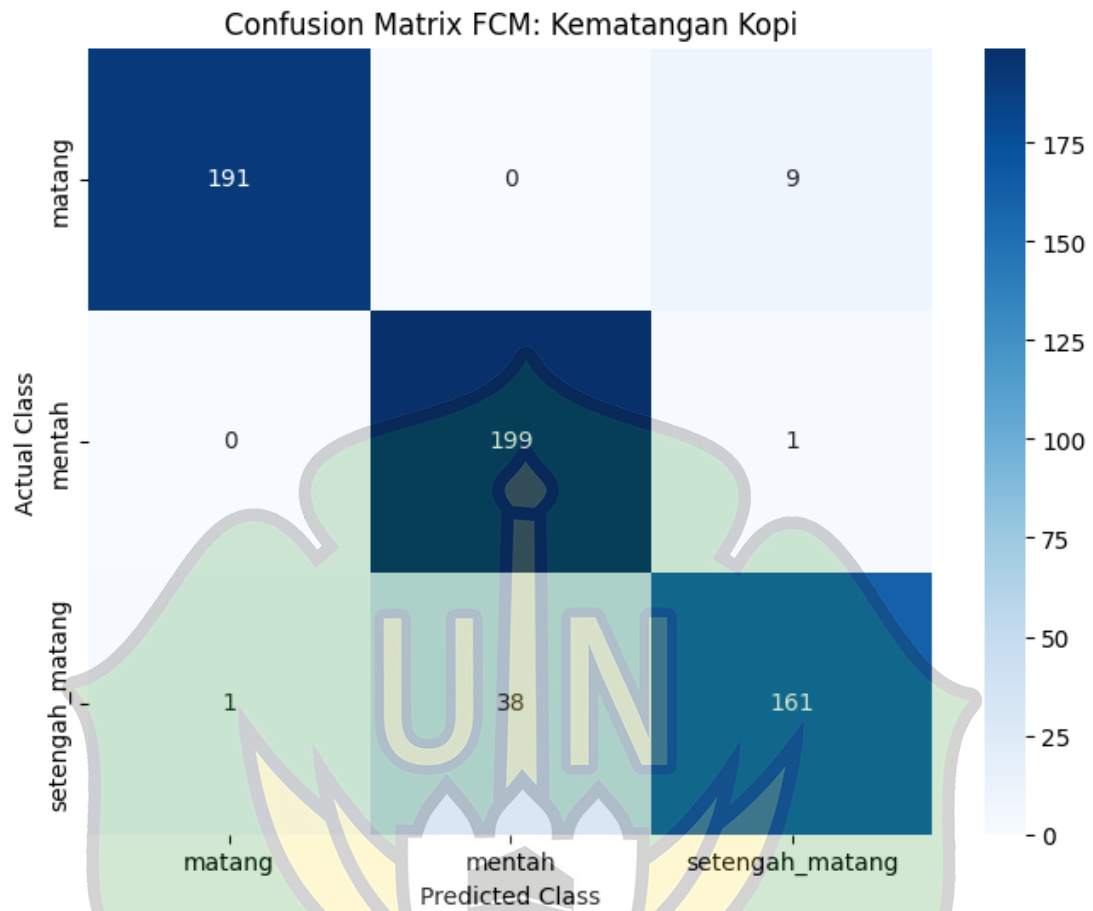
No	Kelas	Recall	precesion	F1-Score	Support
1.	Setengah matang	0.81	0.94	0.87	200
2.	Mentah	0.99	0.84	0.91	200
3.	matang	0.95	0.99	0.97	200
Macro avarage		0.92	0.93	0.92	600
Weighted avarage		0.92	0.93	0.92	600
Accuracy				0.92%	600

Tabel 4.2 menunjukkan bahwa model memiliki performa yang cukup baik dengan akurasi sebesar 0,92 dalam mengklasifikasikan tiga kelas kematangan buah kopi. Kelas matang dan mentah memiliki nilai evaluasi yang tinggi karena keduanya memiliki ciri visual yang jelas dan konsisten, yaitu warna merah pada buah matang dan warna hijau pada buah mentah. Hal ini terlihat dari nilai recall dan precision yang tinggi, yaitu recall 0,99 dan precision 0,84 untuk kelas mentah, serta recall 0,95 dan precision 0,99 untuk kelas matang.

Sebaliknya, kelas setengah matang memiliki nilai recall sebesar 0,81, yang lebih rendah dibandingkan kelas lainnya, meskipun memiliki precision sebesar 0,94. Hal ini disebabkan oleh karakteristik visualnya yang bersifat transisi, yaitu perpaduan warna antara hijau, kuning, hingga kemerahan, sehingga batas antar kelas menjadi tidak tegas. Kondisi ini menyebabkan model cenderung mengalami kebingungan dalam membedakan apakah suatu citra lebih mendekati kelas mentah atau matang, sehingga meningkatkan kemungkinan kesalahan klasifikasi pada kelas tersebut. Selain itu, nilai F1-Score untuk masing-masing kelas yaitu 0,87 (setengah matang), 0,91 (mentah), dan 0,97 (matang) menunjukkan bahwa performa terbaik terdapat pada kelas matang. Secara keseluruhan, nilai macro average dan weighted average masing-masing sebesar 0,92 menunjukkan bahwa model memiliki performa yang stabil dan konsisten dalam mengklasifikasikan seluruh kelas.

4.2.2. Analisa Confusion Matrix

Confusion matrix menggambarkan hasil klasifikasi model terhadap tiga kelas kematangan buah kopi, sehingga dapat diketahui jumlah prediksi yang benar maupun kesalahan klasifikasi pada setiap kelas, seperti yang ditunjukkan pada Gambar 4.3.



Gambar 4. 2 Confusion Matrix Data Test

Pada gambar 4.2 terdapat Confusion matrix, dibaca dengan memperhatikan posisi sumbu, dimana sumbu Y (vertikal) merepresentasikan kelas aktual (*Actual Class*) dan sumbu X (horizontal) merepresentasikan kelas hasil prediksi (*Predicted Class*) dari model. Setiap elemen pada matriks menunjukkan jumlah data dari suatu kelas aktual yang diprediksi ke dalam kelas tertentu oleh model. Nilai yang berada pada diagonal utama menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi. Dengan demikian, confusion matrix tidak hanya memberikan informasi mengenai tingkat akurasi model, tetapi juga menggambarkan pola kesalahan yang terjadi pada masing-masing kelas.

Berdasarkan hasil yang diperoleh, pada kelas matang, sebagian besar data berhasil diklasifikasikan dengan benar, yaitu sebanyak 191 data, sementara 9 data lainnya salah diklasifikasikan sebagai setengah matang. Pada kelas mentah, performa model sangat baik dengan 199 data berhasil diklasifikasikan dengan benar dan hanya 1 data yang salah diklasifikasikan sebagai setengah matang. Sementara itu, pada kelas setengah matang, model menunjukkan performa yang lebih rendah, dimana hanya 161 data yang diklasifikasikan dengan benar, sedangkan 38 data salah diklasifikasikan sebagai mentah dan 1 data sebagai matang.

Hasil ini menunjukkan bahwa model mampu mengenali dengan baik kelas yang memiliki ciri visual yang jelas seperti matang dan mentah, namun masih mengalami kesulitan dalam membedakan kelas setengah matang yang memiliki karakteristik visual transisi antara kedua kelas tersebut. Kondisi ini menyebabkan terjadinya kesalahan klasifikasi yang lebih tinggi pada kelas setengah matang dibandingkan dengan kelas lainnya.

4.2.3. Analisa Visualisasi kopi Setengah Matang

Untuk mengetahui kinerja model secara individu, dilakukan pengujian tunggal pada citra buah kopi, dan hasil prediksi yang diperoleh ditampilkan pada gambar 4.3.



Gambar 4. 3 Buah kopi Setengah Matang

Berdasarkan gambar 3.4 pengujian tunggal pada citra buah kopi, model mengklasifikasikan citra tersebut ke dalam kelas setengah matang dengan tingkat kepercayaan (*confidence*) sebesar 40,36%. Hasil ini menunjukkan bahwa model mengidentifikasi adanya karakteristik visual yang berada di antara kelas mentah dan matang. Secara visual, buah kopi pada citra tersebut memperlihatkan perpaduan warna, yaitu dominasi warna kekuningan hingga kemerahan dengan sedikit nuansa hijau. Ciri ini merupakan karakteristik umum dari fase setengah matang, dimana buah kopi sedang mengalami proses transisi dari kondisi mentah (hijau) menuju matang (merah).

Namun, nilai *confidence* yang relatif rendah (sekitar 40%) menunjukkan bahwa model masih memiliki tingkat ketidakpastian yang cukup tinggi dalam melakukan klasifikasi. Hal ini kemungkinan disebabkan oleh kemiripan fitur visual antara kelas setengah matang dengan kelas lainnya, terutama mentah, sehingga model tidak dapat memberikan keputusan yang sangat yakin. Dengan demikian, hasil pengujian ini menunjukkan bahwa meskipun model mampu mengenali pola umum dari buah kopi setengah matang, masih diperlukan peningkatan performa model agar dapat memberikan prediksi dengan tingkat kepercayaan yang lebih tinggi, khususnya pada kelas yang memiliki karakteristik visual yang ambigu.

4.2.4. Analisa Visualisasi Kopi Mentah

Untuk mengetahui kinerja model secara individu, dilakukan pengujian tunggal pada gambar buah kopi mentah, dan hasil prediksi yang diperoleh ditampilkan pada gambar 4.4.

Hasil Prediksi: MENTAH
Confidence: 63.26%
(Cluster ID: 1)



Gambar 4. 4 Buah kopi Mentah

Berdasarkan gambar 4.4 hasil pengujian tunggal pada citra buah kopi, model mengklasifikasikan citra tersebut ke dalam kelas mentah dengan tingkat kepercayaan (*confidence*) sebesar 63,26% (Cluster ID: 1). Hasil ini menunjukkan bahwa model mampu mengenali karakteristik visual utama dari buah kopi mentah. Secara visual, buah kopi mentah umumnya memiliki warna dominan hijau yang relatif seragam dan belum menunjukkan adanya perubahan warna menuju kekuningan atau kemerahan. Ciri ini menjadi pembeda utama dibandingkan dengan kelas setengah matang maupun matang, sehingga lebih mudah dikenali oleh model. Nilai *confidence* yang cukup tinggi dibandingkan dengan kelas setengah matang menunjukkan bahwa model memiliki tingkat keyakinan yang lebih baik dalam mengidentifikasi kelas mentah. Hal ini disebabkan oleh karakteristik visualnya yang lebih jelas dan konsisten. Dengan demikian, hasil pengujian ini menunjukkan bahwa model telah mampu mengklasifikasikan buah kopi mentah dengan cukup baik, meskipun masih terdapat

peluang untuk meningkatkan tingkat kepercayaan agar hasil prediksi menjadi lebih optimal.

4.2.5. Analisa Visualisasi Kopi Matang

Untuk mengetahui kinerja model secara individu, dilakukan pengujian tunggal pada gambar buah kopi matang, dan hasil prediksi yang diperoleh ditampilkan pada gambar 4.5.



Gambar 4.5 Buah kopi Matang

Berdasarkan gambar 4.5 hasil pengujian tunggal pada citra buah kopi, model mengklasifikasikan citra tersebut ke dalam kelas matang dengan tingkat kepercayaan (*confidence*) sebesar 45,06%. Hasil ini menunjukkan bahwa model mampu mengenali karakteristik visual yang mengarah pada tingkat kematangan penuh, meskipun tingkat keyakinannya masih tergolong sedang. Secara visual, buah kopi matang umumnya memiliki warna merah yang dominan dan merata, yang menjadi ciri utama pembeda dibandingkan dengan kelas mentah maupun setengah matang. Namun, pada beberapa kondisi, pencahayaan atau variasi warna pada citra dapat mempengaruhi hasil ekstraksi

fitur, sehingga model tidak memberikan tingkat kepercayaan yang tinggi. Nilai confidence yang tidak terlalu tinggi ini menunjukkan bahwa masih terdapat kemungkinan kemiripan fitur dengan kelas lain, terutama setengah matang yang memiliki warna transisi menuju merah. Hal ini menyebabkan model belum sepenuhnya yakin dalam menentukan klasifikasi. Dengan demikian, hasil pengujian ini menunjukkan bahwa model sudah mampu mengenali ciri utama buah kopi matang, namun masih diperlukan peningkatan agar tingkat kepercayaan prediksi menjadi lebih tinggi dan lebih konsisten.



BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, maka diperoleh kesimpulan yang menjawab rumusan masalah sebagai berikut:

1. Metode ekstraksi fitur menggunakan DINOv3 yang dikombinasikan dengan algoritma Fuzzy C-Means (FCM) kurang efektif untuk mengelompokkan citra buah kopi berdasarkan tingkat kematangan secara optimal, khususnya pada kelas setengah matang. Berdasarkan pengujian data model menunjukkan performa yang baik pada kelas matang dan mentah karena kedua kelas tersebut memiliki karakteristik warna yang lebih jelas dan mudah dibedakan. Namun, pada kelas setengah matang, model mengalami kesulitan dalam proses pengelompokan karena warna buah kopi setengah matang merupakan gradasi atau peralihan antara warna buah mentah dan matang. Kemiripan karakteristik visual tersebut menyebabkan batas antar cluster menjadi kurang jelas sehingga beberapa data pada kelas setengah matang masih mengalami kesalahan pengelompokan.
2. Model yang dibangun masih kurang cocok digunakan dalam proses klasifikasi tingkat kematangan buah kopi secara optimal. Meskipun model menunjukkan performa yang baik pada kelas mentah dan matang yang memiliki karakteristik visual lebih jelas dan konsisten, model masih mengalami kesulitan dalam mengklasifikasikan kelas setengah matang. Hal ini disebabkan karena karakteristik visual buah kopi setengah matang bersifat transisi atau gradasi antara mentah dan matang, sehingga memiliki kemiripan warna dengan kedua kelas tersebut. Kondisi ini menyebabkan tingkat kesalahan klasifikasi pada kelas setengah matang menjadi lebih tinggi dibandingkan kelas lainnya serta menghasilkan nilai prediksi yang kurang konsisten. Oleh karena itu, model masih memerlukan pengembangan lebih lanjut agar dapat menghasilkan klasifikasi yang lebih akurat dan stabil pada seluruh tingkat kematangan buah kopi.

5.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, penulis mengajukan beberapa saran untuk pengembangan penelitian selanjutnya sebagai berikut:

1. Untuk meningkatkan performa model, disarankan untuk menambah jumlah dataset serta memperkaya variasi data yang digunakan. Penambahan dataset ini dapat mencakup variasi kondisi pencahayaan, sudut pengambilan gambar, latar belakang, serta tingkat kematangan buah kopi yang lebih beragam. Selain itu, proses pra-pemrosesan data (*preprocessing*) seperti normalisasi citra, penghapusan noise, dan segmentasi objek dapat dioptimalkan agar fitur yang diekstraksi lebih fokus pada objek utama. Dengan dataset yang lebih besar dan kualitas data yang lebih baik, diharapkan model dapat mempelajari pola fitur yang lebih representatif sehingga mampu meningkatkan akurasi serta tingkat kepercayaan (*confidence*) dalam proses klasifikasi.
2. Penggunaan metode ekstraksi fitur yang lebih optimal atau penggabungan beberapa metode (*hybrid feature extraction*) juga dapat dipertimbangkan agar model mampu menangkap perbedaan fitur yang lebih detail. Selain DINOv3, penelitian selanjutnya dapat mengombinasikan dengan metode lain seperti Convolutional Neural Network (CNN) atau model berbasis transformer lainnya untuk memperoleh representasi fitur yang lebih kaya. Selain itu, proses *fine-tuning* pada model ekstraksi fitur juga dapat dilakukan agar lebih sesuai dengan karakteristik dataset yang digunakan. Pengembangan pada tahap clustering juga dapat dilakukan, misalnya dengan melakukan optimasi parameter pada algoritma Fuzzy C-Means atau membandingkannya dengan algoritma clustering lain. Dengan pendekatan tersebut, diharapkan hasil klasifikasi menjadi lebih akurat, stabil, serta mampu mengatasi permasalahan pada kelas yang memiliki karakteristik visual yang ambigu seperti setengah matang.
3. Selain itu, peneliti selanjutnya disarankan untuk mengimplementasikan model yang telah dibangun ke dalam bentuk sistem atau alat nyata yang dapat

digunakan secara langsung di lapangan. Sistem ini dapat dikembangkan menjadi alat pemisah otomatis buah kopi berdasarkan tingkat kematangan (mentah, setengah matang, dan matang) yang terintegrasi dengan kamera atau webcam sebagai input citra secara real-time. Dengan adanya alat ini, proses sortasi kopi dapat dilakukan secara lebih cepat, konsisten, dan efisien tanpa bergantung pada penilaian manual, sehingga dapat meningkatkan kualitas hasil panen kopi secara keseluruhan.



DAFTAR PUSTAKA

- Amelia, N., Garonga, M., & Rusman, J. (2023). Penerapan Metode K-Nearest Neighbor (KNN) untuk Klasifikasi Kematangan Buah Kopi. *The Indonesian Journal of Computer Science*, 12(2).
- Baasith, R. H., Sasongko, T. B., Hadinegoro, A., & Saputro, U. A. (2026). Implementation of SSL-Vision Transformer (ViT) for Multi-Lung Disease Classification on X-Ray Images. *Journal of Applied Informatics and Computing*, 10(1), 298-308.
- Elwirehardja, G. N., Liem, S. M., Adjie, M. L., Tjan, F. A., Setiawan, J., Syahputra, M. E., & Muljo, H. H. (2025). Comparative analysis of self-supervised pre-trained vision transformers and convolutional neural networks with CheXNet in classifying lung conditions. *Commun. Math. Biol. Neurosci.*, 2025, Article-ID.
- Gusnanto, F., Rahaningsih, N., Dana, R. D., & Mulyawan, M. (2025). Optimasi Model Klasifikasi Tingkat Kematangan Buah Mangga Dengan Metode Yolo11. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 1773-1780.
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., ... & Tao, D. (2022). A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(1), 87-110.
- Haris, M. (2021). Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid (Implementation of Deep Learning Methods in Predicting Student Performance: A Systematic Literature Review). *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 10(2).
- Izza, M. N., & Kusuma, G. P. (2024). Image classification of green Arabica coffee using transformer-based architecture. *International Journal of Engineering Trends and Technology*, 72(6), 304-314.
- Murugavalli, S., & Gopi, R. (2025). Plant leaf disease detection using vision transformers for precision agriculture. *Scientific Reports*, 15, 22361.
- Pangestu, A., Purnama, B., & Risnandar, R. (2024). Vision Transformer untuk Klasifikasi Kematangan Pisang. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 75-84.

- Purba, S. O. (2024). *Klasifikasi Penyakit Mata pada Manusia dengan Menggunakan Model Arsitektur Vision Transformers* (Doctoral dissertation, Universitas Medan Area).
- Rifqi, M. (2024). *Klasifikasi Penyakit pada Daun Kopi Menggunakan Metode Vision Transformer (Vit)*.
- Saputri, M., Lioe, H. N., & Wijaya, C. H. (2020). Pemetaan karakteristik kimia biji kopi Arabika Gayo dan Robusta Gayo [Mapping the chemical characteristics of Gayo Arabica and Robusta green coffee beans]. *Jurnal Teknologi dan Industri Pangan*, 31(1), 76–85.
- Sary, N., Pasaribu, H. H. S., Situmeang, R. J., & Darus, R. D. (2025). Implementasi Algoritma YOLOv11 Dan Roboflow Untuk Deteksi Tingkat Kematangan Anggur Berbasis Web. *METIK JURNAL (AKREDITASI SINTA 3)*, 9(2), 264-274.
- Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., ... & Bojanowski, P. (2025). Dinov3. *arXiv preprint arXiv:2508.10104*.
- Taufiqurrahman, R. W., Yudistira, N., & Widodo, A. W. (2025). Penerapan Model Vision Transformer dalam Klasifikasi Hama pada Pertanian. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 9(10).
- Uelwer, T., Robine, J., Wagner, S. S., Höftmann, M., Upschulte, E., Konietzny, S., ... & Harmeling, S. (2025). A survey on self-supervised methods for visual representation learning. *Machine Learning*, 114(4), 1-56.
- YUDISTIANSYAH, I. (2025). *PENERAPAN COMPUTER VISION UNTUK KLASIFIKASI PENYAKIT MATA MENGGUNAKAN ARSITEKTUR VISION TRANSFORMERS (VITS) PADA CITRA FUNDUS* (Doctoral dissertation, Nusa Putra University).