BAB II
TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

2.1.1. Klasifikasi Tingkat Roasting Biji Kopi

Kopi merupakan salah satu komoditas pertanian yang memiliki nilai ekonomi tinggi dan dikonsumsi secara luas di berbagai negara. Salah satu faktor utama yang memengaruhi karakteristik rasa, aroma, dan warna biji kopi adalah tingkat *roasting*. Proses roasting menyebabkan perubahan fisik dan kimia pada biji kopi, seperti perubahan warna, tekstur, serta kandungan senyawa volatil. Secara umum, tingkat roasting biji kopi dibedakan menjadi beberapa kategori, antara lain light roast, medium roast, dan dark roast. Perbedaan tingkat roasting ini dapat diamati secara visual melalui warna permukaan biji kopi, sehingga memungkinkan pemanfaatan citra digital sebagai dasar klasifikasi tingkat roasting [6].

Penelitian yang dilakukan pada tahun 2025 oleh Jamil *et al* [7], berfokus pada MobileNetV2. Penelitian ini bertujuan mengembangkan sebuah *framework* klasifikasi citra kopi otomatis bernama X-CoffeeNet yang membedakan jenis kopi (Arabica, robusta, liberica), tetapi juga menyediakan interpretasi hasil keputusan model agar pengguna memahami fitur visual yang memengaruhi prediksi. Dataset penelitian terdiri dari sekitar 3000 gambar biji kopi, hasil evaluasi menunjukkan bahwa X-CoffeeNet mencapai 100% pada semua metrik utama, meliputi akurasi, presisi, *recall,* dan F1-score. Model ini menunjukkan akurasi sempurna dan memiliki potensi aplikasi otomasi industry agrikultur dan makanan yang memerlukan interpretabilitas keputusan model.

Penelitian berikutnya yang dilakukan pada tahun 2022 oleh Pramono *et al* [8], berfokus pada *transfer learning* dengan arsitektur EfficientNet-Lite, sebuah varian ringan dari EfficientNet yang dirancang untuk perangkat seluler dan aplikasi dengan sumber daya terbatas. Arsitektur ini dilatih menggunakan dataset citra biji atau cangkir kopi (dependen konteks dataset) agar dapat mengenali dan membedakan berbagai tipe kopi secara akurat. Detail jumlah kelas atau jumlah gambar tidak dijelaskan secara lengkap dalam sumber yang tersedia, tetapi

konteksnya fokus pada pengenalan tipe kopi (mis. tipe/kematangan) melalui foto yang diambil pengguna atau dari dataset yang disusun untuk penelitian deteksi kopi. Kinerja model akurasi pelatihan (training accuracy) mencapai sekitar 87%, akurasi validasi (validation accuracy) mencapai sekitar 81%. Model EfficientNet-Lite 0 yang dilatih pada dataset kopi berhasil mencapai akurasi 87 % saat pelatihan dan 81% pada tahap validasi, menunjukkan efektivitasnya dalam tugas klasifikasi jenis kopi meskipun belum sempurna.

Penelitian Selanjutnya dilakukan pada tahun 2025 oleh Syaputra *et al* [9], menggunakan dataset USK-Coffee, yang berisi empat kelas objek biji kopi hijau: *defect, longberry, peaberry*, dan premium, dengan foto diambil dalam berbagai kondisi pencahayaan dan sudut sehingga menantang model klasifikasi., peneliti mengevaluasi performa dua arsitektur. Model EfficientNet-B0 mencapai akurasi 87% dengan F1-score sekitar 0.87, sementara ViT-B16 mencapai akurasi 85% dan F1-score 0.85 pada data uji. Selain itu, EfficientNet-B0 juga menunjukkan waktu inferensi per gambar yang jauh lebih cepat (~0,035 detik) dibandingkan ViT-B16 (~0,426 detik), meskipun ViT-B16 memiliki latensi batch yang lebih cepat. Hasil ini menunjukkan bahwa kedua model layak dipakai untuk sistem klasifikasi berbasis edge computing, dengan EfficientNet-B0 unggul dalam akurasi dan inferensi real-time untuk satu gambar.

2.1.2. Penerapan Model Deep Learning untuk Klasifikasi Citra

Penerapan model deep learning dalam klasifikasi citra digunakan untuk mengenali dan mengelompokkan objek berdasarkan pola visual pada gambar secara otomatis. Salah satu metode yang paling banyak digunakan adalah Convolutional Neural Network (CNN) karena mampu mengekstraksi fitur citra seperti warna, tekstur, bentuk, dan pola dengan tingkat akurasi yang tinggi. Dalam penelitian ini, model CNN seperti DenseNet121, MobileNetV2, dan EfficientNetB0 diterapkan untuk melakukan klasifikasi tingkat roasting biji kopi berdasarkan citra digital. Penggunaan model deep learning memungkinkan proses klasifikasi dilakukan secara otomatis, lebih cepat, dan lebih konsisten dibandingkan pengamatan manual.

Penelitian yang dilakukan pada tahun 2022 oleh Sanghvi *et al* [10], model yang digunakan ialah DenseNet-201. Penelitian ini bertujuan untuk

mengembangkan dan mengevaluasi model *deep learning* guna klasifikasi gambar X-ray dada menjadi kategori COVID-19, *pneumonia*, atau normal. Dataset utama adalah gambar radiografi dada (chest X-ray) berlabel yang menangkap kondisi pasien COVID-19, *pneumonia*, dan sehat. Hasil evaluasi model diukur menggunakan metrik *accuracy, precision, recall,* dan *F1-score.* Penelitian menunjukkan bahwa DenseNet-201 adalah arsitektur yang layak pakai untuk klasifikasi otomatis citra X-ray dada, mampu membedakan antara COVID-19 dan *pneumonia* dengan hasil yang baik, mendukung sistem pendukung keputusan medis.

Penelitian berikutnya yang dilakukan pada tahun 2020 oleh Ziliang Zhong *et al* [11], model yang digunakan adalah DenseNet untuk mengidentifikasi kanker metastatik dalam gambar histopatologi berukuran kecil. Model tersebut dirancang agar mampu memberikan prediksi yang lebih akurat dibandingkan model CNN klasik seperti ResNet34 dan VGG19. Metrik evaluasi dengan AUC-ROC dan akurasi untuk mengukur performa klasifikasi. DensetNet dapat meningkatkan akurasi klasifikasi gambar kanker dibanding model konvensional, terutama pemanfaatan teknik data *augmentation*.

Penelitian berikutnya yang dilakukan pada tahun 2022 oleh Aufar Y *et al* [12], model utama yang diteliti adalah MobileNetV2, penelitian ini memiliki tujuan yaitu mendeteksi penyakit daun kopi Robusta. Dataset terdiri dari 1560 gambar daun kopi robusta yang diambil menggunakan kamera *smartphone*. Model dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Model MobileNetV2 untuk dua kelas (sehat vs tidak sehat) mencapai akurasi sekitar 78,57%, pada klasifikasi yang lebih rumit (3 kelas dan 6 kelas), performa turun secara drastis. Metode ini mencapai performa cukup baik untuk klasifikasi dua kelas.

Penelitian selanjutnya yang dilakukan pada tahun 2025 oleh Jodi Wijaya [13], berfokus pada perbandingan dua model CNN, yaitu InceptionResNetV2 dengan MobileNetV2. Tujuan penelitian yaitu membantu proses panen sehingga buah yang matang ideal yang dipanen. Akurasi hasil penelitian terbaik dicapai oleh MobileNetV2 sekitar 98%, sedangkan InceptionResNetV2 mencapai 90%. Kesimpulan utama MobileNetV2 unggul dalam kasus ini, untuk klasifikasi citra

kematangan buah kopi.

Penelitian berikutnya yang dilakukan pada tahun 2023 oleh Muhammad Rosadi *et al* [14], berfokus pada arsitektur EfficientNet untuk membantu petani dalam mengidentifikasi empat tipe penyakit yang menyerang daun kopi. Dataset yang digunakan berjumlah 1464 gambar RGB berukuran 2048 x 1024 piksel, yang dibagi menjadi 1264 data pelatihan dan 400 data pengujian. Beragam varian arsitektur EfficientNet (B0 sampai B4) diuji dengan parameter *resampling* Lanczos, *epoch* 25, *learning rate* 0,0001, fungsi *loss sparse categorical cross entropy,* dan *optimizer* Adam. Hasil eksperimen menunjukkan bahwa EfficientNet-B1 memberikan akurasi terbaik sebesar 97 % dengan *loss* 0,1328 pada data pelatihan maupun pengujian, mengungguli arsitektur lain seperti VGG16, ResNet50, dan MobileNetV2. Penelitian ini menyimpulkan bahwa model EfficientNet-B1 sangat efektif untuk klasifikasi penyakit daun kopi dan merekomendasikan penelitian lanjutan dengan pengaturan *hyperparameter* yang lebih variatif serta pengembangan aplikasi yang dapat digunakan petani secara langsung pada lahan.

2.1.3. Perbandingan Penelitian

Berdasarkan penelitian terdahulu, tantangan umum penggunaan CNN dalam klasifikasi biji kopi adalah perbedaan visual antar kelas sering kali bersifat halus (subtle differences), terutama pada tingkat *medium roast* yang berada di antara *light* dan *dark roast*. Perbedaan warna, tekstur, dan kilap permukaan biji kopi dapat dipengaruhi oleh kondisi pencahayaan, sudut pengambilan gambar, serta variasi alami antar biji kopi. Kondisi ini dapat menyebabkan tumpang tindih fitur visual antar kelas, sehingga menyulitkan model CNN dalam melakukan klasifikasi secara akurat. Secara keseluruhan, tantangan utama penggunaan DenseNet, MobileNet, dan EfficientNet dalam klasifikasi tingkat *roasting* biji kopi meliputi keterbatasan perbedaan visual antar kelas, ketergantungan kualitas dataset, risiko *overfitting*, serta penyesuaian sumber daya dan parameter model. Oleh karena itu, evaluasi komparatif diperlukan untuk menentukan model yang paling sesuai berdasarkan metrik evaluasi.

Berdasarkan tinjauan pustaka yang telah dilakukan, dapat disimpulkan bahwa penerapan deep learning berbasis Convolutional Neural Network (CNN)

telah terbukti efektif dalam berbagai tugas klasifikasi citra, khususnya pada bidang pertanian dan kesehatan tanaman. Beragam arsitektur modern seperti DenseNet, MobileNet, EfficientNet, InceptionResNet, dan Vision Transformer menunjukkan performa yang baik dalam mengidentifikasi objek, jenis, tingkat kematangan, maupun penyakit tanaman, termasuk pada komoditas kopi. Penggunaan transfer learning, data augmentation, serta pemilihan arsitektur yang efisien menjadi faktor penting dalam meningkatkan akurasi model, terutama pada dataset dengan jumlah terbatas. Meskipun demikian, tantangan seperti keterbatasan dataset, variasi kondisi pencahayaan, serta kompleksitas klasifikasi multikelas masih menjadi peluang untuk pengembangan penelitian lebih lanjut. Oleh karena itu, penelitian ini diharapkan dapat berkontribusi dalam menyempurnakan metode klasifikasi berbasis *deep learning* dengan memanfaatkan arsitektur dan teknik yang sesuai guna meningkatkan akurasi dan aplikabilitas sistem di lingkungan nyata. Tabel perbandingan penelitian terdahulu dapat dilihat pada Tabel 2.1.

Tabel 2. 1 Tabel Perbandingan Penelitian

Peneliti	Tahun	Dataset	Model Classifier	Hasil Metrik Evaluasi
SH Wang Y	2020	Dataset citra MRI multiple sclerosis dan normal	DenseNet-121, DenseNet-169, DenseNet-201	DenseNet-201 memperoleh akurasi 98,31%, sensitivitas 98,27%, spesifisitas 98,35%
Ziliang Zhong et al	2020	Dataset gambar histopatologi kanker metastatik	DenseNet	Menggunakan AUC-ROC dan akurasi, hasil menunjukkan performa lebih baik dibanding ResNet34 dan VGG19
Sanghvi et al	2022	Dataset chest X-ray terdiri dari kelas	DenseNet-201	Evaluasi menggunakan accuracy, precision,

		COVID-19, pneumonia, dan normal		recall, dan F1-score dengan hasil klasifikasi sangat baik pada deteksi COVID-19
Aufar Y et al	2022	1560 gambar daun kopi Robusta dengan kelas sehat dan tidak sehat, serta eksperimen multikelas	MobileNetV2	Akurasi dua kelas mencapai 78,57%, performa menurun pada klasifikasi 3 dan 6 kelas
Pramono et al	2022	Dataset citra kopi untuk pengenalan tipe kopi	EfficientNet-Lite0	Akurasi pelatihan 87% dan validasi 81%
Muhammad Rosadi et al	2023	1464 gambar RGB penyakit daun kopi dengan 4 kelas penyakit	EfficientNet-B0 sampai EfficientNet-B4	EfficientNet-B1 memperoleh akurasi terbaik 97% dengan loss 0,1328
Jamil et al	2025	Sekitar 3000 gambar biji kopi dengan kelas Arabica, Robusta, dan Liberica	MobileNetV2 (X-CoffeeNet)	Akurasi, precision, recall, dan F1-score mencapai 100%

Syaputra et al	2025	Dataset USK-Coffee dengan 4 kelas: defect, longberry, peaberry, premium	EfficientNet-B0 dan ViT-B16	EfficientNet-B0 memperoleh akurasi 87% dan F1-score 0,87; ViT-B16 memperoleh akurasi 85% dan F1-score 0,85
Jodi Wijaya	2025	Dataset citra kematangan buah kopi	MobileNetV2 dan InceptionResNetV2	MobileNetV2 mencapai akurasi 98%, sedangkan InceptionResNetV2 90%
Penulis	2026	1200 dataset citra roasting biji kopi	,MobileNetV2 EfficientNetB0	

2.2. Celah Penelitian

Berdasarkan tinjauan pustaka yang telah dilakukan, penelitian yang secara khusus membahas klasifikasi tingkat roasting biji kopi masih terbatas, terutama pada klasifikasi multikelas *light roast, medium roast*, dan *dark roast* yang memiliki perbedaan visual sangat halus. Oleh karena itu, penelitian ini dilakukan untuk membandingkan performa DenseNet, MobileNet, dan EfficientNet dalam klasifikasi tingkat roasting biji kopi guna mengetahui model yang paling baik berdasarkan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.

2.3. Landasan Teori

2.3.1. Tingkat Roasting Biji Kopi

Roasting kopi merupakan proses pemanggangan biji kopi mentah (green bean) pada suhu tertentu untuk menghasilkan aroma, rasa, dan warna khas kopi. Tingkat roasting sangat memengaruhi karakteristik visual biji kopi, terutama warna dan tekstur permukaan. Secara umum, tingkat *roasting* kopi dibagi menjadi tiga

kategori utama (light roast, medium roast, dark roast). Perbedaan tiga kategori tersebut dapat dilihat pada gambar 2.1.



Gambar 2. 1 *Light roast* [15]

Light roast, biji kopi berwarna coklat muda, belum mengeluarkan minyak di permukaan, dan memiliki tingkat keasaman tinggi. Pada tingkat ini, proses pemanggangan masih relatif singkat sehingga senyawa asam masih dominan dan struktur biji belum mengalami karbonisasi yang signifikan. Berikut dapat dilihat pada gambar 2.2. mengenai *medium roast*.



Gambar 2. 2 Medium roast *[16]*

Medium roast ditandai dengan warna coklat sedang dan tekstur biji yang lebih merata, di mana reaksi *maillard* telah berlangsung lebih intens sehingga menghasilkan keseimbangan antara keasaman, aroma, dan rasa. Permukaan biji pada medium roast umumnya masih kering atau hanya sedikit berminyak. Sementara itu, berikut gambar 2. 3 mengenai *dark roast*.

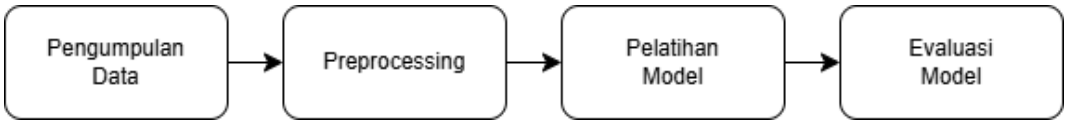


Gambar 2. 3 Dark Roast [15]

Dark roast memiliki warna coklat tua hingga hitam dengan permukaan biji yang berminyak akibat keluarnya minyak alami dari dalam biji kopi. Proses pemanggangan yang lebih lama menyebabkan tingkat keasaman menurun, rasa pahit lebih dominan, serta struktur biji menjadi lebih rapuh. Perbedaan karakteristik visual dan tekstur inilah yang dapat dimanfaatkan sebagai dasar dalam proses klasifikasi tingkat *roasting* kopi berbasis citra digital.

2.3.2. Computer Vision

Computer vision merupakan cabang ilmu kecerdasan buatan yang memungkinkan komputer untuk memperoleh, memproses, dan memahami informasi dari citra digital atau video. Tujuan utama *computer vision* adalah meniru kemampuan penglihatan manusia dalam mengenali objek, pola, dan karakteristik visual [17]. Dalam penelitian ini, *computer vision* digunakan untuk menganalisis citra biji kopi guna mengidentifikasi tingkat roasting berdasarkan karakteristik visual yang terekam dalam gambar. *Computer vision* membantu sistem mengidentifikasi objek, orang, tindakan, atau fitur lainnya dari data visual. Ini merupakan fondasi dari berbagai tugas seperti pengenalan gambar, klasifikasi, deteksi objek, segmentasi, dan pelacakan objek. Berikut alur kerja pada gambar 2.4 *computer vision*.



Gambar 2. 4 *Flowchart computer vision*

Penjelasan dari alur di atas dicantumkan di bawah ini:

1. Pengumpulan Data (Data Gathering)

   Mengumpulkan dataset citra yang relevan dengan tujuan. Dataset ini harus memiliki label yang benar untuk tugas klasifikasi atau anotasi untuk tugas lain seperti deteksi objek.

2. Pra-pemrosesan (Preprocessing)

   Memperbaiki kualitas data sebelum dilatih model, seperti *resizing*, penyesuaian *brightness/contrast*, normalisasi, serta augmentasi (misalnya rotasi, flip) untuk memperluas variasi data.

3. Pelatihan Model (Model Training)

   Model dilatih menggunakan dataset terlabel. Proses ini melibatkan optimasi parameter melalui *backpropagation* dan metode pengurangan kesalahan (loss function) hingga model mampu memetakan *input* ke *output* target secara akurat.

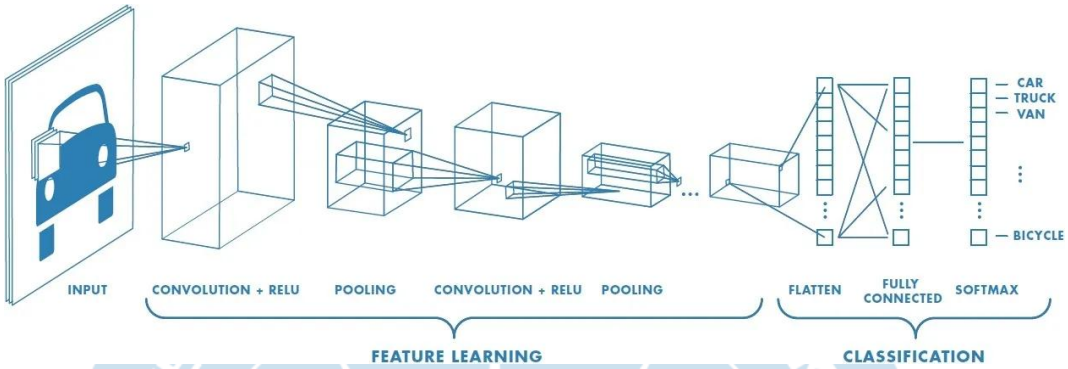
4. Evaluasi Model

   Evaluasi model merupakan tahap yang dilakukan untuk mengukur kinerja model klasifikasi dalam mengenali tingkat roasting biji kopi secara tepat. Proses evaluasi dilakukan dengan menggunakan data uji yang belum pernah dikenali oleh model sebelumnya. Beberapa metrik evaluasi yang umum digunakan yaitu accuracy, precision, recall, dan F1-score. Selain itu, confusion matrix juga digunakan untuk melihat kesalahan klasifikasi antar kategori roasting seperti light roast, medium roast, dan dark roast. Melalui proses evaluasi ini, dapat diketahui seberapa baik performa model dalam melakukan klasifikasi tingkat roasting biji kopi

2.3.3. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah jenis *deep learning* yang sangat efektif untuk tugas-tugas klasifikasi gambar dan visi komputer, karena mampu mengekstraksi pola dari citra secara otomatis tanpa perlu fitur manual seperti metode tradisional [18]. CNN merupakan varian dari jaringan saraf (neural network) yang secara arsitektural dirancang khusus untuk mengenali struktur spasial dalam data visual, seperti piksel, tepi, dan bentuk objek dalam gambar.

Model ini banyak digunakan dalam pengenalan objek, klasifikasi gambar, dan berbagai aplikasi visi komputer lainnya. CNN bekerja dengan melalui beberapa lapisan bertingkat, di mana setiap lapisan mengambil input dari lapisan sebelumnya dan memprosesnya untuk mengekstraksi fitur visual yang semakin kompleks. Secara umum, alur kerja CNN pada gambar 5 adalah sebagai berikut:



Gambar 2. 5 Arsitektur CNN [3]

Berdasarkan gambar 2.5, cara kerjanya dapat diringkas sebagai berikut:

1. Input Gambar, gambar masuk ke CNN dalam bentuk matriks piksel. Misalnya gambar RGB berukuran 32×32 piksel akan direpresentasikan sebagai matriks 32×32×3 (3 channel warna: merah, hijau, biru).
2. Convolutional Layer, setiap filter mendeteksi pola tertentu seperti garis, tepi, tekstur, atau bentuk. Saat filter digeser, dilakukan operasi perkalian dan penjumlahan antara nilai piksel dan bobot filter. Hasilnya adalah feature map (activation map) yang menunjukkan lokasi fitur tertentu pada gambar.
3. Pooling Layer, dilakukan pooling untuk mengecilkan ukuran feature map, biasanya menggunakan Max Pooling, yaitu mengambil nilai terbesar pada suatu area kecil. Contoh dapat dilihat pada Tabel 2.2.

Tabel 2. 2 Contoh Pooling

Sebelum Pooling	Sesudah Max Pooling
2 5	5
3 4	

Tujuannya mengurangi parameter, mempercepat proses komputasi, mengurangi overfitting, dan mempertahankan fitur

4. Pengulangan Conv + Pooling, CNN memiliki beberapa lapisan convolutional dan pooling, semakin dalam lapisannya, layer awal guna mendeteksi garis dan tepi, layer tengah guna mendeteksi tekstur dan bentuk sederhana, layer dalam guna mendeteksi bagian objek (warna, biji kopi), serta layer terakhir guna mengenali objek secara utuh.
5. Flatten, menjadi penghubung antara Feature Extraction Layer dan Fully Connected Layer.
6. Fully Connected Layer, bagian ini semua neuron saling terhubung dengan menggunakan fitur yang telah diekstraksi sebelumnya dan melakukan proses klasifikasi.
7. Output Layer, neuron terakhir menghasilkan probabilitas setiap kelas. Contoh klasifikasi tingkat roasting biji kopi dapat dilihat pada Tabel 2.3.

Tabel 2. 3 Contoh Output

Kelas Roasting	Probabilitas
Light Roast	0.08 (8%)
Medium Roast	0.76 (76%)
Dark Roast	0.16 (16%)

Karena probabilitas tertinggi adalah 76% pada kelas Medium Roast, maka sistem mengklasifikasikan gambar tersebut sebagai Medium Roast.

Secara keseluruhan, CNN mengekstraksi pola visual secara hierarkis dari fitur sederhana di lapisan awal hingga pola kompleks di lapisan lebih dalam, sehingga mampu mengenali objek atau klasifikasi gambar secara akurat. Dari pengembangan tersebut lahirlah berbagai varian CNN, di antaranya DenseNet, MobileNet, dan EfficientNet, yang masing-masing mengadopsi prinsip dasar CNN namun dengan strategi arsitektur yang berbeda. DenseNet berfokus pada penguatan aliran informasi antar lapisan melalui koneksi padat, MobileNet menekankan efisiensi komputasi dengan konvolusi terpisah, sedangkan EfficientNet mengoptimalkan keseimbangan antara kedalaman, lebar, dan resolusi jaringan. Dengan demikian, ketiga arsitektur tersebut dapat dipandang sebagai bentuk pengembangan CNN yang dirancang untuk memenuhi kebutuhan performa dan efisiensi pada berbagai skenario klasifikasi citra.

Deep learning merupakan cabang dari *machine learning* yang menggunakan jaringan saraf tiruan dengan banyak lapisan (deep neural network). Model *deep learning* mampu mempelajari pola kompleks dari data dalam jumlah besar dan telah banyak digunakan dalam berbagai bidang, termasuk pengenalan citra, pengolahan bahasa alami, dan pengenalan suara. Keunggulan *deep learning* terletak pada kemampuannya melakukan pembelajaran fitur secara otomatis tanpa memerlukan perancangan fitur secara manual [19].

Convolutional Layer merupakan komponen utama dalam arsitektur Convolutional Neural Network (CNN) yang berfungsi untuk melakukan ekstraksi fitur secara otomatis dari citra masukan. Berbeda dengan metode pengolahan citra konvensional yang memerlukan proses ekstraksi fitur secara manual, Convolutional Layer mampu mempelajari karakteristik penting dari data melalui proses pelatihan sehingga menghasilkan representasi fitur yang lebih efektif untuk proses klasifikasi.[20]

Proses utama pada *Convolutional Layer* dilakukan dengan menggeser sebuah kernel atau filter berukuran kecil pada seluruh area citra masukan. Pada setiap posisi, dilakukan operasi perkalian elemen demi elemen antara nilai piksel citra dengan bobot kernel, kemudian hasilnya dijumlahkan sehingga menghasilkan satu nilai pada ***feature map***. ***Feature map*** tersebut merepresentasikan keberadaan pola tertentu, seperti tepi (edges), garis, tekstur, maupun bentuk objek yang terdapat pada citra. Semakin banyak filter yang digunakan, semakin banyak karakteristik visual yang dapat dipelajari oleh model.[21] Persamaan operasi konvolusi pada Convolutional Layer ditunjukkan pada Persamaan 2.1.

$$Y(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X\,(i+m, j+n) \times K(m,n) + b \qquad (2.1)$$

- $Y(i,j)$= nilai output (feature map) pada posisi (i, j)
- $X(i+m, j+n)$= nilai piksel citra masukan (input image)
- $K(m,n)$= nilai bobot kernel/filter
- b= nilai bias
- k= ukuran kernel (misalnya 3×3 atau 5×5)
- m, n= indeks elemen pada kernel

Convolutional Layer biasanya diteruskan ke **fungsi aktivasi ReLU *(Rectified Linear Unit)*** yang berfungsi menambahkan sifat non-linear pada jaringan sehingga CNN mampu mempelajari pola yang lebih kompleks. Setelah itu, feature map akan diproses oleh ***Pooling Layer*** untuk mengurangi dimensi spasial tanpa menghilangkan informasi penting. Kombinasi Convolutional Layer, fungsi aktivasi, dan Pooling Layer memungkinkan CNN membangun representasi fitur secara bertingkat (*hierarchical feature learning*). Lapisan awal akan mempelajari fitur sederhana seperti garis dan tepi, sedangkan lapisan yang lebih dalam mampu mengenali bentuk, tekstur, hingga karakteristik objek yang lebih kompleks.

Learning rate merupakan salah satu **hyperparameter** yang paling penting dalam proses pelatihan **Convolutional Neural Network (CNN)**. Learning rate menentukan besar kecilnya langkah (step size) yang digunakan algoritma optimasi dalam memperbarui bobot (weights) model pada setiap iterasi selama proses backpropagation. Pemilihan nilai learning rate yang tepat sangat berpengaruh terhadap kecepatan konvergensi, stabilitas proses pelatihan, serta akurasi model yang dihasilkan. Pada setiap iterasi pelatihan, bobot model diperbarui berdasarkan gradien dari fungsi loss. Secara umum, proses pembaruan bobot dapat dinyatakan dalam persamaan 2.2.[21]

$$W_{t+1} = W_t - \propto \nabla L(w_t) \tag{2.2}$$

dengan:

- W_t= bobot pada iterasi ke-t,
- W_{t+1}= bobot setelah diperbarui,
- α= learning rate,
- $\nabla L(W_t)$= gradien fungsi loss terhadap bobot.

Persamaan tersebut menunjukkan bahwa nilai learning rate mengontrol seberapa besar perubahan bobot pada setiap iterasi. Nilai learning rate yang terlalu besar maupun terlalu kecil dapat memengaruhi performa model CNN.[21]

Nilai learning rate yang terlalu besar (**high learning rate**) menyebabkan pembaruan bobot berlangsung dengan langkah yang besar sehingga proses pelatihan menjadi lebih cepat. Namun, perubahan bobot yang terlalu besar dapat menyebabkan model melewati titik optimum (overshooting), mengakibatkan nilai loss berosilasi, bahkan gagal mencapai konvergensi. Kondisi tersebut dapat menurunkan akurasi model karena proses pembelajaran menjadi tidak stabil. Sebaliknya, learning rate yang terlalu kecil (**low learning rate**) menghasilkan perubahan bobot yang sangat sedikit pada setiap iterasi. Akibatnya, proses pelatihan berlangsung lebih stabil, tetapi membutuhkan waktu yang lebih lama untuk mencapai konvergensi. Selain meningkatkan waktu komputasi, learning rate yang terlalu kecil juga berpotensi membuat model terjebak pada solusi lokal (local minima) atau berhenti belajar sebelum mencapai performa optimal.

2.3.4. DenseNet

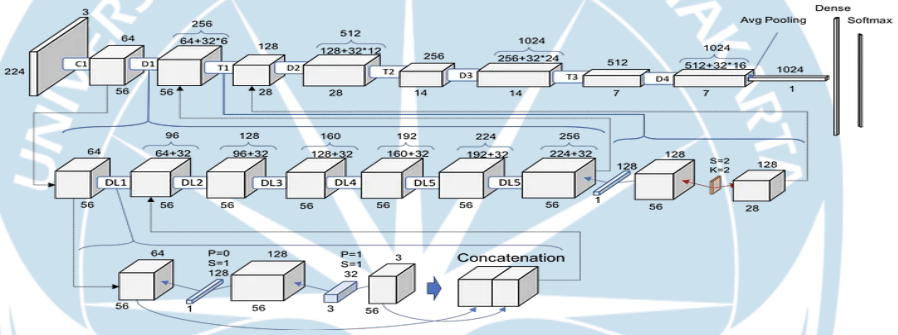
DenseNet121 merupakan salah satu arsitektur Convolutional Neural Network (CNN) yang dikembangkan untuk meningkatkan efisiensi proses pembelajaran pada jaringan saraf dalam. DenseNet (Densely Connected Convolutional Network) memiliki karakteristik utama berupa koneksi langsung antara setiap layer dengan seluruh layer sebelumnya. Melalui mekanisme tersebut, setiap layer dapat menerima informasi fitur dari layer-layer sebelumnya sehingga membantu mengurangi kehilangan informasi selama proses pelatihan model.

Arsitektur DenseNet121 terdiri dari 121 layer yang tersusun atas beberapa dense block dan transition layer. Pada setiap dense block, fitur yang dihasilkan akan diteruskan ke layer berikutnya melalui proses konkatenasi. Hal ini memungkinkan model untuk menggunakan kembali fitur (feature reuse) sehingga jumlah parameter yang digunakan menjadi lebih efisien dibandingkan beberapa arsitektur CNN lainnya. Dalam penelitian klasifikasi tingkat roasting biji kopi, DenseNet121 digunakan karena mampu mengekstraksi fitur visual secara mendalam, seperti perbedaan warna, tekstur, dan pola permukaan biji kopi. Kemampuan tersebut sangat membantu dalam membedakan kategori roasting yang memiliki karakteristik visual yang mirip, terutama pada kelas medium roast. Selain memiliki performa klasifikasi yang baik, DenseNet121 juga mampu mengurangi risiko

overfitting pada proses pelatihan model.

Cara kerja DenseNet121 dimulai dari proses input citra yang akan melewati beberapa tahap ekstraksi fitur. Pada setiap layer, hasil fitur yang diperoleh tidak hanya diteruskan ke layer berikutnya, tetapi juga dikirim ke seluruh layer setelahnya dalam satu dense block.

Berbeda dengan CNN konvensional yang hanya menghubungkan layer secara berurutan, DenseNet menggunakan koneksi padat (dense connection) sehingga setiap layer memperoleh akses langsung terhadap fitur dari seluruh layer sebelumnya. Mekanisme ini membantu model mempertahankan informasi penting selama proses pelatihan. Ilustrasi arsitektur DenseNet121 dapat dilihat pada Gambar 2.6.



Gambar 2. 6 Arsitektur DenseNet121 [22]

Alur kerja arsitektur DenseNet-121 [22]dimulai dengan memasukkan citra ke dalam jaringan sebagai data masukan. Pada tahap awal, citra diproses melalui lapisan konvolusi (convolution layer) berukuran 7×7 yang diikuti oleh Batch Normalization, fungsi aktivasi ReLU, dan Max Pooling. Tahap ini bertujuan untuk mengekstraksi fitur-fitur dasar dari citra, seperti tepi, garis, tekstur, dan pola sederhana yang terdapat pada objek. Selanjutnya, fitur yang telah diperoleh akan diproses melalui serangkaian Dense Block yang menjadi ciri khas arsitektur DenseNet. Berbeda dengan CNN konvensional, setiap layer pada Dense Block menerima masukan tidak hanya dari layer sebelumnya, tetapi juga dari seluruh layer yang telah diproses sebelumnya dalam blok yang sama. Mekanisme ini memungkinkan informasi dan fitur yang telah dipelajari pada layer awal tetap tersedia dan dapat dimanfaatkan oleh layer-layer berikutnya. Dengan demikian, DenseNet mampu mengurangi hilangnya informasi selama proses pelatihan serta meningkatkan efisiensi penggunaan parameter.

Setiap Dense Block terdapat Transition Layer yang berfungsi untuk mengurangi ukuran feature map dan jumlah channel. Transition Layer terdiri dari Batch Normalization, konvolusi 1×1, dan Average Pooling. Pengurangan dimensi ini bertujuan untuk menekan kebutuhan komputasi dan penggunaan memori tanpa menghilangkan informasi penting yang telah dipelajari oleh jaringan.

Arsitektur DenseNet-121 terdiri dari empat Dense Block dengan jumlah layer berturut-turut sebanyak 6, 12, 24, dan 16 layer. Selama proses ini, jaringan secara bertahap mempelajari fitur yang semakin kompleks. Layer awal umumnya mengenali fitur sederhana seperti warna dan tekstur, sedangkan layer yang lebih dalam mampu mengenali pola yang lebih spesifik dan kompleks yang berkaitan dengan objek yang diklasifikasikan. Setelah melewati Dense Block terakhir, feature map yang dihasilkan akan diproses menggunakan Global Average Pooling (GAP). Metode ini menghitung nilai rata-rata dari setiap feature map sehingga menghasilkan representasi fitur yang lebih ringkas. Penggunaan Global Average Pooling membantu mengurangi jumlah parameter serta meminimalkan risiko overfitting dibandingkan penggunaan fully connected layer yang terlalu besar.

Tahap akhir dari arsitektur DenseNet-121 adalah Fully Connected Layer yang berfungsi melakukan proses klasifikasi. Fitur yang telah diringkas kemudian dipetakan ke dalam sejumlah kelas yang telah ditentukan. Pada lapisan output, fungsi aktivasi Softmax digunakan untuk menghasilkan nilai probabilitas pada setiap kelas. Kelas dengan probabilitas tertinggi akan dipilih sebagai hasil prediksi model. Dalam penelitian klasifikasi tingkat roasting biji kopi, DenseNet-121 akan mempelajari karakteristik visual seperti warna permukaan biji, tingkat kegelapan, tekstur, serta pola retakan yang muncul akibat proses penyangraian. Informasi tersebut kemudian diolah melalui Dense Block dan Transition Layer hingga menghasilkan fitur yang representatif. Pada akhirnya, model akan mengklasifikasikan citra ke dalam kategori roasting, seperti Light Roast, Medium Roast, atau Dark Roast, berdasarkan nilai probabilitas tertinggi yang dihasilkan pada lapisan output.

DenseNet121 memiliki beberapa keunggulan yang menjadikannya salah satu arsitektur deep learning yang banyak digunakan dalam tugas klasifikasi citra. Salah satu keunggulan utamanya adalah kemampuannya dalam mengurangi

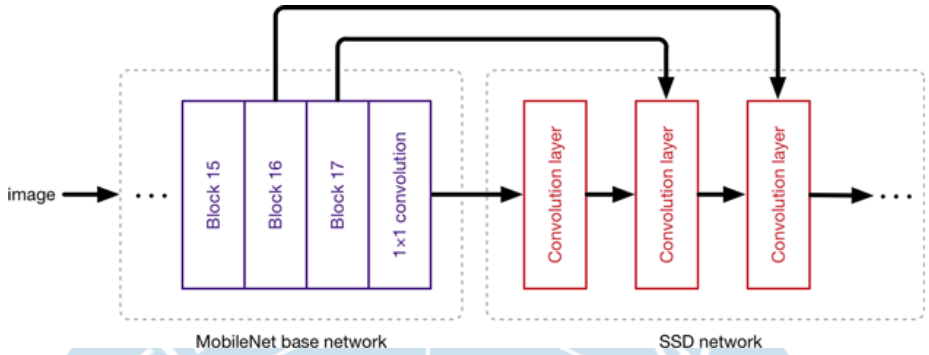
masalah vanishing gradient melalui koneksi langsung antar layer, sehingga aliran gradien selama proses pelatihan menjadi lebih stabil. Selain itu, DenseNet menerapkan konsep feature reuse, yaitu penggunaan kembali fitur yang telah dipelajari oleh layer sebelumnya, sehingga proses pembelajaran menjadi lebih efisien dan informasi penting tidak mudah hilang. Arsitektur ini juga dikenal memiliki jumlah parameter yang relatif lebih sedikit dibandingkan beberapa arsitektur CNN lainnya dengan tingkat kedalaman yang serupa, sehingga penggunaan parameter menjadi lebih efisien tanpa mengorbankan performa. DenseNet121 juga mampu menghasilkan akurasi klasifikasi yang tinggi pada berbagai jenis dataset citra. Keunggulan lainnya adalah kemampuannya dalam membedakan karakteristik visual yang memiliki perbedaan halus, sehingga sangat cocok digunakan untuk klasifikasi tingkat roasting biji kopi yang mengandalkan perbedaan warna, tekstur, dan tingkat kematangan hasil sangrai.

Berbagai keunggulannya, DenseNet121 juga memiliki beberapa keterbatasan. Banyaknya koneksi antar layer menyebabkan kebutuhan memori menjadi lebih besar dibandingkan arsitektur CNN yang lebih sederhana. Selain itu, proses concatenation yang dilakukan pada setiap layer membuat waktu pelatihan (training) cenderung lebih lama karena jaringan harus memproses dan menyimpan lebih banyak informasi. Struktur arsitektur DenseNet yang kompleks juga membuat proses implementasi dan optimasinya lebih menantang dibandingkan model CNN konvensional. Oleh karena itu, untuk memperoleh performa yang optimal, pelatihan DenseNet121 umumnya memerlukan perangkat keras dengan kemampuan komputasi yang baik, terutama GPU yang memiliki kapasitas memori dan kecepatan pemrosesan yang memadai

2.3.5. MobileNetV2

MobileNetV2 merupakan pengembangan dari MobileNet versi pertama dengan penambahan konsep inverted residual dan linear bottleneck. Kedua konsep tersebut membantu model melakukan ekstraksi fitur dengan jumlah parameter yang lebih sedikit serta proses komputasi yang lebih cepat. Arsitektur ini banyak digunakan dalam penelitian klasifikasi citra karena memiliki ukuran model yang ringan namun tetap mampu menghasilkan akurasi yang baik. Pada penelitian

klasifikasi tingkat roasting biji kopi, MobileNetV2 digunakan untuk mengenali perbedaan visual antar kategori roasting berdasarkan citra digital. Ilustrasi arsitektur MobileNetV2 dapat dilihat pada Gambar 2.7.



Gambar 2. 7 Arsitektur MobileNetV2 [23]

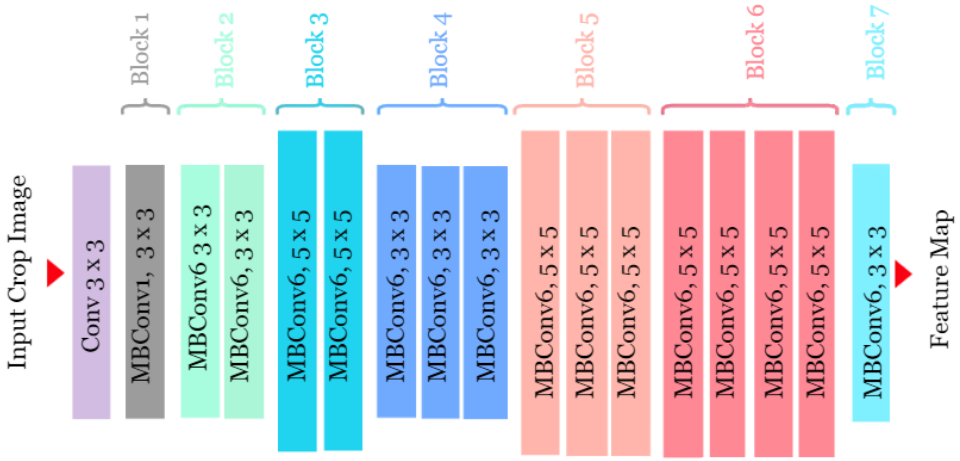
Cara kerja MobileNetV2 [23] bekerja dengan menerima citra sebagai input yang kemudian diproses melalui lapisan konvolusi awal untuk mengekstraksi fitur dasar seperti tepi, warna, dan tekstur. Selanjutnya, fitur tersebut diproses menggunakan Inverted Residual Block, yang terdiri dari tiga tahap utama, yaitu Expansion Layer, Depthwise Convolution, dan Linear Bottleneck. Expansion Layer berfungsi memperbesar jumlah channel agar model dapat mempelajari lebih banyak informasi, sedangkan Depthwise Convolution mengekstraksi fitur pada setiap channel secara terpisah sehingga komputasi menjadi lebih efisien. Setelah itu, Linear Bottleneck mengurangi kembali jumlah channel untuk mempertahankan informasi penting dengan jumlah parameter yang lebih sedikit

Pada beberapa blok, MobileNetV2 menggunakan *residual connection* yang menghubungkan input dan output blok untuk membantu aliran informasi selama proses pelatihan. Setelah seluruh blok diproses, fitur yang diperoleh diringkas menggunakan *Global Average Pooling* dan diteruskan ke *Fully Connected Layer* untuk proses klasifikasi. Pada lapisan output, fungsi Softmax menghasilkan probabilitas setiap kelas, kemudian kelas dengan probabilitas tertinggi dipilih sebagai hasil prediksi. Dalam klasifikasi tingkat roasting biji kopi, MobileNetV2 memanfaatkan fitur seperti warna, tekstur, dan tingkat kegelapan biji kopi untuk membedakan kategori roasting yang berbeda.

Kelebihan utama MobileNetV2 terletak pada arsitekturnya yang ringan sehingga membutuhkan jumlah parameter dan komputasi yang lebih sedikit dibandingkan banyak model CNN lainnya. Penggunaan teknik *depthwise separable convolution* membuat proses pelatihan dan pengujian model menjadi lebih cepat serta cocok diterapkan pada perangkat dengan spesifikasi terbatas seperti *smartphone* atau *embedded system*. Selain itu, MobileNetV2 tetap mampu menghasilkan performa klasifikasi yang baik meskipun memiliki ukuran model yang relatif kecil.\n\nNamun, MobileNetV2 juga memiliki beberapa kekurangan. Karena dirancang untuk efisiensi, kemampuan model dalam menangkap fitur visual yang sangat detail terkadang lebih rendah dibandingkan model CNN yang lebih besar dan kompleks. Pada dataset dengan perbedaan visual yang sangat halus, performa model dapat mengalami penurunan akurasi. Selain itu, kualitas dataset dan pengaturan *hyperparameter* sangat mempengaruhi hasil klasifikasi sehingga diperlukan *proses tuning* agar model dapat bekerja secara optimal

2.3.6. EfficientNetB0

EfficientNetB0 merupakan salah satu arsitektur Convolutional Neural Network (CNN) yang dikembangkan oleh Google dengan tujuan meningkatkan performa klasifikasi citra secara lebih efisien. EfficientNet diperkenalkan menggunakan metode compound scaling, yaitu teknik yang menyeimbangkan peningkatan kedalaman (depth), lebar (width), dan resolusi input gambar secara bersamaan. Pendekatan ini membuat EfficientNet mampu menghasilkan akurasi tinggi dengan jumlah parameter dan kebutuhan komputasi yang lebih efisien dibandingkan banyak arsitektur CNN lainnya. EfficientNetB0 merupakan model dasar (baseline model) dari keluarga EfficientNet yang memiliki struktur lebih ringan namun tetap memberikan performa yang baik dalam berbagai tugas klasifikasi citra. Ilustrasi arsitketur EfficientNetB0 dapat dilihat pada Gambar 2.8.



Gambar 2. 8 Arsitektur EfficientNetB0 [24]

Cara kerja EfficientNetB0 [24] EfficientNet-B0 bekerja dengan menerima citra sebagai input yang kemudian diproses melalui lapisan konvolusi awal untuk mengekstraksi fitur dasar seperti warna, tepi, dan tekstur. Selanjutnya, fitur tersebut diproses oleh beberapa MBConv Block (Mobile Inverted Bottleneck Convolution) yang menjadi komponen utama EfficientNet-B0. Setiap MBConv Block terdiri dari proses ekspansi fitur, depthwise convolution, dan modul Squeeze-and-Excitation (SE) yang membantu model fokus pada fitur yang paling penting. Selain itu, beberapa blok menggunakan skip connection untuk mempertahankan aliran informasi selama proses pelatihan.

Setelah melewati seluruh MBConv Block, feature map diringkas menggunakan Global Average Pooling dan diteruskan ke Fully Connected Layer untuk proses klasifikasi. Pada lapisan output, fungsi Softmax menghasilkan probabilitas setiap kelas, kemudian kelas dengan probabilitas tertinggi dipilih sebagai hasil prediksi. Dalam klasifikasi tingkat roasting biji kopi, EfficientNet-B0 memanfaatkan fitur seperti warna, tekstur, dan tingkat kegelapan biji kopi untuk membedakan kategori roasting secara efisien dengan jumlah parameter yang relatif kecil. EfficientNetB0 memiliki beberapa kelebihan, seperti efisiensi parameter yang tinggi, ukuran model yang relatif ringan, serta kemampuan menghasilkan akurasi yang baik pada berbagai dataset citra. Model ini juga memiliki proses komputasi yang lebih cepat dibandingkan beberapa arsitektur CNN besar sehingga cocok digunakan pada perangkat dengan spesifikasi menengah. Selain itu, metode compound scaling membuat EfficientNetB0 mampu menyeimbangkan performa

dan efisiensi secara optimal. Dalam klasifikasi tingkat roasting biji kopi, EfficientNetB0 efektif digunakan untuk mengenali perbedaan warna dan tekstur antar kelas roasting yang memiliki karakteristik visual cukup mirip.

Namun, EfficientNetB0 juga memiliki beberapa kekurangan. Struktur model yang cukup kompleks membuat proses implementasi dan tuning menjadi lebih sulit dibandingkan CNN sederhana. Selain itu, meskipun efisien, proses pelatihan tetap membutuhkan perangkat keras yang memadai terutama ketika menggunakan dataset berukuran besar. Pada beberapa kasus, model juga dapat mengalami penurunan performa apabila kualitas dataset kurang baik atau jumlah data latih terbatas.

2.3.7. Metrik Evaluasi

Evaluasi model dilakukan untuk mengetahui kemampuan model dalam mengklasifikasikan tingkat roasting biji kopi secara tepat. Pada penelitian ini, performa model dievaluasi menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score. Metrik tersebut diperoleh berdasarkan hasil confusion matrix yang terdiri dari True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

2.3.8. Accurasi

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam melakukan klasifikasi secara keseluruhan. Nilai accuracy menunjukkan perbandingan antara jumlah prediksi benar dengan seluruh jumlah data yang diuji. Rumus dapat dilihat pada persamaan 2.2.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.3}$$

Semakin tinggi nilai accuracy, maka semakin baik kemampuan model dalam melakukan klasifikasi tingkat roasting biji kopi

2.3.9. Precision

Precision digunakan untuk mengukur tingkat ketepatan prediksi positif yang dihasilkan model. Metrik ini menunjukkan seberapa banyak data yang diprediksi Output dari positif benar-benar termasuk dalam kelas tersebut. Rumus dapat dilihat padapersamaan2.3.

$$Precision = \frac{TP}{TP + FP} \tag{2.4}$$

Nilai precision yang tinggi menunjukkan bahwa model memiliki kesalahan prediksi positif yang rendah.

2.3.10. Recall

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data positif yang sebenarnya ada pada dataset. Rumus dapat dilihat pada persamaan 2.4.

$$Recall = \frac{TP}{TP + FN} \tag{2.5}$$

Semakin tinggi nilai recall, maka semakin baik model dalam mendeteksi data pada kelas tertentu.

2.3.11. F1-Score

F1-score merupakan nilai harmonisasi antara precision dan recall. Metrik ini digunakan untuk mengetahui keseimbangan performa model dalam melakukan klasifikasi. Rumus dapat dilihat pada persamaan 2.6.

$$F1 - Score = 2 * \frac{Precision \ * \ Recall}{Precision \ + \ Reccall} \tag{2.6}$$

Nilai F1-score yang tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara precision dan recall dalam proses klasifikasi tingkat roasting biji kopi.