

Comparative Analysis of EfficientNet-B0 and ViT-B16 for Multiclass Classification of Green Coffee Beans

Muh. Rezky Syaputra^{1*}, Majid Rahardi^{2*}

* Informatics, Universitas Amikom Yogyakarta

rezkysyaputra@students.amikom.ac.id¹, majid@amikom.ac.id²

Article Info

Article history:

Received 2025-10-23

Revised 2025-11-23

Accepted 2025-12-10

Keyword:

*EfficientNetB0,
Multi-class Classification,
Transfer Learning,
USKCoffee,
ViTB16*

ABSTRACT

Green coffee bean classification plays an important role in the coffee supply chain, as bean quality has a direct impact on the taste and final quality of the product. The USK-Coffee dataset, which consists of four bean object classes defect, longberry, peaberry, and premium, is photographed under varied lighting conditions and capture angles, thus challenging the accuracy of conventional visual models. Although lightweight CNN models have been used, not many studies have directly compared transformer-based architectures (ViT-B16) and modern efficient CNNs (EfficientNet-B0) for green coffee bean classification under real conditions. With transfer learning strategy, image augmentation (resize, flip, rotation, color jitter, random crop), and normalization, we evaluate the performance of both models on the dataset. ViT-B16 achieved 85% accuracy on the test data (F1-score 0.85), with a fast batch inference latency of 0.0074 seconds per batch. EfficientNet-B0 achieved 87% accuracy (F1-score 0.87), with a slower batch latency (0.0106 seconds per batch). However, EfficientNet-B0 is significantly faster for single image inference (real-time) (0.035 seconds) compared to ViT-B16 (0.426 seconds). This trade-off higher accuracy/faster single inference on EfficientNet-B0 vs. faster batch processing on ViT-B16 shows that both are feasible for edge computing-based classification systems.



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

I. PENDAHULUAN

Pengklasifikasian biji kopi hijau (*green coffee beans*) merupakan tahapan krusial dalam rantai pasok industri kopi global, karena mutu biji secara langsung menentukan nilai jual, cita rasa, dan kualitas akhir produk yang sampai ke tangan konsumen [1]. Kualitas dan harga kopi tergantung pada beberapa faktor, salah satunya adalah ukuran biji yang seragam [2]. Proses klasifikasi biji kopi hijau sangat krusial untuk menjaga standar kualitas dan meningkatkan nilai pasar, terutama melalui pemisahan biji yang cacat atau tidak memenuhi standar.

Tantangan utama dalam klasifikasi biji kopi secara manual terletak pada sifat prosesnya yang rentan terhadap kesalahan, perilaku subjektif, dan ketidakkonsistenan akibat kelelahan atau kondisi lingkungan yang bervariasi [3]. Hal ini dapat menyebabkan potensi kerugian produksi dan reputasi bagi industri kopi. Mengingat volume produksi kopi yang sangat

tinggi, terutama di negara-negara produsen besar, solusi otomatisasi menjadi esensial untuk memastikan efisiensi dan konsistensi.

Di Indonesia, yang merupakan salah satu penghasil kopi utama di seluruh dunia, isu ini menjadi sangat krusial. Indonesia berada di urutan keempat di dunia sebagai negara penghasil kopi terbesar, dengan jumlah produksi yang melebihi 700.000 ton setiap tahunnya [4], dan memiliki ambisi untuk melampaui Vietnam sebagai produsen kopi terbesar kedua secara global [5]. Produksi kopi nasional mencapai sekitar 775 ribu ton pada tahun 2022 [6], dengan luas lahan kopi mencapai lebih dari 1,25 juta hektar dan melibatkan lebih dari 1,8 juta petani [7]. Skala produksi yang masif ini menegaskan urgensi pengembangan sistem klasifikasi mutu biji kopi yang efisien, objektif, dan konsisten.

Penerapan kecerdasan buatan (*Artificial Intelligence*), khususnya *deep learning*, telah menunjukkan potensi besar sebagai solusi untuk mengatasi tantangan tersebut [8].

Berbagai studi sebelumnya telah memperlihatkan potensi *deep learning* dalam klasifikasi biji kopi. Misalnya, Korkmaz et al. [9] membandingkan lima model CNN (Xception, DenseNet201, InceptionV3, InceptionResNetV2, dan DenseNet121) untuk klasifikasi spesies biji kopi, di mana InceptionV3 mencapai akurasi tertinggi 93%. Khusus untuk kopi hijau, Arwatchananukul et al. [10] mengembangkan model untuk klasifikasi cacat biji kopi Arabika Thailand, di mana MobileNetV3 mencapai akurasi 99,84%. Studi lain oleh Nair dkk. [11] melaporkan akurasi 100% pada model EfficientNet-B4 untuk *grading* biji kopi, dan Gonc dkk. [12] menerapkan EfficientNet-B4 pada dataset biji kopi hijau multikelas dan mendapatkan akurasi 0.8844.

Penelitian juga telah membandingkan CNN dan *Vision Transformer* (ViT). Leonardi dkk. [13] menemukan bahwa untuk biji kopi sangrai, ViT-B16 akurasi 99.33% mengungguli model CNN seperti Xception 96.67%. Penelitian Izza dan Kusuma [14] telah meneliti beberapa varian *transformer* (ViT, DeiT, Swin) pada dataset USK-Coffee yang sama dengan penelitian ini, namun hasil untuk ViT-B16 masih terbatas dan belum dibandingkan secara langsung dengan EfficientNet-B0.

Gap analysis penelitian terkait menyoroti dua hal utama: (i) belum ada komparasi langsung dan komprehensif antara arsitektur *transformer* seperti ViT-B16 dan CNN efisien seperti EfficientNet-B0 untuk tugas klasifikasi multikelas biji kopi hijau dalam kondisi pengambilan citra non-kontrol; dan (ii) evaluasi performa seringkali hanya terbatas pada akurasi, tanpa mempertimbangkan metrik penting seperti presisi, *recall*, dan *F1-score* per kelas yang sangat kritis dalam penilaian mutu kopi.

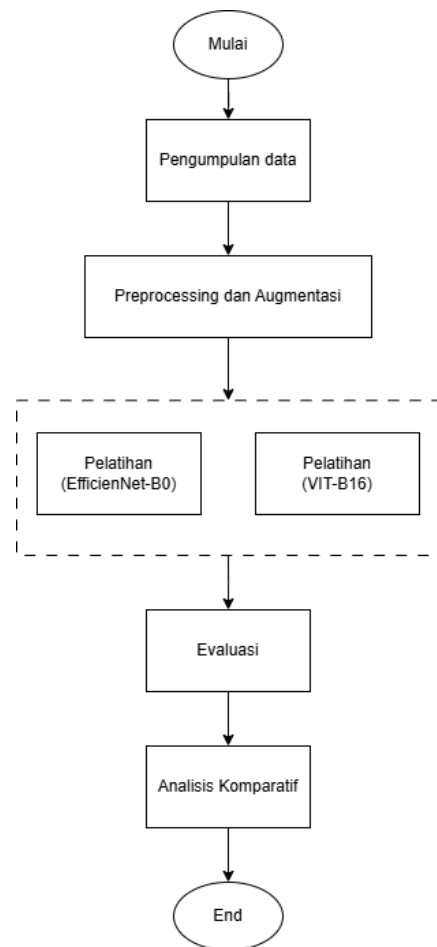
Penelitian ini bertujuan untuk mengatasi kekurangan tersebut dengan membandingkan hasil kinerja model ViT-B16 dan EfficientNet-B0 dalam klasifikasi biji kopi hijau multikelas menggunakan dataset USK-Coffee. Penelitian ini menggunakan strategi *transfer learning* [15], dilengkapi augmentasi gambar (termasuk *resize* ke 224×224 , *flip*, rotasi, *color jitter*, *random crop*), serta normalisasi RGB [16]. Evaluasi dilakukan secara menyeluruh menggunakan akurasi, presisi, *recall*, dan *F1-score* per kelas. Selain itu, analisis efisiensi juga mencakup kecepatan inferensi untuk menilai kesesuaian implementasi di perangkat *edge computing*.

II. METODE

A. Tahapan Penelitian

Penelitian ini membandingkan kinerja ViT-B16 dan EfficientNet-B0 untuk klasifikasi biji kopi hijau pada Dataset USK-Coffee. Prosesnya meliputi: persiapan dan pembagian data, pra-pemrosesan dan augmentasi citra, pembangunan model *deep learning* (ViT-B16 dan EfficientNet-B0) menggunakan *transfer learning* [15], pelatihan model dengan data latih dan validasi, evaluasi kinerja model pada data uji, dan terakhir analisis komparatif hasil untuk kedua arsitektur. Tahapan ini dirancang untuk memastikan validitas dan

reliabilitas hasil perbandingan kinerja kedua model pada kondisi data yang realistis.



Gambar 1. Tahapan penelitian

B. Pengumpulan Data

Tahap awal penelitian berfokus pada persiapan Dataset USK-Coffee. Dataset ini mengandung citra biji kopi yang diklasifikasikan ke dalam empat kategori: defect, longberry, peaberry, dan premium. Citra-citra ini dikumpulkan dalam berbagai kondisi pencahayaan dan sudut pengambilan, merefleksikan variabilitas yang mungkin ditemui dalam lingkungan nyata. Total citra yang digunakan dalam dataset adalah 8000, dengan distribusi yang seimbang antar kelas untuk menghindari bias.

TABEL I
JUMLAH CITRA BIJI KOPI

Kelas	Training	Validasi	Testing	Total
Defect	1200	400	400	2000
Longberry	1200	400	400	2000
Peaberry	1200	400	400	2000
Premium	1200	400	400	2000
Total	4800	1600	1600	8000

Tabel I menyajikan jumlah citra untuk setiap kelas biji kopi (Defect, Longberry, Peaberry, Premium) yang dialokasikan ke dalam set pelatihan, validasi, dan pengujian. Gambar 2 menyajikan empat contoh citra dari setiap kelas biji kopi yang digunakan dalam penelitian ini. (a) Defect, (b) Longberry, (c) Peaberry, dan (d) Premium.



Gambar 2. Citra jenis biji kopi: (a) Defect, (b) Longberry, (c) Peaberry, dan (d) Premium.

C. Preprocessing dan Augmentasi Data

Seluruh citra dalam dataset, baik untuk pelatihan, validasi, maupun pengujian, melalui tahap pra-pemrosesan standar. Tahap ini dimulai dengan mengubah ukuran setiap citra menjadi resolusi seragam 224×224 piksel. Setelah itu, nilai piksel dinormalisasi menggunakan rata-rata (*mean*) [0.485, 0.456, 0.406] dan standar deviasi (*std*) [0.229, 0.224, 0.225], yang merupakan nilai standar dari dataset ImageNet untuk mengoptimalkan penggunaan model *pre-trained*.

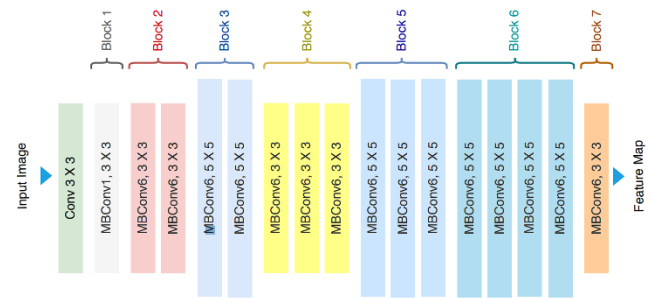
Untuk meningkatkan kemampuan generalisasi model dan mencegah *overfitting* [16],[17], serangkaian teknik augmentasi data diterapkan secara acak hanya pada set data pelatihan. Teknik-teknik yang digunakan meliputi *Random Horizontal Flip* (membalik citra secara horizontal), *Random Resized Crop* (melakukan pemotongan acak pada citra dengan skala antara 80% hingga 100% dari ukuran aslinya, lalu mengubah ukurannya kembali menjadi 224×224 piksel), *Color Jitter* (mengubah properti warna citra secara acak, termasuk kecerahan (*brightness*), kontras (*contrast*), saturasi (*saturation*), dan *hue* dengan faktor 0.1, yang secara khusus mensimulasikan variasi pencahayaan atau perubahan warna kecil akibat tingkat *roasting* atau kekotoran), serta *Random Rotation* (memutar citra secara acak hingga 15 derajat).

Augmentasi ini secara efektif memperkaya keragaman data latih, mensimulasikan berbagai kondisi pengambilan gambar, dan membuat model lebih *robust* terhadap variasi pada data baru.

D. Pembangunan Model

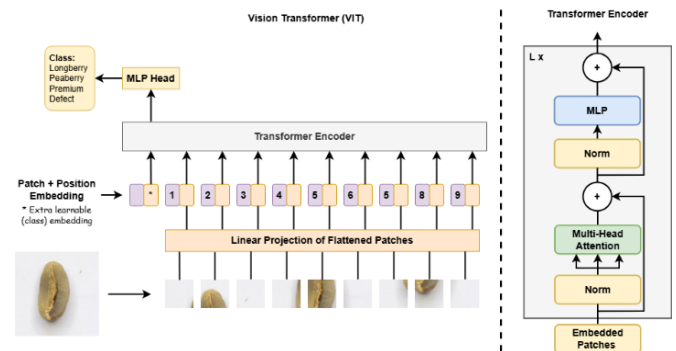
Dua arsitektur deep learning modern dipilih: EfficientNet-B0 dan ViT-B16. Kedua model ini dilatih menggunakan strategi transfer learning [15], memanfaatkan bobot yang telah dilatih pada dataset ImageNet. Dalam implementasi ini, semua lapisan jaringan (*all layers*) dari kedua model di-*fine-tune* (*unfrozen*) untuk menyesuaikan fitur yang telah dipelajari dengan domain biji kopi yang spesifik. Perbedaan jumlah *trainable parameters* juga menjadi faktor efisiensi: *EfficientNet-B0* memiliki 5.3 juta parameter, sedangkan *ViT-B16* memiliki kompleksitas yang jauh lebih tinggi dengan 86 juta parameter.

1) *EfficientNet-B0*: Arsitektur Convolutional Neural Network (CNN) ini terkenal karena efisiensinya, menggunakan pendekatan penskalaan komposit (*compound scaling*) untuk menyeimbangkan lebar, kedalaman, dan resolusi jaringan [18]. Arsitektur dasarnya dibangun di atas blok-blok Mobile Inverted Bottleneck Convolution (MBConv). Untuk menyesuaikan dengan tugas ini, lapisan klasifikasi terakhir dimodifikasi untuk menghasilkan output empat kelas.



Gambar 3. Arsitektur EfficientNet

2) *Vision Transformer (ViT-B16)*: Model ini menerapkan arsitektur Transformer yang awalnya populer di NLP ke dalam tugas pemrosesan gambar. ViT membagi citra menjadi patch berukuran 16×16 piksel yang kemudian diperlakukan sebagai urutan dan diproses menggunakan mekanisme self-attention [19]. Seperti EfficientNet-B0, lapisan klasifikasi akhirnya disesuaikan untuk klasifikasi empat kelas biji kopi.



Gambar 4. Arsitektur Vision Transformer (ViT)

E. Pelatihan Model

Proses pelatihan model diimplementasikan dengan konfigurasi *hyperparameter* yang spesifik untuk memastikan reproduktifitas. Model dilatih menggunakan ukuran *batch* sebesar 32 selama maksimal 30 *epoch* dengan mekanisme *early stopping* yang memiliki *patience* 5 untuk mencegah *overfitting*. Kami menggunakan fungsi *loss* Cross-Entropy Loss dan *optimizer* Adam [20] dengan *learning rate* awal sebesar 1×10^{-4} . Selain itu, diterapkan *learning rate scheduler* StepLR yang mengurangi *learning rate* sebesar 50% (*gamma* 0.5) setiap 5 *epoch* untuk membantu model

mencapai konvergensi yang lebih stabil [21]. Model dengan akurasi validasi tertinggi selama pelatihan disimpan sebagai model final untuk dievaluasi. Seluruh layer pada kedua model dibuka (full fine-tuning). Tidak ada layer yang dibekukan untuk memastikan model dapat beradaptasi sepenuhnya dengan domain citra biji kopi. Hal ini menghasilkan jumlah parameter trainable sebesar 5.3 juta (EfficientNet-B0) dan 86 juta (ViT-B16).

F. Evaluasi Model

Kinerja model dievaluasi pada test set menggunakan metrik akurasi, presisi, recall, F1-score, serta confusion matrix. Metrik-metrik ini memberikan gambaran lengkap tentang kemampuan model dalam mengklasifikasikan setiap jenis biji kopi.

III. HASIL DAN PEMBAHASAN

A. Analisis Kinerja Model

1) Pada *test set*, *EfficientNet-B0* mencapai akurasi keseluruhan sebesar 87%, dengan F1-score rata-rata 0.87. Model ini menunjukkan presisi tertinggi pada kelas 'Longberry' (0.93) dan recall tertinggi pada kelas 'Premium' (0.91), mengindikasikan kinerja yang seimbang di berbagai kelas. Pelatihan berhenti pada epoch ke-23 akibat mekanisme *early stopping*, menandakan model telah mencapai konvergensi optimal. Rincian metrik per kelas disajikan pada Tabel II.

TABEL II
HASIL EVALUASI EFFICIENTNET-B0

Kelas	Presisi	Recall	F1-Score	Support
Defect	0.91	0.79	0.84	400
Longberry	0.93	0.89	0.91	400
Peaberry	0.86	0.90	0.88	400
Premium	0.80	0.91	0.85	400
Macro Avg	0.87	0.87	0.87	1600
Weighted Avg	0.87	0.87	0.87	1600

2) Pada *test set*, *Vision Transformer (ViT-B16)* mencapai akurasi keseluruhan sebesar 85% dengan F1-score rata-rata 0.85. ViT-B16 menunjukkan presisi yang sangat superior pada kelas 'Defect' dan 'Longberry' (keduanya 0.96). Namun, pelatihan berhenti lebih cepat pada epoch ke-12, yang mungkin mengindikasikan model ini lebih cepat konvergen atau lebih rentan terhadap overfitting dengan konfigurasi saat ini. Rincian metrik per kelas disajikan pada Tabel III.

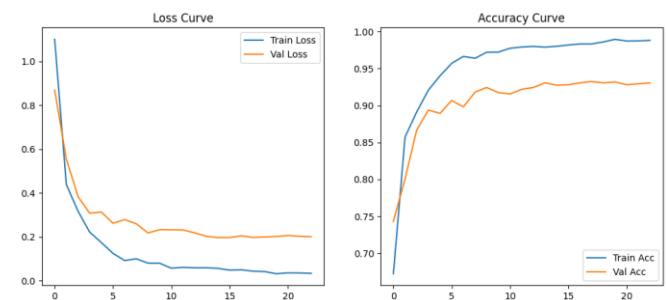
TABEL III
HASIL EVALUASI ViT-B16

Kelas	Presisi	Recall	F1-Score	Support
Defect	0.96	0.79	0.85	400
Longberry	0.96	0.89	0.88	400
Peaberry	0.87	0.90	0.87	400
Premium	0.71	0.91	0.81	400

Macro Avg	0.87	0.85	0.85	1600
Weighted Avg	0.87	0.85	0.85	1600

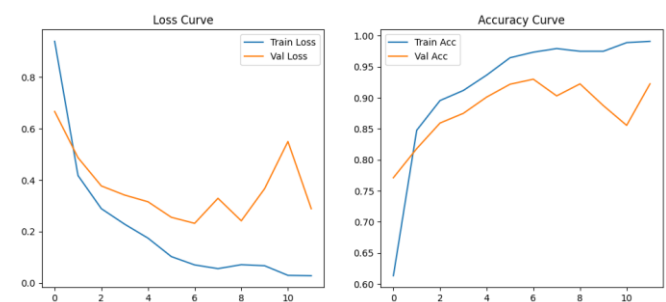
B. Analisis Kurva Pelatihan

Grafik pada Gambar 5 menunjukkan *training loss* *EfficientNet-B0* yang terus menurun dengan mantap, sementara *validation loss*-nya stabil di sekitar nilai 0.2 setelah epoch ke-10. Kondisi ini, bersamaan dengan *validation accuracy* yang mencapai dan stabil ≈ 0.93 , menandakan kemampuan generalisasi model yang baik dan tidak terjadi *overfitting* yang signifikan. Pelatihan berhenti pada epoch ke-23 karena mekanisme *early stopping* (*patience* 5) mendeteksi bahwa *validation loss* telah mencapai konvergensi optimal.



Gambar 5. Grafik Loss Curve dan Accuracy Curve EfficientNet-B0

Sebaliknya, pada Gambar 6, kinerja *Vision Transformer (ViT-B16)* menunjukkan pola *overfitting* klasik. Kurva *Training Loss* terus menurun mendekati nol, sementara *Validation Loss* mulai menunjukkan divergensi yang jelas setelah epoch ke-4 dan meningkat tajam (mencapai ≈ 0.55 di epoch terakhir). Fenomena ini diperkuat oleh *Validation Accuracy* yang mencapai puncaknya di sekitar epoch 6 (≈ 0.93) dan kemudian menurun drastis pada epoch-epoch berikutnya. *Overfitting* ini yang menyebabkan pelatihan berhenti lebih cepat pada epoch ke-12, mengindikasikan *ViT-B16* lebih rentan terhadap ketidakstabilan dan kegagalan generalisasi pada dataset berukuran sedang, yang memvalidasi efektivitas *early stopping* dalam mencegah kerusakan lebih lanjut.

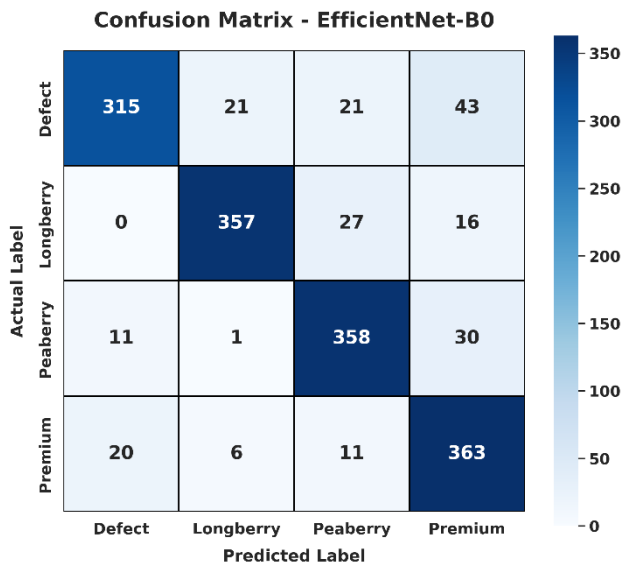


Gambar 6. Grafik Loss Curve dan Accuracy Curve ViT-B16

C. Analisis Confusion Matrix

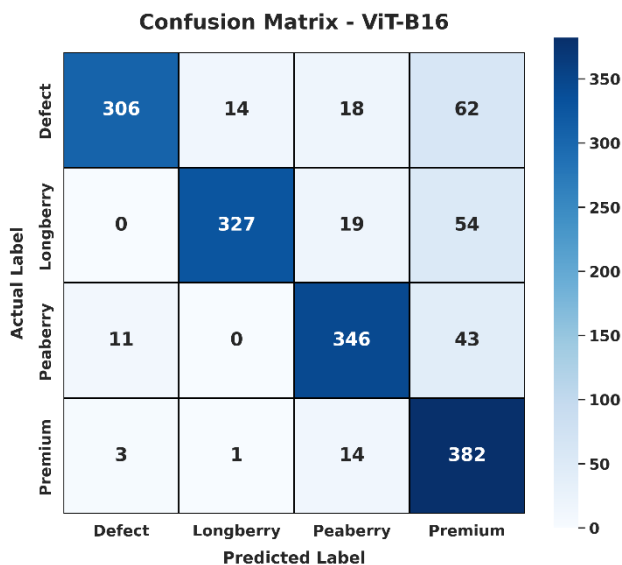
Confusion matrix *EfficientNet-B0* (Gambar 7) menunjukkan tantangan utama pada *recall* kelas 'Defect'

(0.79), di mana 85 dari 400 biji cacat salah diklasifikasikan. Hal ini kemungkinan disebabkan oleh keragaman visual pada biji kopi cacat yang membuat model kesulitan menemukan pola yang konsisten. Selain itu, presisi rendah pada kelas 'Premium' (0.80) menandakan adanya biji non-premium yang lolos sebagai premium, sebuah risiko bagi kontrol kualitas.



Gambar 7. Confusion Matrix EfficientNet-B0

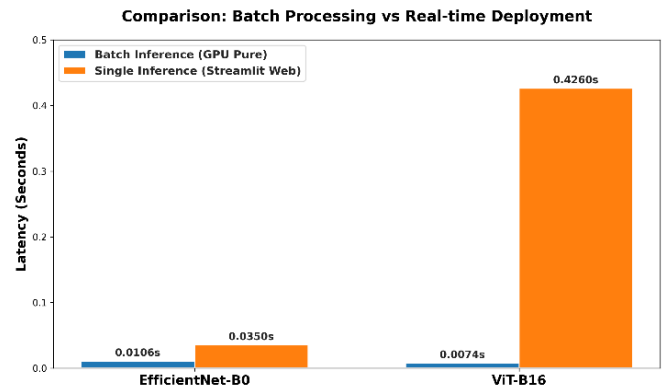
Di sisi lain, *confusion matrix* ViT-B16 (Gambar 8) menunjukkan *recall* yang sangat tinggi pada kelas 'Premium' (0.95), tetapi dengan presisi terendah (0.71). Ini mengindikasikan bahwa ViT-B16 cenderung lebih 'optimis' dalam memprediksi kelas 'Premium', sehingga lebih banyak menghasilkan *false positive* untuk kelas tersebut, yang juga berisiko bagi kualitas produk.



Gambar 8. Confusion Matrix Vision Transformer (ViT-B16)

D. Analisis Efisiensi (Waktu Inferensi)

Analisis efisiensi menunjukkan dua skenario yang kontras. Pengukuran waktu inferensi dilakukan pada perangkat keras GPU NVIDIA Tesla T4 dengan *batch size* 32 dan *input* 224x224. Untuk pemrosesan *batch*, ViT-B16 lebih unggul dengan latensi rata-rata 0.0074 detik per batch dibandingkan EfficientNet-B0 (0.0106 detik). Keunggulan ViT-B16 ini disebabkan kemampuan arsitektur *Transformer* dalam memanfaatkan paralelisasi perangkat keras secara efektif, mengindikasikan *trade-off* kecepatan *batch* yang sebanding dengan kompleksitas modelnya (86 juta parameter).



Gambar 9. Perbandingan Latensi Inferensi Model Batch dan Single

Namun, untuk skenario aplikasi *real-time* yang mengklasifikasikan satu gambar, di mana pengujian dilakukan dengan *deployment* pada *web interface* menggunakan *Streamlit*, *EfficientNet-B0* jauh lebih cepat dengan waktu 0.035 detik, dibandingkan *ViT-B16* yang membutuhkan 0.426 detik. Perbedaan drastis ini disebabkan oleh arsitektur *CNN EfficientNet-B0* yang lebih ringan (5.3 juta parameter) dan operasi *konvolusi* yang sangat teroptimasi untuk inferensi tunggal (*single image inference*), membuatnya ideal untuk aplikasi *real-time* berbasis web seperti yang diuji dengan *Streamlit*. Hal ini berbeda dengan *ViT-B16* yang memiliki parameter jauh lebih besar (86 juta) dan *overhead* komputasi dari mekanisme *self-attention* untuk satu gambar.

Perbedaan performa batch dan single inference penting dalam pemilihan arsitektur. Sistem quality control berbasis edge biasanya melakukan inferensi satu per satu (realtime per citra), sehingga model dengan latensi tunggal rendah lebih layak digunakan. Sebaliknya, sistem cloud yang memproses citra dalam jumlah besar secara paralel akan mendapatkan manfaat dari arsitektur *Transformer*.

IV. KESIMPULAN

Penelitian ini berhasil membandingkan secara langsung performa arsitektur ViT-B16 dan EfficientNet-B0 untuk klasifikasi biji kopi hijau. Hasil menunjukkan bahwa EfficientNet-B0 sedikit lebih unggul dalam hal akurasi 87% dibandingkan ViT-B16 85% pada dataset uji. Dari segi

efisiensi, terdapat pertukaran (trade-off): ViT-B16 lebih cepat untuk pemrosesan data dalam batch, sedangkan EfficientNet-B0 secara signifikan lebih superior untuk inferensi gambar tunggal yang bersifat real-time, seperti yang diverifikasi melalui deployment web interface menggunakan Streamlit.

Kedua model terbukti layak untuk sistem klasifikasi mutu biji kopi berbasis *edge computing*, di mana pemilihan model sangat bergantung pada prioritas aplikasi: akurasi tertinggi dan kecepatan real-time dengan EfficientNet-B0, atau kecepatan pemrosesan *batch* dengan ViT-B16. Keterbatasan utama penelitian ini adalah ViT-B16 yang menunjukkan kerentanan *overfitting* yang lebih tinggi, dan klaim *edge computing* yang belum didukung oleh evaluasi penggunaan memori, konsumsi daya, atau kompatibilitas pada perangkat *edge* nyata.

Untuk penelitian selanjutnya, disarankan untuk melakukan optimasi *hyperparameter* yang lebih ekstensif, terutama untuk ViT-B16 (misalnya, dengan menggunakan *patch augmentation* atau teknik *regularization* tambahan) untuk lebih mengeksplorasi potensi model ini dan mengatasi kerentanan *overfitting*. Selain itu, eksplorasi teknik kompresi model seperti *pruning* dan *quantization* dapat diinvestigasi untuk lebih mengoptimalkan implementasi pada perangkat *edge* dengan sumber daya terbatas. Terakhir, mengingat dataset diambil di satu lokasi geografis, pengujian pada biji kopi dari wilayah lain perlu dilakukan untuk menganalisis dan mengurangi *model bias*, sehingga meningkatkan kemampuan generalisasi model secara global.

DAFTAR PUSTAKA

- [1] J. Liu, "Coffee Bean High Accuracy Classification with eXplainable Artificial Intelligence," pp. 490–497, doi: 10.1145/3659211.3659296.
- [2] A. Widyasari, "Pengaruh Ukuran Biji Kopi Robusta pada Kualitas Citarasa Kopi (The Effect of Robusta Coffee Bean Size on Coffee Taste Quality)," vol. 11, no. 1, pp. 1–14, 2023.
- [3] K. Przybył *et al.*, "Application of Machine Learning to Assess the Quality of Food Products—Case Study: Coffee Bean," *Appl. Sci.*, vol. 13, no. 19, 2023, doi: 10.3390/app131910786.
- [4] K. C. Febrianti *et al.*, "Perbandingan Ekspor Kopi Indonesia Pada Tahun 2021 Dan 2022," *J. Ekon. dan Kewirausahaan West Sci.*, vol. 2, no. 03, pp. 374–384, 2024, doi: 10.58812/jekws.v2i03.1422.
- [5] N. M. Sinta, Z. Alamsyah, and E. Elwamendri, "Analisis Daya Saing Ekspor Kopi Indonesia Dan Vietnam Di Pasar Asean," *J. Ilm. Sosio-Ekonomika Bisnis*, vol. 20, no. 1, p. 3, 2018, doi: 10.22437/jiseb.v20i1.5028.
- [6] M. Rahimi *et al.*, "Statistik Kopi Indonesia 2023," vol. 8, 2023.
- [7] Bibit Bakoh dan Ratri Wibawanti, "Peningkatan Kapabilitas Penanganan OPT Tanaman Kopi," 2023.
- [8] I. V. C. Motta, N. Vuillerme, H. H. Pham, and F. A. P. de Figueiredo, "Machine learning techniques for coffee classification: a comprehensive review of scientific research," *Artif. Intell. Rev.*, vol. 58, no. 1, 2025, doi: 10.1007/s10462-024-11004-w.
- [9] A. Korkmaz, T. Talan, G. I. Science, and S. Kosunalp, "Comparison of deep learning models in automatic classification of coffee bean species," no. April, 2025, doi: 10.7717/peerj-cs.2759.
- [10] S. Arwathananukul, D. Xu, P. Charoenkwan, S. Aung, and R. Saengrayap, "Smart Agricultural Technology Implementing a deep learning model for defect classification in Thai Arabica green coffee beans," *Smart Agric. Technol.*, vol. 9, no. August, p. 100680, 2024, doi: 10.1016/j.atech.2024.100680.
- [11] B. Nair, B. Jayakumari, A. Nanda, and K. Mambilamthoda, "Coffee bean graded based on deep net models," vol. 14, no. 3, pp. 3084–3093, 2024, doi: 10.11591/ijece.v14i3.pp3084-3093.
- [12] R. Gonc *et al.*, "Enhancing green coffee quality assessment through deep learning".
- [13] M. A. Leonardi and A. Y. Chandra, "Analisis Perbandingan CNN dan Vision Transformer untuk Klasifikasi Biji Kopi Hasil Sangrai," vol. 8, pp. 1398–1407, 2024, doi: 10.30865/mib.v8i3.7732.
- [14] M. N. Izza and G. P. Kusuma, "Image Classification of Green Arabica Coffee Using Transformer-Based Architecture," *Int. J. Eng. Trends Technol.*, vol. 72, no. 6, pp. 304–314, 2024, doi: 10.14445/22315381/IJETT-V72I6P128.
- [15] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11141 LNCS, pp. 270–279, 2018, doi: 10.1007/978-3-030-01424-7_27.
- [16] K. Antwi, K. E. Bennin, D. K. Pobi Asiedu, and B. Tekinerdogan, "On the application of image augmentation for plant disease detection: A systematic literature review," *Smart Agric. Technol.*, vol. 9, no. October, p. 100590, 2024, doi: 10.1016/j.atech.2024.100590.
- [17] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
- [18] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.
- [19] Z. Fu, "Vision Transformer: Vit and its Derivatives Swin Transformer: Hierarchical Vision Transformer using," pp. 1–10, 2022.
- [20] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [21] A. Defazio, A. Cutkosky, H. Mehta, and K. Mishchenko, "Optimal Linear Decay Learning Rate Schedules and Further Refinements," 2023, [Online]. Available: <http://arxiv.org/abs/2310.07831>