

Waste Image Classification Using Fine-Tuned MobileNetV2 under Imbalanced Data Conditions

Rahmadini Cahya Demora^{1*}, Muljono^{2*}

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

111202214464@mhs.dinus.ac.id¹, muljono@dsn.dinus.ac.id²

Article Info

Article history:

Received 2026-01-08

Revised 2026-02-16

Accepted 2026-02-27

Keyword:

*Deep Learning,
Fine-Tuning,
Imbalanced Data,
MobileNetV2,
Waste Classification.*

ABSTRACT

Ineffective waste management is a critical global environmental issue where manual sorting is often inefficient and prone to human error. Deep Learning technology, specifically Convolutional Neural Networks (CNN), offers automated solutions to improve classification performance. This study aims to optimize waste image classification by evaluating the effectiveness of the Fine-Tuned MobileNetV2 architecture, specifically addressing the challenges of imbalanced data distribution where recyclable items significantly outnumber residuals. Experiments compared six training scenarios including Basic CNN, CNN with Class Weight, and Transfer Learning using VGG16 and MobileNetV2 with frozen and fine-tuning strategies. Evaluation metrics included Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. Results indicated that the Basic CNN model struggled with the minority class, yielding a low F1-Score of 0.49. Conversely, the optimized MobileNetV2 Fine-Tuned model achieved superior performance, recording a testing accuracy of 96.49% and an F1-Score of 0.92. It is concluded that the fine-tuning strategy on MobileNetV2 is the most effective approach to mitigate data imbalance, providing an optimal balance between high sensitivity toward the minority class and computational efficiency for real-time implementation.



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

I. PENDAHULUAN

Peningkatan volume sampah yang tidak terkendali telah menjadi isu global mendesak yang berdampak serius terhadap keberlanjutan lingkungan dan kesehatan masyarakat. Di Indonesia, tantangan ini diperparah oleh kurangnya kesadaran pemilahan di sumber dan infrastruktur pengelolaan yang belum memadai, sehingga menyebabkan penumpukan sampah yang masif di Tempat Pembuangan Akhir (TPA) [1]. Metode pemilahan sampah secara manual yang masih dominan diterapkan dinilai tidak efisien, memakan waktu lama, dan berisiko tinggi terhadap kesehatan pekerja, sehingga mendorong kebutuhan mendesak akan sistem otomatisasi berbasis teknologi [2]. Penerapan teknologi *Deep Learning*, khususnya algoritma *Convolutional Neural Network* (CNN), menawarkan solusi yang menjanjikan untuk mengklasifikasikan citra sampah secara akurat dan cepat, menggantikan keterbatasan pengamatan manusia dalam membedakan jenis material limbah yang kompleks [3].

Perkembangan arsitektur CNN telah membawa kemajuan signifikan dalam bidang *computer vision*, memungkinkan ekstraksi fitur yang lebih mendalam dibandingkan metode tradisional. Li et al. [4] dalam tinjauannya menjelaskan bahwa CNN memiliki keunggulan dalam weight sharing dan local connectivity yang membuatnya sangat efektif untuk pengenalan pola visual. Dalam konteks klasifikasi sampah, penggunaan model *Transfer Learning* (TL) yang memanfaatkan bobot dari model pra-latih (*pre-trained*) seringkali menjadi pilihan utama untuk mengatasi keterbatasan dataset dan sumber daya komputasi. Penelitian ini urgen dilakukan untuk mengevaluasi efektivitas arsitektur CNN modern dalam menangani variabilitas visual sampah rumah tangga yang seringkali ambigu dan sulit dibedakan [5].

Berbagai penelitian terdahulu telah mengeksplorasi penerapan algoritma pembelajaran mesin untuk klasifikasi sampah dengan berbagai pendekatan. Yasin dan Koklu [6] melakukan analisis komparatif algoritma *Machine Learning* (*Support Vector Machine, K-Nearest Neighbor, Decision*

Tree) menggunakan ekstraksi fitur dari InceptionV3 pada 24.705 citra, di mana *Support Vector Machine* (SVM) mencapai akurasi tertinggi sebesar 96,3%, namun mengalami penurunan performa saat dilakukan reduksi fitur. Sementara itu, Poudel dan Poudyal [7] membandingkan tujuh model CNN pre-trained pada dataset TrashNet yang dimodifikasi, menemukan bahwa DenseNet201 memberikan akurasi validasi terbaik sebesar 95,05% meskipun terdapat variasi kinerja antar kelas. Di sisi lain, Soni et al. [8] membandingkan MobileNet dan ResNet50 pada dataset yang lebih kecil, menyimpulkan bahwa MobileNet lebih unggul dalam efisiensi dengan akurasi 80%, sedangkan Muslihati et al. [9] menunjukkan bahwa model *Transfer Learning Xception* mampu mencapai akurasi 96,43%, jauh mengungguli model CNN dasar yang dibangun dari awal.

Meskipun banyak pendekatan telah diusulkan, terdapat permasalahan mendasar yang sering menghambat implementasi di dunia nyata, yaitu ketidakseimbangan data (*imbalanced dataset*) di mana jumlah sampel sampah daur ulang jauh lebih banyak dibandingkan residu. Kondisi ini menyebabkan model cenderung bias ke kelas mayoritas dan gagal mengenali kelas minoritas, sebuah fenomena yang juga diamati oleh Sulistyowati et al. [10] pada kasus deteksi penyakit tanaman. Selain itu, pelatihan model CNN dasar seringkali kurang optimal dan menghasilkan akurasi yang rendah jika tidak didukung oleh teknik penanganan data yang tepat. Oleh karena itu, diperlukan strategi optimasi melalui pemanfaatan arsitektur yang ringan namun kuat seperti MobileNetV2, yang dikombinasikan dengan teknik *fine-tuning* dan *Class Weight* untuk mengatasi masalah tersebut.

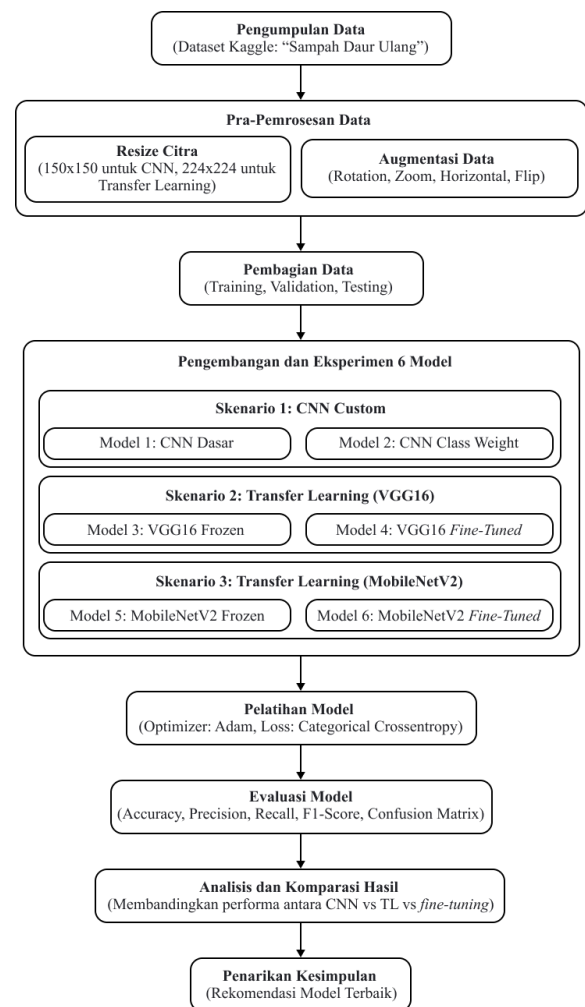
Penelitian ini bertujuan untuk mengoptimalkan klasifikasi citra sampah dengan mengevaluasi efektivitas arsitektur MobileNetV2 *Fine-Tuned* di bawah kondisi data yang tidak seimbang. Sebagai pembanding, penelitian ini juga menguji skenario CNN Dasar, CNN dengan *Class Weight*, serta *Transfer Learning* berbasis VGG16. Kontribusi utama dari penelitian ini adalah mengidentifikasi strategi pelatihan terbaik untuk mencapai keseimbangan antara precision dan recall pada kelas minoritas (residu). Hasil penelitian ini diharapkan dapat memberikan rekomendasi arsitektur yang paling efisien dan akurat untuk pengembangan sistem pemilahan sampah otomatis yang robust dan siap diimplementasikan pada perangkat dengan sumber daya terbatas.

II. METODE

Penelitian ini menerapkan pendekatan kuantitatif dengan desain eksperimental untuk menguji performa algoritma klasifikasi citra secara empiris. Fokus utama metode yang dirancang adalah menjawab tantangan ketidakseimbangan data melalui perbandingan enam skenario model yaitu CNN Dasar dan CNN dengan *Class Weight* sebagai acuan (*baseline*), serta *Transfer Learning* berbasis VGG16 dan MobileNetV2 dengan strategi *Frozen* dan *fine-tuning*. Pemilihan arsitektur ini didasarkan pada tinjauan Li et al. [4] yang menyoroti kemampuan superior CNN dalam

mengekstrak fitur spasial hierarkis, serta temuan Poudel dan Poudyal [7] mengenai efektivitas adaptasi bobot pada lapisan atas model *pre-trained*.

Pelaksanaan penelitian mengikuti alur kerja sistematis yang dimulai dari pengumpulan dataset sekunder, diikuti oleh pra-pemrosesan data yang meliputi penyeragaman dimensi citra dan augmentasi untuk memperkaya variasi visual. Tahap inti melibatkan pengembangan dan pelatihan keenam model menggunakan parameter yang terkontrol ketat, merujuk pada kaidah optimasi yang disarankan oleh Khan et al. [11]. Rangkaian penelitian ini diakhiri dengan evaluasi performa menggunakan metrik statistik standar (*accuracy*, *precision*, *recall*, dan *F1-score*) untuk menentukan arsitektur terbaik. Seluruh tahapan prosedur penelitian ini pada Gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil dataset dari platform open access kaggle yang berisi total 7.014 citra sampah. Sebagaimana dicatat dalam penelitian Oktayessofa et al. [12], pemanfaatan dataset publik memungkinkan evaluasi model yang objektif pada data dengan variabilitas tinggi.

Secara spesifik, dataset ini mencakup enam kategori awal material limbah, yaitu kaca, kardus, kertas, logam, plastik, dan residu. Namun, untuk keperluan klasifikasi biner yang menjadi fokus penelitian ini, keenam kelas tersebut dikategorikan ulang (*re-categorized*) menjadi dua kelas utama. Kategori Daur Ulang (*Recyclable*) merupakan gabungan dari lima jenis material (kaca, kardus, kertas, logam, plastik), sedangkan kategori Tidak Daur Ulang (*Non-Recyclable*) direpresentasikan oleh kelas residu. Visualisasi sampel citra yang memperlihatkan keragaman objek pada kedua kategori tersebut dapat dilihat pada Gambar 2.



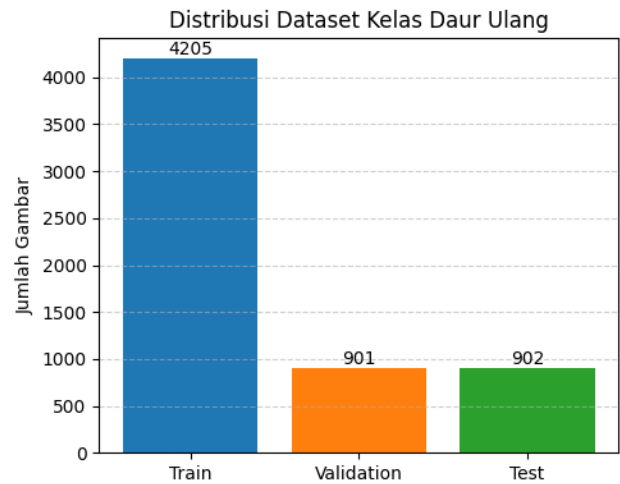
Gambar 2. Contoh Sampel Citra pada Dataset. Baris atas menunjukkan kategori sampah daur ulang, sedangkan baris bawah menunjukkan kategori sampah tidak daur ulang (residu).

Hasil dari proses pengelompokan ulang tersebut menghasilkan distribusi data yang tidak seimbang (*imbalanced*). Kelas Daur Ulang mendominasi dengan total 6.008 citra, sementara kelas Tidak Daur Ulang hanya memiliki 1.006 citra, menciptakan rasio ketimpangan mendekati 6:1. Alih-alih melakukan *undersampling* atau *oversampling* di tahap awal, distribusi asli ini sengaja dipertahankan sebagai bagian dari skenario eksperimental. Pendekatan ini bertujuan untuk menguji ketahanan (*robustness*) model dalam mengatasi bias kelas mayoritas, sebuah tantangan metodologis yang krusial dalam pengembangan sistem klasifikasi sampah cerdas sebagaimana ditekankan oleh Poudel dan Poudyal [7] serta Lilhore et al. [13]. Rincian jumlah citra per kelas setelah dilakukan pembagian dataset (*splitting*) dengan rasio 70% data latih, 15% data validasi, dan 15% data uji disajikan pada Tabel I.

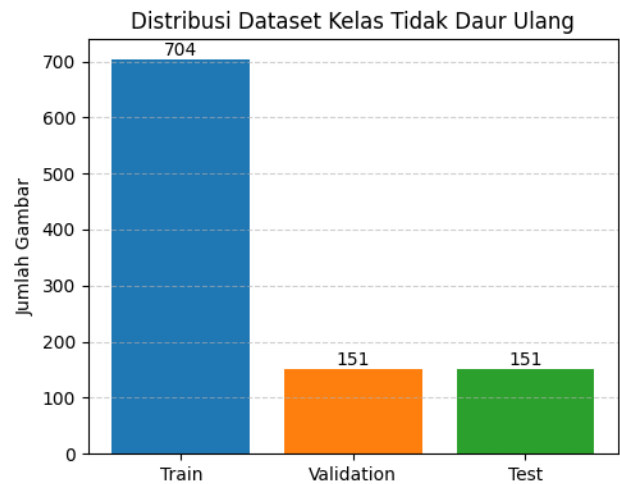
TABEL I
KOMPOSISI DISTRIBUSI CITRA PER KELAS

Kategori	Train (70%)	Validation (15%)	Test (15%)	Total
Daur Ulang	4205	901	902	6008
Tidak Daur Ulang	704	151	151	1006
Total	4909	1052	1053	7014

Visualisasi distribusi jumlah citra pada setiap subset data untuk kelas mayoritas (Daur Ulang) ditunjukkan pada Gambar 3, sementara distribusi untuk kelas minoritas (Residu) disajikan pada Gambar 4.



Gambar 3. Distribusi Dataset Kelas Daur Ulang



Gambar 4. Distribusi Dataset Kelas Tidak Daur Ulang

B. Pra-Pemrosesan Data

Tahap awal pra-pemrosesan dimulai dengan manajemen distribusi data melalui pembagian dataset menjadi tiga subset terpisah menggunakan rasio 70 : 15 : 15, yaitu data latih (*training set*) sebanyak 4909 citra, data validasi (*validation set*) sebanyak 1052 citra, dan data uji (*test set*) sebanyak 1053 citra. Pembagian ini dilakukan secara acak untuk memastikan keterwakilan setiap kelas, sekaligus mencegah terjadinya kebocoran informasi (*data leakage*) antar subset yang dapat membiaskan hasil evaluasi. Dalam penelitian ini, data uji dipisahkan sepenuhnya dan hanya digunakan satu kali pada tahap akhir evaluasi, sebuah prosedur standar yang ditekankan dalam metodologi Lilhore et al. [13] guna menjamin objektivitas pengukuran performa model pada data baru yang belum pernah dilihat sebelumnya selama proses pelatihan.

Setelah pembagian subset, seluruh citra melalui proses penyeragaman dimensi atau *resizing* untuk memenuhi

kebutuhan input arsitektur jaringan saraf. Terdapat perbedaan perlakuan dimensi berdasarkan jenis model yang digunakan. Untuk skenario model CNN yang dibangun dari awal (*scratch*), citra diubah ukurannya menjadi 150×150 piksel, dimensi yang dipilih untuk menyeimbangkan antara retensi detail visual dan efisiensi beban komputasi. Sementara itu, untuk skenario model berbasis *Transfer Learning* (VGG16 dan MobileNetV2), seluruh citra diubah menjadi dimensi 224×224 piksel. Langkah ini bersifat imperatif untuk menyesuaikan dengan spesifikasi input layer dari arsitektur yang telah dilatih sebelumnya (*pre-trained*) pada dataset ImageNet, sebagaimana diterapkan dalam studi klasifikasi material oleh Poudel dan Poudyal [7].

Guna meningkatkan ketahanan model (*robustness*) dan memitigasi risiko *overfitting* akibat ketidakseimbangan kelas, teknik augmentasi data diterapkan secara *real-time* selama proses pelatihan. Transformasi citra yang dilakukan meliputi rotasi acak, pergeseran fokus (*zooming*), dan pembalikan horizontal (*horizontal flip*). Sebagaimana dibuktikan dalam temuan eksperimental Oktayessofa et al. [12], strategi augmentasi ini sangat efektif untuk memperkaya variasi dataset secara sintesis tanpa perlu menambah sampel fisik. Hal ini memungkinkan model untuk mempelajari fitur invariant seperti bentuk dan tekstur sampah tanpa terpengaruh oleh orientasi atau posisi pengambilan gambar, sehingga meningkatkan kemampuan generalisasi model pada data pengujian.

C. Pengembangan dan Eksperimen Model

Pengembangan model dirancang ke dalam tiga skenario eksperimental utama. Pendekatan ini bertujuan untuk membandingkan secara komprehensif antara arsitektur yang dibangun dari awal (*scratch*) melawan arsitektur yang memanfaatkan bobot pra-latih (*Transfer Learning*), serta menguji efektivitas strategi penanganan ketidakseimbangan data.

Skenario pertama berfokus pada pengembangan model CNN Custom yang dibangun dari awal (*scratch*). Arsitektur ini dirancang menggunakan tiga blok konvolusi yang diikuti oleh lapisan *Max Pooling* untuk mereduksi dimensi fitur spasial secara bertahap. Model menerima input citra berukuran 150×150 piksel dan diakhiri dengan lapisan Dense 128 neuron serta *Dropout* 0.5 untuk mencegah *overfitting*. Dalam skenario ini dua varian model diuji yaitu Model 1 (CNN Dasar) yang dilatih secara konvensional, dan Model 2 (CNN *Class Weight*) yang menerapkan pembobotan kelas pada fungsi *loss*. Penerapan *class weight* ini bertujuan untuk memberikan penalti yang lebih besar pada kesalahan klasifikasi kelas minoritas (residu), sebuah teknik yang disarankan dalam studi Lilhore et al. [13] untuk memitigasi bias pada dataset yang tidak seimbang.

Skenario kedua menerapkan metode *Transfer Learning* menggunakan arsitektur VGG16. Arsitektur ini dipilih karena kedalaman dan kemampuannya yang teruji dalam mengekstrak fitur visual kompleks, sebagaimana diimplementasikan dalam penelitian Dacipta dan Putra [14].

Model ini menggunakan input standar 224×224 piksel dengan bobot awal yang berasal dari dataset ImageNet. Eksperimen dibagi menjadi dua pendekatan yaitu Model 3 (VGG16 *Frozen*), di mana seluruh lapisan konvolusi dibekukan (*freeze*) dan hanya lapisan klasifikasi tambahan (*Fully Connected Layer*) yang dilatih serta Model 4 (VGG16 *Fine-Tuned*), di mana blok konvolusi kelima (Block 5) dibuka (*unfreeze*) dan dilatih kembali dengan laju pembelajaran (*learning rate*) yang sangat kecil (1×10^{-5}) untuk mengadaptasi fitur spesifik dari citra sampah.

Skenario ketiga mengeksplorasi arsitektur MobileNetV2, yang dikenal dengan efisiensi komputasinya berkat penggunaan inverted residual blocks. Pemilihan arsitektur ini didukung oleh temuan Oktayessofa et al. [12] yang membuktikan bahwa MobileNetV2 mampu mencapai akurasi tinggi (98,47%) dalam klasifikasi sampah dengan beban komputasi yang ringan. Sama halnya dengan skenario sebelumnya, eksperimen ini menguji dua varian yaitu Model 5 (MobileNetV2 *Frozen*) yang memperlakukan model sebagai ekstraktor fitur statis, dan Model 6 (MobileNetV2 *Fine-Tuned*). Pada varian *Fine-Tuned*, strategi yang diterapkan lebih agresif dengan membuka 30 lapisan teratas (*top layers*) untuk dilatih kembali. Lapisan yang dibuka ini mencakup seluruh bagian *head classification* serta beberapa Inverted residual blocks terakhir pada arsitektur MobileNetV2. Pendekatan ini bertujuan untuk menyempurnakan representasi fitur agar lebih sensitif terhadap karakteristik unik dan detail visual dari dataset sampah, khususnya pada kelas minoritas.

D. Pelatihan Model

Proses pelatihan seluruh model dilaksanakan menggunakan algoritma optimasi Adam (*Adaptive Moment Estimation*). Pemilihan optimizer ini didasarkan pada efisiensi komputasinya dan kemampuannya dalam mengadaptasi laju pembelajaran (*learning rate*) secara dinamis, yang terbukti superior untuk konvergensi arsitektur *Deep Learning* sebagaimana divalidasi dalam studi Khan et al. [11] dan Poudel & Poudyal [7]. Untuk fungsi kerugian (*loss function*), penelitian ini menerapkan *Categorical Crossentropy* mengingat sifat permasalahan klasifikasi yang bersifat *multi-class* (pada tahap awal) atau biner, sebuah pendekatan standar yang juga diterapkan oleh Ouedraogo et al. [15] untuk memastikan pemisahan kategori yang optimal.

Strategi pelatihan dikonfigurasi secara spesifik menyesuaikan karakteristik setiap skenario model. Pada model CNN Dasar dan CNN *Class Weight*, pelatihan dilakukan selama 30 epoch dengan laju pembelajaran awal sebesar 0,001. Sementara itu, untuk model berbasis *Transfer Learning*, diterapkan strategi pelatihan bertahap yang diadopsi dari metodologi Oktayessofa et al. [12]. Pada fase *Frozen*, model dilatih selama 20 epoch dengan laju pembelajaran 0,0005. Khusus untuk varian *Fine-Tuned*, proses dilanjutkan dengan fase kedua selama 10 epoch tambahan, pada fase *fine-tuning* laju pembelajaran diturunkan secara drastis menjadi (1×10^{-5}). Penurunan laju

pembelajaran ini sangat krusial untuk mencegah kerusakan pada bobot pra-latih (*pre-trained weights*) saat model beradaptasi dengan fitur spesifik dataset sampah, sebagaimana disarankan dalam eksperimen Dacipta & Putra [14] dan Poudel & Poudyal [7] guna mencapai stabilitas akurasi. Rincian parameter pelatihan atau *hyperparameters* yang digunakan dalam eksperimen ini diringkas pada Tabel II.

TABEL II
HYPERPARAMETER PELATIHAN MODEL

Parameter	Fase Frozen	Fase Fine-Tuning
Optimizer	Adam	Adam
Learning Rate	(5×10^{-4})	(1×10^{-5})
Epoch	20	10
Loss Function	Categorical Crossentropy	Categorical Crossentropy
Batch Size	32	32

E. Evaluasi Model

Evaluasi kinerja model dilakukan secara ketat menggunakan data uji (*test set*) yang sepenuhnya terpisah dari proses pelatihan dan validasi. Prosedur pemisahan data uji ini merupakan standar metodologis yang ditekankan oleh Lilhore et al. [13] serta Li & Chen [16] untuk menjamin objektivitas dan mencegah bias evaluasi. Mengingat karakteristik dataset yang sangat tidak seimbang (*imbalanced*), penggunaan metrik akurasi semata dinilai tidak memadai. Oleh karena itu, penelitian ini mengacu pada kerangka evaluasi Tan Hor Yan et al. [17] dan Oktayessofa et al. [12], yang memprioritaskan analisis komprehensif menggunakan empat parameter utama yang dihitung berdasarkan komponen Confusion Matrix, dengan definisi operasional sebagai berikut:

- 1) *True Positive (TP)*: Jumlah citra sampah daur ulang yang berhasil diklasifikasikan dengan benar sebagai kelas daur ulang. Nilai ini menjadi indikator utama keberhasilan model dalam mengenali target fitur yang relevan [12].
- 2) *True Negative (TN)*: Jumlah citra sampah residu yang berhasil diklasifikasikan dengan benar sebagai tidak daur ulang. Metrik ini krusial untuk memastikan sistem dapat memilah sampah yang tidak bernilai secara efektif [6].
- 3) *False Positive (FP)*: Jumlah citra sampah residu yang salah diprediksi oleh model sebagai sampah daur ulang. Kesalahan ini disebut juga sebagai *Type I Error* yang dapat menyebabkan kontaminasi pada proses daur ulang [4].
- 4) *False Negative (FN)*: Jumlah citra sampah daur ulang yang salah diprediksi oleh model sebagai sampah residu (terlewat). Kesalahan ini disebut *Type II Error* dan sangat dihindari karena menyebabkan hilangnya potensi ekonomi material daur ulang [6], [12].

Berdasarkan komponen di atas, kinerja model diukur menggunakan persamaan matematis berikut. *Precision* mengukur tingkat ketepatan prediksi positif model (1),

sedangkan *Recall* mengukur sensitivitas model dalam mendeteksi seluruh sampel positif (2).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Untuk mendapatkan gambaran performa yang utuh pada data yang tidak seimbang, digunakan *F1-Score* (3) sebagai rata-rata harmonis antara *precision* dan *recall* [17], serta *Accuracy* (4) untuk melihat rasio prediksi benar secara global terhadap keseluruhan dataset [6], [12].

$$F1 \text{ score} = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (3)$$

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

Analisis hasil kemudian diperdalam menggunakan visualisasi *Confusion Matrix* untuk memetakan distribusi prediksi tersebut, memungkinkan identifikasi spesifik terhadap kelemahan model dalam mengenali kelas minoritas (Residu) dan mendeteksi kecenderungan bias terhadap kelas mayoritas.

III. HASIL DAN PEMBAHASAN

A. Hasil Pelatihan Model

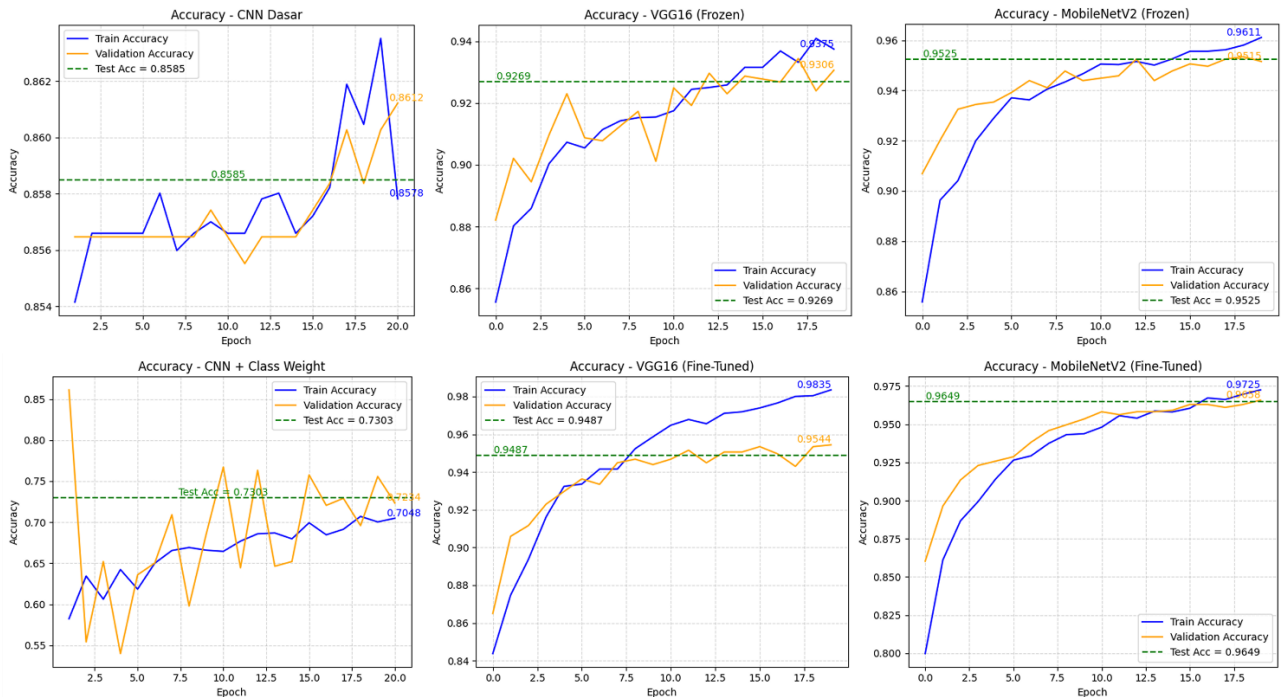
Proses pelatihan keenam model dievaluasi dengan memantau pergerakan grafik akurasi pada data latih (*training accuracy*) dan data validasi (*validation accuracy*) di setiap *epoch*. Pemantauan ini bertujuan untuk mengidentifikasi perilaku model, apakah mengalami *overfitting*, *underfitting*, atau mencapai konvergensi yang baik (*good fit*). Berdasarkan hasil eksperimen yang disajikan pada Gambar 5, terlihat perbedaan karakteristik proses pembelajaran yang signifikan antara model yang dibangun dari awal (*scratch*) dengan model berbasis *transfer learning*.

Pada model CNN Dasar, peningkatan akurasi menunjukkan tren yang stagnan dengan akurasi akhir pada data validasi sebesar 86,12%. Meskipun tidak terjadi *overfitting* yang berat, rendahnya angka akurasi ini mengindikasikan bahwa arsitektur CNN sederhana mengalami *underfitting* dalam menangkap variabilitas visual dataset sampah. Sementara itu, penerapan CNN *Class Weight* justru menghasilkan fluktuasi yang sangat tidak stabil pada kurva validasi, dengan akurasi akhir yang tertahan di angka 72,34%. Hal ini membuktikan bahwa pemaksaan bobot pada kelas minoritas tanpa dukungan arsitektur yang kuat justru merusak kemampuan generalisasi model.

Sebaliknya, skenario *transfer learning* pada baris kedua dan kolom kanan baris pertama menunjukkan lonjakan performa yang drastis dan stabil. Model MobileNetV2 *Frozen* mampu mencapai akurasi validasi 95,15% hanya dengan

memanfaatkan bobot pra-latih *ImageNet*. Peningkatan paling optimal tercatat pada model *MobileNetV2 Fine-Tuned* yang berhasil mendorong akurasi validasi hingga 96,58%. Kesenjangan (*gap*) yang sangat kecil antara kurva pelatihan dan validasi (kurang dari 1%) pada model ini menandakan

terjadinya konvergensi yang sangat baik (*good fit*) dan stabilitas model yang tinggi dibandingkan varian *VGG16 Fine-Tuned* yang mencapai 95,44%.



Gambar 5. Grafik Perbandingan Akurasi *Training* (Garis Biru) dan *Validation* (Garis Oranye) pada Keenam Skenario Eksperimen. Grafik pada kolom pertama memperlihatkan ketidakstabilan pada model *CNN Dasar* dan *CNN Class Weight*, sementara grafik pada kolom kedua dan ketiga menunjukkan hasil yang jauh lebih stabil pada model *Transfer Learning*, khususnya *MobileNetV2 Fine-Tuned*.

B. Evaluasi Kinerja Model

Setelah proses pelatihan selesai, selanjutnya adalah menguji seberapa pintar model dalam menebak data baru yang belum pernah dilihat sebelumnya. Pengujian ini menggunakan 1.053 citra data uji. Hasil kinerja dari keenam model tersebut dirangkum dalam Tabel III, yang mencakup nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

TABEL III
PERBANDINGAN HASIL EVALUASI MODEL PADA DATA UJI

Model	Accuracy	Precision	Recall	F1-Score
CNN Dasar	0.8585	0.74	0.51	0.49
CNN Class Weight	0.7303	0.63	0.75	0.64
MobileNetV2 Frozen	0.9525	0.92	0.88	0.90
MobileNetV2 Fine-Tuned	0.9649	0.96	0.90	0.92
VGG16 Frozen	0.9269	0.93	0.76	0.82
VGG16 Fine-Tuned	0.9487	0.94	0.84	0.88

Berdasarkan Tabel III, terlihat jelas adanya kesenjangan performa yang sangat lebar antara model yang dibangun dari awal (*CNN Scratch*) dengan model yang memanfaatkan metode *Transfer Learning*. Model *CNN Dasar* mungkin memiliki akurasi yang terlihat cukup baik (85%), namun nilai *F1-Score*-nya sangat rendah (hanya 0,49). Hal ini disebabkan oleh nilai *Recall* yang buruk (0,51), yang berarti model ini sering gagal mengenali sampah daur ulang dan keliru menganggapnya sebagai sampah residu.

Sebaliknya, metode *Transfer Learning* terbukti jauh lebih efektif. Model *MobileNetV2 Fine-Tuned* keluar sebagai model terbaik dengan pencapaian tertinggi di seluruh metrik. Model ini mencatatkan akurasi sebesar 96,49% dan *F1-Score* 0,92. Tingginya nilai metrik ini menunjukkan bahwa strategi optimasi melalui *fine-tuning* berhasil menciptakan keseimbangan model yang sangat baik. Model tidak hanya akurat dalam menebak (*High Precision*), tetapi juga memiliki sensitivitas tinggi terhadap kelas minoritas sehingga sangat sedikit sampah residu yang terlewat (*High Recall*). Meskipun model *VGG16* juga menunjukkan kinerja yang positif, *MobileNetV2* tetap lebih unggul, baik dari segi akurasi maupun efisiensi.

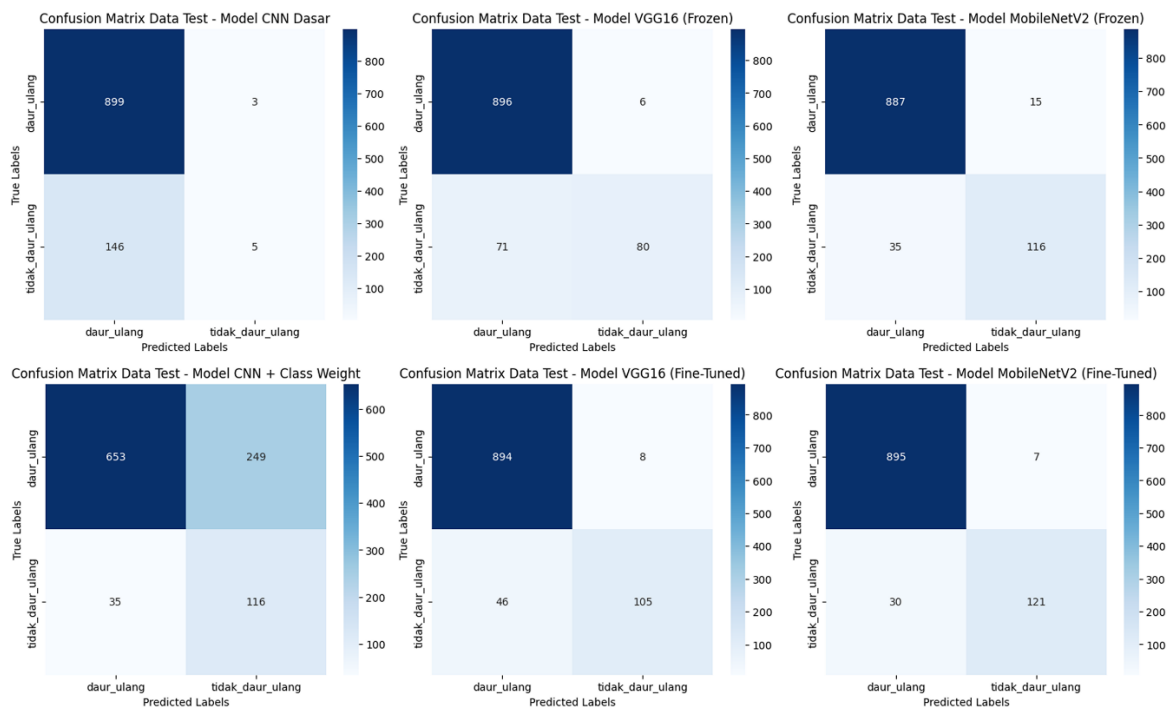
C. Analisis Confusion Matrix

Mengingat karakteristik dataset yang sangat tidak seimbang (*imbalanced*), di mana jumlah sampel kelas daur ulang mendominasi kelas tidak daur ulang (residu), mengandalkan nilai akurasi global saja dapat memberikan gambaran yang kurang menyeluruh. Oleh karena itu, analisis mendalam menggunakan visualisasi *Confusion Matrix* pada Gambar 6 diperlukan untuk menelaah distribusi prediksi model dan mengidentifikasi jenis kesalahan yang paling sering terjadi, khususnya pada kemampuan model mengenali kelas minoritas.

Analisis pada baris pertama Gambar 6 dimulai dengan model CNN Dasar (kiri), yang menunjukkan bias sangat kuat terhadap kelas mayoritas. Model ini mencatatkan angka TP sebesar 899 citra dan FN hanya 3 citra pada kelas daur ulang. Namun, pada kelas residu, model hanya mampu mengenali 5 citra dengan benar (TN) dan melakukan kesalahan prediksi sebesar 146 citra (FP), yang berarti sensitivitas pada kelas minoritas hanya sebesar 3,3%. Selanjutnya, model VGG16 *Frozen* (tengah) mulai menunjukkan peningkatan keseimbangan dengan perolehan TP 896, FN 6, FP 71, dan TN sebesar 80 citra. Peningkatan lebih lanjut terlihat pada

MobileNetV2 *Frozen* (kanan) yang berhasil mencapai TN 116 citra dengan menekan FP menjadi 35 citra, meskipun angka FN pada kelas mayoritas meningkat menjadi 15 citra.

Pada baris kedua Gambar 6, model CNN + *Class Weight* (kiri) menunjukkan upaya penanganan data tidak seimbang yang cukup agresif dengan perolehan TN 116 dan FP 35. Akan tetapi, hal ini berdampak buruk pada kelas mayoritas di mana angka TP turun drastis menjadi 653 citra dan FN melonjak hingga 249 citra. Performa yang jauh lebih stabil ditunjukkan oleh varian *fine-tuned*. Model VGG16 *Fine-Tuned* (tengah) mencatatkan TP 894, FN 8, FP 46, dan TN 105. Terakhir, model MobileNetV2 *Fine-Tuned* (kanan) keluar sebagai model dengan performa paling optimal dan seimbang di seluruh metrik. Model ini sukses mengklasifikasikan 895 citra daur ulang dengan benar (TP) dengan FN hanya 7 citra. Yang terpenting, model ini mencapai angka TN tertinggi yaitu 121 citra (sensitivitas 80,1%) dengan angka FP terendah di kelas minoritas sebesar 30 citra. Keberhasilan ini membuktikan bahwa strategi *fine-tuning* efektif menyempurnakan representasi fitur untuk mengenali karakteristik unik sampah residu tanpa mengorbankan akurasi pada kelas mayoritas.



Gambar 6. Visualisasi *Confusion Matrix* pada Keenam Skenario Eksperimen. Baris atas menyajikan hasil untuk model CNN Dasar, VGG16 *Frozen*, dan MobileNetV2 *Frozen*. Baris bawah menyajikan hasil untuk model CNN + *Class Weight*, VGG16 *Fine-Tuned*, dan MobileNetV2 *Fine-Tuned*. Warna yang lebih gelap pada diagonal utama menunjukkan jumlah prediksi benar yang tinggi, di mana MobileNetV2 *Fine-Tuned* menunjukkan prediksi yang paling seimbang dalam menangani ketidakseimbangan data.

D. Pembahasan

Terdapat berbagai metode yang umum digunakan untuk menangani tantangan ketidakseimbangan data, seperti teknik *oversampling* melalui *Synthetic Minority Over-sampling*

Technique (SMOTE) atau *undersampling*. Namun, pendekatan yang diambil dalam penelitian ini adalah menerapkan teknik *class weighting* dan *fine-tuning* sebagai solusi utama guna mempertahankan integritas distribusi asli dari dataset sekunder. Pemilihan strategi ini didasarkan pada

urgensi untuk menguji ketahanan (*robustness*) arsitektur *deep learning* dalam menghadapi kondisi data yang bersifat *natural* di lapangan, sebagaimana ditekankan oleh Poudel & Poudyal [7] serta Lilhore et al. [13]. Penggunaan data sintesis melalui SMOTE berisiko memperkenalkan *noise* atau pola visual yang sulit direplikasi yang mungkin tidak sepenuhnya representatif terhadap variabilitas sampah rumah tangga yang kompleks. Dengan mempertahankan distribusi asli, hasil penelitian ini memberikan gambaran performa yang lebih objektif dan nyata, sekaligus menghindari risiko *overfitting* yang sering muncul pada metode *oversampling* konvensional.

Efektivitas strategi pelatihan bertahap yang diterapkan dimulai dari fase *frozen* kemudian dilanjutkan dengan *fine-tuning* selaras dengan metodologi yang divalidasi oleh Oktayessofa et al. [12]. Melalui pembukaan 30 lapisan teratas secara selektif, model mampu menyempurnakan representasi fitur untuk mengenali karakteristik unik sampah residu tanpa merusak bobot pra-latih yang sudah stabil. Hal ini terbukti dari peningkatan drastis sensitivitas pada kelas minoritas dari 3,3% pada model *baseline* menjadi 80,1% pada model *MobileNetV2 Fine-Tuned*. Penggunaan *optimizer* Adam juga terbukti signifikan dalam mempercepat konvergensi dan stabilitas akurasi selama fase transisi antar tahap pelatihan tersebut, sebuah pendekatan yang juga didukung oleh Khan et al. [11].

Selain keunggulan dalam akurasi, aspek efisiensi komputasi menjadi poin signifikan yang mendukung klaim kesesuaian model untuk implementasi secara *real-time*. Berbeda dengan arsitektur tradisional seperti VGG16 yang memiliki beban parameter sangat besar mencapai 138,4 juta parameter, *MobileNetV2* dirancang menggunakan struktur *Inverted Residuals* dan *Linear Bottlenecks* yang jauh lebih efisien dengan hanya sekitar 3,4 juta parameter. Penggunaan parameter yang minimal ini secara teoretis berdampak langsung pada penghematan memori dan kecepatan *inference*. Temuan ini selaras dengan penelitian Kumala et al. [18] yang menunjukkan bahwa *MobileNetV2* memberikan keseimbangan performa terbaik untuk klasifikasi sampah dibandingkan VGG16, dengan kecepatan pelatihan mencapai 350 ms/step, dua kali lebih cepat daripada VGG16 yang membutuhkan 700 ms/step. Hal ini memperkuat posisi *MobileNetV2 Fine-Tuned* sebagai solusi yang handal dan ringan untuk diintegrasikan ke dalam ekosistem *edge computing* maupun perangkat *mobile* dengan sumber daya terbatas.

IV. KESIMPULAN

Berdasarkan tujuan penelitian untuk mengoptimalkan klasifikasi citra sampah pada kondisi data yang tidak seimbang, dapat disimpulkan bahwa strategi *fine-tuning* pada arsitektur *MobileNetV2 Fine-Tuned* terbukti menjadi metode yang paling efektif dibandingkan skenario lainnya. Penerapan metode *Transfer Learning* ini jauh lebih unggul daripada pembangunan model *Convolutional Neural Network* (CNN) dari awal (*scratch*) yang mengalami kendala generalisasi signifikan, ditandai dengan rendahnya nilai *F1-Score* (0,49)

dan ketidakmampuan mengenali kelas minoritas (residu). Hasil eksperimen membuktikan bahwa optimasi melalui skenario *fine-tuning* pada *MobileNetV2 Fine-Tuned* berhasil mencatatkan performa tertinggi dalam penelitian ini, dengan *accuracy* pengujian sebesar 96,49%, *Precision* 0,96, dan *Recall* 0,90. Tingginya nilai *Recall* tersebut secara signifikan menjawab tantangan *imbalanced* data, di mana teknik *fine-tuning* mampu meningkatkan sensitivitas model terhadap kelas minoritas secara optimal.

Penelitian ini memiliki beberapa keterbatasan yang perlu diakui. Pertama, dataset yang digunakan berasal dari lingkungan terkontrol dengan latar belakang yang bersih, sehingga performa model pada kondisi lapangan yang dinamis dengan gangguan visual (*noise*) atau pencahayaan ekstrem masih memerlukan pengujian lebih lanjut. Kedua, klasifikasi dalam studi ini baru mencakup dua kategori besar, sedangkan implementasi nyata mungkin membutuhkan identifikasi material yang lebih spesifik. Selain itu, evaluasi efisiensi arsitektur baru dilakukan pada kondisi ideal menggunakan infrastruktur komputasi laboratorium, sehingga validasi langsung pada perangkat *edge computing* dengan sumber daya terbatas masih menjadi ruang terbuka untuk penelitian mendatang.

Sebagai upaya pengembangan, disarankan untuk memperluas dataset dengan variasi visual yang lebih kompleks guna memperkaya kemampuan generalisasi model terhadap kondisi nyata. Ruang lingkup klasifikasi juga dapat ditingkatkan menjadi multi-kelas yang lebih detail untuk membedakan jenis material secara fungsional. Terakhir, mengingat efisiensi arsitektur *MobileNetV2* yang telah terbukti, integrasi model ke dalam ekosistem *edge computing* atau perangkat *mobile* sangat dianjurkan untuk mengevaluasi keseimbangan antara akurasi dan latensi sistem secara *real-time* pada kondisi lapangan yang dinamis.

DAFTAR PUSTAKA

- [1] L. J. Lingga, M. Yuana, N. A. Sari, H. N. Syahida, and C. S. Shahron, "Sampah di Indonesia: Tantangan dan Solusi Menuju Perubahan Positif," vol. 4, pp. 12235–12247, 2024.
- [2] H. Zhang, H. Cao, Y. Zhou, C. Gu, and D. Li, "Urban Climate Hybrid deep learning model for accurate classification of solid waste in the society," *Urban Clim.*, vol. 49, no. February, p. 101485, 2023, doi: 10.1016/j.uclim.2023.101485.
- [3] A. T. Setiawan, "Identifikasi Jenis Sampah Secara Otomatis Menggunakan Metode Convolutional Neural Network (CNN)," pp. 345–357.
- [4] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, and S. Member, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, 2022, doi: 10.1109/TNNLS.2021.3084827.
- [5] F. Ahamed, M. Naznine, and J. Karim, "Knowledge-Based Systems An automated waste classification system using deep learning techniques: Toward efficient waste recycling and environmental sustainability," *Knowledge-Based Syst.*, vol. 310, no. May 2024, p. 113028, 2025, doi: 10.1016/j.knosys.2025.113028.
- [6] E. T. Y. M. Koklu, "A comparative analysis of machine learning algorithms for waste classification: inceptionv3 and chi - square features," *Int. J. Environ. Sci. Technol.*, vol. 22, no. 10, pp. 9415–9428, 2025, doi: 10.1007/s13762-024-06233-z.
- [7] S. Poudel, "Classification of Waste Materials using CNN Based on

- Transfer Learning,” pp. 29–33, 2022, doi: 10.1145/3574318.3574345.
- [8] T. Soni, “MobileNet-Based Garbage Classification: Enhancing Recycling with Machine Learning,” *2024 Int. Conf. Intell. Comput. Emerg. Commun. Technol.*, pp. 1–4, 2024, doi: 10.1109/ICEC59683.2024.10837152.
- [9] S. Sahibu and I. Taufik, “Implementation of the Convolutional Neural Network Algorithm for Classifying Types of Organic and Non-Organic Waste Implementasi Algoritma Convolutional Neural Network untuk Klasifikasi Jenis Sampah Organik dan Non Organik,” vol. 4, no. July, pp. 840–852, 2024.
- [10] T. Sulistyowati, F. Al Zami, and R. A. Pramunendar, “VGG16 Deep Learning Architecture Using Imbalance Data Methods For The Detection Of Apple Leaf Diseases,” vol. 11, no. 1, pp. 41–53, 2023.
- [11] A. Irshad, A. S. Almalaise, Y. B. Abushark, F. Alsolami, A. Almalawi, and A. Marish, “Chemosphere Recycling waste classification using emperor penguin optimizer with deep learning model for bioenergy production,” *Chemosphere*, vol. 307, no. P3, p. 136044, 2022, doi: 10.1016/j.chemosphere.2022.136044.
- [12] E. Oktayessofa, C. A. Sari, E. H. Rachmawanto, and N. M. Yaacob, “Classification Of Organic And Non-Organic Waste With CNN-,” vol. 5, no. 4, pp. 1173–1180, 2024.
- [13] U. Kumar, S. Simaiya, S. Dalal, M. Radulescu, and D. Balsalobre-loriente, “Intelligent waste sorting for sustainable environment: A hybrid deep learning and transfer learning model,” *Gondwana Res.*, no. xxxx, 2024, doi: 10.1016/j.gr.2024.07.014.
- [14] P. N. Dacipta and R. E. Putra, “Sistem Klasifikasi Limbah Menggunakan Metode Convolutional Neural Network (CNN) Pada Web Service Berbasis Framework Flask,” vol. 03, pp. 394–402, 2022.
- [15] A. S. Ouedraogo, A. Kumar, and N. Wang, “Landfill Waste Segregation Using Transfer and Ensemble Machine Learning: A Convolutional Neural Network Approach,” 2023.
- [16] N. Li and Y. Chen, “Urban Climate Municipal solid waste classification and real-time detection using deep learning methods,” *Urban Clim.*, vol. 49, no. February, p. 101462, 2023, doi: 10.1016/j.uclim.2023.101462.
- [17] T. H. Yan, S. Nazrah, M. Azam, Z. Sani, and A. Azizan, “Accuracy study of image classification for reverse vending machine waste segregation using convolutional neural network,” vol. 14, no. 1, pp. 366–374, 2024, doi: 10.11591/ijece.v14i1.pp366-374.
- [18] R. A. Kumala, C. A. Sari, and E. H. Rachmawanto, “A Comparison of MobileNetV2 and VGG16 Architectures with Transfer Learning for Multi-Class Image-Based Waste Classification,” vol. 9, no. 4, pp. 1610–1624, 2025.