PENINGKATAN AKURASI KLASIFIKASI CITRA SAMPAH ORGANIK DAN ANORGANIK MELALUI FINE-TUNING MOBILENETV2 DAN STRATIFIED DATA AUGMENTATION

Muhammad Nur Rizqan[1], Robert Marco[2]

[1] Department of Computer Science, Universitas Amikom Yogyakarta [2,] Department of Computer Science, Universitas Amikom Yogyakarta

Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kab. Sleman, Daerah Istimewa Yogyakarta 55281
[1] muhammadnurrizqan@students.amiko m.ac.id, [2] robertmarco@amikom.ac.id

Abstract

Effective waste sorting is essential for managing growing waste volumes, yet manual sorting depends heavily on human visual judgment and is prone to error. This study develops an image classification model for organic and recyclable waste using MobileNetV2 transfer learning, fine-tuning, and data augmentation, and examines the interpretability of its predictions with Grad-CAM. The Waste Classification Dataset (22,564 images) was divided using a stratified 70:15:15 split, so that the class proportions (55.69% organic, 44.31% recyclable) remained identical across the training, validation, and test sets. Augmentation (random flip, rotation, zoom, and contrast) was applied dynamically to the training data only, which prevents data leakage. Training proceeded in two stages: feature extraction with a frozen ImageNet-pretrained backbone, followed by fine-tuning of the last 30 backbone layers. Fine-tuning raised test accuracy from 95.66% to 96.22%, with precision, recall, and F1-score (weighted) of 96.22% and a ROC-AUC of 0.9928. In an ablation experiment, removing augmentation lowered test accuracy to 95.63%, whereas a random (non-stratified) split gave a nearly identical 96.28%, indicating that stratification mainly safeguards evaluation validity rather than accuracy. Grad-CAM visualizations show that the model bases its decisions on the main object in the image. These findings show that MobileNetV2 with fine-tuning and data augmentation is an effective and interpretable approach for organic and recyclable waste classification.

Keywords : *Deep Learning, MobileNetV2, Transfer Learning, Data Augmentation, Fine-Tuning, Waste Classification, Grad-CAM.*

Abstrak

Pengelolaan sampah yang baik memerlukan pemilahan yang efektif, sedangkan pemilahan manual sangat bergantung pada penilaian visual manusia sehingga rentan keliru. Penelitian ini mengembangkan model klasifikasi citra sampah organik dan recyclable menggunakan transfer learning MobileNetV2, fine-tuning, dan data augmentation, serta menelaah interpretabilitas prediksinya dengan Grad-CAM. Waste Classification Dataset (22.564 citra) dibagi secara stratified dengan rasio 70:15:15, sehingga proporsi kelas (55,69% organik dan 44,31% recyclable) sama pada data latih, validasi, dan uji. Augmentasi (random flip, rotation, zoom, dan contrast) diterapkan secara dinamis hanya pada data latih untuk mencegah kebocoran data. Pelatihan berlangsung dalam dua tahap, yaitu feature extraction dengan backbone berbobot ImageNet yang dibekukan, lalu fine-tuning pada 30 layer terakhir backbone. Fine-tuning meningkatkan akurasi data uji dari 95,66% menjadi 96,22%, dengan precision, recall, dan F1-score (weighted) sebesar 96,22% serta ROC-AUC 0,9928. Pada eksperimen ablasi, penghapusan augmentasi menurunkan akurasi uji menjadi 95,63%, sedangkan pembagian acak (non-stratified) menghasilkan 96,28%, hampir sama, sehingga stratifikasi terutama menjaga validitas evaluasi dan bukan meningkatkan akurasi. Visualisasi Grad-CAM memperlihatkan bahwa model mendasarkan keputusannya pada objek utama dalam citra. Temuan ini menunjukkan bahwa MobileNetV2 dengan fine-tuning dan data augmentation merupakan pendekatan yang efektif dan dapat diinterpretasikan untuk klasifikasi citra sampah organik dan recyclable.

Kata kunci : *Deep Learning, MobileNetV2, Transfer Learning, Data Augmentation, Fine-Tuning, Klasifikasi Sampah, Grad-CAM*

1. PENDAHULUAN

Peningkatan jumlah sampah setiap tahun menjadi salah satu permasalahan lingkungan yang memerlukan perhatian serius. Pertumbuhan penduduk, urbanisasi, dan meningkatnya aktivitas konsumsi menyebabkan volume sampah terus bertambah sehingga diperlukan sistem pengelolaan yang lebih efektif. Salah satu tahapan penting dalam pengelolaan sampah adalah proses pemilahan berdasarkan jenisnya, karena pemilahan yang tepat dapat meningkatkan efisiensi proses daur ulang serta mengurangi dampak negatif terhadap lingkungan[1].

Proses pemilahan sampah masih banyak dilakukan secara manual sehingga memerlukan waktu, tenaga, dan rentan terhadap kesalahan manusia. Perkembangan teknologi kecerdasan buatan, khususnya *Deep Learning*, memberikan peluang untuk mengotomatisasi proses tersebut melalui pengenalan objek berbasis citra. Salah satu metode yang banyak digunakan adalah Convolutional Neural Network (CNN) karena memiliki kemampuan dalam mengekstraksi karakteristik visual secara otomatis tanpa memerlukan proses ekstraksi fitur secara manual[2].

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan integrasi Transfer Learning menggunakan MobileNetV2, teknik Data Augmentasi, serta pendekatan Explainable Artificial Intelligence (XAI) menggunakan Grad-CAM. MobileNetV2 dipilih karena memiliki jumlah parameter yang relatif ringan namun tetap mampu menghasilkan performa klasifikasi yang tinggi [3]. Sementara itu, Data Augmentasi digunakan untuk meningkatkan variasi data pelatihan sehingga kemampuan generalisasi model menjadi lebih baik. Metode Grad-CAM diterapkan untuk memberikan visualisasi terhadap area citra yang menjadi dasar pengambilan keputusan model sehingga hasil klasifikasi dapat dijelaskan secara lebih transparan [4].

Sejumlah penelitian terdahulu telah mengkaji klasifikasi citra sampah dengan berbagai arsitektur. Ma et al. (2023) menerapkan transfer learning MobileNetV2 dan melaporkan akurasi 82,92% [5] . Chhabra et al. (2024) menggunakan CNN yang dimodifikasi dengan akurasi 93,28% [6] , sedangkan Gürcan dan Soylu (2026) membandingkan beberapa arsitektur transfer learning dan memperoleh akurasi 93,90% untuk MobileNetV2 serta 96,95% untuk ConvNeXt_V1 [7]. Pendekatan yang lebih kompleks menghasilkan akurasi lebih tinggi. Islam et al. (2025) menggabungkan DenseNet201, parallel CNN, dan SE attention disertai Grad-CAM hingga mencapai 97,43% [8] , sementara Nguyen et al. (2026) mengombinasikan ResNet50 yang dikustomisasi dengan pendekatan hybrid deep learning dan machine learning hingga 99,72% [9] . Sebagai pembanding, penelitian ini memperoleh akurasi 96,22% pada data uji dengan MobileNetV2 yang memiliki sekitar 2,4 juta parameter. Capaian tersebut lebih tinggi daripada hasil pada [10] , [6] , dan MobileNetV2 pada [7] , tetapi masih di bawah ConvNeXt_V1 pada [7] serta model hibrida yang lebih berat pada [8] dan [9]. Perbandingan ini bersifat indikatif karena setiap penelitian memakai dataset dan skema evaluasi yang berbeda.

Berdasarkan penelitian yang dibandingkan tersebut, sepanjang penelusuran penulis terdapat tiga celah yang masih dapat diisi. Pertama, akurasi akhir umumnya dilaporkan sebagai satu angka tanpa memisahkan kontribusi tahap feature extraction dan fine-tuning, sehingga besarnya peningkatan akibat penyesuaian bobot backbone tidak dapat diketahui. Kedua, peran augmentasi data dan skema pembagian data (stratified atau acak) jarang diuji terpisah melalui eksperimen ablasi, padahal keduanya lazim dicantumkan sebagai bagian dari metode. Ketiga, rincian protokol evaluasi, seperti jumlah citra pada tiap subset dan tiap kelas serta langkah pencegahan kebocoran data (data leakage), tidak selalu dijelaskan sehingga hasilnya sulit direplikasi. Di sisi lain, model dengan akurasi yang lebih tinggi cenderung memiliki arsitektur yang lebih kompleks, sehingga masih relevan menelaah seberapa jauh MobileNetV2 yang ringan dapat dioptimalkan tanpa menambah kompleksitas arsitektur.

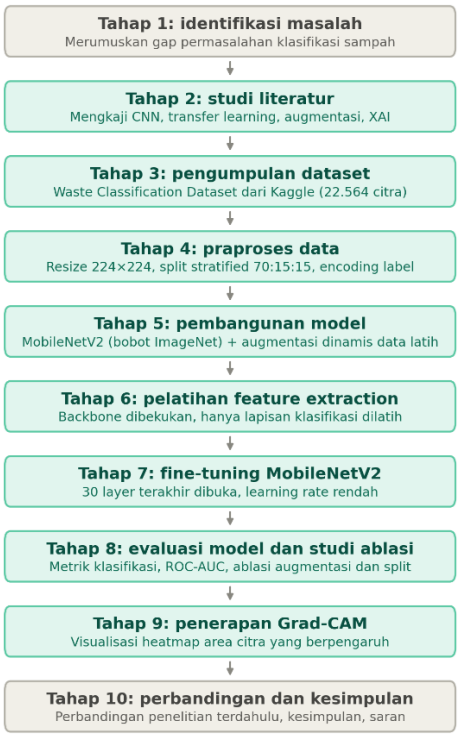
Untuk mengisi celah tersebut, penelitian ini bertujuan mengembangkan dan menganalisis model klasifikasi citra sampah organik dan recyclable berbasis MobileNetV2 dengan transfer learning dua tahap, data augmentation, dan Grad-CAM. Kebaruan penelitian ini terletak pada

analisis yang terukur, bukan pada arsitektur baru, dengan empat kontribusi. Pertama, kontribusi tiap tahap pelatihan diukur terpisah pada data validasi dan data uji sehingga peningkatan akibat fine-tuning dapat dikuantifikasi. Kedua, eksperimen ablasi membandingkan penggunaan dan ketiadaan augmentasi serta pembagian data stratified dan acak untuk menilai peran masing-masing. Ketiga, protokol evaluasi dirancang menjaga validitas hasil melalui pembagian stratified 70:15:15 yang mempertahankan proporsi kelas di setiap subset, augmentasi dinamis yang hanya aktif pada data latih, dan pemeriksaan tidak adanya tumpang tindih antar subset, dengan jumlah citra yang dilaporkan lengkap agar eksperimen dapat direplikasi. Keempat, Grad-CAM digunakan untuk memeriksa apakah keputusan model didasarkan pada objek yang relevan [11] .

Hasil penelitian menunjukkan bahwa fine-tuning meningkatkan akurasi data uji dari 95,66% menjadi 96,22%, dan augmentasi memberikan tambahan sekitar 0,6 poin persentase dibandingkan tanpa augmentasi. Pembagian stratified tidak mengubah akurasi secara berarti dibandingkan pembagian acak, sehingga perannya terutama menjaga validitas dan keterulangan evaluasi, bukan meningkatkan akurasi. Kinerja model dinilai dengan accuracy, precision, recall, F1-score, confusion matrix, dan ROC-AUC, sedangkan Grad-CAM memberikan interpretasi visual atas area citra yang menjadi dasar prediksi.

2. METODOLOGI PENELITIAN

Penelitian ini mengusulkan model klasifikasi citra sampah organik dan anorganik menggunakan pendekatan Transfer Learning berbasis MobileNetV2 yang dikombinasikan dengan teknik Data Augmentation, Fine-Tuning, dan Explainable Artificial Intelligence (XAI) menggunakan Grad-CAM [5]. Tahapan penelitian terdiri atas pengumpulan dataset, praproses data, pembentukan model, pelatihan, evaluasi performa, dan interpretasi hasil klasifikasi. Alur penelitian dapat dilihat pada Gambar 1



Gambar 1. Alur Penelitian

2.1. Dataset

Penelitian ini menggunakan Waste Classification Dataset yang diperoleh dari platform Kaggle. Dataset terdiri atas dua kelas, yaitu Organic dan Recyclable, dengan total 22.564 citra berformat JPG dan PNG, yang terdiri atas 12.565 citra Organic (55,69%) dan 9.999 citra Recyclable (44,31%). Seluruh citra dibagi menjadi data latih, validasi, dan uji dengan rasio 70:15:15 secara stratified (Tabel I).

TABEL I. PEMBAGIAN DATASET

	Organic	Recyclable	Jumlah	Proporsi Organic:Recyclable
Seluruh data	12.565	9.999	22.564	55,69% : 44,31%
Training	8.795	6.999	15.794	55,69% : 44,31%
Validation	1.885	1.500	3.385	55,69% : 44,31%
Testing	1.885	1.500	3.385	55,69% : 44,31%

Pembagian dilakukan dalam dua tahap menggunakan fungsi train_test_split (scikit-learn) dengan parameter stratify dan random_state = 42. Pada tahap pertama, 70% data dipisahkan sebagai data latih, kemudian 30% sisanya dibagi sama besar menjadi data

validasi dan data uji. Daftar file diurutkan terlebih dahulu sebelum pembagian agar hasilnya dapat direproduksi [12] . Dengan skema ini, proporsi kelas pada setiap subset identik dengan proporsi pada dataset asli.

2.2. Praproses Data

Sebelum proses pelatihan dilakukan, seluruh citra melalui beberapa tahapan praproses untuk menyeragamkan format data yang akan digunakan sebagai masukan model.

Tahapan praproses meliputi:

- Penyesuaian ukuran citra. Seluruh citra diubah ukurannya menjadi 224 × 224 piksel sesuai ukuran input MobileNetV2.
- Penskalaan intensitas piksel. Nilai piksel dipertahankan pada rentang 0-255 dalam format float32, kemudian fungsi preprocess_input bawaan MobileNetV2 yang terpasang di dalam model menskalakannya ke rentang [-1, 1]. Skala ini sama dengan skala yang digunakan saat bobot ImageNet dilatih, sehingga bobot awal dapat dimanfaatkan secara optimal.
- Pembagian dataset. Dataset dibagi secara stratified sebagaimana dijelaskan pada subbab 2.1.
- Pengodean label. Label Organic dan Recyclable diubah menjadi representasi one-hot agar dapat diproses oleh fungsi loss categorical crossentropy.

Tahapan tersebut bertujuan menghasilkan data yang seragam sehingga proses pelatihan model menjadi lebih stabil.



Gambar 1. Contoh Citra Organik Dan Anorganik

2.3. Data Augmentasi

Untuk meningkatkan kemampuan generalisasi dan menekan risiko overfitting, penelitian ini menerapkan augmentasi data secara dinamis menggunakan layer preprocessing bawaan Keras yang menjadi bagian awal model, sehingga variasi citra dibangkitkan langsung selama pelatihan [13] . Empat transformasi yang digunakan beserta parameternya disajikan pada Tabel II.

TABEL II. TRANSFORMASI AUGMENTASI

Transformasi	Layer Keras	Parameter	Tujuan
Flip horizontal	Random Flip ("horizontal")	Dibalik dengan peluang 50%	Menambah keragaman orientasi objek
Rotasi	Random Rotation (0.15)	Acak dalam ±15% putaran penuh (±54°)	Mengatasi variasi sudut pengambilan gambar
Zoom	Random Zoom(0.15)	Zoom-in atau zoom-out acak hingga 15%	Mensimulasikan variasi jarak kamera
Kontras	Random Contrast (0.10)	Faktor kontras acak 0,9-1,1	Mengatasi perbedaan kondisi pencahayaan

Jumlah citra sebelum dan sesudah augmentasi. Augmentasi dilakukan secara *on-the-fly* dan tidak menyimpan berkas citra baru, sehingga jumlah citra latih yang tersimpan tidak berubah, yaitu 15.794 citra. Pada setiap epoch, setiap citra latih ditampilkan sekali dengan transformasi acak yang berbeda. Selama 25 epoch pelatihan (15 epoch feature extraction dan 10 epoch fine-tuning), model memproses total 394.850 tampilan citra latih (15.794 × 25) yang masing-masing telah mengalami variasi transformasi berbeda. Angka ini adalah jumlah tampilan, bukan jumlah citra unik.

Pencegahan data leakage**.** Empat langkah diterapkan untuk menjaga validitas evaluasi:

1. Pembagian data dilakukan pada tingkat berkas sebelum proses lain, dan dilakukan pemeriksaan bahwa tidak ada berkas yang sama pada dua subset berbeda.
2. Layer augmentasi hanya aktif saat pelatihan dan otomatis nonaktif saat validasi, pengujian, dan prediksi, sehingga data validasi dan uji tetap berupa citra asli.
3. Pemilihan model (EarlyStopping, ReduceLROnPlateau, dan ModelCheckpoint) hanya memantau

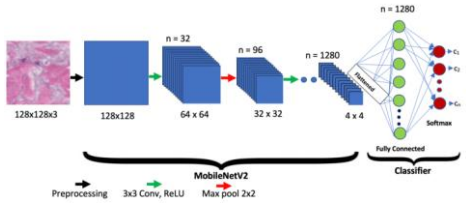
data validasi, sedangkan data uji dipakai hanya untuk pelaporan akhir.
4. Seed acak dan daftar berkas tiap subset disimpan sehingga eksperimen dapat direplikasi.

2.4. Arsitektur MobileNetV2

Penelitian ini menggunakan MobileNetV2 sebagai model dasar (*backbone*) karena memiliki jumlah parameter yang relatif sedikit namun tetap mampu menghasilkan performa klasifikasi yang tinggi. Bobot awal model diperoleh dari dataset ImageNet sehingga proses pelatihan dapat dilakukan menggunakan pendekatan *Transfer Learning [14]*.

Pada tahap awal, seluruh layer MobileNetV2 dibekukan (*freeze*) sehingga hanya lapisan klasifikasi yang diperbarui selama proses pelatihan (*feature extraction*). Setelah model mencapai konvergensi, beberapa lapisan akhir MobileNetV2 dibuka kembali (*unfreeze*) untuk menjalankan proses *fine-tuning [5]*.
Arsitektur model yang digunakan terdiri atas:

1. Input Layer (224 × 224 × 3)
2. Data Augmentation Layer
3. Preprocessing Layer (MobileNetV2 Preprocess Input)
4. MobileNetV2 Backbone
5. Global Average Pooling
6. Dropout (0,30)
7. Dense Layer (128 neuron, ReLU)
8. Dropout (0,20)
9. Output Layer (Softmax, 2 kelas)



Gambar 2. Arsitektur MobileNetV2

2.5. Pelatihan Model

Pelatihan model dilakukan dalam dua tahapan, yaitu Feature Extraction dan Fine-Tuning.

Tahap pertama menggunakan seluruh bobot MobileNetV2 dalam kondisi tidak dapat diperbarui (*freeze*) sehingga model hanya mempelajari parameter pada lapisan klasifikasi. Setelah proses tersebut selesai, beberapa lapisan akhir MobileNetV2 dibuka kembali untuk meningkatkan kemampuan model dalam mengenali karakteristik citra sampah.

Model dikompilasi menggunakan fungsi kerugian **Categorical Crossentropy**, optimizer **Adam**, serta metrik evaluasi berupa **Accuracy**.
TABEL III. PARAMETER PELATIHAN

Parameter	Nilai
Optimizer	Adam
Loss function	Categorical crossentropy
Batch size	32
Ukuran citra	224 × 224
Epoch maksimum feature extraction	15
Epoch maksimum fine-tuning	10
Learning rate feature extraction	0,001
Learning rate fine-tuning	0,00001
EarlyStopping	Memantau val_loss, patience 5, mengembalikan bobot terbaik
ReduceLROnPlateau	Memantau val_loss, faktor 0,2, patience 2, learning rate minimum 10^{-6} (feature extraction) dan 10^{-7} (fine-tuning)
ModelCheckpoint	Menyimpan model dengan val_loss terendah
Seed acak	42

Ketiga callback memantau metrik yang sama, yaitu val_loss, sehingga bobot yang dipakai untuk evaluasi adalah bobot pada epoch dengan val_loss terendah [9]. Pada penelitian ini, bobot terbaik diperoleh pada epoch ke-14 untuk feature extraction dan epoch ke-6 untuk fine-tuning. Pelatihan dijalankan pada Kaggle Notebook dengan GPU NVIDIA Tesla T4 menggunakan TensorFlow 2.20 [9] .

2.6. Evaluasi Model

Kinerja model dievaluasi menggunakan beberapa metrik klasifikasi untuk memberikan gambaran performa secara menyeluruh[15].

Metrik yang digunakan meliputi:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix
- ROC Curve
- Area Under Curve (AUC)

Penggunaan berbagai metrik tersebut bertujuan memastikan bahwa model tidak hanya memiliki akurasi tinggi, tetapi juga mampu memberikan performa yang konsisten pada masing-masing kelas.

Metrik precision, recall, dan F1-score dihitung dengan rata-rata tertimbang (weighted average) terhadap jumlah data tiap kelas. Evaluasi dilakukan pada data validasi dan data uji untuk setiap tahap pelatihan, sedangkan data uji hanya digunakan setelah pelatihan selesai.

2.7. Skenario Ablas

Untuk mengukur kontribusi masing-masing komponen, dilakukan eksperimen ablasi dengan konfigurasi pelatihan yang sama pada Tabel III dan seed 42. Tiga skenario dibandingkan, yaitu (1) augmentasi dengan pembagian stratified (skenario utama), (2) augmentasi dengan pembagian acak tanpa stratifikasi, dan (3) pembagian stratified tanpa augmentasi. Ketiganya dinilai pada data uji setelah tahap feature extraction dan setelah fine-tuning. Selain itu, perbandingan antara hasil sebelum dan sesudah fine-tuning pada skenario utama digunakan untuk mengukur kontribusi fine-tuning.

2.8. Interpretasi Model Menggunakan Grad-CAM

Untuk meningkatkan transparansi hasil prediksi, penelitian ini menerapkan metode Gradient-weighted Class Activation Mapping (Grad-CAM).Visualisasi Grad-CAM membantu menjelaskan proses pengambilan keputusan model sehingga pengguna dapat memahami hubungan antara objek pada citra dengan hasil prediksi yang dihasilkan. Dengan demikian, model tidak lagi dipandang sebagai sistem *black-box*, tetapi memiliki tingkat interpretabilitas yang lebih baik[10].

Heatmap dihitung dari feature map layer konvolusi terakhir backbone (out_relu, berukuran 7 × 7 × 1.280) dengan memanfaatkan gradien skor kelas yang diprediksi. Gradien dirata-ratakan per kanal sebagai bobot, feature map dijumlahkan berbobot, diterapkan fungsi ReLU, dinormalisasi ke rentang 0-1, lalu diubah ukurannya dan ditumpangkan pada citra asli dengan transparansi 40% menggunakan peta warna jet [11] .

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pelatihan Model

Pelatihan berlangsung dalam dua tahap sebagaimana dijelaskan pada subbab 2.5. Pada tahap feature extraction, akurasi validasi naik dari 93,38% pada epoch pertama hingga berkisar 95,3-95,9% pada epoch ke-8 sampai ke-15, sementara val_loss turun dari 0,1699 menjadi 0,1184. Nilai val_loss terendah dicapai pada epoch ke-14, sehingga bobot pada epoch tersebut dipulihkan dengan akurasi validasi 95,75%. Learning rate diturunkan dua kali oleh ReduceLROnPlateau, dari 0,001 menjadi 0,0002 setelah epoch ke-7 dan menjadi 0,00004 setelah epoch ke-11. Pada epoch ke-15, akurasi latih 96,00% dan akurasi validasi 95,83%, selisih hanya 0,17 poin persentase.

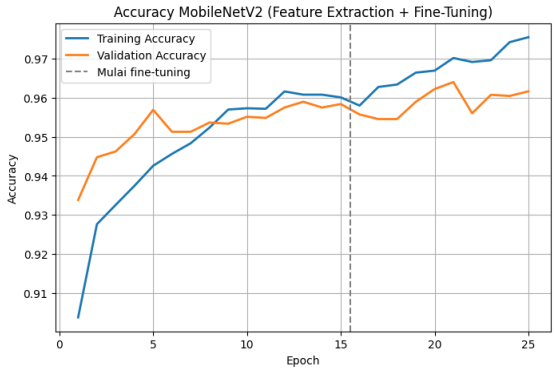
Pada tahap fine-tuning, val_loss turun dari 0,1262 pada epoch pertama menjadi 0,1059 pada epoch ke-6, dengan akurasi validasi 96,40%. Pada epoch ke-7 sampai ke-10 val_loss tidak lagi lebih rendah (0,1089-0,1148), sehingga bobot epoch ke-6 dipulihkan. Learning rate turun menjadi 0,000002 setelah epoch ke-8 dan 0,0000004 setelah epoch ke-10. EarlyStopping tidak terpicu karena hanya terjadi empat epoch berturut-turut tanpa perbaikan, kurang dari patience 5. Akurasi latih terus naik hingga 97,54% pada epoch ke-10, sedangkan akurasi validasi berfluktuasi antara 95,60% dan 96,16%. Selisih keduanya melebar dari 0,61 poin pada epoch ke-6 menjadi 1,38 poin pada epoch ke-10, yang menandakan overfitting ringan pada akhir fine-tuning. Pemulihan bobot terbaik melindungi model dari dampak tersebut.

TABEL IV. KINERJA SEBELUM DAN SESUDAH FINE-TUNING

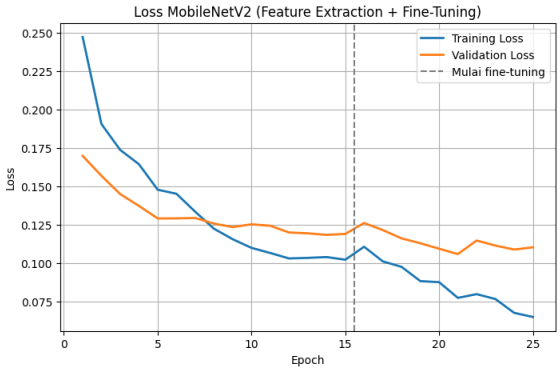
Tahap	Data	Accuracy	Precision	Recall	F1-score	AUC
Feature extraction	Validasi	95,75 %	95,75 %	95,75%	95,75%	0,9910
Feature extraction	Uji	95,66 %	95,66 %	95,66%	95,66%	0,9917
Fine-tuning	Validasi	96,40 %	96,40 %	96,40%	96,39%	0,9928
Fine-tuning	Uji	96,22 %	96,22 %	96,22%	96,22%	0,9928
Selisih pada data uji		+0,56	+0,56	+0,56	+0,56	+0,0011

Metrik precision, recall, dan F1-score menggunakan weighted average. Fine-tuning meningkatkan seluruh metrik pada data validasi maupun data uji. Pada data uji, peningkatan akurasi 0,56 poin persentase setara dengan sekitar 19 citra tambahan yang terklasifikasi benar dari 3.385 citra. Peningkatan ini tergolong kecil, yang wajar karena fitur ImageNet pada backbone sudah cukup representatif untuk dataset ini dan fine-tuning dijalankan dengan

learning rate rendah. Arah peningkatannya konsisten pada ketiga skenario ablasi (subbab 3.5).



Gambar 4. Accuracy MobileNetV2 (Feature Extrction + Fine Tuning



Gambar 5. Loss MobileNetV2 (Feature Extrction + Fine Tuning

Gambar 4 dan 5 menampilkan kurva akurasi dan loss gabungan kedua tahap, dengan garis putus-putus menandai awal fine-tuning. Pada tahap feature extraction, kurva latih dan validasi naik dan turun beriringan dengan selisih kecil. Pada tahap fine-tuning, kurva latih terus membaik sedangkan kurva validasi mendatar dengan fluktuasi kecil, sesuai pola pada paragraf sebelumnya.

3.2 Evaluasi Kinerja Model

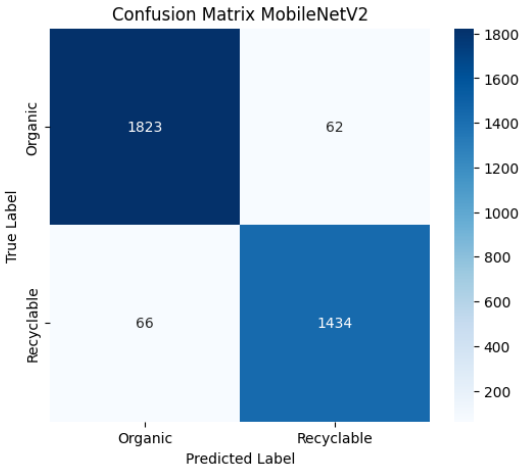
Setelah pelatihan selesai, model diuji pada data uji (3.385 citra) yang tidak digunakan untuk pelatihan maupun pemilihan model [6] . Model setelah fine-tuning memperoleh accuracy 96,22%, dengan precision, recall, dan F1-score (weighted) masing-masing 96,22% (Tabel V). Akurasi pada data uji hanya 0,18 poin di bawah akurasi validasi (96,40%), yang menunjukkan bahwa performa model relatif stabil pada data yang belum pernah dilihat.

Tabel V. Hasil Evaluasi per Kelas pada Data Uji (Setelah Fine-Tuning)

Kelas	Precision	Recall	F1-score	Jumlah data
Organic	96,51%	96,71%	96,61%	1.885
Recyclable	95,86%	95,60%	95,73%	1.500
Macro average	96,18%	96,16%	96,17%	3.385
Weighted average	96,22%	96,22%	96,22%	3.385
Accuracy			96,22%	3.385

Kinerja pada kedua kelas seimbang. Kelas Organic memiliki F1-score 0,88 poin lebih tinggi daripada kelas Recyclable, dengan selisih precision 0,65 poin dan recall 1,11 poin. Precision yang tinggi menunjukkan bahwa sebagian besar prediksi pada tiap kelas sesuai dengan kelas sebenarnya, sedangkan recall yang tinggi menunjukkan bahwa sebagian besar objek pada tiap kelas berhasil dikenali.

3.3 Analisis Confusion Matrix



Gambar 6. Confusion Matrix MobileNetV2

Gambar 6 menunjukkan confusion matrix pada data uji. Dari 1.885 citra Organic, 1.823 diklasifikasikan dengan benar dan 62 keliru diprediksi sebagai Recyclable. Dari 1.500 citra Recyclable, 1.434 diklasifikasikan dengan benar dan 66 keliru diprediksi sebagai Organic. Total kesalahan adalah 128 dari 3.385 citra (3,78%). Untuk mengetahui distribusi prediksi pada setiap kelas, dilakukan evaluasi menggunakan Confusion Matrix.

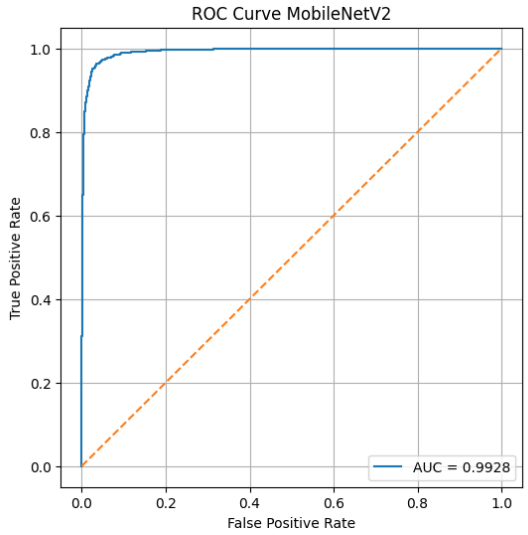
Kesalahan tersebar relatif merata pada kedua arah: 3,29% citra Organic salah diklasifikasikan, dibandingkan 4,40% citra Recyclable. Model tidak menunjukkan kecenderungan kuat terhadap salah satu kelas, meskipun kelas Recyclable sedikit lebih rentan keliru. Penyebab pastinya tidak dianalisis pada penelitian ini, sehingga penelusuran citra yang

salah diklasifikasikan menjadi arah pengembangan lanjutan.

3.4 Analisis ROC Curve

Kurva ROC pada Gambar 7 berada dekat dengan sudut kiri atas, yang menunjukkan bahwa model mampu membedakan kedua kelas pada berbagai nilai ambang klasifikasi. Nilai AUC setelah fine-tuning adalah 0,9928, naik dari 0,9917 pada tahap feature extraction.

AUC menyatakan peluang bahwa model memberi skor lebih tinggi kepada citra Recyclable yang dipilih acak daripada citra Organic yang dipilih acak, sehingga nilai 0,9928 menunjukkan kemampuan diskriminasi yang sangat baik dan tidak bergantung pada satu nilai ambang tertentu.



Gambar 7. ROC Curve MobileNetV2

3.5 Studi Ablasi

Studi ablasi membandingkan tiga skenario dengan konfigurasi pelatihan yang sama (Tabel VI).

TABEL VI. HASIL STUDI ABLASI

Skenario	Tahap	Akurasi validasi	Akurasi uji	F1-score uji	AUC uji
Augmentasi + stratified (utama)	Feature extraction	95,75%	95,66%	95,66%	0,9917
	Fine-tuning	96,40%	96,22%	96,22%	0,9928
Augmentasi + split acak	Feature extraction	95,54%	95,89%	95,89%	0,9904
	Fine-tuning	95,89%	96,28%	96,28%	0,9917
Stratified tanpa augmentasi	Feature extraction	95,48%	95,16%	95,15%	0,9906
	Fine-tuning	95,72%	95,63%	95,63%	0,9925

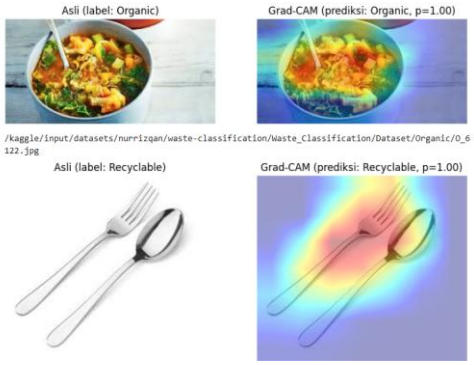
Pengaruh fine-tuning. Fine-tuning meningkatkan akurasi uji pada ketiga skenario: +0,56 poin (utama), +0,39 poin (split acak), dan +0,47 poin (tanpa augmentasi). Konsistensi arah ini memperkuat temuan bahwa fine-tuning memberi peningkatan kecil namun berulang.

Pengaruh augmentasi. Dibandingkan skenario tanpa augmentasi, augmentasi meningkatkan akurasi uji sebesar 0,59 poin setelah fine-tuning (96,22% berbanding 95,63%) dan 0,50 poin setelah feature extraction. Augmentasi juga mengurangi overfitting. Pada epoch ke-10 tahap feature extraction, selisih akurasi latih dan validasi mencapai 2,62 poin tanpa augmentasi (98,40% berbanding 95,78%), tetapi hanya 0,22 poin dengan augmentasi (95,73% berbanding 95,51%). Tanpa augmentasi, val_loss juga mulai naik setelah epoch ke-6 (dari 0,1170 menjadi 0,1448 pada epoch ke-8).

Pengaruh pembagian stratified. Pembagian stratified tidak memberikan perbedaan akurasi yang konsisten dibanding pembagian acak. Setelah fine-tuning, akurasi uji skenario stratified 0,06 poin lebih rendah (sekitar 2 citra), tetapi akurasi validasinya 0,51 poin lebih tinggi. Arahnya berlawanan pada kedua data sehingga tidak dapat disimpulkan bahwa stratified meningkatkan atau menurunkan akurasi. Hal ini dapat dipahami: dengan 3.385 data per subset dan proporsi kelas 55,69% : 44,31%, pembagian acak pun menghasilkan proporsi yang mendekati dataset asli (simpangan baku proporsi sekitar 0,85 poin persentase). Kontribusi stratified pada penelitian ini karena itu terletak pada jaminan bahwa proporsi kelas identik pada seluruh subset, sehingga hasil evaluasi tidak dipengaruhi variasi komposisi kelas. Manfaat ini kemungkinan lebih terasa pada dataset yang lebih kecil atau lebih tidak seimbang.

3.6 Analisis Grad-Cam

Sebagai penerapan explainable artificial intelligence, penelitian ini menggunakan Grad-CAM untuk memvisualisasikan bagian citra yang menjadi fokus perhatian model saat klasifikasi [7]. Dua contoh diambil dari data uji, yaitu citra pertama pada tiap kelas tanpa seleksi tambahan.



Gamba8. Perubahan Gambar Asli vs Grad-Cam

Pada Gambar 8 (kelas Organic, semangkuk sup), aktivasi tinggi (merah-kuning) terkonsentrasi pada isi mangkuk berupa sayuran dan pasta, sedangkan tepi mangkuk dan latar memiliki aktivasi rendah hingga sedang. Model memprediksi kelas Organic dengan probabilitas 1,00. Pada Gambar 9 (kelas Recyclable, sendok dan garpu), aktivasi tinggi berada pada kedua alat makan, terutama di antara kepala garpu dan sendok, sedangkan latar putih hampir tidak berkontribusi. Model memprediksi kelas Recyclable dengan probabilitas 1,00.

Kedua contoh menunjukkan bahwa keputusan model didasarkan pada objek utama dan bukan latar belakang. Namun, dua contoh bersifat ilustratif dan belum menggambarkan seluruh perilaku model, terutama pada citra yang salah diklasifikasikan

3.7 Perbandingan dengan Penelitian Terdahulu

Untuk mengetahui posisi penelitian ini terhadap penelitian sebelumnya, dilakukan perbandingan berdasarkan model yang digunakan dan akurasi yang diperoleh.

TABEL VI. PERBANDINGAN HASIL PENELITIAN

Penelitian	Model	Akurasi/Hasil
Ma *et al.* (2023) [5]	MobileNetV2 (*Transfer Learning*)	**82,92%**
Chhabra *et al.* (2024)[6]	Improved CNN	**93,28%**
Islam *et al.* (2025) [8]	DenseNet201 + Parallel CNN + SE Attention + Grad-CAM	**97,43%**
Nguyen *et al.* (2026) [9]	Customized ResNet50 + Hybrid Deep Learning + Machine Learning	**99,72%**
Gürcan & Soylu (2026) [7]	Transfer Learning (ConvNeXt, MobileNetV2, EfficientNet, ResNet, dll.)	**ConvNeXt_V1: 96,95%; MobileNetV2: 93,90%**
Penelitian ini	**MobileNetV2 + Data Augmentation + Fine-Tuning + Grad-CAM**	**96,22%**

Berdasarkan Tabel VI, model yang diusulkan memperoleh akurasi yang lebih tinggi dibandingkan beberapa penelitian terdahulu. Peningkatan performa tersebut diduga dipengaruhi oleh kombinasi *Transfer Learning*, *Data Augmentation*, dan proses *Fine-Tuning* yang mampu meningkatkan kemampuan model dalam mengenali variasi objek. Selain itu, penggunaan Grad-CAM memberikan keunggulan tambahan berupa interpretabilitas model yang belum banyak dibahas pada penelitian sebelumnya.

4. KESIMPULAN DAN SARAN

Penelitian ini mengembangkan model klasifikasi citra sampah organik dan recyclable menggunakan MobileNetV2 dengan transfer learning dua tahap (feature extraction dan fine-tuning), data augmentation, dan Grad-CAM. Model dilatih dan dievaluasi pada Waste Classification Dataset (22.564 citra) dengan pembagian stratified 70:15:15, sehingga proporsi kelas (55,69% Organic dan 44,31% Recyclable) identik pada data latih (15.794 citra), validasi (3.385 citra), dan uji (3.385 citra).

Fine-tuning pada 30 layer terakhir backbone meningkatkan akurasi data uji dari 95,66% menjadi 96,22%, dan AUC dari 0,9917 menjadi 0,9928. Model akhir mencapai accuracy, precision, recall, dan F1-score (weighted) masing-masing 96,22% pada data uji, dengan 128 kesalahan dari 3.385 citra. Akurasi data uji hanya 0,18 poin di bawah akurasi validasi (96,40%), sehingga model relatif stabil pada data yang belum pernah dilihat. Kinerja pada kedua kelas seimbang, dengan F1-score 96,61% untuk Organic dan 95,73% untuk Recyclable.

Kontribusi tiap komponen dapat diringkas berdasarkan studi ablasi. Pertama, augmentasi data meningkatkan akurasi uji sebesar 0,59 poin persentase (96,22% berbanding 95,63% tanpa augmentasi) dan menekan overfitting: pada epoch ke-10, selisih akurasi latih dan validasi tanpa augmentasi mencapai 2,62 poin, sedangkan dengan augmentasi hanya 0,22 poin.

Kedua, fine-tuning memberikan peningkatan sebesar 0,56 poin pada skenario utama dan konsisten positif pada kedua skenario ablasi (+0,39 dan +0,47 poin). Ketiga, pembagian stratified tidak menghasilkan perbedaan akurasi yang konsisten dibandingkan pembagian acak (96,22% berbanding 96,28% pada data uji), sehingga penelitian ini tidak membuktikan bahwa stratifikasi meningkatkan akurasi. Kontribusi stratified terletak pada validitas evaluasi: proporsi kelas identik pada seluruh subset, tanpa tumpang tindih data antar subset, dengan augmentasi yang hanya aktif pada data latih, sehingga metrik yang dilaporkan mencerminkan performa sebenarnya dan eksperimen dapat direplikasi.

Grad-CAM memperlihatkan bahwa model mendasarkan keputusannya pada objek utama dalam citra, bukan latar belakang, pada kedua contoh yang dianalisis. Secara keseluruhan, MobileNetV2 dengan fine-tuning dan data augmentation merupakan pendekatan yang efektif dan dapat diinterpretasikan untuk klasifikasi citra sampah organik dan recyclable, dengan akurasi yang kompetitif menggunakan arsitektur ringan (sekitar 2,4 juta parameter), sehingga berpotensi mendukung pengembangan sistem pengelolaan sampah cerdas. Kesimpulan ini dibatasi oleh ablasi yang dijalankan dengan satu seed acak, satu dataset dengan dua kelas, dan analisis Grad-CAM yang masih kualitatif.

Penelitian selanjutnya dapat diarahkan pada beberapa hal berikut. Pertama, studi ablasi dapat diulang dengan beberapa seed acak dan dilaporkan dalam bentuk rata-rata dan simpangan baku, agar selisih kecil antar skenario dapat dinilai secara statistik. Pembagian stratified juga perlu diuji pada data yang lebih kecil atau tidak seimbang, karena manfaatnya kemungkinan lebih besar pada kondisi tersebut.

Kedua, dataset dapat diperluas ke lebih banyak kelas, misalnya klasifikasi berdasarkan jenis material (plastik, kertas, kaca, logam, dan organik), serta memasukkan citra dari kondisi lapangan yang lebih beragam sehingga model dapat diterapkan pada skenario pengelolaan sampah yang lebih kompleks.

Ketiga, arsitektur yang lebih mutakhir seperti EfficientNetV2, ConvNeXt, atau Vision Transformer (ViT) dapat dievaluasi dengan protokol yang sama untuk melihat sejauh mana performanya melampaui MobileNetV2, dengan mempertimbangkan kompromi antara akurasi dan kompleksitas.

Keempat, analisis interpretabilitas dapat diperluas dengan menerapkan Grad-CAM pada citra yang salah diklasifikasikan untuk menelusuri penyebab kesalahan, serta membandingkannya dengan metode XAI lain seperti Score-CAM, Eigen-CAM, atau SHAP agar interpretasi model lebih komprehensif.

Kelima, implementasi model pada perangkat edge computing atau perangkat bergerak (mobile devices) dapat diteliti untuk mendukung klasifikasi secara real-time, termasuk pengukuran waktu inferensi dan ukuran model.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Magister Teknik Informatika, Universitas Amikom Yogyakarta, atas dukungan akademik selama pelaksanaan penelitian ini. Penulis juga menyampaikan apresiasi kepada penyedia **Waste Classification Dataset** pada platform Kaggle yang telah menyediakan dataset publik sebagai sumber data penelitian. Seluruh pihak yang telah memberikan dukungan, masukan, dan bantuan selama proses penelitian hingga penyusunan artikel ini turut penulis sampaikan penghargaan dan terima kasih.

Daftar Pustaka

[1] R. Kutho and D. Kota, “Optimalisasi Pengelolaan Sampah: Analisis Bank Sampah Resik Kutho Di Kota Surakarta 1,2,3,” vol. 11, no. 1, 2024.

[2] R. Sistem, “Optimasi Akurasi Metode Convolutional Neural Network untuk,” vol. 1, no. 10, pp. 11–12, 2021.

[3] S. Kornblith, J. Shlens, and Q. V Le, “Do Better ImageNet Models Transfer Better ?”.

[4] T. Otak, M. Cnn, and P. Citra, “(Performance Evaluation of Cnn Models With Various Data Augmentation,” *Teknimedia*, pp. 219–227.

[5] L. Yong, L. Ma, D. Sun, and L. Du, “Application of MobileNetV2 to waste classification,” *PLoS One*, vol. 18, no. 3 March, pp. 1–16, 2023, doi: 10.1371/journal.pone.0282336.

[6] M. Chhabra, B. Sharan, M. Elbarachi, and M. Kumar, “Intelligent waste classification approach based on improved multi-layered convolutional neural network,” *Multimed. Tools Appl.*, vol. 83, no. 36, pp. 84095–84120, 2024, doi: 10.1007/s11042-024-18939-w.

[7] F. Gurcan and A. Soylu, “Automated waste classification for smart recycling: A multi-

class CNN approach with transfer learning and pre-trained models," *Environ. Technol. Innov.*, vol. 41, no. November 2025, p. 104673, 2026, doi: 10.1016/j.eti.2025.104673.

[8] M. Islam, S. M. M. Hasan, and R. Hossain, "Towards sustainable solutions : Effective waste classification framework via enhanced deep convolutional neural networks," pp. 1–49, 2025, doi: 10.1371/journal.pone.0324294.

[9] N. B. Q. Nguyen, T. M. Do, C. T. Phan, and T. T. H. Phan, "Towards accurate and efficient waste image classification: A hybrid deep learning and machine learning approach," *Ain Shams Engineering Journal*, vol. 17, no. 4, p. 104062, 2026, doi: 10.1016/j.asej.2026.104062.

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM : Visual Explanations from Deep Networks via Gradient-based Localization," pp. 618–626, 2017, doi: 10.1109/ICCV.2017.74.

[11] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, 2019, doi: 10.1186/s40537-019-0197-0.

[12] Z. Zuraidah, L. N. Rosyidah, and R. F. Zulfi, "Edukasi Pengelolaan Dan Pemanfaatan Sampah Anorganik Di Mi Al Munir Desa Gadungan Kecamatan Puncu Kabupaten Kediri," *Budimas : Jurnal Pengabdian Masyarakat*, vol. 4, no. 2, pp. 1–6, 2022, doi: 10.29040/budimas.v4i2.6547.

[13] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.

[14] M. Sandler, M. Zhu, and A. Zhmoginov, "MobileNetV2 : Inverted Residuals and Linear Bottlenecks," 2018, doi: 10.1109/CVPR.2018.00474.

[15] I. Nabillah and I. Ranggadara, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," *JOINS (Journal of Information System)*, vol. 5, no. 2, pp. 250–255, 2020, doi: 10.33633/joins.v5i2.3900.