JCOSIS (*Journal Computer Science and Information Syetem*)
E-ISSN : 3032-5676
Volume : 03 | Nomor : 02 | Tahun 2026 | Hal 96-105
DOI : https://doi.org/10.61567

Klasifikasi Sampah Non-Organik Menggunakan EfficientNet-B0 dengan Visualisasi Grad-CAM

Erwin Dwika Putra[1]*, Marissa Utami[2]
[1]Program Studi Informatika, Fakultas Teknik,Universitas Muhammadiyah Bengkulu, Indonesia
[2]Program Studi Sistem Informasi, Fakultas Teknik,Universitas Muhammadiyah Bengkulu, Indonesia
[1] erwindwikap@gmail.com ; [2]marissautami@umb.ac.id

Abstrak.
Tujuan : Penelitian ini mengevaluasi EfficientNet-B0 untuk klasifikasi sampah non-organik dan menganalisis hasil prediksinya menggunakan Grad-CAM.
Metode/Design/Pendekatan: Dataset TACO digunakan untuk mengklasifikasikan lima kategori material, yaitu plastik, logam, kaca, kertas/karton, dan busa/*styrofoam*. Objek diekstraksi menggunakan anotasi *bounding box*. EfficientNet-B0 diterapkan melalui transfer learning dan dievaluasi menggunakan *accuracy*, *precision*, *recall*, serta *F1-score*. *Visualisasi Grad-CAM* digunakan untuk mengamati area citra yang berkontribusi terhadap prediksi.
Hasil/Temuan: Pada 323 objek uji, model memperoleh akurasi 54,49%, *macro precision* 44,19%, macro recall 53,27%, *macro F1-score* 45,75%, dan *weighted F1-score* 56,99%. Kelas plastik memperoleh *F1-score* tertinggi (0,66), sedangkan kaca terendah (0,25). Grad-CAM menunjukkan perbedaan pola visual antara prediksi benar dan salah.
Kebaharuan/Originalitas/Nilai: Penelitian ini menggabungkan evaluasi EfficientNet-B0 dan analisis Grad-CAM untuk mengkaji kinerja klasifikasi serta bukti visual dalam klasifikasi sampah non-organik.

Keywords: klasifikasi sampah non-organik; EfficientNet-B0; *transfer learning*; Grad-CAM; TACO.

Abstract.
Purpose: *This study evaluates EfficientNet-B0 for non-organic waste classification and analyzes its predictions using Grad-CAM.*
Methods/Study design/approach: *The TACO dataset was used to classify five material categories: plastic, metal, glass, paper/cardboard, and foam/styrofoam. Objects were extracted using bounding box annotations. EfficientNet-B0 was implemented through transfer learning and evaluated using accuracy, precision, recall, and F1-score. Grad-CAM visualizations were used to examine image regions contributing to predictions.*
Result/Findings: *On 323 test objects, the model achieved 54.49% accuracy, 44.19% macro precision, 53.27% macro recall, 45.75% macro F1-score, and 56.99% weighted F1-score. Plastic achieved the highest F1-score (0.66), while glass had the lowest (0.25). Grad-CAM revealed differences between correct and incorrect predictions.*
Novelty/Originality/Value: *This study combines EfficientNet-B0 evaluation with Grad-CAM analysis to examine classification performance and visual evidence in non-organic waste classification.*

Keywords*: non-organic waste classification; EfficientNet-B0; transfer learning; Grad-CAM; TACO*

Article history:

Received, 2026-10-20
Revised, 2026-10-23
Accepted, 2026-10-28

*Corresponding author.
Erwin Dwika Putra.
Email addresses: erwindwikap@gmail.com

This is an open access article under the CC BY-SAlicense.

PENDAHULUAN

Pengelolaan sampah merupakan salah satu permasalahan lingkungan yang membutuhkan pendekatan teknologi untuk meningkatkan efektivitas proses pemilahan dan daur ulang. Sampah non-organik, seperti plastik, kaca, kaleng, dan logam, memiliki karakteristik visual yang beragam sehingga proses identifikasi secara manual berpotensi menimbulkan kesalahan, terutama ketika jumlah sampah meningkat dan kondisi objek tidak seragam. Perkembangan *computer vision* dan *deep learning* memberikan peluang untuk mengembangkan sistem klasifikasi sampah berbasis citra yang mampu mempelajari karakteristik visual secara otomatis. Sejumlah penelitian telah mengkaji penerapan *deep learning* untuk klasifikasi, deteksi, dan identifikasi material sampah, termasuk penggunaan CNN untuk pengenalan sampah yang dapat didaur ulang [1], [2], [3], [4]. Kajian tersebut menunjukkan bahwa pemanfaatan citra dan model pembelajaran mendalam menjadi salah satu pendekatan yang relevan untuk mendukung otomatisasi pemilahan sampah.

Convolutional Neural Network (CNN) merupakan salah satu arsitektur yang banyak digunakan dalam klasifikasi citra karena mampu mempelajari fitur visual secara bertingkat. Perkembangan CNN untuk pengenalan citra didukung oleh berbagai arsitektur, mulai dari AlexNet, yang memperlihatkan keberhasilan jaringan konvolusional berskala besar dalam klasifikasi citra [5], hingga VGGNet dan ResNet yang memperkenalkan pendekatan berbeda untuk meningkatkan representasi fitur dan pelatihan jaringan yang lebih dalam [6], [7]. Perkembangan tersebut menunjukkan bahwa pemilihan arsitektur dapat memengaruhi kemampuan model dalam mempelajari pola visual yang kompleks. Namun, klasifikasi sampah memiliki tantangan tersendiri karena beberapa objek dapat mempunyai bentuk, warna, atau tekstur yang mirip. Selain itu, citra sampah di lingkungan nyata dapat memuat latar belakang yang kompleks, perubahan pencahayaan, dan objek yang saling bertumpuk. Oleh karena itu, pemilihan model perlu mempertimbangkan kemampuan klasifikasi sekaligus efisiensi komputasinya.

Salah satu arsitektur yang dapat digunakan adalah EfficientNet, yang diperkenalkan oleh Tan dan Le [8]. EfficientNet menerapkan pendekatan *compound scaling* dengan menyeimbangkan penskalaan kedalaman jaringan, lebar jaringan, dan resolusi citra. Pendekatan tersebut dikembangkan untuk memperoleh keseimbangan antara akurasi dan kebutuhan komputasi. EfficientNet-B0 merupakan arsitektur dasar dalam keluarga EfficientNet dan dapat digunakan sebagai model klasifikasi citra dengan pendekatan *transfer learning*. Secara umum, *transfer learning* memungkinkan pengetahuan dari model yang telah dilatih pada dataset berskala besar dimanfaatkan untuk menyelesaikan tugas baru yang memiliki data lebih terbatas [9]. Dataset ImageNet yang diperkenalkan oleh Deng et al. [10] menjadi salah satu sumber data penting dalam perkembangan model pengenalan citra dan pelatihan awal jaringan visual. Dengan memanfaatkan bobot pralatih, EfficientNet-B0 dapat disesuaikan terhadap kategori sampah yang ditentukan dalam penelitian. Meskipun demikian, kinerja model pada klasifikasi sampah tetap perlu diukur secara empiris karena karakteristik data target berbeda dari dataset pralatih.

Dataset merupakan aspek penting dalam pengembangan model klasifikasi sampah. Proença dan Simões memperkenalkan *Trash Annotations in Context* (TACO), yaitu dataset citra sampah yang dikumpulkan dari beragam konteks lingkungan dan dilengkapi anotasi objek untuk mendukung tugas deteksi serta segmentasi [11]. Berbeda dari dataset klasifikasi yang telah disusun dalam folder berdasarkan kelas, TACO menggunakan anotasi objek yang perlu diproses sebelum dapat digunakan untuk klasifikasi citra. Oleh karena itu, penggunaan TACO dalam penelitian ini memerlukan pemetaan kategori, penentuan objek yang memenuhi kriteria sampah non-organik, serta penyiapan citra yang konsisten dengan labelnya. Kategori akhir harus disesuaikan dengan anotasi yang benar-benar tersedia dan jumlah sampel setiap kategori. Penjelasan mengenai pemetaan kelas, penanganan objek yang memiliki lebih dari satu anotasi, serta pembagian data diperlukan agar eksperimen dapat direplikasi. Kajian mengenai penggunaan CNN untuk identifikasi sampah juga menekankan pentingnya ketersediaan dataset dan konsistensi evaluasi dalam membandingkan model [1], [2], [12].

Meskipun CNN dan EfficientNet dapat menghasilkan prediksi klasifikasi, proses pengambilan keputusan internalnya tidak selalu mudah dipahami. Model dapat memprediksi kelas sampah dengan tingkat keyakinan tinggi, tetapi keluaran tersebut tidak secara langsung menjelaskan bagian citra yang berkontribusi terhadap prediksi. Permasalahan ini berkaitan dengan *black-box model*, yaitu kondisi ketika mekanisme internal model sulit diterjemahkan menjadi penjelasan yang mudah dipahami pengguna. *Explainable Artificial Intelligence* (XAI) berkembang untuk membantu menjelaskan perilaku model dan meningkatkan transparansi hasil prediksi [13], [14]. Pendekatan interpretabilitas dapat dibedakan berdasarkan cara kerjanya, termasuk metode yang menjelaskan prediksi model secara lokal dan metode yang menghasilkan penjelasan berbentuk visual. Ribeiro et al. memperkenalkan LIME sebagai metode untuk menjelaskan prediksi model melalui pendekatan lokal yang dapat diterapkan pada berbagai jenis model [15]. Sementara itu, Lundberg dan Lee memperkenalkan kerangka SHAP yang menggunakan nilai *Shapley* untuk menghubungkan kontribusi fitur dengan keluaran prediksi [16]. Dalam klasifikasi citra, pendekatan visual menjadi relevan karena penjelasan dapat disajikan langsung pada area gambar yang berhubungan dengan prediksi.

Salah satu metode XAI untuk jaringan konvolusional adalah *Gradient-weighted Class Activation Mapping* (Grad-CAM), yang diperkenalkan oleh Selvaraju et al. [17]. Grad-CAM menggunakan gradien skor kelas terhadap aktivasi pada lapisan konvolusional untuk membentuk peta aktivasi yang menyoroti area citra yang berkontribusi terhadap prediksi kelas tertentu. Peta tersebut dapat ditampilkan sebagai *heatmap* dan ditumpangkan pada citra asli sehingga area yang mendapat perhatian model dapat diamati secara visual. Pengembangan terkait, yaitu Grad-CAM++, menawarkan pendekatan pembobotan gradien yang dirancang untuk menghasilkan lokalisasi visual yang lebih baik pada kondisi tertentu [18]. Berbagai kajian XAI juga membahas peran penjelasan visual dalam membantu analisis model, termasuk manfaat dan keterbatasannya [14], [19], [20]. Meskipun demikian, Grad-CAM tidak dapat diperlakukan sebagai bukti mutlak bahwa model memahami objek secara semantik. Area yang disorot merupakan indikasi kontribusi terhadap skor prediksi dan perlu ditafsirkan dengan mempertimbangkan citra asli, kelas target, serta kemungkinan pengaruh latar belakang.

Berdasarkan permasalahan tersebut, penelitian ini menerapkan EfficientNet-B0 untuk klasifikasi sampah non-organik dan menggunakan Grad-CAM untuk menyajikan penjelasan visual terhadap hasil klasifikasi. Tahapan penelitian meliputi penyiapan dataset dan pemetaan kategori, prapemrosesan citra, pelatihan model, evaluasi kinerja klasifikasi, serta visualisasi Grad-CAM. Kinerja model dievaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*, serta *confusion matrix* untuk mengidentifikasi pola kesalahan antarkelas. Visualisasi Grad-CAM dilakukan pada sampel prediksi benar dan salah untuk mengamati bagian citra yang mendapat aktivasi tinggi dan menilai secara kualitatif apakah area tersebut berhubungan dengan objek sampah yang diklasifikasikan. Analisis ini diharapkan dapat memberikan informasi tambahan yang tidak tercakup dalam metrik kinerja numerik saja.

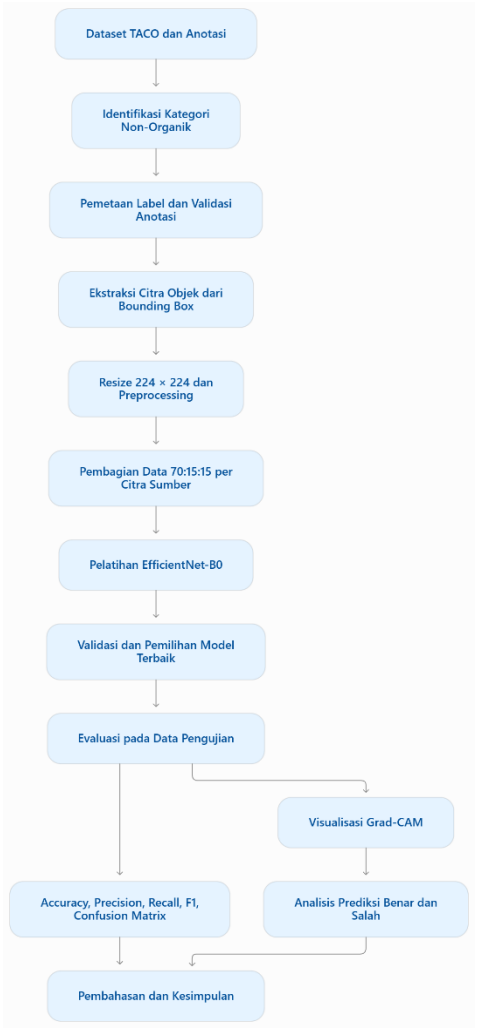
Kontribusi penelitian ini adalah penerapan EfficientNet-B0 yang disertai pemeriksaan visual menggunakan Grad-CAM pada klasifikasi sampah non-organik berbasis citra. Penelitian tidak hanya mengevaluasi hasil klasifikasi, tetapi juga mengamati pola perhatian visual model pada beberapa prediksi untuk membantu mengidentifikasi kemungkinan kesalahan atau ketergantungan terhadap latar belakang. Dengan demikian, penelitian ini diharapkan memberikan gambaran awal mengenai kinerja dan perilaku visual model dalam mengenali kategori sampah non-organik. Hasilnya dapat menjadi dasar bagi penelitian lanjutan yang menguji interpretabilitas secara kuantitatif, misalnya dengan mengukur kesesuaian area aktivasi terhadap anotasi objek yang tersedia.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen kuantitatif untuk mengembangkan model klasifikasi sampah non-organik menggunakan EfficientNet-B0 dan menganalisis hasil prediksinya melalui visualisasi *Gradient-weighted Class Activation Mapping* (Grad-CAM). Tahapan penelitian meliputi pengumpulan dataset, pemetaan kelas, prapemrosesan citra, pembagian dataset, pelatihan model, evaluasi kinerja klasifikasi, dan visualisasi Grad-CAM. Secara keseluruhan, alur penelitian dirancang untuk mengevaluasi kemampuan model dalam mengklasifikasikan sampah non-organik sekaligus memberikan penjelasan visual mengenai area citra yang berkontribusi terhadap prediksi.

Alur penelitian diawali dengan pengumpulan dataset *Trash Annotations in Context* (TACO) beserta anotasi objeknya. Dataset TACO memuat citra sampah dalam beragam kondisi lingkungan dan menyediakan anotasi yang mendukung tugas deteksi serta segmentasi objek [11]. Selanjutnya, dilakukan identifikasi kategori sampah non-organik dan pemetaan label anotasi ke dalam kelas klasifikasi yang digunakan.

Setelah pemetaan selesai, objek diekstraksi berdasarkan koordinat *bounding box*, kemudian dilakukan prapemrosesan citra melalui penyesuaian ukuran dan normalisasi. Dataset yang telah diproses dibagi menjadi data pelatihan, validasi, dan pengujian. Data pelatihan digunakan untuk membangun model EfficientNet-B0, data validasi digunakan untuk memilih model terbaik, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model. Tahap akhir adalah visualisasi Grad-CAM dan analisis hasil prediksi, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1 Alur Penelitian

Dataset TACO dipilih karena menyediakan citra sampah dengan latar belakang dan kondisi lingkungan yang beragam. Berbeda dengan dataset klasifikasi yang telah tersusun berdasarkan kelas, TACO menggunakan anotasi objek yang perlu diproses terlebih dahulu. Oleh karena itu, identifikasi kelas dilakukan dengan memeriksa kategori yang tersedia pada berkas anotasi dan menentukan kategori yang sesuai dengan ruang lingkup sampah non-organik. Kategori target dapat mencakup plastik, logam, kaca, serta kertas atau kardus apabila label tersebut tersedia dan jumlah sampelnya memadai. Pemetaan label dilakukan secara konsisten dengan mempertahankan hubungan antara objek, label asli, dan citra sumber. Jumlah objek pada setiap kelas dicatat untuk mengetahui distribusi data dan menentukan kebutuhan penanganan ketidakseimbangan kelas.

Setelah kelas ditetapkan, setiap objek diekstraksi dari citra sumber berdasarkan koordinat *bounding box* pada anotasi TACO. Apabila anotasi dinyatakan dalam format ((x,y,w,h)), objek dipotong berdasarkan titik awal ((x,y)), lebar (w), dan tinggi (h), dengan memastikan koordinat potongan tidak melampaui batas citra. Sampel dengan anotasi tidak valid, potongan citra kosong, atau label yang tidak dapat dipetakan secara konsisten dikeluarkan berdasarkan kriteria yang ditetapkan. Citra hasil ekstraksi kemudian diubah ukurannya menjadi 224 × 224 piksel agar sesuai dengan konfigurasi masukan EfficientNet-B0. Selanjutnya, citra diproses menggunakan fungsi *preprocessing* yang sesuai dengan implementasi model yang digunakan untuk menghindari normalisasi berulang. Augmentasi data, seperti pembalikan horizontal, rotasi kecil, atau variasi pencahayaan yang wajar, hanya diterapkan pada data pelatihan, sedangkan data validasi dan pengujian tidak diberi augmentasi acak.

Dataset dibagi menjadi data pelatihan sebesar 70%, data validasi sebesar 15%, dan data pengujian sebesar 15%. Pembagian dilakukan berdasarkan identitas citra sumber, bukan berdasarkan potongan objek secara individual. Dengan demikian, apabila satu citra mengandung beberapa objek, seluruh objek dari citra tersebut ditempatkan pada subset yang sama. Strategi ini bertujuan mengurangi risiko kebocoran data akibat kemiripan latar belakang atau karakteristik visual dari citra sumber yang sama. Pembagian juga mempertimbangkan distribusi kelas agar setiap subset memiliki sampel yang memadai. Jumlah citra sumber dan objek pada masing-masing subset dicatat untuk dilaporkan dalam hasil penelitian.

Klasifikasi dilakukan menggunakan EfficientNet-B0 dengan pendekatan *transfer learning*. EfficientNet menggunakan strategi *compound scaling* yang menyeimbangkan kedalaman jaringan, lebar jaringan, dan resolusi citra [8]. Bobot pralatih ImageNet digunakan sebagai titik awal pembelajaran, sedangkan lapisan klasifikasi diganti agar jumlah keluaran sesuai dengan jumlah kelas sampah non-organik. Pelatihan dilakukan dalam dua tahap, yaitu melatih lapisan klasifikasi baru dengan membekukan lapisan ekstraksi fitur pralatih, kemudian melakukan *fine-tuning* pada sebagian lapisan akhir jaringan menggunakan laju pembelajaran yang lebih kecil.

Untuk klasifikasi dengan (C) kelas, model menghasilkan skor bagi setiap kelas yang dikonversi menjadi probabilitas menggunakan fungsi softmax:

$$p_c = \frac{e^{zc}}{\sum_{j=1}^{C} e^{zc}} \qquad (1)$$

dengan (p_c) adalah probabilitas kelas ke-(c), (z_c) adalah skor atau logit kelas ke-(c), dan (C) adalah jumlah kelas. Prediksi akhir ditentukan berdasarkan kelas dengan probabilitas tertinggi. Fungsi kerugian yang digunakan adalah *categorical cross-entropy* untuk label berbentuk *one-hot encoding*, yang dirumuskan sebagai berikut:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \ln(p_{ic}) \qquad (2)$$

dengan (L) adalah nilai fungsi kerugian, (N) adalah jumlah sampel, (y_{ic}) bernilai 1 jika sampel ke-(i) termasuk kelas (c) dan 0 jika tidak, serta (p_{ic}) adalah probabilitas prediksi sampel ke-(i) untuk kelas (c). Jika label direpresentasikan sebagai bilangan bulat, implementasi dapat menggunakan *sparse categorical cross-entropy* dengan prinsip perhitungan yang setara. *Optimizer*, laju pembelajaran, ukuran *batch*, dan jumlah epoch ditetapkan dalam konfigurasi eksperimen. Model terbaik dipilih berdasarkan kinerja pada data validasi, sedangkan data pengujian tidak digunakan untuk menentukan parameter model. Evaluasi dilakukan menggunakan data pengujian dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, *confusion matrix* digunakan untuk mengidentifikasi kelas yang sering tertukar. Dalam klasifikasi multikelas, (TP), (FP), dan (FN) dihitung untuk setiap kelas menggunakan pendekatan *one-vs-rest*. Nilai *precision*, *recall*, dan *F1-score* kemudian dirata-ratakan dengan pendekatan *macro average* agar setiap kelas memiliki bobot yang sama. Perhitungan dilakukan menggunakan persamaan berikut:

$$MacroPrecision = \frac{1}{C}\sum_{c=1}^{C} Precision_c \qquad (3)$$

$$MacroRecall = \frac{1}{C}\sum_{c=1}^{C} Recall_c \qquad (4)$$

$$MacroF1 = \frac{1}{C}\sum_{c=1}^{C} F1_c \qquad (5)$$

dengan (C) merupakan jumlah kelas. Confusion matrix disajikan dalam bentuk matriks berukuran (C\times C), dengan baris menunjukkan kelas aktual dan kolom menunjukkan kelas prediksi. Hasil evaluasi ini digunakan untuk menggambarkan kemampuan model dan menganalisis pola kesalahan klasifikasi antarkelas. Setelah evaluasi kinerja dilakukan, Grad-CAM digunakan untuk menghasilkan penjelasan visual terhadap prediksi EfficientNet-B0 [17]. Metode ini memanfaatkan gradien skor kelas terhadap aktivasi pada lapisan konvolusional untuk menghitung bobot kepentingan setiap kanal fitur. Bobot kanal ke-(k) untuk kelas target (c) dihitung dengan merata-ratakan gradien skor kelas terhadap seluruh posisi spasial:

$$\alpha_k^c = \frac{1}{z}\sum_{i=1}^{H}\sum_{j=1}^{W}\frac{\partial s^c}{\partial A_{ij}^k} \quad (6)$$

dengan (a_k^c) merupakan bobot kepentingan kanal ke-(k), (s^c) adalah skor kelas target sebelum penerapan softmax, ($A_{\{ij\}}^{\{k\}}$) adalah aktivasi kanal ke-(k) pada posisi ((i,j)), (H) dan (W) merupakan tinggi dan lebar peta aktivasi, serta ($Z = H\backslash times\ W$) adalah jumlah posisi spasial. Peta Grad-CAM kemudian dihitung dengan menggabungkan peta aktivasi menggunakan bobot kepentingannya dan menerapkan fungsi ReLU:

$$L_{\{Grad-CAM\}}^c = ReLU\left(\sum_k^k \alpha_k^c A \right) \quad (7)$$

Peta aktivasi yang dihasilkan dinormalisasi, diubah ukurannya agar sesuai dengan citra objek, kemudian ditumpangkan pada citra asli sebagai heatmap. Visualisasi dilakukan pada sampel pengujian yang mencakup prediksi benar dan salah dari setiap kelas. Analisis difokuskan pada apakah area aktivasi berkaitan dengan objek sampah, apakah model lebih memperhatikan latar belakang, dan bagaimana pola visual pada prediksi benar berbeda dari prediksi salah. Grad-CAM digunakan sebagai analisis kualitatif karena area yang disorot menunjukkan indikasi kontribusi terhadap skor prediksi, bukan bukti mutlak pemahaman semantik terhadap objek. Tahap akhir penelitian menggabungkan hasil evaluasi numerik dan analisis visual. Metrik klasifikasi dan confusion matrix digunakan untuk menjelaskan kinerja EfficientNet-B0, sedangkan visualisasi Grad-CAM memberikan informasi tambahan mengenai area citra yang berkontribusi terhadap prediksi. Untuk mendukung reprodusibilitas, penelitian mencatat versi dataset, pemetaan kelas, jumlah sampel, pembagian data, parameter pelatihan, serta perangkat keras dan versi pustaka yang digunakan. Seluruh hasil dilaporkan berdasarkan eksperimen aktual, sehingga kesimpulan mengenai performa model dan pola visualisasinya didasarkan pada data pengujian yang benar-benar digunakan.

HASIL DAN PEMBAHASAN

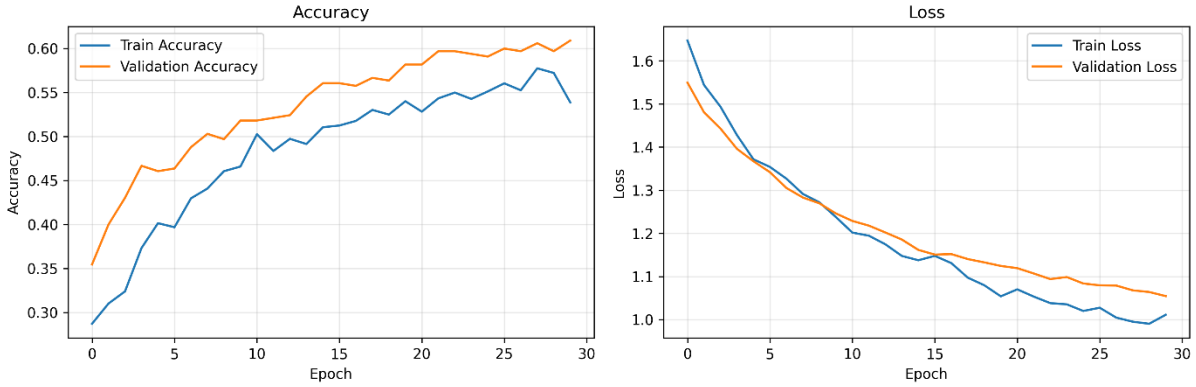
Penelitian ini mengimplementasikan EfficientNet-B0 dengan pendekatan transfer learning untuk mengklasifikasikan sampah non-organik berdasarkan karakteristik material serta menggunakan Grad-CAM untuk menginterpretasikan area citra yang berkontribusi terhadap prediksi model. Dataset penelitian berasal dari TACO (Trash Annotations in Context) yang memiliki anotasi objek berupa bounding box. Berdasarkan proses ekstraksi objek dan pemetaan kategori, diperoleh 2.697 objek hasil ekstraksi, dengan 2.175 objek berhasil dipetakan ke dalam lima kelas material, yaitu Plastik, Logam, Kertas/Karton, Kaca, dan Busa/Styrofoam. Sebanyak 522 objek lainnya tidak termasuk dalam kelas material yang digunakan pada eksperimen ini.

Pembagian dataset dilakukan berdasarkan citra sumber dengan proporsi 70% untuk data pelatihan, 15% untuk validasi, dan 15% untuk pengujian. Hasil pembagian menghasilkan 1.522 objek pada data pelatihan, 330 objek pada data validasi, dan 323 objek pada data pengujian. Berdasarkan citra sumber, sebanyak 694 citra digunakan untuk pelatihan, 149 citra untuk validasi, dan 149 citra untuk pengujian. Pemeriksaan pembagian data menunjukkan bahwa tidak terdapat citra sumber yang tersebar pada lebih dari satu subset. Strategi ini penting untuk mengurangi risiko kebocoran data (data leakage), khususnya karena satu citra sumber dapat memiliki lebih dari satu objek yang diekstraksi. Dengan demikian, objek dari citra sumber yang sama tidak digunakan secara bersamaan dalam proses pelatihan dan pengujian.

Distribusi kelas menunjukkan bahwa Plastik merupakan kelas dengan jumlah objek terbanyak, yaitu 1.190 objek, diikuti Logam sebanyak 388 objek, Kertas/Karton sebanyak 373 objek, Kaca sebanyak 142 objek, dan Busa/Styrofoam sebanyak 82 objek. Distribusi tersebut menunjukkan ketidakseimbangan kelas (class imbalance) yang cukup jelas. Kelas Plastik memiliki jumlah objek jauh lebih banyak dibandingkan kelas Busa/Styrofoam dan Kaca. Kondisi ini dapat memengaruhi proses pembelajaran karena model memperoleh lebih banyak contoh dari kelas tertentu dibandingkan kelas lainnya. Selain itu, karakteristik visual beberapa material dapat memiliki kemiripan, misalnya kemasan plastik dan karton atau benda kaca dan logam yang memiliki pantulan cahaya. Kombinasi ketidakseimbangan data dan kemiripan karakteristik visual tersebut menjadi tantangan dalam membangun model klasifikasi material.

Pelatihan model EfficientNet-B0 dilakukan selama maksimal 30 epoch dengan memantau perubahan akurasi dan loss pada data pelatihan serta validasi. Grafik pelatihan menunjukkan bahwa akurasi pelatihan secara umum meningkat dari sekitar 0,29 pada awal pelatihan hingga mencapai kisaran 0,55–0,58 pada epoch akhir sebelum mengalami penurunan pada epoch terakhir. Sementara itu, akurasi validasi menunjukkan kecenderungan meningkat dari sekitar 0,36 hingga mencapai sekitar 0,61 pada epoch terakhir. Pola ini menunjukkan bahwa model

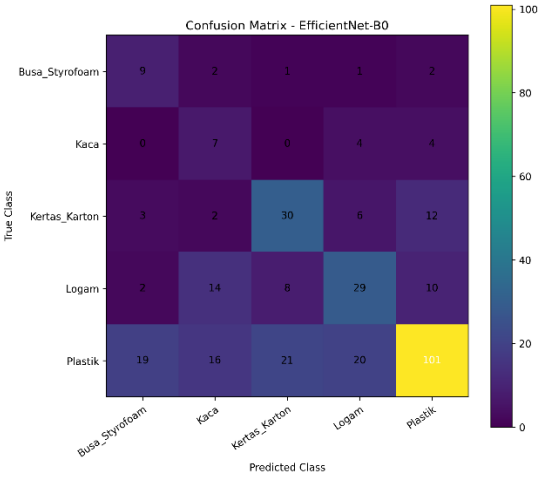
secara bertahap mempelajari karakteristik visual yang membedakan kelima kelas material, meskipun tingkat keberhasilan klasifikasinya masih terbatas.



Gambar 2 Grafik akurasi dan loss pelatihan

Grafik loss memperlihatkan penurunan yang relatif konsisten pada data pelatihan maupun validasi. Training loss menurun dari sekitar 1,65 menjadi sekitar 1,01, sedangkan validation loss menurun dari sekitar 1,55 menjadi sekitar 1,06. Penurunan kedua nilai tersebut menunjukkan bahwa kesalahan prediksi model selama proses pelatihan secara umum semakin berkurang. Selain itu, nilai validation loss tidak memperlihatkan kenaikan tajam menjelang akhir pelatihan, sehingga grafik yang tersedia belum menunjukkan indikasi kuat terjadinya overfitting pada fase tersebut. Meskipun demikian, perbedaan antara akurasi validasi dan hasil evaluasi test set menunjukkan bahwa kemampuan generalisasi model masih perlu diperhatikan. Penentuan model akhir sebaiknya tetap didasarkan pada model tersimpan dengan performa validasi terbaik, bukan semata-mata pada epoch terakhir.

Hasil evaluasi pada data pengujian menunjukkan bahwa model memperoleh test loss sebesar 1,1885 dan akurasi sebesar 54,49%. Nilai macro precision mencapai 44,19%, macro recall sebesar 53,27%, macro F1-score sebesar 45,75%, serta weighted F1-score sebesar 56,99%. Nilai akurasi menunjukkan bahwa model mengklasifikasikan sekitar 54,49% dari 323 objek pengujian secara benar. Sementara itu, nilai macro F1-score yang lebih rendah dibandingkan weighted F1-score mengindikasikan bahwa kinerja model tidak merata pada seluruh kelas. Hal ini berkaitan dengan perbedaan jumlah sampel antarkelas serta variasi kemampuan model dalam mengenali karakteristik visual setiap material.



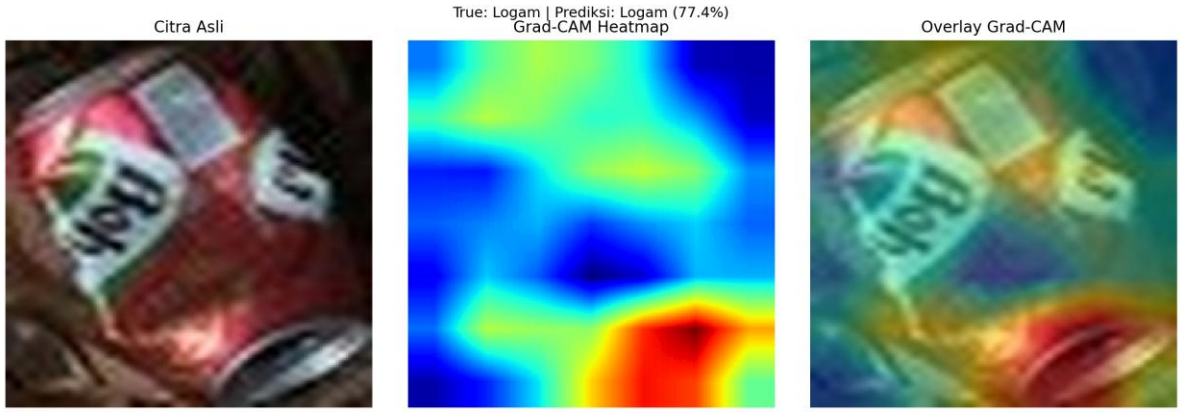
Gambar 3 Matriks konfusi EfficientNet-B0

Berdasarkan classification report, kelas Plastik memperoleh precision sebesar 0,78, recall sebesar 0,57, dan F1-score sebesar 0,66. Nilai precision yang relatif tinggi menunjukkan bahwa prediksi Plastik cukup tepat dibandingkan kelas lainnya, tetapi nilai recall sebesar 0,57 menunjukkan bahwa masih terdapat banyak objek Plastik yang tidak berhasil dikenali sebagai Plastik. Kelas Kertas/Karton memperoleh precision sebesar 0,50, recall sebesar 0,57, dan F1-score sebesar 0,53. Sementara itu, kelas Logam menghasilkan precision sebesar 0,48, recall sebesar 0,46, dan F1-score sebesar 0,47. Hasil tersebut menunjukkan bahwa kemampuan model dalam membedakan material logam dari kelas lainnya masih terbatas.

Kinerja terendah berdasarkan F1-score ditemukan pada kelas Kaca, yaitu 0,25, dengan precision sebesar 0,17 dan recall sebesar 0,47. Kelas Busa/Styrofoam memperoleh F1-score sebesar 0,38, dengan precision sebesar 0,27 dan recall sebesar 0,60. Nilai recall kedua kelas tersebut perlu diinterpretasikan secara hati-hati karena masing-masing hanya memiliki 15 sampel pada data pengujian. Jumlah sampel yang relatif sedikit menyebabkan hasil evaluasi per kelas lebih sensitif terhadap perubahan beberapa prediksi. Meskipun demikian, rendahnya nilai precision menunjukkan bahwa banyak objek yang diprediksi sebagai Kaca atau Busa/Styrofoam sebenarnya berasal dari kelas lain. Temuan ini memperlihatkan bahwa model masih mengalami kesulitan dalam menghasilkan prediksi yang konsisten untuk kelas dengan jumlah data terbatas.

Matriks konfusi memberikan gambaran lebih rinci mengenai pola kesalahan klasifikasi. Dari 177 objek Plastik pada data pengujian, sebanyak 101 objek berhasil diklasifikasikan dengan benar, sedangkan objek lainnya diprediksi sebagai Busa/Styrofoam sebanyak 19 objek, Kaca sebanyak 16 objek, Kertas/Karton sebanyak 21 objek, dan Logam sebanyak 20 objek. Pada kelas Kertas/Karton, sebanyak 30 dari 53 objek diklasifikasikan dengan benar, sementara 12 objek diprediksi sebagai Plastik. Kesalahan tersebut menunjukkan bahwa model belum sepenuhnya mampu membedakan karakteristik visual kemasan atau benda berbahan kertas dan plastik. Untuk kelas Logam, sebanyak 29 dari 63 objek diklasifikasikan dengan benar, sedangkan sebagian objek lainnya diprediksi sebagai Kaca, Kertas/Karton, atau Plastik. Kesalahan klasifikasi tersebut dapat berkaitan dengan kemiripan bentuk, warna, pantulan cahaya, dan kondisi objek dalam citra. Namun, faktor-faktor tersebut merupakan kemungkinan penjelasan yang masih perlu diperiksa melalui analisis citra lebih lanjut, bukan penyebab yang dapat dipastikan hanya dari matriks konfusi.

Visualisasi Grad-CAM digunakan untuk mengamati area citra yang berkontribusi terhadap prediksi EfficientNet-B0. Pada visualisasi yang dihasilkan, heatmap memperlihatkan variasi intensitas aktivasi, sedangkan overlay menggabungkan peta aktivasi dengan citra asli sehingga area yang berkontribusi terhadap prediksi dapat diamati secara visual. Warna dengan intensitas tinggi menunjukkan area yang memiliki kontribusi relatif besar terhadap skor kelas yang sedang dianalisis, sedangkan area berintensitas rendah menunjukkan kontribusi relatif lebih kecil. Interpretasi ini tidak berarti bahwa setiap area berwarna merah merupakan objek yang telah dikenali dengan benar, melainkan menunjukkan bagian citra yang memengaruhi skor prediksi model.



Gambar 4 Visualisasi Grad-CAM pada sampel uji

Pada beberapa sampel, model berhasil memprediksi kelas sesuai label sebenarnya. Sebagai contoh, salah satu objek Busa/Styrofoam diprediksi sebagai Busa/Styrofoam dengan confidence sebesar 34,0%, sedangkan salah satu objek Kaca diprediksi sebagai Kaca dengan confidence sebesar 77,4%. Salah satu objek Logam juga berhasil diklasifikasikan dengan confidence sebesar 77,4%. Visualisasi pada sampel-sampel tersebut memberikan gambaran mengenai distribusi aktivasi yang digunakan model ketika menghasilkan prediksi. Namun, tingkat confidence yang berbeda menunjukkan bahwa keyakinan model bervariasi antarsampel. Prediksi yang benar dengan confidence rendah tetap menunjukkan ketidakpastian dan tidak dapat disamakan dengan prediksi yang benar dengan confidence tinggi.

Visualisasi Grad-CAM juga memperlihatkan sejumlah kesalahan klasifikasi. Salah satu objek Plastik diprediksi sebagai Logam dengan confidence sebesar 72,0%, sedangkan objek Busa/Styrofoam lainnya diprediksi sebagai Kertas/Karton dengan confidence sebesar 35,8%. Pada contoh berbeda, objek Kaca diprediksi sebagai Logam dengan confidence sebesar 31,7%, dan objek Kertas/Karton diprediksi sebagai Plastik dengan confidence sebesar 46,5%. Contoh tersebut memperlihatkan bahwa model masih dapat menghasilkan prediksi yang tidak sesuai dengan kelas sebenarnya. Pada beberapa overlay, aktivasi juga tampak tersebar di area yang luas dan tidak sepenuhnya terkonsentrasi pada bagian objek yang paling mudah dikenali secara visual. Kondisi ini perlu ditelaah

lebih lanjut untuk mengetahui apakah model memanfaatkan karakteristik material yang relevan atau turut dipengaruhi oleh latar belakang dan konteks citra.

Secara keseluruhan, hasil eksperimen menunjukkan bahwa EfficientNet-B0 dengan pendekatan transfer learning mampu mempelajari pola visual untuk klasifikasi lima kelas material sampah non-organik, tetapi kinerjanya pada dataset dan konfigurasi eksperimen ini masih terbatas. Akurasi sebesar 54,49% dan macro F1-score sebesar 45,75% menunjukkan bahwa model belum memberikan performa yang konsisten pada seluruh kelas. Ketidakseimbangan jumlah objek, keterbatasan sampel pada kelas tertentu, serta kemiripan karakteristik visual antarmaterial merupakan aspek yang perlu dipertimbangkan dalam mengevaluasi hasil tersebut. Sementara itu, Grad-CAM memberikan informasi tambahan mengenai area citra yang berkontribusi terhadap prediksi dan membantu mengidentifikasi contoh ketika model menghasilkan klasifikasi yang benar maupun salah. Dengan demikian, penggunaan EfficientNet-B0 dan Grad-CAM dalam penelitian ini memberikan dasar untuk menilai kinerja klasifikasi sekaligus mengamati perilaku model secara visual, tetapi peningkatan performa masih diperlukan sebelum model dapat dipertimbangkan untuk penerapan klasifikasi sampah secara lebih luas.

KESIMPULAN

Penelitian ini menerapkan EfficientNet-B0 dengan pendekatan transfer learning untuk mengklasifikasikan sampah non-organik ke dalam lima kelas material, yaitu plastik, logam, kaca, kertas/karton, dan busa/styrofoam, menggunakan dataset TACO. Hasil evaluasi pada 323 citra uji menunjukkan akurasi sebesar 54,49%, macro precision 44,19%, macro recall 53,27%, macro F1-score 45,75%, dan weighted F1-score 56,99%. Kelas plastik memperoleh F1-score tertinggi sebesar 0,66, sedangkan kelas kaca memperoleh F1-score terendah sebesar 0,25, yang menunjukkan bahwa kemampuan klasifikasi masih bervariasi antarkelas. Visualisasi Grad-CAM memberikan informasi kualitatif mengenai area citra yang berkontribusi terhadap prediksi, baik pada klasifikasi yang benar maupun yang salah, serta memperlihatkan bahwa confidence yang tinggi tidak selalu menghasilkan prediksi yang tepat. Dengan demikian, penggabungan evaluasi kuantitatif dan visualisasi Grad-CAM memberikan gambaran mengenai kinerja sekaligus perilaku visual model. Penelitian selanjutnya perlu berfokus pada peningkatan kualitas dan keseimbangan data, perbaikan kemampuan diskriminasi antarkelas, serta evaluasi interpretabilitas Grad-CAM secara kuantitatif.

REFERENSI

[1] H. Abdu and M. H. Mohd Noor, "A Survey on Waste Detection and Classification Using Deep Learning," *IEEE Access*, vol. 10, pp. 128151–128165, 2022, doi: 10.1109/ACCESS.2022.3226682.

[2] Q. Zhang *et al.*, "Recyclable waste image recognition based on deep learning," *Resour. Conserv. Recycl.*, vol. 171, p. 105636, Aug. 2021, doi: 10.1016/j.resconrec.2021.105636.

[3] Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, and X. Liu, "Waste image classification based on transfer learning and convolutional neural network," *Waste Manag.*, vol. 135, pp. 150–157, Nov. 2021, doi: 10.1016/j.wasman.2021.08.038.

[4] T.-W. Wu, H. Zhang, W. Peng, F. Lü, and P.-J. He, "Applications of convolutional neural networks for intelligent waste identification and recycling: A review," *Resour. Conserv. Recycl.*, vol. 190, p. 106813, Mar. 2023, doi: 10.1016/j.resconrec.2022.106813.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sep. 2020, [Online]. Available: http://arxiv.org/abs/1905.11946

[9] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[11] P. F. Proença and P. Simões, "TACO: Trash Annotations in Context for Litter Detection," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.06975

[12] I. Nasir and G. A. Aziz Al-Talib, "Waste Classification Using Artificial Intelligence Techniques:Literature Review," *Tech. Rom. J. Appl. Sci. Technol.*, vol. 5, pp. 49–59, Jan. 2023, doi: 10.47577/technium.v5i.8345.

[13] W. Ding, M. Abdel-Basset, H. Hawash, and A. M. Ali, "Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey," *Inf. Sci. (Ny).*, vol. 615, pp. 238–292, Nov. 2022, doi: 10.1016/j.ins.2022.10.013.

[14] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?,'" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[16] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017, [Online]. Available: http://arxiv.org/abs/1705.07874

[17] 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," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.

[18] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Mar. 2018, pp. 839–847. doi: 10.1109/WACV.2018.00097.

[19] G. Alicioglu and B. Sun, "A survey of visual analytics for Explainable Artificial Intelligence methods," *Comput. Graph.*, vol. 102, pp. 502–520, Feb. 2022, doi: 10.1016/j.cag.2021.09.002.

[20] M. R. Islam, M. U. Ahmed, S. Barua, and S. Begum, "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks," *Appl. Sci.*, vol. 12, no. 3, p. 1353, Jan. 2022, doi: 10.3390/app12031353.