

Penerapan Vision Transformer Untuk Klasifikasi Sampah Rumah Tangga

Prasista Dhiyaul Haq¹, Bambang Irawan²

¹Teknik Informatika, Universitas Muhadi Setiabudi

Corresponding Author: Name, Surname, e-mail: dhiyaprasista577@gmail.com

Published: February, 2026

ABSTRACT

The increasing volume of Household waste requires an accurate and efficient automatic waste sorting system. This study aims to apply Vision Transformer (ViT) for image-based household waste classification. The dataset was divided into training and validation sets and prepared to match the Vision Transformer architecture. The ViT-Base Patch16-224 model was trained using the AdamW optimizer with a learning rate of 0.0002, batch size of 16, and 15 training epoch. Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Experimental results show that the proposed model achieved an overall accuracy of 95%. The inorganic class obtained a precision of 0.9, recall of 0.96, and F1-score of 0.95, while the organic class achieved a precision of 0.94, recall of 0.93, F1-score of 0.94. These results indicate that self-attention mechanism in Vision Transformer effectively extracts global visual features and improves classification stability. Therefore, Vision Transformer demonstrates strong potential for implementation in intelligent automatic waste sorting systems.

Keywords: Vision Transformer, Waste Classification, Image Processing

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0>

PENDAHULUAN

Peningkatan volume sampah rumah tangga masih menjadi permasalahan lingkungan yang signifikan, terutama akibat rendahnya kesadaran masyarakat dalam melakukan pemilahan sampah sejak dari sumbernya. Limbah rumah tangga yang tidak dikelola dengan baik dapat menurunkan kualitas lingkungan dan meningkatkan risiko pencemaran air serta gangguan kesehatan masyarakat (Kementerian Lingkungan Hidup dan Kehutanan & Kementerian Riset Indonesia, 2023). Selain itu, pertumbuhan populasi dan peningkatan pola konsumsi masyarakat turut mempercepat laju produksi sampah, sehingga menambah beban sistem pengelolaan limbah (Made, Bunga, & Dewi, 2021). Proses pemilahan yang dilakukan secara manual cenderung tidak efisien dan berpotensi menimbulkan kesalahan, sehingga menghambat efektivitas pengelolaan dan daur ulang sampah. Pemanfaatan teknologi kecerdasan buatan berbasis pengolahan citra digital menjadi solusi potensial untuk mendukung klasifikasi sampah secara otomatis dan akurat (Harits et al., 2025).

Pemanfaatan kecerdasan buatan (Artificial Intelligence) dalam klasifikasi sampah telah banyak dikembangkan untuk meningkatkan akurasi dan efisiensi proses pemilahan. Penelitian sebelumnya menunjukkan bahwa metode Convolutional Neural Network (CNN) mampu memberikan performa yang cukup baik dalam mengklasifikasikan jenis sampah rumah tangga berdasarkan citra digital (Aulia, Arwoko, & Asmawati, 2024). Pendekatan transfer learning pada CNN terbukti mampu meningkatkan akurasi klasifikasi jenis sampah dengan memanfaatkan bobot model pra-latih (Udayana et al., 2024). Selain itu, penerapan AI dalam pengelolaan sampah juga berkontribusi dalam pengurangan limbah plastik dan peningkatan efektivitas daur ulang (Listyalina et al., 2024; Kumar et al., 2022).

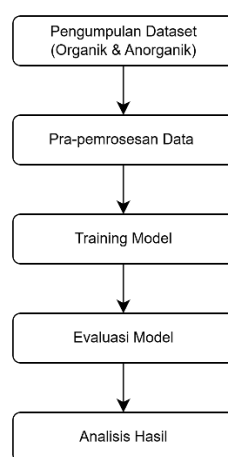
Sistem deteksi dan klasifikasi sampah berbasis deep learning secara real-time telah dikembangkan untuk mendukung proses daur ulang secara berkelanjutan. Beberapa penelitian lain juga mengembangkan sistem klasifikasi sampah otomatis yang mampu meningkatkan kecepatan pemrosesan data serta mengurangi ketergantungan terhadap proses pemilahan manual, sehingga mendukung penerapan konsep smart waste management (Pieters et al., 2025). Namun, pendekatan berbasis CNN masih memiliki keterbatasan dalam menangkap hubungan global antar fitur citra, terutama pada objek sampah yang memiliki karakteristik visual yang mirip.

Sebagai alternatif, Vision Transformer (ViT) mulai dikembangkan dalam klasifikasi citra karena kemampuannya memanfaatkan mekanisme self-attention untuk mengekstraksi informasi global secara lebih efektif. Penelitian menunjukkan bahwa model hybrid berbasis ViT mampu menghasilkan akurasi yang lebih tinggi dibandingkan CNN konvensional dalam klasifikasi sampah (Husein et al., 2025). Selain itu, performa sistem juga dapat ditingkatkan melalui integrasi dengan sensor tambahan seperti infrared dan hyperspectral untuk membedakan jenis sampah dengan karakteristik visual yang serupa (Owen & Yates, 2025). Di samping pengembangan model klasifikasi, sistem pendukung berbasis kecerdasan buatan seperti chatbot edukasi lingkungan dan sistem klasifikasi berbasis prompt AI juga berperan dalam meningkatkan kesadaran masyarakat serta memperluas penerapan smart waste management yang berkelanjutan (Wisesa et al., 2025).

METODE

A. Alur Kerja Penelitian

Alur kerja penelitian ini dimulai dari pengumpulan dataset yang berisi citra sampah rumah tangga dari kaggle. Dataset kemudian melalui tahap pra-pemrosesan untuk menyesuaikan format input model Vision Transformer. Selanjutnya, dataset dibagi menjadi data pelatihan dan data validasi dengan perbandingan 80% dan 20%. Model Vision Transformer kemudian dilatih dan dievaluasi menggunakan data validasi. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, precision, recall, F1-score, dan confusion matrix untuk mengukur performa klasifikasi sampah organik dan anorganik.



Gambar 1 Alur Penelitian

B. Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset sampah rumah tangga dari Kaggle yang terdiri dari dua kelas, yaitu sampah organik dan sampah anorganik. Dataset ini dipilih karena memiliki variasi bentuk objek, kondisi pencahayaan, dan latar belakang citra yang beragam, sehingga sesuai untuk menguji kemampuan generalisasi model Vision Transformer dalam melakukan klasifikasi citra sampah.

Tabel 1 Dataset

Class	Trainin g	Validatio n
Anorgani k	4,560	1,150
Organik	3,744	926
Total	8,304	2,076

C. Pra-pemrosesan Data

Tahap pra-pemrosesan meliputi perubahan ukuran citra menjadi 224x224 piksel sesuai dengan kebutuhan input Vision Transformer, normalisasi nilai piksel, serta augmentasi data berupa rotasi dan horizontal flip. Proses augmentasi bertujuan untuk meningkatkan keragaman data pelatihan dan mengurangi risiko overfitting, sedangkan normalisasi digunakan untuk mempercepat proses konvergensi model selama pelatihan.

Dataset dibagi menjadi dua bagian, yaitu 80% data pelatihan dan 20% data validasi. Data pelatihan digunakan untuk melatih model Vision Transformer, sedangkan data validasi digunakan untuk memantau performa model selama proses training serta mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilatih.

D. Aritektur Vision Transformer

Vision Transformer (ViT) merupakan arsitektur deep learning yang mengadaptasi konsep Transformer dari pemrosesan bahasa alami ke bidang pengolahan citra. Citra input dibagi menjadi beberapa patch berukuran tetap, kemudian setiap patch diubah menjadi vektor embedding. Selanjutnya, embedding tersebut diproses menggunakan mekanisme self-attention untuk menangkap hubungan global antar patch. Output dari Transformer Encoder digunakan untuk melakukan klasifikasi melalui lapisan fully connected

E. Proses Pelatihan Model

Pelatihan model dilakukan menggunakan model Vision Transformer (ViT-Base Patch16-224) yang diimplementasikan dengan framework PyTorch dengan memanfaatkan model Vision Transformer pra-latih (pretrained). Model dilatih selama 15 epoch dengan ukuran batch 16 dan learning rate 0,0002 menggunakan optimizer AdamW. Fungsi loss yang digunakan adalah Cross Entropy Loss untuk meminimalkan kesalahan klasifikasi.

Tabel 2 Hyperparameter

Parameter	Value
Optimizer	Adam / AdamW
Loss	Cross Entropy

Learning Rate	0.0002
Epoch	15
Batch Size	16

F. Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan metrik akurasi, precision, recall, F1-score, serta confusion matrix. Akurasi digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan, sedangkan precision dan recall digunakan untuk mengevaluasi performa klasifikasi pada masing-masing kelas. Confusion matrix digunakan untuk menganalisis pola kesalahan prediksi antar kelas. Hasil evaluasi ini digunakan sebagai dasar untuk menganalisis efektivitas Vision Transformer dalam klasifikasi sampah rumah tangga.

HASIL PENELITIAN

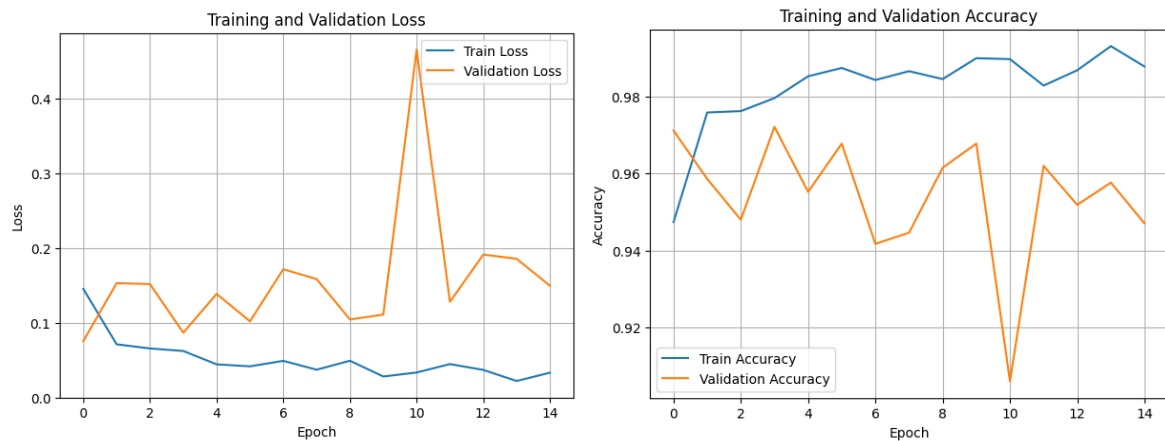
Proses pelatihan model *Vision Transformer* (ViT) dilakukan selama 15 epoch dengan ukuran citra 224x224 piksel, batch size 16, serta learning rate 0,0002 menggunakan optimizer AdamW. Dataset dibagi menjadi 80% data latih dan 20% data validasi secara acak.

Hasil pelatihan menunjukkan peningkatan akurasi yang konsisten pada data latih dan validasi. Pola konvergensi yang stabil mengindikasikan bahwa model mampu mempelajari representasi fitur citra sampah secara efektif tanpa mengalami *overfitting* yang signifikan. Hal ini menunjukkan bahwa mekanisme self-attention pada ViT mampu mengekstraksi informasi visual global secara optimal.

Tabel 3. Training Model ViT

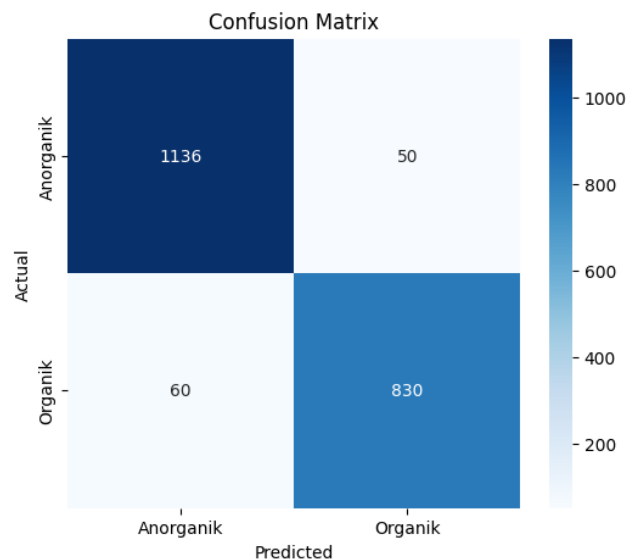
Epoch	Training Loss	Training Accuracy	Validasi Loss	Validation Accuracy
1	0.1452	94.74%	0.0752	97.11%
2	0.0710	97.58%	0.1529	95.86%
3	0.0655	97.62%	0.1516	94.80%
4	0.0621	97.95%	0.0866	97.21%
5	0.0442	98.52%	0.1385	95.52%
6	0.0415	98.74%	0.1018	96.77%
7	0.0489	98.42%	0.1716	94.17%
8	0.0370	98.65%	0.1583	94.46%
9	0.0490	98.45%	0.1043	96.15%
10	0.0279	98.99%	0.1108	96.77%
11	0.0333	98.96%	0.4656	90.61%

12	0.0445	98.28%	0.1279	96.19%
13	0.0368	98.68%	0.1912	95.18%
14	0.0219	99.30%	0.1855	95.76%
15	0.0330	98.77%	0.1494	94.70%



Gambar 2 Grafik Loss dan Akurasi Training

Berdasarkan hasil tersebut, akurasi validasi tertinggi diperoleh pada epoch ke-4 sebesar 97,21% yang menunjukkan performa optimal model sebelum muncul fluktuasi validasi pada epoch berikutnya. Meskipun terjadi peningkatan validasi loss pada epoch ke-11, performa model kembali stabil pada epoch selanjutnya, yang menandakan bahwa model tetap memiliki kemampuan generalisasi yang baik.



Gambar 3 Confusion Matrix

Hasil dari confusion matrix, model Vision Transformer mampu mengklasifikasikan sebagian besar data sampah dengan tingkat ketepatan yang tinggi. Dari total 2.076 data validasi, sebanyak 1.136 citra sampah anorganik dan 830 citra sampah organik berhasil diklasifikasi dengan benar. Nilai akurasi keseluruhan mencapai 95%, yang menunjukkan performa klasifikasi yang sangat baik.

Tabel 4 Hasil Akurasi Saat Training

Kelas	Precision	Recall	F1-Score	Accuracy
Anorganik	0.95	0.96	0.95	0.95
Organik	0.94	0.93	0.94	

Pada kelas anorganik diperoleh nilai presisi sebesar 0,95 dan nilai recall sebesar 0,96, yang menunjukkan bahwa model mampu mengidentifikasi sampah anorganik dengan tingkat ketepatan dan sensitivitas yang sangat baik. Nilai F1-score sebesar 0,95 mengindikasikan keseimbangan yang optimal antara precision dan recall pada kelas tersebut. Pada kelas organik, model menghasilkan nilai precision sebesar 0,94 dan recall sebesar 0,93, dengan F1-score sebesar 0,94. Hasil ini menunjukkan bahwa model juga mampu menghasilkan sampah organik secara akurat meskipun terdapat sedikit penurunan performa dibandingkan dengan kelas anorganik, yang kemungkinan disebabkan oleh variasi bentuk dan tekstur sampah organik yang lebih beragam.

Secara keseluruhan, model mencapai akurasi sebesar 95%, yang menandakan bahwa *Vision Transformer* memiliki kemampuan yang sangat baik dalam melakukan klasifikasi sampah rumah tangga secara otomatis dan konsisten.

KE Simpulan

Kesimpulan dari penelitian ini menunjukkan bahwa penerapan Vision Transformer (ViT) memberikan hasil yang sangat memuaskan dalam klasifikasi sampah rumah tangga berbasis citra digital. Model ViT-Base Patch16-224 yang dilatih menggunakan dataset dua kelas, yaitu sampah organik dan anorganik, mampu mencapai akurasi keseluruhan sebesar 95%, dengan nilai precision, recall, dan F1-score yang tinggi pada kedua kelas. Hasil ini menandakan bahwa arsitektur Vision Transformer memiliki kemampuan yang sangat baik dalam mengekstraksi fitur visual global melalui mekanisme self-attention, sehingga mampu membedakan karakteristik objek sampah secara lebih efektif dibandingkan pendekatan konvensional.

Performa tinggi yang diperoleh pada kelas anorganik maupun organik menunjukkan bahwa model tidak hanya akurat dalam memprediksi kelas dominan, tetapi juga memiliki keseimbangan yang baik antara ketepatan dan sensitivitas prediksi. Stabilitas proses pelatihan yang ditunjukkan oleh kurva loss dan akurasi juga mengindikasikan bahwa model mampu melakukan pembelajaran representasi fitur secara optimal tanpa mengalami overfitting yang signifikan. Dengan demikian, Vision Transformer terbukti mampu menjadi alternatif yang andal dalam pengembangan sistem klasifikasi sampah otomatis berbasis kecerdasan buatan.

Selain memberikan kontribusi pada peningkatan akurasi klasifikasi, penelitian ini juga memperlihatkan potensi besar penerapan Vision Transformer dalam mendukung konsep smart waste management yang berkelanjutan. Sistem klasifikasi otomatis yang akurat dapat membantu mempercepat proses pemilahan sampah, mengurangi kesalahan manusia, serta meningkatkan efektivitas proses daur ulang. Hal ini diharapkan dapat berkontribusi secara langsung terhadap pengurangan pencemaran lingkungan dan peningkatan kualitas pengelolaan limbah rumah tangga. Secara keseluruhan, hasil penelitian ini menegaskan bahwa Vision Transformer merupakan pendekatan yang menjanjikan untuk diterapkan dalam sistem pemilahan sampah otomatis di lingkungan nyata. Dengan pengembangan lebih lanjut, baik dari sisi dataset, arsitektur model, maupun implementasi sistem real-time, teknologi ini berpotensi menjadi solusi inovatif dalam pengelolaan sampah berbasis kecerdasan buatan yang efisien, akurat, dan berkelanjutan.

REFERENSI

- Aulia, D. S., Arwoko, H., & Asmawati, E. (2024). Klasifikasi sampah rumah tangga menggunakan metode convolutional neural network. *Metik*, 8(2), 114–120. <https://doi.org/10.47002/metik.v8i2.956>
- Harits, M. A., Riyadasyah, D., Rohman, S. N., & Irawan, R. D. (2025). Sistem klasifikasi bahan daur ulang dan penanganan berbasis prompt AI. *Prosiding* 265–270.
- Husein, A. M., Harahap, B. B., Simatupang, T. F., & Baeha, K. S. (2025). A hybrid vision transformer model for efficient waste classification. *Journal of Intelligent Systems*, 2, 261–275.
- Kementerian Lingkungan Hidup dan Kehutanan & Kementerian Riset Indonesia. (2023). *Laporan kinerja pengelolaan sampah nasional tahun 2023*. Jakarta: Penulis.
- Kumar, A., et al. (2022). Smart waste management system using deep learning. *IEEE Access*, 10, 21534–21545. <https://doi.org/10.1109/ACCESS.2022.3145823>
- Listyalina, L., Utami, R. R., Arifin, U. F., Putri, N., & Kulit, T. P. (2024). The application of artificial intelligence in waste classification as an effort in plastic waste management. *Telematika*, 21(1), 1–13. <https://doi.org/10.31515/telematika.v21i1.11977>
- Made, N. I., Bunga, N. I. A., & Dewi, S. (2021). Analisa limbah rumah tangga terhadap dampak pencemaran lingkungan. *Lim*, 1159–1164.
- Owen, A. O., & Yates, G. (2025). Combining vision transformers with sensors (e.g., infrared, hyperspectral, or RFID) to improve recognition of waste materials that look visually similar. *Conference Proceedings*.
- Pieters, L. S., Pradita, U., Park, S. B., & Tangerang, K. (2025). Development of automatic waste classification system using CNN based deep learning to support smart waste management. *Jurnal Teknologi*, 10(1), 214–224.
- Udayana, I. G. A. P., et al. (2024). Klasifikasi jenis sampah menggunakan metode transfer learning pada convolutional neural network (CNN). *Jurnal Logika Komputer (JELIKU)*, 12(3), 210–219. <https://doi.org/10.24843/JLK.2023.v12.i03.p11>
- Wisasa, B. A., Putri, V. M., Faristasari, E., Andreanto, S., & Duli, J. (2025). Developing an NLP-based chatbot for waste management education in Sungailiat. *Journal of Ambient Intelligence and Smart Environments*, 5(3), 1245–1253. <https://doi.org/10.30811/jaise.v5i3.7522>