



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 4 (2025) pp: 400-406

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Implementasi Metode Random Forest untuk Klasifikasi Sampah Organik dan Anorganik Berbasis Citra Digital

Melania Velisita Lau¹, Aryani Dwi Handayani², Miyosagusti Yudo Widodo³, Rani Irma Handayani⁴, Risca Lusiana Pratiwi⁵

^{1,2,3}Teknik Informatika, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika, Jakarta

^{4,5}Teknologi Informasi Universitas Nusa Mandiri

melaniavelisita@gmail.com¹, aryanidwi1995@gmail.com², addoytrendy2017@gmail.com³, rani.rih@nusamandiri.ac.id⁴, risca.ral@nusamandiri.ac.id⁵

Abstrak

Pengelolaan sampah merupakan salah satu tantangan utama yang dihadapi lingkungan perkotaan modern. Peningkatan volume sampah setiap hari menuntut adanya sistem yang efisien untuk memilah dan mengelola berbagai jenis sampah, terutama antara sampah organik dan anorganik. Proses pemilahan manual sering kali membutuhkan waktu lama, biaya tinggi, dan bergantung pada tenaga manusia. Oleh karena itu, penerapan teknologi berbasis kecerdasan buatan menjadi alternatif yang potensial untuk membantu proses pengelolaan sampah secara otomatis dan efisien. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi jenis sampah berbasis citra digital menggunakan algoritma Random Forest. Dataset yang digunakan diambil dari platform Kaggle, yang berjumlah 1.000 citra dengan komposisi seimbang, yaitu 500 citra sampah organik dan 500 citra sampah anorganik. Setiap citra berukuran 64x64 piksel dan telah melalui tahap pra-proses berupa normalisasi serta ekstraksi fitur menggunakan metode Histogram of Oriented Gradients (HOG) untuk menghasilkan fitur visual yang merepresentasikan bentuk dan tekstur sampah secara optimal. Dataset dibagi menjadi dua bagian, yaitu 80% data pelatihan dan 20% data pengujian. Hasil eksperimen menunjukkan bahwa model Random Forest mampu mencapai tingkat akurasi sebesar 96%, dengan nilai precision, recall, dan F1-score yang tinggi pada kedua kelas. Hal ini membuktikan bahwa algoritma Random Forest memiliki kemampuan generalisasi yang baik untuk mendeteksi dan mengklasifikasikan jenis sampah. Berdasarkan hasil penelitian, pendekatan ini dinilai efektif dan berpotensi besar untuk diterapkan dalam sistem pengelolaan sampah otomatis berbasis pengenalan citra. Implementasinya dapat mendukung efisiensi proses daur ulang, mengurangi beban kerja manusia, serta berkontribusi terhadap pengelolaan lingkungan perkotaan yang lebih berkelanjutan.

Kata kunci: Klasifikasi, Sampah, Random Forest, Citra Digital, HOG

1. Latar Belakang

Pengelolaan sampah menjadi salah satu permasalahan utama di berbagai negara, termasuk Indonesia. Setiap hari, jumlah sampah yang dihasilkan masyarakat terus meningkat seiring dengan pertumbuhan penduduk dan perkembangan industri. Permasalahan utama terletak pada kurangnya sistem pengelolaan yang efisien, terutama dalam proses pemilahan antara sampah organik dan anorganik. Pemilahan yang masih dilakukan secara manual sering kali menyebabkan ketidakefisienan, membutuhkan waktu yang lama, dan berpotensi menimbulkan kesalahan manusia.

Pengembangan teknologi kecerdasan buatan, khususnya machine learning, telah memberikan peluang besar untuk membantu proses otomatisasi dalam klasifikasi sampah. Salah satu algoritma yang banyak digunakan dalam bidang klasifikasi citra adalah Random Forest, yang memiliki kemampuan untuk mengelola data berdimensi tinggi dan menghasilkan akurasi yang stabil. Menurut Da Silva et al.[1], penerapan algoritma Random Forest mampu mengidentifikasi jenis sampah organik dan non-organik secara efektif berdasarkan fitur warna dan tekstur.

Selain itu, penelitian oleh Akyol dan Karaci [2] menunjukkan bahwa kombinasi antara fitur hasil ekstraksi deep learning (DenseNet201) dan Random Forest Classifier dapat meningkatkan kinerja sistem klasifikasi sampah secara signifikan. Pendekatan ini membuktikan bahwa Random Forest tidak hanya dapat efisien pada data tabular, tetapi juga mampu bekerja optimal pada data berbasis citra digital.

Penelitian lain oleh Sami et al. [3] juga menjelaskan bahwa algoritma machine learning seperti Random Forest memiliki potensi besar untuk mendukung sistem pengelolaan sampah cerdas. Melalui penerapan model klasifikasi yang tepat, sistem dapat membantu masyarakat dan pemerintah dalam memisahkan jenis sampah dengan cepat dan akurat. Berdasarkan hal tersebut, penelitian ini berfokus pada penerapan metode Random Forest untuk klasifikasi sampah organik dan anorganik berbasis citra digital, dengan tujuan untuk menghasilkan model yang memiliki tingkat akurasi tinggi serta dapat menjadi dasar pengembangan sistem pemilahan sampah otomatis di masa depan.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk mengimplementasikan dan menguji algoritma Random Forest dalam mengklasifikasikan citra sampah organik dan anorganik. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data, praproses citra, ekstraksi fitur, pembagian data, pelatihan model, hingga evaluasi hasil klasifikasi. Setiap tahap dirancang agar hasil yang diperoleh dapat merepresentasikan performa model secara objektif.

Dataset Penelitian

Dataset yang digunakan dalam penelitian ini berasal dari situs Kaggle Dataset berjudul "Garbage Classification (6 Classes)" yang dibuat oleh Quang Theng (2020). Yang dapat diakses melalui link: <https://www.kaggle.com/datasets/quangtheng/garbage-classification-6-classes-775class> Dataset ini terdiri dari enam kelas utama, yaitu cardboard, glass, metal, paper, plastic, dan organic, dengan total 775 citra per kelas. Namun, dalam penelitian ini hanya digunakan dua kelas, yaitu organic dan paper, masing-masing sebanyak 500 citra. Pemilihan dua kelas ini didasarkan pada perbedaan karakteristik visual yang cukup jelas, seperti tekstur, warna, dan bentuk objek yang dapat membantu model dalam proses klasifikasi. Seluruh data citra diubah ukurannya menjadi 64x64 piksel agar konsisten dan sesuai dengan kebutuhan komputasi.

Praproses Citra

Tahapan praproses dilakukan untuk meningkatkan kualitas citra sebelum dimasukkan ke proses pelatihan model. Langkah-langkah yang dilakukan meliputi pengubahan ukuran (resizing) citra, normalisasi nilai piksel ke rentang 0-1, serta konversi format warna menjadi RGB. Proses ini bertujuan untuk menyeragamkan ukuran dan skala intensitas citra sehingga model dapat lebih mudah mengenali pola visual tanpa terpengaruh oleh variasi pencahayaan atau ukuran asli gambar. Selain itu, beberapa citra yang rusak atau tidak terbaca juga dihapus agar tidak mengganggu hasil pelatihan.

Ekstraksi Fitur

Ekstraksi fitur dilakukan menggunakan metode Histogram of Oriented Gradients (HOG), yang berfungsi untuk menangkap informasi bentuk dan tepi pada citra. Fitur hasil ekstraksi kemudian direpresentasikan dalam bentuk vektor numerik yang menggambarkan karakteristik setiap citra. Representasi ini digunakan sebagai masukan bagi algoritma Random Forest dalam proses pembelajaran. Pemilihan metode HOG dilakukan karena kemampuannya dalam mempertahankan informasi penting dari gambar sambil mengurangi dimensi data yang terlalu besar.

Pembagian Data

Data yang telah diproses kemudian dibagi menjadi dua bagian, yaitu data pelatihan (training data) sebesar 80% dan data pengujian (testing data) sebesar 20%. Pembagian ini dilakukan secara acak dengan mempertahankan proporsi kelas menggunakan metode stratified sampling agar setiap kelas tetap terwakili secara seimbang pada kedua subset data. Pendekatan ini penting untuk menghindari bias hasil pelatihan terhadap kelas tertentu.

Pelatihan Model Random Forest

Tahapan selanjutnya adalah melatih model Random Forest menggunakan data pelatihan. Algoritma ini merupakan metode ensemble learning yang menggabungkan banyak pohon keputusan (decision trees) untuk menghasilkan prediksi yang lebih stabil dan akurat. Setiap pohon dilatih pada subset data dan fitur yang berbeda,

kemudian hasilnya digabungkan melalui proses voting mayoritas. Pada penelitian ini, jumlah estimator ($n_{estimator}$) ditetapkan sebanyak 150 dengan nilai $random_state$ sebesar 42 untuk menjaga konsistensi hasil. Model kemudian dilatih menggunakan pustaka `scikit_learn` dan disimpan untuk keperluan evaluasi.

Evaluasi Model

Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall dan F1-score berdasarkan hasil prediksi model terhadap data uji. Selain itu, digunakan pula confusion matrix untuk menunjukkan jumlah prediksi benar dan salah pada masing-masing kelas. Berdasarkan hasil pengujian, model Random Forest yang dibangun mencapai akurasi sebesar 96%, yang menunjukkan tingkat keberhasilan tinggi dalam membedakan citra antara kelas organik dan paper. Nilai presisi dan recall juga menunjukkan keseimbangan yang baik antar kelas, menandakan model tidak mengalami bias yang signifikan terhadap salah satu kategori.

Visualisasi Hasil

Hasil evaluasi divisualisasikan dalam bentuk grafik batang yang menampilkan perbandingan nilai presisi, recall dan F1-score pada setiap kelas. Selain itu, confusion matrix divisualisasikan dalam bentuk heatmap untuk memudahkan interpretasi terhadap kesalahan klasifikasi. Visualisasi ini bertujuan untuk memberikan gambaran yang lebih intuitif mengenai performa model serta area yang masih dapat ditingkatkan.

Analisis Hasil

Berdasarkan hasil evaluasi, dapat disimpulkan bahwa algoritma Random Forest memiliki kemampuan generalisasi yang baik terhadap dataset citra sederhana seperti klasifikasi sampah. Akurasi 96% yang diperoleh menunjukkan bahwa kombinasi antara fitur HOG dan metode ensemble mampu memberikan hasil yang optimal. Namun, untuk dataset dengan jumlah kelas yang lebih banyak dan kompleksitas visual yang lebih tinggi, metode ini dapat dikombinasikan dengan pendekatan deep learning untuk peningkatan kinerja.

3. Hasil dan Diskusi

Contoh Citra per Kelas

Pada tahap awal pengujian, dilakukan observasi terhadap contoh citra dari masing-masing kelas untuk memastikan kualitas data yang digunakan dalam pelatihan model. Gambar yang digunakan terdiri dari dua kelas utama, yaitu organik dan paper, masing-masing berjumlah 500 citra. setiap citra mewakili karakteristik visual yang khas sesuai dengan jenisnya.

Citra pada kelas organik umumnya menampilkan objek seperti sisa makanan, kulit buah, atau bahan alami yang memiliki warna tidak seragam dan tekstur kompleks. Sedangkan citra pada kelas paper memperlihatkan benda-benda seperti kertas, kardus, atau cokelat muda.

Proses ini penting karena variasi citra di setiap kelas memengaruhi kemampuan model dalam mengenali pola visual. Dengan melihat contoh citra per kelas, dapat dipastikan bahwa data yang digunakan memiliki representasi yang baik dan tidak menimbulkan bias pada hasil klasifikasi. Selain itu, tahap ini juga membantu peneliti memahami perbedaan visual yang akan menjadi dasar dalam proses ekstraksi fitur menggunakan metode Histogram of Oriented Gradients (HOG).



Gambar 3.1 Contoh Citra per Kelas (Organic dan Paper)

Visualisasi HOG dan Feature Importance Overlay

Setelah proses pra-proses data selesai, dilakukan visualisasi fitur citra untuk memahami pola yang diekstraksi oleh model. Tahap ini bertujuan untuk menunjukkan bagaimana citra asli diubah menjadi representasi fitur yang lebih informatif sebelum digunakan dalam pelatihan algoritma Random Forest. Pada gambar di bawah ini ditampilkan tiga tampilan utama:

Citra asli, yang menunjukkan bentuk awal dari objek sebelum dilakukan ekstraksi fitur.

Visualisasi HOG (Histogram of Oriented Gradients), yang menggambarkan arah dan intensitas gradien pada tepi citra. Bagian ini membantu model mengenali pola bentuk, kontur, serta tekstur objek.

Feature Importance Overlay, yaitu peta visual yang menampilkan area citra yang dianggap paling berpengaruh terhadap hasil klasifikasi oleh model Random Forest. Area yang berwarna lebih terang atau merah menunjukkan bagian yang memberikan kontribusi besar dalam proses pengambilan keputusan.

Visualisasi ini menunjukkan bahwa model fokus pada area dengan tekstur dan pola khas seperti tepi objek atau area dengan perubahan kontras yang signifikan. Hal ini membuktikan bahwa fitur yang diekstraksi melalui HOG mampu menangkap informasi penting yang relevan dengan perbedaan antar kelas citra.

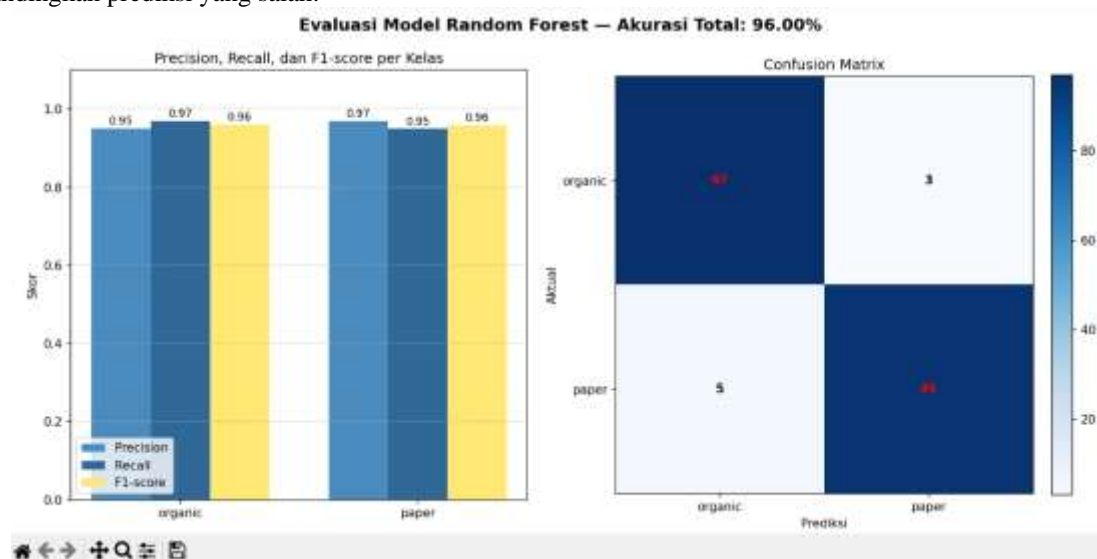


Gambar 3.2 Visualisasi Citra Asli, HOG, dan Feature Importance Overlay

Evaluasi Model (Akurasi, Recall, F1-Score, dan Confusion Matrix)

Tahap evaluasi model dilakukan untuk mengukur seberapa baik performa model klasifikasi dalam mengenali jenis sampah berdasarkan citra digital. Pada penelitian ini, hasil evaluasi diperoleh dari proses pelatihan model di Visual Studio Code dengan menggunakan algoritma Random Forest. Hasil pengujian menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat performa yang sangat baik. Berdasarkan hasil evaluasi, diperoleh nilai akurasi sebesar 96%, yang menunjukkan bahwa mampu mengklasifikasikan sebagian besar data dengan benar. Selain akurasi, digunakan pula metrik precision, recall, dan F1-score untuk mengukur kinerja model secara lebih menyeluruh.

Gambar di bawah ini menampilkan visualisasi grafik hasil evaluasi model yang meliputi nilai akurasi, precision, recall dan F1-score untuk setiap kelas, serta confusion matrix pada sisi kanan. Grafik tersebut memperlihatkan bahwa seluruh kelas memiliki nilai metrik yang tinggi dan relatif seimbang, dengan score precision dan recall mendekati 1. Sementara itu, confusion matrix menunjukkan bahwa jumlah prediksi benar jauh lebih tinggi dibandingkan prediksi yang salah.



Gambar 4.3 Visualisasi hasil evaluasi model random forest

Dari hasil tersebut dapat disimpulkan bahwa model Random Forest memiliki kemampuan klasifikasi yang konsisten dan andal dalam membedakan jenis sampah organik dan anorganik berdasarkan ciri visual citra. Nilai metrik yang stabil pada setiap kelas juga menandakan bahwa model tidak mengalami bias terhadap kelas tertentu dan mampu bekerja efektif meskipun dataset terbatas.

Proses Pelatihan Model Random Forest

Proses pelatihan model dilakukan menggunakan algoritma Random Forest Classifier di lingkungan Visual Studio Code (Vs Code). Dataset yang telah melalui tahap pra-proses dan ekstraksi fitur HOG dibagi menjadi dua bagian, yaitu 80% data latih dan 20% data uji. Selama proses pelatihan, model membentuk sejumlah pohon keputusan (decision trees) yang masing-masing melakukan prediksi secara independen. Hasil akhir klasifikasi kemudian ditentukan melalui majority voting, di mana kelas dengan suara terbanyak menjadi hasil prediksi akhir.

Pendekatan ini membantu mengurangi variansi dan meningkatkan stabilitas prediksi. Selain itu, hasil pelatihan menunjukkan bahwa model mampu mengenali pola visual pada data citra dengan baik dan tidak menunjukkan indikasi overfitting, karena selisih akurasi antara data latih dan data uji tidak terlalu signifikan.

Analisis Kesalahan Klasifikasi

Meskipun model Random Forest menggunakan performa yang tinggi, masih terdapat beberapa kesalahan dalam proses klasifikasi. Berdasarkan hasil confusion matrix, sebagian kecil citra organik teridentifikasi sebagai anorganik (paper).

Hal ini kemungkinan disebabkan oleh kemiripan warna dan tekstur pada beberapa citra, terutama ketika kondisi pencahayaan kurang optimal atau latar belakang tidak seragam. Selain itu, ukuran objek pada citra juga dapat memengaruhi hasil prediksi jika terlalu kecil atau tidak terpusat pada area utama gambar.

Kesalahan semacam ini bersifat wajar dalam sistem klasifikasi berbasis citra, terutama ketika dataset terbatas atau distribusi antar kelas tidak seimbang. Dengan penambahan jumlah data dan variasi citra yang lebih beragam, tingkat kesalahan prediksi dapat diminimalkan di masa mendatang.

Pembahasan Keseluruhan

Hasil penelitian ini menunjukkan bahwa algoritma Random Forest yang dikombinasikan dengan metode Histogram of Oriented Gradients (HOG) mampu memberikan performa klasifikasi yang tinggi terhadap citra sampah organik dan anorganik. Model berhasil mencapai akurasi sebesar 96%, yang menandakan bahwa proses ekstraksi fitur telah berhasil menangkap pola visual penting dari masing-masing kelas citra. Citra organik dan anorganik memiliki karakteristik bentuk serta tekstur yang berbeda, dan hal ini dapat diidentifikasi dengan baik oleh model berkat keunggulan HOG dalam mendeskripsikan tepi dan arah gradien objek pada gambar. Dari hasil pengujian menggunakan data uji sebesar 20% dari total dataset, diperoleh nilai precision, recall, dan F1-score yang relatif seimbang di setiap kelas. Keseimbangan ini mengindikasikan bahwa model tidak bias terhadap salah satu kelas dan mampu memberikan prediksi yang konsisten untuk kedua kategori. Kesalahan klasifikasi sebagian besar terjadi pada citra yang memiliki tingkat kemiripan warna dan tekstur, khususnya antara kelas berwarna kusam dan sampah organik yang cenderung berwarna coklat atau keabu-abuan. Meski demikian, secara keseluruhan model menunjukkan stabilitas yang baik dan tidak mengalami overfitting. Berdasarkan hasil tersebut, dapat disimpulkan bahwa penerapan algoritma Random Forest dengan fitur HOG cukup efektif dan efisien

diterapkan dalam sistem Smart Trash Bin atau aplikasi klasifikasi sampah berbasis citra digital. Model ini juga relatif ringan secara komputasi dibandingkan metode berbasis jaringan saraf dalam, sehingga cocok untuk diimplementasikan pada perangkat dengan sumber daya terbatas. Ke depan, peningkatan akurasi dapat dilakukan dengan memperbesar jumlah dataset, menyeimbangkan distribusi antar kelas, serta melakukan augmentasi citra agar model lebih adaptif terhadap variasi kondisi pencahayaan dan bentuk objek.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest dengan penerapan metode Histogram of Oriented Gradients (HOG) berhasil memberikan hasil klasifikasi yang sangat baik terhadap citra sampah organik dan anorganik. Model mencapai tingkat akurasi sebesar 96%, yang menunjukkan bahwa proses ekstraksi fitur melalui HOG mampu menangkap informasi visual penting seperti bentuk dan tepi objek, sehingga mempermudah proses pembelajaran oleh Random Forest. Kinerja model juga terbukti stabil dengan nilai precision, recall, dan F1-score yang seimbang antar kelas, menandakan bahwa model tidak mengalami bias terhadap salah satu kategori. Hal ini membuktikan bahwa metode Random Forest efektif digunakan untuk permasalahan klasifikasi citra sederhana tanpa memerlukan komputasi kompleks seperti pada model deep learning. Secara keseluruhan, penelitian ini berhasil membangun model klasifikasi citra sampah yang efisien dan memiliki potensi untuk diimplementasikan dalam sistem Smart Trash Bin atau aplikasi pendukung pengelolaan sampah berbasis citra digital.

Referensi

1. Helmina Yeni Da Silva, Alfian Nara Weking, and Dominikus Boli Watomakin, "Implementasi Random Forest Untuk Identifikasi Jenis Sampah Organik Dan Non-Organik," *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 5, no. 2, pp. 216–228, Jul. 2025, doi: 10.55606/teknik.v5i2.7612.
2. K. Akyol and A. Karaci, "Waste Classification Using Random Forest Classifier with DenseNet201 Deep Features," in *Lecture Notes on Data Engineering and Communications Technologies*, Springer Science and Business Media Deutschland GmbH, 2023, pp. 401–409. doi: 10.1007/978-3-031-36118-0_36.
3. K. Nasik Sami, Z. Md Afique Amin, and R. Hassan, "Waste Management Using Machine Learning and Deep Learning Algorithms," 2020.
4. M. Haqqi, L. Rochmah, A. Dwi Safitri, and R. Adhi Pratama, "Implementation Of Machine Learning To Identify Types Of Waste Using CNN Algorithm".
5. Wulansari, A. Setyanto, and E. T. Luthfi, "Systematic Literature Review of Waste Classification Using Machine Learning," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 5, no. 2, pp. 405–413, Jan. 2022, doi: 10.31289/jite.v5i2.6211.

6. V. M. Ochango, G. M. Wambugu, and J. G. Ndia, "Feature Extraction using Histogram of Oriented Gradients for Image Classification in Maize Leaf Diseases," 2022. [Online]. Available: www.ijcit.com54
7. M. I. B. Ahmed *et al.*, "Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management," *Sustainability (Switzerland)*, vol. 15, no. 14, Jul. 2023, doi: 10.3390/su151411138.
8. Y.-H. Chen and C.-H. Chao, "Design and implementation of an intelligent waste classification device," *Computing and Artificial Intelligence*, p. 2331, Mar. 2025, doi: 10.59400/cai2331.
9. Fadlil, R. Umar, Sunardi, and A. S. Nugroho, "Comparison of Machine Learning Approach for Waste Bottle Classification," *Emerging Science Journal*, vol. 6, no. 5, pp. 1075–1085, Oct. 2022, doi: 10.28991/ESJ-2022-06-05-011.