

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sampah merupakan salah satu permasalahan lingkungan yang dihadapi berbagai negara, termasuk Indonesia. Sampah yang tidak terkelola dengan baik dapat menimbulkan berbagai dampak negatif, terutama terkait dengan kebersihan, kesehatan, kenyamanan, dan keindahan (Utami et al., 2023). Menurut Undang-Undang Republik Indonesia Nomor 18 tahun 2008 tentang Pengelolaan Sampah, sampah adalah sisa kegiatan sehari-hari manusia dan/atau proses alam yang berbentuk padat. Kementerian Lingkungan Hidup dan Kehutanan (KLHK) melalui Sistem Informasi Pengelolaan Sampah Nasional (SIPSN) mencatat 28,6 juta ton timbulan sampah pada tahun 2021. Jumlah tersebut bertambah menjadi 38,5 juta ton pada tahun 2022, bahkan mencapai 43,3 juta ton pada tahun 2023. Pada tahun 2024, jumlah timbulan sampah menurun menjadi 34,2 juta ton. Sumber utama timbulan sampah adalah dari rumah tangga yang mencapai 53,7%.

Pengelolaan sampah melalui pemilahan sampah daur ulang dari rumah menjadi salah satu solusi efektif untuk mengatasi masalah sampah (Intan Paradita, 2018). Dengan memilah sampah yang dapat didaur ulang, terutama sampah karton, logam, kertas, kaca, dan plastik dapat mengurangi volume sampah yang berakhir di tempat pembuangan akhir (TPA). Namun, semangat masyarakat untuk memilah sampah masih rendah karena perlu meluangkan waktu dan tenaga serta kurangnya infrastruktur pengelolaan sampah yang memadai. Oleh karena itu, diperlukan teknologi untuk membantu masyarakat memilah sampah dengan lebih efektif.

Convolutional neural network (CNN) adalah sebuah metode yang dirancang khusus untuk mengolah gambar. Metode ini dapat digunakan untuk mengembangkan model yang secara otomatis mengidentifikasi dan mengkategorikan jenis-jenis sampah dari data gambar. Model ini kemudian dapat diintegrasikan ke dalam tempat sampah pintar, robot pemilah sampah, atau pusat daur ulang untuk memungkinkan pemilahan yang lebih efisien dan akurat.

CNN merupakan salah satu jenis jaringan saraf tiruan yang menggunakan fungsi konvolusi untuk menangani data dua dimensi seperti gambar (Adedeji & Wang, 2019). CNN dibagi menjadi dua bagian utama yaitu *feature extraction* dan *classification*. Ciri khas CNN adalah pada bagian *feature extraction* yang berperan dalam melakukan *encoding* dari sebuah citra digital menjadi fitur berupa angka-angka yang merepresentasikan citra tersebut.

CNN telah digunakan pada beberapa penelitian, termasuk yang berkaitan dengan sampah. Sebagai contoh, penelitian dari Kurniawan et al. (2023) yang menggunakan CNN berbasis Xception untuk melakukan klasifikasi sampah anorganik dan memperoleh akurasi sebesar 87,81%. Sementara itu, klasifikasi jenis sampah pernah dilakukan oleh Leonardo et al. (2020) menggunakan metode *support vector machine* dengan fitur *local binary pattern* yang memperoleh akurasi sebesar 87,82%.

Efektivitas CNN sangat dipengaruhi oleh kualitas gambar dan *dataset* yang digunakan. *Dataset* gambar yang bagus bisa dilihat dari beberapa hal, seperti resolusi gambar, variasi dan banyaknya gambar, pemberian label yang akurat, serta tidak adanya gambar *outlier*. *Outlier* pada *dataset* gambar adalah gambar yang tidak memiliki informasi yang cukup untuk pengambilan keputusan akhir atau tidak

termasuk dalam kumpulan data karena spesifikasi tertentu (Islam & Fleischer, 2025). Keberadaan *outlier* dalam *dataset* dapat menyebabkan model mempelajari gambar yang salah dan menurunkan akurasi model CNN.

Pendeteksian *outlier* pada *dataset* gambar dapat dilakukan secara manual, namun akan membutuhkan banyak waktu dan tenaga. Salah satu pendekatan yang dapat dilakukan adalah menggunakan *neighbor embedding* seperti *uniform manifold approximation and projection* (UMAP) untuk menghasilkan plot kluster dua dimensi yang dapat dianalisis secara visual dengan cepat (Islam & Fleischer, 2025). Hasil *embedding* juga dapat dianalisis menggunakan metode statistik seperti *interquartile range* (IQR) untuk menemukan data *outlier* (Perez & Tah, 2020).

Uniform manifold approximation and projection (UMAP) adalah sebuah algoritma *non-linear dimensionality reduction* yang dapat membentuk representasi data berdimensi rendah dari data berdimensi lebih tinggi dengan tetap mempertahankan struktur data asli, baik secara lokal maupun global (McInnes et al., 2020). Representasi data ini kemudian digunakan sebagai masukan untuk mencari *outlier*.

Interquartile range (IQR) adalah metode statistik yang membagi kumpulan data terurut menjadi empat bagian sama besar. Nilai yang memisahkan bagian-bagian ini disebut kuartil pertama ($Q1$), kuartil kedua atau median ($Q2$), dan kuartil ketiga ($Q3$). IQR mendefinisikan *outlier* sebagai observasi yang berada di luar rentang $[Q1 - 1,5(Q3 - Q1), Q3 + 1,5(Q3 - Q1)]$ (Perez & Tah, 2020).

Penelitian ini mendeteksi *outlier* pada *dataset* gambar jenis-jenis sampah menggunakan UMAP dan IQR, kemudian mengidentifikasi dan menghapus gambar yang benar merupakan *outlier*, serta membandingkan kinerja model CNN

dalam melakukan klasifikasi pada data sebelum dan setelah penanganan *outlier*. Penelitian ini diharapkan dapat membantu proses pemilahan sampah daur ulang menjadi lebih efektif dan mengurangi dampak lingkungan dari sampah.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara mendeteksi *outlier* pada dataset gambar?
2. Bagaimana hasil deteksi *outlier* pada dataset gambar jenis sampah?
3. Bagaimana tingkat akurasi pada hasil klasifikasi gambar dengan CNN sebelum dan setelah penanganan *outlier*?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah gambar jenis sampah yang bersumber dari dataset trashnet oleh Thung & Yang (2016).
2. Jenis sampah yang diamati adalah Kardus, Kaca, Logam, Kertas, Plastik.
3. Pengolahan citra dilakukan menggunakan teknik *transfer learning* berbasis *Residual Network* (ResNet).
4. Proses deteksi *outlier* dilakukan menggunakan metode *uniform manifold approximation and projection* dan *interquartile range*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendeteksi *outlier* pada *dataset* gambar.
2. Mengidentifikasi *outlier* pada *dataset* gambar jenis sampah.
3. Menganalisis tingkat akurasi pada hasil klasifikasi jenis sampah dengan metode CNN sebelum dan setelah dilakukan penanganan *outlier*.