Type Deep Learning Model for Multi-Label Waste Classification in Canal Environments: A Comparative Study with CNN Architectures

Najirah Umar[1,*], Billy Eden William Asrul[2], Yuyun[3]

[1,2,3] *Faculty of Computer Science, Universitas Handayani Makassar, Makassar, Indonesia*

[3] *National Research and Innovation Agency (BRIN), Bandung, Indonesia*

(Received: June 15, 2025; Revised: August 15, 2025; Accepted: November 22, 2025; Available online: December 9, 2025)

Abstract

The escalating environmental degradation caused by waste underscores the necessity of developing intelligent and sustainable management systems. This study introduces a deep learning–based framework with proposed a modified ConvNeXt architecture enhanced by a two-layer non-linear MLP classification, specifically designed for multi-object waste classification in canal environments. Specifically, ConvNeXt-CNN is introduced as the primary backbone for extracting visual features from waste images. Then, a modified Multi-Layer Perceptron (MLP) is employed to transform these features into multi-label predictions. To optimize the model's generalization capability in recognizing the complexity of waste images, a hybrid data augmentation technique combining SMOTE and MixUp was applied during training. The proposed approach was then compared with ten fine-tuned Convolutional Neural Network (CNN) architectures, ResNet18, ResNet50, VGG16, VGG19, DenseNet121, MobileNet_v2, and EfficientNet (B0, B1, B2, and B3), and evaluated using accuracy, precision, recall, and F1-score metrics. The experimental dataset comprises 855 waste images containing a total of 2,662 annotated objects across 18 categories, including Bamboo, Beverage Carton, Cardboard, Fabric, Glass Bottle, Inorganic Waste, Kite, Leaf, Metal, Organic Waste, Paper, Plastic, Plastic Bottle, Plastic Cup, Residual Waste, Rubber, Small E-waste, Styrofoam, and Wood. The results show that the fine-tuned ConvNeXt achieved the best performance with an F1-score of 0.99, surpassing DenseNet121 (0.95), ResNet18 (0.91), and VGG16 (0.94). The ConvNeXt model demonstrated its robust capability by achieving consistently high identification scores across majority 18 waste categories. When it came to training efficiency, the fine-tuned MobileNetV2 model proved to be the top performer, outclassing ten other pretrained models, with a training time of 13.35s per epoch. Results exhibit that finetuned ConvNext outperforms in terms of accuracy, recall, precision, and F1-score. In conclusion, Integrating ConvNeXt and MLP for multi-object waste classification effectively supports intelligent waste management, enabling practical real-world deployment in smart bins, Material Recovery Facilities, and IoT-integrated urban waste systems.

Keywords: CNN, ConvNeXt-Tiny, Deep Learning, Multi-label, Waste Classification

1. Introduction

One of the most urban environmental challenges is ineffective waste management. The accelerated growth of urban populations, coupled with increasing food consumption, has led to a surge in waste that overwhelms traditional waste management systems [1]. The main concern lies with canal waste, which can harm water ecosystems and cause blockages in drainage systems. These blockages can then cause flooding, creating a breeding ground for pathogens that facilitate the spread of disease [2], [3]. Quantitative studies have demonstrated the impact of solid accumulation on water channels: in the Isale Koko case study, solid wastes occupied up to 20% of drainage volume [4], while [5] found that near-complete culvert blockage increased flood volume by about 40% under extreme rainfall. In Indonesia, urban canals that connect to major rivers, such as the Citarum, have suffered from severe water quality deterioration [6]. They reported that 54% of river water is heavily polluted and BOD and COD concentrations exceed national standards, mainly due to untreated domestic and textile industrial effluents. Approximately Rp0.8 trillion was required for building solid waste infrastructure over two decades [7]. Then, according to the United Nations Environment Programme (UNEP), more than 80% of marine litter originates from rivers and canals, which act as conduits for mixed household and industrial waste flowing downstream into coastal ecosystems [8]. In Indonesia and other Southeast Asian countries, urban canals function as important drainage infrastructures. However, their performance is frequently

***Corresponding author: Najirah Umar (najirah@handayani.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i1.1066

compromised by accumulated waste such as plastics, organic matter, metal, and e-waste [9]. Thus, canal blockages contribute significantly to recurrent flooding events, with serious socio-economic consequences [10]. Therefore, developing intelligent, automated systems for canal waste classification is essential, both to support sustainable waste management.

Research on CNN-based waste classification has advanced considerably in recent years. Most prior studies have primarily focused on single-label image, in which each image is associated with only one object class and neglecting the complexity of real-world waste occurrences. Such studies commonly relied on datasets collected under controlled environmental conditions, which therefore lack representation of real-world scenarios. Waste objects were typically placed on surfaces with minimal interference from shadows, reflections, or lighting variations. For instance, [11] extended the TrashNet dataset by adding an additional "organic" class con-sisting of seven classes to reach a total of 3,242 images and compared seven CNN models such as InceptionV3, InceptionRes-NetV2, Xception, VGG19, MobileNet, ResNet50 and DenseNet201, reporting the highest validation accuracy of 95.19% with DenseNet201. Similarly, [12] compared AlexNet, GoogLeNet, DenseNet201, InceptionResNetV2, InceptionV3, MobileNetV2, XceptionNet, ShuffleNet, ResNet 18, ResNet 50, and ResNet 101 for the classification of three types of beverage container waste in a Reverse Vending Machine (RVM) system sucha as drink carton boxes, bottles, and aluminium cans with 500 image datasets. Resulted the AlexNet network gave the best F1-score, 97.50%. Other studies have introduced ensemble-based approaches, such as [13], who proposed the EnCNN-UPMWS model by integrating GoogLeNet, ResNet-50, and MobileNetV2. Two datasets used, namely TrashNet and FourTrash, which was constructed by collecting a total of 47,332 common HSW images.

The overall accuracy of the proposed model for the four classes (wet waste, recyclables, harmful waste, and dry waste) was 92.85%. Several works emphasized optimization strategies; for example, [14] applied data augmentation and hyperparameter tuning. applied data augmentation and hyperparameter tuning. They utilized an open Kaggle dataset that provides 25,077 images. The dataset comprises 13,966 organic and 11,111 recyclable images. The Improved-CNN model recorded a high accuracy of 94.40%, outperforming existing backbones including ResNet-50, Inception-V3, and VGG-19. Whereas [15] developed a three-stage system based on DP-CNN, En-ELM, and XAI to classify 36 waste categories. The En-ELM model yielded the highest performance, recording an F1-score of 84.54% and an Accuracy of 85.25%, [16] proposed a deep neural network for Smart systems (LADS) models to identify and classify waste materials. They utilized a waste image dataset comprising nearly 25,077 photos divided into organic and recyclable categories. Their proposed method achieved an accuracy of 94.53%, outperforming the AlexNet, VGG16, and ResNet34 models.

Subsequent research has expanded to multi-class classification scenarios, wherein each image is assigned a single label from several available categories. The majority of these studies have aimed to improve model accuracy and efficiency in differentiating waste categories. For instance, [17] proposed an automatic waste classification system based on deep learning by optimizing the ResNet-34 architecture through multi-resolution feature fusion. Utilizing a waste image dataset of 4,168 online pictures across eight classes, the classification achieved an accuracy of 99%. Similarly, [18] optimized several Deep CNN models for waste classification. The researchers utilized the TrashNet dataset, which contains 2,527 images across categories including glass, paper, cardboard, plastic, and metal. The results showed that the Fine-Tuned DenseNet169 model as the best performer, achieving an accuracy of 96.42% and an F1-score of 96%. Similarly, [19] developed a multi-layer CNN approach for waste classification using two class categories with a dataset of 25,077 images. The results showed that their Improved DCNN model achieved an accuracy of 93.28%, outperforming models based on VGG16, VGG19, MobileNetV2, DenseNet121, and EfficientNetB0. Likewise, [20] proposed the ECCDN-Net technique for classifying organic and recyclable waste. This model shows performance that outperforms other pre-trained models, such as ResNet18, MobileNetV2, InceptionV3, and DenseNet201, with an accuracy of 96.10%. Although positioned as multi-class, these approaches essentially implement single-label classification.

This trend has gradually shifted toward multi-label classification research, which aims to recognize multiple objects within a single image. For example, [21] employed YOLOv8/v5 to detect multiple trash objects accumulating around containers in real-time. Although achieving an accuracy of 0.96, the approach was limited by issues in the speed of automated waste calculation and its inability to account for all types of waste. Subsequently, [22] introduced an efficient

Query2Label (Q2L) framework, powered by the Vision Transformer (ViT-B/16) as its backbone, achieving an accuracy of 92.36% and outperforming the ResNet-101 model. Similarly, [23] proposed a multi-label waste classification model called YOLO-WASTE based on transfer learning. Experimental results showed that their model achieved a mAP value of 92.23%, demonstrating significantly better performance compared to other image classification algorithms. Nevertheless, the datasets they employed remained limited, containing relatively clean and intact images without visual disturbances such as dented bottles, soiled plastics, or wet paper. Consequently, these studies did not fully capture the complexities of real-world scenarios, particularly in canal or drainage environments. In this context, a single image may contain various types of waste under challenging conditions, including heterogeneous or dynamic backgrounds (e.g., turbid water, mud), occlusion (partially hidden by water or soil), physical deformation, and overlapping objects. These factors render traditional single-label methods less effective [24]. Further emphasized that most existing classification models fail to generalize to multi-object scenarios due to the lack of effective multi-label recognition strategies [22].

This addresses the research gap; we propose a deep learning framework for automating multi-label waste classification in canal environments. The framework introduces the ConvNeXt-Tiny model as the feature extraction component. ConvNeXt-Tiny is a small-scale variant within the ConvNeXt model family, which modernizes traditional Convolutional Neural Networks (CNNs). ConvNeXt was selected due to its modern CNN architecture, which incorporates key design elements of Vision Transformers (ViT). It has been demonstrated to outperform many state-of-the-art architectures in classification tasks [25], while maintaining computational efficiency [26] and strong capability in extracting complex visual features [27], [28].

The main contributions of this research are as follows: We present a novel approach for multi-label waste classification in canal environments, which involves heterogeneous objects with occlusion, deformation, and overlap. The proposed model employs ConvNeXt-Tiny has been increased through fine-tuned and hyperparameter optimization techniques, and integrating a modified two-layer non-linear MLP with dropout for improving multi-label classification. A hybrid data augmentation strategy combining SMOTE and MixUp is applied to improve model performance on visually diverse waste. For a comparative analysis, the performance of the proposed model was evaluated with ten other fine-tuned state-of-the-art CNN architectures, including ResNet18, ResNet50, VGG16, VGG19, DenseNet121, MobileNetV2, and EfficientNet (B0–B3).

In addition to the introduction, this article comprises four further sections. Section 2 outlines the methods employed in this study, including existing approaches, a general overview of model architectures, the proposed method ConvNeXt-Tiny with an MLP classifier along with the dataset and performance results. Section 3 provides a detailed description of the experimental settings for canal waste image multi-label classification, followed by an analysis of model performance, accuracy, F1-score, confusion matrix, and the time versus F1-score trade-off, which collectively represent the outcomes of the proposed model. Finally, the conclusion is presented in Section 4.

2. The Proposed Method

2.1. Model Architecture

The architecture of ConvNeXt comprises a sequence of hierarchical ConvNeXt Blocks, each containing a depthwise convolution layer with a large kernel size (7×7) to capture long-range spatial dependencies [25]. The output of this convolution is normalized using Layer Normalization, while Gaussian Error Linear Units (GELU) serve as the activation function, applying a smooth weighting based on a Gaussian distribution [29]. This block design reduces parameters and computational cost compared to standard convolutions, while retaining high representational capacity.

ConvNeXt-Tiny follows a four-stage hierarchical structure, where the resolution of the feature maps is gradually reduced across stages: Step 1, Input images are first passed through a stem convolution (4×4 stride 4), which reduces the resolution to 1/4 of the original image. Features are then processed by several ConvNeXt blocks. Step 2, The feature maps are downsampled by a stride-2 convolution, producing resolution 1/8 of the original image, followed by another stack of ConvNeXt blocks. Step 3, The resolution is further reduced to 1/16, and ConvNeXt blocks extract increasingly abstract representations. Step 4, A final downsampling step yields 1/32 resolution, where the deepest ConvNeXt blocks operate.

 ISSN 2723-6471

Between each stage, downsampling is performed using stride-2 convolutions. Unlike segmentation or detection variants, the classifier form of ConvNeXt-Tiny does not include upsampling; instead, it produces a compact representation for classification tasks. Finally, the output features of the standard ConvNeXt-Tiny processed by a linear layer.

2.2. Proposed Method: ConvNeXt-Tiny with MLP Classifier

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

The proposed model is depicted in figure 1. In the original ConvNeXt architecture, the classifier head consists of a LayerNorm followed by a single linear layer. In this study, the head was restructured into a two-layer MLP, comprising a fully connected layer with 512 hidden units using ReLU activation and dropout (p=0.3), followed by an output layer with the number of neurons corresponding to the classes in the dataset (see table 2). During evaluation, the final logits are passed through a sigmoid activation to enable multi-label classification. The addition of a non-linear head (MLP) has been reported to improve representation quality and transfer performance in downstream tasks across various studies on self-supervised and transfer learning [30], [31], [32].

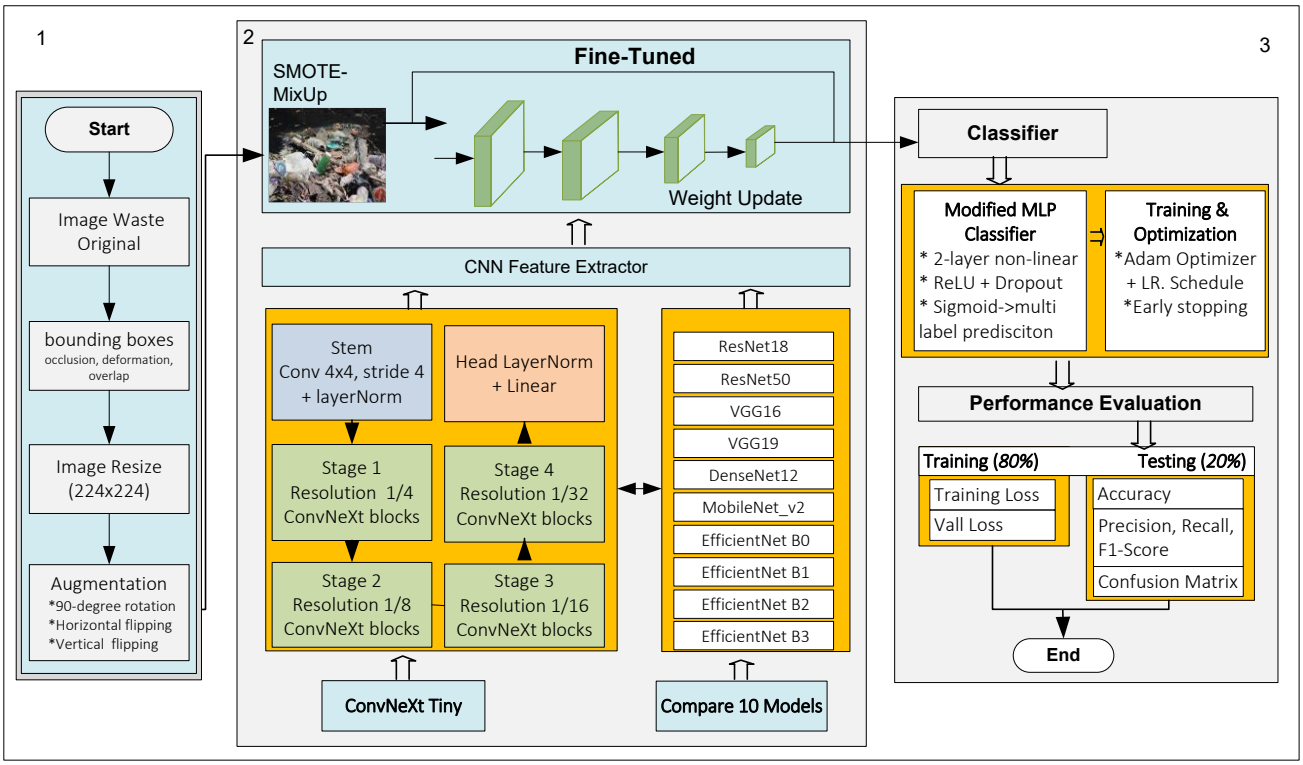


Figure 1. Proposed ConvNeXt and MLP multi-label waste classification in canal.

To train the ConvNeXt model for multi-object waste classification, this system employs a multi-task learning approach consisting of two main components: multi-label classification loss and bounding box regression loss. For the classification component, Binary Cross-Entropy (BCE) Loss with a sigmoid function is applied, which is suitable for multi-label scenarios since a single image can contain more than one waste category [33], [34], formulated as:

$$L_{BCE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{C}\sum_{c=1}^{C}\left[-y_{i,c}\, log\left(\sigma\left(z_{i,c}\right)+\varepsilon\right)-\left(1-\ y_{i,c}\right) log\ \left(1-\ \sigma\left(z_{i,c}\right)+\varepsilon\right)\right] \quad (1)$$

 ISSN 2723-6471

N = number of samples in a mini-batch; C= number of classes; $y_{i,c} \in \{0,1\}$= ground-truth label of class c for sample i; $z_{i,c} \in \mathbb{R}$ = logit output for class c; $\sigma(\cdot)$= logistic sigmoid function; ε = small constant for logarithmic numerical stability. C is the number of classes, $y_c \in \{0,1\}$ is the label for class C, z_c is the output logit for class c, and σ (·) is the sigmoid function. To address class imbalance, a weighting factor, α_c (positive weight), is utilized as shown in Formula 2.

$$L_{cls} = \frac{1}{C}\sum_{c=1}^{C}(-\alpha_c y_c \log\sigma(z_c) - (1 - y_c)\log(1 - \sigma(z_c))) \quad (2)$$

For bounding box regression of (x,y,w,h), Smooth L1 Loss is used, which is more stable, [35] defined as in Formula 3:

$$SmoothL1_\beta(\Delta) = \begin{cases} \frac{1}{2}\frac{\Delta^2}{\beta}, & if\ |\Delta| < \beta, \\ |\Delta| - \frac{\beta}{2}, & other, \end{cases} \quad (3)$$

With $\Delta = \hat{b} - b^{\backslash *}$ which is the difference between the predicted bounding box $\hat{b}$ and the ground truth $b^{\backslash *}$The value of β is typically set to 1. These two loss functions are then integrated within a multi-task framework [36], where the total loss is defined in Formula 4:

$$L_{total} = \lambda_{cls} \cdot \frac{1}{N}\sum_{i=1}^{N} L_{cls}^{(i)} + \lambda_{reg} \cdot \frac{1}{N_+}\sum_{j=1}^{N_+} L_{reg}^{(j)}, \quad (4)$$

N is the number of samples in a mini-batch, N_+ is the number of positive anchors/proposals, λ_{cls} and λ_{reg} serves as the balancing coefficient between the classification and regression losses.

2.3. Model Optimization Through SMOTE-MixUp Balancing

As shown in table 1, the class distribution in the dataset is highly imbalanced, with the number of samples varying across classes. For instance, the Inorganic Waste class contains 392 samples, whereas Kite, Glass Bottle, Paper, cardboard, Beverage Carton, Metal, Rubber, have only 9, 18, 20, 21, 27, 33 and 33 samples, respectively. Such an imbalance can bias the model toward the majority classes, leading to suboptimal performance in recognizing minority classes [37]. To overcome this imbalance, the Synthetic Minority Oversampling Technique (SMOTE) applied to generate synthetic samples in the feature of minority classes [38]. However, since the dataset employed in this study consists of image data, applying SMOTE alone is considered insufficient to capture the complexity and visual variability across classes [39]. Therefore, this study proposes a hybrid strategy that combines SMOTE and MixUp to optimize the model performance by linearly combining two image samples and their respective labels [40]. Table 1 presents a comparison of the class distribution before and after applying the balancing technique, showing an improvement in the representation of underrepresented classes across the dataset.

2.4. Dataset

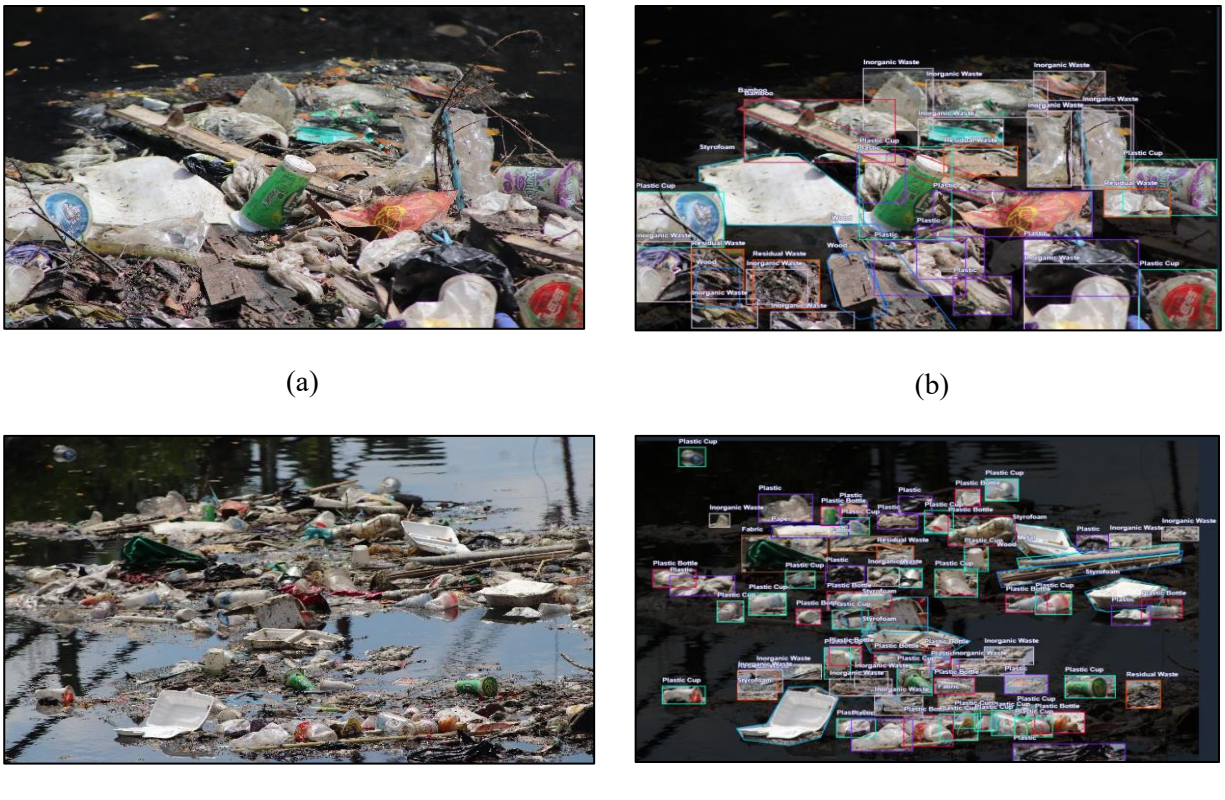
The garbage dataset was collected from two primary canal locations in Makassar City, Indonesia: The Pannampu Canal in Tallo District, located near a traditional market area, and the Jongaya 1 Canal in Rappocini District, situated close to a residential area and housing complexes. These locations were chosen to represent both commercial and residential waste characteristics, thus ensuring the diversity of the collected waste. Data collection was performed during midday and afternoon under clear weather conditions to ensure the dataset's robustness for model training and evaluation. Then images were captured using a camera with a 24.2 MP APS-C CMOS sensor and a DJI Mavic 3 drone, which features a dual-camera system with a 20 MP Hasselblad 4/3 CMOS sensor and a 12 MP 1/2-inch CMOS sensor. This equipment was used to generate high-quality images from the target environment. The collected dataset contains 855 images with a total of 2,662 garbage objects, categorized into 18 classes (see table 1) and each image was resized to 640 x 640 pixels before being used for model training [41], [42].

2.4.1. Data Annotation

In this study, annotations were used to define object boundaries, assign class labels, and ensure the proper representation of multiple objects within a single image. The annotation process was carried out using Roboflow software to create bounding boxes around the garbage objects. Each object in the image was labeled according to predefined categories, including organic and inorganic waste, with the addition of sub-classes where necessary, as shown in figure 2. The annotation results were stored in CSV format [27]. allowing for easy integration into the model training pipeline. During the annotation phase, several challenges were encountered. Garbage objects often have irregular shapes, are subject to occlusion (overlapping), and exhibit variations in size and lighting conditions, all of which complicate the process of defining object boundaries. Overall, the annotation process resulted in a total of 18 classes. The class distribution of the dataset is detailed in table 1.

Table 1. Comparison of the number of objects per class before and after applying the SMOTE–MixUp technique.

No	Class	Before Balancing	After SMOTE-MixUp
1	Bamboo	74	2126
2	Beverage Carton	25	1008
3	Cardboard	17	1092
4	Fabric	100	2661
5	Glass Bottle	15	683
6	Inorganic Waste	315	3120
7	Kite	6	465
8	Leaf	255	3229
9	Metal	29	1054
10	Organic Waste	72	2159
11	Paper	18	801
12	Plastic	281	3178
13	Plastic Bottle	201	3141
14	Plastic Cup	349	3581
15	Residual Waste	114	2475
16	Rubber	29	1023
17	Styrofoam	145	2811
18	Wood	115	2713



(a) (b) (c) (d)

Figure 2. Multi-object image annotation: (a, c) original images, and (b, d) image annotations.

2.4.2. Data Annotation

To enhance the robustness and generalization capability of the proposed model, several image augmentation strategies were applied to the training dataset [43]. The importance of data augmentation to mitigate the risk of overfitting [44], particularly when working with limited datasets [45]. In this study, three primary augmentation operations were utilized: horizontal flipping, vertical flipping, and 90-degree rotation. The steps were performed, such as resizing all images to a uniform resolution of 640 × 640 pixels. Furthermore, pixel value normalization was applied to accelerate convergence during training [46]. The aims to facilitates more balanced learning across diverse waste categories, including plastic, organic waste, paper, and glass [14], [47].

3. Result and Discussion

Here, we present the results of experiments on waste image classification. In this section, the outcomes of the model performance analysis are first described and analyzed. These findings are then evaluated, highlighting the contributions of our proposed approach.

3.1. Result

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

3.1.1. Experimental Settings for Canal waste image multi label classification

The ConvNeXt-Tiny experiments were compared with pretrained CNN backbones, including VGG16, VGG19, ResNet18, ResNet50, DenseNet121, MobileNetV2, EfficientNet-B0, EfficientNet-B2 and EfficientNet-B3 [25], [48], [49], [50], [51], [52], which were employed as feature extractors. To maintain experimental validity, all models were fine-tuned with same hyperparameters and training settings. All models received 224 × 224 input images and were trained with a batch size of 16 for up to 20 epochs using early stopping (patience = 5). We used the Adam optimizer and learning rate is 1×10^{-4}) Each model employed the same two-layer MLP classifier (512 hidden units, ReLU activation, dropout = 0.3) to maintain consistency across backbones (see table 2). All experiments in this study were conducted using Jupyter Notebook within the Anaconda environment. The hardware and operating system setup consisted of a laptop GPU with 8 GB of VRAM, 12 GB of virtual RAM, and Windows 11 as the operating system. The dataset was partitioned into two subsets, comprising 80% of the data for training and 20% for testing. This split was implemented to ensure that the model could learn effectively from the majority of the data.

Table 2. Model parameters and training settings for waste multi label classification

Parameters	Value	Description
Backbone Models	VGG16, VGG19, ResNet18, ResNet50, DenseNet121, MobileNetV2, EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, EfficientNet-B3, ConvNeXt-Tiny	CNN models utilized as the feature extractor.
Input Image Size	224 × 224	Image resolution during the preprocessing.
Batch Size	16	Number of samples per training/validation batch.
Epoch	Maks. 20	Upper bound of training iterations, with early stopping (patience = 5).
Optimizer	Adam (learning rate = 1×10^{-4})	Optimization algorithm for parameter updates.
Loss Function	BCEWithLogitsLoss	Binary Cross-Entropy with Logits, suitable for multi-label classification
Scheduler	ReduceLROnPlateau	Mode = ‘min’, factor = 0.1, patience = 3
Hidden Dim (MLP)	512	Number of hidden units in the intermediate fully connected layer.
MLP Architecture	Fully Connected (Input → 512) → ReLU → Dropout (p = 0.3) → Fully Connected (512)	Two-layer non-linear classifier head. Output layer neurons equal the number of target classes.

 ISSN 2723-6471

DataAugmentation	SMOTE-based MixUp (λ ~ Beta (α, α), α = 0.4)	Synthetic oversampling value via linear mixing of two images and labels, with λ ~ Beta (0.4, 0.4) to improve diversity and balance class distribution.

3.1.2. Model Performance

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

Table 3 presents the performance comparison before and after applying the SMOTE–MixUp technique. The results indicate that the impact of the SMOTE–MixUp technique varies across backbone architectures, showing performance improvements in most CNN models. Models such as vgg16, densenet121, mobilenet_v2, efficientnet_b1, efficientnet_b3, and convnext_tiny achieved improved F1-scores, indicating their robustness in handling synthetic interpolation. Model such as vgg19, resnet18, resnet50, efficientnet_b0, and efficientnet_b2 experienced a slight performance drop. In contrast, in terms of accuracy, only the performance of VGG19 and ResNet50 slightly decreased, while other models showed improvement after applying SMOTE–MixUp.The visualization of model performance as shown in figure 4.

Table 3. Comparison Table of Accuracy, Precision, Recall, and F1-Score before and after balancing

		Before balancing				After SMOTE-MixUp			
Model ID	**CNN-Backbone**	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**
v16	vgg16	0.77	0.97	0.88	0.91	0.80	0.98	0.92	0.94
v19	vgg19	0.72	0.85	0.84	0.83	**0.60**	0.78	0.72	0.74
r18	resnet18	0.82	0.97	0.89	0.92	0.84	0.93	0.89	0.91
r50	resnet50	0.83	0.97	0.86	0.89	**0.77**	0.86	0.80	0.82
d121	densenet121	0.86	0.98	0.89	0.92	0.86	0.98	0.93	0.95
mv2	mobilenet_v2	0.64	0.78	0.70	0.72	0.69	0.73	0.68	0.70
eb0	efficientnet_b0	0.75	0.93	0.82	0.85	0.79	0.91	0.79	0.82
eb1	efficientnet_b1	0.48	0.83	0.62	0.68	0.67	0.88	0.82	0.83
eb2	efficientnet_b2	0.78	0.94	0.86	0.89	0.82	0.87	0.83	0.84
eb3	efficientnet_b3	0.75	0.91	0.83	0.86	0.80	0.91	0.84	0.86
conv-t	**convnext_tiny**	**0.89**	**0.99**	**0.91**	**0.94**	**0.91**	**0.99**	**0.99**	**0.99**

Figure 3, show the learning curves across all models, show a consistent decrease in training loss, indicating effective convergence within 20 epochs. Among them, ConvNeXt-Tiny and ResNet18 demonstrate the most stable performance, with validation loss steadily decreasing in parallel with training loss until both reach very low values (≈0.02). DenseNet12 also converges well, though slightly slower. In contrast, VGG16/19 achieve very low training loss (<0.03) but maintain higher validation loss (~0.04–0.06), indicating mild overfitting. Similarly, MobileNetV2 and EfficientNet-B1/B3 converge quickly at the training level but exhibit larger train–validation gaps, which align with their lower F1-scores. The analysis of loss curves confirms that modern architectures not only outperform in accuracy but also in training stability.

ISSN 2723-6471

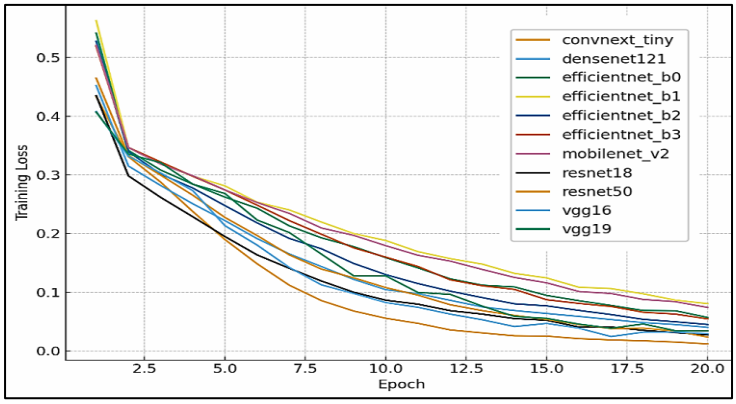


Figure 3. Training loss across CNN backbone (20 epoch)

Table 4 shows a comparison of the F1-scores across 18 waste categories before and after applying the Smote-MixUp technique and indicates that most models achieve high F1-scores across many classes. VGG16, Resnet18, DenseNet121, and ConvNeXt-Tiny can recognize all the classes, while VGG19, ResNet50, MobileNet_V2, EfficientNet_B0, EfficientNet_B1, EfficientNet_B2, and EfficientNet_B3 fail to recognize specific classes (e.g., Cardboard, Kite, Paper, and Rubber with F1 = 0.00). These discrepancies are due to the λ∼Beta (0.4,0.4) probalictic distribution, which produces strong interpolation effects. The applied λ value is not equally suitable for all backbone architectures. Models that successfully recognize all classes have a feature extractor that is more adaptive to data variation. Consequently, they are capable of extracting important patterns from minority classes despite the noise caused by interpolation. Applying Smote-MixUp to the training set yields consistent gains for ConvNeXt-Tiny, most notably on low-support categories. Paper improves from 0.50 to 1.00, while plastic rises from 0.93 to 0.964; Rubber and Organic Waste remain stable at 1.00. Overall, the macro-average across all classes increases from 0.942 to 0.988. Small declines on several majority classes (e.g., Inorganic Waste, Leaf, Plastic Cup, Residual Waste) indicate the usual precision–recall trade-off introduced by minority oversampling, which we mitigate with class-specific decision thresholds.

Table 4. Comparison of F1-scores before SMOTE (B), after SMOTE (A), and Δ improvement across various models.

Name of Class	v16	v19	r18	r50	d121	mv2	eb0	eb1	eb2	eb3	conv-t
	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ	B, A, Δ
Bamboo	0.905	0.909	0.884	0.930	0.952	0.870	0.905	0.837	0.878	0.818	0.976
	0.909	0.818	0.941	0.875	1	0.875	0.971	0.944	1	1	1
	0.004	-0.091	0.057	-0.055	0.048	0.005	0.066	0.107	0.122	0.182	0.024
Beverage Carton	1	1	1	0.889	1	0.857	1	0.857	1	1	1
	1	1	1	1	1	1	1	1	1	1	1
	0	0	0	0.111	0.000	0.143	0	0.143	0	0	0
Cardboard	0.833	0.727	0.833	0.727	0.727	0	0.250	0.250	0.545	0	0.833
	0.857	0	0.857	0	0.667	0	0	0	0	0	1
	0.024	-0.727	0.024	-0.727	-0.060	0.000	-0.250	-0.250	-0.545	0.000	0.167
Fabric	1	0.933	1	1	1	0.980	1	0.979	1	1	0.979
	1	0.872	1	0.974	1	0	1	0.923	1	1	1
	0	-0.061	0	-0.026	0.000	-0.980	0	-0.056	0	0	0.021
Glass Bottle	0.800	0.800	0.800	0.800	0.800	0.500	0.667	0.667	0.800	0.800	0.800
	1	0.800	1	0.800	1.000	0.667	0.857	0.857	1	1	1
	0.200	0	0.200	0	0.200	0.167	0.190	0.190	0.200	0.200	0.200
Inorganic Waste	0.909	0.976	0.963	0.975	0.982	0.963	0.957	0.927	0.957	0.982	1
	0.949	0.912	0.955	0.937	0.942	0.940	0.941	0.930	0.943	0.930	0.937
	0.040	-0.064	-0.008	-0.038	-0.040	-0.023	-0.016	0.003	-0.014	-0.052	-0.063
Kite	1	0,667	1	1	1	1	1	0	1	1	1
	1	0	1	1	1	0	0.500	1	0.800	0.500	1
	0	-0.667	0	0	0	-1.000	-0.500	1	-0.200	-0.500	0
Leaf	0.941	0.872	0.944	0.970	0.948	0.906	0.933	0.857	0.944	0.917	0.993
	0.963	0.882	0.957	0.948	0.955	0.940	0.941	0.896	0.962	0.963	0.963
	0.022	0.010	0.013	-0.022	0.007	0.034	0.008	0.039	0.018	0.046	-0.030

 ISSN 2723-6471

	1	0.800	1	0.667	1	0.667	0.909	0.600	0.909	0.800	1
Metal	1	0.600	1	0.889	0.857	0.857	0.857	1	0.857	0.750	1
	0	-0.200	0	0.222	-0.143	0.190	-0.052	0.400	-0.052	-0.050	0
	0.955	0.939	0.930	0.930	0.930	0.821	0.917	0.757	0.958	0.955	1
Organic Waste	0.957	0.780	0.933	0.957	0.957	0.780	0.909	0.844	0.933	0.957	1
	0.002	-0.159	0.003	0.027	0.027	-0.041	-0.008	0.087	-0.025	0.002	0
	0.500	0	0.500	0.500	0.500	0	0.444	0	0.500	0.500	0.500
Paper	0.667	0.667	0	0.667	1	0	0.667	0.667	0	0.667	1
	0.167	0.667	-0.500	0.167	0.500	0	0.223	0.667	-0.500	0.167	0.500
	0.943	0.933	0.897	0.930	0.923	0.908	0.914	0.815	0.917	0.923	0.930
Plastic	0.949	0.881	0.915	0.957	0.957	0.936	0.950	0.908	0.964	0.936	0.964
	0.006	-0.052	0.018	0.027	0.034	0.028	0.036	0.093	0.047	0.013	0.034
	0.963	0.981	0.990	0.962	1	0.971	0.990	0.898	1	1	1
Plastic Bottle	0.927	0.905	1	0.951	1	0.988	0.976	0.941	0.965	0.976	1
	-0.036	-0.076	0.010	-0.011	0	0.017	-0.014	0.043	-0.035	-0.024	0
	0.952	0.917	0.966	0.976	0.958	0.960	0.947	0.947	0.966	0.959	0.988
Plastic Cup	0.943	0.916	0.949	0.974	0.987	0.962	0.955	0.938	0.949	0.949	0.968
	-0.009	-0.001	-0.017	-0.002	0.029	0.002	0.008	-0.009	-0.017	-0.010	-0.020
	0.896	0.933	0.896	0.958	0.943	0.845	0.932	0.781	0.928	0.882	0.958
Residual Waste	0.889	0.800	0.906	0.897	0.929	0.807	0.926	0.857	0.935	0.926	0.945
	-0.007	-0.133	0.010	-0.061	-0.014	-0.038	-0.006	0.076	0.007	0.044	-0.013
	1	0.833	1	1	1	0	0.667	0.444	0.769	1	1
Rubber	1	0.667	1	0	0.889	0	0.400	0.400	0.857	1	1
	0	-0.166	0	-1	-0.111	0	-0.267	-0.044	0.088	0	0
	1	0.976	1	1	1	0.976	1	0.923	1	0.976	1
Styrofoam	0.984	0.951	0.952	0.984	1	0.984	0.985	0.923	1	0.985	1
	-0.016	-0.025	-0.048	-0.016	0	0.008	-0.015	0	0	0.009	0
	0.938	0.753	1	0.968	1	0.871	0.899	0.862	0.970	1	1
Wood	0.939	0.840	0.958	0.917	0.958	0.894	1	0.936	1	0.980	1
	0.001	0.087	-0.042	-0.051	-0.042	0.023	0.101	0.074	0.030	-0.020	0

Figure 4 shows, applying the balancing technique yields consistent gains for ConvNeXt-Tiny model, most notably on low-support categories. Paper improves from 0.50 to 1.00, while plastic rises from 0.93 to 0.964; Rubber and Organic Waste remain stable at 1.00. Overall, the macro-average across all classes increases from 0.942 to 0.988. Small declines on several majority classes (e.g., Inorganic Waste, Leaf, Plastic Cup, Residual Waste) indicate the usual precision–recall trade-off introduced by minority oversampling, which we mitigate with class-specific decision thresholds.

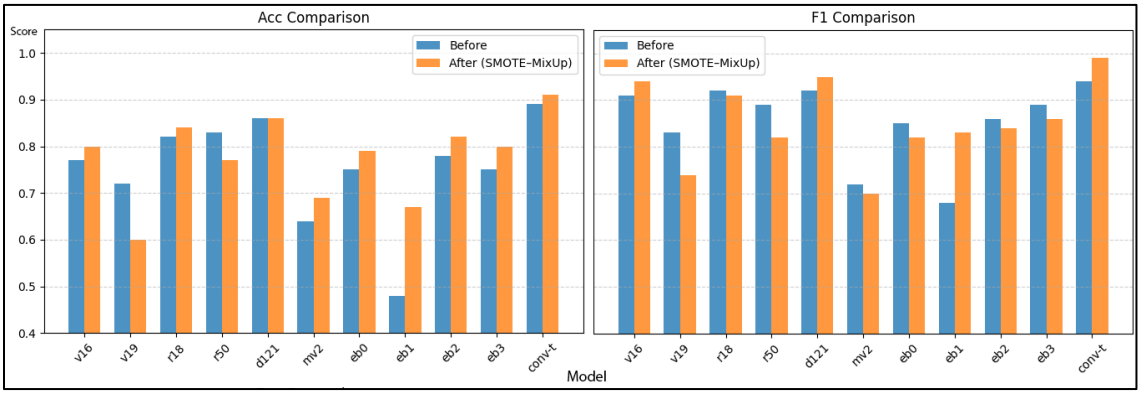


Figure 4. Comparison of Accuracy and F1-score across different CNN Backbones

Figure 5 shows the per-class confusion matrices for the ConvNeXt-Tiny classifier. Overall, performance is strong across most categories, with perfect separation (no errors) for Bamboo (TN = 153, TP = 18), Beverage Carton (TN = 169, TP = 2), Fabric (TN = 151, TP = 20), Glass Bottle (TN = 168, TP = 3), Kite (TN = 168, TP = 3), Paper (TN = 169, TP = 2), Styrofoam (TN = 139, TP = 32), Rubber (TN = 167, TP = 4), and Wood (TN = 146, TP = 25). Several classes exhibit moderate error rates, including Inorganic Waste (TP = 74, FP = 7, FN = 3), Leaf (TP = 65, FP = 1, FN = 4), Plastic (TP = 67, FP = 2, FN = 3), and Plastic Cup (TP = 75, FP = 5). A few categories show pronounced failures characterized by many false negatives and no true positives, namely Cardboard (FN = 4), Organic Waste (FN = 24), and Plastic Bottle (FN = 41). Residual Waste remains robust (TP = 26) but still includes three false negatives. In summary, most categories achieve high true-positive and true-negative counts, whereas a subset, likely those with

 ISSN 2723-6471

stronger visual overlap with other classes or limited training samples, remains a primary source of confusion for the model.

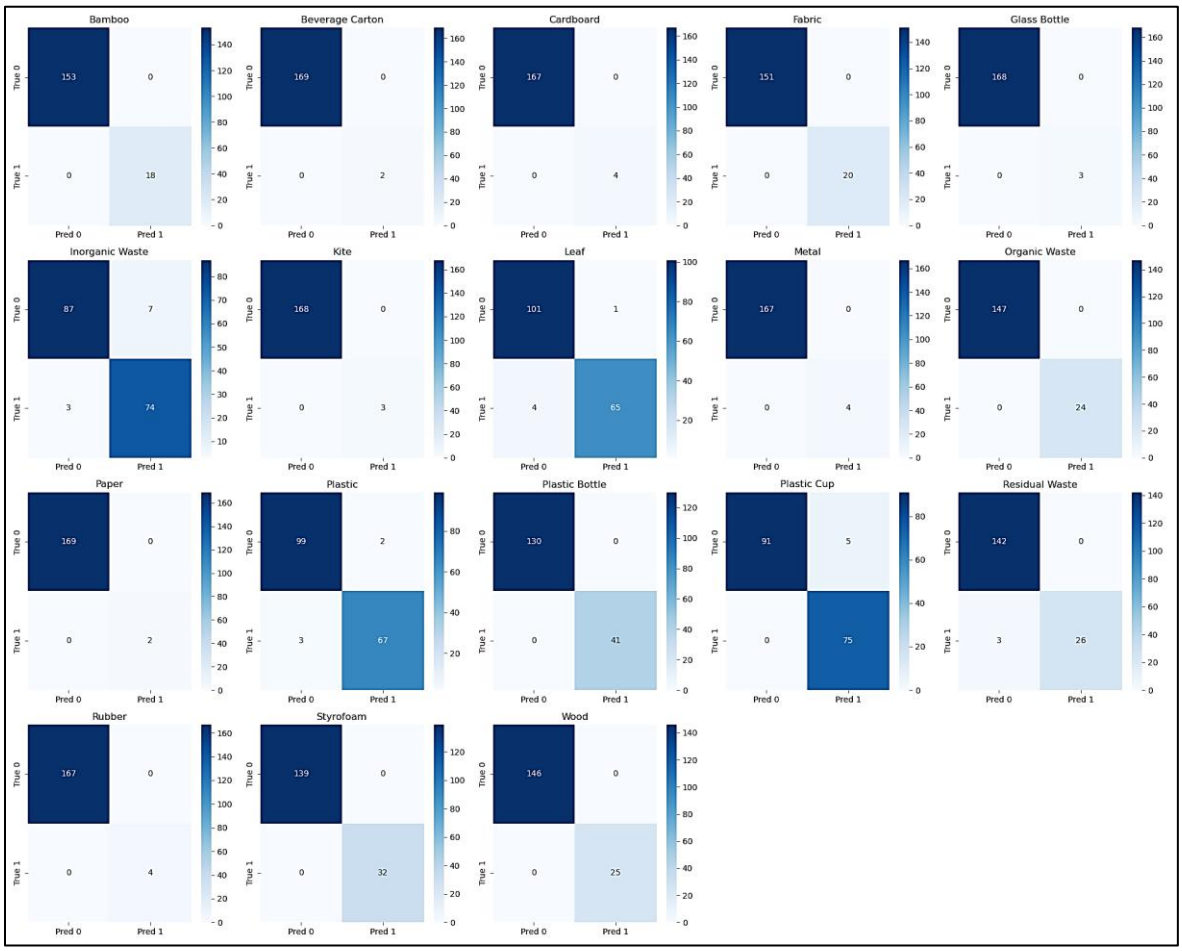


Figure 5. Confusion Matrix for the ConvNeXt-Tiny Backbone by Class.

3.2. Discussion

Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive."

3.2.1. Accuracy, F1-Score and Confusion Matrix

The results demonstrate that ConvNeXt-Tiny delivers the best overall performance for multi-object waste classification in canal environments, achieving an F1-score of 0.94 (see table 3). Its modernized design enables more effective feature extraction compared to classical CNNs such as ResNet18 and DenseNet121, which still performed competitively (F1 = 0.92). VGG16 also showed strong precision but with lower accuracy, while lightweight models such as MobileNet_v2 and EfficientNet-B1 underperformed to represent complex patterns and features from data.

Figure 3 shows rapid initial convergence followed by a widening gap and diminishing improvement, indicating mild overfitting and the influence of class imbalance; in the F1 heatmap, categories such as Beverage Carton, Styrofoam, and Wood remain consistently high, whereas Plastic, Rubber, Organic Waste, and Paper are problematic due to visual similarity, high intra-class variability, and limited positive support, leading to reduced precision or coverage. Practically, this motivates balancing strategies (e.g., SMOTE) for minority classes, hard negative mining near class boundaries, and class-specific thresholding with probability calibration to constrain false positives, complemented by reporting per-class precision–recall curves, annotated failure cases, and confidence intervals for metrics so that the figures convey explanatory insight rather than mere summary.

Table 3 shows in ConvNeXt-Tiny, SMOTE-MixUp provides the strongest synergy: recall rises from 0.91 to 0.99 while precision remains at 0.99, yielding an F1 increase from 0.94 to 0.99 and accuracy from 0.89 to 0.91 indicating the most robust generalization. Beyond that, the effects vary: DenseNet121 also improves (Recall 0.93; F1 0.95; precision remains 0.98) and VGG16 benefits (Acc 0.80; Rec 0.92; F1 0.94). EfficientNet-B1 increases substantially (Acc 0.67; Rec 0.82; F1 0.83), whereas ResNet18 is relatively stable but with reduced precision. EfficientNet-B3 tends to be stable; MobileNetV2 and EfficientNet-B0 show only slight accuracy gains while F1 declines (capacity-limited). In contrast, VGG19, ResNet50, and EfficientNet-B2 exhibit declines.

Role of regression in multi-task learning. Incorporating the bounding-box regression loss $\mathcal{L}_{\text{reg}}$ alongside the multi-label classification loss $\mathcal{L}_{\text{BCE}}$improves spatial grounding and yields a consistent gain in classification, indicating that localization acts as an auxiliary signal that regularizes feature learning. The remaining errors concentrate on Plastic, Rubber, and Organic Waste, where inter-class resemblance and intra-class variability challenge both box placement and label assignment. Future improvements include class-specific detection thresholds, hard-negative mining near class boundaries, and tuning the balancing coefficient λ to trade off localization and classification.

As shown in table 4, these results confirm that feature-space Smote-MixUp effectively addresses class imbalance by lifting the performance of minority or visually ambiguous categories (e.g., Paper, Plastic), with only modest impacts on majority classes. In practice, we pair Smote-MixUp with hard-negative mining at class boundaries and per-class threshold calibration to preserve precision on frequent categories while sustaining improved recall on rare ones.

As shown in Figure 5, the confusion matrix results provide deeper evidence of ConvNeXt-Tiny's robustness for multi-class waste classification in canal environments. The majority of categories are recognized with extremely high reliability, with several classes (Beverage Carton, Styrofoam, Wood) achieving perfect classification. This indicates that this model can effectively capture distinctive shape and texture features when inter-class similarity is low. Despite these issues, ConvNeXt-Tiny demonstrates consistent performance across both majority and minority categories, even in underrepresented classes such as Cardboard, Kite, Rubber, and Paper. Compared to lightweight CNNs that tend to collapse on minority classes.

3.2.2. Time vs F1-Score Trade-off

Table 5 presents a comparison of average training time in minutes across various CNN backbone architectures. The table includes three key metrics: Total_min (the cumulative training time for all epochs), Mean_min (the average training time per epoch), and Min_min (the shortest recorded time for a single epoch). It is evident that VGG19 exhibits the longest total training time at 29.7 minutes, with a relatively high average epoch duration of 0.92 minutes. This suggests that VGG19 is the most computationally high backbone. Similarly, VGG16 also shows a high average epoch time (0.78 minutes), though slightly more efficient than VGG19. In contrast, the MobileNet_v2 model demonstrates computational efficiency, requiring only 8.5 minutes in total, with a Mean_min of just 0.21 minutes per epoch. This is further supported by its Min_min value of 0.19 minutes, indicating not only efficiency but also consistency in training time across epochs. Other lightweight architectures such as ResNet18, EfficientNet_b0, and EfficientNet_b1 also perform well, maintaining both low average and minimum epoch durations. For example, EfficientNet_b0 trains in 9.1 total minutes with a stable average of 0.22 minutes per epoch, and a Min_min of 0.22 minutes, reflecting a consistent runtime.

Table 5. Comparison of average training time in minutes (min) across CNN Backbones

section	Total_min	Mean_min	Min_min
mobilenet_v2	8.5	0.21	0.19
resnet18	8.8	0.22	0.21
efficientnet_b0	9.1	0.22	0.22
resnet50	9.7	0.29	0.27
efficientnet_b1	10.6	0.26	0.25
efficientnet_b2	10.6	0.30	0.28
densenet121	15.4	0.43	0.40
convnext_tiny	17.1	0.42	0.40
efficientnet_b3	18.7	0.46	0.44

vgg16	22.5	0.78	0.40
vgg19	29.7	0.92	0.57

4. Conclusion

This study evaluated eleven CNN backbones for multi-object waste classification in canal environments. Among the models, ConvNeXt-Tiny consistently delivered the highest accuracy a 0.91 supported by F1 a 0.99), demonstrating robustness across both majority and minority classes, as confirmed by the backbone's comparison and confusion matrix analyses. It proved capable of handling class imbalance while maintaining stable recognition for underrepresented categories such as Kite, Glass Bottle, Paper, cardboard, Beverage Carton, Metal, and Rubber. The trade-off analysis further showed that while ConvNeXt-Tiny offers the best predictive reliability, it requires longer training time (17.1 minutes). Based on the training time comparison in table 4, VGG19 and VGG16 are the most computationally less efficient models, requiring the longest training durations. In contrast, MobileNetV2, ResNet18, and EfficientNet-B0 demonstrate the most efficient and consistent training performance, combining low total and average epoch times but with low accuracy. Meanwhile, ConvNeXt-Tiny achieves the highest classification accuracy, making it ideal for applications that prioritize predictive performance over training efficiency, whereas mobilenet_v2 and resnet18 represents the best trade-off for training efficiency. These findings provide a solid baseline for developing automated waste monitoring systems that support sustainable environmental management.

5. Declarations

5.1. Author Contributions

Conceptualization, N.U., B.E.W.A., and Y.; Methodology, N.U. and B.E.W.A.; Software, B.E.W.A. and Y.; Validation, N.U. and B.E.W.A.; Formal Analysis, N.U.; Investigation, B.E.W.A. and Y.; Resources, N.U. and B.E.W.A.; Data Curation, Y.; Writing—Original Draft Preparation, N.U.; Writing—Review and Editing, B.E.W.A. and Y.; Visualization, Y. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The IndoCanalWaste-ML dataset is publicly available upon request at:

https://data.brin.go.id/dataset.xhtml?persistentId=hdl:20.500.12690/RIN/TSL7JL

5.3. Funding

The authors would like to thank the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for the support and funding of this research through Research Contract Number: 130/C3/DT.05.00/PL/2025 through the Fundamental Research Scheme in 2025.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. Salvia, N. Zimmermann, C. Willan, J. Hale, H. Gitau, K. Muindi, E. Gichana, and M. Davies, "The wicked problem of waste management: An attention-based analysis of stakeholder behaviours," *J. Clean. Prod*., vol. 326, no. 12–2020, pp. 1–18, 2021, doi: 10.1016/j.jclepro.2021.129200.

[2] R. B. Pinto, T. H. M. van Emmerik, K. Duah, M. van der Ploeg, and R. Uijlenhoet, "Mismanaged plastic waste as a predictor for river plastic pollution," *Sci. Total Environ.*, vol. 951, no. 5, pp. 1-13, 2024, doi: 10.1016/j.scitotenv.2024.175463.

[3] V. Owatsakul, P. Panput, P. Jaisuda, and D. Rinchumphu, "Impacts of Wastewater Management and Enhancing the Landscape of the Mae Kha Canal: A Quasi-Experimental Study," *Water (Switzerland)*, vol. 17, no. 7, pp. 1–18, 2025, doi: 10.3390/w17071105.

[4] O. A. Mokuolu, A. K. Odunaike, J. O. Iji, and A. S. Aremu, "Assessing the Effects of Solid Wastes on Urban Flooding: A case study of Isale Koko," *LAUTECH J. Civ. Environ. Stud.*, vol. 9, no. 1, pp. 22–30, 2022, doi: 10.36108/laujoces/2202.90.0130.

[5] H. Olsson, S. N. Mugume, and J. Sörensen, "Modelling the effect of sediment and solid waste deposition on pluvial flood risk using in situ high-resolution rainfall and water level measurements," *Water Sci. Technol.*, vol. 91, no. 6, pp. 757–782, 2025, doi: 10.2166/wst.2025.039.

[6] F. F. Fitriana, D. Yudianto, S. Sanjaya, A. F. V Roy, and Y. C. Seo, "The Assessment of Citarum River Water Quality in Majalaya District, Bandung Regency," *Rekayasa Sipil*, vol. 17, no. 1, pp. 37–46, 2023, doi: 10.21776/ub.rekayasasipil.2023.017.01.6.

[7] Asian Development Bank, "Downstream Impacts of Water Pollution in the Upper Citarum River, West Java, Indonesia : Economic Assessment of Interventions to Improve Water Quality," *Downstr. Impacts Water Pollut. Up. Citarum River, West Java, Indones. Econ. Assess. Interv. to Improv. Water Qual.*, vol. 4, no. October, pp. 1–32, 2013, doi: 10.1596/17750.

[8] L. J. J. Meijer, T. van Emmerik, R. van der Ent, C. Schmidt, and L. Lebreton, "More than 1000 rivers account for 80% of global riverine plastic emissions into the ocean," *Sci. Adv.*, vol. 7, no. 18, pp. 1–13, 2021, doi: 10.1126/sciadv.aaz5803.

[9] D. Honingh, T. van Emmerik, W. Uijttewaal, H. Kardhana, O. Hoes, and N. van de Giesen, "Urban River Water Level Increase Through Plastic Waste Accumulation at a Rack Structure," *Front. Earth Sci.*, vol. 8, no. February, pp. 1–8, 2020, doi: 10.3389/feart.2020.00028.

[10] D. A. Abo-Sreeaa, N. M. AboulAtta, and D. A. El-Molla, "The impact of blockage on the performance of canal coverage structures," *J. Eng. Appl. Sci.*, vol. 70, no. 1, pp. 1–21, 2023, doi: 10.1186/s44147-023-00246-0.

[11] S. Poudel and P. Poudyal, "Classification of Waste Materials using CNN Based on Transfer Learning," *ACM Int. Conf. Proceeding Ser.*, vol. 14, no. 22, pp. 29–33, 2022, doi: 10.1145/3574318.3574345.

[12] T. H. Yan, S. N. M. Azam, Z. M. Sani, and A. Azizan, "Accuracy study of image classification for reverse vending machine waste segregation using convolutional neural network," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 1, pp. 366–374, 2024, doi: 10.11591/ijece.v14i1.pp366-374.

[13] H. Zheng and Y. Gu, "Encnn-upmws: Waste classification by a CNN ensemble using the UPM weighting strategy," *Electron.*, vol. 10, no. 4, pp. 1–21, 2021, doi: 10.3390/electronics10040427.

[14] D. Hogan Itam, E. Chimeme Martin, and I. Taiwo Horsfall, "Enhanced convolutional neural network methodology for solid waste classification utilizing data augmentation techniques," *Waste Manag. Bull.*, vol. 2, no. 4, pp. 184–193, 2024, doi: 10.1016/j.wmb.2024.11.002.

[15] M. Nahiduzzaman et al., "An automated waste classification system using deep learning techniques: Toward efficient waste recycling and environmental sustainability," *Knowledge-Based Syst.*, vol. 310, no. 1, pp. 1–29, 2025, doi: 10.1016/j.knosys.2025.113028.

[16] R. Chauhan, S. Shighra, H. Madkhali, L. Nguyen, and M. Prasad, "Efficient Future Waste Management: A Learning-Based Approach with Deep Neural Networks for Smart System (LADS)," *Appl. Sci.*, vol. 13, no. 7, pp. 1–16, 2023, doi: 10.3390/app13074140.

[17] Z. Kang, J. Yang, G. Li, and Z. Zhang, "An Automatic Garbage Classification System Based on Deep Learning," *IEEE Access*, vol. 8, no. 4, pp. 140019–140029, 2020, doi: 10.1109/ACCESS.2020.3010496.

[18] M. Kaya, S. Ulutürk, Y. Ç. Kaya, and O. Altıntaş, "Optimization of Several Deep CNN Models," S*ak. Univ. J. Comput. Inf. Sci.*, vol. 6, no. 2, pp. 91–104, 2023.

[19] 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.

[20] M. S. Islam, M. Shaheenur Islam Sumon, Molla E. Majid, Saad Bin Abul Kashem, Mohammad Nashbat, Azad Ashrafe, Amith Khandakar, Ali K. Ansaruddin Kunju, Mazhar Hasan-Zia, Muhammad E.H. Chowdhury., "ECCDN-Net: A deep learning-based technique for efficient organic and recyclable waste classification," *Waste Manag.*, vol. 193, no. November 2024, pp. 363–375, 2025, doi: 10.1016/j.wasman.2024.12.023.

[21] M. M. Abo-Zahhad and M. Abo-Zahhad, "Real time intelligent garbage monitoring and efficient collection using Yolov8 and Yolov5 deep learning models for environmental sustainability," *Sci. Rep*., vol. 15, no. 1, pp. 1–26, 2025, doi: 10.1038/s41598-025-99885-x.

[22] R. Wu, X. Liu, T. Zhang, J. Xia, J. Li, M. Zhu and G. Gu, "An Efficient Multi-Label Classification-Based Municipal Waste Image Identification," *Processes*, vol. 12, no. 6, pp. 1–14, 2024, doi: 10.3390/pr12061075.

[23] Q. Zhang, Q.Yang, X. Zhang, W. Wei, Q. Bao, J.Su, and X. Liu,., "A multi-label waste detection model based on transfer learning," *Resour. Conserv. Recycl.*, vol. 181, no. 12–2021, pp. 1-15, 2022, doi: 10.1016/j.resconrec.2022.106235.

[24] G. Celik, "Multi-layer feature fusion for high-accuracy solid waste classification using a hybrid deep learning model," *Vis. Comput*., vol. 1, no. 2, pp. 1–23, 2025, doi: 10.1007/s00371-025-04031-3.

[25] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*., vol. 2022, no. June, pp. 11966–11976, 2022, doi: 10.1109/CVPR52688.2022.01167.

[26] B. Madhavi, M. Mahanty, C. Lin, B. O. L. Jagan, H. M. Rai, S. Agarwal and N. Agarwal, "SwinConvNeXt: a fused deep learning architecture for Real-time garbage image classification," *Sci. Rep*., vol. 15, no. 1, pp. 1–16, 2025, doi: 10.1038/s41598-025-91302-7.

[27] D. Ghosh and A. Goswami, "Enhanced deep learning framework for efficient garbage classification in smart waste management systems," *Inf. Sci. (Ny)*., vol. 719, no. 4, pp. 1-12, 2025, doi: 10.1016/j.ins.2025.122462.

[28] W. Qiu, C. Xie, and J. Huang, "An Improved EfficientNetV2 for Garbage Classification," *Lect. Notes Comput. Sci.*, vol. 15864 LNAI, no. 4, pp. 79–90, 2025, doi: 10.1007/978-981-96-9901-8_7.

[29] S. Zhang and G. Ren, "RoSwish: A novel Rotating Swish activation function with adaptive rotation around zero," *Neural Networks*, vol. 192, no. 4–2024, pp. 1-22, 2025, doi: 10.1016/j.neunet.2025.107892.

[30] M. Rezaei, F. Soleymani, B. Bischl, and S. Azizi, "Deep Bregman divergence for self-supervised representations learning," *Comput. Vis. Image Underst*., vol. 235, no. 4–2022, pp. 1-21, 2023, doi: 10.1016/j.cviu.2023.103801.

[31] H. Li, G. Zhang, Q. Su, and L. Ge, "Global representation fine-tuning for federated self-supervised representation learning," *Int. J. Intell. Networks*, vol. 6, no. 3, pp. 224–232, 2025, doi: 10.1016/j.ijin.2025.09.002.

[32] A. Ghanbarzadeh and H. Soleimani, "Self-supervised in-domain representation learning for remote sensing image scene classification," *Heliyon*, vol. 10, no. 19, pp. 1-12, 2024, doi: 10.1016/j.heliyon.2024.e37962.

[33] Q. Guo, C. Wang, D. Xiao, and Q. Huang, "A novel multi-label pest image classifier using the modified Swin Transformer and soft binary cross entropy loss," *Eng. Appl. Artif. Intell*., vol. 126, no. 4, pp. 1-12, 2023, doi: 10.1016/j.engappai.2023.107060.

[34] G. Polat, Ü. M. Çağlar, and A. Temizel, "Class distance weighted cross entropy loss for classification of disease severity," *Expert Syst. Appl*., vol. 269, no. 12–2024, pp. 1–10, 2025, doi: 10.1016/j.eswa.2024.126372.

[35] C. Qin, Y. Cao, and L. Meng, "L1-Smooth SVM with Distributed Adaptive Proximal Stochastic Gradient Descent with Momentum for Fast Brain Tumor Detection," *Comput. Mater. Contin*., vol. 79, no. 2, pp. 1975–1994, 2024, doi: 10.32604/cmc.2024.049228.

[36] J. Zhu et al., "Feature-targeted deep learning framework for pulmonary tumorous Cone-beam CT (CBCT) enhancement with multi-task customized perceptual loss and feature-guided CycleGAN," *Comput. Med. Imaging Graph*., vol. 121, no. 12–2024, pp. 1-12, 2025, doi: 10.1016/j.compmedimag.2024.102487.

[37] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, A survey on imbalanced learning: latest research, applications and future directions, vol. 57, no. 6, pp. 1-12, 2024. doi: 10.1007/s10462-024-10759-6.

[38] B. Kovács, F. Tinya, C. Németh, and P. Ódor, "Unfolding the effects of different forestry treatments on microclimate in oak forests: results of a 4-yr experiment," *Ecol. Appl*., vol. 30, no. 2, pp. 321–357, 2020, doi: 10.1002/eap.2043.

[39] N. Kang, K. Lee, and J. Im, "Calibrated mixup for imbalanced regression on tabular data," *Pattern Recognit*., vol. 172, no. 4, pp. 1-12, 2026, doi: 10.1016/j.patcog.2025.112610.

[40] W.-C. Cheng, T.-H. Mai, and H.-T. Lin, "From SMOTE to Mixup for Deep Imbalanced Classification," *in Technologies and Applications of Artificial Intelligence*, C.-Y. Lee, C.-L. Lin, and H.-T. Chang, Eds., Singapore: Springer Nature Singapore, 2024, pp. 75–96. doi: https://doi.org/10.1007/978-981-97-1711-8_6.

[41] R. Wang, L. Jiao, C. Xie, P. Chen, J. Du, and R. Li, "S-RPN: Sampling-balanced region proposal network for small crop pest detection," *Comput. Electron. Agric*., vol. 187, no. 6, pp. 1-12, 2021, doi: 10.1016/j.compag.2021.106290.

[42] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis, "Efficient High-Resolution Deep Learning: A Survey," *ACM Comput. Surv*., vol. 56, no. 7, pp. 1–19, 2024, doi: 10.1145/3645107.

[43] A. M. L. De Alwis, M. Bazli, and M. Arashpour, "Automated recognition of contaminated construction and demolition wood waste using deep learning," *Resour. Conserv. Recycl*., vol. 219, no. 12–2024, pp. 1-12, 2025, doi: 10.1016/j.resconrec.2025.108278.

[44] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019, doi: 10.1186/s40537-019-0197-0.

[45] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *J. Big Data*, vol. 5, no. 1, pp. 1–30, 2018, doi: 10.1186/s40537-018-0151-6.

[46] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," *Pattern Recognit*., vol. 137, no. 1, pp. 1-12, 2023, doi: 10.1016/j.patcog.2023.109347.

[47] Y. X. Zhao and Y. Lee, "Bird Species Classification Using Image Background Removal for Data Augmentation," *Comput. Mater. Contin*., vol. 84, no. 1, pp. 791–810, 2025, doi: 10.32604/cmc.2025.065048.

[48] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc*., vol. 1, no. 2, pp. 1–14, 2015, doi: https://doi.org/10.48550/arXiv.1409.1556.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.,* vol. 2016, no. Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[50] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 1, no. 2, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[51] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*., vol. 2018, no. 1, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[52] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 3, no. 1, pp. 10691–10700, 2019, doi: https://doi.org/10.48550/arXiv.1905.1194.