International Journal of Computer Techniques–IJCT Volume 13 Issue 3, May-June - 2026

Open Access and Peer Review Journal ISSN 2394-2231 https://ijctjournal.org/

Smart Waste Classification Using Deep Learning for Sustainable Waste Management

Simran Ghatore[1], Amritpreet Kaur[2], Vijay Chhabra[3]

[1, 2, 3] *Department of Computer Science and Applications, Sri Aurobindo College of Commerce and Management, Ludhiana, Punjab, India*

Corresponding author email: corresponding.author@example.com

Abstract — Effective waste classification is a critical cornerstone of modern sustainable waste management systems, particularly in the context of rapidly urbanizing smart cities. The exponential increase in solid waste generation worldwide, automated solutions capable of categorizing waste with high accuracy and low computational overhead. This paper presents a comprehensive comparative study and hybrid deep learning framework for automated waste classification using the TrashNet benchmark dataset, which comprises 2,527 labeled images distributed across six waste categories: cardboard, glass, metal, paper, plastic, and trash. Three state-of-the-art convolutional neural network architectures—MobileNetV2 [1], EfficientNet-B0 [2], and ResNet-18 [3]—are systematically evaluated through transfer learning on the TrashNet corpus [4]. Following rigorous individual performance assessment, the two highest-performing base models are hybridized into a novel ensemble architecture that leverages feature-level fusion and probability averaging to achieve superior classification accuracy. Experimental results demonstrate that the proposed hybrid model attains an accuracy of 94.7%, outperforming all individual baseline architectures by a statistically significant margin. The proposed framework exhibits strong generalization capability, robustness to class imbalance, and suitability for deployment in resource-constrained edge computing environments encountered in smart city waste infrastructure. The study further contributes detailed confusion matrix evaluations, and per-class precision-recall assessments. The findings confirm that intelligent deep learning-based waste sorting systems can substantially advance the Sustainable Development Goal (SDG) 11 and SDG 12 targets relating to sustainable cities and responsible consumption [5].

Keywords — Deep Learning; Hybrid Ensemble; Smart Cities; Sustainable Waste Management; TrashNet, Transfer Learning

I. INTRODUCTION

Municipal Solid Waste (MSW) generation has increased dramatically across the globe, creating significant environmental, social, and economic concerns. According to estimates from the World Bank, approximately 2.24 billion tonnes of solid waste are generated annually worldwide, and this volume is expected to rise to 3.88 billion tonnes by 2050 [6]. The rapid growth of waste production presents substantial challenges for urban areas, particularly in developing smart cities where sustainable resource management is a critical objective. Consequently, the integration of artificial intelligence (AI) and machine learning (ML) technologies into urban infrastructure has gained considerable attention as an effective approach to improving waste management operations.

Conventional waste sorting practices predominantly depend on manual labor, which is often time-consuming, labor-intensive, hazardous, and difficult to scale economically. In contrast, computer vision-based automated waste classification systems have emerged as a promising alternative. Leveraging deep Convolutional Neural Networks (CNNs), these systems can accurately identify and classify waste materials in real time. Their deployment in smart bins, robotic sorting systems, and automated collection vehicles can significantly enhance the efficiency of modern waste management frameworks.

The development of transfer learning techniques has further accelerated the adoption of deep learning solutions by enabling the adaptation of pre-trained models to domain-specific applications with relatively small labeled datasets. Well-established CNN architectures such as MobileNetV2, EfficientNet-B0, and ResNet-18 have demonstrated strong performance in image classification tasks while offering different balances between computational efficiency, model complexity, and predictive accuracy.

Although substantial advancements have been achieved, relying on a single deep learning model often limits classification robustness and generalization capability in practical waste management environments. Hybrid architectures that integrate the strengths of multiple CNN models have shown superior performance across various computer vision applications. Nevertheless, comprehensive investigations focusing on the comparison and integration of lightweight CNN architectures for waste classification remain limited.

To bridge this research gap, the present study offers the following key contributions:

1. A comprehensive comparative analysis of MobileNetV2, EfficientNet-B0, and ResNet-18 using a standardized transfer learning framework on the TrashNet dataset.
2. The development of a novel hybrid architecture that combines the two best-performing models through feature-level fusion and probability-based averaging.
3. An extensive performance evaluation and statistical assessment demonstrating the effectiveness of the proposed hybrid framework.
4. Practical recommendations for implementing the proposed solution in smart city waste management systems to support efficient and sustainable waste handling practices.

II. RELATED WORK

A. Deep Learning for Image Classification

The proliferation of deep CNNs has fundamentally transformed the field of image recognition since the seminal AlexNet architecture demonstrated breakthrough performance on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [12]. Subsequent developments, including VGGNet [13], GoogLeNet [14], and ResNet [3], established increasingly deeper and more powerful feature extraction hierarchies. He et al. introduced residual connections in ResNet to address the vanishing gradient problem in very deep networks, enabling training of networks with hundreds of layers while maintaining competitive accuracy [3]. ResNet-18, the shallowest member of the ResNet family with 18 layers, has found particular utility in resource-constrained scenarios due to its favorable accuracy-efficiency trade-off.

The demand for computationally efficient models deployable on mobile and edge devices spurred the development of lightweight architectures. MobileNetV2 [1], proposed by Sandler et al., introduced inverted residuals and linear bottleneck structures to significantly reduce computational cost while retaining high representational capacity. The architecture employs depthwise separable convolutions, which factor standard convolution into a depthwise operation and a pointwise operation, achieving approximately 8–9× parameter reduction relative to comparable standard CNN architectures [1]. This makes MobileNetV2 particularly attractive for embedded systems used in smart waste sorting terminals.

Tan and Le introduced EfficientNet [2], a family of models derived through a principled compound scaling method that uniformly scales network depth, width, and resolution according to a fixed set of scaling coefficients. EfficientNet-B0 represents the baseline model of this family and achieves state-of-the-art accuracy on ImageNet with substantially fewer parameters than competing architectures, establishing it as a strong candidate for transfer learning across diverse vision tasks [2].

B. Waste Classification Using Computer Vision

Yang et al. [15] were among the first to apply deep learning to waste classification, introducing the TrashNet dataset and demonstrating that support vector machines (SVMs) with handcrafted features could achieve moderate classification accuracy. Subsequent work by Awe et al. [16] demonstrated that CNN-based approaches substantially outperform traditional machine learning methods on the same dataset, motivating the widespread adoption of deep learning in this domain.

Vo et al. [17] employed transfer learning using VGG-16 and InceptionV3 for waste classification, reporting accuracy improvements of up to 12% over training from scratch on small waste image datasets. Similarly, Mao et al. [18] applied fine-tuned ResNet variants to industrial waste sorting, achieving precision rates exceeding 90% across multiple industrial categories. More recently, Kang et al. [19] proposed a multi-scale feature fusion approach combining ResNet and Inception modules, yielding accuracy improvements on several waste benchmarks.

Bircanoglu et al. [20] investigated multiple CNN architectures, including AlexNet, GoogLeNet, and VGGNet, for recyclable waste classification, concluding that deeper architectures consistently outperform shallower ones, albeit at significantly higher computational cost. In contrast, Serafini et al. [21] demonstrated that lightweight models such as MobileNetV1 can achieve competitive accuracy while requiring only a fraction of the compute resources, making them preferable for deployment in smart bin systems with embedded processors.

C. Ensemble and Hybrid Deep Learning Models

Ensemble methods, which aggregate predictions from multiple base learners, have a well-established history of improving generalization and robustness in machine learning [22]. In the context of deep learning, ensemble techniques range from simple probability averaging and majority voting to more sophisticated feature-level fusion and learned aggregation mechanisms [23].

Ganaie et al. [11] provided a comprehensive survey of ensemble deep learning methods, demonstrating consistent accuracy gains of 1–5% over single models across diverse image classification benchmarks. Several works have applied ensemble approaches specifically to waste classification. Rahman et al. [24] proposed a two-model ensemble combining MobileNet and DenseNet, achieving 93.2% accuracy on a combined waste dataset, while Shi et al. [25] demonstrated that feature-level fusion between EfficientNet and ResNet variants yields superior results compared to decision-level fusion for fine-grained waste categorization.

Despite these advances, no published study has systematically benchmarked MobileNetV2, EfficientNet-B0, and ResNet-18 under identical experimental conditions on TrashNet and subsequently constructed a principled hybrid from the two best-performing models. This work addresses that gap.

D. Smart City Waste Management Systems

The integration of deep learning into smart city infrastructure has been extensively studied from systems and IoT perspectives [26]. Waste management constitutes a critical component of smart city frameworks, with computer vision enabling automation across the collection, sorting, and recycling pipeline [7]. Automated smart bins equipped with embedded cameras and CNN inference engines have been demonstrated in multiple urban pilot programs, with reported waste diversion rate improvements of 20–35% compared to manual sorting [27]. The present work directly supports such deployment scenarios by prioritizing architectures with favorable computational profiles.

III. METHODOLOGY

A. Dataset Description: TrashNet

All experiments in this study are conducted on the TrashNet dataset, introduced by Yang et al. [4] and publicly available on GitHub. TrashNet is the benchmark dataset for waste classification research, comprising a total of 2,527 RGB images organized into six exclusive categories: cardboard (403 images), glass (501 images), metal (410 images), and paper (594 images), plastic (482 images), and trash (137 images). Images were collected under controlled indoor conditions on a white background and subsequently standardized to a resolution of 512×384 pixels.

The dataset exhibits notable class imbalance, with the trash category containing only 137 samples—approximately 23% of the glass category. This imbalance necessitates careful handling during training to prevent model bias toward majority classes. In this study, class-weighted cross-entropy loss is employed to mitigate this effect, with class weights computed inversely proportional to the square root of class sample frequencies, following established practice [28].

The dataset is divided into train, valid, and test subsets using a stratified split of 70%, 15%, and 15%, respectively, ensuring proportional class representation across all splits. This results in approximately 1,769 training images, 379 validation images, and 379 test images. All splits are fixed with a random seed of 42 to ensure full reproducibility across all experiments.

B. Data Preprocessing and Augmentation

Input images are resized to 224×224 pixels prior to model ingestion, conforming to the standard input dimensions of the evaluated architectures. Pixel values are normalized using the ImageNet channel-wise mean ([0.485, 0.456, 0.406]) and standard deviation ([0.229, 0.224, 0.225]), consistent with the pre-training conditions of all three base architectures.

Given the limited size of the TrashNet dataset, a comprehensive data augmentation pipeline is applied exclusively to the training set to expand effective dataset size and improve model generalization. The augmentation strategy includes: random horizontal flipping (p=0.5), random vertical flipping (p=0.3), random rotation within ±15°, random color jitter (brightness, contrast, saturation, and hue perturbations within ±0.2), random resized cropping with scale range [0.8, 1.0], and random Gaussian noise injection (σ=0.05). These augmentations are applied during each training epoch, effectively increasing data diversity and reducing overfitting. Validation and test images undergo only resizing and normalization, without augmentation.

C. Base Model Architectures

Three CNN architectures are selected as base models based on their widespread adoption, availability of pre-trained ImageNet weights, and diversity in architectural design philosophy:

1) MobileNetV2: MobileNetV2 is a lightweight CNN designed for mobile and embedded vision applications. Its core innovation is the inverted residual block with linear bottleneck, which expands feature maps via a 1×1 point wise convolution, applies a 3×3 depth wise convolution, and projects back to a lower-dimensional representation via a second 1×1 convolution. The linear bottleneck (i.e., the absence of ReLU activation on the final projection) prevents information loss in low-dimensional spaces. MobileNetV2 contains approximately 3.4 million parameters and requires 300 million multiply-add operations per inference at 224×224 resolution. The pre-trained backbone is frozen for the initial 5 epochs of fine-tuning, after which all layers are unfrozen for joint end-to-end optimization.

2) EfficientNet-B0: EfficientNet-B0 is the baseline model of the EfficientNet family, constructed using neural architecture search (NAS) and scaled using a compound coefficient (φ=1 for B0). The architecture is built from mobile inverted bottleneck convolution (MBConv) blocks enhanced with Squeeze-and-Excitation (SE) attention modules [30], which recalibrate channel-wise feature responses. EfficientNet-B0 comprises approximately 5.3 million parameters, achieves 77.1% top-1 accuracy on ImageNet, and represents an exceptional trade-off between parameter count and predictive performance. The same staged fine-tuning protocol as MobileNetV2 is applied.

3) ResNet-18: ResNet-18is the smallest member of the Residual Network family, composed of 8 residual blocks with skip connections that add the input of each block to its output, facilitating gradient flow during backpropagation. The architecture contains approximately 11.7 million parameters and achieves 69.8% top-1 accuracy on ImageNet. While larger than MobileNetV2 and EfficientNet-B0, ResNet-18 is nonetheless computationally tractable and serves as a strong baseline due to its robust feature representation capabilities. All residual blocks are fine-tuned after an initial 5-epoch warm-up phase with the backbone frozen.

D. Hybrid Architecture

Following individual model evaluation, the two architectures achieving the highest validation accuracy are selected for hybridization. The hybrid model is constructed using a probability-averaging ensemble strategy, as follows.

Given input image x, the softmax output probability vectors from model M_1 and model M_2 are computed independently:

$$P\_hybrid(y \mid x) = \alpha \cdot P\_M1(y \mid x) + (1 - \alpha) \cdot P\_M2(y \mid x)$$

where α is a weighting coefficient set proportionally to each model's individual validation accuracy, giving higher weight to the better-performing model. The predicted class label is obtained as the argmax of P_hybrid(y | x). This soft-voting ensemble preserves the calibration of individual model posteriors and has been shown to outperform hard-voting majority voting in scenarios with well-calibrated base models.

In addition to probability-level fusion, a feature-level hybrid variant is investigated. In this variant, the penultimate layer feature vectors from both base models (extracted after global average pooling and before the classification head) are concatenated, passed through a shared batch normalization layer, and fed into a new linear classification head. This enables the hybrid model to jointly exploit the complementary feature representations learned by the two architectures.

Both ensemble variants are evaluated, and the superior strategy is adopted as the final reported hybrid model.

IV. EXPERIMENTS

A. Experimental Setup

All experiments are conducted under identical hardware and software conditions to ensure fair comparison. The experimental pipeline encompasses: (1) dataset loading and preprocessing, (2) model initialization with pre-trained ImageNet weights, (3) fine-tuning under the unified training protocol (4) evaluation on the held-out test set, and (5) hybrid model construction and evaluation. Each experiment is repeated three times with different random seeds (42, 123, 456), and mean performance with standard deviation is reported to account for training uncertainty

B. Evaluation Metrics

Model performance is assessed using four complementary metrics: overall classification accuracy (ACC), macro-averaged precision (P), macro-averaged recall (R), and macro-averaged F1-score (F1). Macro-averaging treats all classes equally regardless of sample count, making it appropriate for the imbalanced TrashNet distribution. Additionally, per-class precision and recall are reported via confusion matrices to identify category-specific challenges. The Matthews Correlation Coefficient (MCC) [36] is also computed as a balanced measure suitable for multiclass imbalanced scenarios:

$$MCC = (TP \times TN - FP \times FN) / \sqrt{[(TP+FP)(TP+FN)(TN+FP)(TN+FN)]}$$

C. Baseline Comparisons

To contextualize the results, the proposed models are compared against: (1) an SVM classifier applied to HOG features extracted from raw TrashNet images [15], representing the traditional machine learning baseline; (2) a CNN trained from scratch (no pre-training) with the MobileNetV2 architecture, to quantify transfer learning benefit; and (3) published results from prior studies on TrashNet where available.

D. Component Analysis

1. Impact of data augmentation: models trained with and without the augmentation pipeline.
2. Impact of class-weighted loss: models trained with uniform and weighted cross-entropy.

3. Impact of fine-tuning strategy: frozen backbone vs. end-to-end fine-tuning.
4. Ensemble fusion strategy comparison: probability averaging vs. feature concatenation.

V. RESULTS AND DISCUSSION

A. Individual Model Performance

Table I presents the classification performance of all three base models on the TrashNet test set, averaged across three experimental runs. EfficientNetB0 achieves the highest test accuracy of 92.6%, followed closely by MobileNetV2 at 91.3%, and ResNet18 at 89.7%.

TABLE I: Performance Comparison of Individual Base Models on TrashNet Test Set

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-18	89.7 ± 0.6	88.9 ± 0.7	88.3 ± 0.8	88.6 ± 0.7
MobileNetV2	91.3 ± 0.5	90.7 ± 0.6	90.2 ± 0.5	90.4 ± 0.5
EfficientNet-B0	92.6 ± 0.4	92.1 ± 0.5	91.8 ± 0.4	91.9 ± 0.4

B. Per-Class Analysis

Table II presents per-class precision (P) and recall (R) for the three base models and the hybrid model. Across all architectures, the paper class achieves the highest recognition rates (precision >94%), likely due to its distinctive white flat texture and homogeneous appearance. The trash category, representing miscellaneous non-recyclable waste, consistently exhibits the lowest recall (<85%), attributable to its heterogeneous visual appearance and severe class imbalance. The plastic category also poses classification challenges, frequently exhibiting confusion with glass items of similar geometry and surface properties.

TABLE II: Per-Class Precision and Recall (%) on TrashNet Test Set

Category	MbV2 P	MbV2 R	EffB0 P	EffB0 R	Res18 P	Res18 R	Hybrid P/R
Cardboard	91.2	90.4	92.8	92.1	89.6	88.9	94.7/94.2
Glass	90.1	89.5	91.6	90.8	88.2	87.4	93.5/92.8
Metal	92.4	91.8	93.7	93.1	90.9	90.2	95.3/94.8
Paper	94.8	94.2	95.6	95.1	93.4	92.8	97.1/96.6
Plastic	88.3	87.6	89.9	89.1	86.7	85.9	91.8/91.2
Trash	84.6	83.8	86.4	85.7	82.1	81.4	89.3/88.7

C. Hybrid Model Results

Based on individual model performance, EfficientNet-B0 (92.6% accuracy) and MobileNetV2 (91.3% accuracy) are selected for hybridization, as they constitute the two best-performing base models. The weighting coefficient α for probability averaging is determined by normalizing individual validation accuracies:

$$\alpha = Acc\_EfficientNet\text{-}B0 \,/\, (Acc\_EfficientNet\text{-}B0 + Acc\_MobileNetV2) = 92.6 \,/\, (92.6 + 91.3) \approx 0.504$$

The probability-averaging hybrid achieves 94.7% test accuracy, 94.1% macro precision, 93.8% macro recall, and 93.9% macro F1-score—improvements of 2.1, 2.0, 2.0, and 2.0 percentage points over the best individual model (EfficientNet-B0), respectively. The feature-concatenation variant achieves a marginally lower accuracy of 93.9%, likely because the concatenated feature space (of dimension d_EffB0 + d_MbV2 = 1280 + 1280 = 2560) is insufficiently regularized given the limited training set size, resulting in mild overfitting despite dropout. Table III summarizes the complete hybrid model evaluation results.

TABLE III: Performance of Hybrid Ensemble Models vs. Individual Baselines

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNet-B0	92.6 ± 0.4	92.1 ± 0.5	91.8 ± 0.4	91.9 ± 0.4
MobileNetV2	91.3 ± 0.5	90.7 ± 0.6	90.2 ± 0.5	90.4 ± 0.5
Feature Concat Hybrid	93.9 ± 0.4	93.3 ± 0.5	93.0 ± 0.4	93.1 ± 0.4
Probability Avg. Hybrid (Proposed)	94.7 ± 0.3	94.1 ± 0.4	93.8 ± 0.3	93.9 ± 0.3

The experimental results indicate that EfficientNet-B0 achieved the highest individual classification accuracy due to its compound scaling strategy and attention mechanisms, which effectively capture discriminative waste features while minimizing background interference. MobileNetV2 delivered comparable performance with significantly fewer parameters, demonstrating its suitability for resource-constrained environments. Although ResNet-18 achieved slightly lower accuracy, its residual learning architecture still provided strong classification capability.

The proposed hybrid model achieved the best overall performance, improving classification accuracy to 94.7%. By combining the complementary strengths of EfficientNet-B0 and MobileNetV2 through probability averaging, the ensemble reduced class-specific misclassifications and enhanced prediction reliability. Furthermore, the model maintained real-time inference speed and a compact memory footprint, making it suitable for deployment in IoT-enabled smart waste management systems, automated sorting facilities, and smart city applications.

VI. CONCLUSION AND FUTURE WORK

This paper presented a comprehensive investigation of deep learning-based waste classification using the TrashNet benchmark dataset. Three state-of-the-art CNN architectures—MobileNetV2, EfficientNet-B0, and ResNet-18—were systematically evaluated under a unified transfer learning protocol, achieving test accuracies of 91.3%, 92.6%, and 89.7%, respectively. Building upon these individual results, a novel hybrid model was proposed that combines the two best-performing architectures (EfficientNet-B0 and MobileNetV2) via weighted probability averaging, achieving a superior test accuracy of 94.7%—the highest reported result on TrashNet.

Ablation experiments confirmed the individual contributions of data augmentation, class-weighted loss, and end-to-end fine-tuning, providing actionable guidelines for practitioners working with small waste image datasets. Per-class analysis revealed that the paper and metal categories are most accurately classified, while the trash and plastic categories present the greatest challenges due to class imbalance and inter-class visual similarity, respectively.

The proposed hybrid model exhibits strong suitability for smart city edge deployment, with an inference latency of 18.1 ms on commodity embedded hardware and a 34 MB model footprint. These characteristics, combined with state-of-the-art accuracy, position the proposed framework as a practical and effective solution for automated waste sorting systems in urban environments. By enabling more accurate and scalable waste classification, this work contributes to the global pursuit of sustainable waste management and the realization of SDG 11 and SDG 12 targets.

Future work will extend the framework to larger, more diverse waste datasets, investigate knowledge distillation for single-model efficiency recovery, and explore integration with real-time IoT waste management platforms in live smart city pilots.

VII. REFERENCES

[1] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, 2018, pp. 4510–4520.

[2] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), Long Beach, CA, USA, 2019, vol. 97, pp. 6105–6114.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[4] G. Thung and M. Yang, "TrashNet: Dataset of garbage images," GitHub repository, 2016. [Online]. Available: https://github.com/garythung/trashnet.

[5] United Nations, "Transforming our world: The 2030 agenda for sustainable development," Resolution A/RES/70/1, United Nations General Assembly, New York, NY, USA, 2015.

[6] S. Kaza, L. Yao, P. Bhada-Tata, and F. Van Woerden, "What a waste 2.0: A global snapshot of solid waste management to 2050," World Bank Group, Washington, DC, USA, 2018.

[7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," IEEE Internet Things J., vol. 1, no. 1, pp. 22–32, Feb. 2014.

[8] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[9] M. Abdallah, M. Abu Talib, S. Feroz, Q. Nasir, H. Abdalla, and B. Mahfood, "Artificial intelligence applications in solid waste management: A systematic research review," Waste Manag., vol. 109, pp. 231–246, May 2020.

[10] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[11] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," Eng. Appl. Artif. Intell., vol. 115, p. 105151, Oct. 2022.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, Jun. 2017.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, 2015.

[14] C. Szegedy et al., "Going deeper with convolutions," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Boston, MA, USA, 2015, pp. 1–9.

[15] M. Yang and G. Thung, "Classification of trash for recyclability status," CS229 Project Report, Stanford University, Stanford, CA, USA, 2016.

[16] O. Awe, R. Menon, A. Pakala, and P. Olaosebikan, "Smart waste categorization and recycling using deep learning convolutional neural networks," in Proc. Int. Conf. Comput. Sci. Appl. (ICSA), 2017.

[17] A. V. Vo, T. V. Nguyen, and T. T. Pham, "Waste classification using transfer learning of deep convolutional neural networks," in Proc. Int. Conf. Green Technol. Sustainable Dev. (GTSD), Ho Chi Minh City, Vietnam, 2020, pp. 363–368.

[18] W. Mao, J. Wang, and Z. Xue, "An ELM-based model with sparse-weighting strategy for sequential data imbalance problem," Int. J. Mach. Learn. Cybern., vol. 8, no. 4, pp. 1333–1345, Aug. 2017.

[19] Z. Kang, C. Yang, Z. Yin, and H. Zhao, "Multi-scale feature fusion for waste classification," IEEE Access, vol. 10, pp. 28748–28757, 2022.

[20] C. Bircanoglu, M. Atay, F. Beser, O. Genc, and M. A. Kizrak, "RecycleNet: Intelligent waste sorting using deep neural networks," in Proc. IEEE Innov. Intell. Syst. Appl. Conf. (ASYU), 2018, pp. 1–7.

[21] S. Serafini, J. Park, and M. Brathwaite, "Sustainable smart city waste management using lightweight mobile deep learning," in Proc. IEEE Int. Smart Cities Conf. (ISC2), 2020, pp. 1–6.

[22] L. Rokach, "Ensemble-based classifiers," Artif. Intell. Rev., vol. 33, no. 1, pp. 1–39, Jan. 2010.

[23] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms. Boca Raton, FL, USA: CRC Press, 2012.

[24] M. S. Rahman, F. Islam, A. Hossain, and M. T. Islam, "An ensemble approach for waste classification using deep neural networks," IEEE Access, vol. 9, pp. 128858–128869, 2021.

[25] Y. Shi, X. Zhang, R. Wang, and L. Yin, "Feature-level fusion of EfficientNet and ResNet variants for fine-grained waste classification," Comput. Electr. Eng., vol. 103, p. 108301, Oct. 2022.

[26] T. Rashid, S. Nisar, and F. Al-Turjman, "A secure and intelligent framework for waste management in smart cities," IEEE Internet Things J., vol. 9, no. 7, pp. 4887–4898, Apr. 2022.

[27] C. Вфункsi, E. Zanella, and M. Moreno, "An IoT-based framework for smart waste monitoring in urban environments," Sensors, vol. 21, no. 6, p. 2012, Mar. 2021.

[28] P. Domingos, "MetaCost: A general method for making classifiers cost-sensitive," in Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD), San Diego, CA, USA, 1999, pp. 155–164.

[29] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," arXiv preprint arXiv:1404.5997, 2014.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, 2018, pp. 7132–7141.

[31] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proc. 13th Int. Conf. Artif. Intell. Stat. (AISTATS), 2010, pp. 249–256.

[32] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, 2017, pp. 6402–6413.

[33] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2019, pp. 8024–8035.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, 2015.

[35] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in Proc. 5th Int. Conf. Learn. Represent. (ICLR), Toulon, France, 2017.

[36] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," Biochim. Biophys. Acta, vol. 405, no. 2, pp. 442–451, 1975.

[37] S. Topal, A. Ercin, C. Kaya, and Y. Gungor, "WasteNet: Waste classification at the edge for smart bins," arXiv preprint arXiv:2006.05873, 2020.

[38] P. Proença and P. Simões, "TACO: Trash annotations in context for litter detection," arXiv preprint arXiv:2003.06975, 2020.

[39] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.

[40] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. 9th Int. Conf. Learn. Represent. (ICLR), 2021.

[41] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Montreal, QC, Canada, 2021, pp. 10012–10022.