

IMAGE-BASED DETECTION AND DEEP LEARNING METHODS FOR PLASTIC AND ENVIRONMENTAL MONITORING

Terézia Kurimská, Terézia Kurucová, Šimon Horvát, Ľubomír Antoni, Gabriel Semanišin
Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University in Košice,
Jesenná 5, 040 01 Košice, Slovakia
terezia.kurimska@student.upjs.sk, terezia.kurucova@student.upjs.sk

Abstract

Plastic pollution is a major environmental concern, particularly in aquatic and coastal ecosystems. Beyond cleanup efforts, effective prevention requires accurate detection and classification of waste in natural settings. Inspired by prior machine learning tools for waste monitoring (e.g., PlasticNet or DroneDeploy), this study focuses on the publicly available TrashNet dataset. The original six waste categories were re-labeled into two groups according to ecological impact: harmful (plastic, mixed trash) and less harmful (paper, cardboard, glass, metal). Two deep learning models were developed and evaluated: a custom four-block convolutional neural network (CNN) trained from scratch with focal loss, and a transfer learning model based on VGG16 pretrained on ImageNet. Results show that while the CNN achieved reasonable recall for harmful waste, the VGG16 model provided a more balanced trade-off, reaching accuracies above 82% on validation and test data. These findings confirm the potential of deep learning for automated monitoring of plastic pollution, with applications in aerial surveillance, smart waste management, and autonomous cleanup systems.

Keywords: Plastic pollution, waste classification, deep learning, transfer learning, CNN, TrashNet

Preučevanje vplivov plastike na okolje z uporabo zaznavanja na podlagi slik in metod globokega učenja

Izvleček

Plastično onesnaženje je velik okoljski problem, zlasti v vodnih in obalnih ekosistemih. Poleg prizadevanj za čiščenje učinkovito preprečevanje zahteva natančno zaznavanje in razvrščanje odpadkov v naravnem okolju. Navdihnjena s prejšnjimi orodji strojnega učenja za spremljanje odpadkov (npr. PlasticNet ali DroneDeploy) se ta študija osredotoča na javno dostopen nabor podatkov TrashNet. Prvotnih šest kategorij odpadkov je bilo ponovno označenih v dve skupini glede na ekološki vpliv: škodljivi (plastika, mešani odpadki) in manj škodljivi (papir, karton, steklo, kovina). Razvita in ovrednotena sta bila dva modela globokega učenja: prilagojena štiriblokovna konvolucijska nevronska mreža (CNN), učena od začetka z uporabo fokalne funkcije izgube, in model prenosa učenja, ki temelji na VGG16, predhodno naučenem na ImageNetu. Rezultati kažejo, da je CNN dosegel zadovoljiv priklic za škodljive odpadke, medtem ko je model VGG16 zagotovil bolj uravnotežen kompromis ter dosegel natančnost nad 82% na validacijskih in testnih podatkih. Ugotovitve potrjujejo potencial globokega učenja za avtomatizirano spremljanje plastičnega onesnaženja, z uporabo v zračnem nadzoru, pametnem upravljanju odpadkov in avtonomnih sistemih čiščenja.

Ključne besede: plastično onesnaženje, razvrščanje odpadkov, globoko učenje, prenos učenja, CNN (konvolucijska nevronska mreža), TrashNet

1 INTRODUCTION

Plastic pollution has emerged as one of the most pressing environmental challenges of the twenty-first century. The persistence of plastics in aquatic and terrestrial ecosystems contributes to long-term ecological degradation, threatening marine life, disrupting food chains, and impacting human health. According to recent estimates, millions of tons of plastic debris enter the oceans each year, where they fragment into microplastics and accumulate in coastal environments. While cleanup initiatives and regulatory interventions play an important role, the scale of pollution requires complementary technological solutions for continuous monitoring and classification of waste [1], [2], [3].

Recent advances in artificial intelligence, and in particular deep learning, provide new opportunities to automate environmental monitoring tasks. Convolutional neural networks (CNNs) have shown state-of-the-art performance in visual recognition problems [4], and their application to waste detection could accelerate large-scale monitoring. Prior work has demonstrated CNN-based systems for automated classification of municipal waste images [5], as well as robust detection of plastic debris in natural and urban settings [6]. The integration of such methods into aerial surveillance platforms, including drones, offers the possibility of efficient, non-intrusive tracking of plastic waste across large geographic areas.

The main goal of this study is to explore the application of image-based deep learning models for the detection and classification of plastic waste. Specifically, we design and evaluate two approaches: a custom four-layer CNN trained from scratch on the TrashNet dataset, and a transfer learning strategy employing a pre-trained VGG16 model [7]. By comparing a lightweight architecture with a large-scale pre-trained network, we aim to provide insights into the trade-offs between model complexity, training data requirements, and classification performance. The experiments serve as a baseline for further applications in environmental monitoring systems and drone-based waste detection platforms.

The overall contribution of this work is threefold. First, we propose a binary harmful-versus-less-harmful waste classification framework oriented toward practical environmental monitoring applications. Second, we provide a comparative evaluation between a lightweight custom CNN architecture and a substantially larger pretrained VGG16 model, highlighting the trade-offs between predictive performance, computational complexity, and deployment feasibility. Third, we discuss the implications of deep learning model selection for UAV-assisted and edge-based waste monitoring systems, where inference efficiency and hardware constraints are critical practical factors.

The rest of the paper is organized as follows. Section 2 reviews related work on automated waste classification and image-based plastic detection. Section 3 introduces the dataset, data augmentation techniques, and the deep learning models used in this study. Section 4 presents experimental results and discusses model performance. Section 5 and Section 6 conclude the paper and outlines directions for future research.

2 RELATED WORK

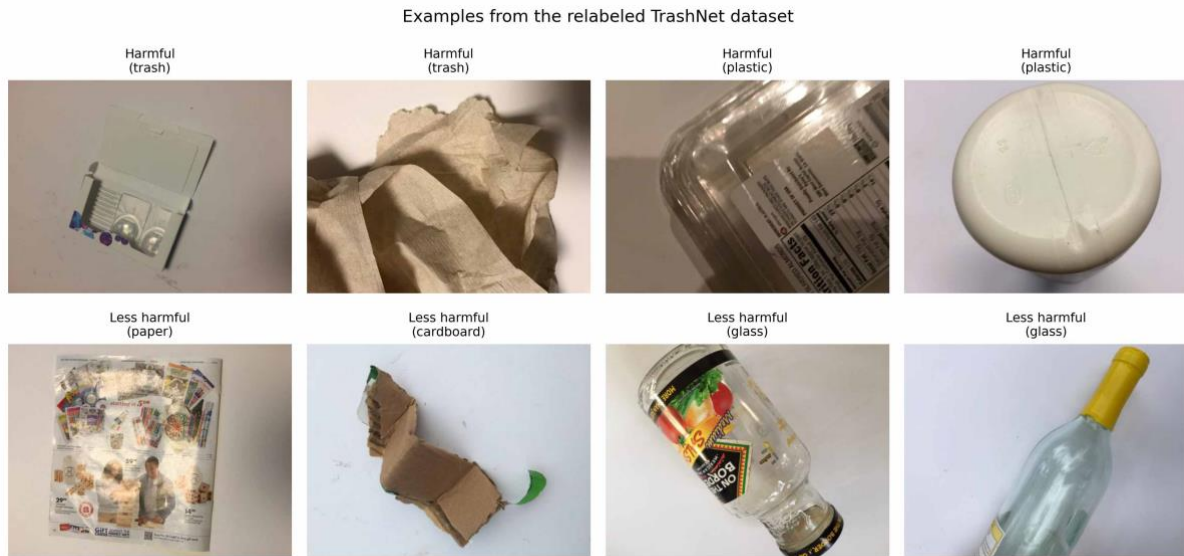


Figure 1: Example images from the relabeled TrashNet dataset. Top row: harmful waste samples (plastic, mixed trash). Bottom row: less harmful waste samples (paper, cardboard, glass, metal). The original six-class categories are indicated in parentheses.

Public, labeled image datasets have been crucial for waste classification research. The *TrashNet* dataset introduced by Thung and Yang has become a de facto benchmark for six common categories (glass, paper, cardboard, plastic, metal, trash) and has been widely reused in academic studies and student projects [8], [9]. Building on such datasets, early deep learning approaches showed that convolutional neural networks (CNNs) substantially outperform traditional feature engineering pipelines.

Several works compared CNN backbones on TrashNet or closely related datasets. [10] reported results across multiple deep learning models at IEEE Big Data 2018, while [11] conducted a systematic comparison of VGG, Inception, and ResNet variants, highlighting the benefit of modern architectures and careful training procedures. Beyond architecture choice, transfer learning from ImageNet typically improves performance when labeled data are limited: [12] combined a ResNet-50 feature extractor with a shallow classifier and achieved strong baselines on TrashNet.

Beyond TrashNet, more domain-specific systems have emerged. *PlasticNet*, a CNN-based spectral classifier developed by Zhu et al. (2023), achieved over 95% accuracy on micro-FT-IR data for detecting microplastics in environmental samples [13]. Similarly, platforms like *DroneDeploy* facilitate aerial imagery collection and real-time environmental monitoring pipelines [14].

Moving from tabletop images toward field deployments, recent studies explore drone (UAV) imagery and remote sensing for monitoring plastic and marine litter. Methodological and operational guidance for UAV-based surveys emphasizes ground sampling distance, flight altitude, and camera setup, which directly affect detectability and classifiability of items [15], [16]. Empirical evaluations demonstrate feasibility and constraints across floating, submerged, and seafloor debris [17]. Application-specific systems have also combined UAV sensors with deep learning to detect river plastics and classify materials in situ [18], [19].

In summary, prior art establishes (i) TrashNet as a common benchmark, (ii) the advantage of CNNs and transfer learning for waste classification with limited data, (iii) the role of PlasticNet and DroneDeploy in inspiring practical applications, and (iv) increasing interest in drone-based monitoring under real-world conditions. However, few papers juxtapose a lightweight from-scratch CNN with a strong transfer baseline (e.g., VGG16) under identical augmentation and training protocols on a single, reproducible dataset. Our study addresses this gap by comparing a compact 4-layer CNN against VGG16 (frozen and fine-tuned) on TrashNet, aiming to clarify accuracy and complexity trade-offs relevant to embedded and aerial applications.

3 METHODOLOGY

Figure 2 summarizes the complete experimental workflow. The original six-class TrashNet dataset was first preprocessed and converted into a binary harmful-versus-less-harmful classification task. The resulting dataset was split into training, validation, and test subsets, with augmentation applied only during training. Two deep learning strategies were then evaluated: a lightweight custom CNN trained from scratch using focal loss, and a VGG16 transfer learning model with ImageNet initialization and fine-tuning. Finally, both models were compared using harmful-class-oriented evaluation metrics and interpreted with respect to practical deployment constraints.

Figure 2: Overall workflow of the proposed harmful waste classification framework.

3.1 Dataset

We used the publicly available *TrashNet* dataset [8], [9], which contains images of six common waste categories: cardboard, glass, metal, paper, plastic, and mixed trash. All images (2527 samples) were resized to 224×224 pixels and converted to RGB format for consistency. Since no predefined validation and test splits were provided, we randomly divided the dataset into 70% training (1773 images), 15% validation (377 images), and 15% testing (377 images, while maintaining class balance (see Figure 1).

The dataset was divided using a stratified 70/15/15 split in order to preserve the original class distribution across the training, validation, and testing subsets. A fixed split strategy was selected to ensure reproducibility and consistent comparison between the evaluated architectures. Given the relatively limited dataset size, the obtained results should be interpreted as indicative rather than fully exhaustive estimates of model generalization performance.

The TrashNet dataset consists of waste objects photographed under controlled laboratory-like conditions, typically placed on a white background. While this setup is suitable for benchmarking image classification models, it does not fully reflect the variability and visual complexity of real-world UAV-based environmental monitoring scenarios. Therefore, the obtained results should be interpreted primarily as a proof-of-concept for lightweight waste classification rather than as a direct validation of deployment-ready aerial monitoring in unconstrained environments.

3.2 Binary Re-labeling Based on Environmental Impact

To transform the dataset into a binary classification task, categories were grouped according to their environmental impact. *Plastic* and *mixed trash* (24.4% of samples) were labeled as *harmful*, due to their persistence, ingestion risk, and toxicity in ecosystems [20], [21], [22]. In contrast, *paper*, *cardboard*, *glass*, and *metal* (75.6%) were assigned to the *less harmful* class, as they are either biodegradable in shorter timescales or inert in natural environments [1], [2],

[3], [23]. This distribution indicates a noticeable class imbalance rather than a fully balanced binary classification setting. Therefore, the imbalance was explicitly considered during model training and evaluation.

3.3 Data Augmentation

Since TrashNet is relatively small, data augmentation was applied to the training subset in order to improve generalization and reduce overfitting. The augmentation pipeline included random horizontal flips, small rotations (up to 8°), zoom transformations (up to 15%), contrast variation ($\pm 10\%$), and brightness adjustments. These transformations were applied online during training, i.e., augmented variants were generated dynamically within each epoch rather than stored as additional permanent images.

Importantly, augmentation was applied only to the training data, while the validation and test sets were kept unchanged. Therefore, the nominal dataset size remained 2527 original images, with 1773 images used for training, 377 for validation, and 377 for testing. Since the augmentations were generated online from the original training images, they did not change the underlying class distribution of the dataset. The binary class distribution therefore remained approximately 24.4% harmful and 75.6% less harmful samples.

3.4 Custom Convolutional Neural Network

The first model was a compact four-block convolutional neural network designed from scratch (Figure 3). Each of the first three blocks consisted of two convolutional layers with increasing filter counts (32, 64, 128), batch normalization, ReLU activation, max pooling, and dropout (0.15–0.25). A fourth block with 160 filters provided additional feature extraction before pooling. Global average pooling was applied to reduce dimensionality, followed by a dense layer of 256 neurons with ReLU activation and dropout (0.35). The final sigmoid output neuron was initialized with a bias term reflecting the empirical class prior to stabilize early training. To mitigate overfitting, all convolutional layers were regularized with L2 weight decay.

Training was performed using the Adam optimizer (learning rate 3×10^{-4} , gradient clipping at 1.0). Unlike standard cross-entropy, we applied a *focal loss* [24], with $\gamma = 2.0$ and $\alpha = 0.75$ to place higher weight on the *harmful* class. This helped address the noticeable class imbalance and placed additional emphasis on recall for the minority harmful class. Early stopping and adaptive learning rate reduction on validation loss were used.

Model: "simple_cnn_binary"

Layer (type)	Output Shape	Param #
input_layer_10 (InputLayer)	(None, 224, 224, 3)	0
conv2d_29 (Conv2D)	(None, 224, 224, 32)	864
batch_normalization_29 (BatchNormalization)	(None, 224, 224, 32)	128
re_lu_21 (ReLU)	(None, 224, 224, 32)	0
conv2d_30 (Conv2D)	(None, 224, 224, 32)	9,216
batch_normalization_30 (BatchNormalization)	(None, 224, 224, 32)	128
re_lu_22 (ReLU)	(None, 224, 224, 32)	0
max_pooling2d_20 (MaxPooling2D)	(None, 112, 112, 32)	0
dropout_21 (Dropout)	(None, 112, 112, 32)	0
conv2d_31 (Conv2D)	(None, 112, 112, 64)	18,432
batch_normalization_31 (BatchNormalization)	(None, 112, 112, 64)	256
re_lu_23 (ReLU)	(None, 112, 112, 64)	0
conv2d_32 (Conv2D)	(None, 112, 112, 64)	36,864
batch_normalization_32 (BatchNormalization)	(None, 112, 112, 64)	256
re_lu_24 (ReLU)	(None, 112, 112, 64)	0
max_pooling2d_21 (MaxPooling2D)	(None, 56, 56, 64)	0
dropout_22 (Dropout)	(None, 56, 56, 64)	0
conv2d_33 (Conv2D)	(None, 56, 56, 128)	73,728
batch_normalization_33 (BatchNormalization)	(None, 56, 56, 128)	512
re_lu_25 (ReLU)	(None, 56, 56, 128)	0
conv2d_34 (Conv2D)	(None, 56, 56, 128)	147,456
batch_normalization_34 (BatchNormalization)	(None, 56, 56, 128)	512
re_lu_26 (ReLU)	(None, 56, 56, 128)	0
max_pooling2d_22 (MaxPooling2D)	(None, 28, 28, 128)	0
dropout_23 (Dropout)	(None, 28, 28, 128)	0
conv2d_35 (Conv2D)	(None, 28, 28, 160)	184,320
batch_normalization_35 (BatchNormalization)	(None, 28, 28, 160)	640
re_lu_27 (ReLU)	(None, 28, 28, 160)	0
max_pooling2d_23 (MaxPooling2D)	(None, 14, 14, 160)	0
global_average_pooling2d_6 (GlobalAveragePooling2D)	(None, 160)	0
dropout_24 (Dropout)	(None, 160)	0
dense_12 (Dense)	(None, 256)	41,216
dropout_25 (Dropout)	(None, 256)	0
dense_13 (Dense)	(None, 1)	257

Total params: 1,541,925 (5.88 MB)
Trainable params: 513,569 (1.96 MB)
Non-trainable params: 1,216 (4.75 KB)
Optimizer params: 1,027,140 (3.92 MB)

Figure 3: Architecture of the custom CNN with four convolutional blocks and focal loss training for harmful waste detection.

3.5 Transfer Learning with VGG16

The second approach used transfer learning with VGG16 [7], pretrained on ImageNet. The convolutional base was initially frozen, and a custom classification head was appended: global average pooling, a dense ReLU layer (256 units), dropout (0.35), and a sigmoid output neuron (Figure 4). After training this configuration for 5 epochs with binary cross-entropy loss and

Adam optimizer (3×10^{-4}), the top convolutional layers of VGG16 were unfrozen for fine-tuning (10 epochs) with a lower learning rate (10^{-5}). Class weights were applied during VGG16 training to balance harmful vs. less harmful examples.

Model: "vgg16_binary"

Layer (type)	Output Shape	Param #
input_layer_12 (InputLayer)	(None, 224, 224, 3)	0
lambda_1 (Lambda)	(None, 224, 224, 3)	0
vgg16 (Functional)	(None, 7, 7, 512)	14,714,688
global_average_pooling2d_7 (GlobalAveragePooling2D)	(None, 512)	0
dropout_26 (Dropout)	(None, 512)	0
dense_14 (Dense)	(None, 256)	131,328
dropout_27 (Dropout)	(None, 256)	0
dense_15 (Dense)	(None, 1)	257

Total params: 41,067,845 (156.66 MB)
Trainable params: 13,110,785 (50.01 MB)
Non-trainable params: 1,735,488 (6.62 MB)
Optimizer params: 26,221,572 (100.03 MB)

Figure 4: Transfer learning setup with VGG16. The base network is pretrained on ImageNet; a custom head is added for binary environmental impact classification.

3.6 Evaluation Protocol

Both models were implemented in TensorFlow/Keras and trained on a v2-8 TPU-enabled Google Colab environment. Performance was evaluated on validation and test sets using accuracy, precision, recall, and F1-score. Importantly, evaluation was reported with the *harmful* class as the positive class, aligning metrics with the primary environmental concern.

4 RESULTS

4.1 Evaluation Metrics

Model performance was evaluated using accuracy, precision, recall, and F1-score. Since the primary goal of this study is to reliably identify environmentally harmful waste, all metrics were computed with the *harmful* class as the positive class.

Table 1 summarizes the differences in model complexity and deployment-related characteristics between the lightweight custom CNN and the VGG16 transfer learning architecture.

Table 1: Comparison of model complexity and deployment-related characteristics.

Model	Total Parameters	Inference Characteristics	Deployment Suitability
Custom CNN	~1.5M	Fast, lightweight	Edge/UAV deployment
VGG16 Transfer Learning	~41M	Computationally intensive	Server/cloud processing

4.2 Custom CNN Results

The custom four-block CNN trained with focal loss achieved moderate overall performance. On the validation set, the model reached an accuracy of 71.1%, with a harmful-class precision of 0.443, recall of 0.717, and F1-score of 0.548. On the test set, accuracy was 70.3%, with harmful precision of 0.430, recall of 0.663, and F1-score of 0.521. These results indicate that

while the CNN can detect harmful waste with reasonable recall, it struggles with precision, often misclassifying less harmful items as harmful.

Table 2: Performance of the custom CNN (positive = harmful).

Metric	Validation	Test
Accuracy	0.711	0.703
Precision	0.443	0.430
Recall	0.717	0.663
F1-score	0.548	0.521

4.3 VGG16 Transfer Learning Results

The transfer learning approach with VGG16 outperformed the custom CNN across both validation and test sets. After fine-tuning, VGG16 achieved an accuracy of 84.1% on the validation set, with harmful precision of 0.646, recall of 0.772, and F1-score of 0.703. On the test set, accuracy was 82.5%, with harmful precision of 0.630, recall of 0.685, and F1-score of 0.656. Compared to the custom CNN, the VGG16 model offered a more balanced trade-off between precision and recall, resulting in significantly higher F1-scores.

Table 3: Performance of VGG16 transfer learning model (positive = harmful).

Metric	Validation	Test
Accuracy	0.841	0.825
Precision	0.646	0.630
Recall	0.772	0.685
F1-score	0.703	0.656

4.4 Comparison of Approaches

The results demonstrate that transfer learning with VGG16 provides a substantial improvement over a custom CNN trained from scratch. While the custom CNN achieved recall above 0.66, its low precision highlights a tendency to over-predict harmful waste, which could be problematic for deployment scenarios with limited resources. In contrast, VGG16 achieved a stronger balance between detecting harmful items and avoiding false alarms. Overall, transfer learning is better suited for small and imbalanced datasets such as TrashNet, where pretrained feature extractors capture more robust representations than compact models trained from scratch.

5 DISCUSSION

The experimental results highlight the practical trade-offs between predictive performance and computational efficiency in waste classification systems intended for environmental monitoring applications. The compact CNN trained from scratch achieved moderate accuracy ($\approx 71\%$) and relatively high recall for the *harmful* class, but at the cost of low precision. This indicates a tendency to over-predict harmful waste, leading to false alarms. While such a conservative strategy might be acceptable in applications where detecting as much plastic as possible is critical (e.g., in coastal cleanup operations), it could reduce efficiency in scenarios where resources for manual verification are limited.

In contrast, the VGG16 transfer learning model demonstrated substantially higher accuracy ($\approx 83\%$) and more balanced precision and recall. By using pretrained ImageNet features, VGG16 generalized better to the relatively small TrashNet dataset, capturing both low-level textures and high-level object shapes. This balance reduces false positives without substantially compromising the detection of harmful items. These results are consistent with prior studies emphasizing the value of transfer learning when domain-specific data are limited [7], [24].

Another noteworthy observation is the effect of using focal loss for the CNN model. By assigning greater weight to the minority harmful class, the model maintained recall above 0.66 on the test set, confirming the utility of class-aware loss functions in imbalanced classification tasks. Nevertheless, even with focal loss, the CNN was unable to achieve the precision–recall balance of VGG16, underscoring the difficulty of training deep networks from scratch with relatively small datasets. Although data augmentation improved the variability of the training samples, it should not be interpreted as a replacement for additional real images of the minority harmful class. Since augmentation was applied online and did not alter the original class proportions, the class imbalance remains an important limitation of the experimental setup.

Another limitation of this study is that the evaluation was performed using a single stratified train-validation-test split rather than repeated cross-validation or multiple independent runs. Although this approach is common in deep learning experiments due to computational cost and training time considerations, performance estimates on relatively small datasets may still be sensitive to the specific partitioning of samples. Consequently, additional evaluation using repeated experiments or k-fold cross-validation could further improve the statistical robustness of the reported findings.

An important aspect of the presented comparison is the substantial difference in model complexity between the proposed custom CNN and the VGG16 architecture. While the custom CNN contains approximately 1.5 million trainable parameters, VGG16 contains roughly 41 million parameters. Consequently, VGG16 requires substantially higher memory consumption and computational resources during both training and inference. Despite its considerably smaller size, the proposed CNN achieved competitive performance, which is particularly important for practical deployment scenarios such as UAV-based environmental monitoring, where inference efficiency, energy consumption, and hardware limitations play a critical role. These results therefore suggest that lightweight architectures may provide a more favorable trade-off between accuracy and deployment feasibility than large-scale generic deep learning models. In real-time aerial monitoring applications, smaller models may also reduce inference latency and enable deployment on edge devices with limited GPU memory and computational power.

Although the proposed models achieved strong classification performance, an important limitation of this study is the nature of the TrashNet dataset, which contains images captured under controlled conditions with relatively clean backgrounds. In practical UAV-based waste monitoring, images are affected by varying illumination, occlusions, shadows, altitude changes, motion blur, and complex natural backgrounds. Consequently, further validation on real aerial imagery acquired in outdoor environments is necessary before reliable deployment in autonomous environmental monitoring systems can be claimed.

Future work will therefore focus on collecting and evaluating real-world UAV datasets containing waste objects in natural environmental conditions.

6 CONCLUSION

This study proposed a binary classification framework for environmental impact assessment of waste using the TrashNet dataset. The original six waste categories were re-labeled into *harmful* (plastic, mixed trash) and *less harmful* (paper, cardboard, glass, metal) based on ecological persistence and degradation profiles. Two deep learning approaches were evaluated: a custom four-block CNN with focal loss and a VGG16 transfer learning model.

The custom CNN achieved reasonable recall for harmful waste but suffered from low precision, whereas the VGG16 transfer learning model provided a more favorable balance, achieving accuracies above 82% and F1-scores of 0.65–0.70 for harmful waste. These results confirm

that pretrained architectures substantially improve performance in small and imbalanced datasets.

In practical terms, the findings support the feasibility of automated harmful waste detection in environmental monitoring applications, such as aerial surveys and smart cleanup robotics. However, they also emphasize the need for larger and more diverse training datasets to further reduce false positives. Future research should investigate the integration of multimodal inputs (e.g., hyperspectral imaging, contextual metadata), domain adaptation across environments, and the deployment of lightweight models on embedded platforms.

Ultimately, deep learning approaches such as transfer learning hold strong potential to accelerate the monitoring and mitigation of plastic pollution, thereby contributing to more sustainable management of aquatic and terrestrial ecosystems.

Despite the promising results, the current evaluation was performed on a controlled benchmark dataset rather than on real UAV imagery. Therefore, additional experiments on aerial environmental datasets are required to verify the robustness and generalization capability of the proposed approach under realistic operating conditions. Such validation represents an important step toward practical deployment in smart environmental monitoring and autonomous waste detection applications. Future work should therefore include repeated experimental runs and cross-validation-based evaluation protocols in order to obtain more robust estimates of model stability and generalization capability.

Although VGG16 represents a classical convolutional baseline rather than the most recent state-of-the-art architecture, it remains widely used in image classification research and provides a well-established reference model for performance comparison. Future work may include evaluation using more recent architectures such as Vision Transformers (ViT) or lightweight transformer-based hybrid models.

7 ACKNOWLEDGMENTS

This student paper (T. Kurimská and T. Kurucová) was co-authored with university teachers (Š. Horvát, L. Antoni, G. Semanišin) as part of the Erasmus+ project *Developing Talents in Artificial Intelligence to Solve Disruptive Environmental Problems* (Project No. 2023-1-PL01-KA220-HED-000166765). The project aims to enhance the quality and impact of AI education while strengthening the digital readiness of higher education institutions and their students.

References

- [1] A. L. Andrady, *Plastics and environmental sustainability*. John Wiley & Sons, 2015.
- [2] United Nations Environment Programme, “From pollution to solution: A global assessment of marine litter and plastic pollution.” 2021. Available: <https://www.unep.org/resources/pollution-solution-global-assessment-marine-litter-and-plastic-pollution>
- [3] S. B. Borrelle, J. Ringma, K. L. Law,..., and C. M. Rochman, “Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution,” *Science*, vol. 369, no. 6510, pp. 1515–1518, 2020.
- [4] F. Chollet, *Deep learning with python*, 2nd ed. Shelter Island, NY: Manning Publications, 2021.
- [5] J. Bobulski and M. Kubanek, “Waste classification system using image processing and convolutional neural networks,” in *International work-conference on artificial neural networks (IWANN)*, Cham: Springer International Publishing, 2019, pp. 350–361.

- [6] S. Majchrowska *et al.*, “Deep learning-based waste detection in natural and urban environments,” *Waste Management*, vol. 138, pp. 274–284, 2022.
- [7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *International Conference on Learning Representations (ICLR)*, 2015.
- [8] G. Thung and M. Yang, “Classification of trash for recyclability status.” CS229 Course Project Report, Stanford University, 2016. Available: <https://cs229.stanford.edu/proj2016/report/ThungYang-ClassificationOfTrashForRecyclabilityStatus-report.pdf>
- [9] G. Thung and M. Yang, “TrashNet: Dataset of images of trash.” 2016. Available: <https://github.com/garythung/trashnet>
- [10] R. A. Aral, Şeref Recep Keskin, M. Kaya, and M. Hacıömeroğlu, “Classification of TrashNet dataset based on deep learning models,” in *2018 IEEE international conference on big data (big data)*, 2018, pp. 2058–2062.
- [11] V. Ruíz, Ángel Sánchez, J. F. Vélez, and B. Raducanu, “Automatic image-based waste classification,” in *From bioinspired systems and biomedical applications to machine learning: 8th international work-conference on the interplay between natural and artificial computation, IWINAC 2019, almería, spain, proceedings, part II*, vol. 11847, in Lecture notes in computer science, vol. 11847., Cham: Springer, 2019, pp. 422–431.
- [12] O. Adedeji and Z. Wang, “Intelligent waste classification system using deep learning convolutional neural network,” *Procedia Manufacturing*, vol. 35, pp. 607–612, 2019.
- [13] Z. Zhu, W. J. Parker, and A. Wong, “Leveraging deep learning for automatic recognition of microplastics (MPs) via focal plane array (FPA) micro-FT-IR imaging,” *Environmental Pollution*, vol. 337, p. 122548, 2023.
- [14] “DroneDeploy: Aerial mapping and analysis platform for environmental monitoring.” <https://www.dronedeploy.com/>.
- [15] U. Andriolo, K. Topouzelis, T. H. M. van Emmerik, and A. Papakonstantinou, “Drones for litter monitoring on coasts and rivers: Suitable flight altitude and image resolution,” *Marine Pollution Bulletin*, vol. 195, p. 115521, 2023.
- [16] U. Andriolo *et al.*, “Beached and floating litter surveys by unmanned aerial vehicles (UAVs): Operational procedures and considerations,” *Remote Sensing*, vol. 14, no. 6, p. 1336, 2022.
- [17] G. Escobar-Sánchez, G. Markfort, M. Berghald, L. Ritzenhofen, and G. Schernewski, “Aerial and underwater drones for marine litter monitoring in shallow coastal waters: Factors influencing item detection and cost-efficiency,” *Environmental Monitoring and Assessment*, vol. 194, no. 12, p. 863, 2022.
- [18] N. Maharjan *et al.*, “Detection of river plastic using UAV sensor data and deep learning,” *Remote Sensing*, vol. 14, no. 13, p. 3049, 2022.
- [19] Y. Jeong *et al.*, “A study on the monitoring of floating marine macro-litter using a multi-spectral sensor and classification based on deep learning,” *Remote Sensing*, vol. 16, no. 23, p. 4347, 2024.
- [20] C. M. Rochman, M. A. Browne, A. J. Underwood, J. A. van Franeker, R. C. Thompson, and L. A. Amaral-Zettler, “The ecological impacts of marine debris: Unraveling the demonstrated evidence from what is perceived,” *Ecology*, vol. 94, no. 7, pp. 1383–1393, 2013.

- [21] NOAA Marine Debris Program, "Can plastic marine debris harm animals?" <https://marinedebris.noaa.gov/can-plastic-marine-debris-harm-animals-0>, 2025.
- [22] NOAA National Ocean Service, "What are the impacts of microplastics?" <https://oceanservice.noaa.gov/education/tutorial-coastal/marine-debris/md04-sub-01.html>, 2024.
- [23] D. Xanthos and T. R. Walker, "International policies to reduce plastic marine pollution from single-use plastics (plastic bags and microbeads): A review," *Marine Pollution Bulletin*, vol. 118, no. 1–2, pp. 17–26, 2017.
- [24] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision (ICCV)*, 2017, pp. 2980–2988.

Terézia Kurimská is a master's student of Computer Science at Faculty of Science at Pavol Jozef Šafárik University in Košice. She specializes in data analysis, predictive modelling, and artificial intelligence techniques. Her academic work concentrates on processing environmental data and evaluating neural network architectures for forecasting tasks. She participates in student research activities and applied AI projects.

Terézia Kurucová is a master's student in Data Analysis and Artificial Intelligence at Faculty of Science at Pavol Jozef Šafárik University in Košice. Her research interests focus on machine learning, time series analysis, and environmental data modelling. She is actively involved in projects related to air quality monitoring and neural network applications.

Šimon Horvát is an assistant professor at the Institute of Computer Science at Faculty of Science at Pavol Jozef Šafárik University in Košice. His research focuses on advanced machine learning and neural network methods. He has contributed to several scientific publications and research projects addressing intelligent data analysis.

Lubomír Antoni is an associate professor at the Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University in Košice. His research interests include artificial intelligence, fuzzy systems, data mining, and applied machine learning.

Gabriel Semanišin is a professor of Computer Science at Faculty of Science, Pavol Jozef Šafárik University in Košice. As part of his research activities, he focuses mainly on algorithmic graph theory and its application in various areas of theoretical and applied informatics. He is a co-guarantor of the study programs Applied Informatics, Data Analysis and Artificial Intelligence and Computer Science. He was a supervisor of 6 PhD students in study programs Computer Science, Discrete Mathematics and Theory of Teaching Informatics.