Article

RealWaste: A Novel Real-Life Data Set for Landfill Waste Classification Using Deep Learning

Sam Single, Saeid Iranmanesh *[ID] and Raad Raad [ID]

School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2522, Australia; ss228@uowmail.edu.au (S.S.); raad@uow.edu.au (R.R.)
* Correspondence: saeidim@uow.edu.au

Abstract: The accurate classification of landfill waste diversion plays a critical role in efficient waste management practices. Traditional approaches, such as visual inspection, weighing and volume measurement, and manual sorting, have been widely used but suffer from subjectivity, scalability, and labour requirements. In contrast, machine learning approaches, particularly Convolutional Neural Networks (CNN), have emerged as powerful deep learning models for waste detection and classification. This paper analyses VGG-16, InceptionResNetV2, DenseNet121, Inception V3, and MobileNetV2 models to classify real-life waste when trained on pristine and unadulterated materials, versus samples collected at a landfill site. When training on DiversionNet, the unadulterated material dataset with labels required for landfill modelling, classification accuracy was limited to 49.69% in the real environment. Using real-world samples in the newly formed RealWaste dataset showed that practical applications for deep learning in waste classification are possible, with Inception V3 reaching 89.19% classification accuracy on the full spectrum of labels required for accurate modelling.

Keywords: classification; machine learning; deep learning; convolution neural networks; dataset; landfill waste; waste management; sustainability

1. Introduction

Solid waste encompasses all materials that are produced as a result of human and societal activities but have lost their utility or desirability [1]. This waste may be items belonging to three primary groups: (i) recyclable inorganics fit for repurposing, i.e., plastics, metals; (ii) divertible organics from which energy and fertilizer can be derived, i.e., food and vegetation; and (iii) inorganic materials requiring landfill, i.e., ceramics, treated wood. Improper management of the disposal of waste poses significant risks to environmental sustainability and human health, resulting from toxic wastewater byproducts and the global warming potential of methane in landfill gas [2–4]. For example, in 2020, 3.48% of global greenhouse gas production was attributed to the waste disposal sector, with the methane component of landfill gas accounting for 20% of worldwide methane emissions in the following year [5,6]. The chemical composition and production of both wastewater and landfill gases are heavily dependent on the organic content within the waste processed, with significant variations arising in the presence of food and vegetation materials [7–10]. Detection of the waste types under meticulously refined material classes in both an accurate and timely manner is therefore essential in sustainable waste management, by ensuring accountability for both seasonal variations and recycling uptake [11–13].

In the literature, various approaches for the detection and classification of solid waste have been explored, including both traditional and machine learning methods. Traditional approaches, such as visual inspection, weighing and volume measurement, and manual sorting, have been widely used for waste detection. Visual inspection relies on the expertise of human operators to visually assess and classify waste based on its appearance. Weighing and volume measurement techniques quantify waste by measuring its weight or volume,

providing valuable information for waste estimation and management planning. Manual sorting, commonly employed in recycling facilities, involves the physical separation of different types of waste by workers. While these traditional approaches have their utility, they are limited by subjectivity, scalability, and labour requirements.

On the other hand, machine learning approaches have emerged as powerful tools in waste detection and classification. Image processing and computer vision techniques, combined with machine learning algorithms, enable automated waste detection and classification based on visual characteristics. These approaches analyse images or video footage to identify and categorize different types of waste, enhancing waste sorting, recycling efforts, and landfill operations [14–18]. As an alternative to visual analytics, sensor-based systems that are integrated with machine learning algorithms utilize Internet of Things (IoT) devices or embedded sensors in waste bins and collection vehicles to detect abnormalities belonging to material types that are not allowed in the given waste stream [19]. Real-time analytics provided by these systems offer valuable insights for decision-making in waste management.

One prominent deep learning architecture used in waste classification is the Convolutional Neural Network (CNN). CNNs are specifically designed for processing and analysing visual data, making it an ideal choice for classifying landfill waste diversion. Landfills contain a wide range of waste materials, and accurately categorizing them is essential for effective waste management. By leveraging CNNs, the classification process can be automated, significantly improving the efficiency and accuracy of waste diversion strategies such as recycling, composting, and proper disposal. CNNs excel at image analysis and feature extraction, allowing them to capture intricate details and patterns from waste images. The convolutional layers in CNNs are able to identify edges, textures, and other relevant characteristics, enabling the network to learn and leverage these features for accurate waste classification. Pooling layers nested throughout the architecture down-sample produced feature maps, allowing models to generalize object specific characteristics into differing contexts. The hierarchy of the aforementioned layers enables models to extract low-level features early in the architecture and derive semantical information in later layers forming the basis of classifications at the output (Figure 1).
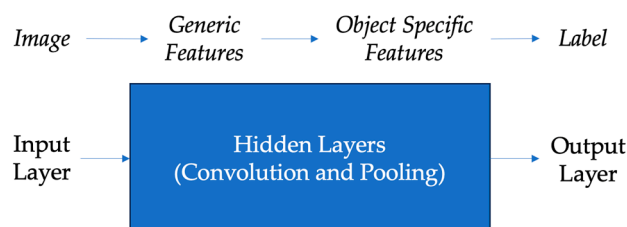
Figure 1. Generic convolution neural network architecture and classification process.

Despite the promising potential of machine learning approaches, they also present challenges. Sensor-based systems may be limited in their ability to detect all material types, providing no information on other items present in the waste stream, which can affect their suitability for comprehensive waste classification. Image processing and computer vision techniques heavily rely on the quality and representativeness of their datasets, and biases in the dataset labels will impact the accuracy of trained models in real-world waste environments. Furthermore, the current state of deep learning models used in waste classification often focuses on a limited number of waste material classes, failing to adequately represent the full range of detectable material types found in real-world waste scenarios. Additionally, these models often rely on object representations that assume the pristine forms of materials, disregarding the diverse and degraded states commonly observed in waste materials. These limitations indicate the need for more diverse and comprehensive datasets that accurately represent the appearance and characteristics of waste materials.

Based on the above limitations, this paper proposes a comprehensive dataset called RealWaste that covers various classes of landfilled waste for sustainable waste management. Furthermore, five deep learning models are applied over RealWaste, and other existing datasets and we provide a critical analysis over the results to see the impact of quality of the dataset and more detailed classes of the waste. Hence, the main contributions of this paper are as follows:

- We created the first dataset, RealWaste, to comprehensively cover more classes of landfilled waste required for sustainable waste management. It includes three primary material types for divertible organics, recyclable inorganics, and waste, with meticulously refined labels for food, vegetation, metal, glass, plastic, paper, cardboard, textile trash, and miscellaneous trash. There are 4808 samples captured with the resolution of 524×524 from the Whyte's Gully Waste and Resource Recovery Centre's landfill site located in Wollongong, New South Wales, Australia, where waste items from municipal waste collection comingle and contaminate one another.

- The evaluation and analysis of five deep learning models over the RealWaste dataset and the datasets existing in the literature. The selection of models used has been intentionally made broad with respect to their design motivations to draw generalised outcomes on the larger input image resolution. Moreover, our objective is to evaluate the performance of the model when type of material over different items is important to be detected. The outcome shows that waste detection is indeed achievable for the meticulously refined classes required in sustainable waste management, with every model reaching above 85% classification accuracy, with the best performer at 89.19%.

The remainder of this paper is structured as follows: Section 2 discusses the related works. The RealWaste dataset and the proposed models for evaluation are detailed in Section 3. Section 4 discusses the results and outcomes and Section 5 concludes the paper.

2. Related Work

The impact of the training phase of a CNN model on performance is well recognized. Overfitting occurs when datasets lack diversity in the object features present in labels, causing models to become proficient at classifying training data but lose the ability to generalize label features. Deep learning often faces the challenge of small datasets, which makes it difficult to effectively model the problem domain. To address this, large-scale datasets with thousands of labels and millions of images have been assembled, such as ImageNet by Russakovsky et al. [20]. Training models on such datasets with diverse features has led to the exploration of heterogenous transfer learning in deep learning applications. This supposition has been extensively investigated in the literature. For example, He et al. [21] examined the potential of heterogeneous transfer learning in hyperspectral image classification. Their novel heterogenous approach outperformed four popular classifiers significantly across four datasets, with an improvement of 5.73% in accuracy compared to the closest performer on the Salinas dataset. Heterogenous transfer learning becomes a vital consideration for applications on smaller datasets to ensure success.

The relationship between dataset labels and size, and the depth and width of fully connected layers has been studied by Basha et al. [22]. Their findings indicate that, regardless of dataset size, fully connected layers require fewer parameters in deeper networks, while shallower architectures require greater width to achieve the same results.

The connections and organization of hidden layers within networks have also been analysed by Sharma et al. [23]. They compared three popular architectures: AlexNet, GoogLeNet, and ResNet-50. The performance accuracy on the CFIR-100 dataset revealed significant differences, with AlexNet scoring 13.00%, GoogLeNet scoring 68.95%, and ResNet-50 scoring 52.55%. These results demonstrate that deeper networks are not always the best option, and that connections and network layouts are often more important for the specific task at hand.

Data preprocessing and augmentation play a crucial role in deep learning. Data augmentation, which generates new images from existing samples, is particularly important

when training on smaller datasets. It improves a model's generalization capacity and learnable features by increasing the diversity of training data. Shijie et al. [24] evaluated the effectiveness of different data augmentation techniques on the AlexNet model trained on the CFIR-10 and ImageNet datasets. Geometric and photometric transformations were found to boost performance, with geometric transformations performing the best. This result was confirmed by Zhang et al. [25] in their study on leaf species classification. Geometric transformations outperformed photometric transformations across the same model and dataset, indicating their effectiveness in improving performance and generalization.

Data augmentation also helps address the performance degradation caused by imbalanced datasets. Lopez de la Rosa et al. [26] explored this issue in semiconductor defect classification. By applying geometric augmentations to scale an imbalanced dataset, the study achieved a significant improvement in the mean F1-score, demonstrating the effectiveness of geometric augmentation in handling imbalanced datasets.

Image resolution is another factor that can affect feature extraction. Sabottke and Spieler [27] investigated the impact of resolution on radiography classifications. Higher resolutions provided finer details for feature extraction and led to increased classification accuracy. However, trade-offs between smaller batch sizes and higher resolutions had to be considered due to hardware memory constraints. The study highlights the importance of choosing the optimal resolution for specific applications.

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC), introduced by Russakovsky et al. [20], has been a driving force in advancing CNNs. Models developed for the ILSVRC have tackled challenges related to dataset size and real-world applicability. For example, the AlexNet model, developed by Krizhevsky et al. [28] to win the ILSVRC in 2012, addressed the computational time requirements by spreading its architecture across two GPUs for training. Subsequent advancements, such as the VGG family of architectures proposed by Simonyan and Zisserman [29], GoogLeNet (Inception V1) developed by Szegedy et al. [30], and ResNet architectures developed by He et al. [31], pushed the depths of networks even further and achieved high object detection and classification accuracies. DenseNet architectures, as described by Huang et al. [32], offered an alternative approach by connecting convolution and pooling layers into dense blocks. InceptionResNet, combining inception blocks with residual connections and developed by Szegedy et al. [33], demonstrated improved convergence and lower error rates compared to the Inception and ResNet families of models from which it was inspired. MobileNetV2, developed by Sandler et al. [34], addressed the computational complexity issue by replacing standard convolution layers with depth wise convolutions, enabling processing on hardware-restricted systems.

In waste classification literature, various models from the aforementioned architectures have been implemented and evaluated on the TrashNet dataset [35]. The combination of a Single Shot Detector (SSD) for object detection and a MobileNetV2 model trained via transfer learning achieved high accuracy, outperforming other models [14]. Moreover, optimization techniques applied to baseline models, such as the Self-Monitoring Module (SMM) for ResNet-18, showed significant improvements in performance [15]. However, studies have shown that under the right conditions, popular baseline models can achieve similar or even better results compared to more complex implementations [16,17]. Nonetheless, the major issue in the literature is the lack of organic labels in the TrashNet dataset, limiting its suitability for waste auditing and landfill modelling.

To address the limitations of the TrashNet dataset, Liang and Gu [18] proposed a multitask learning (MTL) architecture that localized and classified waste using the WasteRL dataset. Their specialized network achieved high accuracy compared to other architectures. However, the dataset's labelling does not capture the spectrum of waste required for accurate modelling, as the absence of organic breakdown and inorganic recyclable breakdown limits its practical application.

Outside of the CNN approach, advancements in deep learning transformers have shown significant promise to image classification. Originally designed for natural language processing tasks to address the weighted context of individual words in the meaning

of sentences through self-attention mechanisms, transformers have achieved significant performance at lower computational complexities. Yu et al. demonstrated their particular suitability towards long-term predictions on time series data, showing their efficiency at representing complex problem domains [36]. Dosovitskiy et al. demonstrated how this ability translates to computer vision, matching the performance of modern CNNs across an array of benchmark datasets [37]. Within the waste classification space, transformers have begun to find particular use in recognising construction waste composition based off mixed samples in images. Dong et al. showed significant improvements within this scope for semantic segmentation and fine-grained classification over baseline results when applying transformers to the task [38].

In summary, the related work has investigated various aspects of deep learning, including the impact of dataset size, network architecture, data preprocessing and augmentation, image resolution, and the influence of the ILSVRC on CNN advancements. The studies have provided insights into improving performance, handling imbalanced datasets, and addressing computational constraints. However, the limitations in the datasets and the lack of comprehensive labelling in waste auditing studies remain challenges for practical applications in waste composition data and landfill modelling.

3. Methodology

This section has been structured towards addressing the major gap in the waste classification literature, which pertains to the limitations in datasets and the lack of comprehensive labelling in waste auditing studies [14–18]. To overcome these challenges and improve the accuracy and practicality of waste classification models, we collect and preprocess data to be utilized by different models. Hence, we aim to (i) evaluate the effectiveness of pristine, unadulterated material datasets for training models to classify waste in the real environment; (ii) assess the impacts of training on real waste samples to compare the dataset approaches and reveal the suitability of each; and (iii) determine the optimal model for waste classification in the live setting.

An experimental study has been conducted to analyse the performance of five popular CNN models, on labelling waste across two datasets. Specifically, VGG-16 has been selected for its shallow design [29], DenseNet121 for pushing layer depth [32], Inception V3 for its grouping of hidden layers [30], InceptionResNet V2 for combining techniques [33], and finally MobileNetV2 for its lightweight design [34].

In terms of dataset, the first dataset is DiversionNet, that has been assembled by combining the TrashNet dataset with elements from several opensource datasets to populate the labels which TrashNet lacks and represent the approach in the literature largely based on pristine and unadulterated objects for model training [14–18]. The second dataset is called RealWaste and we collected the samples during the biannual residential waste audit at the Wollongong Waste and Resource Recovery Centre's landfill. The samples were taken from bins of municipal, recycling, and organic waste streams. To clearly demonstrate the quality of the two datasets, DiversionNet relies on object samples in their pristine and unadulterated forms, shown in Figure 2a. Conversely, RealWaste is made up of items arriving at landfill, where comingling occurs between material types and objects undergo structural deformation, shown in Figure 2b. The comingling of materials is attributed to the presence of organic waste, where its decomposition and remains in food packaging contaminate other items.

For evaluation purposes, a test dataset was assembled prior to the curation of RealWaste, consisting of 10% of the total images collected from each class during the residential waste audit, selected at random. Both RealWaste and DiversionNet training applied an 80:20 training to validation split.

Table 1 provides insights into the distribution of images across different waste categories and highlight potential imbalances or variations within the datasets. For instance, in the DiversionNet dataset, the label with the highest image count is Paper (594 images), followed by Plastic (482 images) and Glass (501 images). On the other hand, the label with the

lowest image count in the DiversionNet dataset is Miscellaneous Trash (290 images). These variations in image counts suggest that certain waste categories may be overrepresented or underrepresented in the dataset.

(**a**)   (**b**)
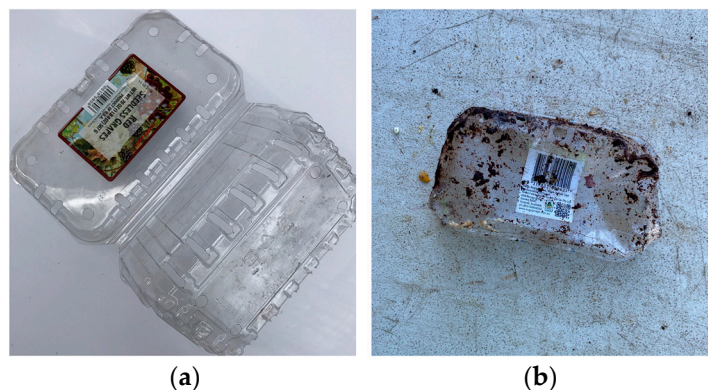
Figure 2. Training dataset samples from the unadulterated and practical environment approaches: (**a**) DiversionNet, sourced from [35]. (**b**) RealWaste (right).

Table 1. Labels and image count of datasets.

Label	DiversionNet	RealWaste	Test
Cardboard	403	417	46
Food Organics	311	370	41
Glass	501	378	42
Metal	410	711	79
Miscellaneous Trash	290	445	50
Paper	594	497	55
Plastic	482	831	92
Textile Trash	417	286	32
Vegetation	519	392	44

3.1. Data Preprocessing

From the findings of the literature, two strategies have been implemented to treat the datasets before commencing training: relatively large image sizes; and data augmentations.

3.1.1. Image Size

Each dataset has been scaled to 524 × 524 image resolutions to better distinguish the finer object features and reach better classification accuracies [27]. The selection was made to accommodate several factors present in waste classification: the comingled state of material types; transparent objects in plastic and glass classes; and similarities between specific objects (e.g., glass and plastic bottles). Although the resolution is relatively high, the smaller datasets and mini-batch sizes meet hardware memory requirements, and transfer learning, the process of initializing models with pretrained weights from generalized classification tasks, allows for manageable training times.

3.1.2. Data Augmentation

Data augmentation methods have been applied to the training datasets, using Python and the Augmentor library. Two sets of augmentations outlined in Table 2 have been selected and applied to each image, respectively, tripling the total number of images within each training dataset.

Both sets from Table 2 consist of a combination of geometric transformations and are aimed at reducing the effects of overfitting to the training dataset. Geometric transformations have been selected due to their enhancements on deep learning vision-based

tasks [14–16,25] and effectiveness on imbalanced datasets [26]. The techniques were combined to further increase performance as found in [24].

Table 2. Training dataset data augmentation.

Set No.	Augmentation Techniques
1	Horizontal flip and elastic distortion
2	Rotate and shear

Each combination contains one transformation on the orientation of the object to provide its features in a different context (horizontal flipping and rotation) and another distorting objects to increase the feature space diversity, to account for the varied structural state of objects received at landfill (elastic distortion and shear). Samples have been included for an original image, elastic distortion and horizontal flipping, and rotation and shearing in Figure 3a–c, respectively.

(a)   (b)   (c)
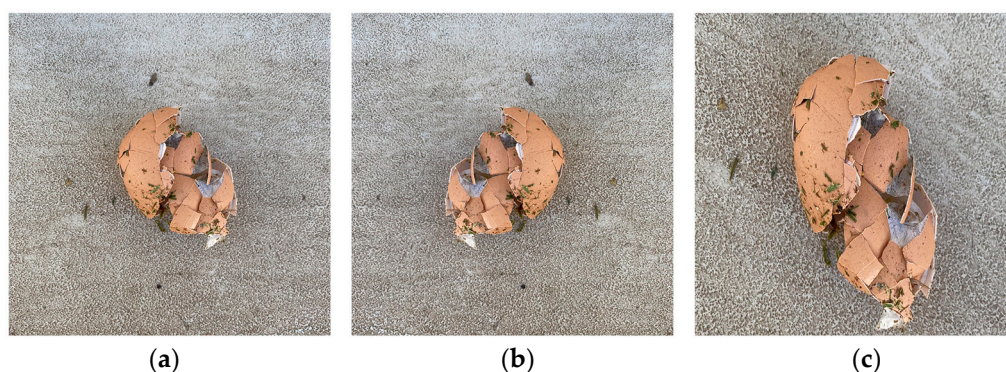
Figure 3. Data augmentations: (**a**) Original image. (**b**) Elastic distortion and horizontal flip. (**c**) Rotate and shear.

3.2. Model Training

Prior to training, each model was loaded with pretrained weights from the ImageNet dataset [20]. To preserve the feature extraction capability imparted through transfer learning, models were trained in progressive stages. Initially, the fully connected layers were trained to adapt the model to the waste domain before fine tuning the feature extraction layers. Different learning rates were applied across the stages where required to prevent destabilization of the pre-learned feature extraction.

Hyperparameter Specification

The hyperparameter selection of batch sizes and learning rate has been detailed in Table 3. A mini-batch approach was taken for batch sizing, whilst relatively low learning rates were selected to ensure stability. No preset epoch size was set as each model displays varied convergence behaviour and rather, model training was cut off once performance on the validation dataset ceased to improve.

Table 3. Hyperparameter specifications.

Model	Batch Size	Learning Rate for Fully Connected Layers	Learning Rate for Feature Extraction Layers
VGG-16	4	1×10^{-5}	1×10^{-5}
DenseNet121	16	1×10^{-4}	1×10^{-5}
Inception V3	32	1×10^{-5}	1×10^{-5}
InceptionResNet V2	32	1×10^{-6}	1×10^{-6}
MobileNetV2	4	1×10^{-5}	1×10^{-5}

4. DiversionNet versus RealWaste

This section details the results achieved by the five models over DiversionNet and RealWaste during training and testing. Further to this, the best performing model and dataset combination has been considered with respect to its confusion matrix to reveal greater insights on deep learning waste classification.

4.1. Model Training

To analyse the learning behaviour of each model, the training and validation accuracies have been plotted against the epoch span in Figures 4–8.

(**a**)   (**b**)
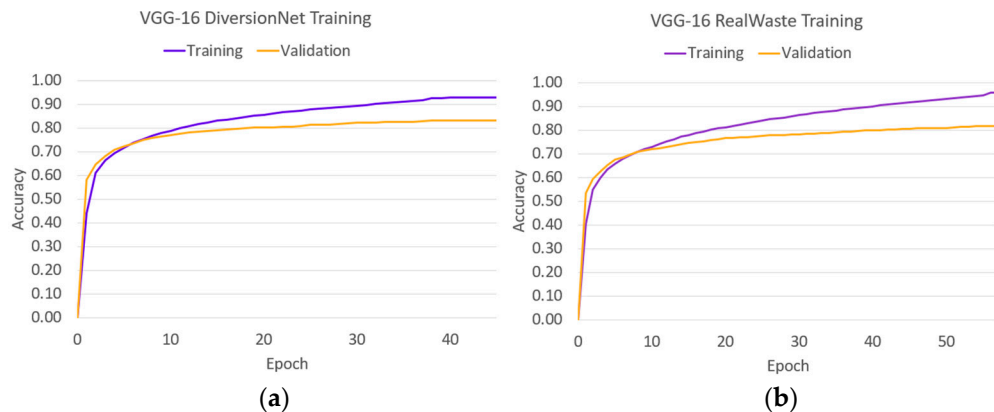
Figure 4. VGG-16 training performance: (**a**) DiversionNet dataset. (**b**) RealWaste dataset.

(**a**)   (**b**)
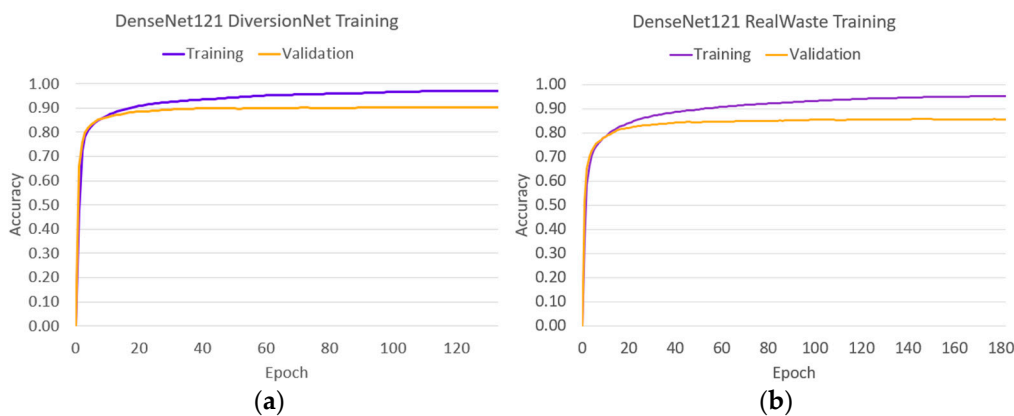
Figure 5. DenseNet121 training performance: (**a**) DiversionNet dataset. (**b**) RealWaste dataset.

(**a**)   (**b**)
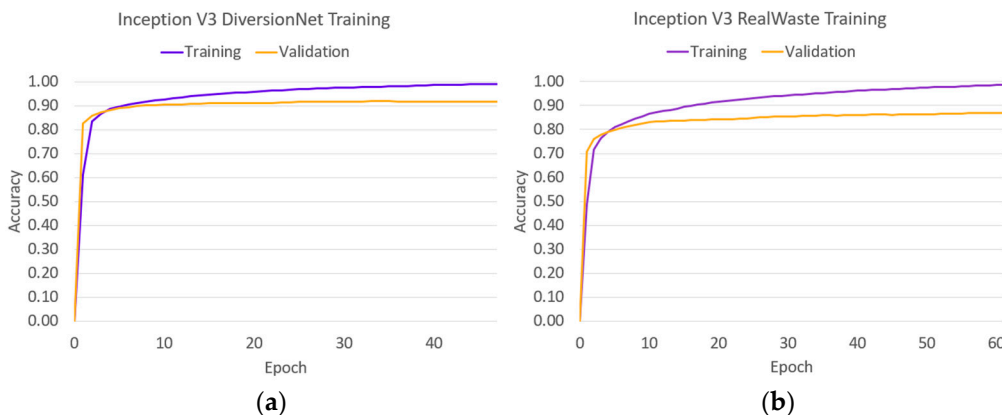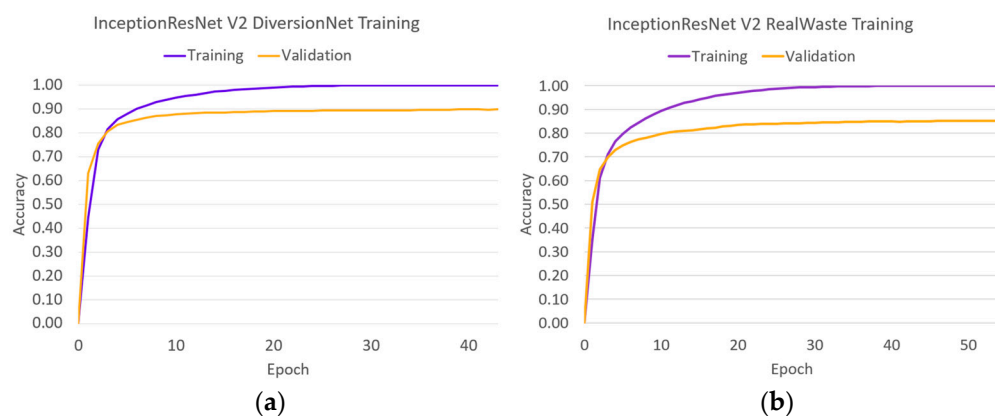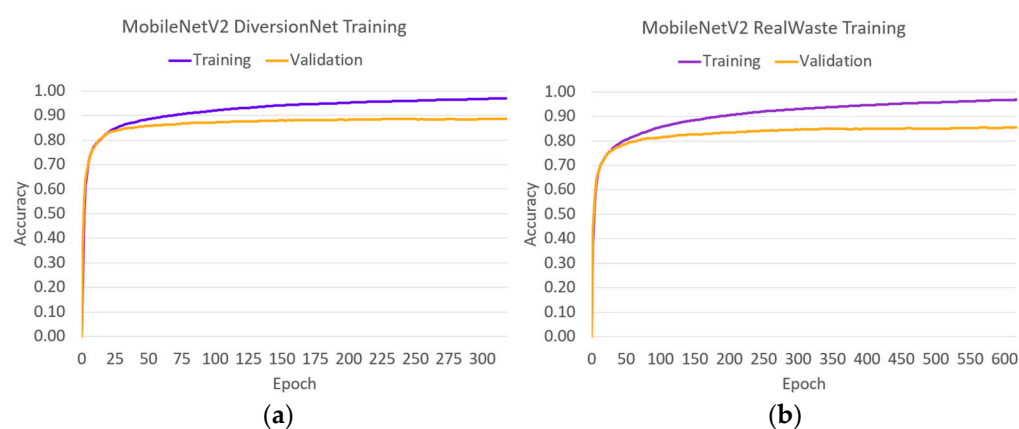
Figure 6. Inception V3 training performance: (**a**) DiversionNet dataset. (**b**) RealWaste dataset.

Figure 7. InceptionResNet V2 training performance: (**a**) DiversionNet dataset. (**b**) RealWaste dataset.

Figure 8. MobileNetV2 training performance: (**a**) DiversionNet dataset. (**b**) RealWaste dataset.

The convergence behaviour suggests that the authentic waste materials in RealWaste present a more diverse and complex feature space than their pristine counterparts in DiversionNet, as training and validation accuracies were greater in every model when trained on the latter. Furthermore, the differences between the metrics increased under RealWaste training by 4% for VGG-16 in Figure 4a,b, 3% for DenseNet121 in Figure 5a,b, 4% for Inception V3 in Figure 6a,b, 5% for InceptionResNet V2 in Figure 7a,b, and 3% for MobileNetV2 in Figure 8a,b. Aside from the differences between the training datasets, the individual performance of models exposes insights on deep learning waste classifiers. On both datasets, VGG-16 performed the worst reaching lower training and validation accuracies compared to the deeper models. Although increasing the network depth is not a certain way to success [23], this outcome suggests that some degree is required for accurate waste classification due to the complex feature space present in the waste material.

4.2. Testing Performance

The performance of the models trained on DiversionNet and RealWaste has been evaluated on the testing dataset with the results presented in Table 4.

The results from Table 4 are very similar to the training performance of the models. Both Inception V3 and DenseNet121 models achieved 89.19% classification accuracy when trained on RealWaste; however, the former performed the best by achieving the highest results in the other metrics. Specifically, the 91.34% precision, 87.73% recall, and 90.25% F1-score reached by Inception V3 were more than 0.5% greater than the next closest DenseNet121 model for each metric. With respect to accuracy between RealWaste and DiversionNet training in Table 4, VGG-16 showed an improvement of 58.83%, 48.86% in DenseNet121, 31.50% in Inception V3, 42.62% in InceptionResNet V2, and finally, 60.50% in MobileNetV2.

Table 4. Performance results on testing dataset.

Model	Accuracy	Precision	Recall	F1-Score
DiversionNet Training				
VGG-16	26.82%	28.80%	25.99%	27.32%
DenseNet121	40.33%	44.80%	34.93%	39.25%
Inception V3	49.69%	55.43%	39.29%	45.98%
InceptionResNet V2	44.70%	52.83%	40.75%	46.01%
MobileNetV2	27.65%	28.50%	24.95%	26.61%
RealWaste Training				
VGG-16	85.65%	87.74%	84.82%	86.26%
DenseNet121	89.19%	90.06%	86.69%	89.62%
Inception V3	89.19%	91.34%	87.73%	90.25%
InceptionResNet V2	87.32%	89.69%	85.03%	88.49%
MobileNetV2	88.15%	89.98%	85.86%	87.87%

In every single case from Table 4, the RealWaste trained networks greatly outperformed their DiversionNet counterparts. Any model trained on data from a certain sample space has a bias towards classification within said sample space. The significantly greater performance in every metric shows the negative impact from the training samples collected outside the authentic waste environment. The results show that the learning imparted by the pure objects is far too biased for their characteristics to be generalized in the real-life landfill setting, indicating that waste objects possess a much more complex feature space.

In both datasets detailed by Table 4, VGG-16 performed the worst with a classification accuracy of 26.82% when trained on DiversionNet and 85.65% on RealWaste. Like the training results, this can be attributed to the lack of layer depth within the architecture. MobileNetV2 performed well on RealWaste training with 88.15% accuracy, but like VGG-16, rather poorly on DiversionNet with 27.65%. Interestingly, the difference in classification accuracy between the worst performing shallow and lightweight designs compared to the best performing Inception V3 model was much larger when trained on DiversionNet rather than RealWaste, with a difference of around 20% in the former case, and only 4% in the latter. Even though the results have indicated that the pristine and unadulterated objects within DiversionNet misrepresent real-life waste conditions, the deeper, more complex models are shown to be better at generalizing the available features to the live environment.

With respect to the individual datasets in Table 4, the remaining DenseNet121 and InceptionResNet V2 architectures performed well relative to the others, with classification accuracies of 40.33% and 44.70% on DiversionNet, and 89.19% and 87.32% on RealWaste, respectively. These classifications accuracies along with the results of Inception V3 give some credit towards increased layer depth and complexity being required for waste classification in real-life conditions. Specifically, the computational complexity within the models can be considered through the number of parameters in the network. These parameters refer to weights and biases applied to the inputs of neurons within the hidden layers and are updated during training to encode the learned feature representations. Therefore, larger numbers of parameters require more computational resources and thus increase the time of processing the images. Although more parameters may surmise a greater ability to capture feature representations, the structuring of the layers themselves is often more important to the specific task at hand [23].

The results demonstrated this notion, where MobileNetV2 outperformed InceptionResNet V2 in the RealWaste experiment group with the former having just over 2 million parameters and the latter 58 million, meaning that network specific classification techniques outweigh this factor. Inception V3 outperforming InceptionResNet V2 emphasizes this finding, where the latter combines the inception block design of the former with residual connections. In the past, the more complex InceptionResNet V2 has been shown as the better out of the two [33]; however, the model's ability to classify waste feature space has proven otherwise.

Of further note is the extremely similar performance compared to the complexity of DenseNet121 and Inception V3, with the former having around 7 million parameters, and the latter a much larger amount at 26 million. Although VGG-16 has the second largest number, totalling close to 34 million, its lack of performance is attributed to its architectural constraints with limited depth.

Aside from raw accuracy, the RealWaste trained models performed extremely well in precision, recall, and F1-scores in Table 4. The DiversionNet trained models performed significantly worse, furthering the indication that waste classification requires training samples from the real environment. However, both experiment groups have the tendency towards a higher precision than recall. The difference on average for RealWaste was 3.74% and 8.89% for DiversionNet. In practice, this shows that models will tend to minimise the false positives rather than false negatives. In the case of RealWaste, this is less of an issue given the smaller difference and high F1-scores showing a good balance; however, it is still notable. Minimising the number of false negatives to increase recall is the more desirable outcome. Classifiers need to be able to predict the correct label for any object input to the system, given the almost endless diversity in the actual objects belonging to each waste label.

4.3. Analysis

To gain a deeper understanding of its classification performance, a detailed analysis was conducted using the confusion matrix, as shown in Figure 9. The confusion matrix provides insights into the model's performance in correctly classifying different waste categories and reveals the presence of false negatives and false positives.
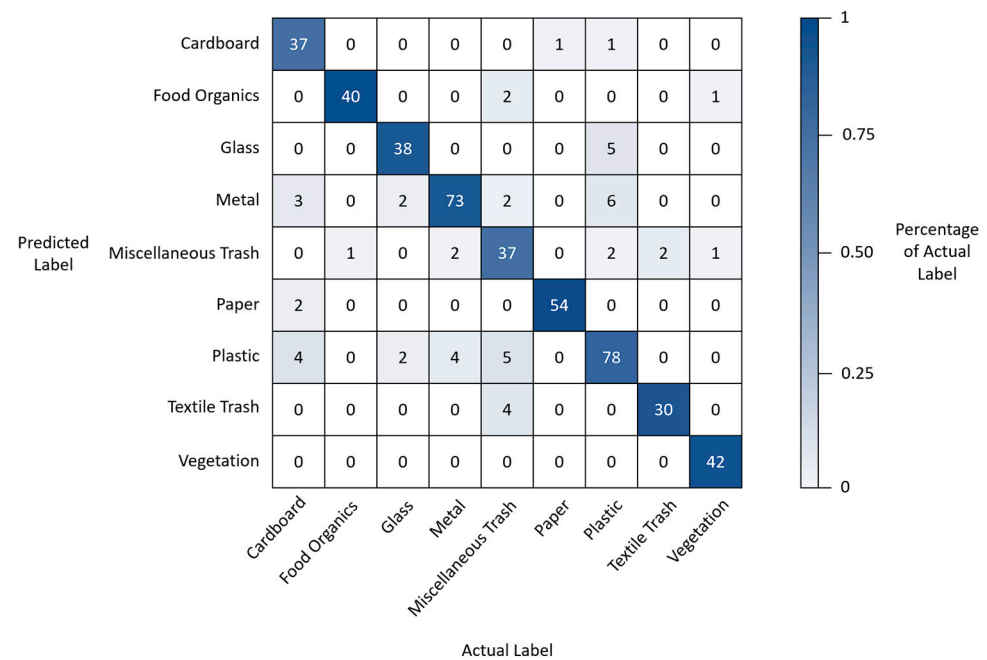

Figure 9. Confusion matrix for Inception V3 best model testing.

Out of all the labels in Figure 9, the confusion between classes was the most prominent with miscellaneous trash objects. The accuracy for this category was only 74.00%, indicating a significant challenge in correctly identifying these objects. The class encountered 13 false negatives, where actual miscellaneous trash items were mislabelled, and eight false positives, where other waste types were incorrectly classified. The diverse and sparse feature space of miscellaneous trash, characterized by different shapes, textures, and colours, makes it difficult to encode identifiable features into the convolution layers. This issue is somewhat unavoidable due to the label's purpose of containing objects that do not fit into any other material types. However, the poor performance highlights the need for further

refinement in labelling objects belonging to the miscellaneous trash category to decrease feature space diversity and improve overall accuracy.

The cardboard label also exhibited relatively poorer performance, with only 80.44% of the total sample size correctly labelled. An analysis of the labels responsible for false negatives and positives revealed confusion with other waste categories. Samples from the metal, paper, and plastic classes, which can possess branding and graphical designs similar to certain cardboard products, caused mislabelling. This confusion is evident in Figure 10, where cardboard was labelled as metal, paper, or plastic. The lower image count of cardboard compared to other labels suggests that increasing the number of samples could help the model learn better representations and prevent misclassifications.

(a)   (b)   (c)
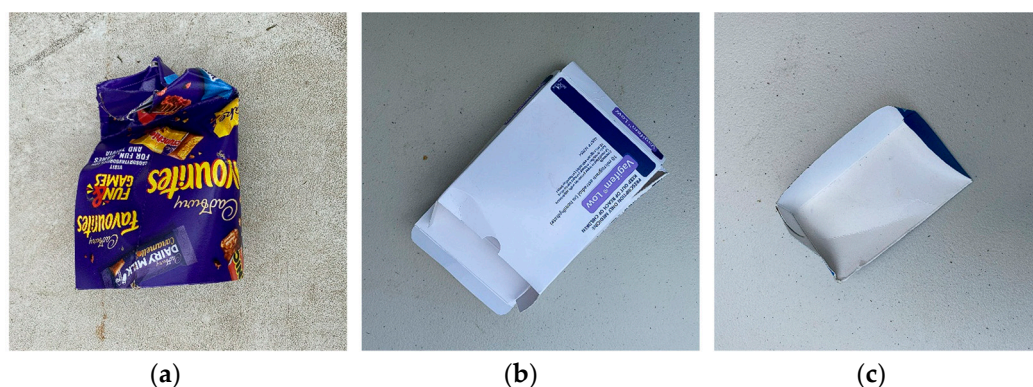
Figure 10. Confusion between cardboard, metal, paper, and plastic: (**a**) Cardboard labelled as metal. (**b**) Cardboard labelled as paper. (**c**) Cardboard labelled as plastic.

Similar confusion was observed between the metal and plastic labels. The model correctly classified 92.41% of the total metal objects but mislabelled four objects as plastic. The plastic label itself performed slightly poorer, accurately classifying 84.78% of the actual objects in its category. Six items were mislabelled as metal, and one item was mislabelled as cardboard. The similarities between certain objects in the plastic, metal, and cardboard waste categories pose challenges for the model, indicating the need for better differentiation between these materials.

Interestingly, the paper label demonstrated the best performance overall, with an accuracy of 98.18% when correctly identifying paper objects. Only a single image was mislabelled as cardboard. This superior performance may be attributed to the distinctive structural properties of paper, which can become more deformed compared to other materials, providing distinct features for differentiation.

The plastic class encountered confusion with the glass label, where five plastic objects were mislabelled as glass. This confusion arises from the fact that plastic and glass containers share a similar feature space, particularly with transparent objects, as shown in Figure 11a,b. The glass label achieved 90.48% accuracy, with two false negatives occurring with the metal and plastic classes. The misclassifications with the metal class are interesting as both false negative objects contain a small amount of metal material, causing confusion due to their transparency, as seen in Figure 12a,b. This transparency issue poses a significant challenge for practical applications.

The final inorganic material type, textile trash, achieved an accuracy of 93.75%. Although the model performed well, two false negatives occurred with the miscellaneous trash label. The decision to separate the textile and miscellaneous trash labels was motivated by the uniqueness of the textile feature space compared to miscellaneous trash. However, the limited number of images in the textile trash class may contribute to the false negatives. Increasing the number of samples could improve the model's differentiation between these labels.
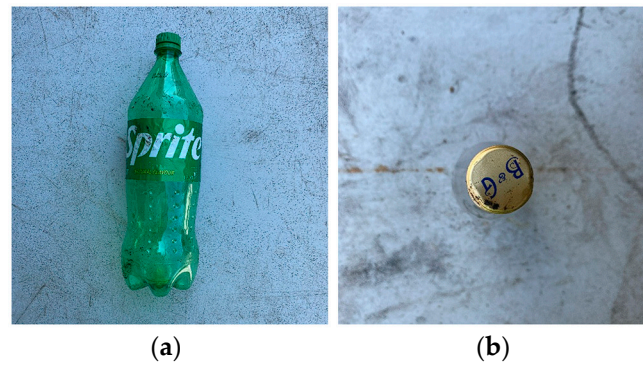
Figure 11. Confusion between plastic and glass bottles: (**a**) Plastic labelled as glass. (**b**) Glass labelled as plastic.
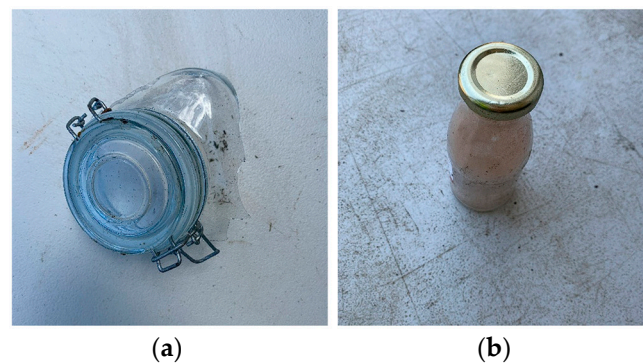
Figure 12. Confusion between glass and metal: (**a**) Metal wiring. (**b**) Metal-appearing bottle top.

Turning to the organic waste types, the food organics label performed second best overall, with 97.56% accuracy. Only a single false negative occurred, where a food organics item was mislabelled as miscellaneous trash, as shown in Figure 13a. The diverse feature space of miscellaneous trash may lead to false negatives for objects at the fringes of the food organics category. Similarly, the vegetation label, which shares similarities with food organics in terms of leafy green objects, achieved 95.45% accuracy. A single false negative occurred with the miscellaneous trash label, as shown in Figure 13b, again indicating the challenge of distinguishing objects at the edges of different waste categories. Figure 13c also shows a false negative, where a vegetation object was mislabelled as food.
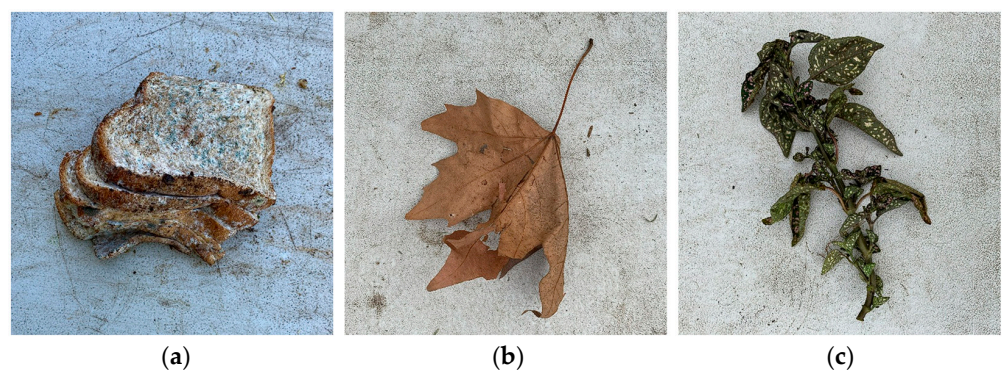
Figure 13. Confusion between organic labels: (**a**) Food organic labelled as miscellaneous trash. (**b**) Vegetation labelled as miscellaneous trash. (**c**) Vegetation labelled as food organic.

Interestingly, no other false negatives or positives were encountered between food organics and vegetation objects with any of the other labels apart from miscellaneous trash. This indicates that the model can handle the compositing issue of organic waste materials when provided with sufficient feature space to learn from.

Outside of Inception V3, most models showed a similar trend for classifying organic material types. InceptionResNet V2 encountered two false positives labelling food organics as vegetation, whilst MobileNet V2 had one. A notable exception to this was DenseNet121 that did not encounter any confusion between the two. All models displayed the same trend between glass and plastic, with MobileNet V2 proving the worst by labelling six plastic objects as glass. Similar signs were shown between cardboard, metal, paper, and plastic with all models performing slightly worse than Inception V3. The largest deviation in performance from Inception V3 was the false positives arising when labelling plastics as miscellaneous trash, with the other models encountering five to six errors compared to two in Inception V3.

In summary, the analysis of the confusion matrix for the Inception V3 model trained on the RealWaste dataset reveals important insights into the strengths and weaknesses of waste classification. It highlights the challenges of correctly identifying miscellaneous trash objects due to their diverse feature space, the potential confusion between different waste materials such as cardboard, plastic, metal, and glass, and the superior performance of the model in distinguishing paper and vegetation objects. These findings underscore the need for refined labelling, increased sample sizes, and feature space differentiation to enhance the accuracy and practicality of waste classification models.

5. Conclusions

In this paper, we have shown that machine learning approaches based on CNN deep learning models provide a practical alternative to traditional manual methods for waste classification and are therefore suited towards alleviating the subjectivity, scalability, and labour requirement issues of the latter. The analysis of the VGG-16, InceptionResNet V2, DenseNet121, Inception V3, and MobileNetV2 models has proven the ability of deep learning to classify waste across the full spectrum of labels required for accurate modelling with a classification accuracy reaching 89.19% on the Inception V3 architecture. However, these outcomes have revealed the requirement of sampling training data from the real-life environment, exposing the unsuitability of models relying on waste materials represented in their pristine and unadulterated states with classification accuracy being at best limited to 49.69%.

The potential for the Inception V3 within waste classification has been proven as significant, based off the outcomes of training on RealWaste. However, it has been shown to encounter difficulty in differentiating between cardboard, paper, plastic, and metal waste materials. Although the confusion is understandable given the labelling appearing on these material types, this places a limitation on the accuracy of landfill modelling projections. As all other models showed the same trend to a worse degree, this appears more of an issue of complexity within the waste feature space than the model's classification ability itself. Therefore, future work is suggested to consider expanding images within RealWaste to account for unbalanced labels and explore the implications of further label refinement within the dataset.

Author Contributions: S.S. contributed to programming, experiments, analysis, and writing. S.I. provided analyses and writing. R.R. conceptualized and supervised the project and reviewed. All authors have read and agreed to the published version of the manuscript.

References

1. Hoornweg, D.; Bhada-Tata, P. What a waste: A global review of solid waste management. *Urban Dev. Ser. Knowl. Pap.* **2012**, *15*, 87–88.
2. Öman, C.B.; Junestedt, C. Chemical characterization of landfill leachates—400 parameters and compounds. *Waste Manag.* **2008**, *28*, 1876–1891. [CrossRef]
3. Themelis, N.J.; Ulloa, P.A. Methane generation in landfills. *Renew. Energy* **2006**, *32*, 1243–1257. [CrossRef]
4. Goverment, A. National Greenhouse and Energy Reporting Regulations 2008. 2008. Available online: https://www.legislation.gov.au/Details/F2020C00673 (accessed on 1 January 2020).
5. Climate Watch Historical GHG Emissions (1990–2020); World Resources Institute: Washington, DC, USA. Available online: https://www.climatewatchdata.org/ghg-emissions (accessed on 1 January 2020).
6. United Nations Environment Programme and Climate and Clean Air Coalition. *Global Methane Assessment: Benefits and Costs of Mitigating Methane Emissions*; United Nations Environment Programme: Nairobi, Kenya, 2021.
7. El-Fadel, M.; Bou-Zeid, E.; Chahine, W.; Alayli, B. Temporal variation of leachate quality from pre-sorted and baled municipal solid waste with high organic and moisture content. *Waste Manag.* **2002**, *22*, 269–282. [CrossRef] [PubMed]
8. Karanjekar, R.V.; Bhatt, A.; Altouqui, S.; Jangikhatoonabad, N.; Durai, V.; Sattler, M.L.; Hossain, M.S.; Chen, V. Estimating methane emissions from landfills based on rainfall, ambient temperature, and waste composition: The CLEEN model. *Waste Manag.* **2015**, *46*, 389–398. [CrossRef] [PubMed]
9. Cruz, F.B.D.L.; Barlaz, M.A. Estimation of Waste Component-Specific Landfill Decay Rates Using Laboratory-Scale Decomposition Data. *Environ. Sci. Technol.* **2016**, *44*, 4722–4728. [CrossRef] [PubMed]
10. Lee, U.; Han, J.; Wang, M. Evaluation of landfill gas emissions from municipal solid waste landfills for the life-cycle analysis of waste-to-energy pathways. *J. Clean. Prod.* **2017**, *166*, 335–342. [CrossRef]
11. Kamran, A.; Chaudhry, M.N.; Batool, S.A. Effects of socio-economic status and seasonal variation on municipal solid waste composition: A baseline study for future planning and development. *Environ. Sci. Eur.* **2015**, *27*, 16. [CrossRef]
12. Khan, F.; Ahmed, W.; Najmi, A. Understanding consumers' behavior intentions towards dealing with the plastic waste: Perspective of a developing country. *Resour. Conserv. Recycl.* **2019**, *142*, 49–58. [CrossRef]
13. Parizeau, K.; von Massow, M.; Martin, R. Household-level dynamics of food waste production and related beliefs, attitudes, and behaviours in Guelph, Ontario. *Waste Manag.* **2015**, *35*, 207–217. [CrossRef]
14. Melinte, D.O.; Travediu, A.-M.; Dumitriu, D.N. Deep Convolutional Neural Networks Object Detector for Real-Time Waste Identification. *Appl. Sci.* **2020**, *10*, 7301. [CrossRef]
15. Zhang, Q.; Zhang, X.; Mu, X.; Wang, Z.; Tian, R.; Wang, X.; Liu, X. Recyclable waste image recognition based on deep learning. *Resour. Conserv. Recycl.* **2021**, *171*, 105636. [CrossRef]
16. Mao, W.-L.; Chen, W.-C.; Wang, C.-T.; Lin, Y.-H. Recycling waste classification using optimized convolutional neural network. *Resour. Conserv. Recycl.* **2021**, *164*, 105132. [CrossRef]
17. Aral, R.A.; Keskin, Ş.R.; Kaya, M.; Hacıömeroğlu, M. Classification of TrashNet Dataset Based on Deep Learning Models. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2058–2062. [CrossRef]
18. Liang, S.; Gu, Y. A deep convolutional neural network to simultaneously localize and recognize waste types in images. *Waste Manag.* **2021**, *126*, 247–257. [CrossRef] [PubMed]
19. Funch, O.I.; Marhaug, R.; Kohtala, S.; Steinert, M. Detecting glass and metal in consumer trash bags during waste collection using convolutional neural networks. *Waste Manag.* **2021**, *119*, 30–38. [CrossRef] [PubMed]
20. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
21. He, X.; Chen, Y.; Ghamisi, P. Heterogeneous Transfer Learning for Hyperspectral Image Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3246–3263. [CrossRef]
22. Basha, S.H.S.; Dubey, S.R.; Pulabaigari, V.; Mukherjee, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing* **2020**, *378*, 112–119. [CrossRef]
23. Sharma, N.; Jain, V.; Mishra, A. An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Comput. Sci.* **2018**, *132*, 377–384. [CrossRef]
24. Shijie, J.; Ping, W.; Peiyi, J.; Siping, H. Research on data augmentation for image classification based on convolution neural networks. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 4165–4170. [CrossRef]
25. Zhang, C.; Zhou, P.; Li, C.; Liu, L. A Convolutional Neural Network for Leaves Recognition Using Data Augmentation. In Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, UK, 26–28 October 2015; pp. 2143–2150. [CrossRef]
26. de la Rosa, F.L.; Gómez-Sirvent, J.L.; Sánchez-Reolid, R.; Morales, R.; Fernández-Caballero, A. Geometric transformation-based data augmentation on defect classification of segmented images of semiconductor materials using a ResNet50 convolutional neural network. *Expert Syst. Appl.* **2022**, *206*, 117731. [CrossRef]

27. Sabottke, C.F.; Spieler, B.M. The Effect of Image Resolution on Deep Learning in Radiography. *Radiol. Artif. Intell.* **2020**, *2*, e190015. [CrossRef] [PubMed]

28. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [CrossRef]

29. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]

30. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 12 June 2015; pp. 1–9. [CrossRef]

31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385. [CrossRef]

32. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]

33. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 12 February 2017.

34. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [CrossRef]

35. Yang, M.; Thung, G. TrashNet, Github. 2016. Available online: https://github.com/garythung/trashnet (accessed on 20 November 2023).

36. Yu, C.; Wang, F.; Shao, Z.; Sun, T.; Wu, L.; Xu, Y. DSformer: A Double Sampling Transformer for Multivariate Time Series Long-term Prediction. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 21–25 October 2023. [CrossRef]

37. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.

38. Dong, Z.; Chen, J.; Lu, W. Computer vision to recognize construction waste compositions: A novel boundary-aware transformer (BAT) model. *J. Environ. Manag.* **2022**, *305*, 114405. [CrossRef]

39. Single, S.; Iranmanesh, S.; Raad, R. RealWaste, Electronic Dataset, Wollongong City Council, CC BY-NC-SA 4.0. Available online: https://github.com/sam-single/realwaste (accessed on 20 November 2023).