A Comparative Analysis for Trash image classification using Deep Learning

MSc Research Project
Data Analytics

Prithvi Mysore Dayananda
Student ID: x19242204

School of Computing
National College of Ireland

Supervisor: Dr. Christian Horn

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Prithvi Mysore Dayananda
Student ID:	x19242204
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Christian Horn
Submission Due Date:	30/01/2022
Project Title:	A Comparative Analysis for Trash image classification using Deep Learning
Word Count:	6273
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	30th January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A Comparative Analysis for Trash image classification using Deep Learning

Prithvi Mysore Dayananda
x19242204

Abstract

Garbage classification is the most essential tool in the given era. Over 3.40 billion tonnes of waste is predicted to be produced by the year 2050. At this rate the toll garbage production has on the environment, health, marine, and wildlife are enormous. In this fast-moving technological era, we civilians must follow the three R's of waste management, recycle, reuse and reduce. Climate change for one is a great example that we take precautions and try to build a better future. In recycling waste, the most crucial part is of segregating the right material. Since a mixture of materials is of no use. Currently, there are sensors used to divide materials but the machines fail to categorize all of them which then leads to manual handpicking of trash. To solve this we have implemented multiple convolutional neural networks such as VGG16, ResNet50, and a custom model MLH-CNN. A comparative study has been done on all of the models mentioned. Poor or random initialization of parameters may lead to longer training or vanishing gradient problems. Hence in this research, many experiments have been implemented and the results have been compared. The main objective of this research is to compare the performance of the light-weighted custom MLH-CNN model with VGG16 and ResNet. Since the MLH-CNN model was trained on TrashNet data with only 2575 images only the simple structure worked well, the dataset used in this research is 15728 images and by multiple experiments, we see that the model's accuracy has been depreciated and ResNet50 outperforms Custom MLH-CNN and VGG16 model with an accuracy of 82.20% with an overall precision and recall of 85.30% and 80.36% respectively.

1 Introduction

Over 2.01 billion tons of waste has been reported by the municipal solid waste department in the year 2021 all over the world. It is expected to be 3.40 billion tons of waste production by the year 2050. When numbers are dug in we see that the income level of the country is directly correlated to waste production. In the US about 75% of waste is recyclable and only 34 percent are recycled in real. This can be improved by educating people about the importance and the effects on the environment because of waste production. It is seen that most recyclable materials tend to get mixed up with the non-recyclable things which complicate the process of waste management. There are several ways that wastes can be divided such as their properties, their effects on the environment, and so on. [1]It is important that we the civilians must follow the three R's of waste management.

[1]`https://4waste.com.au/rubbish-removal/5-types-waste-know/`

We must reduce shifting for new boxes, plastic bags rather than use the sustainable ones which help us produce less waste. Rather than buying new things every time we shop, we must learn how to reuse the things we have by donating, making something creative out of it. [2] Plastic, Glass, Paper, and many more must be recycled so that we stop the environmental impact. In this paper, we are going to see how the sorting process can be speeded up by using deep learning methods. Once the recyclable materials are collected they are then sorted on their size, color, and type.
Plastics in particular are of many categories, the article Dr. Anne Meyer (n.d.) provides us a detailed overview of them. Non-recyclable materials are then used in energy Recovery which will be utilized for producing electricity, useable heat, and so on. In the process of segregating the recyclable materials, we can use deep learning's image classification which helps fasten the process. Even though the machines use censors to classify them there are further steps that involve manual segregation. This slower the process and deteriorates the process. By using CNN powerful algorithms we can classify waste into recyclable and non-recyclable parts either in the initial garbage collection point or in the waste management factories as well. In this paper, we can see that pre-trained image classification CNN models such as VGG16, and ResNet50 have been implemented and evaluated for quicker classification.

1.1 Background and Motivation

The process of waste management involves collecting the waste from various sources such as domestic, medical, industrial, irrigation and Mining. The collected trash is sectioned based on recyclable and non-recyclable ones. The recyclable ones are further then divided into categories such as glass, plastic, clothes, aluminum, and so on. These materials once when collected in bulk are compressed into 1000 - 1500 pounds of blocks called a bale. They are then sold to other companies to manufacture new pieces by further going through the process of cleaning and heating. In this process, some materials get mixed with recyclable ones but aren't recyclable such as thin plastic bags, plastic boxes with food on it. Since they end up in the process the cost spend on them is of no use and is thrown away. These are either converted into some form of energy or are dug inside the land. Figure 1

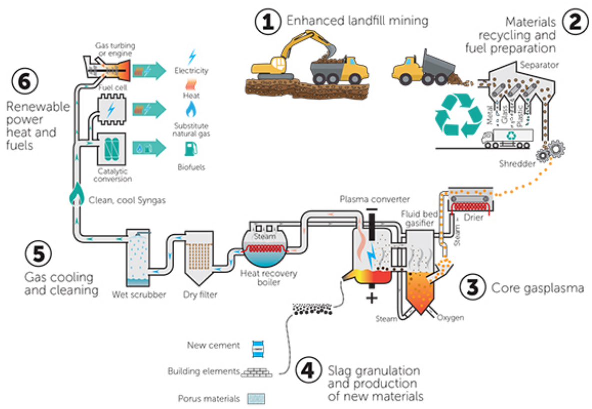

Figure 1: Waste Management on Landfill process

[2] https://roguedisposal.com/resources/education/recycling/exploring-the-three-rs-of-waste-manageme

gives us a overview of the process. The trash decomposed inside land is dangerous since they emit hazardous chemicals methane gas out to the public, hence there are pipes installed with filters that remove the hazardous gases before letting it out to the public. There are sensors and strong magnets used in the real world to classify these objects. This raises the motivation to introduce strong CNN models which help in classifying images and help to fasten the process of segregating different categories of waste.

1.2 Research Question

"To what extent Deep Learning models help in segregating trash and prediction which helps in waste management?"

This research aims in solving the above mentioned question by performing classification on openly available dataset. We can see that in order to achieve good accuracy on classification, many pretrained models such as VGG16, ResNet50 and a custom model MLH-CNN inspired by Shi et al. (2021) have been implemented and tested using appropriate metrics.

1.3 Research Objectives

Table 1: Overview of research objectives

Objectives	**Description**
Objective 1	A detailed investigation on trash management
Objective 2	Collection of Data, and pre-processing of images
Objective 3	Implementation of State-of-Art models
Objective 4	Comparision of models such as VGG16, ResNet50, and Inception V3 with and without image augmentation
Objective 5	Evaluation of models using accuracy, validation accuracy, and confusion matrix
Objective 6	Representing the obtained result in better visualization tools

The paper provides information by this following, Section 2 elaborates more on the most recent work done on waste management classification. Section 3 provides details on the methodology followed in this research, Section 4 gives us the design flow taken in this research, Section 5 is where the models mentioned earlier are implemented, Section 6 explains the performance of each model and in Section 7 we have provided conclusion for this research.

2 Related Work

2.1 Deep Learning for Waste Classification

Deep Learning has been utilised in solving many real world problems in the same way as humans from natural language to speech recognition we can see the varities of application of deep learning in healthcare, autonomous driving, military, wildlife and so on. Using CNN models there are many experiments implemetned in waste classfication. A fully deesinged hardware smart trashcan has been designed from (Kang et al.; 2020)which consists of a solar panel , battery, raspberry pi for the computing part and a camera for classifying objects. Cifar-10 dataset has been used to test all the state of art algorithms for image classification. It is observed that Resnet and Inception works well after performing 200 steps of iterations. Considering the calculation speed Resnet-34 has been used as base model. Multiple feature extraction layers has been introduced so to not lose the integrity of the images captured. A dataset of 4168 iamges has been collected manually and to avoid the large eigen value production in Relu activation function, a combination of softplus and softsign has been implemented to speed up the convergence. The model has performed exceptionally well with an accuracy of 99%
A proposal (Guo et al.; 2021) where EfficientNet has been used to extract features and performed attention mechanism like squeeze, Excitation and Attention to reduce computational time. We know that Batch Normalization is used to normalize data to a statistical form with a mean of 0 and variance of 1. Since it works well on huge batch size and not on the smaller ones. Here in this research instead of regular batch Normalization to avoid gradient descent the authors have tested out group normalization. The model is trained to predict on 40 kinds of garbage, Effiecient net has an accuracy of 93.47 to 98.3 after 30 epochs

Squeeze net is light weight model which uses only 1/50 of the parameters of Alexnet, which basically is made up of fire modules. Around 4373 images collected have been tested out on Resnet 50 which produces 93.6%, Inception V3 of an accuracy 96%. Squeezenet on the other hand produces 83.5% of accuracy. Though Resnet and Inception provides 90 and above accuracy, (Wentao et al.; 2020) the author has went ahead with squeeze net because of its simple structure and robustness. The model has been further improved by Data enhancement and Tansfer Learning the model boost upto an accuracy of 87.7. The idea of having a light weight model further led to discovery of the paper Shi et al. (2021). The main objective of the paper was to produce better accuracy on simple architure model than a dense model such as VGG16, ResNet50, Inception V3. The authors have used the dataset collected by Stanford student called TrashNet which is widely experimented and we will further see a few papers on it. A custom model called MLH-CNN has been implemented with 4 block of convolutional layers varying in units size. The model appears to outperform all dense pre trained model with an accuracy of 92.60%. Since the size of the dataet is quite small here in this research paper a comparision analysis will be performed to evaluate the simple model with pre trained models.

Using python crawler the author has gathered images of over 30 categories with 6000 images in total. (Cao and Xiang; 2020) By data augmentation the authors have produced over 20000 images. Transfer Learning for 30 categories has been carried out holding Inception V3 as a base model. By experimenting with multiple learning rates the author have finalized on a parameter of 0.2227 which provides an accuracy of 92.5%.

With an increase in population we can see that the Municipal solid waste has rose drastically since the decade. (Ziouzios et al.; 2020) The authors have proposed a model which classifies garbage in real time by the help of cloud server with a hardware installation of raspberry pi 3B+ model to capture images. The cloud server hosts a neural network model which helps in classification of five categories, i.e; Paper, glass, plastic, metal, carton and garbage. All server requests are encrypted to maintain safety. The authors have also created a platform using PHP and MySQL for supervisors to monitor and improvise models to improve their efficiency. Trashnet data has been implemented on MobileNet model published by Google for classification. After running the model for over 20 epochs the authors have gained an accuracy level of 96.57% and confusion matrix has also been taken into consideration as a part of evaluation metric.

A combination of 2 datasets from TACO containing 1403 images and Kaggle's images of 2500 have been combined together. Here in finding the researchers (Patil et al.; 2021) have used the combined dataset and have compare three different models. A simple CNN model which gives an accuracy of 96.73% of accuracy on Dump classification with a loss of 11.25. The same model has been used for Trash classification which has an accuracy of 90.68% with a higher loss of 29.77. We see that the authors have compared the model with state-of-art-models VGG16 and ResNet50 as well. 94.29% with a loss of 11.82 was observed in Dump classification while on Trash classification the model had 93.46% of accuracy and 21.49 % of loss. We can see that higher the model's accuracy is, higher the loss rate is. Finally ResNet50 has been applied for Dump classification only since, the model did not make a worth full performance on Trash classification. The accuracy of the model which is 88.59% with a loss of 19.92% is observed to be low when compared to other models. Object detection is localizing the objects position and detecting multiple objects. This has been implemented in the paper by using self-collected images from the official Beijing website. The authors have implemented Mask RCNN and Mask Scoring RCNN to overcome the accuracy issues by using MaskIOU head module. Using Mask scoring algorithm as base the model has achieved 65.9 percentage. As the complexity of the image provided increases the model's prediction level decreases drastically. Li et al. (2020) This can be improved by providing more data to the model since the model has been trained only on 5000 images.

2.2 Survey on existing solutions

A Research was conducted by Teampanpong (2021) in Thailand where the authors ran a real time tracking for endagered animals. Kaeng Krachan National Park in Thailand has been used to test out the effect of dumspters on animals. It was observed around 30 differnt animals were seen on the cameras installed in the dumpsters. Since the wild animals are fed on trash,a nd the trash was observed to be correlated to the number of tourist. This raises another essential nees for sorting out trash so that the animals do not feed on waste materails. Another general seervey was taken in Mexico by Velazquez et al. (2020) where the authors personally managed the houshold Trash of over 58 houses and sotred out manually, it was seen that most of the waste could be didvide as Organic, Inorganic and domestics waste. A large percentage of waste has been seen from food-waste. The rest from Plastic, Glass, carboard and so on. The idea that the suthors have mentioned is that sorting is a huge task and since the trash was collected from rural

places, it is necesseary for the public to understand the effect of waste on climate and the hazardous reactions it has on the planet. Hence there is a need that schools must have a healthy discussion on these topics before it's too late.
Specially designed for identifying plastic waste Bobulski and Kubanek (2021). using a RGB camera and a micro computer 120*120 and 227*227 input image size has been experimented with base model opted for AlexNet which recognizes PET, PE-HD, PS, and PP types of plastics. The data has been obtained from WaDaBa database. The model provides an accuracy of 97.43%. Two types of custom models i.e., 15 layer works well on 120*120 pixels and 23 layered architecute works well on 227*227 input size. Sun and Xiao (2020)889 images of 5 classes iamges were collected manually by the student of usinveristy of beijign. Cameras have been set up across all the 11 trash collection point on the college campus and the iamges captured were trianed simulatneoulsy on real time. Each trash bin is identofied by unique annoted binary vector. The state-of-art models sich as VGG16, ResNet50 and DenseNet169 are applied using cross validation methods and is seen that ResNet50 outperforms the rest of the model by 92.40%. But the research uses a very slow learning rate of 0.00001. This cause overload and slowers the process of training.
Anh H. et al. (2019) proposes a custom model called DNN-TC inspired by the work on Xie et al. (2016) where the input shape of images are of 224*224 Where the last layer of softmax uses a log version to produce the probabilities which helps in gradient optimization. The model was tested out of TrashNet with a combination VN dataset with domestic waste. With a learning rate of 0.0001 the model performs an accuracy of 98% when ran for 100 epochs. Since the learning rate is less, the time taken to train the model will be high. To avoid this our research explains various experiments.

3 Methodology

This research follows the structure as seen in the Figure 7. All the research papers discussed in the 2 are ought to follow the same methodology too. The data is collected and is understood deeply. Once the data is well understood, it is then pre-processed, augmented which includes performing various methods on images, various models are applied, and using appropriate metrics these models are evaluated and appropriate epochs are learned for each model after performing multiple training sets. We are not going to deploy this model.

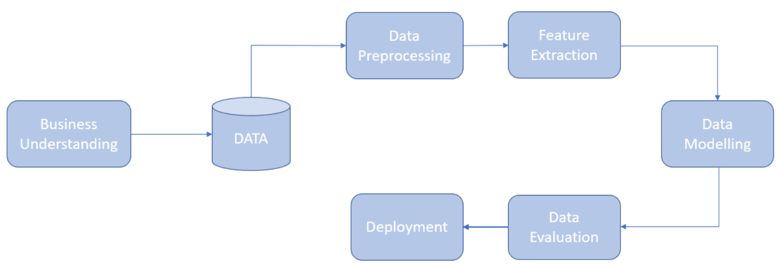

Figure 2: Modified KDD Methodology

3.1 Business Understanding

There is an indispensable need in the world to manage the waste produced. Mainly to reduce, recycle and reuse things rather than disposing of them. As we saw in 1 we can minimize by making small changes in our daily life routine. When segregating the collected waste then comes the picture of us introducing deep learning methods using image classification for better, strong, and quick division. This is the motivation for the research and to improve the sorting process of waste. The following section will provide details on how and where the data was experimented in this research.

3.2 Data Acquisition and Understanding

The data used in this research has been downloaded by a publically available data website called Kaggle [3]. The dataset holds around 15515 images belonging to various categories of waste which can be recycled. There are 12 different categories of waste which will help us cover a few issues raised in 1 where underfitting has been seen as a problem because of limited data. Figure 7 gives us the count of images in each 12 of the categories. We can see that the number of images in clothes is more than the rest of the categories. This might result in the overfitting of any model applied. This can be solved by image augmentation by producing many replicates of the same images by adding a few different features.

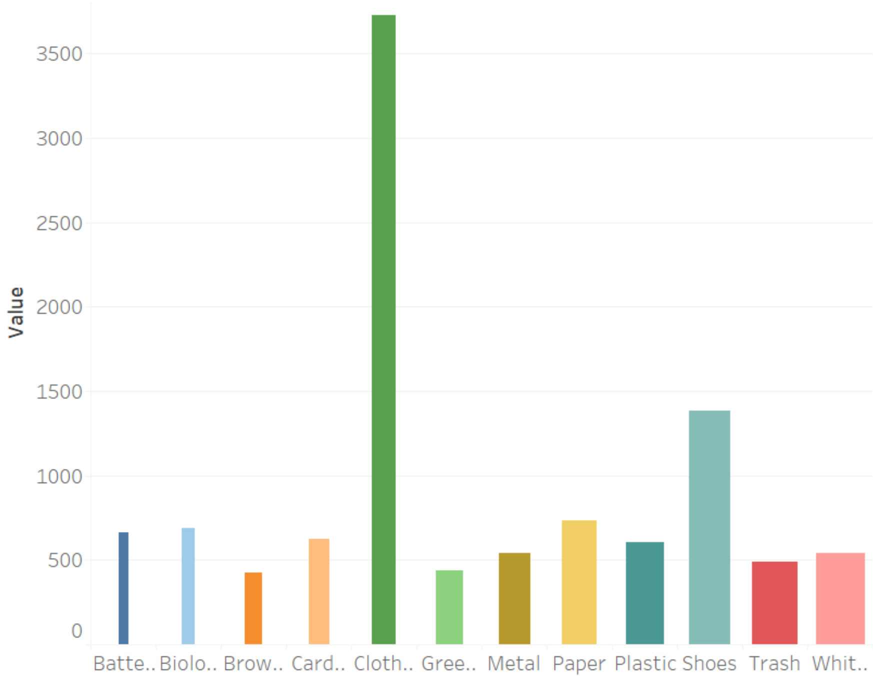

Figure 3: Image count of each category

[3]Kaggle Garbage Dataset: `https://www.kaggle.com/mostafaabla/garbage-classification`

3.3 Data Preparation and Processing

The data when downloaded is not split into train, validation and test folders to fit our model. Using python's library splitfolders, the data here has been divided in the ratio 7:2:1 for training, validation and testing respectively. we can see a few images in the dataset from the Figure 4 The images are scaled to values between 0 and 1 to make the

Figure 4: Image count of each category

model understand better. The images are reshaped to a size of 64*64 in RGB channels. By using python's ImageDataGenerator package the labels have been converted into one hot encoding which helps the model to understand the category representation of wastes since the model cannot understand the default labels. Before fitting the model, all the images are shuffled so that the model trains on different and random patterns.

3.3.1 Splitting of data

Around 10854 images are used to train all our models with 3100 images as the validation set. It is good practice to test our models on validation tests since we can figure out the best model using the validation accuracy. Knowing the best fit model we can further test our model with unseen data, in this scenario, we have around 1561 images as the test set.

3.3.2 Data Augmentation

It is usually a good idea to use python's image library "Image Generator" since it reduces a lot of manual work and memory. During the training of the model, each epoch will receive and be trained on variants of images that undergo many preprocessing functions such as rotation, brightness, filliping, standardization which generally avoids a model overfitting and prepares the model for unseen data. In this research, we have experimented with all the above features with different values and tested models with augmentation and without. Rotation range of 40, which aids the model in recognizing the same images in different angles. Wide shift range and height shift range have been implemented with a range of 0.2. Since the model can behave poorly on the same images, zooming in and out has been applied with a horizontal shifting set as true.

4 Design Specification and Solution Development

A Design workflow of this research can be explained in the following three steps as seen in Figure 5. It mainly consists of Data Preparation, Model implementation, and interpreting the results with graphs and other visualization tools for better impact.

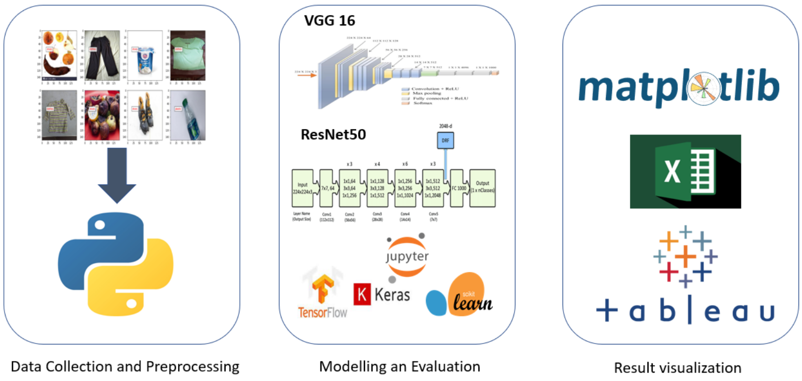

Figure 5: Garbage classification- Design Flow

- Data Preparation stage involves all the processes implied on data from data gathering, extraction, feature selection, splitting, augmenting, and enriching the existed data. The data used has been exported from the website Kaggle platform. The size of the original dataset is 238 MB. The data when downloaded is not split into testing and training. This has been perfromed using python library "splitfolders". It is important the labels of data are one hot encoded that the computer recognizes.

- The next stage involves modeling the clean data, here in this research we have applied a custom model called MLH-CNN from Shi et al. (2021). Multiple convolutional neural models have been implemented such as VGG16 and ResNet 50. This stage involves parameter tuning, model optimization, and evaluation using suitable parameters such as accuracy, precision, and recall.

- Hence, we can now step into the final stage where we interpret the above calculation results into meaningful output presenting it in graphs using visualization tools.

5 Implementation

5.1 Custom MLH-CNN model:

A model from Shi et al. (2021) has been implemented as the accuracy score of the model on the TrashNet dataset is higher than any other pretrained model. The model is simple enough than the dense models as the pre-trained ones. The Figure 6 Shi et al. (2021)

gives us the structure of the model implemented. The model takes in an input size of 64*64. BatchNormalization and Convolution layer are mainly used in Feature Extraction from the images. Since the BN helps in faster convergence, Versloot (2020) it is really useful in building the model.

To help in regularization and gradient descent issues Batch Standardization has been

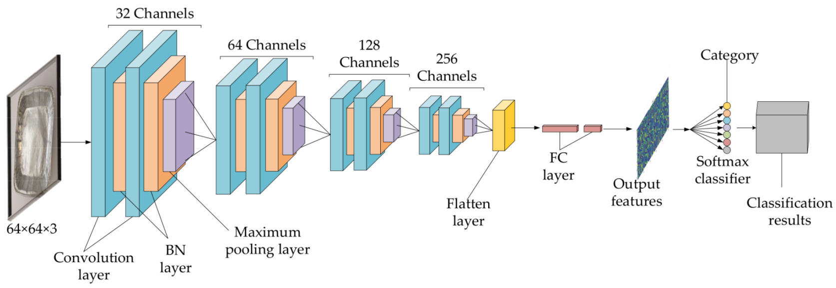

Figure 6: Architecture of custom MLH-CNN model

placed for each feature map. Batch Normalization for feature map of size a*b will be equivalent to normalizing the feature map of size m=—β —=mab. It is known that in VGG16 architecture, two 3*3 convolution layers have the same perceptual field of view as one 5*5 convolutional layer. The authors have therefore considered 3*3 convolutional layers since it reduces the number of trainable parameters. The structure has 4 main blocks of convolutional layers built with a mix of BatchNormalization and Maxpooling Layers. The first block consists of two convolutional layers with 32 units of neurons and a kernel size of 3*3 has been initialized. A batch normalization layer has been built for every convolutional layer to extract features from the respective input. The same block has been implemented for various numbers of neurons as followed, 64, 128, and 256. Every block has been passed through a Maxpooling layer just to extract the small edges and features. The obtained output from the 256 neuron block is then passed through a Flattened Layer and then through two Fully connected layers of size 128 and 64 neurons. The activation function for all the layers has been initialized to the Relu activation function. The model is then completed by adding a softmax layer of the required number of categories since this is a multi-class prediction model.

5.2 VGG16

Karen Simonyan and Andrew Zisserman from Oxford University in the year 2014 proposed a very deep convolutional network for a large-scale image recognition model called VGG16. This was found a solution for the annual image classification competition held by "ImageNet Large Scale Visual Recognition Challenge (ILSVRC)" which has over 14 million images with 1000 categories. The images of Imagenet are of the standard size of 224*224 with RGB channels. The input shape provided to this model must be a size of (224*224*3). The output of 1000 categories consists of the probability of each class. The figure Figure 7 represents the calculation of the softmax layers. Simonyan and Zisserman (2015) Unlike Alexnet the architecture of VGG16 is more uniform and simple. To overcome the vanishing gradient problem caused by the Alexnet activation function VGG16 has the Relu activation function. All the convolutional layers have fixed parameters i.e., a filter size of 3*3, a stride of 1, and padding of the same which essentially means zero

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Figure 7: Softmax Function

padding. In Maxpooling layers, a filter size of 2*2 and a stride of 2 is set. We can see that the architecture has 2 fully connected layers with 4096 units each followed by a softmax layer of 1000.

5.3 ResNet50

ResNet 50 is the improved version of ResNet 152. As with any model the predefined input shape of this model is 224*224*3. ResNet50 is a deep neural model which is quite difficult to understand. Figure 8 gives us a detailed overview of ResNet50 architecture Matsuyama (2020). ResNet model can be implemented by using Keras pretrained version by downloading the weights, here in this research the model has been built by scratch and a few layers of choice have been added to make a better prediction. The model has been built looking into the original architecture. As VGG16, ResNet as well uses the Relu activation function to avoid gradient descent.

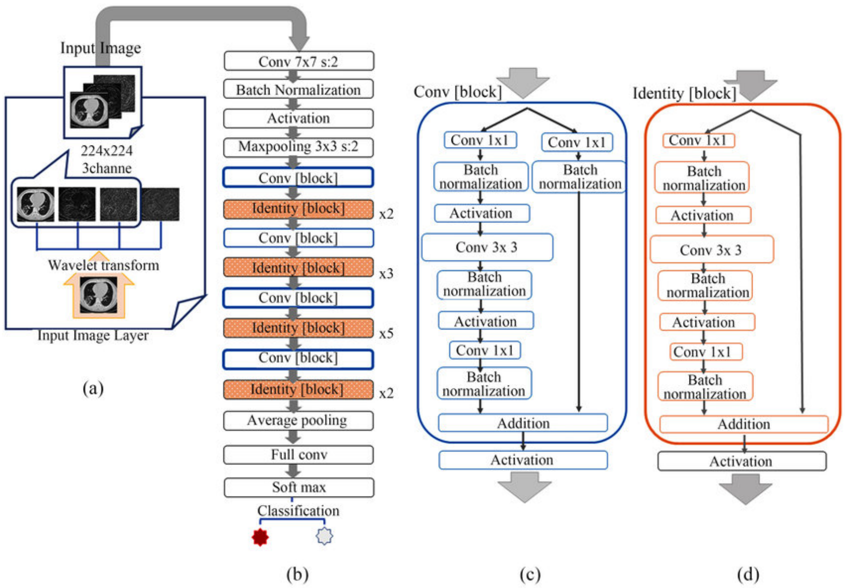

Figure 8: ResNet50 architecture

6 Evaluation and Results Analysis

To know how good a model is, it important for us to measure the model's behavior. Using appropriate methods such as Accuracy, Precision, and Recall we have evaluated the models mentioned above. Precision metric has been used from the multi-class confusion matrix built to determine how well the model has predicted the observed values. The Experiment 6.1 and 6.2 are implemented using the pre-trained models as we see in [paper something] with regularization techniques. In 6.3 we see the evaluation of custom MLH-CNN model.

6.1 Experiment with Custom MLH-CNN model

The model as we saw was proposed by Shi et al. (2021) whose main aim was to reduce the complex structure of the model to be trained on. We see that with a learning rate of 0.001 the model's training accuracy is 70.53% with a validation accuracy of 63.25%. The loss of this model was observed to be near 0.89. Although the model performed well on training, the testing accuracy of the model was observed to be 60.06%. Hence the model needed to learn the environment better.

Table 2: Result of custom MLH-CNN model

Epoch	Training Acc	Validation Acc	Loss
1	54.39%	60.45%	1.44
3	64.65%	63.83%	1.06
5	68.15%	65.14%	0.94
7	70.37%	68.75%	0.83
9	72.30%	68.03%	0.81
10	72.95%	67.38%	0.79

The following Table 2 represent the model performed on a learning rate of 0.01. The slower the learning, the better a model learns from its input. We see that the model performs better with an accuracy of 72.95% on training and an accuracy of 67.38% on the testing set. The table gives us the improvement on each epoch. Figure 11, .

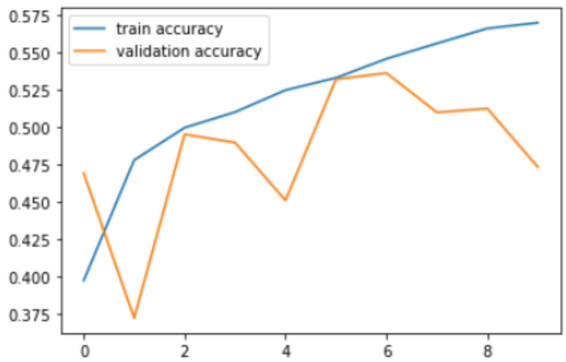

Figure 9: Accuracy of MLH-CNN model

Figure 10 and Figure 9 gives us the accuracy and confusion matrix of the final model implemented. The overall Precision of the model is about 70.89 and the overall recall is about 63.011. The ability of a classification model to recognize only the relevant data. Given in mathematical terms, precision is the number of true positives divided by the number of true positives plus the number of false positives. It was observed that the

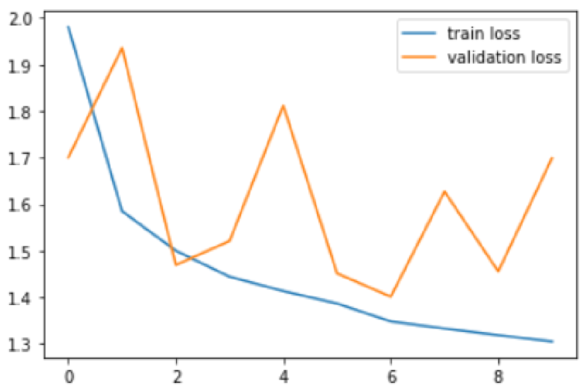

Figure 10: Loss of MLH-CNN model

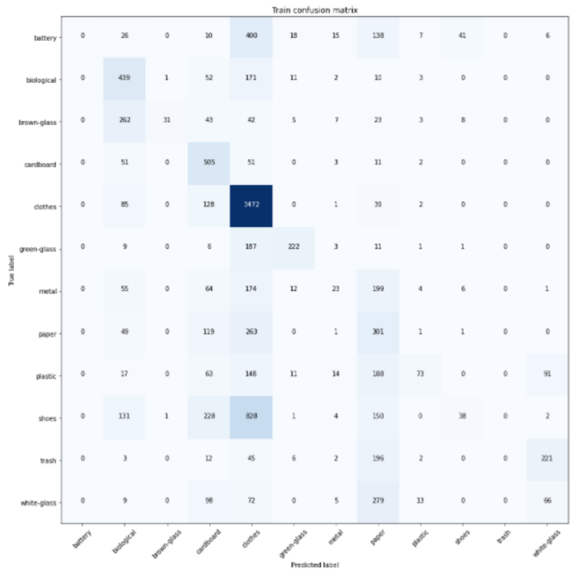

Figure 11: Confusion Matrix of MLH-CNN model

model took around 57 minutes to be trained on the training images. We can from the graph that the loss has been decreased eventually which is always a good sign, and the accuracy has been increased. Recall on other hand is the ability of a model to find all the relevant cases within a data set. We define recall as the number of true positives divided by the number of true positives plus the number of false negatives. We can say that Precision plays an important role in our study.

6.2 Experiment with ResNet-50 model

ResNet 50 is a simple version of ResNet 152 which is mainly used to skip connections which helps in learning an identity function that makes sure that the higher layer performs better than the lower layer of the model. While constructing the model these are the few steps taken into consideration.

- Constructing the ResNet50 model from scratch.

- Adding custom network layers on top of the baseline network.
- Training the model on different learning rates, batch normalization momentum, and patience level.

A batch size of 32 ran for over 10 epochs. With a learning rate of 0.0001 the model took over seven hours to train since the dataset is huge and ResNet is a heavy model. The obtained accuracy of the training dataset was 90.62% with a validation accuracy of 63% which was quite low compared to the custom model. When the model was tested on the test set the accuracy obtained was around 65%. Since the model took quite a long time to perform, a learning rate of 0.01 was used to fasten the process, the result of this parameter has been recorded in the Table 3 . along with the accuracy. From Figure 12

Table 3: Result of ResNet-50 model

Epoch	**Training Acc**	**Validation Acc**	**Loss**
1	32.15%	38.84%	0.3215
3	59.88%	31.35%	0.3135
5	66.25%	55.65%	0.5565
7	71.89%	55.84%	0.5584
9	79.04%	54.52%	0.5984
10	82.20%	55.74%	0.5574

we see that after 3 epoch there is a huge change in the accuracy of the model. Figure 12 and loss of the model.

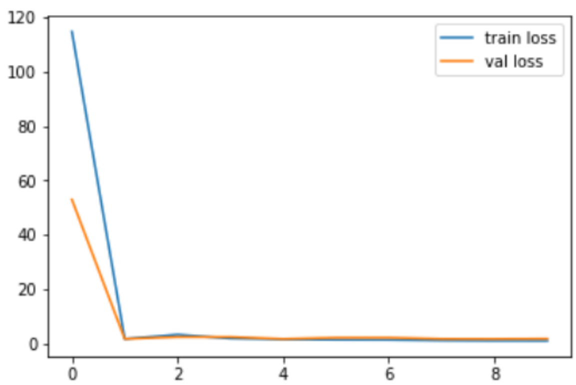

Figure 12: Loss of ResNet50 model

Figure 13 We see a drastic drop on loss from 1st epoch which is expected since the model takes time to understand the features on the first epoch. We know that calculating loss is necessary to understand the performance of models. It is important that we have less loss and more accuracy as loss represents us the error in finding out the true classes. The validation accuracy on the other hand varies quite often.

Figure 14 presents the confusion matrix of ResNet50 model where we see a high precision and recall percentage of 85.30% and 80.36%

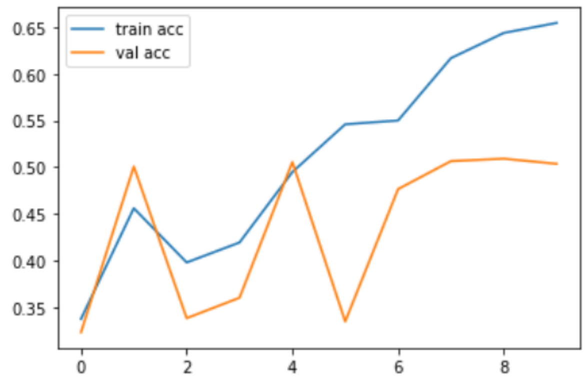

Figure 13: Accuracy of ResNet50 model

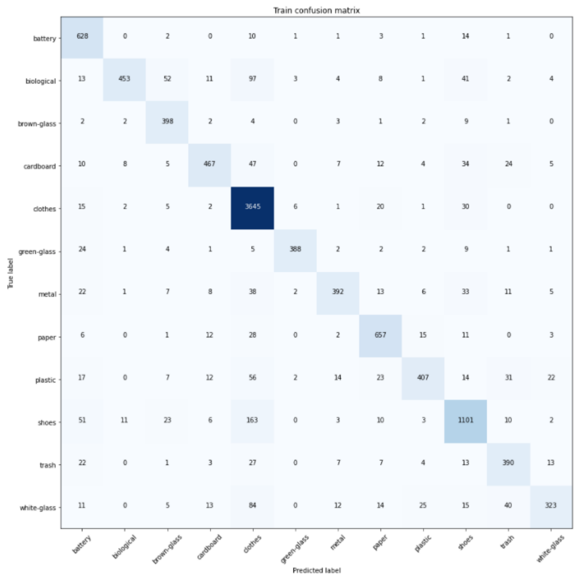

Figure 14: Confusion Matrix of ResNet50 model with data augmentation

6.3 Experiment with VGG16 model

Convolutional Networks state of art models can be utilized for a wide variety of computer vision solutions. In this research transfer learning approach has been implemented using the VGG16 pre-trained model. The model has been extracted from the Keras applications and has been modified for our convenience. The architecture has about 138 million training parameters which essentially give 92.7% of accuracy on the ImageNet dataset of 1000 categories. Here in this research, the data has 12 different categories, the model's last softmax layer has been modified and 12 output layers have been implemented for appropriate results. The model has been tested out for different parameters of learning rates and epochs. The best model has been presented in this paper. The Table 4 represents the training process of the model with parameters of the following, learning set to 0.01 and epochs of 10. Sheh (n.d.) has provided a great explaination on initialization of batch size and how learning rate and batch size effect the model training and performance. The model has a higher accuracy of 80.47% with a validation accuracy of 84.60%

Table 4: Result of transfer learning VGG16 model

Epoch	Training Acc	Validation Acc	Loss
1	65.80%	78.29%	5.99
2	73.35%	80.14%	4.96
3	76.22%	84.67%	5.32
4	78.83%	82.10%	4.76
5	79.19%	85.03%	4.83
6	79.74%	85.74%	5.06
7	80.47%	84.60%	4.97

from Figure 15 . The model has been ran for 10 epochs and we see that around 3rd epoch

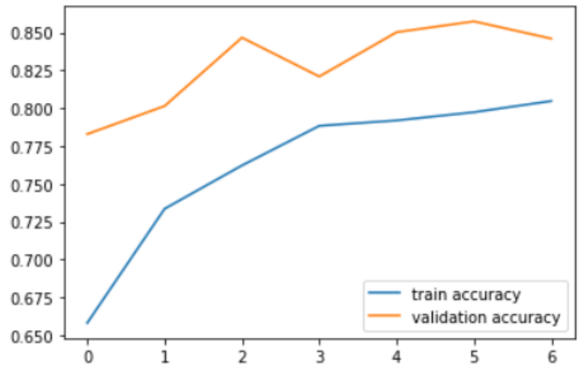

Figure 15: Accuracy of VGG16 model

the accuracy takes a curve and at 5 very slow improvement is seen. The accuracy can be increased by running the model for more number of epochs.

Figure 16 gives the training loss observed. As the accuracy is high, we can also see

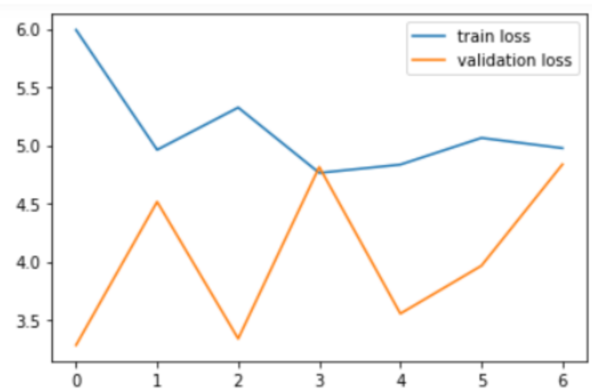

Figure 16: Loss of VGG16 model

the loss of the model is around 4.97 which is high and this represents that the model recognizes the images well enough but with few small errors.

The time taken by the model train on this huge dataset was around 5.36 hours, though it is expected the same since the model has around 138 million training parameters. When we have a look into the confusion matrix Figure 17

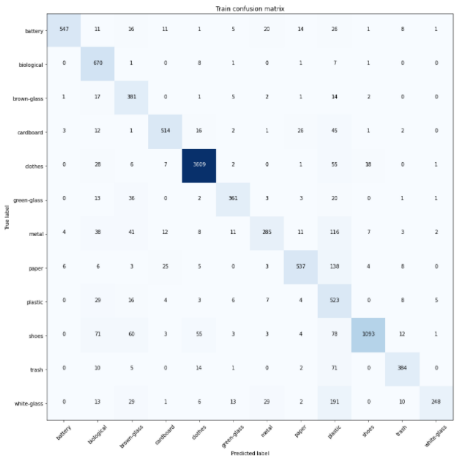

Figure 17: Confusion Matrix of VGG16 model with data augmentation

of the model we see that the category of "Plastic" has been often miscategorized in all the experimented models. The model has a precision of 83.27% and recall of 78.94%. As we consider precision in evaluating our study, we can say that a model of 83% precision is well enough.

Since the model has neither been overfitted nor under fitted, we can say that the quality of the images gathered might be the reason for the miscategorization.

7 Discussion

This study was experimental shown to perform well on classifying trash. The huge data of approximately 15,000 images of 6 categories have been trained on multiple state-of-art models and custom models built by Shi et al. (2021). The main motivation of this study was to compare the performance of the custom model MLH-CNN with state-of-art models on this novel dataset. It is observed that the model MLH-CNN outperforms the existing state-of-art models on the TrashNet dataset. Since the number of images from the TrashNet dataset is 2527, the data is rather small and the model used might not be suitable for the huge dataset. The same architecture has been applied to the data provided by the Kaggle platform along with pre-trained models such as VGG16 and ResNet50. As we saw in Section 6.1 we see that the MLH-CNN model gives an accuracy of 57.00% with validation of 47.36% when trained on a learning rate of 0.01, the accuracy is seen to be increased when slower steps are taken i.e, with a learning rate of 0.0001. The accuracy of this model was about 72.95% with a validation accuracy of 67.38%. When we consider the model's performance by taking into consideration the time taken, accuracy and loss we can say that it is better if we choose the model with a learning rate of 0.01. To compare this model's performance we have experimented on a pre-trained VGG16 model and see that the model had an accuracy of 80.47% on a learning rate of 0.01. The model took around five hours and thirty-six minutes to be trained. The same model had been implemented on a learning rate of 0.0001 since the model took rather too long to execute it has not been mentioned and the accuracy of the model was not that complimenting either. The model further can be improved by training it on more epochs, due to computational power we have it trained on 10 epochs. The original VGG16 paper for the ImageNet classification, the model has been trained for 74 epochs and generates an accuracy of 92.7%. Hence, we can say that the model at only 10 epochs can produce an accuracy level of 80.47%. If it is trained for more epochs we certainly can outperform VGG16 accuracy level. To finish off the data was also trained on ResNet 50 which again is a pre-trained model, the model has been built from scratch rather than importing it from Keras applications. An accuracy of 90.62% with a validation loss of 82.20% has been observed which took about 6 hours to train on while we got quite a validation accuracy. While we ran the model again on a learning rate of 0.01 we saw that the model had an accuracy of 65.46%. We see from the confusion matrix of Section 6.1 and Section 6.2 that often the model has wrong interpretations on the category plastic, this is not caused due overfitting or underfitting of the model. Since the images are captured from a background color of white, it can be confusing for the model to understand the difference between the background and white plastics. A solution to this would be to have a proper image dataset with bounding boxes so that model gets trained on the exact location of the object. This helps the model in better understanding the patterns when trained on RGB colors. Figure 18 gives us information on all the model's accuracy performance for a concise understanding. To conclude on

Models	ResNet50	MLH-CNN	VGG16
Training	82.20%	70.53%	80.47%
Validation	55.74%	63.25%	84.60%
Precision	85.30%	54.04%	83.27%
Recall	80.36%	44.47%	78.94%

Figure 18: Overview of models training and testing accuracy

which model would do better is explained further in the conclusion section.

8 Conclusion

Waste when is not managed well, can lead to many problems such as climate, health risk, and harm wildlife. We can already see this happening in many ways. It is known that out of the 75% of recyclable materials the US recycles only 34% of them. When a developed country as the US is seen recycling at this low percentage, we can imagine the scenario of other developing countries like India and many others. To recycle the important task is to segregate materials into their respective categories. An immense problem in waste management factories goes through even after the machine sensors separate waste is that a manual process of segregation has to be done since most of the machine fails to recognize and separate materials. Here comes the picture of implementing deep learning. We can use cameras to classify materials, this research uses the data on waste from the Kaggle platform and presents a better way than the present state-of-art models. The obtained data has been explored carefully, pre-processed, and reshaped as per the applied model requirements. The MLH-CNN model presented by Shi et al. (2021)

has a simple structure on TrashNet than the heavy models such as VGG16 and ResNet50. The same structure has been implemented to our dataset, Since the data here is huge we have studied the performance variance on each model. From the model MLH-CNN, we see that the training accuracy of 57% is observed with a test accuracy of 46.52%. The pretrained model VGG16 produced a training accuracy of 80.47% and a validation accuracy of 81.80%. The overall precision and recall percentage of VGG16 was 83.27% and 78.94% respectively. The model has surpassed the performance by –. ResNet50 has been experimented with and has an accuracy of 82.20% which was the highest of all. We can say that ResNet50 architecture outperforms all of the models tested with a precision and recall values of 85.30% and 80.36% respectively. The study's main objective was to compare the custom MLH-CNN model with other state-of-art models and from the obtained results we can say that even though the model's architecture is simple enough this works well on a small dataset. Hence the performance on TrashNet of 2527 images was accurate enough. Since the dataset used in this research was huge it is seen that the model performs well on Dense Networks like VGG16 than the custom model. The model implemented can be further implemented in either the garbage initial collection point or the waste management outlets to segregate proper waste and improve the experience of living for humans.

9 Acknowledgement

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Christian Horn for the continuous support, insightful comments, and hard questions throughout the research work. For 13 weeks, my supervisors extended their valuable assistance to the success of this study. Thanks to Mostafa Mohamed for providing the dataset on Kaggle to the public and my appreciation also extends to my family for their exceptional love, constant understanding. I also acknowledge thanks to my friends who always inspired me with positive motivations.

References

Anh H., V., Le Hoang, S., Minh Thanh, V. and Tuong, L. (2019). A novel framework for trash classification using deep transfer learning., *IEEE Access* **7**: 178631 – 178639.

Bobulski, J. and Kubanek, M. (2021). Vehicle for plastic garbage gathering, *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pp. 1–5.

Cao, L. and Xiang, W. (2020). Application of convolutional neural network based on transfer learning for garbage classification, pp. 1032–1036.

Dr. Anne Meyer, Dr. Oleg Panchenko, D. T. G. (n.d.). Understanding plastics and polymers - the different types of plastic, Available at `https://www.azom.com/article.aspx?ArticleID=17477`.

Guo, Q., Shi, Y. and Wang, S. (2021). Research on deep learning image recognition technology in garbage classification, *2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, pp. 92–96.

Kang, Z., Yang, J., Li, G. and Zhang, Z. (2020). An automatic garbage classification system based on deep learning, *IEEE Access* **8**: 140019–140029.

Li, S., Yan, M. and Xu, J. (2020). Garbage object recognition and classification based on mask scoring rcnn, *2020 International Conference on Culture-oriented Science Technology (ICCST)*, pp. 54–58.

Matsuyama, E. (2020). A deep learning interpretable model for novel coronavirus disease (covid-19) screening with chest ct images, *Journal of Biomedical Science and Engineering* **13**: 140–152.

Patil, A., Tatke, A., Vachhani, N., Patil, M. and Gulhane, P. (2021). Garbage classifying application using deep learning techniques, *2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)*, pp. 122–130.

Sheh, K. (n.d.). Effect of batch size on training dynamics. Accessed: 2021-11-30.
URL: *https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e*

Shi, C., Tan, C., Wang, T. and Wang, L. (2021). A waste classification method based on a multilayer hybrid convolution neural network, *Applied Sciences* **11**(18).
URL: *https://www.mdpi.com/2076-3417/11/18/8572*

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, p. 141.

Sun, A. and Xiao, H. (2020). Thanosnet: A novel trash classification method using metadata, *2020 IEEE International Conference on Big Data (Big Data)*, pp. 1394–1401.

Teampanpong, J. (2021). Improper garbage management attracts vertebrates in a thai national park, *Écoscience* **28**(2): 107–113.
URL: *https://doi.org/10.1080/11956860.2021.1872264*

Velazquez, L., Anaya, C., Contreras-Aganza, O. and Munguia, N. (2020). A household waste stream analysis in a rural mexican location, *The Central European Review of Economics and Management (CEREM)* **4**(3): 53–65.
URL: *http://hdl.handle.net/10419/229833*

Versloot, C. (2020). How to use batch normalization with keras?
URL: *https://www.machinecurve.com/index.php/2020/01/15/how-to-use-batch-normalization-with-keras/*

Wentao, Z., Lan, G. and Zhisong, Z. (2020). Garbage classification and recognition based on squeezenet, *2020 3rd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM)*, pp. 122–125.

Xie, S., Girshick, R. B., Dollár, P., Tu, Z. and He, K. (2016). Aggregated residual transformations for deep neural networks, *CoRR* **abs/1611.05431**.
URL: *http://arxiv.org/abs/1611.05431*

Ziouzios, D., Tsiktsiris, D., Baras, N. and Dasygenis, M. (2020). A distributed architecture for smart recycling using machine learning, *Future Internet* **12**: 141.