Waste Classification by Using Deep Learning and Transfer Learning

MSc Research Project
MSc Data Analytics

Suchal Pote
Student ID: X21121397

School of Computing
National College of Ireland

Supervisor: Prof Athanasios Staikopoulos

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Suchal Pote
Student ID:	X21121397
Programme:	MSc Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Prof Athanasios Staikopoulos
Submission Due Date:	15/12/2022
Project Title:	Waste Classification by Using Deep Learning and Transfer Learning
Word Count:	8543
Page Count:	17

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:	Suchal Pote
Date:	15th December 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):	

Waste Classification by Using Deep Learning and Transfer Learning

Suchal Pote
X21121397

Abstract

Improper waste management directly impacts numerous habitats and species as well as air pollution and climate change. If waste can be identified at an early stage and classified, it can then be recycled or disposed of depending on the type of waste. This waste segregation helps to lower the cost of treatment, uses fewer land resources, and is also beneficial for social, economic, and ecological factors. This project aims to efficiently categorize the most common types of waste at the early stage by using deep learning frameworks on waste datasets. This separated waste can be further recycled or disposed of. The TrashNet dataset was previously used in many studies that classified only six types of waste. This research is being carried out in order to broaden the waste categories for the classification of other common types of waste. This is achieved by customizing the TrashNet dataset with additional classes. The transfer learning method is used to construct ImageNet pre-trained deep learning algorithms Vgg16, InceptionV3, Xception, and DenseNet201 to classify the 10 various forms of waste. The model's performance is evaluated in order to determine accuracy and loss. All the models suffered from overfitting in the first attempt. The 'Dropout Regularization' and 'Batch Normalization' in Keras are used for all deep learning algorithms to prevent overfitting of the model. Overfitting was not observed in the second attempt, also the accuracies of each model were increased by 4% to 5%. The performance of DenseNet201 is satisfactory with a validation accuracy of 89%, whereas Xception has achieved a validation accuracy of 87%. With a validation accuracy of 84%, InceptionV3 has good performance. Additionally, Vgg16 performs averagely with a 74% validation accuracy. The performance of all four models has been compared with each other to decide the most efficient algorithm. Several common types of waste can be successfully categorized for recycling or disposal using the suggested method.

Keywords- Deep learning, Waste, Transfer Learning, Vgg16, InceptionV3, Xception, DenseNet201.

1 Introduction

The waste nowadays produced by humans may contain harmful chemical elements that can directly affect human lives since it's not classified. The landfilling, incinerating, and stacking are the major ways to dispose of waste. It is challenging to avoid the infiltration of dangerous toxic elements by entering the whole ecosystem and causing polluted water as well as air pollution due to the explosion of toxic gases and serious impacts on plants and animals as stated in study (Shent et al.; 1999). Eventually, all this directly affects human health. Additionally, disposing of waste uses a significant amount of money. However, researchers in this study (Austin et al.; 1993) found that a sizable portion of the waste is made of recyclable materials. Mainly wastes include materials made of cardboard, paper, glass, plastic, and metal which make up around one-third of household, commercial and medical wastes that can be recycled or disposed of after classifying them. Even though there has been much research done to classify different trash types, they have not included the majority of common waste types.

The classification task has been carried out in many studies with the help of convolutional neural networks on the TrashNet dataset having only a few classes of cardboard, glass, metal, plastic, and trash, which do not cover other common waste categories. The authors in this study (Khubchandani et al.; 2020) have found that, during and after the COVID-19 epidemic, there has been a sharp increase in the one-time use of medication, masks, and gloves, also a shortage of semiconductors and e-chips in the global market as per researchers finding in study (Voas et al.; 2021), gave notion to incorporate these types of wastes. These types of e-waste can be recycled and made them into use again as well as medical wastes can be disposed of. Thus, this study is attempted to fill this gap by modifying the TrashNet dataset with additional classes.

Therefore, the prime aim of this study is to achieve better performance for the classification of numbers of most common household, commercial, and medical wastes. To classify the 10 different types of waste, the ImageNet pre-trained deep learning algorithms InceptionV3, Xception, DenseNet201, and Vgg16 are built using the transfer learning method with the help of TensorFlow Keras. The dataset includes cardboard, electronic chips, glass objects, surgical gloves, several types of face masks, pharmaceuticals, metal objects, paper, plastic, and trash. The overfitting of all the models has been noticed in the first experiment. To avoid overfitting of models the data augmentation techniques have been implemented, additionally, drop-out and batch normalization layers have additionally introduced in all proposed deep learning models.

Even though all models have attained excellent performance, there are some limitations. The photographs from the dataset used here were taken against a white background. Therefore, in reality, if the models lose their natural characteristics and shape, for example, if they become dusty or broken, the model's accuracy will drop. Due to identical shapes and features among different waste categories, there is a high probability of misguiding models. The moderate dataset size has an impact on how well the model works. By including more samples in the dataset, the performance of the models can be improved even further. The number of classes in the dataset has extended from 6 to 10, but there are still other waste categories that can be included to make the model more productive in practice.

This research paper is outlined in different sections. Section 2 describes tools, techniques, results, and limitations of previous studies conducted in the area of waste classification. Section 3 gives all the information about the methodology followed and the development of the models in this study. The architecture of all the models and different employed techniques are briefly discussed in section 4. Section 5 provides all the details about the implementation process that was carried out, different tools used, the python version used, and information about the machine's hardware specification on which research is conducted. The outcomes of each model in this research are outlined in section 6. It consists of the model evaluation techniques to measure the performance and accuracy of each approach. Section 7 discusses the conclusion of the study and interpretations of the model's result. This section also comprises the future work of this research.

2 Related Work

This section provides a critical assessment and analysis of the research in the waste classification fields, which helps comprehend the necessity to look at the effectiveness of deep learning models for various waste classes.
The production of solid waste in metropolitan areas is a major issue that, if not effectively handled, could lead to environmental pollution and be dangerous to human health. By adopting garbage classification and management systems, this problem can be reduced to a minimum. In order to streamline this process and eliminate human intervention, researchers conducted a study (Adedeji and Wang; 2019) to classify the waste. The Convolutional Neural Network (CNN) model implemented in this study is a combination of ImageNet pre-trained ResNet50 architecture and Support Vector Machine (SVM), where ResNet50 works as a feature extractor from the images and SVM works as a classifier to classify the images according to their class. The trash image dataset is employed here, which consists of 1989 records having 4 different classes glass, paper, plastic, and metal. The input image is resized to 224 x 224. The ResNet50 model implemented here is pre-trained on the ImageNet dataset, so the weights were already decided. But the last layer has been disabled by using $'include\_top = False'$, hence only extracted features came out and fed to the multiclass SVM architecture to perform the classification tasks. As the size of the dataset was very small, the data augmentation technique was implemented to increase the size. The final fully connected layer of ResNet50 architecture has been replaced by SVM for classification. The colour augmentation and batch normalization were implemented between convolutional and activation layers. To update, the weights and biases of Stochastic Gradient Descent with Momentum (SGDM) are employed here. The early stop method is used to abort training of the model, as validation loss was not reducing anymore after 12 epochs. The final validation accuracy achieved in this study is 87%. The authors draw the conclusion that the accuracy of the model can be improved by enlarging the dataset and including more photos. This paper's future work expresses the implementation of additional waste types. This paper's future work expresses the implementation of additional waste types. The author provides the knowledge of a model developed clearly. ResNet50 and SVM were combined in a method that produced successful results. However, in this work, the other model evaluation processes have not been carried out. Additionally, only 4 waste categories were used in this study, other most common waste categories could be included for classification to make the system more productive.

The increase in the generation of waste has an impact on the economy as well as contributes to climate change. To reduce the problems caused by garbage, it can be recycled. In this study Aral et al. (2018), a number of deep learning models have been implemented to address this issue. The classification of the various waste types is the main goal of this study in order to facilitate recycling. The TrashNet dataset was employed in this study along with the Densenet121, DenseNet169, InceptionResnetV2, MobileNet, and Xception architectures. Adam and Adadelta were preferred as the neural network model's optimizers. The TrashNet dataset used in this study has 6 categories of garbage including cardboard, glass, metal, paper, plastic, and trash. The dataset has been divided in a ratio of 70%-17%-13% for training, testing, and validation. Here, several experiments have been carried out with the help of the Keras library with TensorFlow (version 2.1.4). In the pre-processing stage, data augmentation steps were carried out due to the small size of the dataset including horizontal and vertical flipping, and image rotation. Even when employing data augmentation, researchers in this study encountered overfitting of some of the models in experiments, even though some of the models performed satisfactorily. Using weights from the pre-trained model on the ImageNet dataset, researchers fine-tuned several models through trials. The batch size for the DenseNet model training experiments was chosen to be 8. Similar to prior training experiments, a batch size of 32 was chosen. Here, the image input size of 224x224 was chosen for the training experiments for DenseNet, Xception, and MobileNet, in other side the image input size of 299x299 was used to train the InceptionResnetV2 model. With 100 epochs in the training phase, the InceptionResnetV2 network has achieved a test accuracy of 89%. Similarly, with the 150 epochs, DenseNet-169 and MobileNet network have been able to get 84% test accuracies. The researchers in this study conclude that the use of adam optimizer provides good accuracy than Adadelta and after applying the fine-tuning to the models the performance of the models can be increased. Some models have overfitting issues even after utilizing data augmentation and fine-tuning, which could be avoided by adopting transfer learning instead. Although it has been discussed in future development, the current system should be improved by adding more trash categories. While all the experimental information in this work is sufficiently detailed and provides a clear understanding of pre-tuning, other models that were not used in the experiments have been mentioned, which has caused additional confusion.

Another study Bircanoğlu et al. (2018) has carried out in 2018 to develop intelligent deep convolutional neural network architectures to effectively classify garbage. The pre-trained ResNet50, MobileNet, Inception-V4, DenseNet, Xception, and another novel approach named RecycleNet were implemented in this study. Some of the models were not built with any weights that had already been trained, While other models were created via transfer learning and then fine-tuned using weights from ImageNet. Adam and Adadelta were preferred as the neural network model's optimizers. Since this model takes some time to produce output predictions, they have changed the connection patterns of the skip connections within dense architecture to improve the model's ability to anticipate predictions. The trash image dataset is employed here, which consists of 1989 records having 4 different classes glass, paper, plastic, and metal. For models ResNet50, MobileNet, DenseNet, and RecycleNet input images have been resized to 224x224, while Inception-v4 and Xception have been set to 229x229. After analyzing the performances of various models that have been used, researchers found that when using the DenseNet family of networks that have been pre-trained on the ImageNet dataset, they achieved optimum classification results. But these architectures have one drawback, it takes a bit longer time for a prediction. To overcome this issue, they have changed the connection patterns of the skip connections inside dense blocks to improve the model's ability to anticipate outcomes. Here, several experiments have been carried out with the help of the Keras library with TensorFlow (version 2.1.4). They have also carried out fine-tuning tests on some of the best-performing models in order to benefit from model capacity. Asper Keras model definitions, the default learning rates for Adam and Adadelta optimizers have been chosen to be 0.001 and 1, respectively, without weight decay. The pre-trained model's weights were adjusted for the transfer learning experiments using the ImageNet dataset. A few data augmentation steps were performed such as image flipping and rotation to enhance the capabilities of deep learning models. The batch size has been set to 32. The models developed from scratch such as InceptionResNetV2, and DenseNet121 with the Adadelta optimizer have achieved maximum 90% and 84% test accuracies respectively, whereas, with the Adam optimizer, they were able to achieve maximum 80% and 85% test accuracies respectively. The ResNet50, MobileNet, and Xception using Adam optimizer have got 75%, 76%, and 85% test accuracies respectively. On other hand, the models that were implemented by using a fine-tuning method such as DesneNet121 and InceptionResNetV2 have achieved test accuracies of 95% and 87% respectively. A novel approach introduced in this research, RecycleNet was able to get 81% test accuracy. Researchers draw the conclusion that losing the actual shape and characteristics of the objects in the garbage may lower the accuracy of deep learning models in reality. Here, in some instances, the overfitting of the model was not prevented by using data augmentation techniques. The author's approach in this study is clear towards building the deep learning algorithms with modification in connection to creating novel architecture. The authors' strategy for developing deep learning algorithms with modifications in connection to providing unique architecture is evident in this study. Additionally, the process of developing models utilizing transfer learning and fine-tuning and comparing the outcomes provides a greater understanding of the importance of the two approaches.

The key concern is solid waste management, particularly in developing nations like Indonesia. To address the issue of waste management, many attempts were performed. The Indonesian government began a campaign to separate various waste types in September 2019. Using automatic garbage sorting can be beneficial for this program. The research paper submitted at IEEE in 2020 addresses this issue. In this research, paper Rismiyati et al. (2020) researchers have built three ImageNet pre-trained deep learning algorithms, Vgg16, Xception, and ResNet50. The TrashNet dataset is used to categorize various waste categories using the transfer learning technique. Normalization techniques were used in the pre-processing stage and the input image is mapped to the range of 0 to 1 according to the value of pixels. The input size for images has been set to 224x224 in this study. The training of the dataset was accomplished by maintaining epoch sizes of 20 and 50. In this experiment, the accuracy of the ResNet-50 model is increased by 1.4%, but the accuracy of the VGG-16 and Xception models are unaffected by this increase in epochs. Therefore, the ResNet-50, VGG-16, and Xception model's ultimate testing accuracy are 85.5%, 84.16%, and 87.13%, respectively. The graph of loss and accuracy indicates the generalization of all the models. Each model's confusion matrix revealed that no model could accurately predict the trash category. The author mentions a lack of samples in the dataset for the Trash class as the cause of this poor forecast. The similarity between trash and other categories and the imbalance of the dataset, according to researchers, deceives the models and reduces their accuracy. The pre-processing techniques used in this research article are not covered in detail in this paper. Also, the study's methodology was not revealed by the paper's authors. Additionally, the document gives no explicit information regarding the built-in layered architecture or its features.

A study was carried out in 2019 to detect several types of waste for recycling purposes. In this study (Sousa et al.; 2019), to detect waste and classify them, the hierarchical deep learning method is utilized. The main objective of this study is to identify and classify various waste materials from used food trays. The suggested two-step method keeps the benefits of modern object detectors (such as Faster R-CNN) while enabling the classification task to be handled in bounding boxes with better resolution. The dataset used here contains 1002 samples and was created by taking photographs of used food trays from different canteens, hotels, and cafes. The dataset has been further extended with an additional 180 samples from the food dataset. Thus, the dataset contains a total of 19 classes including a glass bottle, paper box, metal can, paper cup, plastic glass, plastic straw, plastic cutlery, tissue paper, etc. The classes in the dataset were divided into two groups by size and shape to identify bounding boxes of waste objects in trays and label them according to their belonged class. In order to expand the training dataset, certain data augmentation techniques were applied, such as brightness adjustment and image flipping. The Standard baseline method used here entails creating a model that utilizes all pictures and concurrently recognizes the bounding box and waste class for each input. They have used the cutting-edge Faster R-CNN method for object recognition and categorization. Faster R-CNN utilizes a similar network as CNN for region proposal as well as to identify the object and make training and forecasting in less time as compared to previous versions with the enhanced result. As faster R-CNN reduces the image resolution, so information gets lost while resizing, to overcome this issue an improved baseline approach is used. In this step, the model is again trained with 19 classes by keeping the bounding boxes. Further, this coordinates of bounding boxes and original images were resized to 500X500 and provided to CNN in order to perform the classification task. The author claims that this approach provides improved accuracy while preserving image information. On other hand, the proposed hierarchical approach is implemented in two stages. In the first stage preserving the bounding box, classification, and root class depend on their shape and material. In the second stage bounding boxes taken from original images were cropped, resized, and provided to CNN. A batch normalization step has been performed and the drop-out layer is also added and set to 0.5. The R-CNN is a trained material-based approach and shape-based approach, where the material-based approach contains 4 prime categories such as glass, plastic, paper, and metal, whereas the shape-based approach is sorted by their shape such as cup, bottle, plate, food waste, etc. The performance of the model is evaluated by mean average precision (mAP). In Cascaded Region Proposal and Classification approach, 81.4% mAP was achieved with faster R-CNN, whereas in the Hierarchical method, material-based and shape-based approaches obtained 80.9% and 86% mAP with faster R-CNN respectively. The future scope of this paper expresses that this dataset will be extended to test it with other CNN models. The goal of the study of achieving better performance with faster R-CNN than standard deep learning architectures has been successfully achieved. Also, the concept of categorizing waste of used food trays is very sensible. However, the structure of the paper and the information given about connections between approaches are very unclear and complicated to understand. Additionally, only the mAP technique is used to assess the model's performance; training and test accuracy and loss were left unplanted in order to better comprehend the model's performance.

The study Zhang et al. (2021) submitted to the journal of Resources, Conservation, and Recycling seeks to increase the precision of garbage sorting and to build an intelligent waste classification system through deep learning using mobile devices or other smart devices. Here, the deep learning residual network model is combined with a self-assessing module to utilize the particular features of channel graphs and reduce the spatial dimension data by having a global receptive field. However, the number of channels is kept the same, allowing the model to enhance the feature map's ability to represent data and automatically extract features from various forms of waste images. In this study, a model for waste identification and classification from images has been proposed, which is named as a 'classification of waste for recyclability model'(CTR). The residual network, which enables intelligent waste classification via mobile terminals, is used in this study to improve the model, whereas ResNet18 is used to optimize the CTR model. The ResNet18 model implemented here is made up of 18 layers including 17 convolutional layers and one FC (fully connected layer). This model undergoes through a convolutional layer, a pooling layer as well as through 4 small bunch of layers which are made up of 4 convolutional layers each. The pooling layer has been eliminated as it does not need parameter learning. The self-monitoring module (SMM) has been developed to extract the waste image's feature and to integrate with the ResNet18 network to build the CTR model. This module was consistently modifying the weights of waste image features to get closer to the global optimal solution and hasten network convergence. To classify the waste images TrashNet dataset has been utilized here, having a total of 6 classes including cardboard cartons, glass, plastic, metal, paper as well as trash. In this research, several data augmentation steps were performed to prevent form overfitting and to make the model more generalized in the data pre-processing stage. The data augmentation steps include the conversion of the RGB image into a single greyscale image followed by gaussian blur, randomly selecting various size and height ratios for the image to be cropped, image rotation to a random direction, and at the end cropping of the image to size 256X256 RGB image followed by image normalization. The model in this study has been trained with the help of the PyTorch framework. To optimize parameters, stochastic gradient descent (SGD) is employed. The batch size was selected to 95, where the learning rate was set to 0.01 and decreased by 10 times after 90 cycles. The epoch size was kept at 300 for the model training. Some model evaluation steps have been performed here with the help of confusion matrices, Receiver Operating Characteristic (ROC) curve, and Area Under the Curve (AUC), The loss value and accuracy were also calculated to evaluate the performance of the model. The proposed model has obtained 95.87% accuracy. The performance comparison of the proposed model with the previously developed model has been also contrasted at the end. The experiments conducted for this research using the ResNet18 model also obtained a classification accuracy of 88.66%. According to researchers, this study might be expanded in the future to test the proposed model using different datasets and improve classification efficiency and accuracy. This paper provides sufficient and in-depth information about the model development process. This article's information is logically organized. Every step that was taken has been supported by evidence. Sufficient model evaluation techniques have been conducted to determine the model's performance. The goal of this research is achieved by integrating a self-monitoring module with the deep learning model by obtaining 95.87% accuracy.

This work Chu et al. (2018) proposed the automatic categorization of trash that people throw out in public spaces by using a hybrid deep learning approach. Using a multilayer hybrid deep-learning system (MHS) to automatically categorize trash that people throw out in public spaces. This technology used sensors to find other important feature information and a high-definition camera to record trash images. The 5000 photos were created from the 100 RGB captured photographs of garbage. The main categories of the datasets were paper, plastic, metal, glass, kitchen waste, and organic waste. By using the Keras library the features of the images were enhanced and further resized from 640x480 pixels to 240x240 pixels. The data augmentation steps were performed such as rotation of images, shifting of height and width, image rescaling, and image zooming. Waste classifications were carried out using a multilayer hybrid method (MHS), which comprises many subsystems. Convolutional neural networks (CNN) as well as multilayer perceptrons (MLP) make up the system's basic components. The AlexNet was implemented here to classify the waste. This multilayer hybrid system (MHS) was made up of a number of interconnected subsystems, such as the image system, the sensor system as well the central back-end classification system. The camera and sensors in the hybrid system were turned on when a trash item is introduced for observation. The imaging system comprises a camera that takes pictures that CNN then analyzes. The sensor system acted simultaneously to gather numerical data from the items. The MLP system, whose inputs were the 22 CNN outputs and the numerical data from sensors, was used to generate the final findings (binary output). In this way, the CNN model's weight, as well as bias parameters, were preserved while the MLP system was trained separately. The input to MLP got from CNN's output, these 2 models actually worked together to produce the classification results of the binary. In the MHS model, which was tested under two distinct conditions, the object was put with fixed and random orientation and trained using 5000 data samples. The performance of the model was contrasted with that of a CNN model which has been used simply the image as input. For the first and second tests, MHS obtained accuracy rates over 90%, which is 10% better than the individual CNN approach. Additionally, the MHS model yields greater precision rates of 98.5%, 97.1% to 88.6%, and 85.9%, respectively, demonstrating its efficacy in forecasting recyclables. The researchers conclude that the MHS model was able to obtain better classification performance in two distinct testing situations with 98.2% and 91.6% accuracies.

This paper Vo et al. (2019) suggested a reliable model for automatic trash classification using deep neural networks, which can be used in intelligent waste sorting systems. The first collection, the TrashNet dataset [22], comprises 2527 photos with six different categories, paper, cardboard, plastic, metal, as well as trash ware gathered using smartphones. The VN-trash dataset is another dataset that was gathered for this project in order to address the issue of garbage classification. Three different waste categories are represented in this dataset: organic, inorganic, and medical wastes. The 5904 photographs in this collection were obtained through both photographing nature and exploring online resources. The research is conducted on the Pytorch framework. A solid framework using Deep Neural Networks for Waste Categorization (DNN-TC) is presented in this research. In order to reduce redundancy, the study employs ResNext as a base model and modifies the ResNext-101 design by adding 2 fully connected layers with output 1024 and N class dimensions following the global average pooling layer. The brightness levels of the input images are normalized during the data pre-processing stage to provide values ranging from 0 to 1. In order to generate more images for the training and testing phases, the input photos will then be subjected to a number of pre-processing procedures, including the horizontal flip and random crop around the image with 224 224 sizes. The input photos that represent each distinct waste class are fed into the suggested architecture throughout the training process. The confidence for each label is determined in the final layer using the log SoftMax function. First, the suggested model loaded the pre-trained weights from the original ResNext-101 onto the ImageNet dataset to initialize the weight model. The model then goes through a fine-tuning process to learn the features of the different types of garbage from the trash dataset, and the most accurate model is chosen by making an estimation on the validation set. They pre-trained and adjusted the ResNext-101 model by swapping out the final fully connected layers. In these procedures, the ResNext-101 model with a learning rate of 0.0001 is optimized using the Stochastic Gradient Descent (SGD) algorithm. The learning rate for the Adam optimizer's hyper-parameters for the proposed model is set to 0.001. Then, over the following 100 epochs, SGD is applied as an optimization tool for the fine-tuning model with a learning rate of 0.0001. Additionally, this study examines the model's performance on the validation set for each epoch during training processing using a mini-batch size of 8 with 100 epochs. In the testing phase, researchers evaluated the best model on the testing set to predict the final output for each input image. During testing, they assessed the top model on the testing set to forecast the result for each input image. For the TrashNet dataset, the DNN-TC model performs better than other methods. In terms of accuracy, DNN-TC obtained 94%, whereas Densenet121 Aral, RecycleNet, and ResNet Ruiz obtained 91%, 68%, and 72%, respectively.

All the research carried out previously in the area of waste classification with different deep learning models gave knowledge about building the model, steps performed in the pre-processing stage, different techniques performed to avoid overfitting of the model to achieve better accuracy, and different performance metrics techniques. Even though the TrashNet dataset, which has six trash categories, was used in nearly all studies. The future scope of these studies suggests that the research will be expanded by enlarging the dataset by including new categories. The dataset presented in this paper is attempted to address this problem by modifying the TrashNet dataset by adding other common waste types such as multiple types of medical wastes and electronics waste (e-waste).

3 Methodology

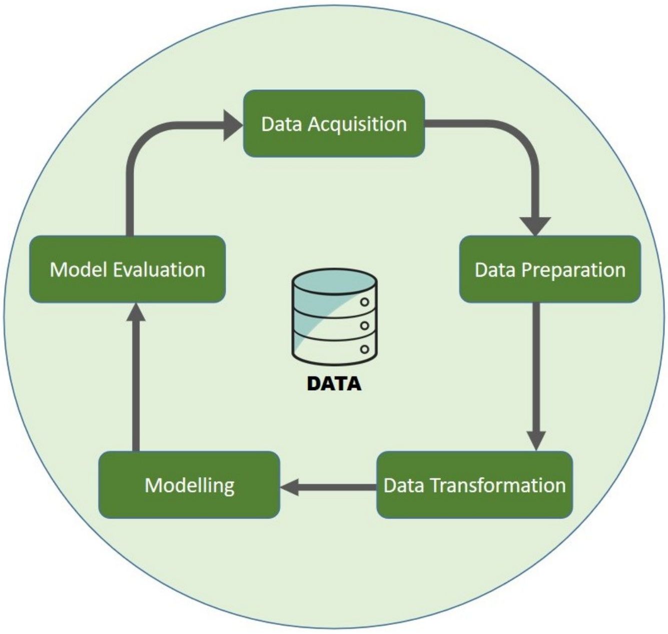


Figure 1: Process Flow

To perform the classification task for several types of waste and to achieve the aim of better classification performance, four pre-trained deep learning algorithms are identified. As shown in fig 1, the five-step research technique used in this study is described below.

3.1 Data Acquisition:

In the first stage, The TrashNet dataset with different classes of recyclable or disposable waste is obtained in order to categorize various waste types. Although this dataset is unlabelled, it covers the main different forms of waste as compared to the other available datasets including cardboard, papers, glass, plastic, metal, and trash. This dataset comprises 6 categories and 2427 images of waste objects that were photographed on a plane and white background. The size of all images was reduced to 512x384. To increase the effectiveness of the model and classify other common forms of waste, the TrashNet dataset is modified by additional categories including e-waste, used masks, used gloves, and unused or expired pills. Therefore, the final dataset contains a total of 10 categories and 4157 images. The sample from each category is showcased in fig. 2.

Figure 2: Sample images from each category of dataset

3.2 Data Preparation:

The dataset used in this study was already divided into training and test sets in ratio of 50:50. Some of the images were found to be corrupted, which are eliminated in this stage. The input images are resized to 224x224 as per model requirement and fed to the data transformation stage. The number of images in the test and test set for each category is showcased through the bar plot in fig.3.

3.3 Data Transformation:

In the data transformation stage, some data augmentation techniques are performed in order to prevent overfitting of the models and to make the model generalize. The ImageDataGenerator class has been used to conduct data augmentation steps. In this, the input normalization is performed by rescaling the input image on both train and test

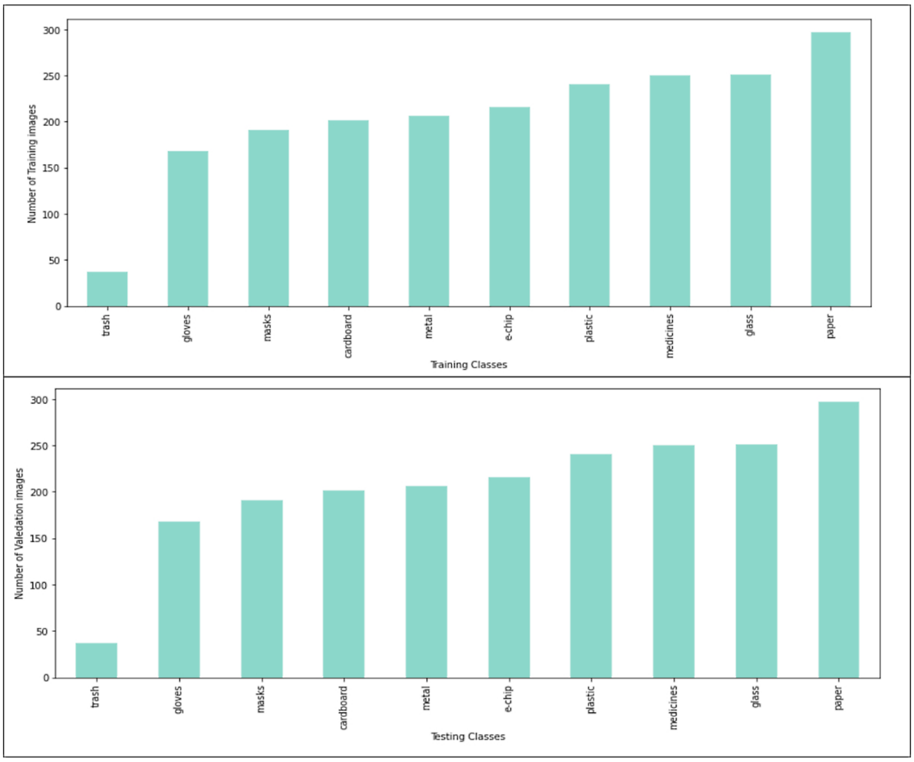


Figure 3: Total number of images in training and test set

sets to dynamically resize pixel values to the range [0,1] from the range of [0,255]. Due to image rescaling, the images will contribute to the total loss more evenly. Along with rescaling, image shearing is carried out with a range of 0.2. Additionally, image zooming with a range of 0.2 and random horizontal flipping is also performed. The input images after have been resized to 224x224.

3.4 Modeling and Evaluation:

In this stage, the proposed pre-trained deep learning models including Vgg16, Inception, Xception, and DenseNet201 are designed by using transfer learning techniques to train and test processed data. The performance of the models was then evaluated in the model evaluation stage. In order to calculate training accuracy and loss as well as accuracy and loss the accuracy and loss graph has been plotted for all deep learning models that are proposed in this study.

4 Design Specification

The goal of this research is to develop a more accurate waste classification deep learning system for the most common domestic, commercial and medical wastes by customizing the TrashNet dataset with the addition of more classes. In order to achieve the best classification performance four deep learning models including Vgg16, Inception, Xception and DenseNet201 are designed by using the transfer learning method. The architectural design of these models is described below.

4.1 Vgg16 Architecture

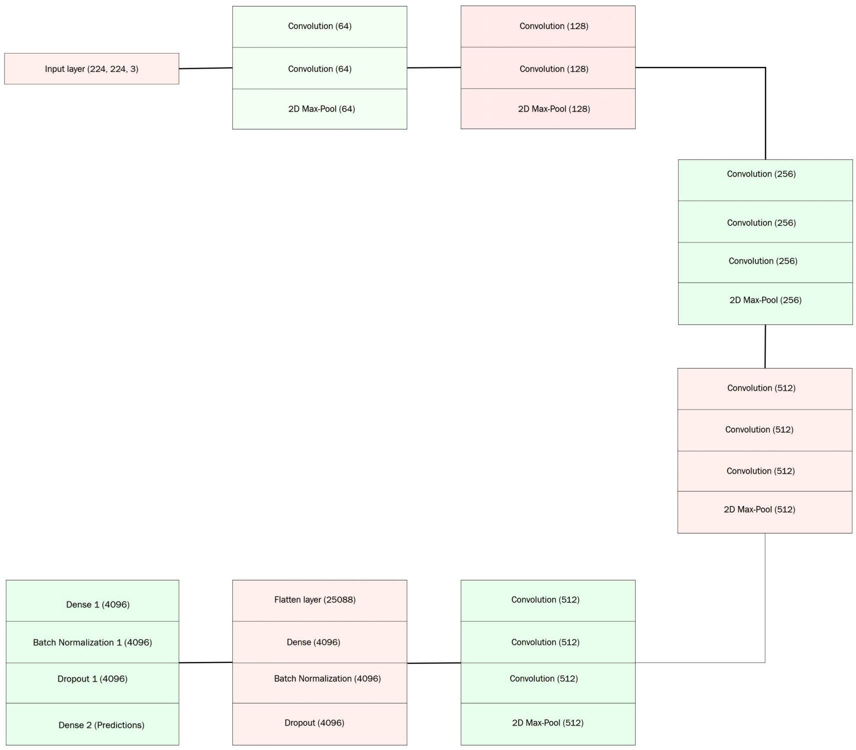


Figure 4: Vgg16 Architecture

The Vgg16 is a convolutional neuron network designed here in such a way that each layer of the layer provides its feature maps to the next block of layers. This is a pre-trained model on the ImageNet dataset, hence the already decided weights are used and the last layer of the model has been removed Adedeji and Wang (2019). Starting from the top, the model typically consists of an Input layer with an input image size fixed to 224x224 and with 3 RGB channels. The input image of size 224x224x3 is processed using two convolutional layers, each with a relatively small filter size of 3x3. The stride of the convolution layer is set to 1 pixel. The spatial pooling is accomplished by five max pooling layers which are followed by a group of convolution layers having a filter size of 3x3. These max pool layers have a filter size of 2x2 pixels with a stride set to 2. The convolution layers have a relatively low number of channels. The first layer contains 64 channels and raises with X2 after every max pool layer, where X is the number of channels. The number of channels each convolution has 64, 128, 256 and reaches to 512 channels Simonyan and Zisserman (2014).

The next layer is a flatten layer with 25088 channels, in order to flatten the input image for this layer. The output of flatten layer is fed to the dense layer, batch normalization layer, and dropout layer with 4096 channels in each layer. The output of these layers is again passed to the dense layer, batch normalization layer, and dropout layer with 4096 channels. In order to avoid overfitting the model and make the model generalize as well as faster, two batch normalization layers have been added to this neural network. The two drop-out layers are also introduced in the model to prevent the model from overfitting. The final layer is a dense layer that is responsible for predictions and the number of units in this layer is equal to the number of classes. The last layer is the SoftMax layer.

Every hidden layer has rectification non-linearity unit (ReLU) implemented Simonyan and Zisserman (2014). Fig. 4 shows the architecture of the Vgg16 model implemented in this study.

4.2 InceptionV3 Architecture

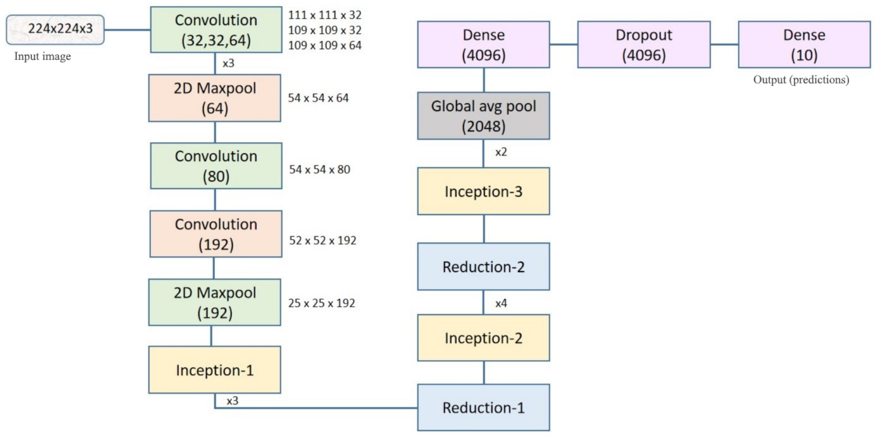


Figure 5: InceptionV3 Architecture

Another deep learning neural network is used here, InveptionV3, to categorize wastes Szegedy et al. (2015). The standard input image size for InceptionV3 is 299x299x3, but in this study, the input image size is set to 224x224x3 as same as the Vgg16 model. The architecture of InceptionV3 is illustrated in fig.5. The model consists of three convolutions layers at the beginning with filters of 32, 32, and 64 which resize images to 111x111, 119x119, and 119x119 respectively. The filter size of maxpool layer is 54x54x64 which feds the image to another set of convolutional layers. These convolutional layers have 80 and 192 filters and provide output images with sizes of 54x54 and 52x52 respectively. The second maxpool layer has a filter size of 25x25x192 and gives input to the first inception block. The first inception block consists of seven convolutional layers and one maxpool layer as showcased in fig.6. The architecture contains three such inception-1 blocks, followed by the reduction-1 block. The reduction-1 block contains four convolutional layers and one maxpool layer with a filter size of 3x3 as shown in fig.6. The output of the reduction-1 block is passed to the inception-2 block. It consists of ten convolutional layers and one average pool layer with a filter size of 3x3. This architecture is having four inception-2 blocks followed by a reduction-2 block. The reduction-2 block is having a network of one maxpool layer with a 3x3 filter size and seven convolutional layers. The resulting image is sent to the Inception-3 block, which has a network with nine convolutional layers and one maxpool layer with a 3x3 filter size. There is a network of two inception-3 blocks in this architecture. After each convolutional layer, batch normalization layers are added to make the model more general and prevent overfitting. Each batch normalization layer is followed by an activation layer. The output of Inception-3 is fed to the average pool layer. The additional drop-out layer has been added to avoid overfitting the model.

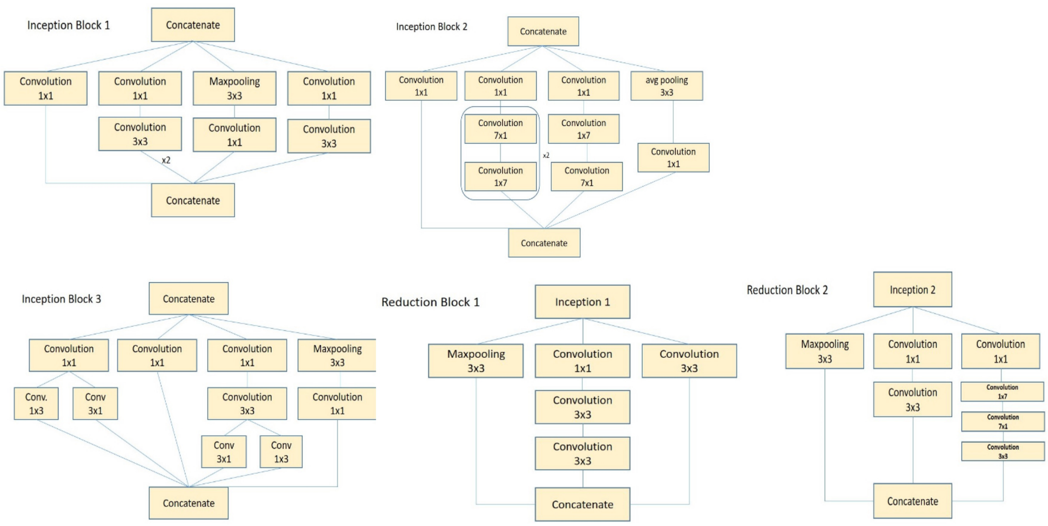


Figure 6: Architecture of Inception and Reduction blocks

4.3 Xception Architecture

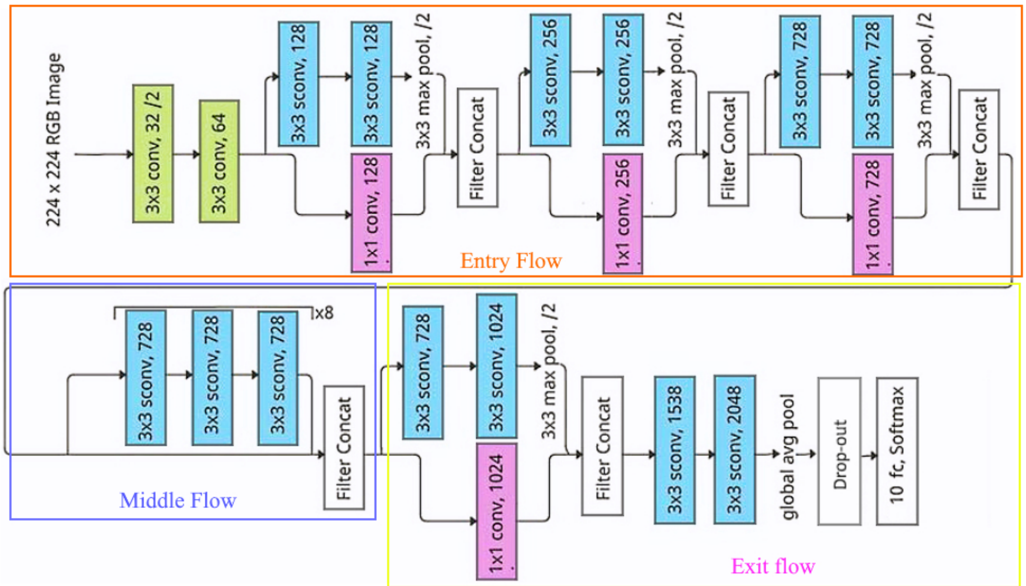
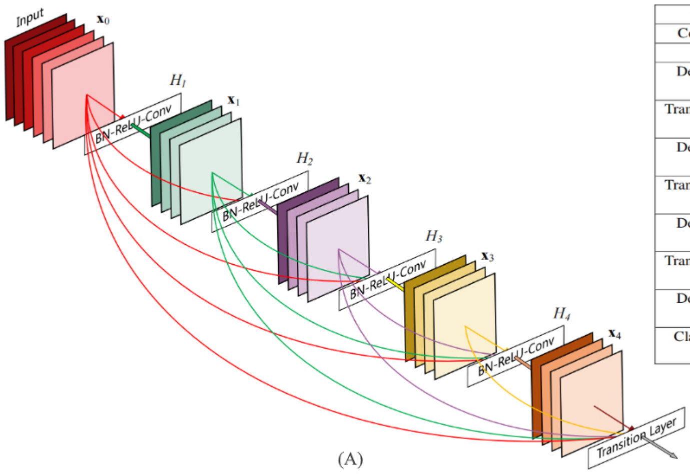


Figure 7: Xception Architecture

François Chollet is the developer of the Xception model Chollet (2017). The depth-wise Separable Convolutions method in the Xception is employed in this study. The depth-wise Separable Convolutions are a variation to regular convolutions that are intended to be faster to compute. In depth-wise separable convolutions, two primary processes take place: a) Depth-wise convolution, and b) Pointwise convolution. The first stage is pointwise convolution, while the second is depth-wise convolution. Fig.7 represents the architecture of the Xception model. This architecture is divided into three prime parts entry flow, middle flow, and exit flow. The Xception model has 36 convolutional layers that serve as the network's feature extraction foundation. The image size of 224x224x3 is fed to the entry flow stage, passes through the middle flow stage where it gets executed 8 times, and then through exit flow to produce predictions. The entry flow contains five convolutional layers, six separable convolutional layers, and three max pool layers. The middle flow consists of a set of three separable convolutional layers which is executed for 8 times. The exit flow contains one convolutional layer, four separable layers, a max pool layer, an average layer, and a dense layer for predictions. The drop-out layer has also been added to avoid the model from overfitting. Each convolutional layer and separable layer have been followed by the Batch Normalization layer in the Xception architecture. The softmax activation function is used for predictions in the dense layer.

4.4 DenseNet201 Architecture



(A)

Layers	Output Size	DenseNet-201($k = 32$)
Convolution	112×112	7×7 conv, stride 2
Pooling	56×56	3×3 max pool, stride 2
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv
	28×28	2×2 average pool, stride 2
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv
	14×14	2×2 average pool, stride 2
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Transition Layer (3)	14×14	1×1 conv
	7×7	2×2 average pool, stride 2
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$
Classification Layer	1×1	7×7 global average pool
		fully-connected, softmax

(B)

Figure 8: DenseNet201 Architecture

The architecture of DenseNet201 is showcased in fig.8(A). As we can observe from the architecture, unlike other convolutional neural networks, the Dense Convolutional Network (DenseNet) establishes a feed-forward connection between each layer and each subsequent layer Huang et al. (2017). The DenseNet makes the direct connection in L(L2 +1)/2 manner, where L is the layer number. The feature maps of all previous layers served as inputs for each layer, and feature maps of their own served as inputs to all succeeding levels. The layers having the particular size of feature maps are lined with another layer having the corresponding feature map size. The DenseNet201 is made with four dense blocks. Here, the input image of size 224x224x3 is fed to the input layer and passes through succeeding blocks. Each layer receives extra inputs from all earlier layers and transmits its very own feature maps to all succeeding layers in order to maintain the feed-forward structure. Each block is connected through transition layers which consist of a 1x1 convolution layer, 2x2 average pool layer, rectified linear unit (ReLU), and batch normalization. The layered network of dense blocks is shown in fig.8(B) Each dense block contains a 1x1 convolution layer and a 3x3 convolution layer. Block 1 is executed 6 times. Similarly, block 2, block 3 and block 4 executed 12, 24, and 16 times respectively. All the convolution layers have been followed by a sequence of batch normalization and ReLU. The final classifier layer contains a 7x7 global average pool, a drop-out layer, and a fully connected dense layer. The drop-out layer is added to minimize the effect of overfitting the model.

5 Implementation

In this research paper pre-trained deep learning models including Vgg16, InceptionV3, Xception, and DenseNet201 have been implemented by using Keras with the TensorFlow framework. This research is carried out on the Jupyter notebook with python version 3.9.7. The input image size is scaled to 224x224x3 before passing through to all proposed convolution networks. The last layer of each model has been disabled. The transfer learning technique is used in this research; thus, weights are taken from the pre-trained model using the ImageNet dataset. All the models have been loaded and batch normalization, as well as drop-out layers, have been added to models. The batch normalization and drop-out layer are introduced to prevent models from overfitting. The modified dataset has unlabelled images and comprises 4157 total images across 10 categories. The dataset was already equally divided into train and test set; hence no splitting steps is performed. The 'adam' optimizer and '*categorical_crossentropy*' as a loss function have been used to compile models before training. To avoid models from overfitting data augmentation techniques have been conducted. The ImageDataGenerator class has been used to perform data augmentation steps on the test set. The data augmentation steps consist of rescaling the image, image shearing, image zoom, and horizontal flipping. Only image rescaling is performed on the test set. The batch size is kept at 32, whereas the target image size is set as 224x224 for both the train and test sets. For all models, an epoch size was selected as 15 for training. DensNet201 and the Xception model employed the EarlyStopping approach by keeping the minimum epoch size as 10 since the validation loss increased after a specific number of epochs. Every model required a different amount of time to train the models. The Vgg16 and DenseNet201 models took longer time to train, about 3 hours, whereas Xception and InceptionV3 took less time to train, about 1 hour and 45 minutes. This experiment is carried out on Microsoft Windows 11 computer with an 11th Gen. Intel i5 processor and 8GB RAM.

6 Evaluation

The aim of this research is to classify the most common types of waste by using different pre-trained deep learning models using the transfer learning technique. After the training, all models are saved, so they can be loaded afterward if required. This section describes the results of all the models after the training of the models that were carried out. To assess the performance, the accuracy and loss graphs for both the training and test set are plotted. The epoch size is selected as 15 for all models for training. Fig. 9 represents a graph plotted for training and test accuracy and loss. The X-axis indicates the number of epochs and the Y-axis indicates accuracy/loss. As we can observe in the figure, the results are satisfactory as the raised graph of training and validation accuracy for all models is a sign of increased accuracy after each epoch. On other hand, the loss graph is decreased for all models after each epoch which indicates a low error rate. The Vgg16 has achieved 74% test accuracy whereas, InceptionV3 has achieved 84% test accuracy. The Xception and DenseNet201 models have obtained 87% and 89% test accuracies respectively.

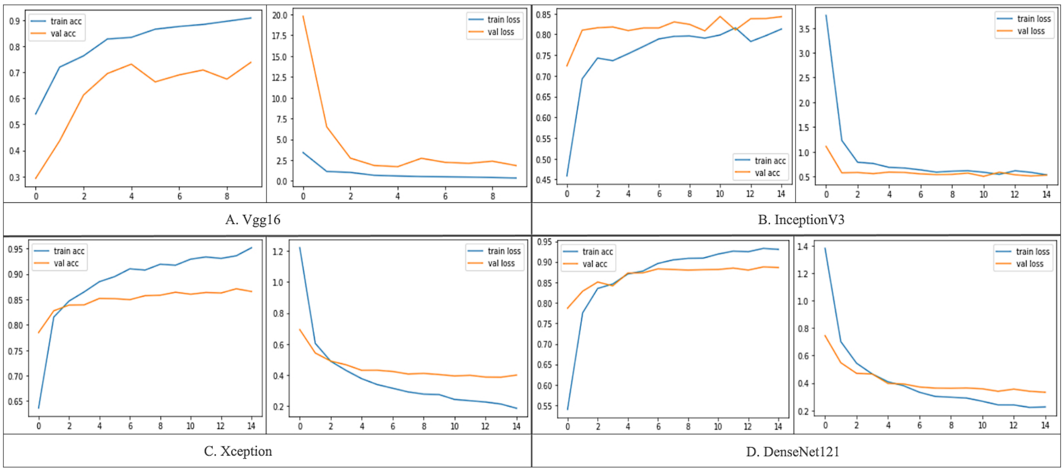


Figure 9: Graph of Accuracy and Validation loss for all models

6.1 Discussion

The results as well as interpretations of the research are discussed in this section. The comparison of the performance of each model is also carried out in this part. The TrashNet dataset is modified for this study, and some of the most typical commercial, domestic, and medical wastes are also taken into account. The weights of the pre-trained models are set to the ImageNet dataset and the last layer have excluded as the transfer learning method is supposed to utilize. After the Vgg16, InceptionV3, Xception, and DenseNet201 models were trained, overfitting for all of the models was noticed because the validation loss for each model grew quickly after a specific number of epochs due to the smaller dataset size. Thus, the dropout and batch normalization were added in models implemented in this research to reduce the effect of overfitting. In DensNet201 and the Xception, EarlyStop is also utilized to minimize validation loss. The training for all the models has performed again with an epoch size of 15. As shown in fig the validation loss for all the models has not been raised in the second experiment, which indicates a lower error rate in this attempt. The accuracy graph for the train and test set is also raised as per expectations. After the drop-out layer and batch normalization test accuracy of each model is observed to be increased by 4%-5% The fig.10 given below summarizes the comparison in terms of train and test accuracy as well as the amount of consumed training time for each model. DenseNet201 achieves the highest level of accuracy out of all, followed by the Xception approach. The Vgg16 approach has obtained the lowest test accuracy. The more training time required for DenseNet201, and least time required for the InceptionV3 approach.

Models	Epoch	Accuracy		Training time (Minutes)
		Training	Test	
Vgg16	15	92%	74%	182
InceptionV3	15	81%	84%	105
Xception	15	95%	87%	110
DenseNet121	15	93%	89%	184

Figure 10: Comparison of all models

7 Conclusion and Future Work

The use of the pre-trained convolutional neural network to classify the most common types of waste with more accurate classification performance is the motive of this research. This classified waste can be further recycled or disposed of. All the studies conducted previously to classify the wastes, where TrashNet dataset with few categories of wastes was commonly used. To extend the study of waste classification, other common types of waste have also been considered here by modifying the original TrashNet dataset. The Vgg16, InceptionV3, Xception, and DenseNet201 models with the use of transfer learning are developed here to categorize cardboard, electronic chips, glass objects, surgical gloves, several types of face masks, pharmaceuticals, metal objects, paper, plastic, and trash. After the model's training overfitting was noticed for all approaches, which is later avoided by drop-out and batch normalization methods. Although all the approaches have performed satisfactorily there are high chances of the model's misguiding due to similar kind of shapes and properties across waste categories. For example, a cup or bottle can be found in the glass, paper, or metal categories, also paper can be found in the paper and trash categories. Additionally, the model's efficiency has also been impacted by small dataset size and imbalanced data across categories. In practice, these wastes might be in a different situation or lose their shape and properties, which could affect the accuracy of the models and results would be different. Overall, this study was successful in reaching its objective of effective waste classification. This study can be extended in the future by adding noise in the images of the dataset as well as by changing the original property and shape of waste objects to achieve efficient performance in the reality also. Additionally, the dataset size could be increased, and different other deep learning algorithms could be implemented.

References

Adedeji, O. and Wang, Z. (2019). Intelligent waste classification system using deep learning convolutional neural network, *Procedia Manufacturing* **35**: 607–612.

Aral, R. A., Keskin, R., Kaya, M. and Hacıömeroğlu, M. (2018). Classification of TrashNet Dataset Based on Deep Learning Models, *2018 IEEE International Conference on Big Data (Big Data)*, pp. 2058–2062.

Austin, J., Hatfield, D. B., Grindle, A. C. and Bailey, J. S. (1993). Increasing recycling in office environments: The effects of specific, informative cues, *Journal of Applied Behavior Analysis* **26**(2): 247–253. Publisher: Wiley Online Library.

Bircanoğlu, C., Atay, M., Beşer, F., Genç, and Kızrak, M. A. (2018). RecycleNet: Intelligent waste sorting using deep neural networks, *2018 Innovations in intelligent systems and applications (INISTA)*, IEEE, pp. 1–7.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258.

Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S. and Xiong, X. (2018). Multilayer hybrid deep-learning method for waste classification and recycling, *Computational Intelligence and Neuroscience* **2018**. Publisher: Hindawi.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. (2017). Densely connected convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Khubchandani, J., Saiki, D. and Kandiah, J. (2020). Masks, gloves, and the COVID-19 pandemic: rapid assessment of public behaviors in the United States, *Epidemiologia* **1**(1): 16–22. Publisher: MDPI.

Rismiyati, Endah, S. N., Khadijah and Shiddiq, I. N. (2020). Xception Architecture Transfer Learning for Garbage Classification, *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, pp. 1–4.

Shent, H., Pugh, R. J. and Forssberg, E. (1999). A review of plastics waste recycling and the flotation of plastics, *Resources, Conservation and Recycling* **25**(2): 85–109.
URL: *https://www.sciencedirect.com/science/article/pii/S0921344998000172*

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* .

Sousa, J., Rebelo, A. and Cardoso, J. S. (2019). Automation of waste sorting with deep learning, *2019 XV Workshop de Visão Computacional (WVC)*, IEEE, pp. 43–48.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015). Going deeper with convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.

Vo, A. H., Vo, M. T., Le, T. and others (2019). A novel framework for trash classification using deep transfer learning, *IEEE Access* **7**: 178631–178639. Publisher: IEEE.

Voas, J., Kshetri, N. and DeFranco, J. F. (2021). Scarcity and global insecurity: the semiconductor shortage, *IT Professional* **23**(5): 78–82. Publisher: IEEE.

Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X. and Liu, X. (2021). Recyclable waste image recognition based on deep learning, *Resources, Conservation and Recycling* **171**: 105636. Publisher: Elsevier.