National College of Ireland

Image Based Trash Classification using Machine Learning Algorithms for Recyclability Status

MSc Research Project
Data Analytics

Mandar Satvilkar
x17108837

School of Computing
National College of Ireland

Supervisor: Prof. Noel Cosgrave

National College of Ireland
Project Submission Sheet – 2018/2019
School of Computing

Student Name:	Mandar Satvilkar
Student ID:	x17108837
Programme:	Data Analytics
Year:	2018
Module:	MSc Research Project
Lecturer:	Prof. Noel Cosgrave
Submission Due Date:	13/08/2018
Project Title:	Image Based Trash Classification using Machine Learning Algorithms for Recyclability Status
Word Count:	7000

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 encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	17th September 2018

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. 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):	

Contents

Image Based Trash Classification using Machine Learning Algorithms for Recyclability Status

Mandar Satvilkar
x17108837
MSc Research Project in Data Analytics

17th September 2018

Abstract

This research paper talks about Trash classification which comes across as a social and unique cause where Image processing through Analytics can help with better management of garbage and waste materials. A standout amongst the most proficient approach to process trash as indicated by its recyclability is to comprehend what class that waste has a place. Analytics can make this herculean task achievable through image classification into its proper categorizations. The goal of this task is to take pictures with single items, segregate them and group those pictures into five particular classifications of cardboard, glass, metal, paper and plastic. We will utilize a dataset of around 400-500 pictures from every class indicated previously. This dataset was made accessible freely and the creators of the dataset give full access to it through their GitHub store. The model utilized will be a Convolutional Neural Network (CNN) because of its impressive and efficient capabilities in the world of machine learning, particularly in image classification. Further, an Extreme Gradient Boosting will be implemented and checked if it can perform better than a CNN. Multiple methods were juxtaposed in order to zero in on the best method or combination of methods which can prove beneficial and efficient for image classification.

Keywords: Image Classification, Trash Classification, Support Vector Machines, Convolutional Neural Networks, Random Forest, Extreme Gradient Boosting, k Nearest Neighbours.

1 Introduction

Trash classification is a topic which is less trodden by researchers and is an uncharted territory having a whole lot of potential to bring change in society. In this paper it is aimed to bridge this gap between image classification and trash classification. The motivation to pursue this topic was the way in which our environment and mind-set is changing every day and thousands of tons of trash is dumped into water and land unconsciously. The same waste if recycled and put to use can benefit ecology and human race in ineffable ways. The first step in doing so is to identify the right type of material for apt requirement which as of now we do with the help of limited human knowledge manually. But image classification using machine learning can help detect the right material type

in a trash which can then be put to use in an appropriate way. Secondly humans collecting garbage if are not educated properly can mix multiple materials together without its proper identification, which again can be prevented by an interface which just classifies the trash using a picture. Images are generally in the size of 512 x 384 pixels but image of such a size created 589824 columns and such a huge dataset was too heavy to process with limited computational resources and hence images were scaled down to 10% of their original size which means to 51*38 pixels yielding 5814 columns. Since the data dealt over here was in the form of images, it was difficult to decide whether any particular pixel can become the principal component for modelling. So, dimensionality reduction of images was the only technique used to limit the size of the dataset. When it comes to image processing, the main features taken into consideration are HOG (histogram of oriented gradient), SIFT (Scale Invariant feature transform), Gabor filters and Fischer Kernels and deep learning techniques have the ability to learn these features on their own by reading rich feature representation from intensities of pixels in an image. Talking about the performance of complex models, Convolutional Neural Networks portray state-of-the-art performance for image recognition and classification, material classification and semantic labelling Krizhevsky Alex and E (2012). Due to the complexity and reputation of being one the most proven classification algorithm, SVM was used in our study and its performance was compared with CNN. In case of a multiclass SVM, a one versus all classification is used where the model classifies the test dataset with greatest margin. SIFT features of an image were exploited by SVM where SIFT created a blob in an image describing image in 128 numbers. This multidimensional data is segregated into hyperplane by SVM which sets the base for SVM classification. Adaboost was then used in this paper to boost the capabilities and performance of SVM by optimizing the features of the dataset. This paper aims to compare the capabilities of SVM, Random Forest (RF), XGBoost and CNN for trash image classification and we will discuss these approaches, their implementation and evaluation in detail in the research below.

2 Related Work

2.1 Data Collection

Trash classification essentially come under the under the umbrella of image classification with only garbage and trash related images being as subject of interests. Having said that, multiple databases were taken into consideration for this research which can be defined as follows

1. CURET Standard Image dataset with 61 different textures and 205 different lighting conditions for image [Liu et al. (2010)]

2. KTH- TIPS2 Introduced to create intra-class variation, this dataset has 11 different textures and 4 samples for every category photographed under varied conditions [Liu et al. (2010)]

3. PASCAL VOC Realistic databases of images having multiple objects and people and various interactions between them [Uvarov (2017)]

4. NUS WIDE Real world pictures in multiple combinations taken under various environmental conditions [Uvarov (2017)]

5. Garbage In Images (GINI) Bing search API was used to crawl the web for images of garbage and locations where patches of garbage were used for classification from the taken picture. [Mittal et al. (2016)]

6. Flickr Material Database Created from Flickr.com, multiple images from multiple categories are chosen to make a database [Yang and Thung (2016)]

But since our topic was trash classification, all the above databases were studied thoroughly to scrape trash related images but no relevant images were found and some of the databases in other researches were private databases which needed authorizations, hence Bing search was carried out on web to scrap trash images through Bing search API. In order to make a good classifier it is important to collect data from legitimate source and the data in that dataset needs to be feature rich to train the model efficiently. There was no direct connection of these image databases with trash classification as they contained minimal information that could help in classifying the recyclable trash. For instance, Flickr image database contained images which were in sound and undamaged state whereas trash would be in a ruffled and damaged state. Hence specific database was created for images related only to trash and made it public to be used by other researchers to carry out their research. The dataset consists of 400-500 images of each category of trash and making it a rich repository of 2390 images in total. The data acquisition involved using a white poster in the background and collected pictures from various locations around them. They used separate lighting and positions for each picture making it distinct in posture and made sure that there is variation in the dataset. This variation was necessary as to make sure all angles of observations are covered and prediction of any new image would be relatively healthy. Post collection, the images were augmented to increase the size of each class in the dataset. Random operations were performed on the image to rotate, adjust brightness, translation, scaling and shearing of the image. Below shown are the examples of images contained in the five different categories of the dataset [Yang and Thung (2016)].

Figure 1: Cardboard

Figure 2: Glass

Figure 3: Metal

Figure 4: Paper

Figure 5: Plastic

2.2 Classification of Images using machine learning:

Myriad of methods and techniques have been attempted at optimized classification of images. Image classification is essentially segregating the image to a belonged class by subjecting the image data to classification algorithms. Classification can be further bifurcated into Parametric and non-parametric classification- Parametric classification involves SVM and boosting algorithms whereas non parametric algorithms constitutes of Bayesian network , Random Forest and so on. A basic and coherent non parametric approach to classification is NBNN (Naïve Bayesian nearest neighbour)
NBNN yet is ineffective because of the following reasons

1. The complexity of the algorithm while testing is at peril.
2. For accuracy with NBNN, datasets have to be balanced.
3. Independence of descriptors hinders the classification.

NBNN are proposed in spite of NBNN having a fair performance. NBNN classification suffers from a detrimental assumption of conditional independence and yet no work has been done to deal with it. Hence in Sun (2007), dependencies between local features will be evaluated and layer structure will be introduced for the same. Bayesian Network will be remodelled to overcome the weakness for conditional independence assumption. No training of the dataset is done to preserve the discriminative power of the dataset parameters and its ample and ambient with high volume datasets Naïve Bayes has been used in text classification where all the words in the document were treated independently but thats not the case when it comes to real world problem, for in real world a combination of words makes more sense logically than independent words and this exactly is the assumption of conditional independence for Naïve Bayes classifier. Although with advancement in computer and technology a lot of work has been done to improvise NBNN such as SIFT (Scale Invariant Feature Transform) developed by David Lowe which takes into consideration the key points in the images and features which describe the key point. Image classification according to this literature is based on such local features which are dependent on each other and can together help build a layered architecture of local dependencies to overcome the problem of conditional independence in NBNN.

According to Dash and Mrutyunjaya Panda (2016), Image classification and identification is one of the most intriguing topics in machine learning. Although a lot of work has been done on classifying images, it needs continuous evolution as thousands of pictures categorized mainly into satellite, medical and scenic images are uploaded to the web based platform like YouTube, Facebook, Twitter and taking out meaningful information out of them remains a challenging task. Image classification is also making ways for video classification and video analysis. Multiple methodologies were critiqued and three categories of images named as scenery, medical and satellite were all segregated into normal image, normal image corrupted by Gaussian noise and noisy image applied to a filter. For training, normal images were used whereas noisy and filtered images were subjected to testing dataset. Features were extracted from the image using Region of interest (ROI) where cancer and scenery and satellite images yielded different attributes and same were used for their classification. Primarily Naïve Bayes, Decision Tree and Random Forest were used as methods to classify these images and results were obtained in terms of accuracy and RMSE (Root mean squared error) which clearly stated Random

forest is best for the study done here and in some noisy images Nave Bayes gave better results but overall Random forest is preferred for classification of images presented in this paper.

Liu et al. (2010) presents the idea of material classification. According to the author, material recognition is an important aspect of visual recognition as we encounter wide variety of materials on a day to day basis and classifying materials can benefit us in many ways such as disease identification, road safety and so on. Finding good features in an image is important than finding object or texture like in other visual recognition objectives. Rather than focussing on high level features like colour and SIFT, low level and middle features such as curvature of edges, histogram of oriented gradient (HOG) along and perpendicular to edges were developed to better understand the material and better classify it. LDA (Latent Dirichlet Allocation) was then used to distribute words as images were already converted to bag of words using quantised features. LDA is upgraded to aLDA (augmented LDA) as features from various dictionaries were concatenated together to better understand the cluster of words which can help in maximum material recognition rate through optimal combination of features. Unlike instance database such as CURET or database with fewer samples per class such as KTH-TIPS2 , Flickr database was developed and used which has close up and object level images of the roughly 10 different materials which helped in better insights to the Naïve Bayes algorithm results which was applied for classification in this research. Following features of material recognition was taken into consideration while executing the aLDA leveraging the Bayesian Framework.

- Colour and Texture
- Micro-texture
- Outline shape
- Reflectance based features.

2.3 Machine Learning for Trash Classification:

Mittal et al. (2016) submitted a paper which introduced a garbage app designed for android which is one of its kind niche studies which takes the foundation step of classifying garbage related images from the real world images of various locations and types. The technique used to classify images involves a machine learning algorithm which is Convolutional Neural Network that has been named as GarbNet which detects the garbage related part from the image automatically. The another benefit of GarbNet is unconstrained images, generally image of the trash needs to be loaded but this algorithm allows to take any picture and it can spot garbage in that picture. GarbNet takes any randomly sized image as input and gives out a segmentation of image which is coarse grained highlighting patches of garbage. GarbNet also is optimized enough to perform with constrained resources and hence it is optimum enough to run on smartphone app. The study introduces GINI dataset which is a garbage specific dataset having geo tagged images taken in real world and GarbNet runs on these GINI dataset images to classify garbage in that image. Detecting garbage as well as marking approximate regions of garbage in an image was the main objective of this research and it was achieved by training the neural network on the training dataset from GINI and testing it on test image. Patches of fixed size were generated from parts of images which did not have only garbage in the picture

were excluded from the entire machine learning process to ensure no wrong information is fed to the model in determining characteristics of garbage. Images were cropped into various sizes to train the model on multiple scale, sizes and context. Stride of 10% as the compressed image size was taken with patch size of 10%, 20%, 40% and 80%. to carry out Poisson-disk image subsampling and then further rotations are done on it so that sample size is big enough to avoid over fitting of the model. One of the already built models called AlexNet was used to pre-train the GarbNet model in this research. 512 and 256 neurons are respectively present in two fully connected layers of networks following 5 convolutional layers comprising of 4096 neurons each. The results obtained were in favor of GarbNet model which was 11 times faster than Nave Bayes sliding window CNN and 6 times faster than normal image processing. The accuracy obtained after optimization of GarbNet was 87.69% with 93.45% specificity.

Yang and Thung (2016) observed trash classification as a niche topic which can be categorized under image classification, for which a new database was created with 400-500 images for each class of trash image. SVM and CNN have been implemented multiple times for image classification but less work has been done on trash classification. One such CNN framework implemented is the AlexNet which won the ImageNet large scale visual recognition challenge. Also, in the 2016 Techcrunch Auto Hackathon, Auto Trash, which was an automatic trash sorting algorithm was implemented exploiting Googles Tensorflow. Previous work has also been done on Bing image search and flickr material database for spotting garbage and garbage identification but spotting garbage and classifying trash are two different areas, single images with white background for each type of trash was collected using Bing API search. Features such as Random scaling , random shearing , random brightness and so on were used to augment the image and data quality. SVM was implemented using SIFT features in the image and CNN was implemented with 11 layers resulting in Non normalized log softmax for 5 classes of images. SVM performed better than CNN giving 63% accuracy whereas CNN gave only 22% accuracy and hence this paper presented optimization of these techniques for trash classification as the future work to be continued.

3 Methodology

3.1 Data Preparation and Pre-Processing

As the data used in this experiment is a collection of trash related images, there needed to be performed some pre-processing on them as to convert the data in the format that can be fed to the machine learning models. The original images were of size 512 x 384 pixel each which if taken at this size will make the processing expensive, the time to process the data would be too long with pc resources. Thus, the images will be compressed to 10% of its original size so that the time taken by the images to be processed is relatively lower. Once parsed, the dataset will contain columns as much as the product of X-dimension, Y-dimension and channel size. The first column will be the name of the category, the second column will be the category code for the category in which that image belongs, third column will be the name of file, fourth will be the path on which the image is stored in the local machine and from the fifth column onwards was the value of R, G and B pixels of the image. While these images are converted into features, the brighter pixels will have more weightage than the normal pixels and number of estimators should always be greater to get a better picture [Pujari (2015)]. readJPEG function from the EBImage

package will be used to extract features from these images and extract the Red, Green and Blue values of every pixel in the image.
Partitioning of data will be done where 75% of data will be contained in the training set and residual 25% will be utilized for testing the performance of the model built. Calibration of index is important part of splitting, processing the data and simplifying the factors in the dataset. Factor Analysis is a data pre-processing activity which lets us know how suitable the features of the dataset are for the machine learning model which is impending to be applied on the data. To do a comprehensive Factor Analysis, dataset will be explored and checked whether Factor Analysis is well suited for the data. Two tests namely KMO test (Kaiser Meyer Olkin Test) and Bartletts test of sphericity will be performed on the dataset to check its eligibility for Factor Analysis.

Kaisen Meyer Olkin Test is a test to validate the sampling adequacy of each variable and whole dataset as well [Kaiser (1974)]. Common variance between variables is considered to calculate the proportion of variance so the lower the proportion, better the data suited for Factor Analysis.

KMO returns values between 0 and 1 which can be interpreted as:

- The values which are in the range of 0.8 to 1 are considered highly preferable
- The values which are less than 0.6 are to be ignored and remedial action should be taken.
- Values nearing zero have large partial correlation than the sum of these correlations.

Bartletts test will be performed to check the significance of data by calculating its p value and degree of freedom for the dataset [Bartlett (1950)]. If the tests display favourable results to perform a PCA on the dataset, PCA will be performed on the dataset to check if the dataset contains any principal components which will further help in performing dimensionality reduction on the dataset. Plots will be used to graphically see the most influential factors in the dataset and combination of how many components can reduce the variance to the minimum in the given dataset [Jolliffe (1972)].

3.2 Modelling

3.2.1 Support Vector Machines

Problems in real world require algorithms and machine learning methods that can handle the complexity of the problems but these algorithms are difficult to analyse and require uncanny knowledge and skills to implement. The Support vector algorithm saves the day as it contains class of complex kernels like RBF (Radial Bias Function), Class Neural Network and polynomial classifiers but at the same time SVM is easier to analyse, for SVM in a high dimensional feature which is connected to input space nonlinearly, corresponds to a linear method but does not require any complex calculations in that high dimensional space. All relevant calculations and computations take place in input space with the help of Kernels [Amari and Wu (1999)].
It can be demonstrated that Optimal Hyperplane, which precisely is the maximum range of separation distance of two classes possesses the minimum capacity. This hyperplane is constructed using a quadratic function which makes it optimized enough to contain

the most optimized patterns lying on its margin. These patterns are named as support vectors that contain classification related information. Support Vector Machines are convenient to use it because of its kernels available for different types of data. With its default implementation, it separates two linearly separable classes on the basis of a hyperplane. This kind of SVM is the LSVM (Linear SVM). All the available training vectors are split into two classes by considering the extremes of a dataset and the hyperplane is selected such that the support vectors are at the minimum distance from the hyperplane. If the hyperplane is not constructed properly, it is not possible to classify the classes in a dataset correctly. So, according to SVMs, only these support vectors are important to classify any class rather than the complete training examples. The distance between the support vectors and the hyperplanes is usually denoted by D+ and D- whereas the margin of the separating hyperplane is the sum of both these distances.
In this situation, expectation of the data to be linearly separable was not there due to it being a multiclass classification problem. In such situations, we can use a function to transform our data into a higher dimensional space. An easy polynomial function can be applied to the data available to transform it into a parabola of data points. But this process can be computationally much expensive to follow and thus a kernel trick can be used in such cases. This involves using a function that takes the vectors in the original space as its input and results into a dot product of the vectors in the feature space. This eventually transforms the vectors in a nonlinear space into a linear space [Foody and Mathur (2004)].

3.2.2 Random Forest

Random Forest is a popular classifier used for multiclass classification, it comprises of n varied trees and randomization is at work at every growing or grown tree. Ballpark figure of every distribution over class of image is labelled as leaf nodes of each tree. Image is classified when it is sent down at every node and tree and aggregated value is calculated at the end of distributions of leaves. Randomization is part of the algorithm in two ways; one is by subsampling the dataset in training partition and by selection of node tests. Sampling strategy plays an important role in the result classification. Millard and Richardson (2015) provided a case study with three aspects which were sample size, spatial autocorrelation and proportions of classes within the training sample. Image Classification through Random Forest has shown sensitivity to factors like proportions of classes, size of sample and characteristics of training data. RF classifications should be replicated for optimising performance and accuracy even when it already is an ensemble approach to regression modelling and classification. Every algorithm has its own advantages and disadvantages.
Advantages of Random Forest include:

1. Can be juxtaposed to SVM and Boosting algorithms with easy to use parameters and it is less sensitive to those parameters.

2. Lesser problem of overfitting compared to individual decision trees and hence pruning of trees can be avoided.

3. Automatic detection of outliers and important variables takes the accuracy higher and hence RF is comparatively easier to use

However, each advantage comes with its own set of limitations as well. Limitation of RF which have been explored as yet is that, because of regression trees, prediction is

restrained till a particular range of response values in training dataset and hence it becomes almost a prerequisite that training data consists of full range of response variables and all samples should have all range of response data values.

3.2.3 eXtreme Gradient Boosting

Any tree traversing algorithm can be boosted using boosting methods and one the most popular of them is XGBoost. The popularity and scalability which XGBoost enjoys is because of the ability to scale up in almost every problem statement. Many optimizations have been done on the algorithm of XGBoost such as,

1. It has been designed to handle sparse data using a unique tree learning method
2. It can handle multiple instance weights in estimated tree traversing due to weighted quantile sketch procedure.
3. Model can be explored faster because of parallel and distributed computing.

These advantages can be combined together to create an algorithm which is scalable in mining massive datasets with least number of resources required for clustering. There is some data processing required for implementing this model. The data accepted by this model needs to be in the form of an XGB matrix which is prepared using a function that considers the data and the category in which the data is to be categorized into. The categorical variable that is used for classification need to be converted into numeric form. Performance of the model will be optimized by tuning the parameters available in the function while error will be recorded using mlogloss [Chen and Guestrin (2016)].

3.2.4 k Nearest Neighbours

Another method which was used to compare the results of best known classifiers is kNN (K Nearest Neighbours). We compared the kNN model to Random Forest, SVM and a boosted approach called XGBoost. Conventional approaches towards kNN classify an image based on k images from the training dataset which resemble the most to the image in question, or check the measure of most similar images comparing it to each image through analysis of class weighted frequency. Alternatively analysis and classification based on local features of an image can be used to perform kNN, these local features such as SIFT (Scale Invariant feature transformation) and SURF (Speed Up Robustness Feature) are generated over interest points. Local features of the images were compared for similarity rather than the whole image which allowed for better and wider strategies to explore [Amato and Falchi (2010)].
The kernel methods for SVM and RVM rely on the Image to Image and Image to class distance methods which can be done on small dataset like we use for our experiment as they generalize the capabilities of classifying an image [Boiman et al. (2008)].

Finally all the models listed above were compared to state of the art Convolutional Neural Network model which was implemented as part of this research as CNN and Deep Learning are considered best machine leaning models when image classification is defined in problem statement.

3.2.5 Convolutional Neural Network

Li et al. (2014) proposed a research titled Medical Image Classification with Convolutional Neural Network with a convolutional layer of kernel size of 7x7 pixels and 16 output channels followed by a pooling layer of 2x2 kernel size and neuron count of 100-50-5. The ReLU activation function improved the performance and the learning rate of their model. [Ciresan et al. (2011)] designed a CNN by not designing a pre-wired connection but considering the learning of data in a supervised manner while experimenting on MINST, NORB and CIFAR 10 images. For different datasets, different structure of layers was implemented. Using the guidelines of previous researches, CNN will be implemented for classification of the image data available. The structure of the network developed is as follows.

1. Convolution Layer: This layer takes care of convoluting the kernel grid across the whole image by applying the filters specified for it. This also acts as an input layer where the input dimensions of the image are to be specified.

2. Max-Pooling layer: This layer takes care of lowering the down sampling and processing time. Over fitting of the model is taken care at this layer by making sure no extra parameters are added to the model.

3. Dropout Layer: This layer drops out the unwanted random set of activations by setting them to 0. This process takes into consideration only the training data and not the testing or validation data.

4. Flattening Layer: This layer is used to convert the dimensions of the layer above it into a single dimension by taking a product of all the dimensions in it.

5. Dense Layer: This acts as an output layer with the number of categories as the units supplied to it with a specific activation function which yields to an optimized result producing model.

4 Implementation

Post preparation of data and converting it into a feature vector, the dataset generated contained 5817 columns and 2390 rows. Each of the values were a result of the R, G and B values in every pixel in the image. Once data was prepared, tests to check the variance and specificity among the features contained inside the data. Implementing KMO test in R required a function called KMO which was directly used in code to perform this test on the dataset. KMO test yielded the value 0.51 which means it is of no use to perform a factor analysis on the data available. Another test for validating the eligibility of Factor Analysis is Bartletts Test for sphericity. This test is used to cite a comparison between identity matrix and correlation matrix (Pearsons Correlations). Both the test indicated unfavourable conditions to conduct Factor Analysis. For this research, due to the unfavourable conditions presented by the Preliminary tests, we conducted a Principal Component Analysis (PCA) but did not use it for any of the data processing. It was only used for analysing the distribution of data and selection of proper kernels for multiple models.

4.1 Models At Work

4.1.1 Support Vector Machines

Without any parameter consideration, a vanilla SVM was implemented using no parameters other than selecting the radial kernel. The selection of radial kernel was decided upon analysing the distribution of data in a plot. The accuracy of this model was considerably lower than it should be. So, in order to increase the accuracy of this model, tuning of SVM was performed and the model was tuned to obtain the best values for Cost and Gamma. Tuning of SVM increased the robustness of the model by yielding the most optimised values for cost and gamma and these values were used for fitting the SVM model on our dataset.

4.1.2 Random Forest

Being the most trustworthy algorithm for any kind of classification, Random Forest classification was used to solve our classification problem. Using the Category code available in our dataset, we classified the image data available with us. As random forest is a tree traversing algorithm and performs best when the number of trees are optimum, multiple values for number of trees was tested starting from 500 to 6000 increasing the number of trees by 500 in every iteration, the maximum accuracy of the model was attained when the number of trees was 5000. Post building the model and predicting its performance against the test dataset, a confusion matrix was created where it displayed an accuracy of 70.1%.

4.1.3 eXtreme Gradient Boosting

XGBoost was used to boost the performance and accuracy of applied Random Forest in this research as it boosts the tree traversing algorithm. It works on tabular and structural multiclass dataset, and is quite popular among machine learning algorithms for boosting the accuracy. One the most important prerequisite to prepare the data for XGBoost is to convert all features from categorical to numeric and in this study, One Hot Encoding was used to convert category code for images into numeric values. This technique uses function to.categorical to convert all variables to numeric. After the encoding is done, parameters are outlined for the model. MultiSoftprob is used as objective parameter which defines the objective of the classification and uses fuzzy clustering logic to estimate the probability for each class and mlogloss is used as Evaluation Metric parameter in this model. Hence XGBoost was implemented after Random Forest and the improvement in accuracy recorded. The booster parameter used for the model was a gbtree and the objective function was a multi:softprob due to the multiclass nature of the problem. The error metric used is mlogloss. The model was run for 1000 rounds, with a 10-fold cross validation and stratified sampling set to true.

4.1.4 k Nearest Neighbour

The simplest form of classification can be said to be an algorithm that classifies objects on the basis of its surrounding objects. K nearest neighbours is one such algorithm that classifies on the basis of its neighbouring objects. Being the simplest to implement, this algorithm was implemented to check if it can be a suitable option to classify such data. Post modelling, it was found that the algorithm performed adequately for such data. So

a tuned version of kNN was used called kkNN which is a kernal specific modification to the regular kNN. In kkNN, we specified an optimal value of 'k' and running it across 10 folds for validation data.

4.1.5 Convolutional Neural Network

Considering the performance aspects of neural networks and the nature of the problem, it was decided to implement a convolutional neural network on the data available, mainly due to its faster execution and efficiency over a regular neural network. A basic neural network would have led to more than billion parameters depending on dimensions considered for the image, however CNN resulted in some million parameters for the same scenario and hence it was deemed as the preferred choice amongst the best-known Neural Networks. To start with, the dataset required by CNN is not the same as we used for other models. It requires a list of all matrices merged together as a vector of data with dimensions of the vector as number of data, dimension over X-axis, dimension over Y-axis and colour scheme of the image. The data and its expected response were split into train and test which was then fed to the network. Due to the constraints of hardware of the available system, we decided to scale down the image set and use only 20% of the images available for each category. They were read using the readJPEG function and the features of the images were utilized without extracting the feature vectors of those images. The splitting of images was done using the data partitioning function and test and train dataset for modelling were generated. In order to categorize the images efficiently, the labels were converted to numerical by the method of One Hot Encoding.

The model used is the sequential model from the keras package in R. A CNN maps the three colour layers present in the images separately. Due to the small size of the image, the kernel size used is relatively small. Activation function used was Rectified Linear Unit activation function with 32 filters in the first 2-dimensional convolution layer providing the input shape as the X and Y dimensions of the image along with the number of channels contained by the image (51, 38, 3). Post convoluting the image at the convolution layer, in order to reduce the down sampling, processing time and over fitting, a pooling layer with the pool size of 2x2 was included. A dropout layer with a rate of 0.25 was included in the network after pooling. Flattening is performed on the data inside the network which is basically converting the 3 dimensions into 1 dimension so that the data can be used in a fully connected neural network. A dropout layer with the rate of 25% is applied every time a layer is built in order to regularize the model with the softmax function at the last dense layer as the output layer with number of classification categories as the units to be regularized into. In order to compile the model, a compiler block is used with categorical_crossentropy as the loss function as we have multiclass categorical values as our output. To optimize the results, a stochastic gradient descent optimizer is used [Karpathy et al. (2014)]. The learning rate applied to this function is 0.01, momentum of 0.9 while the metric used for evaluating the performance was accuracy.
The above built model is then fit using the train data which has all the independent variables and dependent variables from train labels available with us for the images. The iteration count was kept as 60 with a batch size of 32 and a validation split of 20% of the training data.

5 Evaluation

Machine learning algorithms generally work best on massive datasets which are difficult to mine using basic data analysis techniques and hence their performance are measured using multiple parameters. Accuracy, precision, recall, F1-score, ROC and kappa statistic can be used for evaluating any machine learning model and understand how it has performed for analysing the data present with the individual. Usage of these metrics depends upon how the problem statement is to be evaluated. First step involves splitting the data into training and testing data where training data majorly is used to train the model with correct classification outputs and then test data is used to test the performance of model which it has gained on training dataset. A Confusion matrix is then created for each model which carries the correct number of predictions or classification a model can make based on positive and negative scenarios. This matrix along with other statistical parameters of each algorithm can be used to determine the performance of a model.

5.1 Accuracy

This is the most used performance metric in machine learning. It considers all the correctly classified classes without considering the labels available in the dataset.

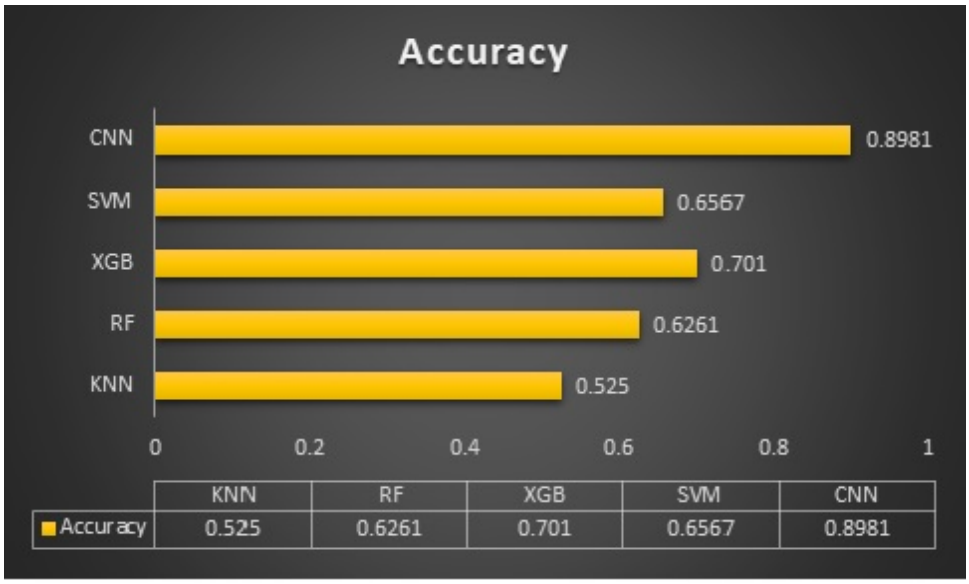

Figure 6: Accuracy Graph for Models

As per the accuracy metric for multiple models plotted in Figure 1, it was found that CNN had the highest performance among all with an accuracy of 89.81%. But considering the limited number of images considered for the experiment, I would like to make it a point to consider XGBoost to be the highest accurate algorithm with an accuracy rate of 70.1% for this experiment. CNN always works leniently with small datasets and we can expect high accuracy from it if it is trained with a small number of images.

5.2 Sensitivity

Sensitivity is the statistical measure of the actual positives after the analysis of data and implementation of model. For example in this study sensitivity is right classification of

trash for a particular category when the trash actually belonged to that category , that is, categorizing and labelling a trash image to plastic when it actually is plastic. Here 5 models were compared and contrasted to find out the best fit to test sensitivity to dataset.

	Cardboard	Glass	Metal	Paper	Plastic
KNN	0.584158416	0.515873016	0.233009709	0.469798658	0.801652893
RF	0.801980198	0.698412698	0.485436893	0.838926174	0.652892562
XGB	0.772575251	0.636138614	0.647509579	0.795309168	0.680672269
SVM	0.683168317	0.563492063	0.398058252	0.899328859	0.652892562

Table 1: Sensitivity

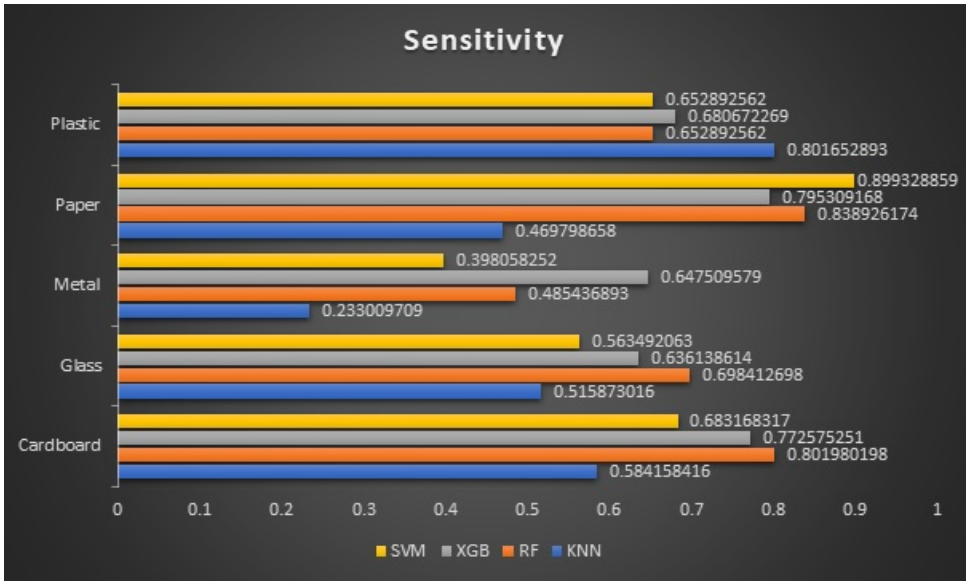

Figure 7: Sensitivity of Models for all categories

Using the graph for Sensitivity of multiple models shown in Figure 2, it can be observed that plastic was the most sensitive to kNN, paper is most sensitive to SVM, metal is most sensitive to XGBoost, Glass is most sensitive to Random Forest and Cardboard is most sensitive to Random Forest.

5.3 Specificity

Specificity on the other hand is the measure where actual negatives are compared to negative values in the dataset which exactly can be explained by the same example as sensitivity, if a material is not plastic and model also categorizes into non-plastic class then model is said to be specific for that particular data value. Both these measures were compared for 5 models in our problem of image classification.

	Cardboard	Glass	Metal	Paper	Plastic
KNN	0.923847695	0.742616034	0.971830986	0.944567627	0.82045929
RF	0.923847695	0.877637131	0.949698189	0.933481153	0.945720251
XGB	0.952380952	0.914862915	0.909744931	0.945495836	0.917655269
SVM	0.96993988	0.890295359	0.927565392	0.820399113	0.954070981

Table 2: Specificity

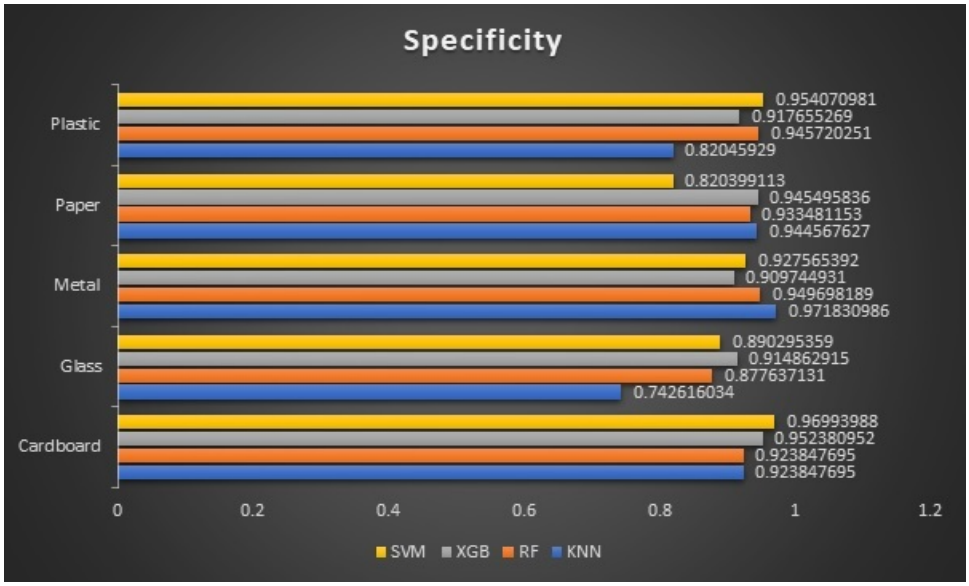

Figure 8: Specificity of Models for all categories

5.4 Precision

Precision of any machine learning algorithm is to proportion of total number of actual positives to total number of true positives and false positives.
Precision = actual positives / (true positives + false positives)

	Cardboard	Glass	Metal	Paper	Plastic
KNN	0.608247423	0.347593583	0.631578947	0.736842105	0.530054645
RF	0.680672269	0.602739726	0.666666667	0.806451613	0.752380952
XGB	0.764900662	0.685333333	0.550488599	0.838202247	0.673130194
SVM	0.821428571	0.577235772	0.532467532	0.623255814	0.782178218

Table 3: Presision

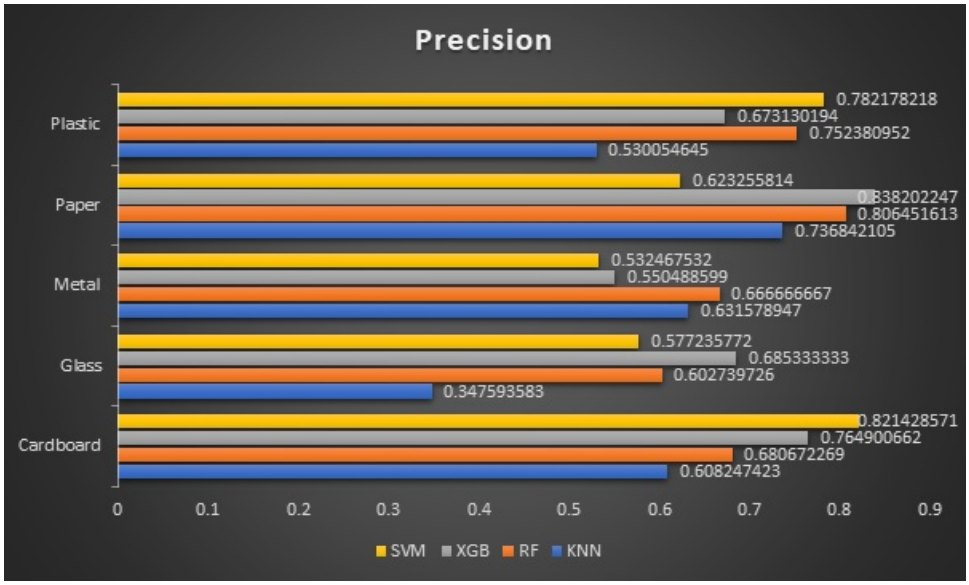

Figure 9: Precision of models for all categories

5.5 Recall

Recall on the other hand is proportion of actual positives figured out after implementation of algorithm to the sum of true positives and false negatives which is Recall = true positive/ (true positives + false negatives)

	Cardboard	Glass	Metal	Paper	Plastic
KNN	0.584158416	0.515873016	0.233009709	0.469798658	0.801652893
RF	0.801980198	0.698412698	0.485436893	0.838926174	0.652892562
XGB	0.772575251	0.636138614	0.647509579	0.795309168	0.680672269
SVM	0.683168317	0.563492063	0.398058252	0.899328859	0.652892562

Table 4: Recall

Figure 10: Accuracy Graph for Models

6 Conclusion and Future Work

According to the research conducted and the results gathered, it can be clearly seen that a Neural network can outperform the performance of almost every model built so far. Boosting any algorithm and validating it with Cross Validation schemes with multiple folds, the performance of any model can be elevated. Extreme Gradient Boosting comes right after CNN in terms of performance and that can be observed from the fact that we applied a booster to a tree traversal algorithm which resulted in better accuracy than a vanilla tree traversing algorithm. Tuning the parameters of Support Vector Machine improvised its overall performance and resulted in more robust and better output. To summarize we can conclude that multiple classification algorithms were tried and tested as part of this research which required algorithm to rightly classify trash into its respective category by taking its image as data and their overall performance was observed with parameters such as accuracy, specificity, sensitivity and so on.

To exploit the full potential of any model, maximum utilization of all features of a dataset is considered as the industry best practice. Without altering the dimensions of images in the dataset, a high performance system could have been made, however lack of hardware and software resources compelled the compression of images to 20% of its original dimensions for implementing almost all models in our research. Specifically for CNN because of its resource heavy prerequisites, we could make use of only 20% of the entire dataset which was further split into train and test data and model performance was only a result of processing CNN on this small subset of data. Thus our suggestion and consideration for future work would involve exploring the capabilities of models implemented on a fully functional resourceful system which can be used for stress testing of these models without loosing its performance and robustness. Further, we suggest to implement the trash classification model for a multilabel image as this is more complicated and require more of time and effort.

7 Acknowledgement

I am highly grateful to my supervisor Prof. Noel Cosgrave for his unrelenting support and encouragement. I would also like to acknowledge the contribution of National College of Ireland for providing me with all relevent skills to carry out my research. I would like to take some time to acknowledge Yang Mindy and Thung Gary who collectively put their efforts and created a publicly available dataset of images for research. last but not the least, I would like to thank my parents, friends and my colleague Miss Urvashi Goyal in supporting me, mentoring me in working up with things correctly and believing in me that I could fall true to their expectations.

References

Amari, S.-i. and Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions, *Neural Networks* **12**(6): 783–789.

Amato, G. and Falchi, F. (2010). kNN based image classification relying on local feature similarity, ACM Press, p. 101.
URL: *http://portal.acm.org/citation.cfm?doid=1862344.1862360*

Bartlett, M. S. (1950). Tests of significance in factor analysis.

Boiman, O., Shechtman, E. and Irani, M. (2008). In defense of Nearest-Neighbor based image classification, IEEE.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System, ACM Press, pp. 785–794.

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M. and Schmidhuber, J. (2011). Flexible, High Performance Convolutional Neural Networks for Image Classification, p. 6.

Dash, S. K. and Mrutyunjaya Panda (2016). Image Classification using Data Mining Techniques.

Foody, G. and Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines, *IEEE Transactions on Geoscience and Remote Sensing* **42**(6): 1335–1343.

Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. i: Artificial data, *Applied statistics* pp. 160–173.

Kaiser, H. (1974). An index of factor simplicity. psychometrika.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks, *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725–1732.

Krizhevsky Alex, Sutskever, I. and E, H. G. (2012). Imagenet classification with deep-convolutional neural networks.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D. and Chen, M. (2014). Medical image classification with convolutional neural network, pp. 844–848.

Liu, C., Sharan, L., Adelson, E. H. and Rosenholtz, R. (2010). Exploring features in a Bayesian framework for material recognition, IEEE.
URL: *http://ieeexplore.ieee.org/document/5540207/*

Millard, K. and Richardson, M. (2015). On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping, *Remote Sensing* **7**(7): 8489–8515.

Mittal, G., Yagnik, K. B., Garg, M. and Krishnan, N. C. (2016). SpotGarbage: smartphone app to detect garbage using deep learning, ACM Press, pp. 940–945.

Pujari, A. (2015). Retail product image classication, p. 14.

Sun, M. (2007). Bayesian network for image classification, p. 31.

Uvarov, N. (2017). Multi-label classification of a real-world image dataset, p. 87.

Yang, M. and Thung, G. (2016). Classification of trash for recyclability status, pp. 1–6.