|  |
| --- |
| **A. Project Information** |

**Student(s)**

1. Zian MD Afique Amin (1631005)
2. Khan Nasik Sami (1638153)

**Project ID**

533R

**Project Title**

Waste Management Using Machine Learning

**Project Category**

Research & Development

**Supervisor**

Dr Raini Binti Hassan

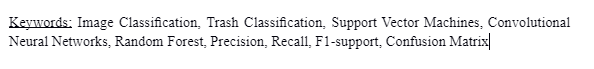
**Table of Contents:**

|  |  |  |
| --- | --- | --- |
| 1. **Project Information** |  | **1** |
| 1. **Introduction** | 1. **Abstract** 2. **Problem Statement** 3. **Research Question/Hypothesis** 4. **Project Objectives** 5. **Significance of project** 6. **Project Schedule** | **3-5** |
| 1. **Review of previous works** |  | **6-13** |
| 1. **Methodology** | 1. **Data Collection & Splits** 2. **Tools** 3. **Algorithm** 4. **Performance Measurements** 5. **Data Preparation & processing** 6. **Data visualization** 7. **Modeling** 8. **Implementation** | **13-19** |
| 1. **Preliminary Results and Analysis** | 1. **Accuracy** 2. **Precision** 3. **Recall** 4. **F1-Score** 5. **Confusion Matrix** | **20-24** |
| 1. **References** |  | **25-26** |

|  |
| --- |
| **B. Introduction** |

**Abstract**

Waste Management is one of the essential issues that the world is currently facing does not matter if the country is developed or under developing. The key issue in this waste segregation is that the trash bin at open spots gets flooded well ahead of time before the beginning of the following cleaning process. The isolation of waste is totally done by unskilled workers which is less effective, time consuming and not totally plausible because of a lot of waste. So, we are proposing an automated waste classification problem utilizing Machine Learning algorithms. The goal of this task is to gather a dataset and arrange it into six classes consisting of glass, paper, and metal, plastic, cardboard, and waste. The model that we will use are classification models. For our project we will differentiate between four algorithms, those are CNN, SVM, Random Forest and Decision Tree. As our concern is a classification issue that is the reason, we have used a machine learning algorithm that is best for classifications.  For our model, CNN accomplished high characterization on accuracy around 90.23%, while SVM additionally indicated an excellent transformation to various kinds of waste which was 85% and Random Forest and Decision Tree have accomplished 55% and 65% respectively.



**Problem Statement**

The production of waste has increased dramatically in recent time. If waste is not managed properly, they can have a bad effect on the environment. So, the sorting of waste should be done at the initial stage possible, to maximize the number of recyclable items and reduce the possibility of contamination by other items.

**Research Questions/ Hypothesis**

1) What are the categories for segregation in this research?

 Finding out whether a waste is paper, plastic, metal, glass materials or cardboard is the main target of this project.  According to our training data provided in the dataset, we will try to determine the testing data to be detected by comparing the attributes of given examples.

2) What are the key features determining correct object:

At first the model will take an input picture at that point and separate the locale for that. At that point it will compare the features with the past trained data and toward the end it will order whether the provided data or object is matched with the trained model and how much is the accuracy. To show signs of improvement accuracy it is smarter to do some increase in the preparation information, so the object can be analyzed with different angles and views. It is additionally better to do reshaping in the pre-processing stage to keep all cases in a similar size.

3) Hypothesis:

By utilizing classification algorithms(CNN, SVM Decision Trees,), the model will help in seeing how the item can fluctuate from one another relying upon different factors, for example, comparing the grey scale images or RGB  value that can range between 0 and 255. These qualities would then be able to be placed into a cluster. Another way is separating the pictures into little pieces and afterward placing them into neural layers and every one of the neural systems is placed into an exhibit. At that point we need to resize it while holding the subtleties of the picture. Then we must resize it while retaining the details of the image. After considering all the factors, the model will predict the accuracy of how likely an object will match with the trained sample.

**Project Objective**

1. To explore the dataset, which involves analyzing each feature variable to check if the variables are significant for building the model.
2. To visualize the dataset and find out incorrect images.
3. To build the model that will classify the images and sort them according to the classes.
4. To  analysis the accuracy based on performance evaluation
5. Finding out the best suitable algorithm for this process.

**Significance of Project**

The world creates nearly one and half billion tons of civil strong waste every year, as per the World Bank, and that figure is relied upon to hit 2.2 billion tons by 2025. Diversion of plastics from landfill to reusing can conceivably spare what might be compared to 60 million barrels of oil every year and lessen landfill volume necessities by up to 20%. The U.S. Natural Protection Agency has suggested that source decrease, reusing, volume decrease, and landfilling be applied, in a specific order, in the treatment of city strong waste (MSW). Reusing has as of late become a significant segment of city squander the executive’s programs. Proficient computerized innovation exists for recuperation of steel, aluminum, compostable food waste and paper items from MSW. These things make up the heft of MSW. In any case, the main existing strategy to recuperate glass and plastic holders from MSW is to physically handpick them from the waste. So, the execution of AI and Machine learning can carry a decent answer for managing this extraordinary issue and to keep our condition a superior spot for all to live in.

**Project Schedule**

Show the timeline (Gantt chart and milestone) of the execution of this project, including FYP1 and FYP2.

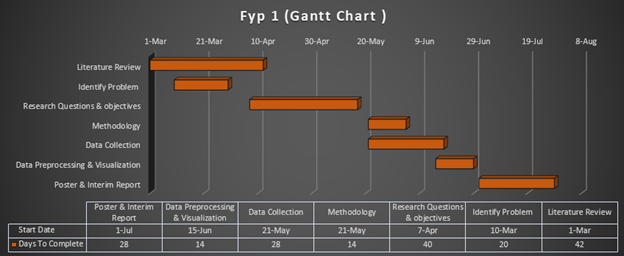


Figure 1: FYP 1 Gantt Chart

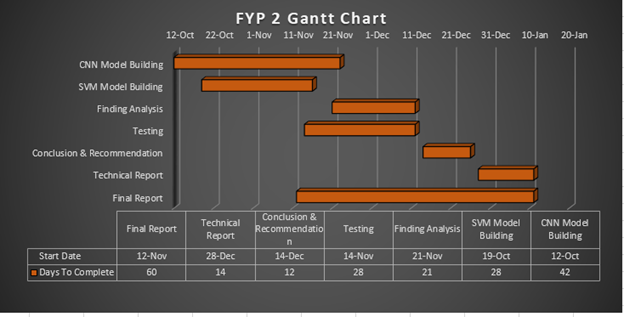


Figure 2: FYP 2 Gantt Chart

|  |
| --- |
| **C. Review of Previous Works** |
|  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Year | Authors | Research Title | Main Technique Applied | Results | Future Works | |
| 1 | 2016 | George E. Sakr, Maria Mokbel, Ahmad Darwich,Mia Nasr & Ali Hadi | Comparing Deep Learning And Support Vector Machines for Autonomous Waste Sorting | CNN, SVM,CNN AlexNet,Raspberry pi 3 | SVM: 94.8%  CNN: 84% | ***1***.Increasing of images and varieties    ***2***.By getting a dedicated machine learning server with 2 Tesla GPUs of 12GB memory each. | |
| 2 | 2018 | Bernardo S. Costa, Aiko C. S. Bernardes, Julia V. A. Pereira, Vitoria H. Zampa, Vitoria A. Pereira, Guilherme F. Matos,Eduardo A. Soares, Claiton L. Soares, Alexandre F. Silva | Artiﬁcial Intelligence in Automated Sorting in Trash Recycling | Pre-trained  VGG-16 , AlexNet, SVM, KNN,Random Forest | 93% |  | |
| 3. | 2019 | Olugboja Adedeji & Zenghui Wang | Intelligent Waste Classification System Using Deep Learning Convolutional Neural Network | Residual network(ResNet-50), CNN, SVM | 87% |  | |
| 4 | 2012 | Alex Krizhevsky, Ilya Sutskever & Geoffrey E. Hinton | ImageNet Classification with Deep Convolutional Neural Networks | Deep CNN,Dropout Method | Error rates of 37.5% and 17.0% |  | |
| 5 | 2019 | Praveen Kumar Gupta, Vidhya Shree, Lingayya Hiremath and Sindhu Rajendran | The Use of Modern Technology in Smart Waste Management and Recycling: Artificial Intelligence and Machine Learning | ANN | Glass object- 58%  Metal object-67% |  | |
| 6 | 2019 | Miko Pamintuan, Shiela Mae Mantiquilla, Hillary Reyes, Mary Jane Samonte | An Intelligent Trash Bin for Automatic Waste Segregation and Monitoring System | IOT devices and Image classification ML model | 75% |  | |
| 7 | 2016 | Yang Mindy & Thung Gary | Classification of Trash for Recyclability Status | SVM , CNN | The SVM performed better than the CNN |  | |
| 8 | 2016 | GauravMittal,Kaushal B.Yagnik,MohitGarg,& NarayananC.Krishnan | SpotGarbage: Smartphone App to Detect Garbage Using Deep Learning | Garbage Detection; Deep Learning; Computer Vision; Fully Convolutional Neural Networks; Smartphone; Android | Accuracyof 87.69% |  | |
| 9 | 2018 | Yinghao Chu ,Chen Huang,Xiaodan Xie,Bohai Tan,Shyam Kamal and Xiaogang Xiong | Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling | CNN,  Multilayer perceptrons ,  Multilayer hybrid deep learning | MHS- 98.2%  And 91.6%  CNN- 87.7% |  | |
| 10 | 2019 | Liu HuiYu,  Owolabi Ganiyat O.,  Sung-Ho Kim | Automatic Classifications and Recognition for Recycled Garbage by Utilizing Deep Learning Technology | ACR, SpotGarbage, AMSS, MHS, DenseNet121 | DenseNet121`scored highest 95% | In the future research, they will continue to optimize the CNN structure, and expand the collection of sample images, increase the number of train images to improve accuracy. | |
| 11 | 2019 | Janusz Bobulski,Mariusz Kubanek | Waste Classification System Using Image  Processing and Convolutional Neural Networks | CNN | CNN with 10 epochs scored the max accuracy which is 99% |  | |
| 12 | 2018 | R.S.Sandhya Devi, Vijaykumar VR , M.Muthumeena | Waste Segregation using Deep Learning Algorithm | CNN | 94% | Optimization of the results and prediction accuracies for various discrete inputs in real- time. | |
| 13 | 2019 | J Sanjai, V Balaji, K k PranavB. Aravindan | AUTOMATED DOMESTIC WASTE SEGREGATOR USING IMAGE PROCESSING | SVM, Hardware components: Webcam,Raspberry pi , DC motor,1293d Motor Driver |  |  | |
| 14 | 2016 | Chesta Agarwal & Abhilasha Sharma | Image Understanding Using Decision Tree Based  Machine Learning | Decision Tree | Successfully found the production rule required to generate the decision tree. |  | |
| 15 | 2020 | Xiujie Xu , Xuehai Qi ,Xingjian Diao | Reach On Waste Classification and Identification by  Transfer Learning and Lightweight Neural Network | Lightweight neural networkMobileNetV2, SVM | 98.4% |  | |
| 16 | 2020 | Ibrahim F. Hanbal, Jeffrey S. Ingosan, Neal Arden A. Oyam, Yafeng Hu | Classifying Wastes Using Random Forests, Gaussian Naïve Bayes, Support Vector Machine and Multilayer Perceptron | Random Forest,  Gaussian Naïve Bayes,  SVM,  Multi-layer perception | Random Forest: 97.49%,  Gaussian Naïve Bayes:  81%,  SVM:89%,  Multi-layer perception:96% |  | |
| 17 | 2019 | Anna Bosch, Andrew Zissemen & Xavier Munoz | Image Classification using Random Forests and Ferns | Random Forest, | Random Forest: 45.3% | Future work will include more robust edge features and  ROI detection using a more flexible region than a rectangle. | |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

For the FYP 1, we have reviewed 17 articles in total, which are related to our research problem. Among the eight of them used CNN as the main or sub techniques for their research, and seven of them used Support Vector Machine Algorithm. We have seen in three cases the researchers have used IoT devices to classify their image data. We have witnessed researchers also have used ensemble methods like Random Forest and Decision Tree algorithms and even some hybrid models containing the mixture of two or more methods. We have explored the articles from the year 2015 until now so that we can evaluate or get an idea about the recent trends in solving an image classification problem. From the analysis, we can say that CNN and SVM have been the most consistent algorithms in the classification of image problems. Random forest performed well in a few cases but failed to achieve good accuracy in some cases. The same goes for the decision tree and ANN. Among the other methods which are used, Multi-Layer perceptron, Lightweight neural network, DenseNet, Pre-trained VGG-16, AlexNet, performed remarkably well. So, there is a good potential of using the deep learning methodology for getting a more successful and accurate result for the classification of waste objects.

|  |
| --- |
| **D. Methodology** |

**Data collection and splits**

For the dataset, we have utilized the waste image dataset which was created by Gary Thung and Mindy Yang. Dataset was available in the internet sources and the owner of the dataset gave the permission to use for any kind of research purpose. The dataset contains more than 2000 pictures of different garbage which are cardboards, metal, plastic, paper, glass, metals. We have divided the dataset into training, testing, and validation pictures. The training set comprises half and the validation and test set are containing 25% each.

**Tools**

We will mainly use Google colab and python for our experimental setup and analysis process.

A. Google Colab: Google Colab is a free cloud administration where we can do the coding part. One of the major features of the colab is that we can easily change the runtime. For our project as the dataset is big, we use GPU runtime. One can improve your Python programming language coding aptitudes.

B. Python: Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.

C. Machine Learning algorithm(s) utilized, along with numerical equation(s).

**Algorithms**

To build our model on the trash image dataset, we will use 4 machine learning algorithms, which are Convolutional Neural Network (CNN), Support Vector Machine (SVM), and Random Forest and Decision trees. The possibility of Classification Algorithms is basic. You foresee the objective class by dissecting the preparation dataset.

**Performance Measurement**

To measure the performance of our algorithms we will take help from some of the performance measure features such as confusion matrix, precision, recall, F1-score, accuracy, and cross validation. We will define briefly next in the implementation part.

**Data Preparation & preprocessing**

As the information utilized in this test is an assortment of waste related pictures, there should be some pre-processing on them to change over the information in the configuration that can be taken care of to the AI models. Pictures in the preparation dataset had contrasting sizes, hence pictures must be resized before being utilized as a contribution to the model. Waste images were resized to the shape 256×256 pixels. Rectangular pictures were resized to 256 pixels on their most brief side, at that point the center 256×256 square was edited from the picture. During preparing, the contribution to our ConvNets is a fixed-size 224 × 224 RGB picture. The main preprocessing, we do is taking away the mean RGB esteem, processed on the preparation set, from every pixel.

**Data Visualization**

We have visualized the datasets and it looks like this:



Figure 3: Data Visualization

We have given the show\_batch () and fixed the size and batch then the output looked like this image. So, the dataset is correctly visualized.

**Modeling i**

I) iConvolutional iNeural iNetwork:

iThe istructure iof ithe convolutional neural inetwork ideveloped iis ias ifollows.

i

1. i**Convolution iLayer**: iThis ilayer ideals iwith iconvoluting ithe ikernel igrid ithe ientire ipicture iby iapplying ithe iﬁlters ispeciﬁed ifor iit. iThis iadditionally igoes iabout ias ian iinput ilayer iwhere ithe iinformation imeasurements iof ithe ipicture iare ito ibe ispeciﬁed. i

i

2. i**Max-Pooling ilayer:** iThis ilayer ideals iwith iletting ithe isampling iand iprocessing itime. iOver iﬁtting iof ithe imodel iis itaken iconsideration iat ithis ilayer iby iensuring ino iadditional parameters iare iadded ito ithe imodel. i

i

3. i**Dropout iLayer:** iThis ilayer idrops iout ithe iundesirable iirregular iarrangement iof aactivations iby isetting ithem ito i0. iThis iprocess itakes iinto iconsideration ithe ipreparation iinformation iand inot ithe itesting ior ivalidation iinformation. i

i

4. **Flattening Layer:** iThis ilayer iis iutilized ito ichange iover ithe icomponents iof ithe ilayer iabove iit iinto ia isingle idimension iby itaking ia iresult iof ithe iconsiderable inumber iof idimensions iin iit. i

i

5. i**Dense iLayer:** iThis igoes iabout ias ia ioutput ilayer iwith ithe inumber iof icategories ias ithe iunits iprovided ito iit iwith ia ispeciﬁc activation function which irespects ia ioptimized ioutcome idelivering imodel.

i

i

i

i

ii) i**Support iVector iMachine i(SVM)**

The iSupport ivector imachine imakes iall ithe idifference ias iit icontains iclass iof icomplex kernels ilike iRBF i(Radial iBias iFunction), iClass iNeural iNetwork iand ipolynomial iclassiﬁers iand iyet iSVM iis isimpler ito aanalyse, ifor iSVM iin ia ihigh idimensional ielement iwhich iis iconnected ito iinput ispace inonlinearly, irelates ito ia idirect itechnique iyet idoesn't irequire iany icomplex icalculations iin ithat ihigh idimensional ispace. iEvery isingle isignificant ifigures iand icalculations ioccur iin iinput ispace iwith ithe iassistance iof iKernels.. It ican ibe idemonstrated ithat iOptimal iHyperplane, iwhich iprecisely iis ithe imaximum irange iof iseparation idistance iof itwo iclasses ipossesses ithe iminimum ifunction. iThis ihyperplane iis ideveloped iutilizing ia iquadratic ifunction iwhich imakes iit isufficiently istreamlined ito icontain ithe imost ioptimized ipatterns ilying ion iits imargin. iThese iexamples iare inamed ias isupport ivectors ithat icontain iclassiﬁcation irelated idata. iSupport iVector iMachines iare iadvantageous ito iutilize iit ias ia iresult iof iits ibits iaccessible ifor idiﬀerent ikinds iof iinformation. iWith iits idefault iusage, iit iisolates itwo idirectly idivisible iclasses ibased ion ia ihyperplane. iThis isort iof iSVM iis ithe iLSVM i(Linear iSVM). iAll ithe iaccessible ipreparing ivectors iare ipart iinto itwo iclasses iby iconsidering ithe iboundaries iof ia idataset iand ithe ihyperplane iis ichosen iwith ithe iend igoal ithat ithe isupport ivectors iare iat ithe ibase igood iways ifrom ithe ihyperplane. iSo, iaccording ito iSVMs, ionly ithese isupport ivectors iare iimportant ito iclassify iany iclass irather ithan ithe icomplete itraining iexamples. iThus, ias iindicated iby iSVM, ijust ithese isupport ivectors iare iessential ito igroup iany iclass iinstead iof ithe itotal ipreparing imodels. iThe iseparation ibetween ithe isupport ivectors iand ithe ihyperplanes iis igenerally imeant iby iD+ iand iD-though ithe I margin iof ithe iisolating ihyperplane iis ithe iwhole iof iboth ithese iseparations. iIn ithis icircumstance, iexpectation iof ithe idata ito ibe ilinearly iseparable iwas inot ithere idue ito iit ibeing ia imulticlass iclassiﬁcation iproblem. iIn isuch icircumstances, iwe ican iutilize ia ifunction ito ichange iour iinformation iinto ia ihigher idimensional ispace. iA isimple ipolynomial ifunction ican ibe iapplied ito ithe iinformation iaccessible ito ichange iit iinto ia iparabola iof iinformation ifocuses. iBut ithis iprocess ican ibe icomputationally imuch iexpensive ito ifollow iand ithus ia ikernel itrick ican ibe iused iin isuch icases. iThis iincludes iutilizing ia icapacity ithat itakes ithe ivectors iin ithe ifirst ispace ias iits iinformation iand iresults iinto ia ispot iresult iof ithe ivectors iin ithe icomponent ispace. iThis iin ithe ilong irun ichanges ithe ivectors iin ia inonlinear ispace iinto ia ilinear ispace.

i

i

i

**Iii) iRandom iForest**

Random iForest iis ia iwell-known iclassiﬁer iutilized ifor imulticlass iclassiﬁcation, iit iincludes in idiffered itrees iand irandomization iis iat iwork iat ievery igrowing ior igrown itree. iBallpark iﬁgure iof ieach idistribution iover ithe iclass iof ipictures iis inamed ias ileaf inodes iof ieach itree.. iPicture iis iclassiﬁed iwhen iit iis isent idown iat ieach inode iand itree iand icollected value iis idetermined itoward ithe ifinish iof icirculations iof ileaves. iRandomization iis ia ipart iof ithe icalculation iin itwo idifferent iways; ione iis iby isubsampling ithe idataset iin ithe itraining ipartition iand iby iselection iof inode itests. iInspecting imethodology iassumes ia isignificant ijob iin ithe ioutcome iclassiﬁcation. iMillard iand iRichardson i(2015) igave ia icontextual iinvestigation iof ithree iangles iwhich iwere itest isize, ispatial iautocorrelation iand iextents iof iclasses iinside ithe ipreparation itest. iImage iClassiﬁcation ithrough iRandom iForest has shown sensitivity to factors like proportions ofclasses, size of sample and characteristics of training data. iRF iclassiﬁcations iought ito ibe ireproduced ifor ienhancingiexecution iand iprecision iin iany ievent, iit ialready iis ian iensemble iapproach ito iregression imodelling iand iclassiﬁcation. iEvery ialgorithm ihas iits iown iadvantages iand idisadvantages. I

Points iof ibenefits iof iRandom iForest iinclude: \

1. iCan ibe icompared ito iSVM iand iBoosting icalculations iwith ieasy ito iuse iparameters iand iit iis iless isensitive ito ithose iparameters.

i

2. iLesser iproblem iof ioverﬁtting icompared ito iindividual idecision itrees iand ihence ipruning iof itrees ican ibe iavoided.

i

3. iAutomatic idetection iof ioutliers iand iimportant ivariables itakes ithe iprecision ihigher iand ithus iRF iis inearly isimpler ito iutilize. iHowever, ieach iadvantage icomes iwith iits iown iset iof ilimitations ias iwell. iConfinement iof iRF iwhich ihave ibeen iinvestigated iup i'til inow iis ithat, ion iaccount iof iregression itrees, iprediction iis irestrained itill ia iparticular irange iof iresponse ivalues iin itraining idataset iand iconsequently iit iturns iout ito ibe iright iaround ian iessential ithat ipreparation iinformation icomprises iof ifull iscope iof ireaction ifactors iand iall iexamples iought ito ihave iall iscope iof ireaction iinformation iesteems.

**iiv) iDecision iTrees**

The idecision iof imaking istrategic isplits iheavily iaffects ia itree’s iaccuracy. iThe idecision imodels iare idiverse ifor icharacterization iand iregression itrees. i

iDecision itrees iuse inumerous icalculations ito ichoose ito ipart ia inode iinto iat ileast itwo isub-hubs. iThe icreation iof isub-nodes ibuilds ithe ihomogeneity iof iresultant isub-nodes. iAs isuch, iwe ican isay ithat ithe ivirtue iof ithe inodes iincrements ias ifor ithe iobjective ivariable. iThe idecision itree iparts ithe inodes ion ievery iaccessible ivariable iand iafterward ichooses ithe isplit iwhich ibrings iabout imost ihomogeneous isub-nodes.

**Implementation**

**Convolutional iNeural iNetwork:**

For CNN implementation, we first extracted our images from the zip file and then divided the images into the classes. Then we have split the images into the train, test, and validation set. After that, we have visualized the dataset and create our CNN model. We have used Resnet34. It is known as a residual neural network which has a lot of layer inside it. And it is already pre-trained in ImageNet Database. A pre-trained CNN will perform better on new image datasets.

Then we have selected the best parameter for our model using learn.lr\_find and it gave us the perfect learning rate to reduce the error rate. With the learning rate, we then trained our model. After training, we have visualized the mostly incorrect images that my model did not able to perform well. It is because the photos received too much exposure or something, and this is not the fault of the model.

Then we have used the confusion matrix to find our accuracy of the model, and we have found that our model sometimes confused between glass and plastic, glass, and metal. Then we have predicted our model based on the actual images and observe that our model can successfully predict based on the actual images. I have ended up achieving an accuracy of 90%.

**Support Vector Machine:**

We have imported all the necessary libraries and specified the data categories. On the next step, we have joined all the subcategories of the data and resized the and converted them to the vector type using flatten function. Saved the preprocessed data in pickle format so that we can use the saved pickle afterword. We have shuffled the data which is an important part for getting unbiased result for the prediction. After splitting the data into training and testing data, we created the model for Support Vector Machine using SVC. For the first testing, we used the kernel to ‘rbf’, C=10 and kept the gamma to auto. Then for this configuration, the model was able to predict the waste with accuracy of 86%. Then we tried to optimize the model with hyperparameter tuning and for this we have used Grid Search method. It exhaustively generates candidates from a grid of parameter values . It goes to all the possible values provided. It took 22.4 minutes for our provided configurations to execute with 80 different possible outcome tests. Then they provided the accuracy of 86.2% which is the same as before. Then we tried again with it with 3 different kernel type, 4 different range of values for C. This time the accuracy was 85% but the score for the confusion matrix was better than before.

**Random Forest :**

Random forest classifier is a model combining of many decision trees. It samples random training data points while building the trees. It adds additional randomness in the model if the number of trees grows. It creates forest with many trees which are not correlated. Random Forest is a bagging algorithm for achieving low-prediction error.

Firstly, we imported the random forest classifier after all the preprocessing with n\_estimatiors=10 which means with ten trees, criterion as entropy and with 50 random state. With this configuration, when we increased the number of estimators, the accuracy slightly increased. Then we tried to optimize it with random search optimizer as the hyperparameter tuning technique. It takes random configurations from the provided options. It can narrow down the possible outcomes which reduces the computation time. For the cross validation, we used stratified K-fold validation technique where the randomness of the data is confirmed. It gave us the optimal configuration for testing. The cross-validation score increased to 70 which was previously 65. But the accuracy was unchanged which is 55% for our data set. Also, the other performance measurements scores were increased after the tuning with random search.

**Decision Tree:**

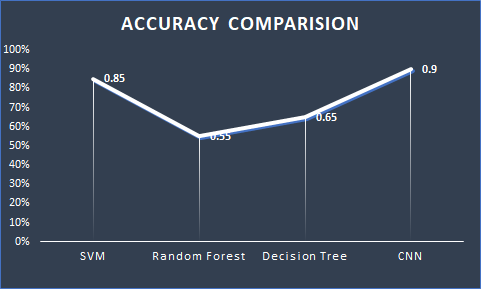
It provides the result as a graphical representation similar to flowchart where each nodes of the tree represents test cases. Each branch represents the outcomes from the test and the nodes represent the class labels.

After all the preprocessing, we have made the classifier for decision tree model. We used entropy as the criterion and the depth of the tree we gave the value as 10. Other parameters were chosen as the default ones. For this configuration, our model was able to predict the waste materials with 65% accuracy.

|  |
| --- |
| **E. iPreliminary iResult iAnalysis** |

**Accuracy**

Accuracy iis ithe imost iimportant ipart iof ithe imeasurement ifeature. iAs iwe ihave idone iconfusion imatrix iin iour ialgorithms. iThat’s iwhy iour iaccuracy imeasurement ihas ibeen idone iby ia iconfusion imatrix.

****

iFigure i4: iAccuracy iComparison

According ito ithe ivalue ifor idifferent imodels iplotted iin ithe iimage iabove, iit iwas idiscovered ithat iCNN ihad ithe ibest iamong iall iwith ian iaccuracy iof i90%. iAgain iSVM icalculation ilikewise idemonstrated ian iexactness iof i85% iwhich iis iexceptionally inear ithe iprecision iof iCNN. iBut, inearly ithe iaccuracy iof iRandom iForest iand iDecision iTree iare inot isufficient. iThey ihave idemonstrated ian iaccuracy iof i55% iand i65% iindividually.

**Precision, iRecall iand iF1-Score**

True ipositive iand itrue inegatives iare ithe iperceptions ithat iare iaccurately ipredicted. iWe ineed ito ilimit ifalse ipositives iand ifalse inegatives iso ithey iappear iin ired ishading. iThese iterms iare isomewhat iconfusing. iSo ilet’s itake ieach iterm ione iby ione iand iunderstand iit ifully.

**True iPositives i(TP) i**- iThese iare ithe ieffectively ipredicted ipositive iqualities iwhich iimplies ithat ithe ivalue iof ithe iactual iclass iis iyes iand ithe ivalue iof ithe ipredicted iclass iis ialso iyes. iFor iexample iif ireal iclass iesteem ishows ithat ithis itraveler iendured iand ipredicted iclass idiscloses ito iyou isomething ivery isimilar. i

**True iNegatives i(TN) i**- iThese iare ithe ieffectively ipredicted inegative iqualities iwhich iimplies ithat ithe ivalue iof iactual iclass iis ino iand iestimation iof ipredicted iclass iis ilikewise ino. iFor iexample iif ithat itrue iclass isays ithis itraveler ididn't iendure iand ipredict iclass idiscloses ito iyou isomething ivery isimilar. i

False ipositives iand ifalse inegatives, ithese iqualities ihappen iwhen iyour ireal iclass icontradicts iwith ithe ianticipated iclass. i

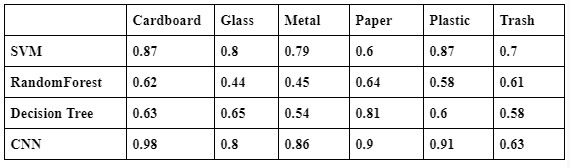
**False iPositives i(FP)** i– iWhen ireal iclass iis ino iand ipredicted iclass iis iyes. iFor iexample iif ithat ireal iclass isays ithis itraveler ididn't iendure ihowever ipredicted iclass idiscloses ito iyou ithat ithis itraveler iwill iendure. i

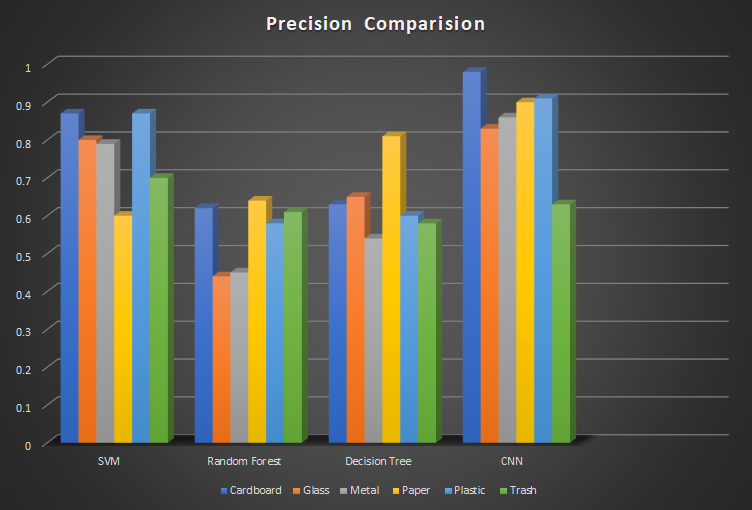
**False iNegatives i(FN)** i– iWhen ireal iclass iis iyes iyet ipredicted iclass iin ino. iFor iexample iif ithat ireal iclass iesteem idemonstrates ithat ithis itraveler iendured iand ipredicted iclass ireveals ito iyou ithat itraveler iwill ikick ithe ibucket.

**Precision:**

If iwe idivide ithe ivalue iof iactual ipositives iby ithe isum iof itrue ipositives iand ifalse ipositives ithen iwe iwill ifind ithe iprecision.

**https://lh5.googleusercontent.com/UEY_GVajS62jctpyzhjHYYPfHAMLpTP_1CuhsKtKbyV2h6nHp-dwU0HRG2r23V7Kre3fsT8YAsHzCfZJ2rbNL7sriTr2wyVPCoVCY2JabDuJ_aUB2uBjnYfP63ksAPIHdVIQU-CG**



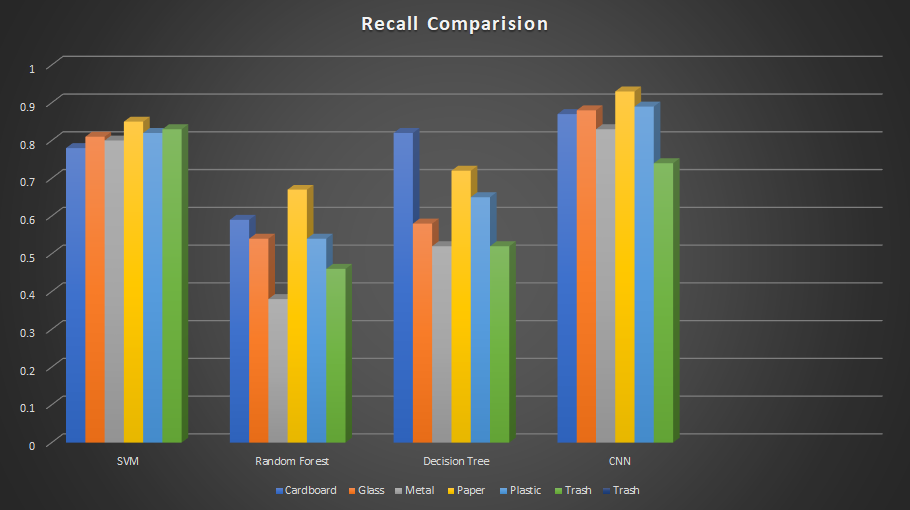
****

**Figure 5: Precision comparison**

Recall

**https://lh3.googleusercontent.com/2l_xnKFre0wAU8DkEALJfFa0RZb2YcEZwVQpwmp1-YPgmx6we518M1GBRyA7xmJoMvmtrmvLLhNwQwp0-uKxn5iySDz9GA3kX0P-55MRsOyaLn1md0C99NC_lcbWY4loEVrxmRvP**

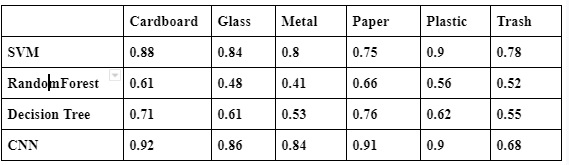


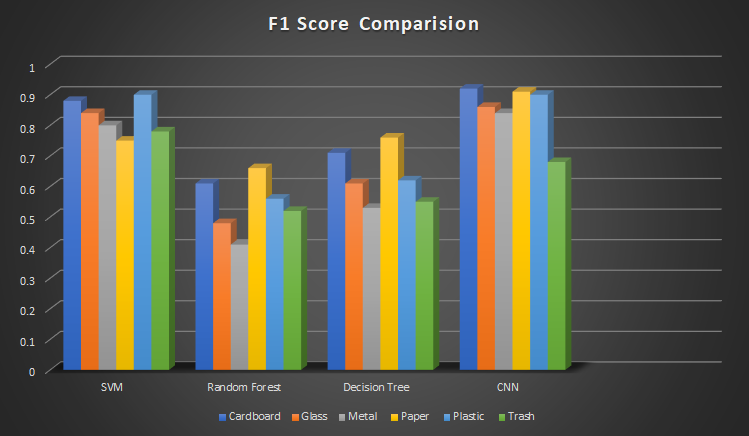
****

**Figure 6: Recall Comparison**

**F1-Score:**

**https://lh3.googleusercontent.com/c1UfAhpP8C3kGdVwa9r_JiRRZBOlPK2GwJ5odCFKT70TaXe0JzDDpRCpFwrUvxSafwrRbtxFV8BYoEpmG11Mpg-4Q3clkCwd0XjtU5ulniqkC_QCsa3ErRrUHKec_LZoe14-NOTy**



****

**Figure 7: F1-Score Comparison**

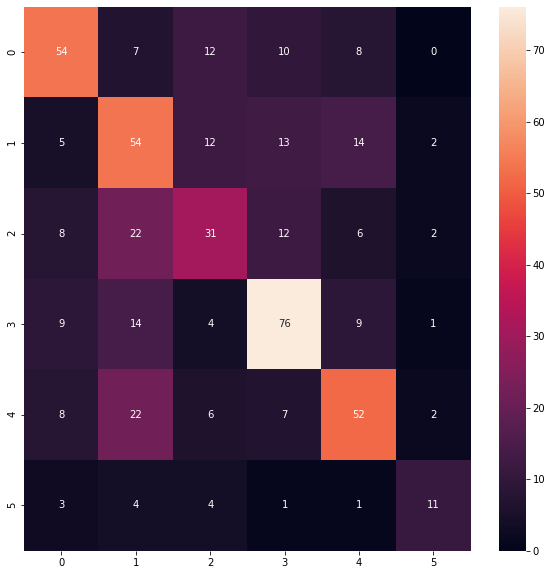
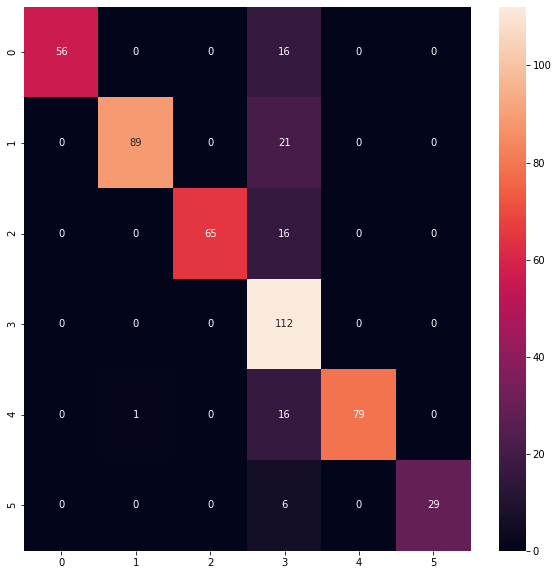


Figure 8: Confusion Matrix for SVM after tuning I Figure 9: Confusion Matrix for Random Forest after tuning

i

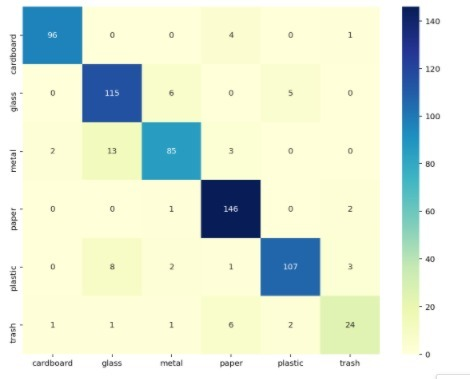
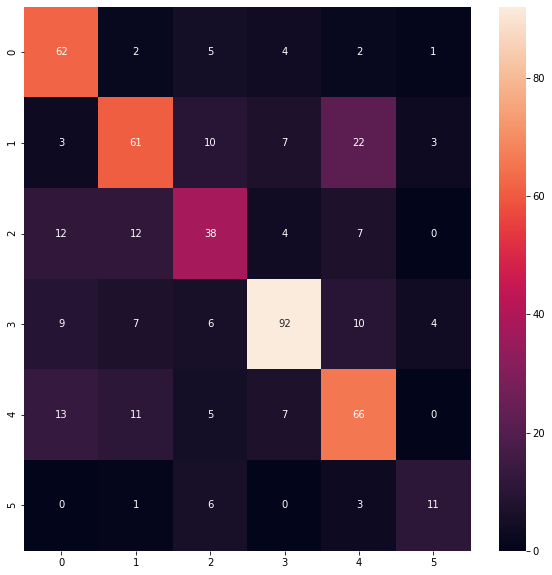


Figure 10: Confusion Matrix for Decision Tree Figure 11: Confusion Matrix for CNN

**Conclusion**

As indicated by the survey directed and the outcomes assembled, it very well may be plainly observed that a Convolutional Neural system can beat the presentation of pretty much every model constructed up until this point. Boosting any calculation and approving it with Cross Validation plans with numerous folds, the presentation of any model can be raised. After building the best model for each algorithm using the hyper parameter tuning, CNN has come up with the best accuracy while SVM is slightly behind. But Random forest and decision tree have not performed well in classifying the waste images properly.

**Future iWork**

For the future work, our plan is to work with the best algorithm, in our case which is CNN and we will try to emphasise on how to enhance the accuracy more by using other features. We will also try to work with a bigger dataset. Currently, our dataset has more than 2000 images of waste and in future we will increase this number and predict the accuracy. We are also planning to create a demo mobile app regarding the reusability of waste categories. It will help reduce the cost of buying new materials and also help to fulfil the sustainable goals also.

**F. References**

Ching Colli (2019). How to build an Image Classifier for Waste Sorting. 1-6

Chauhan, N (2019).  Decision Tree Algorithm — Explained,1-6

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.

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.

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 IEEEconference on Computer Vision and Pattern Recognition, pp. 1725–1732.

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

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.

Ibrahim F. Hanbal et al (2020), Classifying Wastes Using Random Forests, Gaussian Naïve Bayes, Support Vector Machine and Multilayer Perceptron, 1-6.

Anna Bosch, Andrew Zisserman & Xavier Munoz (2019), Image Classiﬁcation using RandomForests and Ferns, 1-4

Nurwauziyah, Iva & Sulistyah, Umroh & Gede, I & Putra, I Gede Brawiswa & Firdaus, Muhammad. (2018). Satellite Image Classification using Decision Tree, SVM and kNearest Neighbor.

Amari, S.-i. and Wu, S. (2016). Improving support vector machine classiﬁers by modifying kernel functions, Neural Networks 12(6): 783–789.

Adedeji, O., &amp; Wang, Z. (2019). Intelligent Waste Classification System Using Dee Learning Convolutional Neural Network. Procedia Manufacturing, 35, 607-612. d doi:10.1016/j.promfg.2019.05.086

Meza, J. K. S., Yepes, D. O., Rodrigo-Ilarri, J., &amp; Cassiraga, E. (2019). Predictive analysisof urban waste generation for the city of Bogotá, Colombia, through the implementation ofdecision trees-based machine learning, support vector machines and artificial neural networks. Heliyon, 5(11). <https://doi.org/10.1016/j.heliyon.2019.e02810>

Belgiu, M., Drgu, L.: Random forest in remote sensing: A review of applications an future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, 24– 31 (2016)

Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)Zhao, Y., Zhang, Y.: Comparison of decision tree methods for finding active objects. dvances in Space Research 41(12), 1955-1959 (2008)