March- April 2020
ISSN: 0193-4120 Page No. 1162 - 1168

Comparative Study on Classification of Mobile Images Using Supervised Machine Learning Algorithms

Swapnali Lotlikar
Poornima University, Jaipur
skhanolkar@rediffmail.com

Dr. Ajay Khunteta
Poornima University, Jaipur

Article Info
Volume 83
Page Number: 1162 - 1168
Publication Issue:
March - April 2020

Article History
Article Received: *24 July 2019*
Revised: *12 September 2019*
Accepted: *15 February 2020*
Publication: *14 March 2020*

Abstract

Mobile-waste is a growing problem in India. One of the factors that mobile has shorter life is rapidly changing technology. Most of the people throw their unwanted mobiles into scrap. Such scraped mobiles are hazardous to living life and environment as it contains lead, mercury, cadmium which affects living life and environment. It becomes necessary to recycle the mobile waste. According to the directives of Government of India, mobile manufacturer should recycle their own mobile-products which create e-waste. The classifying scrap mobile devices is a challenging task. This paper proposes work focus on automatic classification of damaged mobile as per the brands. A comparative study of different classifier algorithms such as Decision tree, Naive Bayes' theorem, and support vector machine, K nearest neighbor, Convolution Neural Network and inception model was carried out. Accurately classifying damaged mobiles is an essential task for the manufacturer. This paper compares the performance of six machine learning algorithms. The primary objective is to evaluate the performance in classifying data with respect to classification test accuracy, precision and recall.

Index Terms— *Mobile waste, Machine learning algorithms, precision, recall, accuracy.*

I. INTRODUCTION

Electronic waste are electronic and electrical items which are not in working condition or have been replaced by the new version. Computers, mobile phones, refrigerators, air conditioners, televisions are common electronic products, which eventually becomes e-wastes.

Many toxic components are present in E-waste like cadmium, lead, mercury etc. The presence of lead, mercury, chromium, cadmium is hazardous to health and environment.

Figures in TABLE 1 shows that mobile sales have grown exponentially in India.

Recycling is the only key to reduce the e waste. Approximately, 95% of e-waste in India is processed by the unorganized sectors because of cheap labour and weak legislation system [5].

March- April 2020
ISSN: 0193-4120 Page No. 1162 - 1168

TABLE 1: MOBILE SALES IN INDIA [NUMBERS ARE IN MILLIONS [20]

Year	Q1	Q2	Q3	Q4
2010	25.53	38.63	40.08	45.76
2011	62.4	42.8	44.92	29.88
2012	48.8	51.24	42.82	57.14
2013	60.73	62	60.44	61.83
2014	60.05	63.21	72.5	74.24
2015	55.4	60.8	57	58.5
2016	52.8	61.2	72.3	58.7
2017	62.22	62	80	56

According to the Government of India directive, it is the responsibility of mobile manufacturer to state the process of collection and channelization of mobile e-waste and its appropriate treatment and disposal methods. After visiting recycling company, it is observed that recycling is the successful key for e-waste management. Hence it is important that recycling companies are able to identify the mobiles they may receive for recycling. In the process of recycling, identifying the damaged mobile is a challenge.

II. DATA COLLECTION

For the study purpose, mobile images of 5 manufacturers, HTC, iPhone, Motorola, Samsung and Xiaomi considered. Mobile model identification can be done by looking at front and the rare side of it and hence for each model 2 images were captured one for front and another of rare side. In total 1050 mobile images of five mobile manufacturers and 105 images for each side. 86% of mobile images are trained for the different machine learning model and 14% of mobile images are tested for the different machine learning model. The total images for training data set is for one model is 90. The total images for testing data set for individual model is 15. And damaged images used for validation are 5 for each model

III. METHODOLOGY

Block diagram shown in Fig. 1 shows the methodology we adapted

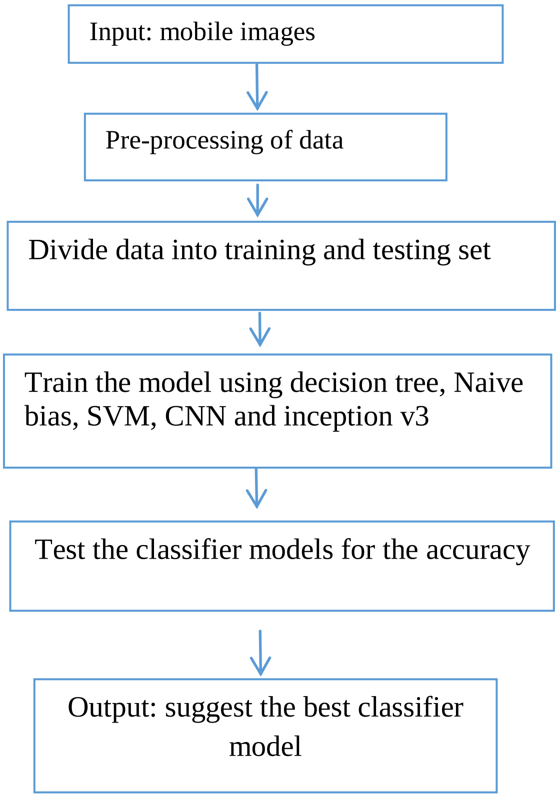

Figure 1. Methodology adopted

Step1: Input mobile images

Step 2: Data preprocessing: Before applying any classification model, preprocessing of data is important. Removal of unwanted information, re-sizing all images, conversion of RGB image to greyscale image are some of the techniques used.

Step 3: The System takes 105 mobile images of of each manufacturers, HTC, Iphone, Motorola ,Samsung and Xiaomi. 86% of mobile images were used in training data set and 14 % mobile images are used in testing data set (testing is done for training samples). Other than this we used 4 % damaged mobile images are used for validation.

Step 4 Image processing allows you for the detection of region especially homogeneous region. Image segmentation helps to identify certain features of image. The image size considered as 128* 128*2

Step 5: In binary classification, accuracy can be calculated by measuring the correctly classified and incorrectly classified data as positives and negatives:

Accuracy percentage = [tp+tn)/(tp+tn+fp+fn)]*100

Where

tp (True positive): When the data point is classified as positive and it is actually positive

tn (True Negative): When data point is not classified properly and you actually do not have it

False Positive: When values in data set is positive but these values are classified as negative.

False Negative: When values in data set is classified as negative and these values are classified as positive.

The harmonic mean of recall and precision is derived from f1-score. [19].

Recall= TP(true positive)/(false negative true positive)

Precision =TP(true positive)/(false positive+ true positive)

F1=2*[(recall*precision)/(recall +precision)

The scores calculated for every class will give us the accuracy of the classifier. It helps in classifying the data in that particular class.

The classification process is implemented using python programming language

Machine learning models are implemented using Tensorflow. One of the major advantages of using Tensor flow is its versatility with respect to data; irrespective of whether the input being provided consists of images or text, it works efficiently and provides excellent results [7].

IV. ALGORITHM FOR MOBILE IMAGE IDENTIFICATION

This section explains the machine learning algorithm used for identification of mobile image. Table 2 shows the algorithms and its advantages and disadvantages

V. RESULT AND DISCUSSION

Algorithms discussed in above section 3.0 are trained in order to find the best suitable classifier. 14% of the images are used for the testing
Detailed analysis of the algorithms

a) KNN:

A class is assigned based on the classes of the nearest neighbor data points. the nearest neighbor is the data point which is close/similar to a given data point. Here the euclidian distance formula is used.

$$d(X_i, X_j) = \sqrt{\left((x_{i1} - x_{j1})^2 + \cdots + (x_{ip} - x_{jp})^2\right)}$$

Following fig 2 shows 28% accuracy for K=2 which is very poor.

TABLE 2: ALGORITHMS FOR MOBILE IMAGE IDENTIFICATION

Algorithm	Method of Implementation	Advantages	Disadvantages
Naïve Bayes Classifier	The Naive Bayes theorem assumes the conditional independence on training data set. It helps to decrease the complexity of Baye's theorem.	Naive Bayes is a probabilistic based classification model that can provide better classification with a	Since it is a linear classifier it does not perform as good as a linear classifiers like SVM. They are typically used in text classification

	If we consider n different attribute values, then the probability can be written as: $P(X_1 \ldots X_n \mid Y) = \prod_{i=1}^{n} P(X_i \mid Y)$	minimum training data and time.	[11]
Nearest Neighbor (kNN) Algorithm	It is mandatory to select k values in KNN. Several tests are done with different values of k. This ensures an optimal value of k with good accuracy	Algorithm does not make ant assumption of data distribution since it is non parametric	Choosing value of k Error rate almost twice than Bayes error rate[13] KNN gives poor results when there is small changes in rotation, translation, viewpoint of image. Using raw pixel intensities as inputs yield poor results [22]
Decision tree classifier	It is a supervised learning algorithm that creates a model that predicts a variable based on a set of decision/conditional rules.	Nonlinear relationship among parameters, has no impact on tree performance	Duplication of sub-tree in different paths is possible [14]
SVM - Support Vector Machine	It is a machine learning based classification tehnique. It implements the supervised model for learning and it is widely used for cancer diagnosis and prognosis field [6].	Work efficiently with small data set Work even if number of features are greater than number of samples Effective in high dimensional data sets Memory Efficient — uses subset of training data (only support vectors are considered) Can control over fitting and under fitting	Choosing a “good” kernel function is not easy Pair-wise classifications can be used for a multi-class classification, that is one class against all others, for all classes [14]
CNN	Convolutional Neural Networks (CNN)s which are characterized by feature extraction and abstraction with the help of various interlayers. As a result of this approach, parameter space extends and allows higher accuracy, but also requires a large number of training samples and sessions [10]	No human supervision required for detecting important features [3]	High Computation cost due to large training data and GPU requirements.
Inception model	Multi-level features can extracted in the inception model.	Performance improved because of multiple features from	Requires significant amount of

	It extracts the features by computing 1x1, 3x3, 5x5 convolutions. The results are then concatenated and hence it behaves as multiple convolution filters.	multiple filters	labeled training data and compute power

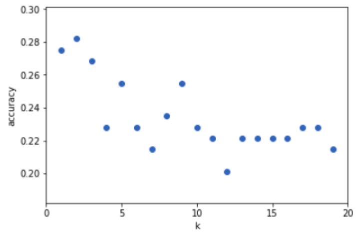

Figure 2. KNN accuracy for various K values

Confusion matrix for KNN on evaluation test data set:

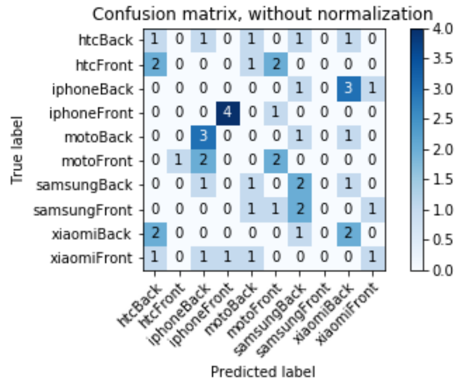

Figure 3. Confusion matrix for KNN

Figure 4.

b). Decision tree classifier

In this algorithms CART and ID3 are used to classify the image data. Table 3 shows the accuracy for the training and the testing data. The classifier shows around 30% accuracy which shows that this model fails to correctly classify the data

TABLE 3: DT CLASSIFIERS ACCURACY

	Accuracy rate in %(training data set)	**Accuracy rate in %(testing data set)**
CART (gini index)	32	34.22
ID3 (Entropy)	30	29

Confusion matrix of decision tree on evaluation data set or test data set:

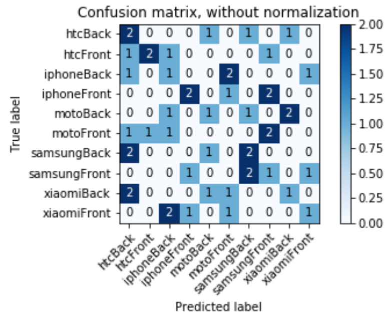

Figure 5. Confusion matrix for Decision tree classifer

c) CNN:(Convolution Neural Network)

A neural network (CNN) consists of atleast two layers, namely an input and output layer. The input layer usually consumes a sum of normalized input values which could be interpreted as an external stimulus.

use a neural network with two input nodes i1 and i2, two hidden neurons h1 and h2, two output neurons 01 and o2. Here's the basic structure

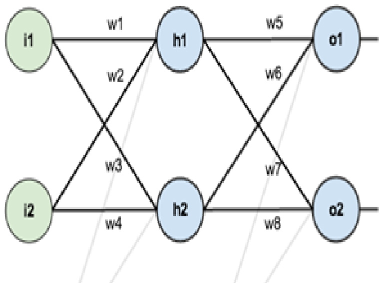

Figure 6. Basic structure of CNN

Activation function used is relu. Max Pooling layer of size is 2*2

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 250, 250, 32)      2432
_________________________________________________________________
activation_1 (Activation)    (None, 250, 250, 32)      0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 125, 125, 32)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 125, 125, 32)      25632
_________________________________________________________________
activation_2 (Activation)    (None, 125, 125, 32)      0
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 62, 62, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 62, 62, 64)        51264
_________________________________________________________________
activation_3 (Activation)    (None, 62, 62, 64)        0
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 31, 31, 64)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 61504)             0
_________________________________________________________________
dense_1 (Dense)              (None, 500)               30752500
_________________________________________________________________
activation_4 (Activation)    (None, 500)               0
_________________________________________________________________
dense_2 (Dense)              (None, 10)                5010
_________________________________________________________________
activation_5 (Activation)    (None, 10)                0
=================================================================
Total params: 30,836,838
Trainable params: 30,836,838
Non-trainable params: 0
```

Figure 7. CNN Architeture

d) Inception model

In inception module, the multiple convolution filters are used. The results are merged. Multi-level feature extraction is the main advantage of this model. For example, it extracts local (1*1) and general (5*5) features at the same time.

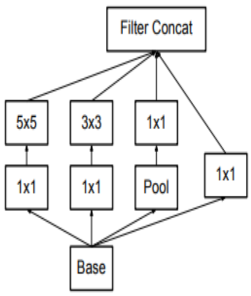

Figure 8. Basic structure of Inception model

Confusion matrix, without normalization

[[5 0 0 0 0 0 0 0 0 0]
[0 4 0 1 0 0 0 0 0 0]
[0 0 5 0 0 0 0 0 0 0]
[0 1 0 4 0 0 0 0 0 0]
[0 0 0 0 5 0 0 0 0 0]
[0 0 0 3 0 2 0 0 0 0]
[0 0 0 0 0 0 5 0 0 0]
[1 1 1 1 0 0 0 1 0 0]
[0 0 0 0 0 0 1 0 4 0]
[0 2 0 0 0 1 0 2 0 0]

Figure 9. Confusion matrix for inception model

Table 5 shows the training models accuracy. Accuracy is measure of how correctly the system classifies the unseen data into a correct class. Greater the accuracy more reliable is the model.

TABLE 6: Accuracy % of testing data set

	Comparison of accuracy percentage on testing data set					
Algorithm	Naive Bayes	Nearest Neighbor	Decision Tree	Support vector machine	NN	Inception Model
	24	24	26	22	82	70

TABLE 5: Accuracy table for the detecting the manufacturer of damaged mobile:

From the above table it is observed that CNN and inception model gives better accuracy. Other models - Naive Bayes, KNN, Decision tree, SVM are less accurate. Between CNN and Inception the accuracy for CNN is better for all the mobile models except Xiaomi back.

	Evaluation test accuracy in %																	
	Naïve Bayes			**K nearest neighbour**			**Decision Tree classifier**			**Support vector machine**			**CNN**			**Inception Model**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
HTC-B	.00	.00	.00	.17	.20	.18	.22	.40	.29	.29	.40	.33	**1.00**	**1.00**	**1.00**	.83	1.00	.91
HTC-F	.67	.80	.73	.00	.00	.00	.67	.40	.50	.75	.60	.67	**.62**	**1.00**	**.77**	.50	.80	.62
Iphone-B	.00	.00	.00	.00	.00	.00	.17	.20	.18	.25	.20	.22	**1.00**	.80	.89	.83	**1.00**	**.91**
Iphone-F	.25	.80	.38	.80	.80	.80	.50	.40	.45	.33	.20	.25	**1.00**	**1.00**	**1.00**	.44	.80	.57
Moto-B	.67	.00	.00	.00	.00	.00	.25	.20	.22	.67	.40	.50	**1.00**	**1.00**	**1.00**	1.00	1.00	1.00
Moto-F	.00	.00	.00	.33	.40	.36	.00	.00	.00	1.00	.20	.33	**.80**	**.80**	**.80**	.67	.40	.50
Samsung-B	.00	.00	.00	.25	.40	.31	.33	.40	.36	.00	.00	.00	**1.00**	**.80**	**.89**	.83	1.00	.91
Samsung-F	.00	.00	.00	.00	.00	.00	.17	20	.18	.10	.20	**.13**	**1.00**	**.20**	**.33**	.33	.20	.25
Xiaomi- B	.40	.80	.53	.25	.40	.31	.25	.20	.22	.00	.00	.00	.62	**1.00**	.77	1.00	.80	**.89**
Xiaomi-F	.00	.00	.00	.33	.20	.25	.33	.20	.25	.00	.00	.00	**.80**	**.80**	**.80**	.00	.00	.00

*p=precision, *R=recall

VI.Conclusion

The objective of the above paper is to identify the image of damaged mobile. The data distribution across training and testing data sets are 86% and 14% respectively. Different machine learning algorithms are discussed and implemented in this paer .Six popular machine learning algorithms are applied for mobile manufacturer detection. The main methodology of each algorithm is described.

The performance for all the models is as follow:-

The above result shows that the convolution neural network model is the best suitable model for classifying manufacturer of damaged mobile

References

[1] Vardhani, P. R., Priyadarshini, Y. I., & Narasimhulu, Y. (2019) CNN data mining algorithm for detecting credit card fraud in Soft Computing and medical Bioinformatics (pp. 85-93). Springer, Singapore.

[2] Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of

March- April 2020
ISSN: 0193-4120 Page No. 1162 - 1168

statistical learning: data mining, inference and rediction. The Mathematical Intelligencer, 27(2), 83-85.

[3] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018)

[4] Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9(4), 611-629.

[5] Kim, K., Hong, H., Nam, G., & Park, K. (2017). A study of deep CNN- based classification of open and closed eyes using a visible light camera

[6] sensor. Sensors, 17(7), 1534.

[7] Awasthi, A. K., Zeng, X., & Li, J. (2016). Environmental pollution of electronic waste recycling in India: A critical review. Environmental pollution, 211, 259-270.

[8] Sinha-Khetriwal, D., Kraeuchi, P., & Schwaninger, M. (2005). A comparison of electronic waste recycling in Switzerland and in India. Environmental Impact Assessment Review, 25(5), 492-504.

[9] Electronic waste recycling in Switzerland and in India. Environmental impact Assessment Review, 25(5), 492-504.

[10] H. You and G. Rumbe, "Comparative study of classification techniques on breast cancer FNA biopsy data," International Journal of Artificial Intelligence and Interactive Multimedia, vol. 1, no. 3, pp. 6-13, 2010

[11] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," Artificial Intelligence in Medicine, vol. 23, no. 1, pp. 89-109, 2001.

[12] https://www.infoworld.com/article/3278008/what-is-tensorflowhe-machine-learning-library-explained.html

[13] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks,” in Proc. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-4.

[14] Ragupathy, R., & Maguluri, L. P. (2018). Comparative analysis of machine learning algorithms on social media test. International Journal of Engineering and Technology.

[15] Starzacher, A., & Rinner, B. (2008, December). Evaluating KNN, LDA and QDA classification for embedded online feature fusion. In 2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (pp. 85-90). IEEE.

[16] Cover, T. M., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.

[17] [15] Suganya, A., Mohanapriya, N., & Kalaavathi, B. (2015). Lung Nodule Classification Techniques for Low Dose Computed Tomography (LDCT) Scan Images as Survey. International Journal of Computer Applications, 131(14), 12-15.

[18] Vishal, T. V., Srinidhi, S., Srividhya, S., Kumar, S. V., & Swathika, R. (2016). A Survey and Comparison of Artificial Intelligence Techniques for Image Classification and Their Applications. International Journal of Science and Research (IJSR), 5(4).

[19] https://towardsdatascience.com/applied-deep-learning-part-4 onvolutional-neural-networks-584bc134c1e2

[20] [18]https://www.analyticsvidhya.com/blog/2017/06/transfer-learning-the-art-of-fine-tuning-a-pre-trained-model/

[21] Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In European Conference on Information Retrieval (pp. 345-359). Springer, Berlin, Heidelberg.

[22] Lotlikar, S., & Vakharia, S. A. (2018). Neural Network Model Prediction A prediction model for Toxin release in environment from E-Waste. Recent Trends in Programming languages, 4(3), 16-20.

[23] Park, D. C. (2016). Image classification using Naïve Bayes classifier. Int J. Comp Sci Elec Eng, 4(3), 135-139.

[24] [21]https://www.pyimagesearch.com/2016/08/08/k-nn-classifier-for-image-classification/