JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

GARBAGE CLASSIFICATION USING RANDOM FOREST ALGORITHM

[1]Dr.S. Ramesh, [2]Dr.J.S. Kanchana, [3]K.P. Siranjeevi , [4]M.N. Yashwanth Kumar.

[1]Associate Professor, [2]Profesor, [3]IT Student, [4]IT Student.

Information Technology,

K.L.N.College Of Engineering, Sivaganga, India.

Abstract: In today's fast-paced world of the internet age with all the amenities and latest gadgets, the major urban cities in the world are still struggling with trash management. Only a few countries use the recycling of wastes but most of them dump all the trash to the landfills. The quantity of generated trash in day to day life is affecting land, water, and air which causes a serious threat to the environment.The objective is to develop a system that can classify the trash images into their correct categories with the help of Machine learning and Deep learning methodologies.In this work the system is going to classify the Garbage into nine categories Aluminium, Carton, Organic Waste, Other plastics, Paper and Cardboard,Plastic,Textiles,Wood. The classification are going to be done using Decision Tree Random Forest Classification algorithm. Extensive experiments on various real-world datasets demonstrate the effectiveness of our method.

Index Terms - Decision Tree, Random Forest, Garbage Classification.

I. Introduction:

Waste generation has expanded vastly around the world in current decades, and there are no signs and symptoms of it slowing down. By 2060, global municipal solid waste generation is anticipated to have extended by roughly 70 percentage to 3.4 billion metric tons. This is due to a variety of factors, such as population growth, urbanization, and monetary growth, as nicely as purchaser purchasing habits. Every year, human beings produce tens of millions of heaps of waste, and this is an increasing number of turning into a most important issue worldwide. With such monstrous volumes of waste arising, the need for authorities to grant enough waste cure and disposal offerings has turn out to be ever more important. However, much less than 20 percentage of waste is recycled every year, with large portions still dispatched to landfill sites. Waste is additionally frequently disposed of at hazardous open dump sites, specifically in growing nations. Richer nations produce extra waste than poorer countries, however generally have higher waste administration to assist deal with these issues.Machine Learning **:** It is the process of making computers learn stuff by themselves. In this, a computer is expected to use algorithms and statistical models to perform specific tasks without any instructions,Machine learning applications provide results based on past experience.Big Data: It can be defined as a concept used to describe a large volume of data, which are both unstructured and structured , and that gets increased day by day by any system or business.There are five innate characteristics of big data known as the "5 V's of Big Data" which help us to better understand the essential elements of big data.The five V's are Velocity, Volume, Value, Variety, and Veracity.

II. Literature Review:

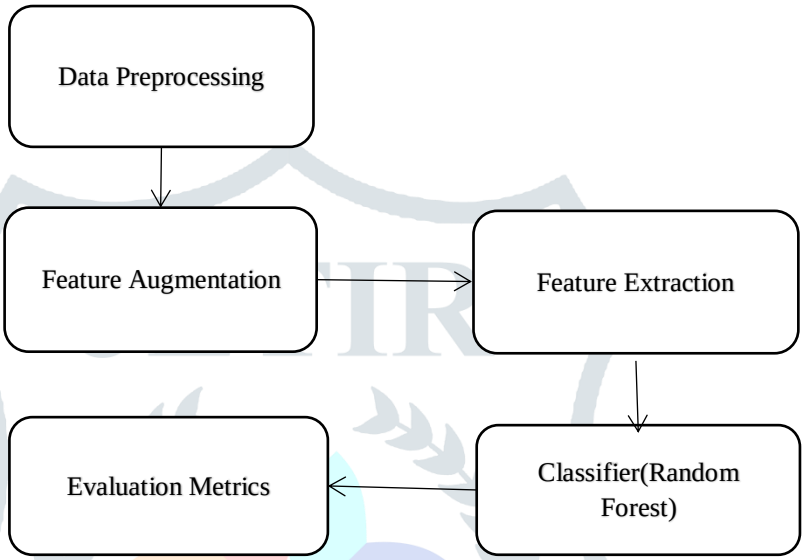
The increase in Land Pollution has instigated many researchers to develop a more efficient and reliable trash classification system. from Studying and clear observation of the previous researchers, we see that all the algorithms like KNN, CNN, SVM, etc.. perform in different manner, A list of a few literature studies associated with this subjected are given below.

Table 1. Garbage classification research and accuracy using different machine learning algorithm.

S.no	**Paper Title**	**Models/Algorithm**	**Results**
1.	Deep Learning for Garbage Classification Based on Visual Images	CNN	Accuracy: **94.10%**
2.	A Neural Network-Based Garbage	DNN	Accuracy: **94.0%**

	Collection management System		
3.	A Novel Frame work for Trash Classification	DNN,CNN	CNN: Accuracy: **94.56%** DNN: Accuracy: **96%**
4.	A Unified Learning Framework for Robust Garbage Classification	KNN	Accuracy:**93.8%**

III. Proposed Methodology



3.1 Machine learning algorithm used:

3.1.1 Decision Tree :

In Decision Tree Algorithm, we have two nodes, the Decision Node and Leaf Node. Decision Node being used for making decisions and erecting more multiple branches while Leaf Nodes are the output of aforementioned Decision Nodes. This Algorithm shows easy to understand output, a tree like structure resulting in effective consortium and interrelation among datasets and can be easily understood so it is usually associated with human thinking as it is quite easy to understand. Splitting, Pruning etc. are some keywords users should familiarize with. As a result, they are quite easy to prepare and require less data cleaning. In the context of trash detection, the nodes are usually required to be classified . The steps given below are few steps involved in working of this algorithm are Choosing a root node,Laying out dataset to probe the best attribute,Splitting into subsets and producing tree nodes, which include the exceptional attributes. Repeating the procedure till the node can't be further categorized or split, ensuing in the best result.

3.1.2 Random Forest Classifier:

Another popular ML algorithm that falls under the supervised learning category, it is used for Classification Problems. It is one of the more advanced learning and flexible algorithms. It makes use of multiple decision trees to form a group of trees for classification methods, the output of this algorithm heavily relies on this simple yet multiplex and compounded method. And its robustness is seen when used it in a large datasets.

IV. Proposed Methodology

The proposed methodology contains 4 modules they are given below

4.1 Data Preprocessing:

Data Preprocessing in Machine Learning refers to the technique of preparing the raw data to make it suitable for a building and training Machine Learning models.Important steps in Data Preprocessing are Acquire dataset, Import Important Libraries, Image Cropping, Splitting Dataset into Test and Train.

4.2 Feature Extraction:

Feature extraction refers to the method of remodeling data into numerical features which will be processed while preserving the knowledge within the original data set. It yields better results than applying machine learning on to the raw data .This process includes conversion of color images into 3D array, i.e., The array contains pixel value of an image.

4.3 Feature Argumentation:

Feature Augmentation is a method of making certain alteration to the existing dataset and adding it to the original dataset to generate slightly modified and multiplied data. It can take all the samples available in the dataset and modify them several times in a many different ways to get the larger dataset.

4.4 Evaluation Metrics:

The Accuracy of this system is calculated using the formula

Accuracy = TP+TN / TP+TN+FP+FN

True Negatives : Number of outcomes that are actually negative and are predicted negative
False Positives : Number of outcomes that are actually negative but predicted positive.
True Positives : Number of outcomes that are actually positive and are predicted positive
False Negatives : Number of outcomes that are actually positive but predicted negative.

V. Experiment Results:

The work was implemented on a ryzen 5 processor(equivalent to i5 processor) with 2.30 Ghz speed, 6 GB RAM, 1 TB of HDD and all tests and experiments on the classifiers and algorithms explained and laid out thus far in this paper have been conducted using libraries from JUPYTER notebook, Python 3, Tensorflow and using scikit learning machine. In this study, the dataset is split in the ration of 70:30. 70% for training and 30% for testing.

Table 2. Comparison with KNN model

S.no.	Model/Algorithm	Accuracy	Sensitivity
1.	K-Nearest Neighbour Algorithm	93.85%	89%
2.	Random Forest Algorithm (Proposed System)	98.09 %	92%

5.1 Graphs:

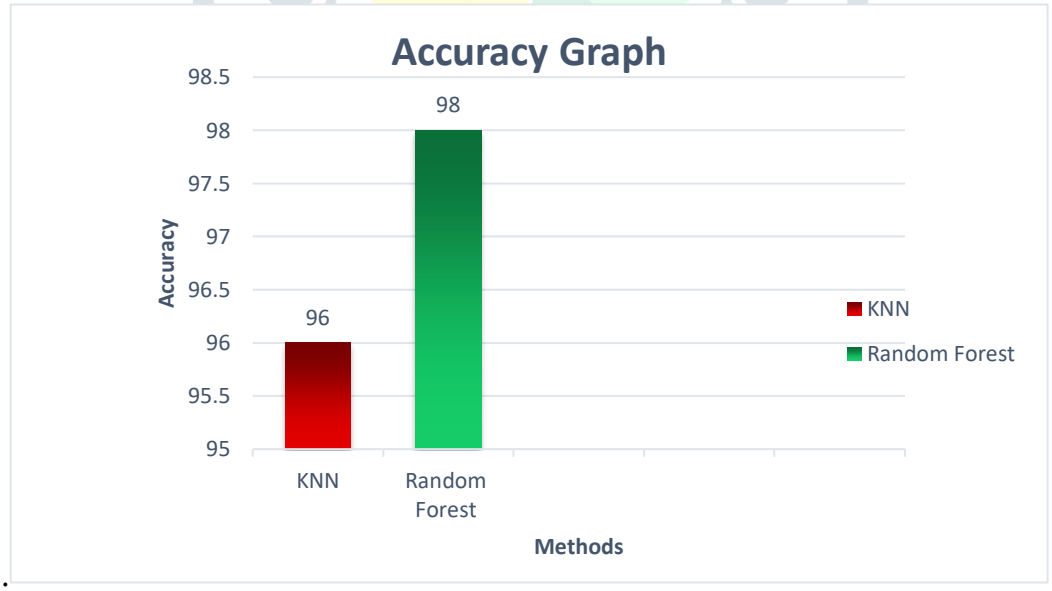


Fig.1. Accuracy graph of existing and proposed system

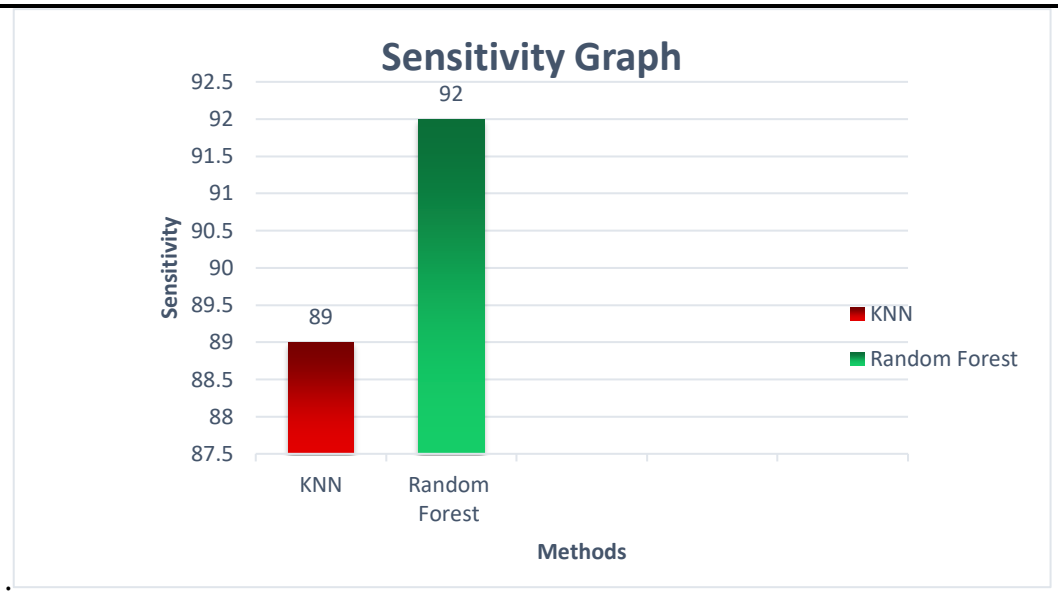


Fig.2. Sensitivity graph of existing and proposed system

VI. Conclusion:

This paper proposed a Random forest algorithm for Trash classification, The proposed system contains total of 5070 images.Referring to Table 1,Table 2, Figure 1 and Figure 2. It can be seen that the KNN have low accuracy and sensitivity, unlike other model that is proposed. Whereas the accuracy obtained by Random Forest Classification is 98% and its sensitivity is 92%, making it better than KNN , which have accuracy of 96% and 89% sensitivity , Random Forest algorithm also performed better in terms of classification error, i.e, correctly classified instances and incorrectly classified instances.Based on the study it can be concluded that Random Forest classifier performs far superior.

VII. Future Work:

In future, the system will be continuously developed to improve the effectiveness of the proposed system, and several segmentation techniques will be applied to Pre-processing to improve the performance of Garbage classification. In addition, the number of images in each garbage category usually unbalanced, therefore, a deep model for trash classification in imbalance scenarios would be studied.

VIII. Reference

[1]. Jianfei Yang , Zhaoyang Zeng, Kai Wang, Han Zou, and Lihua Xie, "Garbage Net: A Unified Learning Framework for Robust Garbage Classification," in IEEE Access, August. 2021.
[2]. S.Meng, N. Zhang, and Y. Ren, "X-DenseNet: Deep learning for garbage classification based on visual images," in IEEE Access, Jan.2021.
[3]. K.Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region attention networks for pose and occlusion robust facial expression recognition,"IEEE Trans. Image Process., vol. 29, pp. 4057–4069, Jul.2020.
[4] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," in Proc. AAAI Conf. Artif. Intell., vol. 33, 2019.
[5] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in Proc. IEEE Conf. Computer.Vis. Pattern Recognition., 2020.
[6] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classifification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5,pp. 2811–2821, May 2018.
[7] G. Cheng, P. Zhou, and J. Han, "Duplex metric learning for image set classifification," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 281–292, Jan. 2018.
[8] B. Frénay and M. Verleysen, "Classifification in the presence of label noise: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 845–869, May 2014.
[9] T. Le, B. Vo, H. Fujita, N. T. Nguyen, and S. W. Baik, ''A fast and accurate approach for bankruptcy forecasting using squared logistics loss with GPU-based extreme gradient boosting,'' Inf. Sci., vol. 494, pp. 294–310, Aug. 2019.
[10] T. Le, M. Y. Lee, J. R. Park, and S. W. Baik, ''Oversampling techniques for bankruptcy prediction: Novel features from a transaction dataset,'' Symmetry, vol. 10, no. 4, p. 79, 2018.
[11] I. B. S. Salimi Dewantara and I. K. Wibowo, ''Visual-based trash detection and classification system for smart trash bin robot,'' in Proc. ESKCIC, 2018, pp. 378–383. [21] Y. Chu, C. Huang, X. Xie, B. Tan, S. Kamal, and X. Xiong, ''Multilayer hybrid deep-learning method for waste classification and recycling,'' Comput. Intell. Neurosci., vol. 2018, Nov. 2018, Art. no. 5060857.
[12] M. Yang and G. Thung, ''Classification of trash for recyclability status,'' Mach. Learn., Stanford, CA, USA, Project Rep. CS229, 2016
[13] J. Kim, H. J. Kim, and H. Kim, ''Fraud detection for job placement using hierarchical clusters-based deep neural networks,'' Appl. Intell., vol. 49, no. 8, pp. 2842–2861, 2019. [14] Y. Song, J. W. Lee, and J. Lee, ''A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction,'' Appl. Intell., vol. 49, no. 3, pp. 897–911, 2019.
[14] Google, "MobileNets: Open-Source Models for Efficient On-Device Vision," Research Blog. [Online]. Available: https://research.googleblog.com/2017/06/mobilenets-open-sourcemodels-for.html. [Accessed: 08-Dec-2017].
[15] M. Harvey, "Creating insanely fast image classifiers with MobileNet in TensorFlow," Hacker Noon, 11-Jul-2017. [Online]. Available: https://hackernoon.com/creating-insanely-fast-image-classifiers-withmobilenet-in-tensorflow-f030ce0a2991. [Accessed: 08-Dec-2017].