Waste Classification

Maral Amiri, Krinal Gujarati, Meghana Hegde, Sejal Gupta, and Yukta Mehta

Department of Applied Data Science

San Jose State University

DATA 245: Machine Learning Technologies Section 21

Vishnu S. Pendyala, Ph.D.

November 12, 2023

Table of Contents

Index of figures

Project Evaluation Criteria

Table 1

Project Evaluation Criteria

Criteria	**How Criteria Met**
Code Walkthrough	The presentation will provide a comprehensive guide through our code, guaranteeing clarity and adherence to industry best practices.
Presentation Skills	Includes time management. The slides are well organized, visually appealing and will convey key project aspects.
Discussion / Q&A	We have practiced our presentation and are left with time for discussion and Q&A.
DEMO	Visualization includes exploratory analysis (heat maps and other visuals).
Report	Format, completeness, language, plagiarism, whether Turnitin could process it (no unnecessary screenshots), etc. The project report follows correct formatting, maintains completeness, and adheres to language standards, avoiding plagiarism and unnecessary inclusion of screenshots.
Version Control	Use of Git / GitHub or equivalent; must be publicly accessible. We have created an account on GitHub, and the link is `https://github.com/yourusername/yourrepository`.

Continued on the next page

Table 1 – Continued from previous page

Criteria	**How Criteria Met**
Relates to Sustainability	The work is toward one or more of the 17 SDGs of the United Nations. Describe in the report how this criterion is met. See also: `https://sdgs.un.org/goals`, `https://unesdoc.unesco.org/ark:/48223/pf0000247444`.
Lessons Learned	Included in the report and presentation. Prospects of winning competition/publication, innovation, evaluation of performance, teamwork, technical difficulty. Practiced pair programming. Used GitHub Copilot, if possible, and describe the experience using screenshots. Practiced agile/scrum (1-week sprints). Submit evidence on Canvas - meeting minutes, sprint backlog, and any other artifacts. Use tools such as `https://trello.com/en-US/pricing` (Free license available). The meeting minutes are captured in `https://drive.google.com/drive/folders/1-31U84u_WOgwXj98QhxIKWPKtqkBqXCm?usp=sharing`.
Used Grammarly / Other Tools for Language	Grammarly free version is sufficient. Can use other tools as well. Submit a report screenshot on Canvas.
Slides	Saving the model for a quick demo. See `https://www.kaggle.com/prmohanty/python-how-to-save-and-load-ml-models`. Upload the model file if it is < 2MB. Otherwise, save it in the cloud accessible to the ISAs and provide the URL in the report.

Continued on the next page

Table 1 – Continued from previous page

Criteria	How Criteria Met
Used LaTeX	Upload .tex file (it should indicate that the IEEE LaTeX template was used and not generated from doc or other formats). Using editors such as Lyx is fine.
Used Creative Presentation Techniques	Use of Generative AI is okay here. Try animation, effects, newer features such as those offered by Prezi, etc.
Literature Survey	1. Did not miss out on any important existing work that is relevant to the project. 2. Literature survey is organized into meaningful subsections. 3. All references are cited and follow the standard notation used in the template.

Abstract

The explosion of trash generation is brought on by the hasty acceleration of urbanization and companies with a variety of plastic, garden waste, paper, glass, and more. The abundance of solid waste poses a severe danger to the city's ecological environment and has resulted in a number of problems, including contamination and unlawful disposal. Consequently, the management of solid waste has emerged as a critical global environmental challenge, particularly in developing countries. Traditional waste sorting methods have proven labor-intensive and error-prone, leading to inefficient resource allocation and significant environmental hazards. Therefore, it is anticipated that ML techniques will be enhanced in the modeling, prediction, and optimization processes in waste classification. Various kinds of waste need to be handled in diverse ways for

effective waste management. In this context, our proposed application aims to classify waste into 2 fundamental categories: Organic and Non-organic waste, based on captured images of the waste. To train the model, a dataset of 22,564 photos from each category was collected. Machine learning classification techniques, including Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF) and K-Nearest Neighbors (KNN) algorithms can be used to train our proposed methodology. The accuracy will be examined using machine learning algorithms, with 70% of the data allocated for training and 30% for validation. The results of our test will show that the algorithm successfully recognizes photos of both organic and non-organic waste. By effectively addressing the growing trash dilemma and reducing its negative environmental effects, this application presents a viable path for changing waste management procedures.

Motivation

Waste management can have a widespread impact on human life, society, and nature. Some of the motivations for researching waste management using Machine Learning are: Carbon footprint: People want to reduce their carbon/emissions footprint, but today there is a scarcity of proper waste management and treatment centers. There are some waste separation and management infrastructures, but those are very restrictive and rely on people separating recycling and landfill waste. We need to come up with a new infrastructure that can offload this workflow of waste separation and its further treatment in a widespread way. Pollution: Organic waste decays slowly and releases methane and carbon dioxide, also known as landfill gas, contributing to 20% of the methane emissions globally. Methane is a greenhouse gas, one of the reasons for global warming. Thus, there is a need to divert organic waste from landfills and treat them separately. Waste to Energy: If we have an appropriate waste treatment method, we can generate energy from it. Organic waste can be processed using anaerobic digestion. This process produces biogas, a renewable energy. Composting is an aerobic process that converts organic waste into fertilizer. Nature: Waste often ends up in water bodies, oceans, or grounds where dumping is illegal. The Great Pacific Garbage Patch is an example of massive waste accumulated in the Pacific Ocean. Recycling and Reusing: Some of the daily waste that can be recycled and reused is plastic, metals, glass, cardboard, and paper. Increasing recycling and reusing would limit the extraction of elements like silicon, nickel, etc. from the earth and prevent deforestation. Automation using machine learning techniques: Human labor is costly. Also, most of the waste segregation happens at the source of waste. That means people themselves are responsible for segregating waste. But this brings in a lot of behavioral factors. Offloading this task to an automated machine learning process can make it more affordable and effective eventually. Machine learning techniques can be applied to classify waste. Other advantages are processing large volumes of data rapidly, accurate classification, and saving cost by reducing the need of manual sorting.

Literature Survey

A variety of waste management studies have been conducted under various perspectives These studies have focused on the environmental impacts of waste, the potential health risks of improper disposal of waste, and the economic implications of waste management. The findings of these studies can be used to develop strategies for efficient and effective waste management. The waste classification CNN algorithm is proposed to classify the biodegradable waste that can be used for real-time detection of recyclable material in the garbage and contribute to reducing pollutionChhabra, Sharan, Gupta, and Astya (2022). For separating plastic bottles from waste, the researchers in the paper use SVM as a classification technique. Various classifications were used, such as separating caps from bottles and classifying colored recycled bottles separately. Additionally, overlapping adjacent bottles are segmented using distance transformation as well as segmentation to achieve 95.7% accuracyWang, Peng, Huang, and Sun (2019) .The research uses three perspectives for classifying the handcrafter and non-handcrafted features using the combination of CNN and SVM for training and testing purposes. It uses a computer vision system combining deep CNN-based features with traditional handcrafted features for generic image classification. It outperforms standard approaches, achieving significantly improved accuracy (p-value of 0.01) when fusing handcrafted and novel featuresNanni, Ghidoni, and Brahnam (2017) . The research has significant implications for improving construction and demolition waste (CDW) management, enabling efficient waste sorting, and reducing environmental impact.

However, the study notes the need for more diverse training data and potential enhancements using additional sensors or data sources to further improve accuracy and reliability Nežerka, Zbíral, and Trejbal (2023). The paper uses the two-class classification of the waste segregation technique for classification one at the individual level and the second at the society level. The sensors were used to send the alert messages by checking the type of waste and concentration of poisonous gases. The main focus of the research was to reduce the pollutants and conserve the resources used Dubey, Singh, Yadav, and Singh (2020). The paper attempts to achieve efficient and vigorous waste management techniques by predicting the level of trash in the bins. They used the LoRa module for data transformation from bins to systems which contributes to the best route to collect the waste from bins Anh Khoa et al. (2020). For classification purposes, the paper gathered the classes of glass, paper, metal, plastic, cardboard, and waste. Researchers collected and sampled data using machine learning algorithms and converted images to RGB format using machine learning algorithms Sami, Amin, and Hassan (2020). The paper aims to correctly identify the quantity of domestic waste (DW) produced by families daily. They performed the analysis produced on the waste produced between 2010-2019 and family-generated (11 members) waste for one month the results predicted that family contributed about 1.5kg of waste daily and the average rate generation in 2010-2019 was around 1.7-7.9kg.Lakhouit et al. (2023). The paper introduced the concept of automated ML for connecting waste management with real-life problems. The paper focuses on the concept of when the bins installed are filled and the sensor will send the measurements to the system. They used binary classification techniques along with various data-driven models.Rutqvist, Kleyko, and Blomstedt (2020). The research was conducted using SVM and PCA with 6 different scenarios to extract a Scale Invariant Feature Transform (SIFT) feature for waste classification of glass, paper, cardboard, metal, and plastic. The highest accuracy of 62% was achieved using SVM and a conclusion was made that increasing the PCA reduced the accuracy of the model Puspaningrum et al. (2020). The paper focused on the segregation of solid waste generated by municipal corporations. The Berkley Method of composting and computer vision technique was used to increase the efficiency of waste classification. It was observed that SVM gave a performance of 85% in the overall process Behera, Sr, Vasundhara, Saisudha, and Priya (2020). The paper identifies the ability of SVM in waste image classification of cans, cigarette butts, plastic bottles, and cartons, by introducing a boosting algorithm to optimize the SVM that would lead to forming of boosting classifierWeifeng, Baobao, ZhiQiang, FangZhi, and Qiang (2021). The research was conducted under 2 different conditions in CNN for the classification of waste images it proposes a system for classifying waste with the following classes: polyethylene terephthalate, high-density polyethylene, polypropylene, and polystyrene resulting in an effective mechanism for classification. Using CNN, the research was conducted under 2 different conditions for analyzing waste images. It proposes a method for classifying waste into four different classes: polyethylene terephthalate, high-density polyethylene, polypropylene, and polystyrene, resulting in a mechanism that is effective Bobulski and Kubanek (2019). The CNN and SVM were used to classify the images using computer vision algorithms to classify the recyclability status of waste.Yang and Thung (2016).

Methodology

Data Collection

We get our dataset from Kaggle, platform for getting dataset and competitions. The dataset we used was create by someone named Sekar (2019) and has a collection of 22,500 images for different types of waste, with use of classification waste divided to the organic or non-organic. These pictures were uploaded about four years ago and have been shared on Kaggle for researchers like us to explore.

Number of Images: 22,500

Categories: Organic, and Non-organic.

Image Format: JPG

Total Size: 448 MB

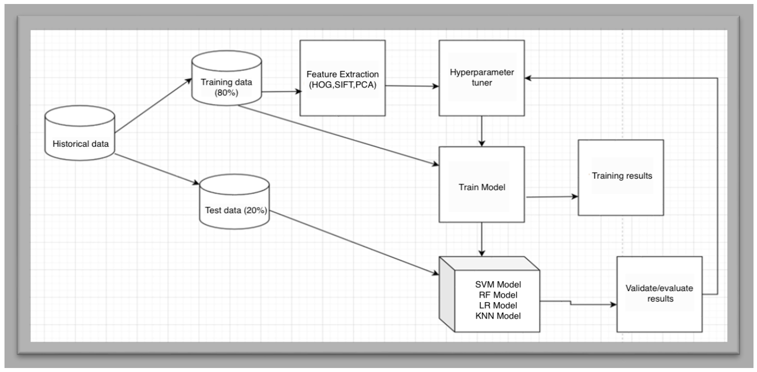
Dataset is divided into train data (85%) and test data (15%)

Training data - 22564 images Test data - 2513 images

Architecture of the Project

Figure 1

Architecture of the project



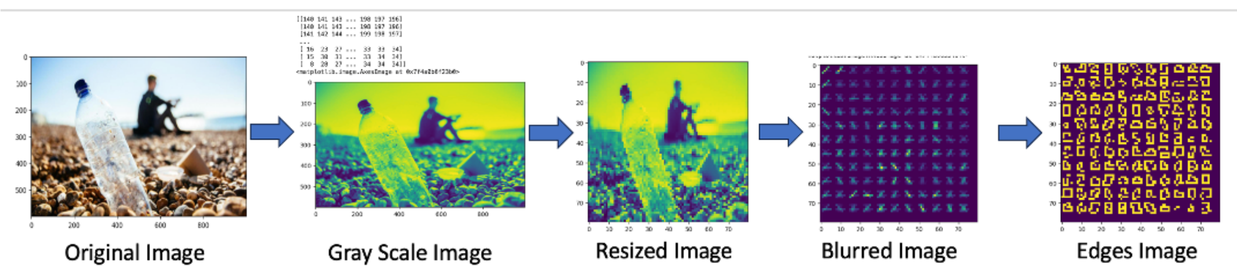
Data Preprocessing

Feature Engineering

Histogram of Oriented Gradients: HOG is an image feature descriptors to describe the image based on the gradient's directions and magnitudes. At the current time, this project supports calculating the following: Horizontal and vertical gradients. We have applied the featuring engineering to all the datasets and followed the steps shown in the figure below. In Figure 2It was observed that the images at the blur stage gave better accuracy than those at the edges.

Figure 2

Histogram of Oriented Gradients(HOG)



Feature engineering for SVM using PCA: Feature engineering is employed to preprocess raw image required for SVM classification. We first resized all the images and then flattened them and extracted the pixel values, which are features (X values) and retrieved the name associated with image as (Y label). Using these features for SVM classification is NP-Hard problem the run time would be forever. To deal with this problem we had to normalize the features and reduce the dimensionality of features. We used standard scaler to normalize features and incremental PCA for dimensionality reduction, using PCA we reduced the dimension of multi-dimensional data to 7.5K features. We also had to run these images in batches and extract PCA to optimize the run time, thus our runtime was reduced and efficient by 80%.

Model Development:

Support Vector Machine Support Vector Machines (SVM) is supervised machine learning algorithms that can be used for both classification as well as regression. Our dataset is labeled and for image classification SVM works best as it can handle nonlinear data efficiently. Below are the steps we followed to build a model.

1) Extract features of images using PCA.
2) Train and Test data (80% train and 20% test).
3) Choosing hyperparameters (C= 1) and Kernel function (RBF).
4) Experiment with different hyperparameters and choose the best parameter.
5) Fit model using SVM
6) Evaluate performance of model.
7) Predict whether the given image is organic or recyclable.

K-Nearest Neighbours It relies on the idea that similar data points tend to have similar labels or values. It our experiment we are using KNN for the classification of images using HOG for feature engineering. The performance was tested using various parameters of k. It was found that K=20 gave the highest accuracy. KNN was testing using brute, kd_tree, and ball_tree, kernels. It was observed that the accuracy for all the kernels was the same, whereas brute gave the best computational speed. Therefore, we finalized working with brute. After using PCA for dimensionality reduction using 2k components the accuracy was increased by 68%.

1) Feature engineering using HOG with different image sizes, blur intensity, and bin size.
2) Training and Test data(80% train and 20% test).

3) Choosing Hyper parameter K =[3,5,7,10,15,20]

4) Fit the model for each K

5) Testing with kernel brute, kdd_tree, ball_tree

6) The KNN using PCA

7) Fit the model for each K=20

8) Predict whether the given image is organic or recyclable.

Logistic Regression: Logistic Regression is a supervised machine learning algorithm used for binary classification tasks. In the context of image classification, we Use this Model to see if it can be effective for handling linear relationships in the data.

1.Feature engineering using HOG with different image sizes, blur intensity, and bin size.

2.Split the data into training and testing sets (80% train and 20% test).

3.Choose hyperparameter max_iter.

4.Experiment with different max_iter values and choose the best, which was found to be 5.

5.Fit the model using using the scaled features (X_scaled) and the target variable (y_train).

6.Evaluate the performance of the model on the test set:

7.Predict whether the given image is organic or recyclable using the trained model.

Random Forest Random Forest is known for its robustness, capacity to manage high-dimensional datasets, and its resistance to overfitting. It proves to be versatile across various applications, including image classification, where the data often demonstrates intricate patterns and non-linear relationships. Below are the steps used for Random Forest:

1.Feature engineering using HOG with different image sizes, blur intensity, and bin size.

2.Training and Test data (80% train and 20% test).

3.Choosing Hyperparameters including max_features, n_estimators, max_depth and max_leaf_nodes

4.Fit the model after assigning values for each hyperparameter.

5.Evaluate the performance of the model on the test set.

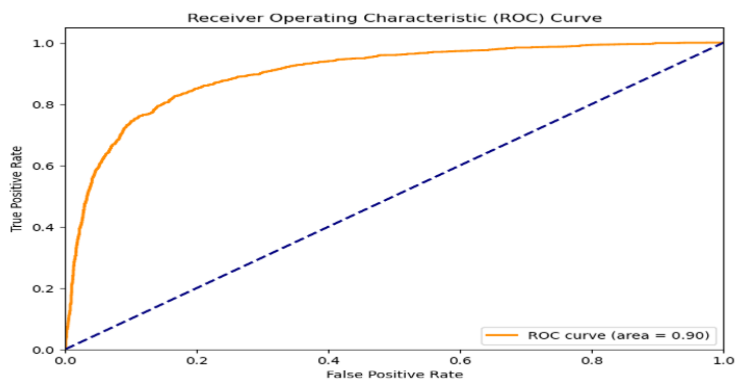
6.Predict whether the given image is organic or recyclable.

Model Evaluation and Testing:

We have evaluated the performance of model using various evaluation metrics such as Accuracy, Precision, Recall, F1-score, Cohen's kappa statistic, ROC-AUC curve, Specificity and MCC (Mathew's correlation coefficient).

Figure 3

ROC – AUC for SVM



Hyperparameter Tuning:

To improve the performance of the model, tuning the parameters is very important. So, we chose the parameter that resulted in the best accuracy

Table 2

Model Hyperparameters and Test Accuracy

Model	Hyperparameter	Kernel Function	Test Accuracy
SVM using PCA	$C = 1$	Radial basis function	82.64%
Logistic regression	Max_iter = 7	_	87.97%
KNN using PCA	$K = 20$	Brute	68%
SVM with HOG and PCA	$C = 1$	Radial basis function	70.95%
Random Forest	max_features, n_estimators	Linear kernel	66.42%
Random Forest	max_depth, max_leaf_nodes	Linear kernel	67.26%

Results:

With the machine learning models trained following the above process, we were able to provide the following enhanced capabilities

- Prediction of image whether it is organic or recyclable using SVM, KNN, RF and LR.
- Improved accuracy of model choosing hyperparameters and various feature extraction techniques such as HOG (Histogram of oriented gradients), PCA.
- Evaluated the performance of each model using various metrics such as Confusion Matrix, Accuracy, Precision, Recall, F1-score, ROC-AUC etc. Based on the results of each model we chose SVM – PCA for predicting new image because it has highest accuracy compared to other models and Cohen's kappa statistic also is satisfactory.

Figure 4

Test Result: Organic

Figure 5

Test Result:Recyclable

Technical Difficulties

In our project, we faced some technical challenges that required us to get a bit creativity to solve it. First, the project needed a lot of RAM and memory, and to tackle that, we turned to Google Colab's online services. Not only did this take the pressure off our local resources, but it also made collaboration a easier for us. Then, there was the issue of our dataset being quite heavy. The file size is about 448 MB with 22,564 JPG files. To handle this, we found a reliable storage, Google Drive. Next on the list were time complexity headaches, pushing us to dive into multiprocessing techniques to distribute and execute tasks efficiently. Despite our best efforts, we faced into system crashes, which has made us to fine-tune the multiprocessing parameters by scaling down the size of pools. Lastly, keeping shared resources like data and files in check was vital for easy collaboration, and Google Colab stepped up as our trusty solution. These smart moves helped us to fixing the problem and making sure it progressed without a hitch despite the initial tech hiccups.

Innovation

We believe our implementation has a variety of novel features. We significantly reduced (80%) the runtime of model training by integrating PCA for reducing dimensions without much accuracy loss. We experimented with a lot of hyperparameter tuning and settled down on the best set of hyperparameters. We studied and experimented with various feature extraction methodologies like SURF, SIFT, PCA, HOG etc and eventually chose PCA and HOG as it offered the best accuracy.

Prospects of winning competition / publication

Our initial implementation has shown quite encouraging outcomes. Enhancing the model’s performance and training it with additional data are the tasks that need to be completed. We think we can publish our findings in journals and the technical community with additional work, time, and peer review.

Lessons learned

- To optimize the run time performance of training images (nearly 25k images), we need to reduce the image size and flatten it otherwise the run time will be forever and use dimensionality reduction techniques such as PCA.

- Image data have more dimensions in our case after PCA it was 7500 features, which means that RBF kernel works best in SVM model, but processing time is slightly increased when compared to kernel function using polynomial.

- At the time of feature engineering selected appropriate hyperparameters is a tedious task, the model accuracy varies after each stage of feature engineering. It is not necessary to get higher accuracy from the final stage of feature engineering. We can get it from one step prior too.

- Whenever the images size was increased for testing purposes at the feature engineering stage a higher value of K is needed. The KNN doesn't perform that well with the larger image dataset in this case. Furthermore, when KNN using PCA was used with 2K features it provided better performance than before.

- Accuracy increased slightly for Random Forest classification after normalizing the data. Performing a multi-fold cross validation helps in finding the most accurate performance of the model. Performance of the model further elevated on tuning various hyperparameters including max_features, n_estimators, max_depth, max_leaf_nodes, etc. though tuning some of these parameters led to a larger used of resources.

- Random forest did relatively well when the depth and breadth of the tree was not very huge.

Future Scope

We can add more advanced models to improve the accuracy score and more computational resources to fine-tune the hyperparameters. In the future scope, the models will be deployed on real-time devices to segregate the organic and non-organic waste and reduce the carbon footprint in the environment. Also using organic waste classification can further help the environment to get good fertilizers for agriculture.

Relates to sustainability

Waste has numerous negative effects on our world, our health, and many other areas. The goal of choosing this subject is to present positive features of waste classification. The primary cause of waste is the rise in health issues and the invitation of various diseases in the body and environment by physical waste sorting, which also takes a lot of time and energy. Our project intends to automate the trash classification process through the use of machine learning models. As a result, it will speed up the classification process while protecting employees from harm. Therefore, the goal of our project is in line with sustainability principles by minimizing labor work and ensuring the safety of the environment and works.

Practiced pair programming

We used Google Collab and teams to discuss our issues and had a call to resolve it together. Using the google collab both the teams were able to see the file and using team were able to execute and debug the error.

Conclusion

After exploring the various accepts of feature engineering and dimensionality reduction we were able to identify the important features and roles of dimensionality reduction. Using PCA for dimensionality reduction helped in increasing the accuracy measures. With all the test results it was observed that SVM using PCA gave the maximum accuracy results of 83% followed by logistic regression at 72%. It was learned during the process that whenever the image size was increased for testing purposes at the feature engineering stage a higher value of K was needed. In this case, K=20 gave the best results with 2000 dimensions. In random forest and SVM using SIFT the accuracy of the model performance was average. So we further select SVM using PCA as our final model for evaluation.

- Observations: Whenever the images size was increased for testing purpose at the feature engineering stage the higher value of K is needed.

References

Anh Khoa, T., Phuc, C. H., Lam, P. D., Nhu, L. M. B., Trong, N. M., Phuong, N. T. H., & Duc, D. N. M. (2020). Waste management system using IoT-based machine learning in university. *Wireless Communications and Mobile Computing*, 1–13. doi: 10.1155/2020/6138637

Behera, S. K., Sr, A. B., Vasundhara, Saisudha, N. H., & Priya, S. (2020). AI based waste classifier with thermo-rapid composting. In *2020 international conference on power, energy, control and transmission systems (ICPECTS).*

Bobulski, J., & Kubanek, M. (2019). Waste classification system using image processing and convolutional neural networks. In *Advances in computational intelligence* (pp. 350–361). Springer International Publishing.

Chhabra, M., Sharan, B., Gupta, K., & Astya, R. (2022). Waste classification using improved CNN architecture. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4157549

Dubey, S., Singh, P., Yadav, P., & Singh, K. K. (2020). Household waste management system using IoT and machine learning. *Procedia Computer Science*, *167*, 1950–1959. doi: 10.1016/j.procs.2020.03.222

Lakhouit, A., Shaban, M., Alatawi, A., Abbas, S. Y. H., Asiri, E., Al Juhni, T., & Elsawy, M. (2023). Machine-learning approaches in geo-environmental engineering: Exploring smart solid waste management. *Journal of Environmental Management*, *330*(117174). doi: 10.1016/j.jenvman.2022.117174

Nanni, L., Ghidoni, S., & Brahnam, S. (2017). Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, *71*, 158–172. doi: 10.1016/j.patcog.2017.05.025

Nežerka, V., Zbíral, T., & Trejbal, J. (2023). Machine-learning-assisted classification of construction and demolition waste fragments using computer vision: Convolution versus extraction of selected features. *Expert Systems with Applications*, *238*(121568). doi: 10.1016/j.eswa.2023.121568

Puspaningrum, A. P., Endah, S. N., Sasongko, P. S., Kusumaningrum, R., Khadijah, R., &

Ernawan, F. (2020). Waste classification using support vector machine with SIFT-PCA feature extraction. In *2020 4th international conference on informatics and computational sciences (ICICoS).*

Rutqvist, D., Kleyko, D., & Blomstedt, F. (2020). An automated machine learning approach for smart waste management systems. *IEEE Transactions on Industrial Informatics*, *16*(1), 384–392. doi: 10.1109/tii.2019.2915572

Sami, K. N., Amin, Z. M. A., & Hassan, R. (2020). Waste management using machine learning and deep learning algorithms. *International Journal on Perceptive and Cognitive Computing*, *6*(2), 97–106. doi: 10.31436/ijpcc.v6i2.165

Sekar, S. (2019). *Waste classification data (version 1).* `https://www.kaggle.com/datasets/techsash/waste-classification-data/data`. (Kaggle Dataset)

Wang, Z., Peng, B., Huang, Y., & Sun, G. (2019). Classification for plastic bottles recycling based on image recognition. *Waste Management (New York, N.Y.)*, *88*, 170–181. doi: 10.1016/j.wasman.2019.03.032

Weifeng, W., Baobao, Z., ZhiQiang, W., FangZhi, Z., & Qiang, L. (2021). Garbage image recognition and classification based on HOG feature and SVM-Boosting. *Journal of Physics. Conference Series*, *1966*(1), 012002. doi: 10.1088/1742-6596/1966/1/012002

Yang, M., & Thung, G. (2016). Classification of trash for recyclability status. *CS229 project report*, *2016*(1), 3.

Appendix

Appendix

Sr. No.	Purpose	ML Algorithm	Waste Classification
1	To classify biodegradable waste for reducing pollution	CNN:88.54%	Biodegradable
2	To classify plastic bottles based on position and color during recycling	SVM:94.7%	Plastic Bottles
3	To use computer vision system for classification of handcrafted and non-handcrafted features	CNN:>90%, SVM:>92%	Handcrafted and non-handcrafted
4	To extract features using GB and MLP and fragments CDW using RGB images	MLP:91.3%, GB:92.3%, CNN:85.9%	Construction and demolition
5	To classify household waste at individual and community level	KNN:93%	Household
6	To collect waste from bins when full by defining the path towards the bins	LR:97%	Smart Bins
7	To collect and organize glass, paper, metal, plastic, cardboard, and waste images	SVM:85%, RF:55%, DT:65%, CNN:90%	Glass, paper, metal, plastic, cardboard
8	To find amount of waste generated daily from families between 2010-19	LiR:87%, SVM:88%	Glass, paper, plastic
9	To collect information from smart bins and notify management to empty bins using automated ML process	KNN:96.3%, LR:96.6%, SVM:96.8%, DT:96.6%, RF:97.2%	Smart Bins
10	To find out SFIT-PCA features for waste classification using feature extraction	SVM:62%	Glass, paper, cardboard, metal, plastic
11	To classify waste using AI-based waste classifier with thermo-rapid composting	SVM:>85%, CNN:>85%	Industrial, biodegradable, non-biodegradable
12	To optimize SVM using boosting methods to classify different types of waste images	SVM:95%	Cans, cigarette butt, plastic bottle, carton
13	To classify images based on different layers of CNN	CNN	Plastic
14	To classify trash for recyclability status using computer vision	SVM:63%	Glass, paper, metal, plastic, cardboard, trash

Table A1

Waste Classification using Machine Learning Algorithms

Table A2

Model Performance

Model	Model Accuracy (%)	Sensitivity (%)	Precision (%)	F1-Score (%)	Cohen's kappa st
SVM using PCA	82.64	77.47	82.11	79.72	64.56
Logistic Regression	72.20	78.3	53.6	63.6	32.3
KNN	65.62	R=99 O=0 Overall=45.21	R=44.57 O=0 Overall=0.01	R=61.65 O=0 Overall=62.27	18
KNN using PCA	68.58	R=67.98 O=69.06 Overall=65.11	R=64.46 O=72.33 Overall=65.88	R=66.18 O=70.66 Overall=65.06	36.88
Random Forest	66.71	R=57 O=74	R=64 O=69	R=60 O=71	32.93
SVM using SIFT	56.15	1.36	46.15	2.64	0.12
SVM using HOG	67.76	67.76	67.84	67.79	35
SVM with HOG and PCA	70.95	70.95	71.28	71.01	41.79

Table A3

Software Resources and Configurations

Software Resource	Configuration	Purpose
Python	Version 3.10.0	Data processing, analysis, machine learning implementation
Microsoft 365	Version 16.78.3 (23102801)	Utilize PowerPoint for presentations, Teams for meetings and team alignment, and Excel for
Overleaf	Online Platform	Collaborative LaTeX editor for writing and editing project reports, research papers, and academic documents with version control.
Zoom	Version 5.16.2	Virtual team meetings and discussions
Google Drive	Online Platform	Cloud-based storage and collaboration platform for storing, sharing, and collaborating on documents, spreadsheets, presentations, and other project-related files.
Google Colab	Online Platform	Cloud-based storage and collaboration platform for storing, sharing, and collaborating on documents, spreadsheets, presentations, and other project-related files.
Azure DevOps	Online Platform	Essential for project management and task tracking
GitHub	Online Platform	Codebase publicly accessible
Prezi	Online Platform	Make creative and innovative presentations
Grammarly	Version 2.0.0	Spelling and grammar checking

CRediT author statement

Sr. No.	Tasks	Krinal	Maral	Meghana	Sejal	Yukta
1	Data Collection	20%	20%	20%	20%	20%
2	Data Preprocessing(HOG)	5%	5%	5%	5%	80%
2	Data Preprocessing(PCA)	5%	5%	80%	5%	5%
2	Data Preprocessing(Normalization)	5%	5%	80%	5%	5%
3	Exploratory Data Analysis	20%	20%	20%	20%	20%
4	Model Selection	20%	20%	20%	20%	20%
5	Model Development	20%	20%	20%	20%	20%
6	Model Training and Testing	20%	20%	20%	20%	20%
7	Performance Evaluation	20%	20%	20%	20%	20%
8	Fine-tuning and Optimization	20%	20%	20%	20%	20%
9	Final Testing and Validation	20%	20%	20%	20%	20%
10	Deployment	20%	20%	20%	20%	20%
11	Report Writing and Documentation	12.5%	50%	12.5%	12.5%	12.5%
12	Presentation Preparation	20%	20%	20%	20%	20%

Table A4

File Information and Links

Sr. No	File Name	Type	Usage	Links for Collab Notebo
1	Exploratory Data Analysis.ipynb	python	Exploring the data distribution	`https://github.com/webrockerz2020/waste_classifcation_traditional_machine_learning/blob/main/Exploratory_Data_Analysis.ipynb`
2	waste_classification_ml.ipynb Feature Engineering (HOG), KNN Model Designing, KNN Evaluation, Logistic Regression Modeling, Random Forest Modeling, SVM	python		`https://github.com/webrockerz2020/waste_classifcation_traditional_machine_learning/blob/main/waste_classification_ml.ipynb`
3	Waste_Classification_SVM_PCA.ipynb Feature Engineering (PCA), Support Vector Machine, Evaluation Metrics, Model Testing/Prediction, ROC – AUC curve	python		`https://github.com/mhegde95/Waste-Classification-ML/blob/main/Waste_Classification_SVM_PCA.ipynb`

Continued on the next p

Table A4 – Continued from previous page

Sr. No	File Name	Type	Usage	Links for Collab Notebo
4	model_testing.ipynb	python	To Evaluate the models using pickle files	https://github.com/webrockerz2020/waste_classifcation_traditional_machine_learning/blob/main/model_testing.ipynb
5	Experiment_SIFT_SVM.ipynb	python	SVM using SIFT Experimental purpose	https://github.com/mhegde95/Waste-Classification-ML/blob/main/Experiment_SIFT_SVM.ipynb
6	best_logistic_regression_model.pkl	pickle	Logistic model	https://drive.google.com/drive/folders/1RPKEwGgR5eEu16w2Rp5-Puz9r1cy?usp=sharin
7	knn_model_brute.pkl	pickle	KNN parameters	
8	hog_features_train.pkl	pickle	HOG parameters	
9	train_features.pkl	pickle	KNN train features	
10	svm_hog_model.pkl	pickle	SVM mode using HOG	
11	SVM_UsingPCA.pkl	pickle	SVM model using PCA	
12	randomForestClassifier.pkl	pickle	Random Forest using PCA	
13	Group1_waste_classification Report	Latex		https://github.com/mhegde95/Waste-Classification-ML

Continued on the next p

Table A4 – Continued from previous page

Sr. No	File Name	Type	Usage	Links for Collab Notebo
14	Group1_wate_classification.pdf	pdf	Slides	

The complete project is uploaded on the
https://github.com/mhegde95/Waste-Classification-ML.

Google Drive folder: https://drive.google.com/drive/folders/1KlhxN6_5MA3q6Q4xJM7OIjY_6h7a61rM?usp=sharing.