# Garbage Classification using Traditional Machine Learning Approaches

Lai Ka Chung 22080062d

*Abstract*—**Proper garbage classification is crucial for environmental sustainability. This paper presents a traditional machine-learning approach to classify garbage into 10 categories: battery, biological, cardboard, clothes, glass, metal, paper, plastic, shoes, and trash. We implement and compare three feature extraction methods (HOG, LBP, and SIFT) with three classifiers (SVM, KNN, and Random Forest) to identify the most effective combinations. The HOG+SVM model achieved the highest individual accuracy at 51.31%, while an ensemble approach improved performance to 63.25%. Testing on complex scenes revealed significant challenges, including multi-object scenes, background interference, and feature extraction limitations. We propose improvements through scene segmentation, enhanced feature engineering, advanced ensemble methods, and data augmentation. This research demonstrates both the potential and limitations of traditional machine learning for garbage classification in real-world scenarios.**

## I. Introduction

Garbage classification is a critical component of effective waste management systems worldwide. Proper waste segregation enables efficient recycling, reduces landfill usage, and minimizes environmental impact. While manual sorting remains common, automated classification systems offer promising solutions to improve accuracy and processing speed. Traditional machine-learning approaches provide valuable insights into the fundamental challenges of waste classification. Unlike deep learning methods that may obscure underlying processes, traditional approaches enable a clear analysis of feature extraction techniques, classification algorithms, and their interactions. This transparency is particularly valuable for understanding classification failures and identifying targeted improvements. In this paper, we address the challenge of garbage classification using traditional machine learning methods. We explore how different feature extraction techniques capture distinct aspects of waste items, how various classifiers perform with these features, and how robust these approaches are when faced with complex, real-world scenarios. Our research focuses on several key objectives:

Creating a balanced dataset representing common waste categories Implementing and comparing different feature extraction methods (HOG, LBP, and SIFT) Evaluating the performance of multiple classification algorithms (SVM, KNN, and Random Forest) Testing model robustness in complex, cluttered scenes Identifying specific challenges and proposing practical improvements

By restricting our approach to traditional machine learning techniques rather than deep learning, we aim better to understand the fundamental requirements for effective garbage classification. This understanding can inform future research using traditional or deep learning approaches by highlighting critical challenges and potential solutions in this domain.

## II. Dataset

### A. Dataset Collection and Organization

There are 10 various materials which are commonly found in daily life. The original size of the dataset is 15214 samples. The dataset was organized into the following categories: battery with 945 samples, biological with 985 samples, cardboard with 891 samples, clothes with 5325 samples, glass with 1800 samples, metal with 769 samples, paper with 1050 samples, plastic with 865 samples, shoes with 1977 samples and trash with 697 samples. To prevent the training bias problem, every class will only take 200 samples to train the model randomly, which is 2000 samples.

### B. Data Preprocessing

For consistency and better performance for the result, all the images have been processed by the following preprocessing step:

1. Resize: All images were resized to a dimension of 128x128 pixels to maintain uniformity and reduce computational load.

2. Color conversion: All the images were converted to RGB format and converted to grayscale for feature extraction.

3. Normalization: The pixel values were normalized for equal feature contribution and to improve the classifier accuracy.

### C. Data Splitting

The dataset was split into training and testing sets: 80% for training (160 samples per class) and 20% for testing (40 samples per class).

### D. Dataset Visualization

Fig-1 shows the distribution of the original dataset and the distribution after sampling.

## III. Feature Extraction

Three feature extraction algorithms were implemented to capture different aspects of the images: a histogram of Oriented Gradients (HOG) for shape features, Local Binary Patterns (LBP) for texture features, and a Scale-Invariant Feature Transform (SIFT) for point features.

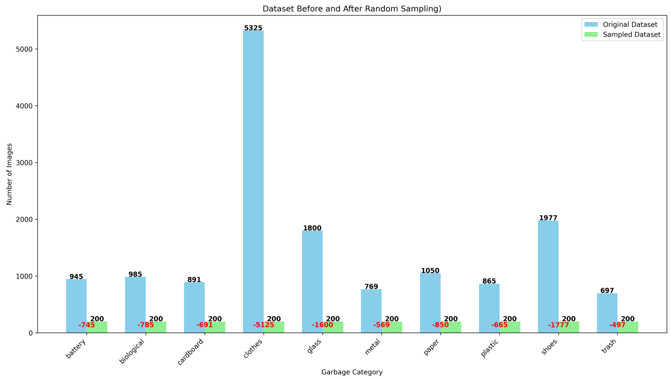



Fig. 1. Dataset Before and after sampling

### A. Histogram of Oriented Gradients (HOG)

HOG features capture the disturbance of gradient direction in an image, extracting the global feature of the image, which is effective in detecting shape-based patterns. In the implementation, we use eight orientation bins to provide sufficient angular resolution without over-segmenting the gradient directions, 16×16 pixels per cell to capture broader shape characteristics, 2x2 cells per block and L2-Hys for Block normalization. HOG features were particularly effective for distinguishing items with distinct shapes, such as cardboard boxes, glass bottles, metal cans, and batteries.

### B. Local Binary Patterns (LBP)

LBP features capture global texture patterns, which are common internal characteristics of the object surface. These patterns contain information about the structure and its relationship to the surrounding object. In the implementation, the radius of 3 with 24 sampling points provides a balance between local detail and computational efficiency. Also, uniform LBP reduces the feature dimensionality, which retains the most important texture patterns. LBP features were effective for items with distinct patterns, such as paper, clothes, biological waste, and shoes.

### C. Scale-Invariant Feature Transform (SIFT)

SIFT features to detect and describe local features in images invariant to scaling and rotation and partially invariant to illumination changes. To implement it, Bag of Visual Words combines 100 visual words with a maximum of 100 key points per image. For the feature vector, the Histogram of visual word occurrences is chosen. SIFT features were effective for items with distinct patterns and point features, such as plastic packaging with logos, metal cans with labels, and glass with reflective surfaces.

### D. Feature Visualization and Analysis

These features are visualized for representative images from each category to better understand the extracted features. The figure shows the original image with their HOG, LBP, and SIFT feature visualizations.

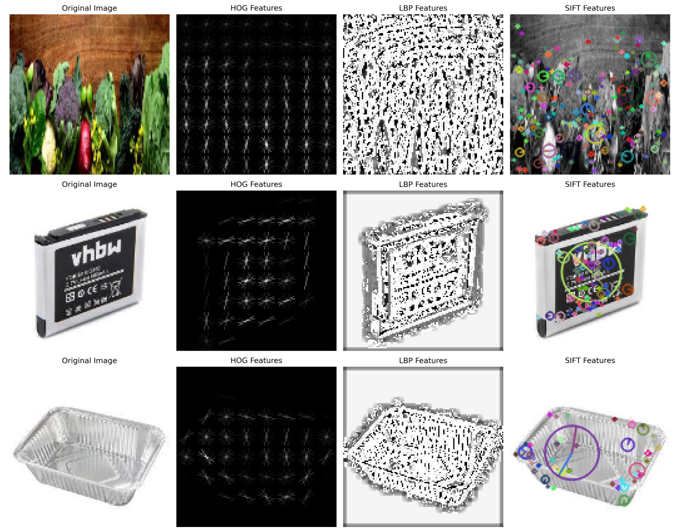



Fig. 2. Feature Visualization

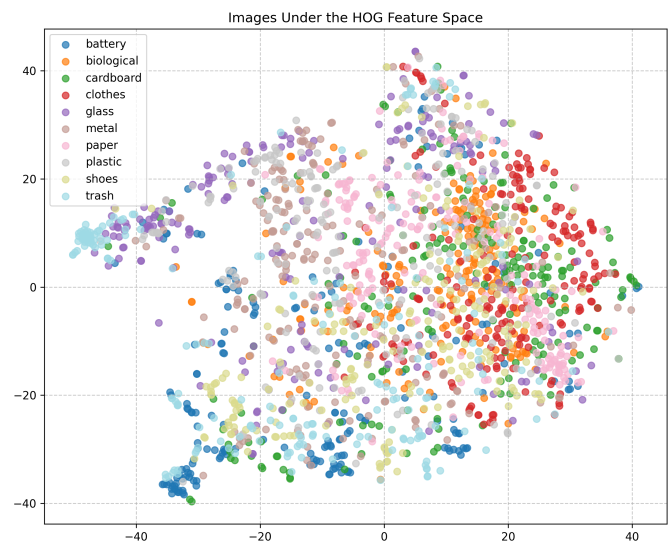



Fig. 3. HOG feature space

Moreover, t-SNE(t-Distributed Stochastic Neighbor Embedding) was implemented to project the high-dimensional feature spaces onto 2D planes for visualization.

Figure 3 show the HOG feature space. Some clusters are for certain categories, particularly for items with well-defined shapes, like batteries and metal cans. However, there is significant overlap between categories with similar structural properties, such as paper and cardboard, or between clothes and shoes. While HOG effectively captures shape-based characteristics, it struggles to differentiate categories where shape alone is insufficient.

Figure 4 demonstrates a more complex distribution with multiple sub-clusters within each category. It suggests that texture patterns within waste categories can vary significantly. The visualization reveals a more defined separation between organic materials (biological waste) and manufactured items,

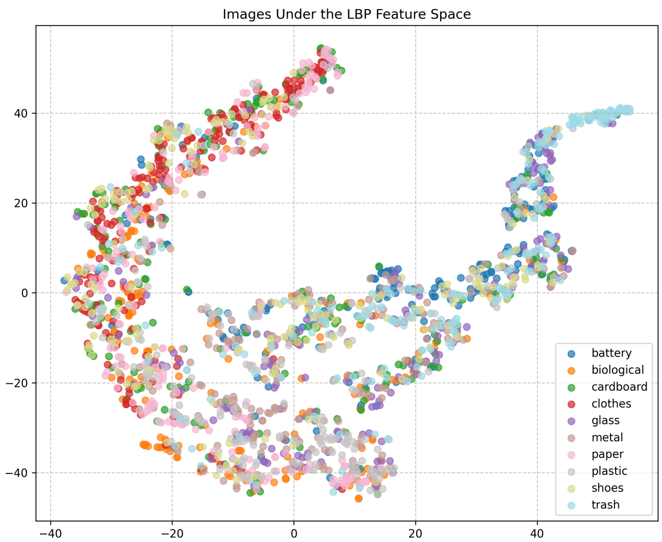



Fig. 4. LBP feature space

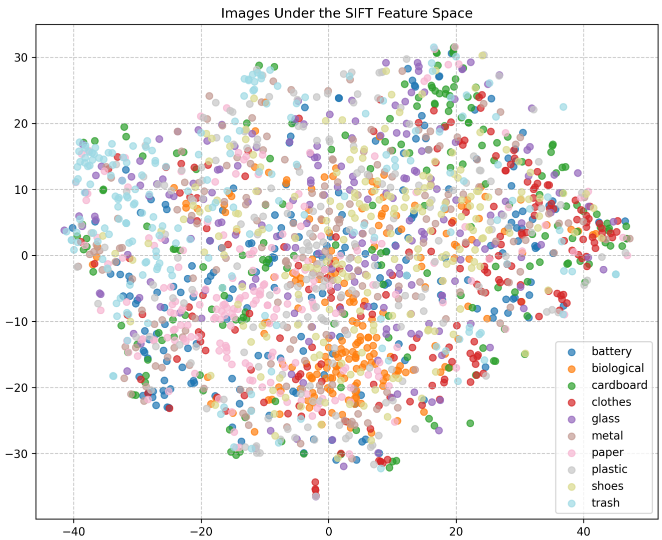



Fig. 5. SIFT feature space

Better discrimination between paper and plastic than the HOG feature space and Significant overlap between categories with similar texture patterns (clothes and shoes).

Figure 5 shows the most dispersed distribution of the three methods. Unlike HOG and LBP, which form somewhat coherent clusters, SIFT creates a more scattered representation with numerous small clusters. This reflects SIFT's focus on distinctive local key points rather than global patterns.

## IV. Methodology

### A. Classifier Selection and Implementation

Three different classifiers were implemented and evaluated:

*1) Support Vector Machine (SVM):* SVM finds an optimal hyperplane that maximizes the margin between different classes in the feature space. To implement SVM, Grid searches for the best parameter of the kernel (linear or Radial Basis Function), C values (0.1, 1, 10, 100) and gamma (scale, auto, 0.01, 0.1, 1).

*2) K-Nearest Neighbors (KNN):* KNN classifiers are based on the majority class among its nearest k neighbors in the feature space. To implement KNN, a Grid search is used to find the best parameter of K-values(3, 5, 7, 9, 11), Distance metrics(Euclidean or Manhattan), and Weighting(Uniform or distance-based).

*3) Random Forest (RF):* Random Forests combine multiple decision trees to improve generalization and robustness. To implement Random Forest, Grid search for the best parameter of the Number of trees(50, 100, 200), Maximum depth(None, 10, 20, 30), and Minimum samples for split(2, 5, 10).

### B. Feature Scaling and Normalization

Before training the model, all the feature vectors were standardized using StandardScaler from scikit-learn, standardizing features by removing the mean and scaling to the unit. This preprocessing helps with fair comparisons of different models and more accurate results.

### C. Hyperparameter Optimization

GridSearchCV with 5-fold cross-validation was used to find the best hyperparameter for each model. This approach automatically finds the best parameter combination from the parameter space to maximize the classification accuracy on the validation set.

### D. Ensemble Method

Create an ensemble by combining all the strengths of different feature types and classifiers, which is a soft voting classifier based on the probability for the ensemble. Also, the weighting from all models is equal, which can enhance and improve the robustness by combining diverse classification models with their strengths and biases.

### E. Evaluation Metrics

There are four metrics to evaluate classifier performance:

- **Accuracy:** Overall percentage of correctly classified samples
- **Precision, Recall, and F1-score:** Per-class and weighted average metrics
- **Confusion Matrix:** To analyze patterns of misclassifications
- **Cross-validation score:** To assess model stability and generalization

Comparing these metrics will help evaluate the best classifier.

### F. Testing on Complex Scenes

To evaluate the applicability of our model in the world, a test set of four complex scene images:

- **Messy kitchen with grabage**
- **Office desk with trash**
- **Restaurant table leftovers**

- **Beach pollution**

These scenes present significant challenges compared to the controlled training dataset, as they contain:

- Multiple waste types in a single image
- Environmental context (background elements)
- Varied lighting conditions
- Occlusion (objects overlapping)
- Scale variations

## V. Results and Analysis

### A. Classifier Performance on Test Set

TABLE I
Model Performance Comparison

| Model/Classifier | Accuracy | Precision | Recall | f1-score |
|---|---|---|---|---|
| hog_svm | 51.31% | 0.55 | 0.55 | 0.55 |
| hog_knn | 46.19% | 0.47 | 0.47 | 0.46 |
| hog_rf | 46.44% | 0.47 | 0.47 | 0.47 |
| lbp_svm | 43.25% | 0.43 | 0.42 | 0.42 |
| lbp_knn | 39.38% | 0.37 | 0.37 | 0.36 |
| lbp_rf | 41.31% | 0.39 | 0.39 | 0.38 |
| sift_svm | 38.50% | 0.37 | 0.37 | 0.36 |
| sift_knn | 28.38% | 0.32 | 0.30 | 0.28 |
| sift_rf | 39.44% | 0.40 | 0.40 | 0.39 |
| Ensemble | 63.25% | 0.64 | 0.63 | 0.63 |

Table I shows the confusion for all the models with all the feature extraction methods. The combination of HOG with SVM achieved the highest individual accuracy with 51.31%The ensemble method further improved to 63.25%.

### B. Performance on Complex Scenes

Table II summarizes the predictions across all models for each complex scene.

### C. Analysis of Individual Complex Scenes

*1) Beach Pollution Scene:* The image of beach pollution primarily contains plastic waste (bottles) mixed with other debris on the sand. This scene presents a significant challenge for our classifiers:

- **Inconsistent predictions:** Models produced varied predictions, including cardboard, paper, biological, clothes, and shoes.
- **Misleading confidence:** HOG+KNN predicted ”paper” with 100% confidence, despite the image clearly containing plastic waste.
- **SIFT models:** Both SIFT+SVM and SIFT+KNN classified the scene as ”shoes” with relatively high confidence (36.53% and 42.38%).
- **Ensemble model:** Predicted ”paper” with 30.03% confidence, failing to correctly identify plastic waste.

The failure to correctly identify plastic waste suggests our models may have difficulty with reflective surfaces and irregular shapes typical of plastic waste on beaches.

*2) Messy Kitchen Scene:* The messy kitchen scene contains a cluttered environment with various waste types:

- **Feature-dependent predictions:** HOG-based models predicted paper/cardboard, while LBP and SIFT models consistently predicted biological waste.
- **Highest confidence:** SIFT+KNN achieved 86.09% confidence for biological waste, showing strong consensus among SIFT-based models.
- **Ensemble model:** Predicted ”paper” (30.07%), following HOG-based models rather than the more consistent LBP/SIFT models.

This scene demonstrates how different feature extraction methods can capture distinct aspects of complex scenes, leading to varied predictions.

*3) Office Desk Scene:* The office desk image predominantly contains paper waste:

- **HOG models:** Generally predicted ”paper” correctly (except HOG+KNN → cardboard)
- **LBP models:** All predicted ”clothes,” suggesting textural confusion
- **SIFT models:** Split between biological waste and paper
- **Ensemble model:** Correctly predicted ”paper” with 39.81% confidence

This scene shows that HOG features might be more effective at capturing the structural elements of paper waste in office environments.

*4) Restaurant Leftovers Scene:* This scene contains food waste on plates and cups:

- **Consistent predictions:** Most models correctly classified this as ”biological” waste
- **Outliers:** HOG+KNN predicted ”cardboard” and SIFT+RF predicted ”paper”
- **Highest biological confidence:** SIFT+KNN (55.97%) and Ensemble (46.01%)

This scene demonstrates the highest consensus among models, suggesting that biological waste features are more consistently captured across different feature extraction methods.

### D. General Observations

- **Confidence vs. correctness:** High confidence predictions (e.g., HOG+KNN at 100% for beach scene) were not always correct.
- **Feature-specific strengths:** HOG features performed better for structured waste (paper), SIFT features excelled with biological waste, while LBP features showed mixed results.
- **Classifier comparison:** KNN models often produced higher confidence predictions than SVM or RF for the same features.
- **Ensemble limitations:** The ensemble model did not consistently outperform individual models, often defaulting to paper prediction.

TABLE II
SUMMARY OF MODEL PREDICTIONS ON COMPLEX SCENES

| Model | Beach Pollution | Messy Kitchen | Office Desk | Restaurant Leftovers |
|---|---|---|---|---|
| HOG+SVM | cardboard (28.57%) | paper (67.11%) | paper (70.39%) | biological (42.09%) |
| HOG+KNN | paper (100.00%) | cardboard (33.88%) | cardboard (66.82%) | cardboard (67.39%) |
| HOG+RF | cardboard (14.25%) | paper (28.43%) | paper (22.21%) | biological (33.08%) |
| LBP+SVM | biological (29.68%) | biological (28.50%) | clothes (23.68%) | biological (29.01%) |
| LBP+KNN | clothes (27.43%) | biological (33.27%) | clothes (36.94%) | biological (37.08%) |
| LBP+RF | paper (27.50%) | biological (29.00%) | clothes (32.50%) | biological (44.00%) |
| SIFT+SVM | shoes (36.53%) | biological (28.85%) | biological (45.64%) | biological (18.19%) |
| SIFT+KNN | shoes (42.38%) | biological (86.09%) | biological (43.55%) | biological (55.97%) |
| SIFT+RF | biological (18.73%) | biological (27.38%) | paper (24.88%) | paper (26.04%) |
| Ensemble | paper (30.03%) | paper (30.07%) | paper (39.81%) | biological (46.01%) |

## VI. DISCUSSION

### A. Challenges in Complex Scene Classification

Our experimental results highlight several critical challenges in garbage classification within complex scenes:

*1) Multi-object Scenes:* Real-world garbage scenes typically contain multiple waste items, often of different categories. It fundamentally challenges our classification approach, which was trained on single-object images. The models attempt to make a single prediction for the entire scene rather than identifying individual waste items.

*2) Background Interference:* Environmental contexts (beach sand, kitchen surfaces, office materials) introduce visual elements not present in the training data. These backgrounds can significantly influence feature extraction, as demonstrated by the beach scene, where the textural information from sand may have contributed to misclassifications.

*3) Feature Extraction Limitations:* Different feature extraction methods showed varying effectiveness across scenes:

- **HOG features:** Performed reasonably well with paper waste but struggled with irregular objects like plastic bottles and food waste. It suggests HOG's focus on gradient patterns works better for waste with consistent structural elements.
- **LBP features:** Demonstrated inconsistent performance across scenes. The textural information captured by LBP appears sensitive to background elements and lighting variations.
- **SIFT features:** Showed strong performance with biological waste but produced unexpected classifications in other scenes (e.g., classifying beach waste as ”shoes”). This suggests SIFT key points may detect similarities between visually distinct categories.

*4) Classifier Behavior:* We observed distinct patterns across classifiers:

- **SVM models:** Generally provided moderate confidence predictions (20-70%), showing reasonable balance between confidence and accuracy.
- **KNN models:** Often produced higher confidence predictions, sometimes reaching 100%, but these high-confidence predictions were occasionally incorrect, suggesting possible overfitting to training data.
- **Random Forest models:** Typically showed lower confidence (15-40%), perhaps reflecting the inherent uncertainty in complex scenes.

*5) Ensemble Model Performance:* The ensemble model did not consistently improve classification accuracy as expected. In particular:

- It often defaulted to predicting ”paper” across multiple scenes
- It sometimes ignored the consensus among several models in favour of feature types that may have been more dominant in the training data
- The simple voting mechanism may not adequately combine the strengths of different feature-classifier combinations

### B. Proposed Improvements

Based on our findings, we propose several approaches to improve garbage classification in complex scenes:

*1) Scene Segmentation:* Implementing an object detection or segmentation step before classification could significantly improve performance by:

- Isolating individual waste items from the background
- Allowing multiple classifications within a single image
- Reducing interference from background elements

*2) Feature Engineering:* Our results suggest potential improvements to feature extraction:

- **Combined features:** Creating hybrid feature vectors that leverage the strengths of HOG (structure), LBP (texture), and SIFT (distinctive points)
- **Color features:** Incorporating colour histograms, especially for distinguishing certain waste types (e.g., biological waste from paper)
- **Context-aware features:** Developing features that account for environmental context

*3) Advanced Ensemble Methods:* More sophisticated ensemble approaches could improve performance:

- Weighted voting based on model performance for specific waste types
- Stacking classifiers with a meta-learner to combine predictions more intelligently
- Context-specific model selection based on scene characteristics

*4) Data Augmentation:* Expanding training data to include:

- Waste items on various backgrounds
- Multiple waste items in a single image
- Varied lighting and perspective conditions
- Partially occluded waste items

### C. Real-world Applications and Limitations

Our findings have important implications for real-world garbage classification systems:

*1) Automated Waste Sorting:* For automated waste sorting applications, our results suggest:

- Current traditional machine learning approaches may be insufficient for complex, uncontrolled environments
- Controlled conditions (consistent lighting, isolated items) would significantly improve performance
- A hierarchical approach (first segregating waste from the background, then classifying) may be necessary

*2) Educational Applications:* For educational garbage classification systems:

- The system could provide reasonable guidance for clearly defined waste categories (e.g., food waste)
- Users should be cautioned about potential misclassifications in complex scenes
- Interactive systems allowing user correction could improve performance over time

*3) Comparison with Deep Learning:* While this assignment focused on traditional machine learning approaches, it's worth noting:

- Deep learning approaches (particularly CNNs) might handle complex scenes more effectively due to:
  - Hierarchical feature learning
  - Greater robustness to background variation
  - Better handling of multi-object scenes
- However, traditional methods offer advantages in:
  - Computational efficiency
  - Interpretability of features and decisions
  - Performance with limited training data

## VII. Conclusion

Our experimental investigation into garbage classification using traditional machine-learning approaches has yielded valuable insights into these methods' capabilities and limitations in real-world applications.

Through a systematic evaluation of three feature extraction techniques (HOG, LBP, and SIFT) combined with three classification algorithms (SVM, KNN, and Random Forest), we achieved a maximum individual model accuracy of 51.31% with HOG+SVM and an improved ensemble accuracy of 63.25%. While these results are promising for controlled images, our testing on complex scenes revealed significant challenges that must be addressed for practical implementation.

The feature space analysis demonstrated that each extraction method captures fundamentally different aspects of waste items: HOG effectively represents structural elements, LBP captures texture patterns, and SIFT identifies distinctive local features. This complementarity explains why our ensemble approach outperformed individual models, though not consistently across all complex scenes.

Several critical challenges emerged in our experiments with complex, real-world scenes:

1) **Multi-object environments** fundamentally challenge single-label classification systems trained on isolated waste items. Real-world garbage typically consists of multiple waste types in a single scene.
2) **Background interference** significantly impacts feature extraction, as environmental contexts (like beach sand or kitchen surfaces) introduce visual elements absent from training data.
3) **Feature extraction limitations** varied across waste types, with HOG performing better for structured waste, SIFT excelling with biological waste, and LBP showing mixed effectiveness depending on the scene.
4) **Classifier behaviour patterns** revealed that SVM provided balanced confidence-accuracy, KNN often produced overconfident predictions, and Random Forest typically showed appropriately lower confidence reflecting inherent uncertainty.
5) **Ensemble limitations** became apparent when our voting classifier sometimes defaulted to paper predictions or ignored the consensus of multiple models.

Based on these findings, we propose several promising directions to enhance garbage classification performance:

1) Implementing **scene segmentation** as a preprocessing step to isolate individual waste items from backgrounds and enable multiple classifications within a single image.
2) Developing **enhanced feature engineering** approaches, including hybrid feature vectors that combine the strengths of HOG, LBP, and SIFT, along with colour features, to distinguish certain waste types better.
3) Adopting **advanced ensemble methods** such as weighted voting based on category-specific performance or stacking classifiers with meta-learners to combine predictions intelligently.
4) Expanding training data through **augmentation techniques** to include waste items on various backgrounds, multiple items per image, varied lighting conditions, and partially occluded items.
5) Exploring **incremental learning approaches** that allow systems to improve over time as they encounter new examples.

For practical applications, our findings suggest that current traditional machine learning approaches require either controlled conditions or additional preprocessing steps to achieve robust performance. The accuracy gap between test set performance and complex scene results indicates that real-world garbage classification remains challenging, though targeted improvements could significantly enhance capabilities.

While deep learning approaches might offer higher baseline accuracy for complex scenes, our traditional machine learning investigation provides valuable transparency into specific failure modes and improvement opportunities. The interpretability of these models allows for more targeted enhancements and a better understanding of classification decisions.

In conclusion, while traditional machine learning approaches face significant challenges in complex garbage classification scenarios, they offer a foundation that can be systematically improved through the targeted enhancements we've identified. With these improvements, such systems could potentially achieve practical utility in educational contexts, semi-automated sorting, and other applications where perfect accuracy is not required, but interpretability and computational efficiency are valuable.

## APPENDIX

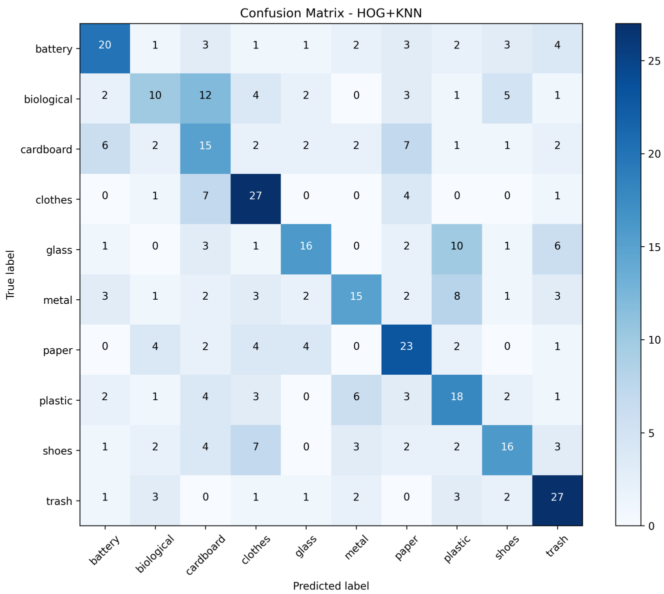



Fig. 6. Confusion matrix of HOG+KNN

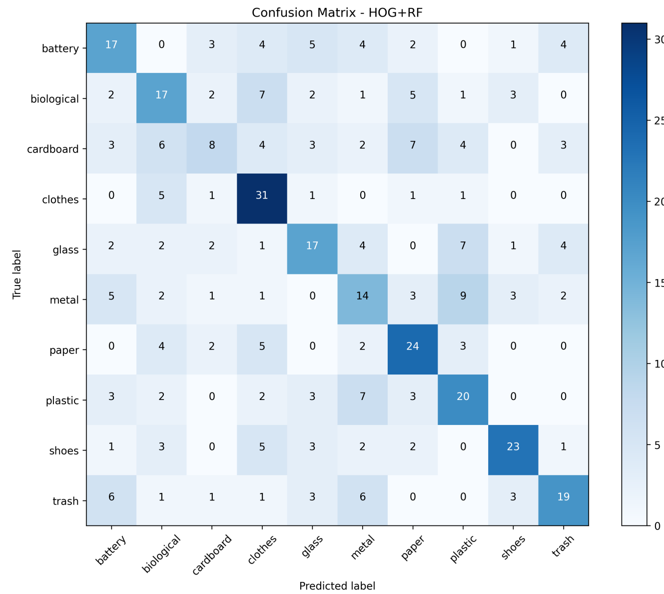



Fig. 8. Confusion matrix of HOG+RF

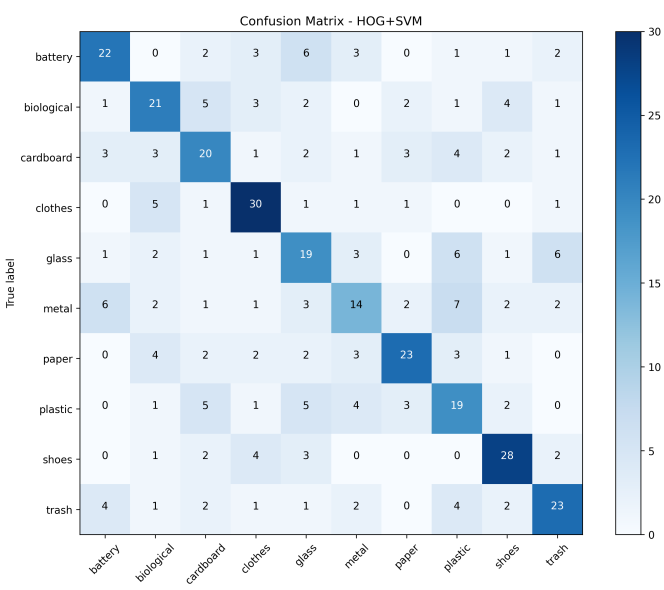



Fig. 7. Confusion matrix of HOG+SVM

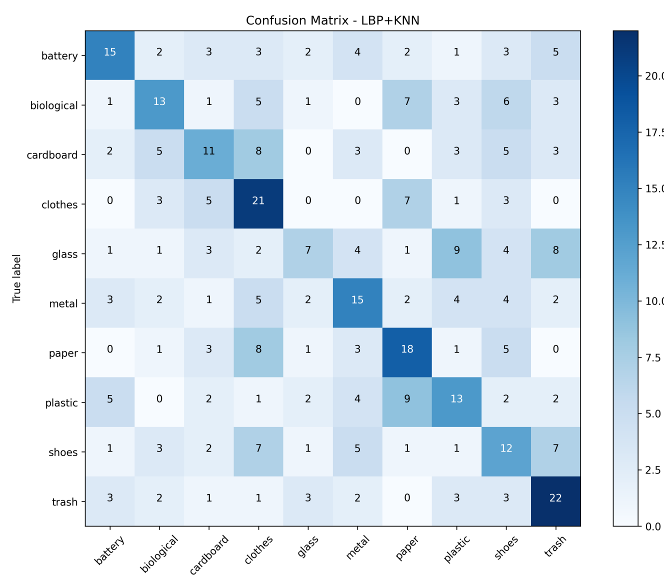



Fig. 9. Confusion matrix of LBP+KNN

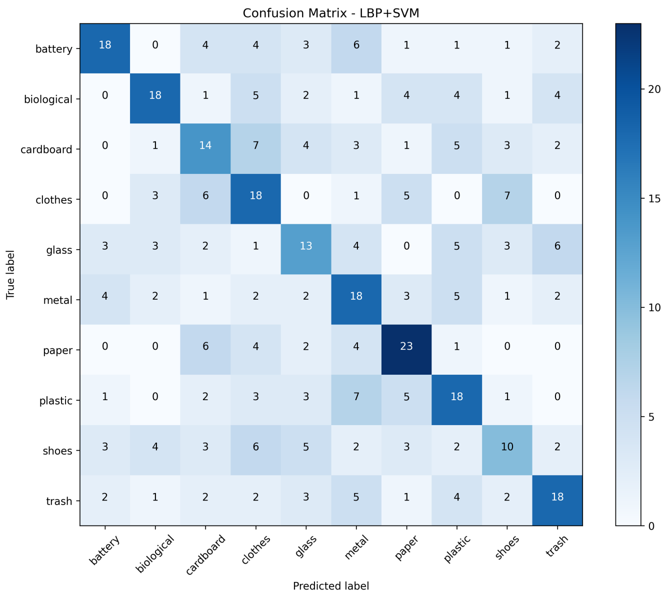



Fig. 10. Confusion matrix of LBP+SVM

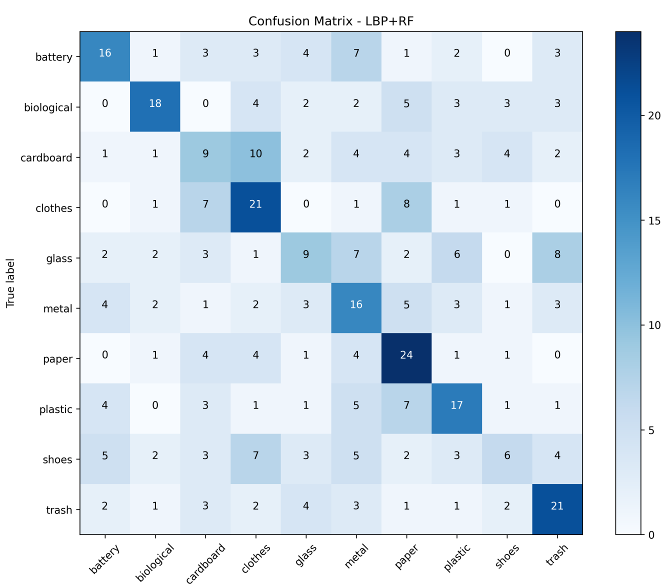



Fig. 11. Confusion matrix of LBP+RF

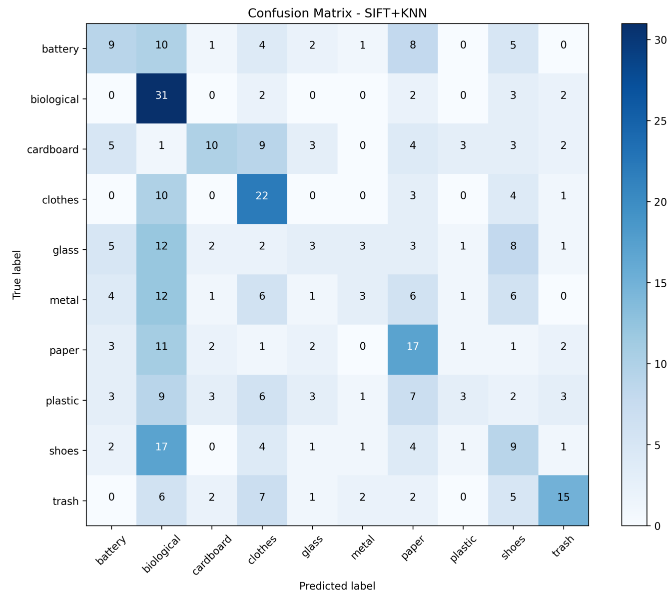



Fig. 12. Confusion matrix of SIFT+KNN

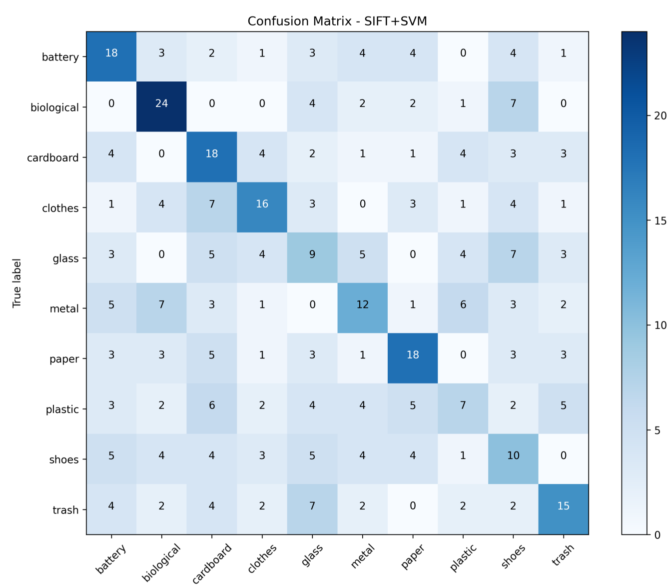



Fig. 13. Confusion matrix of SIFT+SVM

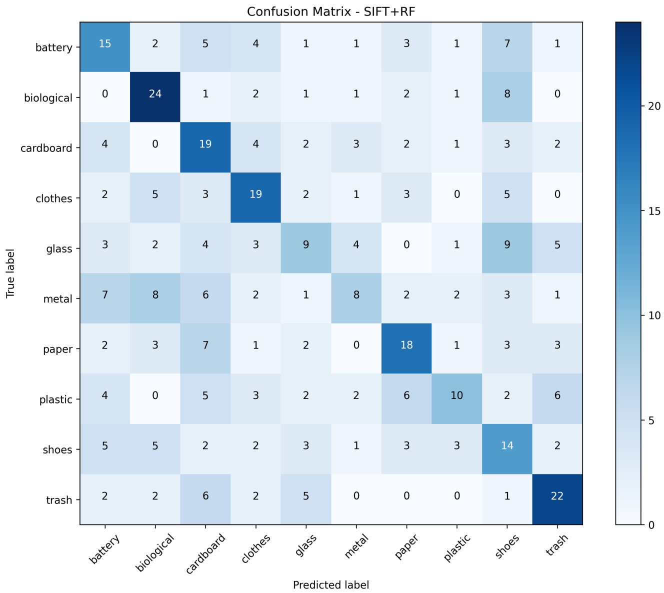



Fig. 14. Confusion matrix of SIFT+RF

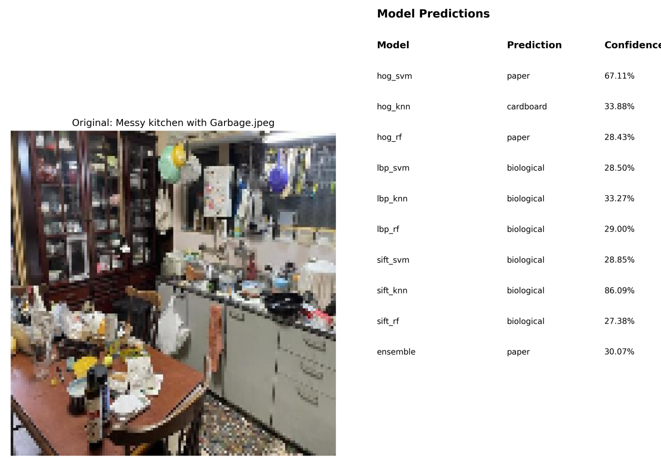



Fig. 16. Complex Scene model comparison

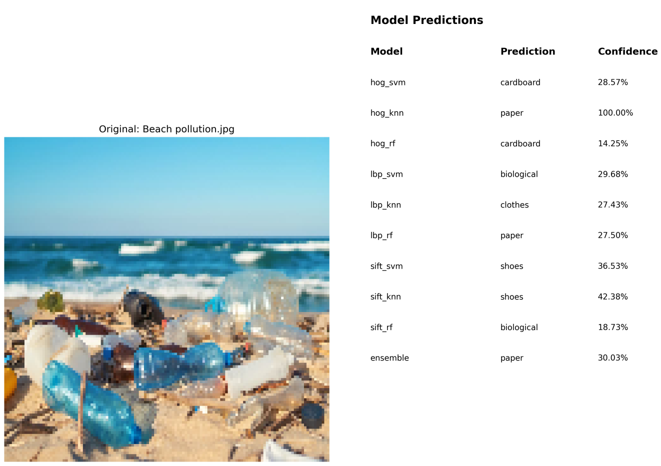



Fig. 15. Complex Scene model comparison

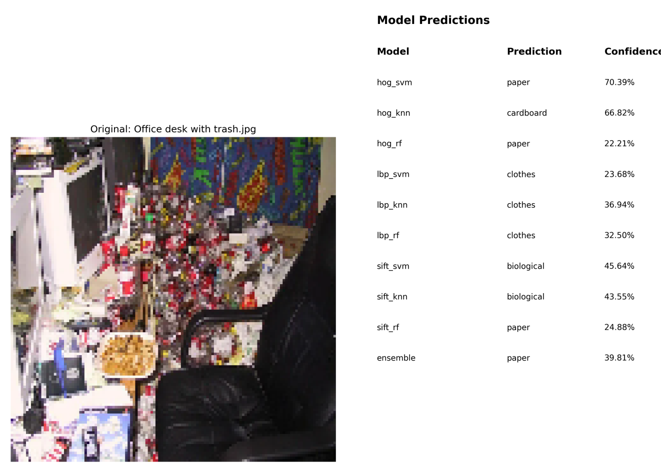
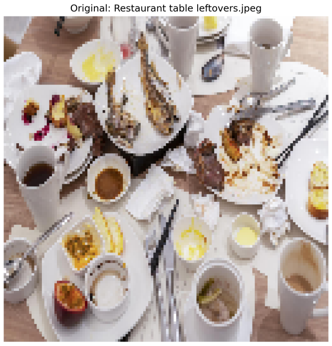



Fig. 17. Complex Scene model comparison

Original: Restaurant table leftovers.jpeg



**Model Predictions**

| Model | Prediction | Confidence |
|---|---|---|
| hog_svm | biological | 42.09% |
| hog_knn | cardboard | 67.39% |
| hog_rf | biological | 33.08% |
| lbp_svm | biological | 29.01% |
| lbp_knn | biological | 37.08% |
| lbp_rf | biological | 44.00% |
| sift_svm | biological | 18.19% |
| sift_knn | biological | 55.97% |
| sift_rf | paper | 26.04% |
| ensemble | biological | 46.01% |

Fig. 18. Complex Scene model comparison