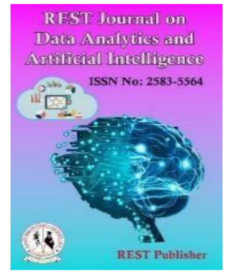
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/5



Material Classification for Efficient Recycling

P S Abhishek, O Prakyath Rao, B Leorah, Abdul Ahad
School of Engineering Anurag University, Hyderabad, India.
**Corresponding Author Email: psabhiishek@gmail.com*

Abstract. *Effective waste management is a critical challenge in today's world, requiring innovative technological solutions to optimize waste classification and recycling processes. This project introduces an advanced automated waste classification system that utilizes deep learning models to accurately categorize waste into seven distinct classes: cardboard, e-waste, glass, medical, metal, paper, and plastic. The system is designed to improve waste segregation, enhance recycling efficiency, and contribute to environmental sustainability. Two state-of-the-art deep learning architectures, ResNet-152 and Vision Transformer (ViT-B/16), form the backbone of the classification system. Extensive hyperparameter tuning is performed using Optuna to maximize the performance of these models. To further enhance accuracy, an ensemble learning strategy is implemented, leveraging the strengths of both models for superior classification results. Additionally, a confidence-based threshold mechanism is incorporated to flag uncertain predictions, ensuring robustness and reliability in real-world applications. The system is deployed through a user-friendly web interface built with Streamlit, enabling users to upload images of waste and receive real-time classification results. This interactive platform is designed for practical deployment in waste management facilities, offering scalability, efficiency, and ease of use. By integrating cutting-edge AI technologies, the proposed system significantly improves the accuracy and reliability of waste classification, reducing human effort and enhancing overall waste management practices. This project highlights the potential of deep learning in addressing global waste management challenges, promoting sustainable environmental practices, and paving the way for intelligent, automated waste segregation solutions.*
Keywords—waste management, deep learning, CNNs, ensemble learning, vision transformer.

1. INTRODUCTION

With the rapid increase in global waste generation, effective waste management has become a pressing concern. The improper disposal of waste contributes significantly to environmental degradation, leading to pollution, health hazards, and resource depletion. Proper segregation of waste materials—such as plastic, metal, paper, and medical waste—is essential for efficient recycling, reducing landfill accumulation, and minimizing environmental impact. However, traditional waste classification methods rely heavily on manual labor, which is time-consuming, costly, and prone to human error. To address these challenges, automated waste classification systems driven by artificial intelligence (AI) and deep learning have emerged as a promising solution.

In recent years, machine learning, particularly deep learning, has revolutionized image classification tasks, making it highly suitable for waste identification. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in processing and classifying image data, while transformer-based architectures such as the Vision Transformer (ViT) have further enhanced classification performance through self-attention mechanisms. This project leverages these advanced deep learning techniques to develop a robust and efficient waste classification system capable of accurately categorizing waste into seven distinct classes: cardboard, e-waste, glass, medical, metal, paper, and plastic.

To maximize performance, hyperparameter tuning is conducted using Optuna, an automated optimization framework that fine-tunes model parameters for improved accuracy. Additionally, an ensemble learning approach is employed, combining the strengths of multiple models—ResNet-152 and ViT-B/16—to achieve superior classification results. To handle real-world challenges, a confidence-based threshold mechanism is incorporated to detect low-confidence predictions and flag images that may not belong to any predefined category, ensuring robustness in practical applications.

The system is designed for real-world deployment through a user-friendly web interface built with Streamlit, allowing users to upload waste images and receive real-time classification results. This interactive platform facilitates practical implementation in waste management facilities, recycling plants, and smart cities, providing a scalable, efficient, and reliable solution for automated waste segregation. By integrating state-of-the-art AI techniques, this project aims to contribute to sustainable environmental practices by improving waste sorting accuracy, reducing manual labor, and optimizing recycling processes.

2. LITERATURE SURVEY

Notable works

This literature survey explores various approaches to automated waste classification, highlighting the datasets, methodologies, and performance metrics used in recent studies. The research spans diverse waste types and classification challenges, demonstrating the growing interest in leveraging computer vision and machine learning for efficient waste management.

Several studies have benchmarked general waste classification using the TrashNet dataset, a common resource for this task. For example, study [1] evaluated various CNN architectures on TrashNet and found that ResNet18 achieved the highest accuracy at 88.66%, demonstrating the effectiveness of residual networks. In another study using TrashNet [4], a comparative analysis between Support Vector Machines (SVM) and CNNs was conducted. The results showed that while SVM achieved 63% accuracy, CNNs underperformed significantly at 22%. This discrepancy underscores the importance of proper CNN architecture and training, suggesting potential implementation issues in the CNN approach of study [4].

Addressing a specific type of general waste, study [5] focused on detecting and classifying floating waste in water by introducing AquaVision. They created the AquaTrash dataset, a combination of TrashNet and TACO datasets, and using a RetinaNet model with a ResNet50-FPN backbone, achieved a mean Average Precision (mAP) of 0.8148, showcasing deep learning's potential for tackling water pollution.

Beyond general waste, research has also focused on specialized waste streams. Study [2] concentrated on office waste classification, creating a custom dataset from online and Raspberry Pi images, categorized into six common office items. Using Inception-V3, they achieved a high accuracy of 95.33%, indicating the efficacy of tailored systems for specific waste environments. For bulky waste, study [3] tackled the complexity of 95 classes, utilizing a dataset with 500 images per class. They fine-tuned VGG19 and experimented with hybrid models to address class imbalance, with the best hybrid model reaching 86.19% accuracy, demonstrating strategies for complex, imbalanced datasets. In contrast to multi-class problems, study [6] simplified the task to a binary classification of biodegradable versus non-biodegradable waste. Employing an Inception-V3-based model, they achieved 83.30% accuracy, showing CNN effectiveness even in focused binary problems. Furthermore, study [7] developed a system to classify waste into specific material categories like glass, paper, metal, and plastic. They employed a 50-layer residual network CNN for feature extraction combined with SVM for classification, achieving 87% accuracy, highlighting the use of deep learning for feature extraction followed by traditional classifiers.

Alternative methodologies and sensor technologies have also been explored. Study [8] at Luleå University of Technology moved away from image-based classification and developed a mechanical identifier for metal waste recycling, using chemical and mechanical methodologies to determine chemical composition. For material identification outside of waste streams, study [9] focused on fabric material identification from surface images using a dataset from Flickr. They utilized a Bayesian classifier, achieving a recognition rate of 44.6%, demonstrating alternative image-based methods and classifiers. Addressing real-world challenges, study [10-13] investigated the use of Near-Infrared (NIR) and RGB cameras for waste classification in realistic conveyor belt scenarios. They found NIR information largely sufficient, and introduced a sensor-fusion algorithm to enhance accuracy in complex cases, highlighting the importance of sensor selection and data fusion in real-world systems.

Existing Systems

Manual Sorting Systems: Traditional waste management facilities often rely on human labor for sorting different types of waste [14-18]. Workers manually separate recyclable materials from waste streams based on visual inspection. Although this method is widely used, it is time-consuming, prone to human error, and inefficient in handling large-scale waste sorting demands.

Conventional Image Recognition Systems: These systems use basic computer vision techniques, such as edge detection, color segmentation, and shape analysis, combined with classical machine learning models (e.g., decision trees, support vector machines) and early convolutional neural networks (CNNs) like AlexNet. While AlexNet and other early CNNs marked a significant advancement in image recognition tasks, these systems are often limited in their ability to handle complex and diverse waste streams. They generally require large, labeled datasets for training and struggle with varying environmental conditions, such as changes in lighting or image orientation. While they improve on traditional methods, they still fall short in terms of the scalability and adaptability needed for real-world waste management applications.

Infrared and X-Ray Sorting Systems: Some waste management facilities utilize infrared (near-infrared or NIR spectroscopy) or X-ray technologies to identify and sort materials based on their chemical composition. These systems are effective for distinguishing between different types of plastics or metals. However, they require expensive equipment and can be limited in scope, as they may not handle all types of materials or complex waste compositions.

Semi-Automated Systems: Semi-automated waste sorting systems combine mechanical and manual processes. Waste is transported on conveyor belts, where basic sensors and cameras assist in pre-sorting materials before final manual sorting. These systems aim to increase efficiency but still require significant human oversight, limiting scalability and accuracy. They may also be hindered by environmental factors, such as lighting conditions or contamination.

Limitation of existing system

Limited Waste Types: Existing systems often classify only a limited range of waste types, typically focusing on common materials such as plastic, paper, and metal. Specialized categories, such as e-waste or medical waste, are often neglected, resulting in improper sorting and inefficiencies in recycling processes

Broad and Unfocused Methods: Many current methods employ general approaches that lack a specialized focus on the unique characteristics of different waste types. This broad approach reduces accuracy and effectiveness, as it fails to account for the specific visual and structural differences between various materials.

Manual Feature Engineering: Traditional systems frequently rely on manual feature extraction, where human input is required to define the important characteristics for classification. This process is labor-intensive, time-consuming, and prone to errors, limiting the system's ability to automatically learn from data and generalize to new waste types.

Lower Accuracy: Current systems often struggle with accurately classifying diverse waste items, particularly in complex, real-world environments. Variations in lighting, angles, and background conditions can significantly affect performance, resulting in lower classification accuracy and reduced reliability in real-time waste sorting tasks.

Inability to Handle Contaminated or Mixed Waste: Many existing systems have difficulty accurately classifying waste when items are contaminated or composed of mixed materials, such as food-soiled packaging or composite products. This limitation leads to incorrect sorting decisions, which increases contamination in recyclable materials and reduces the overall effectiveness of the recycling process.

Gaps Identified

Limited Public Engagement and Awareness: Most systems do not provide user-friendly interfaces or tools that engage the public in proper waste disposal practices. This gap represents a missed opportunity to integrate waste classification technologies into everyday consumer behavior, promoting broader recycling efforts through mobile apps or smart bins.

Insufficient Data for Rare Waste Types: There is often a lack of sufficient labeled data for less common waste types, such as hazardous materials or certain types of e-waste. This gap makes it difficult to train robust models capable of accurately identifying these rare categories, leading to misclassifications.

Slow Adaptation to New Waste Materials: As new types of waste materials (e.g., biodegradable plastics or advanced composites) emerge, existing systems are slow to adapt. This gap highlights the need for models that can quickly incorporate new waste categories without requiring extensive retraining or manual updates.

Minimal Use of Advanced Optimization Techniques: There is a gap in the use of advanced optimization techniques, such as automated hyperparameter tuning or neural architecture search (NAS). These methods could help improve the performance of waste classification models but are often underutilized in current systems.

Over-Reliance on Manual Feature Engineering: Traditional systems often depend on manual feature extraction, which is time-consuming and prone to human bias. This limits the automation potential and scalability of the system, making it less efficient compared to modern deep learning methods that can automatically learn features from data.

Problem Statement

Effective waste classification is a critical challenge in modern waste management systems, particularly as the volume and diversity of waste materials continue to increase. Existing solutions are often limited in scope, focusing on a narrow range of waste types and struggling to classify specialized categories such as e-waste, medical waste, and mixed materials. Furthermore, traditional systems are highly reliant on manual sorting, feature extraction, and outdated machine learning techniques, which result in lower classification accuracy and inefficiency, especially in real-world conditions where waste is contaminated or presented in complex forms.

Current waste classification methods also face limitations in adaptability, often performing poorly under varying environmental conditions such as changes in lighting, orientation, and background, which hinders their practical implementation in diverse waste management settings. Additionally, many systems fail to integrate advanced automation, real-time processing, or scalable architectures, limiting their effectiveness in large-scale recycling efforts. The lack of accurate and automated solutions in waste classification contributes to increased contamination rates, inefficient recycling processes, and missed opportunities for improved resource recovery and sustainability.

This project aims to address these gaps by developing an advanced, automated waste classification system that leverages cutting-edge deep learning techniques, specifically convolutional neural networks (CNNs) and transformer-based models, to classify a wide variety of waste types more accurately and efficiently. By integrating an ensemble approach, robust preprocessing, and user-friendly deployment, the system will contribute to more sustainable waste management practices.

Objectives

Develop an Automated Waste Classification System: Build a system that uses advanced deep learning techniques, including convolutional neural networks (CNNs) and transformer-based models, to classify waste into predefined categories, such as cardboard, e-waste, glass, medical waste, metal, paper, and plastic.

Enhance Classification Accuracy: Improve the accuracy of waste classification by implementing an ensemble learning approach that combines the strengths of both ResNet-152 and Vision Transformer (ViT-B/16), optimizing performance across diverse waste types and environmental conditions.

Automate Feature Extraction: Replace manual feature extraction methods with automated processes that allow the models to learn directly from the data, reducing human intervention and improving scalability.

Handle Diverse Waste Types and Real-World Conditions: Ensure that the system can accurately classify waste items under varying environmental conditions (e.g., different lighting, orientations, and contamination levels) and handle a wide range of waste types, including complex and mixed materials.

Deploy a User-Friendly Interface: Implement the system using a user-friendly platform like Streamlit, allowing users to easily upload waste images for classification and receive real-time predictions with high confidence scores.

Optimize Model Hyperparameters: Utilize Optuna or other automated hyperparameter tuning techniques to optimize model parameters, ensuring the best possible performance of the system.

Incorporate Real-Time Uncertainty Handling: Implement a mechanism to flag uncertain classifications, providing the option to either request further input or process the image as uncertain, to maintain system reliability.

3. PROPOSED SYSTEM

Architecture And Methods

In this project, we designed and implemented a waste classification system using machine learning models, specifically focusing on convolutional neural networks (CNNs) and transformer-based models. The primary architecture consists of an ensemble approach combining ResNet-152 and Vision Transformer (ViT-B/16), which leverages both deep feature extraction and attention mechanisms for classification tasks. The system aims to classify images of waste into seven predefined categories, including Cardboard, E-Waste, Glass, Medical, Metal, Paper, and Plastic.

Convolutional Neural Networks (CNNs): ResNet-152 was chosen for its proven success in image classification tasks, especially on large datasets. ResNet employs residual connections that help to mitigate the vanishing gradient problem in deep networks. The architecture consists of repeated residual blocks that allow the model to learn deeper representations without degradation in performance.

Pretrained Weights: We used pretrained ResNet-152 weights from ImageNet to take advantage of the model's learned feature extraction capabilities.

Modified Fully Connected Layer: The final fully connected (FC) layer was modified to accommodate the 7-class output. Additionally, dropout was incorporated to reduce overfitting.

Optimizer: The AdamW optimizer was chosen for ResNet to provide adaptive learning rates and weight decay for improved generalization.

Learning Rate Scheduler: This was used to adjust the learning rate dynamically, helping the model converge more effectively during training.

Transformer-Based Models: Vision Transformer (ViT-B/16) was integrated to capture long-range dependencies in images, a significant benefit for visual tasks involving complex spatial relationships. ViT divides images into patches and applies a transformer mechanism to learn relationships across these patches, unlike CNNs which rely solely on convolutional filters.

Patch Embedding: The image was split into non-overlapping patches (16x16) and then embedded into a sequence of tokens.

Self-Attention Mechanism: The transformer architecture applies self-attention to capture contextual relationships between image patches.

Modified Classification Head: Similar to ResNet, we modified the classification head for 7 output classes and incorporated dropout to enhance generalization.

Optimizer: We used the AdamW optimizer for ViT as well, with specific hyperparameters chosen after extensive experimentation.

Learning Rate Scheduling: The learning rate was controlled using the ReduceLROnPlateau scheduler to adapt during training based on validation loss.

Ensemble Learning: To maximize classification accuracy, an ensemble learning approach was adopted by combining the predictions of ResNet and ViT models. Ensemble learning enhances predictive performance by leveraging the strengths of different models.

Soft Voting: Initially, soft voting was applied, where the probability distributions from both models were averaged, and the class with the highest average probability was selected as the final prediction.

Hard Voting: We also experimented with hard voting, where the class predicted by the majority of the models was chosen as the final output.

Weighted Voting: The most effective approach was found to be weighted voting, where ResNet and ViT predictions were weighted differently based on their individual accuracies. For instance, assigning a higher weight to ResNet (due to its better standalone performance) and a lower weight to ViT led to the highest test accuracy in the ensemble.

Data Preprocessing and Augmentation

A critical component of the system was robust data preprocessing, ensuring that the models could generalize well across different input variations:

Resizing: All images were resized to 224x224 to be compatible with both ResNet and ViT input requirements.

Data Augmentation: To improve model robustness, the following augmentations were applied during training: Random Horizontal Flips, Random Rotation, Random Perspective Transformations, Normalization

Optuna Hyperparameter Tuning: Optuna, an automated hyperparameter optimization framework, was employed to fine-tune the models. Both ResNet and ViT models underwent multiple trials to identify the best hyperparameters for learning rate, weight decay, batch size

Streamlit Deployment: The entire system was deployed using Streamlit, a user-friendly framework for creating interactive web applications. Users can upload waste images for classification, and the ensemble model predicts the waste category along with the probability/confidence score.

Uncertainty Handling: A threshold mechanism was introduced to handle uncertain classifications. If the highest predicted probability was below a certain threshold, the system classified the image as "uncertain."

Image Upload: Users could upload single or multiple images, and the system processes each image, displays the predictions, and tracks the count of classifications.

Requirements And Specifications

Client Requirements: Automated Waste Classification: The system should be able to automatically classify waste into distinct categories like plastic, glass, metal, cardboard, paper, medical waste, and e-waste.

High Accuracy: The classification model should achieve high accuracy (at least 93%) to ensure reliable sorting of waste.

Image Upload Capability: Users should be able to upload images of waste items for classification through a simple interface.

Support for Multiple Waste Types: The system must handle various types of waste and ensure correct classification across all the seven predefined waste classes.

Uncertainty Handling: In cases where the system is unsure about the classification (based on a probability threshold), it should label the waste type as "Uncertain" rather than forcing a prediction.

Software Requirements

The following software components are necessary

Programming Language: Python 3.x for implementing the machine learning models and the backend.

Machine Learning Framework: PyTorch for model training, inference, and deployment. The framework should support both CPU and GPU for flexibility.

Streamlit: For deploying the web-based interface that allows users to upload images and see the classification results.

Libraries/Packages

Torch and torchvision: For pre-trained models like ResNet and Vision Transformer (ViT) and their associated weights.

Optuna: For hyperparameter optimization during the model training phase.

Operating System: Windows 10/11, Linux, or macOS, compatible with Python 3.x and deep learning frameworks.

Integrated Development Environment (IDE): Any Python-compatible IDE such as VS Code, PyCharm, or Jupyter Notebook to develop and debug the code.

Hardware Requirements

For training and deploying the waste classification model, the following hardware is required:

CPU: A multi-core processor (Intel i7 or equivalent) for running the development environment and the Streamlit app.

GPU: A high-performance GPU with at least 8-12 GB of VRAM (e.g., NVIDIA RTX 2080, 3080, or Tesla V100). The project involved training models like ResNet152 and Vision Transformer, which require a powerful GPU to reduce training time. For inference purposes lightweight GPU’s are sufficient

RAM: At least 16 GB of RAM for smooth operation during model training and inference tasks.

Storage: A minimum of 50 GB of disk space to store datasets, trained models, and other project files.

Cloud Infrastructure (Optional): Cloud platforms offer an easier alternative for the user to train perform their training and inferencing tasks. If the user lacks insufficient resources, they can use cloud platforms. This project made use of Google Colab and Vast.ai

4. DESIGN

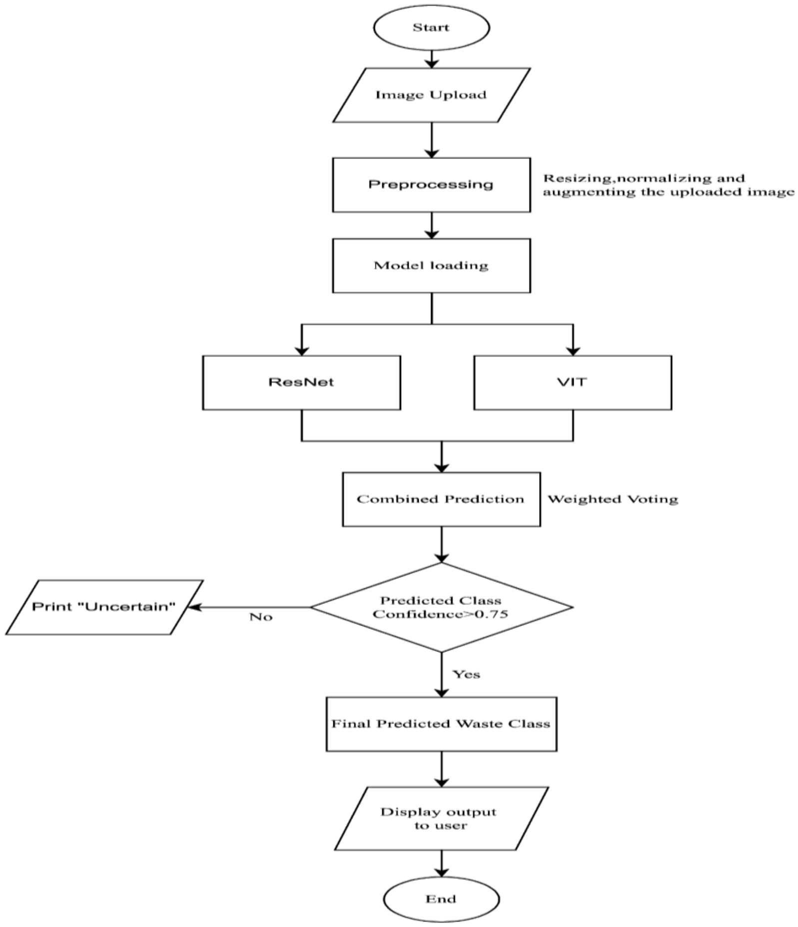


FIGURE 1. Application Flowchart

Module Design and Organization.

The waste classification system is organized into several key modules, each responsible for a specific part of the process. Below is an overview of the module design and organization, detailing the functionality of each module and how they interact to create the complete system. This module is responsible for preparing the input data for the model. The images uploaded by the user undergo several transformations to ensure they are in the correct format for the model's inference process.

Functions:

Resize images to a fixed size (224x224) for compatibility with pre-trained models like ResNet and (ViT). Normalize the images using mean and standard deviation values based on ImageNet pre-trained models. Convert images to tensors, ensuring they are in the correct format for the model's input layer. Handle any special cases such as palette images or images with transparency.

Model Loading and Initialization Module

This module is responsible for loading and initializing the machine learning models (ResNet and Vision Transformer) with pre-trained weights and necessary modifications.

Functions:

Load pre-trained ResNet and ViT models from PyTorch's model zoo. Modify the final classification layers (fully connected layers) to accommodate the 7 waste classes in the dataset. Apply dropout layers to prevent overfitting during training and inference. Load the saved model weights that were fine-tuned during training for both ResNet and ViT.

Ensemble-Prediction-Module

This module handles the ensemble learning approach by combining predictions from both ResNet and VisionTransformer

Functions:

Perform inference using both the ResNet and ViT models to generate probability distributions for the 7 waste-classes. Implement weighted voting to combine the predictions models. Allow for uncertainty handling, where the predicted class is labeled as "Uncertain" if the combined probabilities do not meet a specified threshold.

Streamlit

This module handles the user interface and interaction with the system, allowing users to upload images get results

Functions:

Provide a simple interface for image uploads using Streamlit's file uploader widget. Display the uploaded image for verification before classification. Output the predicted waste class and the corresponding probability/confidence. Track the number of images classified and the distribution of predictions across different waste classes, logging this data into a CSV file for further analysis. Handle multiple image uploads at once to classify multiple waste images in one go.

Classification-Statistics

This module tracks and logs the classification data

Functions:

Maintain a CSV file that logs the number of images classified, and the count of images predicted for each class. Update the CSV file with each new prediction to ensure a real-time log of classification statistics.

Hyperparameter Optimization Module

During the model training process, hyperparameter optimization was performed using Optuna to find the best parameters for each model. This module is responsible for conducting those optimization trials.

Functions:

Define the hyperparameter search space for parameters such as learning rate, weight decay, dropout rate, and batch size. Run trials to evaluate different combinations of hyperparameters and select the best-performing set. Use the results of the hyperparameter tuning to train the final models that are used in the ensemble.

Model Training Module

This module was used to fine-tune the pre-trained models on the waste classification dataset.

Functions:

Load the dataset, apply transformations, and split it into training and validation sets. Train the models (ResNet and ViT) with the selected hyperparameters. Implement early stopping and learning rate scheduling to optimize training. Save the model weights that yielded the best validation performance. Each of these modules works in tandem to build the final waste classification system. This modular design allows for flexibility and scalability, enabling future enhancements, such as integrating new models or adding additional waste classes.

5. IMPLEMENTATION & TESTING

Technology Used

Programming Language: Python: Python is a high-level, interpreted programming language known for its readability and versatility, widely used in web development, data analysis, and automation.

Deep Learning Framework: PyTorch: PyTorch is an open-source machine learning library for deep learning, known for its dynamic computation graphs, tensor operations, and automatic differentiation. It's widely used for research and production due to its flexibility and ease of use.

- ✓ Models: Pretrained Architectures from PyTorch
- ✓ Final implementation makes use of
- ✓ ResNet-152
- ✓ Vision Transformer (ViT-B/16)
- ✓ Optuna for Hyperparameter Optimization
- ✓ Deployment Framework:

Streamlit: Streamlit is an open-source framework for building interactive web applications quickly and easily, primarily for data science and machine learning projects.

- ✓ Matplotlib for Visualization

6. PROCEDURE

Data Collection and Preprocessing: Data Collection: Two datasets were identified, Garbage Classification dataset from Kaggle, this dataset has 6 classes - cardboard, glass, metal, paper, plastic and trash. From these the “trash” class was dropped as it consisted of generalized objects that count as everyday trash and might conflict with other classes as well. Examples of such images includes plastic wrappers. The data distribution was not even and the number of images in each class varied from a minimum of 127 images (in “trash” class) to 584 images (in “paper” class). In addition to this dataset we have also taken the TrashBox dataset from GitHub. This dataset is much vaster and contains a total of 17785 images divided into seven classes - glass, plastic, metal, e-waste, cardboard, paper, medical waste. This was combined with the Garbage classification dataset thereby adding two additional classes (e-waste and medical waste). Once combined the images were randomly dropped ensuring that the final dataset has 1500 images per class. Image Preprocessing: Each image was resized to 224x224 to maintain consistency across all models, including ResNet, Vision Transformer (ViT), DenseNet, and VGG16. To enhance the model's generalization, random transformations like horizontal flipping, rotation, and perspective distortion were applied to the training dataset. This reduced the risk of overfitting by simulating a diverse set of image variations. Normalization: All images were normalized using the mean and standard deviation values of ImageNet's dataset (mean: [0.485, 0.456, 0.406], std: [0.229, 0.224, 0.225]) to ensure compatibility with pre-trained models. Splitting the Dataset: The dataset was split into training, and validation sets in an 80:20 ratio. The training set underwent augmentation, while the validation and test sets only underwent resizing and normalization. The TrashBox dataset also consisted of a separate test split that was utilized as the test dataset in this project

Model Selection and Training: ResNet-152: Initially, ResNet-152 model was chosen due to its deep architecture and strong performance in image classification tasks. Due to the presence of skip connections that bypass one or more layers it allows for effective learning without a significant increase in computational complexity. The fully connected layer was modified to match the seven-class classification task. Hyperparameters like learning rate, weight decay, dropout, and batch size were optimized using Optuna. AdamW optimizer was chosen for its ability to handle sparse gradients and large datasets efficiently. Vision Transformer (ViT-B/16): The Vision Transformer (ViT) Base model, with a 16x16 patch size, was also implemented. The model’s heads were modified to suit the classification task, and after initial trials, Optuna was used to tune the hyperparameters. Unlike ResNet, ViT processes images differently by converting them into patches. ViT accepts the same image preprocessing as that of ResNet, thereby allowing us to maintain consistency. VGG16: For VGG16, a similar preprocessing pipeline was used, and Optuna was employed to determine the best hyperparameters for the model. Although VGG16 is less computationally intensive than ResNet, it was still capable of achieving strong classification performance on the waste dataset after fine-tuning. DenseNet-121: DenseNet was included for its efficient feature propagation through dense connections between layers. The same preprocessing and training strategies were employed as with the other models.

Evaluation and Hyperparameter:

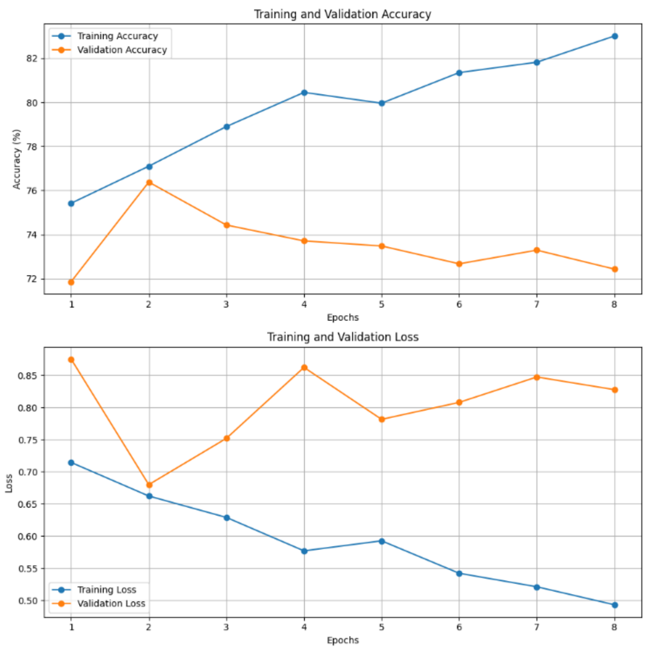


FIGURE 2. ResNet Training without hyperparameter tuning

Optuna Trials: Optuna was extensively used to tune hyperparameters across all models. Pruning strategy was also set so that unpromising trials are terminated. Each trial consisted of 5 epochs and 30 trails were performed for each model (ResNet, ViT, DenseNet, and VGG16), Optuna explored different optimizers (Adam, AdamW, RAdam), learning rates, weight decays, dropout rates (0.2- 0.5), and batch sizes (32, 64). The result of the trials was the most promising hyperparameters for that model. A common trend that was observed is that AdamW is preferred over Adam thanks its decoupled approach to weight decay. Early Stopping was implemented to halt training when the validation loss stopped improving, preventing overfitting. Models were evaluated based on accuracy, loss on the test set also custom inputs were given to see if the model has good generalization. Weighted voting was then introduced, where ResNet and ViT were assigned different weights (e.g, 0.5-0.5, 0.6-0.4) based on their performance, achieving better accuracy (up to 94.87%).

Deployment:

Streamlit Integration: The ensemble model was deployed using Streamlit, allowing users to upload images for classification. The system processed the uploaded image, passed it through both ResNet and ViT models, and then used the ensemble voting method to predict the class of the waste. The probabilities were also displayed to give users confidence in the classification result. Handling Uncertainty: To handle cases where an image might not belong to one of the seven predefined classes, a probability threshold was introduced. If the highest probability returned by the model fell below the threshold (e.g. 0.7), the result was flagged as uncertain.

Testing & Validation

Design Test Cases and Scenarios: The system was tested under various conditions to ensure reliability. The following test cases were designed to validate the model's behaviour under different scenarios, including image classification accuracy, handling of uncertain predictions, and performance of the ensemble approach.

Some of the test cases included –

- ✓ Upload an image belonging to one of the seven waste classes (e.g., Cardboard, Metal)
- ✓ Upload an image not belonging to any of the seven waste classes
- ✓ Upload multiple images in a batch
- ✓ Upload low-resolution or noisy images
- ✓ Implement confidence threshold for uncertain predictions

Validation

The validation process was carried out to ensure that the model's performance generalizes well on unseen data and handles different real-world scenarios.

Validation Process: Dataset Split: The dataset was split into training (80%) and validation (20%) sets. After model tuning, a separate test dataset was used for final evaluation to ensure that the model showed results that are consistent with that of the validation set.

We evaluated the model using the following metrics - Accuracy: The primary metric used to measure the percentage of correctly classified images.

Loss: Cross-entropy loss was used to track model performance during training and validation, with the goal of minimizing it.

Confidence Scores: Probabilities for each class were tracked, checking how they were consistent across various examples.

Ensemble Model Validation: The ensemble approach, combining ResNet and ViT predictions, was validated using:

Soft Voting: Combining predicted probabilities and selecting the class with the highest average probability.

Hard Voting: Selecting the class that receives the most votes from individual models.

Weighted Voting: Assigning different weights to the models' predictions, with ResNet weighted higher due to better individual performance.

Uncertain Predictions: A threshold-based approach was introduced to manage predictions with low confidence. If the confidence score for the predicted class fell below the defined threshold, the prediction was marked as "uncertain," improving the system's reliability in ambiguous cases.

7. RESULTS

Output

Throughout the project, multiple machine learning models were trained, evaluated, and tested on the waste classification task. The final output of the project includes the successful deployment of a machine learning model using an ensemble of ResNet and Vision Transformer (ViT) architectures. Below are the key results from the individual models and the ensemble:

ResNet-152: Achieved a test accuracy of 93.75% on the waste classification task after extensive training and hyperparameter optimization. The model excelled at identifying key features in the dataset, contributing substantially to the ensemble's success.

ViT-B/16 (Vision Transformer): The Vision Transformer (ViT) model achieved a test accuracy of 91.52%. Despite its complexity and need for more computational power, it added a different approach by focusing on the transformer architecture, which models global dependencies in the image data.

VGG16: The VGG16 model, known for its simplicity and effectiveness, achieved a validation accuracy of 88.19% using the optimal hyperparameters from the Optuna trials. Although it performed slightly lower than ResNet and ViT, it was still a strong performer.

DenseNet-121: DenseNet, known for its feature reuse through dense connections, achieved a test accuracy of 89.86% after fine-tuning with optimal hyperparameters. As it was outperformed by ResNet and ViT, it was not considered for the ensemble model

Ensemble Model (ResNet + ViT):

Hard Voting: The ensemble using hard voting achieved an accuracy of 92.75%. This method combined the class predictions from both models and chose the most frequent class. Soft Voting on the other hand achived a score of 91.3%

Weighted Voting: The ensemble model using weighted voting (with ResNet contributing 0.5 and ViT contributing 0.5) produced the best result, with a test accuracy of 94.87%.

Result Analysis

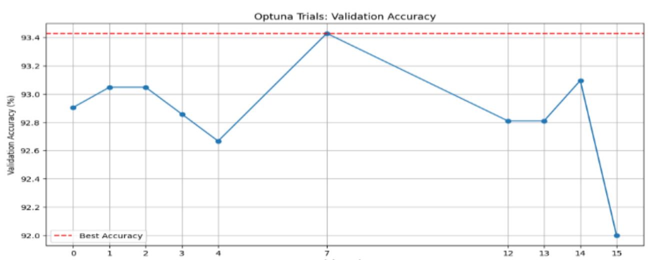


FIGURE 3. Optuna Trial for ResNet

The results obtained from training and evaluating the different models show the varying strengths of the deep learning architectures used. A comparative performance analysis was conducted for all models, including ResNet-152, Vision Transformer (ViT), VGG16, and DenseNet-121. The following key points summarize the result analysis:

ResNet-152 consistently demonstrated superior performance in classification tasks, achieving a high accuracy of 93.75%. This can be attributed to its deep architecture and residual connections, which help it learn robust features even in deeper layers.

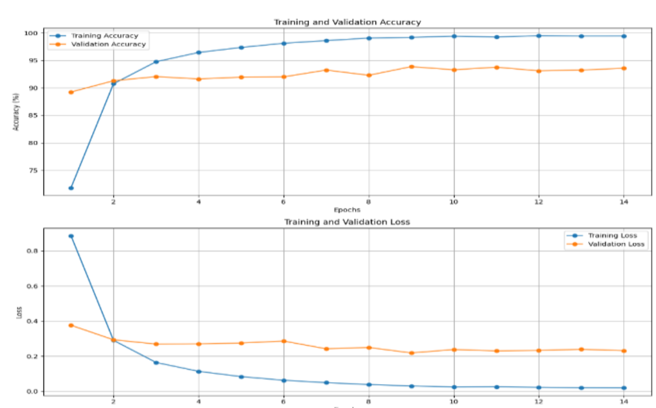


FIGURE 4. Final Training of ResNet

ViT-B/16 provided strong performance despite having lower accuracy than ResNet. Its transformer architecture focuses on long-range dependencies in images and proved valuable in an ensemble setup, where its unique strengths could complement those of ResNet.

VGG16 showed solid results with 88.14% validation accuracy, demonstrating the value of simpler architectures for image classification. However, it underperformed compared to ResNet and ViT due to the absence of more complex feature extraction mechanisms like residual connections or transformers.

DenseNet-121 delivered competitive results, achieving 86.19 % test accuracy. DenseNet's dense connectivity allows for feature reuse, which improved its generalization capabilities. However, due to computational constraints, it was not included in the ensemble approach.

Ensemble Approach The ensemble approach significantly improved the model's performance. Using weighted voting, the combination of ResNet and ViT yielded a test accuracy of 94.87%. By assigning equal weights to both models, we harnessed the complementary strengths of each architecture. The ensemble proved that leveraging multiple models can mitigate the weaknesses of individual architectures. ResNet's strong performance on high-level features and ViT's ability to capture global dependencies resulted in a more robust final model.

Handling Uncertainty: An uncertainty threshold was introduced to handle edge cases where images did not belong to one of the seven defined classes. If the predicted confidence for any class fell below a predefined threshold, the system flagged the prediction as "uncertain." This feature added reliability and robustness to the classifier in real-world use cases.

The ensemble model outperformed the individual models by combining their strengths, demonstrating that the hybrid approach provides better generalization on the waste classification task.

TABLE 1. Comparison of Results

Model	**Test Accuracy (%)**	**Validation Accuracy (%)**	**Test Loss**
ResNet-152	93.75	93.43	0.2225
Vision Transformer (ViT-B/16)	91.52	92.0	0.2688
VGG16	88.14	86.19	0.34
DenseNet-121	86.19	85.4	0.345
Ensemble (ResNet + ViT)	94.87	94.53	0.201

8. CONCLUSION

This project highlights the successful development of a waste classification system using state-of-the-art deep learning models, aimed at automating waste sorting and improving waste management efficiency. The system explored several architectures, including ResNet-152 and Vision Transformer (ViT), as well as an ensemble approach combining both models. ResNet-152 demonstrated the strongest individual performance with a test accuracy of 93.75%, leveraging its deep convolutional structure to effectively classify waste images. The Vision Transformer, while slightly less accurate at 91.52%, still showcased the potential of transformer-based models in handling image data. The most significant improvement came from the ensemble approach, which used weighted voting to combine the predictions of both models, achieving a final test accuracy of 94.87%. This result underscores the power of ensemble learning in leveraging the complementary strengths of diverse architectures. Key components of the system, such as hyperparameter tuning using Optuna, dropout layers for preventing overfitting, and uncertainty handling mechanisms, contributed to its robustness and real-world applicability. In conclusion, the proposed system not only achieved high accuracy but also demonstrated the effectiveness of combining advanced deep learning techniques with ensemble methods and uncertainty management. This makes it a viable solution for real-world deployment, where it can automate waste sorting processes and contribute to more sustainable waste management practices.

9. FUTURE WORK

While the current project has demonstrated a high level of success in classifying waste into seven categories, there are several areas for potential improvement and expansion. Future work can focus on the following key aspects:

A. Incorporating Additional Waste Categories: The current model is limited to seven waste categories. Future work can involve extending the classification to more diverse waste types, particularly those commonly encountered in urban waste management, such as textiles, wood, and hazardous materials.
B. Object Detection for Multiple Waste Items: In real-world waste management systems, images may contain more than one waste item. Implementing an object detection model, such as YOLOv10 or Faster R-CNN, would allow for multi-object detection and classification within a single image, providing more granular and actionable data.
C. Exploration of Advanced Ensemble Methods: While weighted and hard voting ensemble methods were used, future work could explore more sophisticated ensemble techniques such as stacking or blending, where the outputs of individual models (ResNet, ViT, etc.) are fed into a meta-model for final prediction. This may further improve classification accuracy.
D. Real-time Waste Classification Deployment: The current deployment is limited to batch image classification via a web interface. A potential extension would be to integrate real-time waste classification using video streams from cameras placed in recycling centres or waste collection sites. This would require optimization to improve inference speed without sacrificing accuracy.
E. Transfer Learning and Fine-Tuning with Domain-Specific Data: While pre-trained models were fine-tuned for waste classification, more specific pre-training on a large-scale waste management dataset could lead to even better results. Transfer learning from models trained on closely related domains (e.g., industrial objects) could also boost performance.
F. User Feedback Integration: Future iterations of the system could integrate a feedback mechanism where users can correct misclassifications. This feedback could be used to improve the model by dynamically updating and retraining on new labeled data.

By pursuing these avenues of future work, the waste classification system could become even more robust, versatile, and scalable, ultimately contributing to more efficient waste management processes and environmental sustainability.

10. REFERENCES

[1]. V. Ruiz, A. S ´anchez, J. F. V ´ elez, and B. Raducanu, ´ "Automatic image-based waste classification," in International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer, 2019, pp. 422–431.
[2]. J.-w. Feng and X.-y. Tang, "Office garbage intelligent classification based on inception-v3 transfer learning model," in Journal of Physics: Conference Series, vol. 1487. IOP Publishing, 2020, p. 012008
[3]. Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01
[4]. H. Wang, Y. Li, L. M. Dang, J. Ko, D. Han, and H. Moon, "Smartphone-based bulky waste classification using convolutional neural networks," Multimedia Tools and Applications, vol. 79, no. 39, pp. 29 411–29 431, 2020.
[5]. M. Yang and G. Thung, "Classification of trash for recyclability status," CS229 Project Report, vol. 2016, 2016
[6]. Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., "Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection", Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409
[7]. H. Panwar, P. Gupta, M. K. Siddiqui, R. MoralesMenendez, P. Bhardwaj, S. Sharma, and I. H. Sarker, "Aquavision: Automating the detection of waste in water bodies using deep transfer learning," Case Studies in Chemical and Environmental Engineering, p. 100026, 2020.

[8]. F. Shaikh, N. Kazi, F. Khan, and Z. Thakur, “Waste profiling and analysis using machine learning,” in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2020, pp. 488–492
[9]. O. Adedeji and Z. Wang, “Intelligent waste classification system using deep learning convolutional neural network,” Procedia Manufacturing, vol. 35, pp. 607–612, 2019
[10]. S. Zhang and E. Forssberg, “Intelligent liberation and classification of electronic scrap,” Powder technology, vol. 105, no. 1-3, pp. 295–301, 1999.
[11]. C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, “Exploring features in a bayesian framework for material recognition,” in 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010, pp. 239–246
[12]. E. Strollo, G. Sansonetti, M. C. Mayer, C. Limongelli, and A. Micarelli, “An AI-Based approach to automatic waste sorting,” in International Conference on HumanComputer Interaction. Springer, 2020, pp. 662–669.
[13]. S. M, S. M, S. Doss and P. K. P, "Deep Learning based Waste Material Classification," 2023 IEEE Silchar Subsection Conference (SILCON), Silchar, India, 2023, pp. 1-6
[14]. Manoranjan Dash, “Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment,” International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.
[15]. R. Faria, F. Ahmed, A. Das and A. Dey, "Classification of Organic and Solid Waste Using Deep Convolutional Neural Networks," 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India, 2021, pp. 01-06
[16]. Nikhil Kumsetty, Amith Nekkare, November 15, 2022, "TrashBox", IEEE Dataport
[17]. CCHANG, "Garbage Classification,"Kaggle, 2018.
[18]. P. S. Silpa et al., "Designing of Augmented Breast Cancer Data using Enhanced Firefly Algorithm," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 759-767, doi: 10.1109/ICOSEC54921.2022.9951883.