ISTP

RESEARCH ARTICLE

Optimizing Waste Management with Squeeze-and-Excitation and Convolutional Block Attention Integration in ResNet-Based Deep Learning Frameworks

Saiprasad Potharaju,[1] **Swapnali N. Tambe,**[2] **Satya Kiranmai Tadepalli,**[3] **Shobarani Salvadi,**[3] **T.C. Manjunath,**[4] **and A. Srilakshmi**[5]

[1]Department of CSE, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India
[2]Department of Information Technology, K. K.Wagh Institute of Engineering Education & Research, Nashik, MH, India
[3]Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
[4]Department of Computer Science & Engineering, RajaRajeswari College of Engineering, Bengaluru, Karnataka, India
[5]Department of IT, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

Abstract: Effective waste classification is crucial for sustainable waste management, yet existing automated models face challenges such as misclassification of visually similar waste materials, dataset imbalance, and poor generalization to real-world conditions. This study addresses these limitations by integrating squeeze-and-excitation (SE) and convolutional block attention module (CBAM) mechanisms into ResNet-50, a deep learning model, to improve waste classification accuracy. Utilizing the TrashNet dataset (six waste categories), we have employed the synthetic minority over-sampling technique to balance the dataset, increasing samples from 2,527 to 3,564. Experimental results demonstrate significant improvements: ResNet-50 achieves 74.42% test accuracy, SE+ResNet-50 improves accuracy to 93.47%, and CBAM+ResNet-50 reaches 95.74%. The findings highlight that attention-based deep learning models can significantly enhance feature extraction, improve classification accuracy, and optimize waste segregation processes, contributing to more efficient recycling and waste management automation. Future work includes extending the model to classify hazardous and mixed waste, ensuring broader applicability in real-world waste management systems.

Keywords: attention mechanism; automated waste classification; convolutional block attention module (CBAM); deep learning; squeeze-and-excitation (SE)

I. INTRODUCTION

Global waste production was greatly increased by the exponential growth in urbanization and industrialization, which poses incredible problems in waste management. Environmental degradation and rising public health risks, as well as cost and resource wastage [1], have resulted from the mismanagement of waste, and specifically municipal solid waste. Systematic collection, sorting, and recycling of waste depend on the effective waste classification, which is important along the line of these processes.

A broad waste typology is necessary to initiate recycling processes and lower the dependency on landfill. But the manual sorting or simple mechanical processes dominating current waste segregation methods are commonly labor intensive, inconsistent, and prone to mistake [2]. Thus, the automated waste classification with advanced technologies, such as deep learning, has become popular. With high accuracy, scalability, and efficiency, these methods promise for sorting waste, especially for recycling materials of plastic, metal, glass, and organic waste [3].

To an extent, waste classification and materials were important to the proper management of waste and continued lean

manufacturing techniques based on sustainable environmental practices. It allows for accurate treatment and segregation of waste into recyclable, organic, hazardous, and non recyclable to ensure effective recycling and also less amount of waste sent to landfill. Waste mismanaged is a major contributor of environmental pollution, resource depletion, and greenhouse gas emissions. Classification can properly determine contents leading to improved material recovery, greater resource efficiency, and corresponding circular economy objectives [4]. Moreover, it reduces the contamination of recyclable materials thereby improving the quality and recycling of such materials. Accurate waste classification is also necessary to minimize the health hazards from improper disposal of dangerous waste, like chemicals and e waste. Taking all of this together, the classification of waste is the first step in a sustainable waste management process impacting environment conservation, public health, and resource sustainability.

Manual methods of waste classification are given advantages to automate waste classification. Using technological knowhow such as machine learning (ML) and image recognition, it boosts accuracy and efficiency, taking out errors and inconstancies of humans. Sorting is accelerated by automation, making it possible to deal with large amounts of waste, especially in urban environments, where we are dealing with high waste formation. It also cuts back on the manual labor that is employed, lowering operational

Corresponding author: Saiprasad Potharaju (e-mail: psaiprasadcse@gmail.com)

costs and lessening the risk to workers exposed to hazardous waste. Real-time data analysis can also be automated and integrated into automated systems to boost decision-making and increase recycling workflow efficiency. Automation guarantees precise separation of waste streams and therefore helps achieve higher recycling rates, reducing reliance on landfills, and diminishes environmental contamination [5]. In the end, automation is a scalable and cost-effective answer to the needs of up to date waste administration techniques, fostering sustainability and resource effectiveness.

With numerous advancements in ML and artificial intelligence (AI), existing automated waste classification models continually struggle with persistent limitations preventing their effectiveness and achieving real world applicability [6,7]. However, there is very often insufficient accuracy when working with complex waste classes. Basic convolutional neural networks (CNNs) and pre-trained models can be conventional models that often fail to capture the subtle and nuanced differences of different material wastes, such as being able to distinguish visually similar recyclable and non recyclable items [8]. As a result, materials typically occur widely misclassified, especially in the cases when heterogeneous materials are often co-existing.

Data scarcity, as well as the lack of large, balanced datasets [9], is another big barrier. For instance, most of the datasets available publicly are small, often missing representation of particular categories of waste, including hazardous materials or "trash" in general. In particular, this imbalance further exaggerates model bias, as lack of training leads to poor performance of underrepresented classes that degrades the system's reliability in practical applications. Additionally, the small dataset size prevents the training of complex models since such models need lots of data to avoid overfitting [10].

The challenge of overfitting and generalization further complicates the issue. Deep learning models trained on limited datasets tend to memorize training data rather than generalize to unseen scenarios, leading to reduced classification performance in real-world environments [11]. This problem is particularly critical for waste management systems, which must operate in dynamic and unpredictable settings.

This research accepts the challenges related to classification of a waste and proposes to leverage advance deep learning architectures and attention mechanisms to greatly improve classification accuracy. The main part of the approach is employing ResNet50, a robust pretrained CNN model [12]. ResNet50 is a strong base line for tasks of classification by handling the issue of vanishing gradient through its residual connection. The fact that these connections allow for deeper network architectures to retain effective gradient flow (i.e., gradient flow 'through' the network) during backpropagation helps improve feature learning and representation [13].

Based on this foundation, the research is extended to include state-of-the art attention mechanisms to facilitate the model's concentration on salient features in waste images. To recalibrate channel wise feature responses the network uses squeeze-and-excitation (SE) blocks [14], which enable the network to focus on more informative features and suppress the less useful ones. Such capability significantly improves classification performance, especially with heterogeneous waste materials. The architecture also includes the convolutional block attention module (CBAM) mixed in to enable spatial and channel wise attention [15] additional to the focus. This dual mechanism is important so that the model can efficiently learn complex patterns in images, for instance, identify these between two materials that share similar colors or textures.

Since small and imbalanced datasets are used in our approach, we employ transfer learning and data augmentation techniques to resolve the problem they pose. Transfer learning uses networks that were trained on large datasets such as ImageNet to determine weights they already know and to adapt those already learned weights to the waste classification task [16]. It reduces data scarcity challenges, by reducing data dependency on in-depth domain-specific training data. Additionally, data augmentation techniques such as rotation, flipping, scaling, and cropping are applied in order to artificially augment the amount of training samples, which increases diversity in the training samples, thus increasing model robustness and generalization [17].

Despite advancements in deep learning for waste classification, existing models struggle with misclassification of visually similar waste types, dataset imbalance, and poor generalization to real-world waste scenarios. Traditional CNNs often fail to capture fine-grained distinctions between materials, leading to inaccurate classifications. Addressing these challenges requires advanced feature extraction mechanisms, such as attention modules that refine feature selection and enhance discriminative learning.

Current AI-based waste classification models suffer from several limitations, including suboptimal feature extraction, high misclassification rates due to visually similar materials, and lack of adaptability to real-world waste scenarios. Simple CNNs fail to differentiate between recyclable and nonrecyclable items accurately. Furthermore, dataset imbalance exacerbates model bias, leading to poor classification of underrepresented waste categories. To overcome these issues, this study proposes an enhanced ResNet-50 model with integrated SE and CBAM attention mechanisms, which improves feature extraction and enhances classification accuracy.

Unlike traditional deep learning models, this study introduces a novel approach that integrates SE and CBAM into ResNet-50 for improved waste classification. SE enhances channel-wise feature recalibration, while CBAM further refines both spatial and channel attention, leading to superior classification accuracy. The model is trained on a balanced dataset using SMOTE to address data imbalance issues.

In this study, an extensive literature survey that is conducted to identify challenges in waste classification and to explore advancements in deep learning techniques is presented in Section II. In Section III, a robust methodology is proposed, incorporating ResNet50 with attention mechanisms like SE and CBAM, supported by transfer learning and data augmentation. The methodology is implemented, and experimental results demonstrate significant improvements in classification accuracy over baseline models produced in Section IV. The research concludes with future recommendations in Section V.

II. LITERATURE REVIEW

The researchers focused on waste classification as a critical component of sustainable waste management by integrating deep learning models with IoT technology [18]. It addressed challenges in waste classification, such as variability in waste characteristics (e.g., differences in color, texture, and thickness of materials like glass and plastic) and the limitations posed by imbalanced datasets, which restricted the generalization capabilities of ML models. The methodology employed a CNN, specifically ResNet34, achieving a classification accuracy of 95.3% through fine-tuning of a pretrained model. Waste was divided into two primary categories: digestible and indigestible, with further classification of indigestible waste

into five classes (cardboard, glass, plastic, metal, and paper). However, limitations were observed, such as difficulties in distinguishing visually similar classes like plastic and paper, and challenges arising from imbalanced datasets that influenced model predictions [19].

Preprocessing techniques, including resizing images, were implemented to simplify model complexity, while dropout layers and batch normalization were used to mitigate overfitting. Training was conducted using an 80%–10%–10% split for training, validation, and testing, with cyclical learning rate scheduling employed to optimize learning. Although the model achieved high accuracy, reliance on a small dataset and its limited capability in handling complex mixed-waste scenarios were identified as notable drawbacks. Furthermore, while ResNet34 performed effectively, it was suggested that larger datasets and more advanced architectures could enhance classification accuracy. This investigation demonstrated the potential of deep learning in waste management while highlighting areas for improvement in methodology and data availability [19].

Researchers used deep learning for classifying plastic waste into four categories: PET, HDPE, PP, and PS. Using CNNs, two network architectures were compared: a 15 layer custom CNN and a 23 layer AlexNet inspired model. To address dataset imbalance, the research used rotation-based augmentation and the WaDaBa database with ~33,000 – 37,000 images per class. By using only 4 epochs and a 15 layer CNN, the accuracy reached 97.43% for low resolution images (rows 120×120), whereas the 23 layer CNN reached a higher accuracy of 99.23%, but it took much longer to train and thus is not practical for deployment in real applications. Despite the use of a controlled experimental setup with single object images, the applicability of the model to mixed waste scenarios was limited. At the same time, while data augmentation leveled the field, the model's reliance on a prior database prevents scalability to new environments. These results suggest we must develop lightweight, efficient architectures that generalize well onto real-world conditions [20].

To increase recycling efficiency, authors proposed a sensor-based hybrid plastic waste sorting system using CNNs. A camera with six sensors, including near infrared (NIR) and gas detectors, were used for the classification of waste into material type, cleanliness and dimensions. Inception-v3 scores the highest accuracy (78%) of all CNN models tested but struggled with residual waste and dirt contaminated plastics. Incomplete quality assessment points of the system included low accuracy for dark plastics, difficulties for mixed plastics (bottles with caps or labeling) as well as misclassification of flattened or reshaped objects caused by sensor misunderstanding. The deep learning model was further constrained by a paucity of size and diversity in the custom waste data augmented for diversity. To fill such gaps in plastic waste sorting efficiency and contamination reduction [21] this research focused on the importance of more robust data integration of the spectral and visual data.

To address inefficiencies of waste classification, researchers introduced a smart garbage management system using deep learning. In this paper, two deep learning architectures, ResNeXt and ResNet50, were combined using modified forms, including vertical and horizontal blocks, for the purpose of image classification enhancement. Traditional CNN showed an accuracy of 98.9%, and this was superior to the system's categorization of waste into seven classes. Real-time implementation involved an ultrasonic and thermal sensors-based hardware setup that uses a Raspberry Pi with a Pi camera and an LCD display to monitor the waste levels. The system had limitations; however, it required a large dataset to deal with the diversity of the waste materials and the misclassification of mixed waste situations. The dependence on a controlled experimental environment limited its scalability. It concluded with the potential augments to waste classification accuracy and efficiency derived from combining advanced CNN networks with hardware innovations [22].

The authors evaluated and compared the performance of four deep learning models—ResNet50, InceptionV3, Xception, and GoogleNet—for waste material classification using a dataset with 1,451 images across four classes: glass, cardboard, metal, and trash. We achieved an accuracy of 95%, with very slight misclassification in the glass category using ResNet50. Other models that we tried included Xception and GoogleNet, the latter achieved a strong performance but struggled to differentiate glass from trash. InceptionV3 performed the best, with a perfect accuracy of 100%. Dataset limitations were addressed via data preprocessing such as resizing and augmentation, but small size of dataset prevented model generalization. Misclassified materials were limited, however, due to the lack of diversity in training data, and reliance on controlled experimental conditions, the results may not generalize to the real world. The study showed that to achieve high accuracy and scalability in waste classification, it would be worthwhile to think bigger by exploring larger datasets, fine tuning the architectures, and using ensemble approaches [23].

In the proposed [24], improved DCNN framework for binary waste classification and organic and recyclable household waste is considered. To account for the computational inefficiencies and overfitting, the model incorporates architectural improvements such as the LeakyReLU activation function and drop out regularization. A dataset of 25,077 high-resolution images was obtained and was spilt into 70% training and 30% testing. First, the proposed model achieved an accuracy of 93.28 %, better than VGG16, VGG19, MobileNetV2, DenseNet121, and EfficientNetB0 pretrained models. The limitations were due to the dataset's lack of diversity and complex background and inter-class variability. Furthermore, the model's generalization to realistic scenarios where noise or lighting variation exists is unexplored. The future recommendations are dataset augmentation, experimentation with EfficientNet Lite for larger datasets, and class imbalance via more advanced hyperparameter tuning. However, the work here showed that DCNNs might automate waste management, albeit there was room for improvement in their scalability and robustness.

In an extensive review of intelligent waste classification techniques with AI using ML and deep learning (DL) model, researchers studied the methods. They proved to be a dominant approach in waste classification with image-based information because of their high accuracy and low complexity. Among the methods tried such as DenseNet169, ResNet50 and MobileNetV2, classification accuracies from 87% to 99% for the dataset and task were achieved. Second, the paper also mentioned the success of transfer learning, preparing models with pretrained features and fine tuning them to give the best results even when faced with small and imbalanced datasets. Limitations were identified in dependency on controlled datasets with limited diversity and lack of generalization in real-world mixed waste environments. There was in addition underrepresentation of certain waste types like organic and hazardous waste, which impact on accuracy of classification for these types. The study underscored the need for larger, diverse datasets and the exploration of lightweight architectures for

scalable applications, emphasizing the potential of transfer learning and ensemble techniques for further performance improvements [25].

Different interpretations for the TrashBox dataset with seven categories such as cardboard, glass, metal, plastic, and medical waste were experimented by using the deep learning models for the sake of research. ResNeXt-101 performed the best among ten tested models with a test accuracy (89.62%) and F1 score (89.66%) far greater than ResNeXt-50. Performance comparable to that demonstrated by models like ShuffleNetV2 was shown, but at lower cost in terms of computation, making them ideally suited for lightweight applications. However, some limitations were known, including moderate accuracy for underrepresented waste classes, dependence on computational resources for high performing architectures (e.g. ResNeXt), and potential overfitting in complex models. There were also challenges related to practical deployment on top of the limited real-world complexity and the lack of mixed waste representation in the dataset. The research showed the tradeoffs between the model complexity and performance, claiming that enhanced datasets, sophisticated architectures, and federated learning frameworks are needed to construct scalable, precise and unnecessary waste placing frameworks for the wise city purposes [26].

The summary of above research is given in Table I.

From this summary, it is observed that authors implemented ResNet34, 15-layer custom CNN, 23-layer AlexNet-inspired CNN, Inception-v3, ResNeXt, ResNet50, DenseNet169, MobileNetV2, Xception, GoogleNet, LeakyReLU-enhanced DCNN, ResNeXt-101, ResNeXt-50, ShuffleNetV2, and ResNet-34. Maximum authors tested their research with the ResNet model and CNN. Considering the limitations of existing surveys in this research, we decided to take advantage of the attention mechanism of SE and CBAM with ResNet-50.

III. METHODOLOGY

The primary issues highlighted in the existing survey include difficulties in handling class imbalance, misclassification of visually similar waste categories, and poor generalization in real-world waste scenarios. Attention mechanisms address these by enhancing the model's focus on critical features.

ResNet-50 is already well-documented as a reliable backbone in several research papers. Combining it with SE and CBAM directly targets feature optimization, making the model more robust in differentiating subtle differences in waste materials. The Proposed Workflow is as per Fig. 1.

The flowchart in Fig. 1 illustrates the workflow for a deep learning-based waste classification system, divided into four key stages: **Data Acquisition**, **Data Preprocessing**, **Model Training**, and **Model Evaluation**. In the **Data Acquisition** stage, raw images of various waste materials, such as cardboard, glass, metal, plastic, and trash, are collected to build a diverse and representative dataset. The Data Preprocessing stage processes these raw images with labeling that assigns class categories, data augmentation to ensure increased dataset diversity while dealing with class imbalance and segmentation so that areas of interest could be selected that would lead to better extraction of the features.

Once the data are preprocessed, it moves to the Model Training stage, where three deep learning models are employed: ResNet-50 based as the base model, ResNet-50 with SE (Squeeze-and-Exit) channel attention, ResNet-50 with CBAM, channel and spatial attention. They extract and embed significant features from data to classify the waste materials with accuracy. In the final stage of the modeling pipeline, referred to as Model Evaluation, trained models are evaluated against metrics such as training accuracy, validation accuracy, and test accuracy. These evaluations choose the best

Table I.   Summary of literature survey

| Reference	Activity	Methods Applied	Limitations Observed
1	Waste classification using deep learning and IoT integration.	ResNet34 fine-tuned on a small dataset with preprocessing, dropout, and batch normalization.	Struggled with visually similar classes, imbalanced datasets, and limited generalization for mixed scenarios.
2	Classification of plastic waste into four categories (PET, HDPE, PP, PS).	Compared 15-layer custom CNN and 23-layer AlexNet-inspired CNN using WaDaBa dataset with augmentation.	Limited to single-object images, high computational cost for complex models, and scalability challenges.
3	Hybrid sorting of plastic waste combining sensor technology and deep learning.	Inception-v3 model with near-infrared (NIR) and gas sensors for classification.	Poor accuracy for dark plastics, issues with mixed materials, and dataset limitations in size and variation.
4	Smart garbage classification into seven categories with real-time implementation.	Integrated ResNeXt and ResNet50 with architectural modifications and hardware (Raspberry Pi, sensors).	Required larger datasets, faced misclassifications in mixed waste, and relied on a controlled setup.
5	Performance evaluation of four deep learning models for waste classification.	Tested ResNet50, InceptionV3, Xception, and GoogleNet on a small dataset of 1,451 images with preprocessing.	Small dataset limited generalization; challenges in distinguishing glass and trash; reliance on controlled data.
6	Binary classification of organic and recyclable household waste using improved CNNs.	Enhanced DCNN with LeakyReLU activation and dropout regularization on a dataset of 25,077 high-resolution images.	Lacked dataset diversity; struggled with real-world scenarios involving noise and lighting variations. (90.01%)
7	Intelligent waste classification techniques using machine learning and deep learning.	Evaluated CNN models like DenseNet169, ResNet50, and MobileNetV2 with transfer learning.	Dependency on controlled datasets; poor representation of hazardous waste and generalization in mixed waste.
8	Comparison of deep learning models for waste classification using the TrashBox dataset.	Tested ResNeXt-101, ResNeXt-50, ShuffleNetV2, and ResNet-34 with federated learning framework proposed.	Moderate accuracy for underrepresented classes; dataset lacked real-world complexity; resource-intensive.
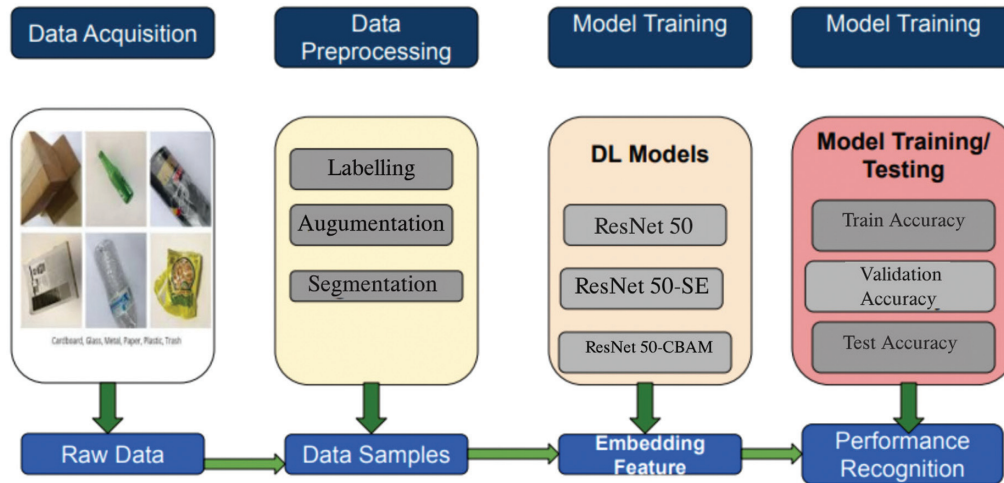
Fig. 1. The proposed workflow.
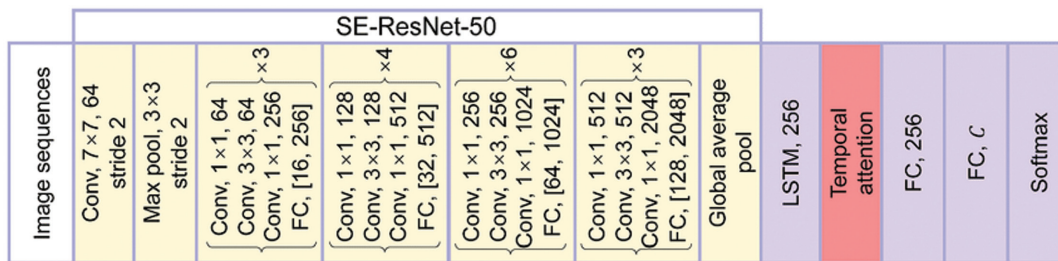
Fig. 2. SE+ResNet-50 architecture.

model, optimized classification performance. The structured workflow exemplifies efficient data processing, robust training, and comprehensive evaluation to enhance system capability of classifying waste materials with good accuracy. Fig. 2 demonstrates that the SE+ResNet-50 architecture.

Figure 2 illustrates the SE enhanced ResNet-50 architecture, which improves waste classification performance by optimizing feature selection within deep convolutional layers. The model builds upon the ResNet-50 backbone but integrates SE blocks, which introduce an adaptive recalibration mechanism for feature maps.

A. STEP-BY-STEP BREAKDOWN OF ARCHITECTURE OF SE+RESNET-50 ARCHITECTURE

1. **Initial Convolution and Pooling**:
   - ○ Convolution with a large kernel (7×7) and stride 2 is applied to extract low-level features.
   - ○ Max pooling reduces the spatial resolution for computational efficiency.
2. **Residual Block**:
   - ○ Each residual block has:
     - ▪ Three convolutions (1×1, 3×3, 1×1) to reduce, process, and restore the dimensions.
     - ▪ An SE Block:

- ▪ **Squeeze**: Global average pooling reduces each feature map to a single scalar value.
- ▪ **Excitation**: Fully connected layers with sigmoid activation compute channel importance weights.
- ▪ **Scaling**: Reweights the feature maps by multiplying with the computed weights.
- ▪ A skip connection adds the input features to the scaled features, allowing efficient gradient flow.
3. **Global Average Pooling**:
   - ○ Converts the final feature maps into a compact vector by averaging across spatial dimensions.
4. **Fully Connected Layers**:
   - ○ Dense layers refine the extracted features and perform classification with a softmax activation to output probabilities.

Pseudo-code of the same is given below with detailed explanation how it functions step by step

 # Input Image Processing
Input: Image (I)
Perform initial convolution:
$X1 = Conv2D(I, kernel\_size=(7,7), filters=64, strides=2, padding='same')$
$X2 = ReLU(X1)$
$X3 = MaxPooling2D(X2, pool\_size=(3,3), strides=2, padding='same')$

Pass Through Residual Blocks with SE Mechanism

Input: Feature Map (X)
For each residual block:

1. *Perform 1×1 convolution to reduce dimensions:*

   X_reduce = Conv2D(X, kernel_size=(1,1), filters=ReducedDim, activation='relu')

2. *Perform 3×3 convolution to extract spatial features:*

   X_spatial = Conv2D(X_reduce, kernel_size=(3,3), filters= SpatialDim, activation='relu', padding='same')

3. *Perform 1×1 convolution to restore dimensions:*

   X_restore = Conv2D(X_spatial, kernel_size=(1,1), filters= RestoredDim, activation='relu')

 # Squeeze-and-Excitation (SE) Block

4. *Squeeze operation:*

   Z = GlobalAveragePooling2D(X_restore)

5. *Excitation operation:*

   S = Dense(units=ReducedDim, activation='relu')(Z)

   S = Dense(units=RestoredDim, activation='sigmoid')(S)

6. *Scale the feature maps:*

   X_scaled = Multiply()([X_restore, S])

7. *Add skip connection:*

   X = Add()([X_scaled, X])

   X = ReLU(X)

Perform Global Average Pooling

X_pooled = GlobalAveragePooling2D(X)
 # Fully Connected Layers for Classification

1. *Pass through a dense layer:*

   X_dense = Dense(units=256, activation='relu')(X_pooled)

2. *Perform softmax classification:*

   Output = Dense(units=Number_of_Classes, activation='softmax') (X_dense)

Final Output

Return: Class Probabilities (Output)
The CBAM+ResNet-50 architecture is shown in Fig. 3.
Pseudo-code of the same is given below with a detailed explanation how it functions step by step

Input Image Processing

Input: Image (I)
Perform initial convolution:
   X1 = Conv2D(I, kernel=7×7, filters=64, stride=2)
   X2 = ReLU(X1)
   X3 = MaxPooling2D(X2, kernel=3×3, stride=2)

Pass Through Residual Blocks with CBAM Mechanism

Input: Feature Map (X)
For each residual block:

1. *Perform 1×1 convolution to reduce dimensions:*

   X_reduce = Conv2D(X, kernel=1×1, filters=ReducedDim)

2. *Perform 3×3 convolution to extract spatial features:*

   X_spatial = Conv2D(ReLU(X_reduce), kernel=3×3, filters= SpatialDim)

3. *Perform 1×1 convolution to restore dimensions:*

   X_restore = Conv2D(ReLU(X_spatial), kernel=1×1, filters= RestoredDim)

 # Apply CBAM Block

4. *Channel Attention Module:*

   4.1. *Perform global average pooling:*

      Z_avg = GlobalAveragePooling2D(X_restore)

   4.2. *Perform global max pooling:*

      Z_max = GlobalMaxPooling2D(X_restore)

   4.3. *Pass Z_avg and Z_max through a shared dense layer:*

      M_channel = Sigmoid(Dense(Z_avg) + Dense(Z_max))

   4.4. *Scale the feature maps:*

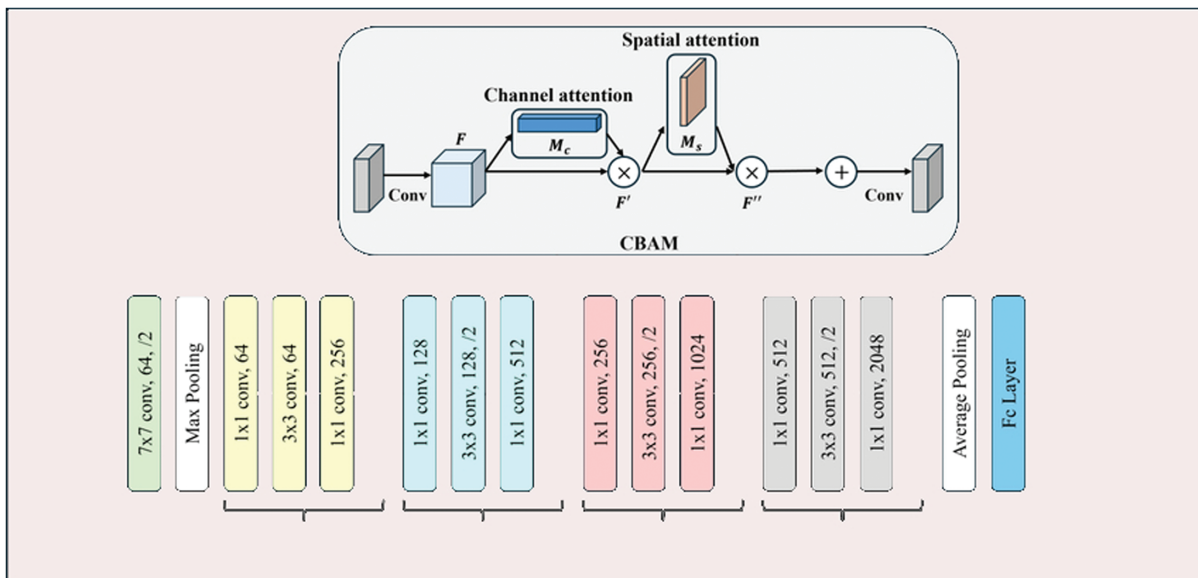   *X_channel = X_restore * M_channel*

Fig. 3. CBAM+ResNet-50 architecture.

5. *Spatial Attention Module:*

5.1. *Compute spatial descriptors using average and max pooling across channels:*

$Z\_avg\_spatial = Mean(X\_channel, axis=ChannelDim)$

$Z\_max\_spatial = Max(X\_channel, axis=ChannelDim)$

5.2. *Concatenate Z_avg_spatial and Z_max_spatial:*

$Z\_concat = Concatenate(Z\_avg\_spatial, Z\_max\_spatial)$

5.3. *Pass Z_concat through a 7×7 convolutional layer:*

$M\_spatial = Sigmoid(Conv2D(Z\_concat, kernel=7×7))$

5.4. *Scale the feature maps:*

$X\_spatial = X\_channel * M\_spatial$

6. *Add skip connection:*

$X = ReLU(X + X\_spatial)$

 # Perform Global Average Pooling
X_pooled = GlobalAveragePooling2D(X)
 # Fully Connected Layers for Classification

1. *Pass through a dense layer:*

$X\_dense = ReLU(Dense(X\_pooled, units=256))$

2. *Perform softmax classification:*

$Output = Softmax(Dense(X\_dense, units=Number\_of\_Classes))$

 # Final Output
Return: Class Probabilities (Output)

B. Step-by-Step Breakdown of CBAM Integration

1. INITIAL CONVOLUTION AND POOLING.

- Convolution (7×7) with stride 2 extracts basic features from the input image.
- Max pooling reduces spatial dimensions for computational efficiency.

2. RESIDUAL BLOCK.

- Each residual block includes three convolution layers:
  - **1×1 Convolution** reduces dimensions for computational efficiency.
  - **3×3 Convolution** captures spatial features.
  - **1×1 Convolution** restores dimensions.
- The CBAM module is applied after these convolutions.

3. CBAM MECHANISM.

a. Channel Attention:

- Extracts channel-wise importance using both global average pooling and global max pooling.
- Combines these two descriptors using shared dense layers to produce channel attention weights.
- Scales the input feature maps by these weights, prioritizing significant channels.

b. Spatial Attention:

- Extracts spatial importance by pooling across channels (average and max pooling).
- Concatenates the pooled spatial descriptors and applies a 7×7 convolution.
- Produces spatial attention weights to scale the feature maps, focusing on important spatial.

4. SKIP CONNECTION.

- Adds the scaled output from the CBAM module back to the input of the residual block, ensuring efficient gradient flow.

5. GLOBAL AVERAGE POOLING.

- Reduces the final feature maps to a compact vector by averaging across spatial dimensions.

6. FULLY CONNECTED LAYERS.

- Dense layers with softmax activation generate class probabilities for classification.

Integrating SE and CBAM with ResNet-50 has the following advantages.

Advantages of ResNet-50:

- Efficient feature extraction using residual connections, avoiding vanishing gradient issues.
- Handles deep architecture effectively due to skip connections, allowing better gradient flow.
- Pretrained on large datasets (e.g., ImageNet), making it robust for transfer learning.
- Provides a good balance between computational efficiency and performance.
- Suitable as a backbone for various computer vision tasks like classification, detection, and segmentation.

Advantages of SE+ResNet-50:

- Dynamically recalibrates channel-wise features using the SE block, focusing on important channels.
- Enhances model interpretability by prioritizing meaningful features while suppressing irrelevant ones.
- Effective on small or imbalanced datasets, as the SE mechanism improves feature discrimination.
- Lightweight addition to ResNet-50 with minimal increase in computational cost.
- Improves classification accuracy by boosting the representational power of the network.

Advantages of CBAM+ResNet-50:

- Integrates both channel and spatial attention, allowing the network to focus on both what and where to prioritize in feature maps.
- Further improves the focus on discriminative features compared to SE blocks alone.
- Enhances performance in datasets where spatial and channel features both play a critical role (e.g., in object or waste classification).
- Flexible design that can be easily integrated into ResNet-50 and other architectures.
- Slightly better accuracy than SE blocks in tasks requiring both spatial and channel attention.
- Robust against noisy or cluttered backgrounds by learning to ignore irrelevant spatial regions while enhancing critical areas.

IV. EXPERIMENTAL RESULTS

To experiment with the proposed flow, the TrashNet dataset on Kaggle is a collection of images designed for training ML models in waste classification tasks [27]. It includes images categorized into six classes: cardboard, glass, metal, paper, plastic, and trash. The class

Table II.   Data distribution

Class	Before balancing	After balancing
Cardboard	403	594
Glass	501	594
Metal	410	594
Paper	594	594
Plastic	482	594
Trash	137	594
Total	**2527**	**3564**

Table III.   ResNet 50 performance

Epoch	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
1	1.5649	0.3636	1.2275	0.5705
2	1.3292	0.5008	1.1132	0.6212
3	1.2761	0.5143	0.9731	0.6719
4	1.1955	0.5509	0.9496	0.6578
5	1.2046	0.5454	0.9063	0.7093
6	1.1347	0.562	0.8486	0.7272
7	1.135	0.5669	0.8069	0.7178
8	1.1078	0.5755	0.7886	0.7225
9	1.0638	0.593	0.761	0.7405
10	1.0967	0.5854	0.7526	0.7467

distribution of various images are in Table II. The dataset is initially undergoing a data imbalance problem. So, it was addressed using a SMOTE (Synthetic Minority over sampling) technique.

The distribution of the dataset before and after preprocessing is given below.

The balanced dataset is now split into Train, validation, and Test with 70:20:10 ratio, and then it is applied to ResNet-50 model. Here is the tabulated summary of the training and validation metrics of ResNet50 in Table III. Its graphical representation is shown in Fig. 4. The final test accuracy **74.42% is recorded and** resnet50_model.pth is saved in the local drive.

The training process for the model over 10 epochs demonstrates steady improvement in both training and validation metrics. The initial train loss of 1.5649 with an accuracy of 36.36% improves to a final train loss of 1.0967 and accuracy of 58.54%. Similarly, the validation loss decreased from 1.2275 to 0.7526, while validation accuracy increased from 57.05% to 74.67%. These results indicate consistent learning and generalization throughout the training. The model achieves a final test accuracy of 74.42%, and the trained model was saved to resnet50_-model.pth for further evaluation or deployment.

To enhance the model performance SE is integrated with ResNet50. Its performance is tabulated in Table IV. Similarly its graphical representation is shown in Fig. 5.

The integration of the SE module with ResNet-50 demonstrates a significant improvement in both training and validation metrics over 10 epochs. The model begins with a training loss of 1.5096 and accuracy of 52.36%, along with a validation loss of 1.4821 and accuracy of 54.62%. As training progresses, the training loss decreases steadily to 0.5992 with an accuracy of 89.47%, while the validation loss reduces to 0.4279 with an
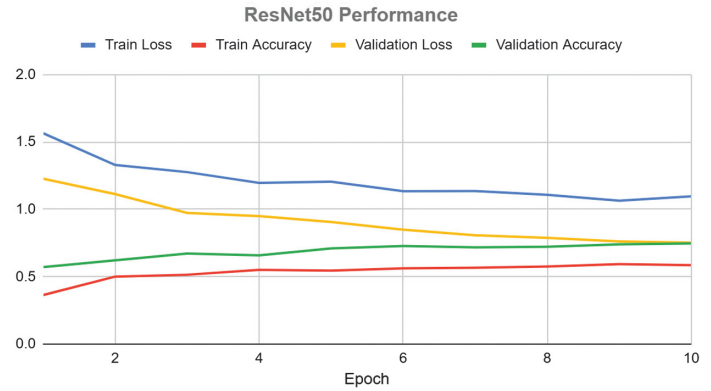

Fig. 4.   Graphical representation of ResNet 50 performance.

Table IV.   SE integration with ResNet 50 performance

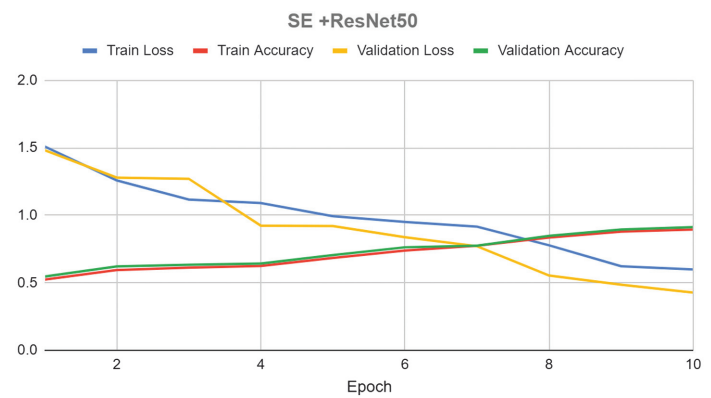| Epoch	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
1	1.5096	0.5236	1.4821	0.5462
2	1.2587	0.5945	1.279	0.6215
3	1.1169	0.6127	1.2702	0.6341
4	1.091	0.6245	0.9228	0.6421
5	0.9936	0.6845	0.9199	0.7054
6	0.951	0.7389	0.8376	0.763
7	0.9162	0.7742	0.7708	0.7748
8	0.7775	0.8354	0.5539	0.8475
9	0.6225	0.8786	0.4854	0.8946
10	0.5992	0.8947	0.4279	0.9126

Fig. 5.   Graphical representation of SE+ResNet 50 performance.

accuracy of 91.26%. These results highlight the effectiveness of the SE module in enhancing feature extraction and improving model generalization. The SE integration allows the model to achieve robust performance, demonstrating its potential for complex classification tasks. The SE+ResNet50 model achieved a final test accuracy of 93.47%, and the trained model is saved to SE_resnet50_model.pth for further evaluation or deployment.

The ResNet 50 model is accelerated with CBAM integration and its performance is articulated in Table V and graphical representation is shown with Fig. 6.

Table V.   CBAM integration with ResNet 50 performance

Epoch	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
1	1.2757	0.5565	1.3985	0.5621
2	1.1361	0.5946	1.2037	0.6015
3	1.0627	0.6207	1.1256	0.6413
4	0.8976	0.6719	1.0407	0.6578
5	0.8198	0.7124	0.9537	0.7541
6	0.7909	0.7823	0.9487	0.7845
7	0.7362	0.8225	0.9011	0.8431
8	0.5131	0.8642	0.7711	0.8775
9	0.4662	0.8941	0.6863	0.9034
10	0.4386	0.9124	0.6566	0.9345
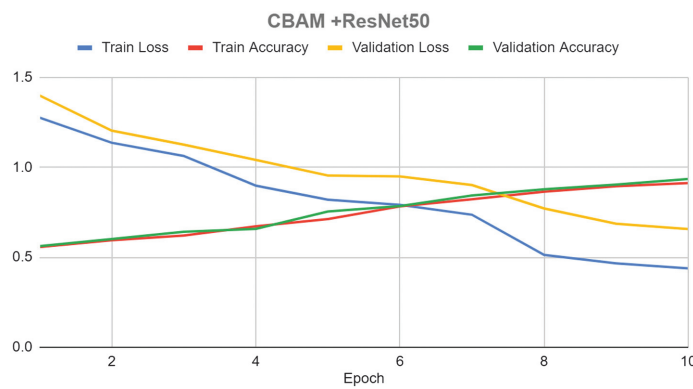
Fig. 6.   Graphical representation of CBAM+ResNet 50 performance.

The CBAM integration with ResNet-50 demonstrates consistent improvements in both training and validation metrics over 10 epochs. The training loss steadily decreases from 1.2757 to 0.4386, with the training accuracy improving from 55.65% to 91.24%. Similarly, the validation loss reduces from 1.3985 to 0.6566, while validation accuracy increased from 56.21% to 93.45%. These results highlight the effectiveness of the CBAM module in enhancing feature refinement and model generalization. The integration allows the model to achieve robust performance, making it suitable for complex classification tasks.

The CBAM+ResNet50 model achieves a final test accuracy of 95.74%, and the trained model was saved to CBAM_resnet50_-model.pth for further evaluation or deployment. These .pth files can be used to create any web based application using Flask or Django to classify the unknown waste image for classification in real time.

Comparison with exited models is shown in Table VI.

V. DISCUSSION

The baseline ResNet-50 model, without attention mechanisms, achieved 74.42% accuracy, indicating that the model struggled to differentiate between visually similar waste categories such as plastic and paper. The integration of the SE module enhanced feature recalibration, leading to a 19.05% increase in accuracy (93.47%), demonstrating that emphasizing channel-wise features helped resolve misclassification issues. Furthermore, the CBAM module improved accuracy to 95.74%, an additional 2.27% boost, by refining both spatial and channel attention. This suggests that waste classification benefits significantly from spatial awareness, allowing the model to distinguish objects with similar textures but different spatial patterns (e.g., crushed plastic vs. metal cans).

The high accuracy of CBAM-enhanced ResNet-50 (95.74%) suggested that integrating attention mechanisms in deep learning-based waste classification systems can significantly reduce errors in automated sorting facilities. This was particularly useful in real-world applications where manual sorting is prone to inconsistency and inefficiency. With an improvement of over 21% from the baseline, the model could be deployed in smart waste management systems, potentially reducing landfill waste by improving the identification of recyclable materials.

Despite achieving state-of-the-art accuracy, the model relies on data augmentation (SMOTE) to balance the dataset, which may not perfectly reflect real-world conditions where waste distribution is naturally imbalanced. Additionally, the model's computational complexity increases with attention mechanisms, potentially making deployment in low-resource settings challenging. Future work should explore lightweight versions of SE and CBAM, or knowledge distillation techniques to reduce model size without sacrificing accuracy

VI. CONCLUSION

This research addressed critical challenges in automated waste classification, including data imbalance, misclassification of visually similar classes, and generalization to real-world conditions. The proposed system employed ResNet-50 as the base model and incorporated attention mechanisms SE for channel recalibration and CBAM for both spatial and channel attention. Experimental results showed progressive improvement: ResNet-50 achieved 74.42% test accuracy, SE+ResNet-50 enhanced performance to 93.47%, and CBAM+ResNet-50 reached 95.74%. These

Table VI.   Comparison between the of proposed model and existing models

Model	Accuracy (%)	Remarks
ResNet34 [1]	95.78%	Slightly better than CBAM+ResNet-50, but lacks attention mechanisms.
Inception-v3 [3]	92.60%	Inferior accuracy, indicating that depth and feature selection in ResNet variants are more effective.
Integrated ResNeXt and ResNet50 [4]	94.80%	Performs well, but CBAM+ResNet-50 surpasses it due to better feature recalibration.
DCNN with LeakyReLU [6]	90.01%	Lower accuracy, suggesting that a deeper network with attention mechanisms (CBAM) extracts features better.
Proposed CBAM+ResNet-50	95.74%	Nearly matches ResNet34, with added attention mechanisms enhancing spatial and channel-wise feature extraction.

improvements demonstrated the efficacy of attention mechanisms in refining feature extraction and mitigating classification errors in heterogeneous datasets. The CBAM module's ability to address spatially distributed and channel-specific patterns proved especially effective for complex waste classes. The results indicated significant progress in creating scalable, efficient, and accurate waste classification models.

Despite the promising results, this study had some limitations. The model's reliance on SMOTE for dataset balancing may not have fully captured real-world waste distribution, leading to potential biases. Additionally, the integration of SE and CBAM attention mechanisms increased computational overhead, making real-time deployment on low-resource devices challenging. While the model generalized well on the TrashNet dataset, further validation was needed on diverse, real-world waste classification datasets. Future research should focus on developing lightweight attention modules to reduce computational costs, incorporating self-supervised learning to minimize data dependency, and adapting the model for real-time waste classification in smart waste bins and automated sorting systems. Expanding the study to include hazardous and mixed waste categories could further enhance its applicability in industrial and municipal waste management.

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] S. Nanda and F. Berruti, "Municipal solid waste management and landfilling technologies: a review," *Environ. Chem. Lett.*, vol. 19, no. 2, pp. 1433–1456, 2021.

[2] M. Abdullah and M. Z. Abedin, "Assessment of plastic waste management in Bangladesh: a comprehensive perspective on sorting, production, separation, and recycling," *Results Surf. Interfaces*, vol. 15, p. 100221, 2024.

[3] B. Fang et al., "Artificial intelligence for waste management in smart cities: a review," *Environ. Chem. Lett.*, vol. 21, pp. 1959–1989, 2023.

[4] W. L. Mao et al., "Recycling waste classification using optimized convolutional neural network," *Resour. Conserv. Recycl.*, vol. 164, p. 105132, 2021.

[5] M. Reza and M. Hassan, "AI-Driven solutions for enhanced waste management and recycling in urban areas," *Int. J. Sustain. Infrastruct. Cities Soc.*, vol. 8, no. 2, pp. 1–13, 2023.

[6] D. B. Olawade et al., "Smart waste management: a paradigm shift enabled by artificial intelligence," *Waste Manage. Bull.*, vol. 2, no. 2, pp. 244–263, 2024.

[7] D. Patil et al., "Machine learning and deep learning: methods, techniques, applications, challenges, and future research opportunities," In *Trustworthy Artificial Intelligence in Industry and Society*, Maharashtra: Deep Science Publishing, 2024, pp. 28–81.

[8] G. Goyal et al., "Design and implementation of VGD6-NET framework for waste segregation using 3-tier convolutional neural networks," *J. Mater. Cycles Waste Manag.*, vol. 27, pp. 223–240, 2024.

[9] I. H. Rather, S. Kumar, and A. H. Gandomi, "Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets," *Artif. Intell. Rev.*, vol. 57, no. 9, p. 226, 2024.

[10] M. M. Bejani and M. Ghatee, "A systematic review on overfitting control in shallow and deep neural networks," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 6391–6438, 2021.

[11] M. Mundt et al., "A wholistic view of continual learning with deep neural networks: forgotten lessons and the bridge to active and open world learning," *Neural Netw.*, vol. 160, pp. 306–336, 2023.

[12] W. Islam et al., "Improving performance of breast lesion classification using a ResNet50 model optimized with a novel attention mechanism," *Tomography*, vol. 8, no. 5, pp. 2411–2425, 2022.

[13] M. K. Panda et al., "Encoder and decoder network with ResNet-50 and global average feature pooling for local change detection," *Comput. Vis. Image Underst.*, vol. 222, p. 103501, 2022.

[14] S. Zhu et al., "SECA-Net: squeezed-and-excited contextual attention network for medical image segmentation," *Biomed. Signal Process. Control*, vol. 97, p. 106704, 2024.

[15] M. Zakariah and A. Alnuaim, "Recognizing human activities with the use of convolutional block attention module," *Egypt Inf. J.*, vol. 27, p. 100536, 2024.

[16] H. Santoso et al., "A hybrid model for dry waste classification using transfer learning and dimensionality reduction," *Int. J. Inform. Visualization*, vol. 8, no. 2, pp. 623–634, 2024.

[17] A. Mumuni and F. Mumuni, "Data augmentation: a comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, 2022.

[18] A. Sleem, "Enhancing sustainability through automated waste classification: a machine intelligence framework," *Sustainable Mach. Intelligence J.*, vol. 5, no. 6, pp. 1–8, 2023.

[19] M. W. Rahman et al., "Intelligent waste management system using deep learning with IoT," *J. King Saud University-Computer Inf. Sci.*, vol. 34, no. 5, pp. 2072–2087, 2022.

[20] J. Bobulski and M. Kubanek, "Deep learning for plastic waste classification system," *Appl. Comput. Intelligence Soft Comput.*, vol. 2021, no. 1, p. 6626948, 2021.

[21] M. Dokl et al., "A waste separation system based on sensor technology and deep learning: a simple approach applied to a case study of plastic packaging waste," *J. Cleaner Prod.*, vol. 450, p. 141762, 2024.

[22] J. Gunaseelan, S. Sundaram, and B. Mariyappan, "A design and implementation using an innovative deep-learning algorithm for garbage segregation," *Sensors*, vol. 23, no. 18, p. 7963, 2023.

[23] I. B. Al-Mashhadani, "Waste material classification using performance evaluation of deep learning models," *J. Intell. Syst.*, vol. 32, no. 1, p. 20230064, 2023.

[24] M. Chhabra et al., "Intelligent waste classification approach based on improved multi-layered convolutional neural network," *Multimed. Tools Appl.*, vol. 83, pp. 84095–84120, 2024.

[25] I. Nasir and G. A. Aziz Al-Talib, "Waste classification using artificial intelligence techniques:literature review," *Technium*, vol. 5, pp. 49–59, 2023.

[26] H. Ahmed Khan et al., "Enhancing trash classification in smart cities using federated deep learning," *Sci. Rep.*, vol. 14, p. 11816, 2024.

[27] Kaggle, "TrashNet," https://www.kaggle.com/datasets/feyzazkefe/trashnet/data