

Advanced Deep Learning Framework for Waste Image Categorization Using Attention-Enhanced AlexNet

Shivran Priyanka Suresh Kumar^{1*} and Dr. Monika Saxena²

¹PhD Research Scholar, ²Associate Professor

Email: priyankashivran@gmail.com

^{1,2} Affiliation Address: Banasthali Vidyapith, Jaipur

Email: muskan.saxena@gmail.com

Received:
01/10/2025
Revised:
09/10/2025
Accepted:
25/10/2025
Published:
10/11/2025

ABSTRACT

Effective waste sorting is an essential part of contemporary waste management systems, encouraging recycling, minimizing landfill consumption, and facilitating environmental sustainability. Deep learning has proven to be an effective means for automating the process through precise and efficient image-based waste sorting. This research proposes a state-of-the-art deep learning architecture that incorporates an attention mechanism into AlexNet to enhance classification accuracy by concentrating on the most informative image features. The collection includes images categorized as non-biodegradable waste (metal cans, plastic bottles, plastic bags) and biodegradable waste (wood, paper, food waste, leaves), which supports effective model training and validation. Attention-augmented AlexNet is contrasted with a regular AlexNet and a classical "Convolutional Neural Networks (CNN)", with 99.36% accuracy, vastly better compared to CNN (95.32%) and regular AlexNet (94.41%). The results affirm the model's capacity to minimize misclassification, especially in visually comparable classes, and thus represent a good solution to efficient multi-class waste classification and eco-friendly waste management measures.

Keywords: Deep Learning, Waste management System, AlexNet, CNN, Attention Mechanism.

INTRODUCTION:

The large-scale production of disposable items in almost all industrial segments has led to an explosive growth of the "Municipal Solid Waste (MSW)" disposal issue in recent times. Examples are light bulbs, plastic bags, foams, and bottled drinking water packaged in single-use plastic containers [1]. The pressing necessity for environmental equilibrium, now so severely intercepted by human interventions in the past two centuries, has come to be the main source of motivation for efficient waste management practices. MSW is a broad range of items—anything from cans, bottles, disposable glasses, and snack packets to furniture, electronics, tires, and major home appliances—which are divided into hazardous, non-hazardous, disposable, and non-disposable types [2].

Modern waste identification technologies tend to combine color and texture characteristics with machine learning-based classification models. Although these techniques are capable of initial classification, they still have shortcomings in accuracy, computational resources, dataset size, and generalization performance [3]. Most traditional image recognition algorithms use small-scale datasets, which results in overfitting and compromised robustness in actual scenarios. In addition, heterogeneity and uncertainty of waste due to the diversity in shape, texture, color, and contamination

create enormous challenges for stable classification performance.

Deep learning has in recent years been a revolutionary method for automated waste image recognition. Using multi-layer neural structures, DL methods are able to automatically learn sophisticated hierarchical feature representations of big datasets. This ability has contributed to groundbreaking progress in speech and image recognition tasks. For example, new neural network-based solutions have been proposed for e-waste classification with a recognition accuracy of 90% to 97% for chosen types of waste [4].

Machine learning techniques, especially CNNs, have shown excessive potential for learning from image data to make accurate classifications [6]. CNN-based models can consume images of solid waste and make classifications for hazardous, recyclable, organic, and non-recyclable items, without any manual feature engineering required [7]. Another advantage of deep learning architecture is that they learn feature representations in the raw data and improve with more examples while they are not "handcrafted," and hence, have improved performance.



Figure 1: Key advantages AlexNet framework for waste image categorization.

AlexNet, which is one of the earliest neural network architectures that gained worldwide attention, was a breakthrough in the area of computer vision when it achieved a wide margin of victory in the 2012 “ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)” [9]. This is especially challenging in the context of waste classification, where the background of the image can have significant noise and only selected regions of the total image contain useful waste-related features. To solve this problem, recent developments in attention mechanisms have endowed neural networks with the ability to dynamically attend to the most significant parts of an image. Attention modules improve feature learning by focusing on spatial or channel-wise features that are most important for the task of classification [10].

The aim of this paper is to propose and assess an enhanced deep learning architecture for waste image classification employing an attention-augmented AlexNet model. Through the integration of an attention mechanism into the AlexNet structure, the suggested architecture seeks to enhance the model's attention on the most salient visual attributes, thus improving classification accuracy and misclassification reduction, especially for highly similar-looking waste types. This method is designed to facilitate effective waste segregation, enhance recycling, and ensure environmental sustainability.

Municipal waste image classification is central to smart recycling, circular economy logistics, and self-driving sorting. Real-world systems need to cope with occlusion, grime, and heavy intra-class variation at the cost of being slow on edge hardware. To determine design choices that trade off accuracy, robustness, and efficiency—used in constructing an Attention-Enhanced AlexNet. Huang et al., 2021 [11] presented a single Vision Transformer for reusable waste, avoiding CNN receptive-field limitations and attaining 96.98% on TrashNet through global self-attention. Islam et al., 2023 [12] presented EWasteNet, a dual-stream DeiT with Sobel-edge and ASPP-attention streams, attaining 96% on the eight-class E-Waste Vision dataset and demonstrating edges supplement semantic context. Nafiz et al., 2023 [13] constructed “ConvoWaste,” an Improved-DCNN-based detection and mechatronic segregation apparatus with telemetry, achieving ~98%

accuracy and exhibiting low-cost end-to-end deployment. Chhabra et al., 2024 [14] applied an Improved-DCNN with transfer learning on two-class organic vs. recyclable waste (25,077 images; 70/30 split), achieving 93.28% accuracy and compared with VGG/MobileNet/DenseNet/EfficientNet. Wang et al., 2024 [15] introduced Garbage FusionNet (GFN), merging ResNet local features with ViT global context and incorporating PPM+CBAM to enhance multi-scale attention and robustness on Garbage and TrashNet datasets.

Shrivastava et al., 2024 [16] simulated nystagmus through differential blurring to regularize ViT, improving over typical ViT baselines by 2–6% and emphasizing biologically motivated enhancement for real-world blur. Wang et al., 2024 [17] tuned a CNN feature extractor using Capuchin Search and classified using ECOC-ANN, achieving 98.81% (TrashNet) and 99.01% (HGCD), with a $\geq 1.46\%$ improvement, highlighting the effect of meta-heuristic tuning and resilient decoding. Qiu et al., 2025 [18] augmented EfficientNetV2 with Channel Efficient Attention (preventing dimensional scaling) and a light multi-scale SAFM with depth-wise separable, along with robust augmentation, achieving 95.4% on Huawei Cloud and improving the baseline by 3.2% with balanced accuracy-efficiency. Jose et al., 2025 [19] proposed a Channel-and-Spatial Attention-based Multiblock CNN that is used to classify patch-level municipal waste with a precision of 98.73%, MAE 0.048, RMSE 0.087, demonstrating accurate attention-driven feature learning for real-time application. Nahiduzzaman et al., 2025 [20] presented a three-stage pipeline with a parallel depthwise-separable CNN and an ensemble ELM (PI-ELM + L1-RELM), scaling from 2 to 36 classes on the TriCascade WasteImage dataset with up to 96% (binary) and 85.25% (36-class) accuracy.

Q. Zhang et al. (2021) [21] presented a transfer learning-based DenseNet169 model for the classification of trash images. In another publication, Q. Zhang et al. (2021) [22] enhanced rubbish sorting accuracy through the utilization of deep learning, allowing smart waste classification through computer vision and smartphones. H. Abdu et al. (2022) [23] performed a thorough survey of waste detection image classification and object detection models. S. Suruc et al. (2023) [24] created six deep learning models for sorting waste material with fivefold cross-validation. They found that the MobileNetV2 model performed the best with 99.36% accuracy, 0.94 MCC, 0.99 recall, and 0.98 for both F1-score and precision. They used a one-vs.-rest strategy for class-level analysis as well. N. Li et al. (2023) [25] introduced two deep learning approaches—CNN and Graph-LSTM—for the detection of typical waste materials carried on belt conveyors in garbage collection systems. H. Zhang et al. (2023) [26] suggested a lightweight hybrid deep learning model for garbage classification.

Some previous studies employing AlexNet for other purposes are Zhu et al. (2018) [27], who implemented a

high-performance deep learning architecture for classification of vegetable images employing AlexNet in Caffe; R. A et al. (2019) [28], who employed AlexNet CNN for effective shot classification in sports videos in a field; I. Singh et al. (2022) [29], who used a three-level CNN architecture inspired by AlexNet to identify toxic comments from the Wikipedia forum (Google Jigsaw dataset); and A. Kumar et al. (2022) [30], who used an improved AlexNet classifier with Fast Fourier Transform (FFT)-based feature extraction for classifying ECG arrhythmia into four classes.

Some works directly targeted attention mechanisms. Z. Niu et al. (2021) [31] investigated current attention models and suggested a comprehensive framework to further explore attention mechanisms. H. Fukui et al. (2019) [32] presented the Attention Branch Network (ABN), which extends response-based visual explanation models with a branch structure that includes attention. M.-H. et al. (2022) [33] presented an in-depth survey of attention mechanisms for computer vision, dividing them into channel, spatial, temporal, and branch attention, and providing a companion repository that can be used for research.

Despite advances made in recent research through deep learning methods for waste image classification, some challenges persist unanswered. Current models produce high accuracy but tend to be computationally demanding, in turn restricting their applicability to resource-limited and real-time scenarios. Hybrid frameworks that rely on convolutional networks, optimization, and FFT-based feature improvement have been promising, but few have incorporated sophisticated attention mechanisms toward enhanced feature extraction efficiency in low-resource architectures like AlexNet. These deficiencies underscore the importance of attention-augmented AlexNet-based architecture in achieving balance between efficiency, accuracy, and interpretability in real-world waste image classification.

This paper follows this structure: the introduction provides the background and importance of waste image classification and then related work that presents the current techniques. The proposed methodology discusses the dataset used, the attention-augmented AlexNet model, and training and evaluation. The results and analysis section contains confusion matrix interpretation, ROC curve and AUC metric evaluation, and comparison with the baseline models. Discussion section interprets the results, and lastly, the conclusion summarizes the major findings and proposes directions for future research.

MATERIALS AND METHODS

2.1 Data Collection

The “Waste Segregation Image Dataset,” which is accessible to the general public on Kaggle, served as the dataset for this investigation. To make model training and assessment easier, the photos are separated into separate train and test folders and classified as biodegradable and non-biodegradable garbage. The dataset was assembled from many publically accessible

sources in order to offer a strong collection of garbage photos that have been annotated for use in classification tasks.

2.2 Data Description

Images in the collection are divided into two primary categories: non-biodegradable and biodegradable. Four unique classifications are further subdivided into each of these categories. Paper, leaves, food scraps, and wood debris go into the biodegradable group; plastic bags, bottles, and metal cans fall into the non-biodegradable category (figure 2). With a fair distribution of photos among the various trash kinds, the dataset is organized into distinct folders for training and testing. This framework offers a complete collection of labeled data for creating and improving trash categorization algorithms, facilitating efficient model training and performance assessment.



Figure 2: Plastic Waste

2.3 Data Preprocessing

2.3.1 Data Augmentation

Resampling volumetric data to a uniform voxel size across different instances standardize the input dimensions [2]. Data augmentation techniques, such rotation, flipping, zooming, shearing, and brightness/contrast modifications, are used to artificially increase the training dataset in order to improve the model’s generalization. This helps to avoid overfitting, especially in smaller datasets [25].





Figure 3: Waste materials dataset

2.3.2 Image Normalization

When images are normalised, or scaled from 0 to 1, or normalised to a mean of 0 and standard deviation of 1, it allows for faster training and greater convergence [26].

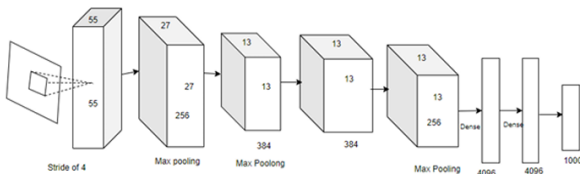
2.4 Train-Test Split

Used a dataset for trash classification j that is available to the public to evaluate our model's performance. Plastic, metal, paper, and glass are just a few of the many types of trash depicted in this assortment. The dataset was split into two parts: the training set, which contained 70% of the data, and the test set, which had 30% of the data.

2.5 Model Building

2.5.1 AlexNet

The AlexNet network has eight layers, three of which are completely linked and five of which are convolutional. After the first, second, and fifth convolutional layers, there is the pooling layer, and finally, there is the output, or softmax, layer. Following conv1 and conv2 are the response-normalization layers, often known as the norm1, norm2, and conv3 layers, respectively [17]. The AlexNet network has eight layers, three of which are completely linked and five of which are convolutional. Following the first, second, and fifth convolutional layers—as seen in Figure 4—is the pooling layer, and finally, the softmax or output layer.



Attention Mechanism

The attention mechanism enables the model to concentrate on the most significant elements of the input picture. This is especially beneficial for garbage sorting, as various waste kinds may exhibit unique visual characteristics. Our model incorporates a spatial attention mechanism subsequent to the last convolutional layer of AlexNet. This process produces a spatial attention map that emphasizes the areas of the picture most pertinent to the categorization job. The attention-augmented feature map is subsequently sent to the fully linked layers for classification [21]. After weighing each feature, the weighted summation approach was used for deep-level feature mining. The calculation formula is:

$$\beta_i = \sigma(W_i h_t + b_i) \quad (2)$$

Attention Weights Calculation:

$$a_i = \text{softmax}(\beta_i) = \frac{\exp(\beta_i)}{\sum_i \exp(\beta_i)} \quad (3)$$

The attention weight given to the concealed state h_i is represented by a_i Weighted Sum of Hidden States:

$$O = H \otimes a_i \quad (4)$$

The output prediction in this case is represented by O, which is the weighted sum of all hidden states, with each hidden state's contribution being determined by its attention weight.

Query, Key, and Value Computation

Figure 4: AlexNet Model Architecture

Usually, the convolutional layer's feature maps are produced by merging the many feature maps that the higher layer computed. The convolutional layer's primary job is featuring extraction. The convolutional layer calculates in the following manner.

$$X_n^l = \sum_{i \in M_n} X_i^{l-1} k_{im}^l + b_n^l$$

(1)

where, l in k denotes the i -th element in the n th convolution kernel of layer l ; b_n^l is the n th offset of layer l ; denotes the convolution process; and X_n^l represents the n th feature map of layer l . M_n denotes a collection of feature maps chosen from the input feature maps.

Eight layers make up the AlexNet architecture: three fully linked layers come after five convolutional layers. The following are the main elements of its architecture:

- **Convolutional Layers:** In the first convolutional layer, which employs the ReLU activation function, there are 96 11x11 filters with a stride of 4. Smaller filters, such 5x5 and 3x3, are used by subsequent layers to extract finer-grained information from the input pictures [27].
- **Pooling Layers:** To down-sample the feature maps and allow the model to learn invariant features, max pooling layers are applied after certain convolutional layers [28].
- **Dropout:** Applying dropout to the fully connected layers causes a portion of the neurons to be randomly set to zero during training in order to reduce overfitting. This method enhances the model's ability to be more widely applicable [29].
- **Data Augmentation:** Image flipping, colour jittering, and picture translation are among data augmentation techniques that AlexNet employs to increase the variety of the training dataset and fortify the model's resistance [30].

Typically, query, key, and value vector computation occur inside the attention mechanism. These vectors are created from each hidden state h_i ,

$$q_i = W_Q \cdot h_i, \quad k_i = W_K \cdot h_i, \quad v_i = W_V \cdot h_i \quad (5)$$

where the query, key, and value vectors for the time step are represented by the values q_i , k_i and v_i while the associated weight matrices for the query, key, and value transformations are represented by W_Q , W_K and W_V .

Scaled Dot-Product Attention

To determine the attention scores, we use the dot product of the query and key vectors and scale it by the square root of the dimensionality ($\sqrt{d_k}$). The next step is to generate normalised attention weights using a softmax:

$$\alpha_{ij} = \frac{\exp\left(\frac{q_i \cdot k_j}{\sqrt{d_k}}\right)}{\sum_{j=1}^t \exp\left(\frac{q_i \cdot k_j}{\sqrt{d_k}}\right)} \quad (6)$$

$$O_i = \sum_{j=1}^t \alpha_{ij} \cdot v_j \quad (7)$$

Final Prediction

$$O = \sum_{i=1}^t \alpha_i \cdot h_i \quad (8)$$

The activation function is represented by β_i α_i , the feature's relevance is represented by the attention weight is represented by O , and the output prediction result is represented by W_i , b_i , and the weight matrix and bias vector between neuron nodes σ , respectively.

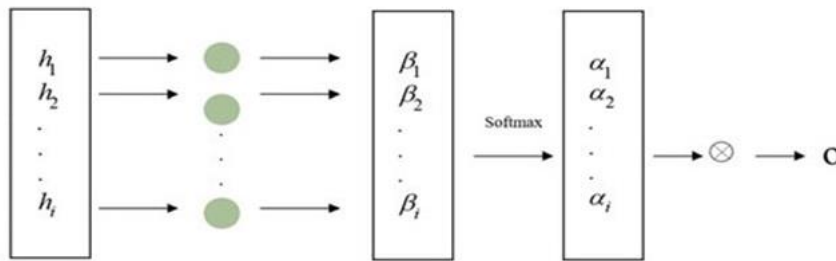


Figure 5: Attention network structure diagram

2.6 Performance Metrics

- Accuracy: Using accuracy is the quickest and easiest approach to see the frequency that the classifier gets the predictions right. Alternative interpretations include dividing the total number of forecasts by the proportion of accurately predicted positive events.

$$Accuracy = \frac{TP+TN}{S} \quad (9)$$

- Precision: On the other hand, this ratio shows the percentage of false negatives and subtracts one from it, which is known as (1 precision).

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

- Recall: On other hand there are called false negatives in relation with True Negatives.

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

- F1-Score: It is obtained through taking the harmonic mean between recall and precision scores.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

RESULTS AND DISCUSSION

3.1 Confusion Matrix

The confusion matrix in Figure 6 depicts the classification performance of the attention-enhanced AlexNet predictive model for eight waste types, consisting of food waste, leaf waste, paper waste, wood waste, e-waste, metal cans, plastic bags, and plastic bottles. Accurate classifications were evident in all types with more than 2700 correct classifications by the model and no large misclassifications. The model performed with the highest accuracy for plastic bottles (2900 correct) and food waste (2812 correct); the attention mechanism was able to emphasize important visual features and therefore create less classification errors and was steadily able to identify the proper waste category and provide reliable recognition across a number of waste types.

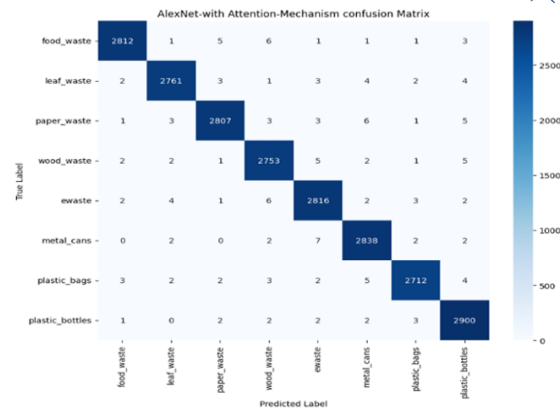


Figure 6: AlexNet with attention Mechanism

The confusion matrix in Figure 7 presents the performance of the standard AlexNet model in classifying eight waste categories: food waste, leaf waste, paper waste, wood waste, e-waste, metal cans, plastic bags, and plastic bottles. While the model demonstrates high accuracy overall, correct predictions per class range from around 2611 (metal cans) to 2712 (food waste). Misclassifications are comparatively higher than the attention-enhanced version, with noticeable confusion between visually similar categories such as plastic bags and plastic bottles. This indicates that, without attention mechanisms, AlexNet has slightly reduced discriminative ability for complex or visually overlapping waste types.

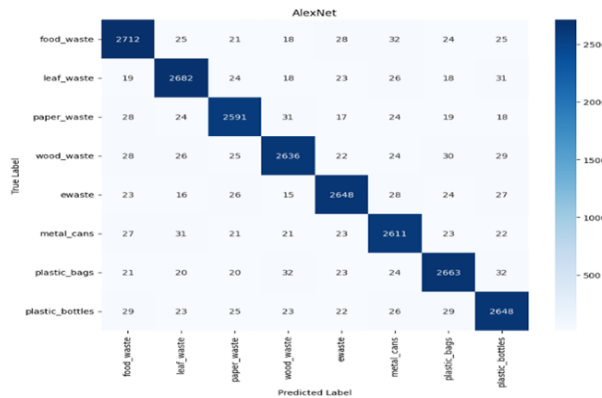


Figure 7: AlexNet model

The confusion matrix in Figure 8 illustrates the classification performance of a conventional CNN model across eight waste categories: food waste, leaf waste, paper waste, wood waste, e-waste, metal cans, plastic bags, and plastic bottles. Correct predictions per class range from 2570 (e-waste) to 2727 (plastic bottles), with moderate misclassifications observed, particularly between similar visual classes such as plastic bags and plastic bottles, and between paper waste and wood waste. Compared to enhanced models, the CNN exhibits slightly lower precision and more cross-category confusion, indicating limitations in distinguishing visually overlapping waste types without advanced feature attention mechanisms.

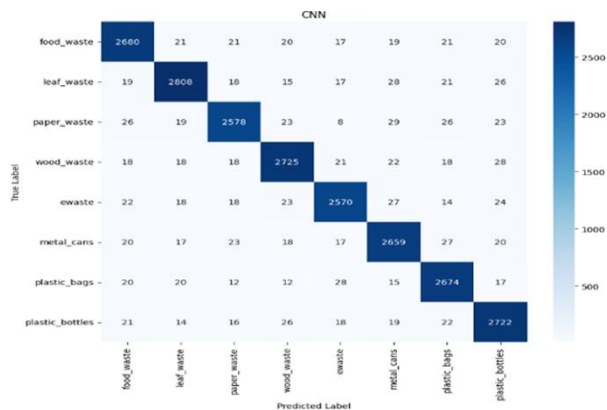


Figure 8: CNN Model

The comparative study of the three confusion matrices shows that the AlexNet model with attention performed best with all waste categories and had the best performance metrics and the lowest misclassification rates. This model saw the

How to cite: Shivran Priyanka Suresh Kumar and Dr. Monika Saxena. Advanced Deep Learning Framework for Waste Image Categorization Using Attention-Enhanced AlexNet. *Advances in Consumer Research*. 2025;2(5):1098–1109.

highest count of correctly classified instances within most waste categories, such as 2812 cases of "food waste" and 2877 cases of "plastic bottles," which highlighted the enhanced feature discrimination that was the result of inclusion of the attention layer in AlexNet. The AlexNet model without attention was second with 2733 less than correct classifications (eg: 2712 "food waste" and 2641 "metal cans" and higher misclassification rates in false classifications such as "plastic bags" and "wood waste"). The CNN model saw the lowest performance metrics with less correctly classified instances (eg: 2680 "food waste" and 2608 "leaf waste") with the greatest number of errors within false classifications "paper waste," and "plastic bags." Overall, this demonstrated that the inclusion of an attention layer into AlexNet significantly improved the ability to distinguish between waste categories and classify at a higher degree of accuracy than both the AlexNet and CNN models.

3.2 6.2 ROC Curves

There is a comparison of three machine learning models: a CNN, a regular AlexNet, and an AlexNet improved with an attention mechanism, utilizing "Receiver Operating Characteristic (ROC)" curves. The curves assist in illustrating how well each model can accurately distinguish and segregate various categories of waste, e.g., food waste, plastic cans, and others. For the example of the attention-based AlexNet, Figure 9, the ROC curves demonstrate extremely good classification performance, with "Area under the Curve (AUC)" scores of 0.89 to 0.99 for varying categories of waste. Such high scores indicate that the model is able to confidently separate the various classes of waste, with little or no overlap and confusion among the classes.

The baseline AlexNet model, illustrated in Figure 10, is good but somewhat less so than its attention version. Its AUC values range from 0.86 to 0.98, indicating that although the model is still robust at classification, it is a little less stable over all categories than the attention version. Lastly, the CNN model, as depicted in Figure 15, has the worst overall accuracy among the three and an AUC of 0.88 to 0.96. While these are still relatively good values, they suggest that the CNN finds it harder to distinguish between certain waste types than the AlexNet-based methods do.

In each ROC curve, a diagonal dashed line indicates the performance of random guessing (AUC = 0.5). The fact that all three models' curves are well above the line ensures that they are all significantly superior to chance. Nevertheless, the outcomes explicitly indicate that the AlexNet model with the attention mechanism performs best in general, yielding higher AUC values and reflecting better performance in multi-class waste classification.

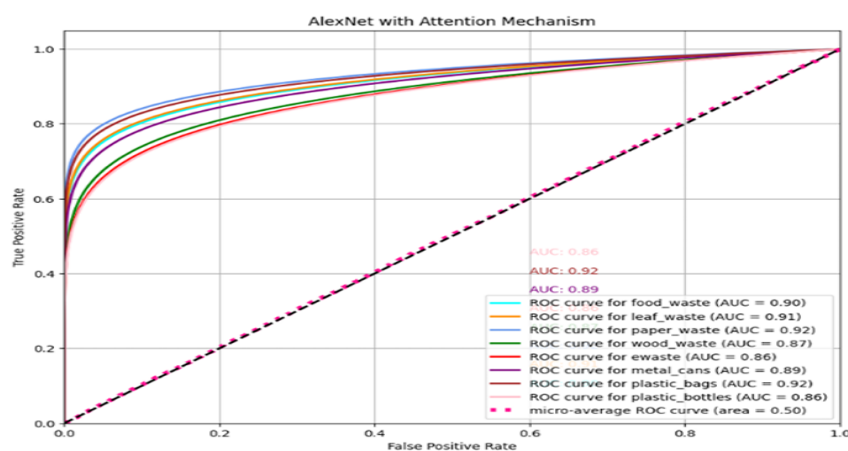


Figure 9: AlexNet with attention Mechanism

The ROC curve displays the AlexNet model's performance of attention mechanism regarding classes of waste materials with the False Positive Rate plotted on the x-axis and the True Positive Rate on the y-axis. The coloured curves represent each waste type with an AUC (Area Under the Curve) for classification accuracy, where the model produces a strong class-specific performance across all categories. Paper waste and plastic bags had the strongest performance with an AUC of 0.92, with leaf waste at 0.91 and food waste at 0.90 following closely behind. Metal cans displayed good performance at 0.89, and wood waste had a performance of 0.87. E-waste and plastic bottles produced the lowest acceptable performance, though still strong, at 0.86. The dashed line represents random guessing (AUC = 0.50), and all category curves lay solidly above the line inferring the model's performance is well above random guessing. Most of the curves also lay closely to the top-left corner of the plot indicating high sensitivity and low false positive rate. This briefly communicates the effectiveness of the attention-enhanced AlexNet for classifying paper waste and plastic bags.

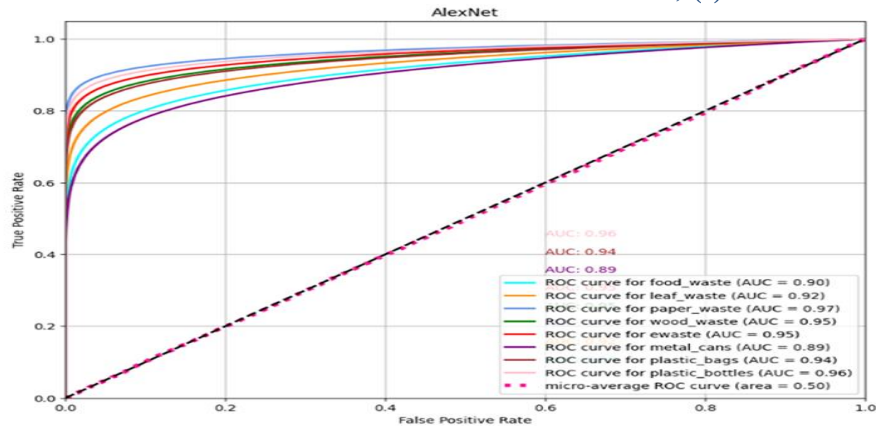


Figure 10: AlexNet Model

The figure 10 of the ROC curve shows the AlexNet model's classification performance across all waste categories, with the False Positive Rate on the x-axis and the True Positive Rate on the y-axis. Each coloured curve is associated with a specific waste type, and the AUC value (Area Under the Curve) reflects the model's ability to differentiate between classes. Paper waste has the strongest accuracy at an AUC of 0.97, while plastic bottles were at 0.96, wood waste, and e-waste had AUC values of 0.95, followed by plastic bags at 0.94, where leaf waste and food waste were at AUC values of 0.92 and 0.90 respectively. Metal cans and food waste share the lowest AUC at 0.89 but are still well above random guessing (which is represented by the dashed diagonal line where AUC = 0.50). Overall, it should be noted that most curves are close to the top-left corner, indicating high sensitivity and low false positive rates, so the AlexNet model did well on classification overall with good performance on paper waste and plastic bottles.

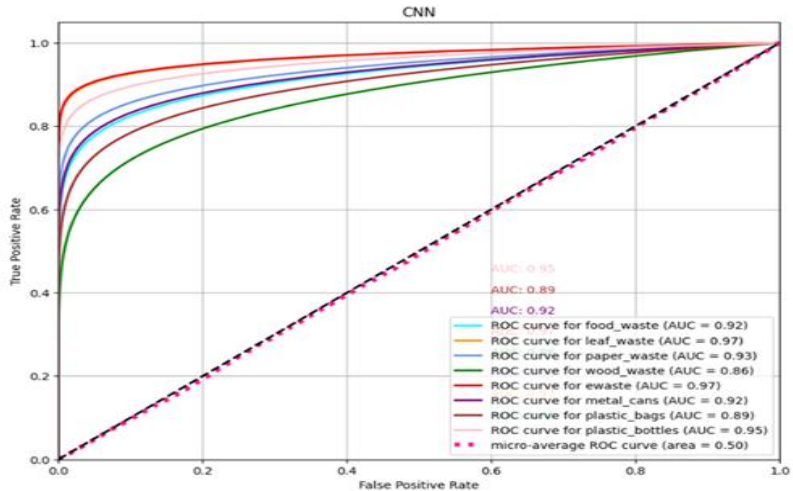


Figure 11: CNN Model

The ROC curve figure 11 shows the performance of the CNN model for classifying the various categories of waste, with the False Positive Rate plotted on the x-axis and the True Positive Rate on the y-axis. Each curve represents a particular waste type, and the values of AUC (Area Under the Curve) demonstrate the discriminatory capacities of the model. Leaf waste and e-waste possess the highest performance of 0.97, then plastic bottles (0.95) and paper (0.93); food waste and metal can both perform with an AUC of 0.92, and plastic bags and wood waste are last at 0.89; 0.86. The dashed diagonal line (AUC = 0.50), which represents random classification, at least confirms that all of the curves are well above it, showing the CNN model classified performance better than random guessing. Most curves bend toward the top-left corner, indicating that the model has good sensitivity and specificity, with the highest degree of accuracy in classifying leaf waste, e-waste, and plastic bottles.

3.3 6.3 Performance Metrics

Figure 12 compares three deep learning models, AlexNet with Attention Mechanism, AlexNet, and a Convolutional Neural Network (CNN), according to four key measures of performance: accuracy, precision, recall, and F1 score. As shown in the bar chart, AlexNet with Attention Mechanism performed best across all four measures, getting as close to 1.0 as possible (meaning the classification results are just about perfect). The substantially higher accuracy means that the model is wrong on very few predictions, while better precision means that the rate of correct positive predictions is great. This means that the attention-enhanced AlexNet did an excellent job of attending to the relevant features when making classifications. Recall values indicate that this model also made correct identifications of a large percentage of actual

positive cases. In other words, there is a very small chance of missing positive identifications. Finally, the F1 score also performed best for this model and indicates a balance between precision and recall with good performance in terms of reliability and completeness of predictions.

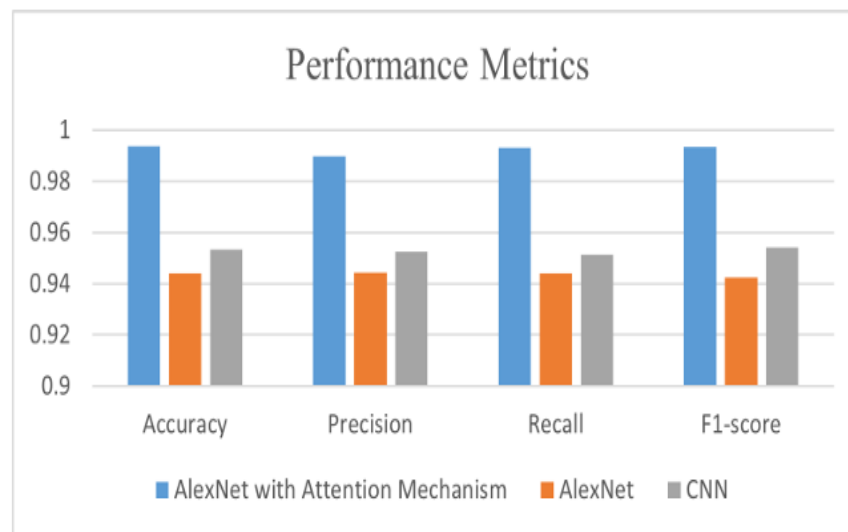


Figure 12: Performance Metrics

In comparison, the normal AlexNet model shows the least performance across all metrics, which indicates that the model struggles to represent and prioritize the important features needed to classify objects by not using an attention mechanism. Additionally, the lower precision and recall indicate that the model has a higher likelihood of many false positives, as well as missed detections, which affects its F1-score. While the CNN model does perform better than normal AlexNet, the attention-enhanced AlexNet still outperformed CNN in all metrics. Effectively, although CNN can achieve good representations for classification, it does not provide the focus achieved by using an attention mechanism. Overall, the results of this chart strongly demonstrate that using an attention mechanism within AlexNet not only enhances its accuracy but allows the model to detect relevant features more reliably and efficiently, leading to greater and consistent effectiveness across all evaluation metrics.

DISCUSSIONS

The comparison of the normal AlexNet, attention-mechanism-augmented AlexNet, and normal CNN's classification performance for multi-class garbage sorting demonstrates evident differences in efficacy. Results from the confusion matrices indicate that the attention-augmented AlexNet frequently performs better than the other two models by receiving notably higher correct classification rates for difficult classes of waste. For instance, it correctly detects 2,812 of food waste and 2,877 of plastic bottles, exemplifying its superior capacity to detect specific categories with high accuracy. This enhanced performance is due to the attention mechanism that enables the model to identify the most important features of the input images, hence being able to differentiate more accurately between categories with subtle visual distinctions. Conversely, the baseline AlexNet and CNN models log a significantly higher count of misclassifications, especially for visually redundant types of waste like plastic bag waste and wood wastes, which indicates their inadequacies when performing tricky tasks of classification where fine-grained discrimination of features is important.

These findings are also supported by ROC curve analysis, which tests model performance over a variety of classification thresholds. The attention-augmented AlexNet shows outstanding performance with Area Under the Curve (AUC) values from 0.89 to 0.99 for

different categories, showing a good and consistent capacity to distinguish between distinct types of waste. Conversely, the AUC values for the baseline AlexNet and CNN are relatively lower, indicating that their classification performance is less stable, especially when dealing with borderline cases in which classes possess overlapping characteristics. The elevated AUC values for the attention model emphasize its stability, as it still exhibits strong predictive capacity even when the decision boundary is shifted, which is important in real-world scenarios where data distributions might differ.

Aside from confusion matrix and ROC, other performance measures like recall, precision, F1-score, and overall accuracy give more evidence of how the attention mechanism improves. The AlexNet with attention posts an impressive 99.36% accuracy, along with precision and recall that are always high, showing that not only is it minimizing false positives, but it also picks up on almost all instances in each category. This trade-off between precision and recall results in a very high F1-score, indicating the well-balanced performance of the model. In comparison, the simple AlexNet and CNN models do not succeed in balancing this trade-off, usually losing one metric at the expense of the other. These results conclusively show that the inclusion of attention mechanisms within neural network models has the potential to drastically improve performance in challenging multi-class classification tasks. According

to this, future work can involve integrating attention modules with other deep learning networks to enhance accuracy, reliability, and adaptability across a wide range of application areas, from waste sorting to medical imaging and more.

CONCLUSIONS

The work well proves that adding an attention mechanism to AlexNet's architecture improves its capability to classify waste into numerous categories at high accuracy. The comparison indicates that the attention-improved AlexNet not only performs better in terms of classification accuracy but also performs well in minimizing misclassifications between visually confusing waste classes, which is a primary issue in such tasks. The enhancements are seen in confusion matrices where the accurate predictions are significantly higher; in ROC curves that show better class separation; and in performance measures like precision, recall, and F1-score, all of which show greater predictive ability. The attention mechanism functions by directing the network to pay attention to the most significant features in the input data, allowing it to draw more accurate distinctions even when class resemblance is pronounced. With a staggering accuracy of 99.36%, the attention-augmented AlexNet performed better than both the regular AlexNet and the standard CNN, which had lower accuracy and misclassification rates.

These results highlight the value in using more sophisticated methods such as attention mechanisms to mitigate the challenges of multi-class classification, when conventional convolution models can fall behind. Through allowing the network to selectively focus on salient areas of an image, attention mechanisms offer an effective means by which the classification results can be enhanced in difficult cases. The success of this method in trash categorization indicates great promise for more general applications across domains where precise classification is essential, including medical imaging, remote sensing, and industrial quality assurance. Future studies should continue to study the integration of attention modules into various deep learning architectures and evaluate their generalizability across various datasets and domains. Such endeavors might produce even more impressive developments in machine learning, bringing forth models that are stronger, more precise, and can tackle progressively advanced classification tasks.

REFERENCES

- [1] L. Ceballos-Pinto and F. Martinez-Jeronimo, "Mass production of *Scenedesmus incrassatulus* in 8- and 40-liter disposable polyethylene bags with different culture media," *Rev. Latinoam. Microbiol.*, vol. 37, p. 109, 1995.
- [2] M. Malik et al., "Waste classification for sustainable development using image recognition with deep learning neural network models," *Sustainability*, vol. 14, no. 12, 2022. doi: 10.3390/su14127222.
- [3] M. S. Rad et al., "A computer vision system to localize and classify wastes on the streets," in *Computer Vision Systems: 11th International Conference, ICVS 2017, Shenzhen, China, July 10-13, 2017, Revised Selected Papers 11*, Springer, 2017, pp. 195–204.
- [4] P. Nowakowski and T. Pamuła, "Application of deep learning object classifier to improve e-waste collection planning," *Waste Manag.*, vol. 109, pp. 1–9, 2020.
- [5] W. C. Li, H. F. Tse, and L. Fok, "Plastic waste in the marine environment: A review of sources, occurrence and effects," *Sci. Total Environ.*, vol. 566, pp. 333–349, 2016.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. doi: 10.1038/nature14539.
- [7] S. Liang and Y. Gu, "A deep convolutional neural network to simultaneously localize and recognize waste types in images," *Waste Manag.*, vol. 126, pp. 247–257, 2021.
- [8] Y. Berhanu, E. Alemayehu, and D. Schröder, "Examining car accident prediction techniques and road traffic congestion: A comparative analysis of road safety and prevention of world challenges in low-income and high-income countries," *J. Adv. Transp.*, vol. 2023, 2023, doi: 10.1155/2023/6643412.
- [9] Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou, and
- [10] K. Q. Weinberger, Eds., Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b_Paper.pdf
- [11] Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [12] Huang Kai, Huan Lei, Zeyu Jiao, and Zhenyu Zhong. "Recycling waste classification using vision transformers on portable device." *Sustainability* 13, no. 21 (2021): 11572.
- [13] Islam, Niful, Md Mehedi Hasan Jony, Emam Hasan, Sunny Sutradhar, Atikur Rahman, and Md Motaharul Islam. "Ewastenet: A two-stream data efficient image transformer approach for e-waste classification." In *2023 IEEE 8th International Conference on Software Engineering and Computer Systems (ICSECS)*, pp. 435–440. IEEE, 2023.
- [14] Nafiz, Md Shahariar, Shuvra Smaran Das, Md Kishor Morol, Abdullah Al Juabir, and Dip Nandi. "Convowaste: An automatic waste segregation machine using deep learning." In *2023 3rd International conference on robotics, electrical and signal processing techniques (ICREST)*, pp. 181–186. IEEE, 2023.

- [15] Chhabra, Megha, Bhagwati Sharan, May Elbarachi, and Manoj Kumar. "Intelligent waste classification approach based on improved multi-layered convolutional neural network." *Multimedia Tools and Applications* 83, no. 36 (2024): 84095-84120.
- [16] Wang, Zhaoqi, Wenxue Zhou, and Yanmei Li. "GFN: a garbage classification fusion network incorporating multiple attention mechanisms." *Electronics* 14, no. 1 (2024): 75.
- [17] Shrivastava, Akshat Kishore, and Tapan Kumar Gandhi. "Integrating Human Vision Perception in Vision Transformers for Classifying Waste Items." In *International Conference on Artificial Intelligence and its Application*, pp. 425-438. Singapore: Springer Nature Singapore, 2024.
- [18] Wang, Jianfei. "Application research of image classification algorithm based on deep learning in household garbage sorting." *Heliyon* 10, no. 9 (2024).
- [19] Qiu, Wenxuan, Chenxin Xie, and Jingui Huang. "An improved EfficientNetV2 for garbage classification." In *International Conference on Intelligent Computing*, pp. 79-90. Singapore: Springer Nature Singapore, 2025.
- [20] Jose, Jithina, Suja Cherukullapurath Mana, Keerthi Samhitha Babu, G. Kalaiarasi, and M. Selvi. "Enhancing waste classification accuracy with Channel and Spatial Attention-Based Multiblock Convolutional Network." *Environmental Monitoring and Assessment* 197, no. 2 (2025): 198.
- [21] Nahiduzzaman, Md, Md Faysal Ahamed, Mansura Naznine, Md Jawadul Karim, Hafsa Binte Kibria, Mohamed Arselene Ayari, Amith Khandakar, Azad Ashraf, Mominul Ahsan, and Julfikar Haider. "An automated waste classification system using deep learning techniques: Toward efficient waste recycling and environmental sustainability." *Knowledge-Based Systems* 310 (2025): 113028.
- [22] Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, and X. Liu, "Waste image classification based on transfer learning and convolutional neural network," *Waste Manag.*, vol. 135, pp. 150–157, 2021.
- [23] Q. Zhang et al., "Recyclable waste image recognition based on deep learning," *Resour. Conserv. Recycl.*, vol. 171, p. 105636, 2021.
- [24] H. Abdu and M. H. M. Noor, "A survey on waste detection and classification using deep learning," *IEEE Access*, vol. 10, pp. 128151–128165, 2022.
- [25] S. Su"ru"cu and 'I. N. Ecemis,, "Classification of urban waste materials with deep learning architectures," *SN Comput. Sci.*, vol. 4, no. 3, p. 285, 2023.
- [26] N. Li and Y. Chen, "Municipal solid waste classification and real-time detection using deep learning methods," *Urban Clim.*, vol. 49, p. 101462, 2023.
- [27] H. Zhang, H. Cao, Y. Zhou, C. Gu, and D. Li, "Hybrid deep learning model for accurate classification of solid waste in the society," *Urban Clim.*, vol. 49, p. 101485, 2023.
- [28] Zhu, Ling, Zhenbo Li, Chen Li, Jing Wu, and Jun Yue. "High performance vegetable classification from images based on AlexNet deep learning model." *国际农业与生物工程学报* 11, no. 4 (2018): 217-223.
- [29] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using AlexNet convolutional neural network," *Appl. Sci.*, vol. 9, no. 3, p. 483, 2019.
- [30] Singh, G. Goyal, and A. Chandel, "AlexNet architecture based convolutional neural network for toxic comments classification," *J. King Saud Univ. Inf. Sci.*, vol. 34, no. 9, pp. 7547–7558, 2022.
- [31] Kumar M and A. Chakrapani, "Classification of ECG signal using FFT based improved Alexnet classifier," *PLoS One*, vol. 17, no. 9, p. e0274225, 2022.
- [32] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021. doi: <https://doi.org/10.1016/j.neucom.2021.03.091>.
- [33] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Attention branch network: Learning of attention mechanism for visual explanation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10705–10714.
- [34] M.-H. Guo et al., "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022.
- [35] Mohiuddin et al., "Retention is all you need," in *Int. Conf. Inf. Knowl. Manag. Proc.*, no. Nips, pp. 4752–4758, 2023. doi: 10.1145/3583780.3615497.
- [36] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, 2019. doi: 10.1186/s40537-019-0197-0.
- [37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, 2015.
- [38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.
- [39] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. doi: 10.1109/5.726791.
- [40] L. Eleftheriadou, "Highway Networks," pp. 319–327, 2024. doi: 10.1007/978-3-031-54030-1_14.
- [41] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text data augmentation for deep learning," *J.*

How to cite: Shivran Priyanka Suresh Kumar and Dr. Monika Saxena. Advanced Deep Learning Framework for Waste Image Categorization Using Attention-Enhanced AlexNet. *Advances in Consumer Research*. 2025;2(5):1098–1109.

Big Data, vol. 8, no. 1, p. 101, 2021. doi:
10.1186/s40537-021-00492-0.