Neural Computing and Applications (2025) 37:4567–4583
https://doi.org/10.1007/s00521-024-10855-2

ORIGINAL ARTICLE

Enhancing waste sorting and recycling efficiency: robust deep learning-based approach for classification and detection

Faizul Rakib Sayem[1] · Md. Sakib Bin Islam[2] · Mansura Naznine[3] · Mohammad Nashbat[4] · Mazhar Hasan-Zia[4] · Ali K Ansaruddin Kunju[4] · Amith Khandakar[5] · Azad Ashraf[4] · Molla Ehsanul Majid[6] · Saad Bin Abul Kashem[7] · Muhammad E. H. Chowdhury[5]

Received: 22 August 2024 / Accepted: 29 November 2024 / Published online: 23 December 2024

Abstract

Given the severity of waste pollution as a major environmental concern, intelligent and sustainable waste management is becoming increasingly crucial in both developed and developing countries. The material composition and volume of urban solid waste are key considerations in processing, managing, and utilizing city waste. Deep learning technologies have emerged as viable solutions to address waste management issues by reducing labor costs and automating complex tasks. However, the limited number of trash image categories and the inadequacy of existing datasets have constrained the proper evaluation of machine learning model performance across a large number of waste classes. In this paper, we present robust waste image classification and object detection studies using deep learning models, utilizing 28 distinct recyclable categories of waste images comprising a total of 10,406 images. For the waste classification task, we proposed a novel dual-stream network that outperformed several state-of-the-art models, achieving an overall classification accuracy of 83.11%. Additionally, we introduced the GELAN-E (generalized efficient layer aggregation network) model for waste object detection tasks, obtaining a mean average precision (mAP50) of 63%, surpassing other state-of-the-art detection models. These advancements demonstrate significant progress in the field of intelligent waste management, paving the way for more efficient and effective solutions.

Keywords Intelligent waste management · Sustainable waste processing · Urban solid waste · Deep learning for waste management · Waste image classification · Object detection in waste management

1 Introduction

The global increase in urbanization and consumption has put unprecedented strain on waste management systems, resulting in environmental degradation, resource depletion, and public health issues. Efficient waste management and sorting are crucial in today's world as growing populations and consumer preferences continue to challenge the resilience of our ecosystems [16]. Traditional trash disposal procedures, particularly in low-resource settings, are unable to handle the growing volume of solid garbage. Manual sorting is labor-intensive, prone to human error, and inefficient, limiting the widespread adoption of sustainable techniques. Proper waste disposal reduces environmental degradation, conserves resources, and minimizes pollution. Sorting garbage from its origin is essential because it allows for the separation of recyclable items, organic waste, and non-recyclable materials, allowing it easier to properly handle and dispose of them. To be able to effectively manage waste management, people and communities need to implement an extensive strategy that encompasses three key components: reduction, reuse, and recycling [15]. As consequently, solid waste management (SWM) is producing significant challenges, leading to the depletion of air, land, and water quality, which has detrimental impacts on natural ecosystems and the well-being of society [33]. Sustainable solid waste management (SSWM) (T. D. [8],T.-D. [9, 36]) is an innovative approach to processing solid waste that aims to enhance operational quality and achieve the objectives of reducing, reusing, and recycling garbage.

The most significant concern associated with the accumulation of garbage is a reduction in innocuous organic waste and a rise in chemically reactive substances inside

Extended author information available on the last page of the article the waste. Plastic waste has significantly altered the situation due to its non-biodegradable nature [14]. While it is feasible to recycle it, there is currently no sufficient infrastructure in place for its storage. Robotic conveyor lines are often used for this specific objective. These robots are equipped with industrial manipulators and video cameras capable of identifying and sorting specific types of waste [17, 31]. Recycling plays a pivotal role in this transformation, enabling innovative research and promoting eco-friendly business practices. Despite these advances in technology, there continues to be a critical problem that requires urgent action—the requirement for precise separation of waste based on the capacity to decompose naturally. The implementation of efficient waste categorization encounters significant obstacles, including limited public knowledge and variations in waste classification standards across different regions [13]. Dependence only on manual tree sorting amplifies these challenges, resulting in elevated labor expenses and inefficiencies, as well as posing hazards to the health of people. For tackling these issues, advanced technologies such as Artificial Intelligence (AI, deep learning (DL and machine learning (ML might be used (Abdallah et al., 2020a [12, 28, 35],). These advancements have the potential to completely transform trash management by enhancing the precision of sorting, increasing economic efficiency, and fostering environmental sustainability. Existing deep learning-based solutions, while promising, frequently necessitate large computational resources and human feature engineering, limiting their practical utility.

Artificial intelligence, a rapidly advancing technology, is currently being embraced across numerous sectors, with waste management being an especially significant area of emphasis (Abdallah et al., 2020b). ML algorithms have the ability to acquire knowledge from datasets that are categorized and consist of images depicting different types of waste products [19]. These algorithms then analyze the data to detect patterns and patterns, allowing them to categorize novel waste items that have not been previously seen. However, conventional machine learning methods sometimes include substantial individuals feature engineering, a process that may be both time-consuming and specialized to a particular subject. Deep neural networks possess a hierarchical structure that enables them to automatically extract characteristics from raw picture input [24]. Convolutional Neural Networks (CNNs) excel at performing this task, making them highly efficient in applications involving computer vision such as waste sorting and detection.

The principal impetus for this research is the escalating global issue of waste management, propelled by swift urbanization, rising trash production, and the necessity for more efficient, scalable, and automated waste sorting systems. Conventional manual sorting techniques are laborious and susceptible to errors, demanding the urgent use of AI-driven automation in this field. This paper introduces an innovative AI-driven trash classification system that addresses the shortcomings of conventional and current machine learning methods. The system employs Convolutional Neural Networks (CNNs) to autonomously learn features from raw picture inputs, therefore obviating the necessity for manual feature extraction. The aim is to create a precise and scalable model that can categorize various waste categories, including recyclable and non-recyclable materials, without the need for manual intervention often involved in sorting procedures. Our technology enhances prior research by amalgamating diverse cutting-edge techniques, hence augmenting classification accuracy and efficiency in both intricate and straightforward waste streams. Our suggested approach attains enhanced automation and precision, creating the way for future incorporation into practical waste management frameworks, especially in areas confronting significant waste management difficulties. The scarcity of data in waste detection systems, particularly in recycling waste materials, presents a significant obstacle in training and deploying deep learning models effectively. A lack of large and standardized datasets for waste detection restricts the development of precise algorithms for waste recognition. In this article, we have used a large dataset to develop robust waste classification and detection models for efficient waste sorting and recycling. The main contributions of our paper are shown below:

i) A total of 28 categories of waste images with reliable performance for both classification and detection tasks are reported in this study.
ii) A novel dual-stream network is proposed, which comprises of DenseNet-201 and multi-axis vision transformer model. By concatenating the feature space from both streams, a superior output performance compared to other studies and state-of-the-art models is achieved.
iii) Several object detection models were utilized for waste detection tasks and achieved the best performance using GELAN-E model outperforming the state-of-the-art models.

The paper is organized into several essential sections to deliver a comprehensive summary of the research. Section 2 provides a review of pertinent literature, correlating prior studies with the aims of our research. Section 3 explores the dataset and methodology, providing a thorough elucidation of the tiered implementation procedure. Section 4 contains a comprehensive analysis and discussion of the research findings. Section 5 concludes by summarizing the principal findings and emphasizing the significant discoveries of the inquiry.

2 Related works

Automated garbage sorting is a key element of effective recycling procedures. Computer vision methods have shown great promise in enhancing the precision and effectiveness of garbage sorting systems in recent years. Multiple investigations explored the use of deep learning algorithms for the purpose of trash categorization utilizing datasets consisting of images. The WaRP (Waste Recycling Plant) dataset addresses this need by offering an assortment of tagged images taken in an actual waste recycling facility setting. The WaRP dataset is a great resource for academics working on computer vision algorithms for automated trash sorting tasks due to its diverse waste categories and the inclusion of occlusions, deformations, and varied lighting circumstances.

Ogrezeanu et al., [22] employed the WaRP-C dataset, comprising segmented picture regions from the WaRP-D collection, annotated with class labels. Numerous deep learning architectures were examined, including CNN, LeNet5, AlexNet, VGG16, Mobile-Net-v2, Inception, and DenseNet, to evaluate their efficacy in accurately classifying trash items. The study employed the YOLO v8 architecture for waste detection, recognized for its efficacy in recognizing items inside chaotic settings. [25] introduces a dual-phase deep learning system aimed at improving waste identification and categorization, tackling the limitations of conventional approaches. The proposed framework has two primary modules: an object detection module and an object categorization module. The object detection module utilizes a three-tier design comprising a Backbone, Neck, and Head to proficiently identify trash things in photos. Riyadi et al., [29] emphasizes the notable improvements in detection accuracy and speed attained with YOLOv8 in comparison with its predecessor, YOLOv5, recognized for its multi-scale detection capability. The findings demonstrate that YOLOv8 surpasses YOLOv5 in detection speed and resolution, with precision and recall metrics of 0.478 and 0.569 for YOLOv5, and 0.442 and 0.513 for YOLOv8, respectively. Deep learning methods, as utilized in [4, 5, 21, 23], involve the adaption of pre-existing pre-trained models [23] to optimize trash categorization. Moreover, a comprehensive evaluation of several classification techniques, such as Support Vector Machines (SVMs) and SoftMax, implemented in conjunction with the CNN model, has been performed. The TrashNet dataset, as examined in [32], is used in a model for garbage image categorization. A more effective detection strategy has been developed to classify garbage using a multilayer hybrid CNN. This method has a simpler structure and fewer parameters, resulting in an impressive accuracy of up to 92%. Another instance by [20] is using Efficient Net classification models to analyze waste picture images obtained from the ImageNet dataset. [18] used a deep learning algorithm to implement a novel technique for classifying garbage, with the aim of improving its efficient utilization and disposal.

Hybrid techniques that include CNNs and multilayer perceptron's have been developed. These systems use data from other sensors, as described. [11] introduced a multilayer hybrid DL system that has the ability to autonomously categorize rubbish discarded by persons in metropolitan public spaces. The MLP approach was used to combine picture characteristics and additional feature information, resulting in excellent classification performance. This enhances the quality; however, it may not always be technically viable in actual implementation. A significant advancement in trash identification was accomplished by [3]. An inherent drawback of this task is that the used dataset included photos with a consistent backdrop. This occurrence rarely occurs inside the industrial setting of recycling facilities. Regarding object segmentation, several supervised real-time segmentation models exist, such as convolutional models like DeepLabv3 + (F. [38]) or transformer models like SegFormer [40]. In the research, [6] examined several neural network techniques for the segmentation of deformable objects in complex settings. The research presents in [27], provides a comprehensive assessment of the use of artificial intelligence (AI) technology in the management of waste classification. The main focus is on the actual implementation and pertinent research outcomes of AI in the procedures of waste disposal, collection, and categorization. The objective of this study is to assess the practicality of incorporating artificial intelligence (AI) technology into the waste categorization management system at H University in China [27]. According to (Abdallah et al., 2020c), frequently employed AI models for garbage categorization include Artificial Neural Network, Support Vector Machine, Linear Regression, Decision Trees, and Genetic Algorithm. Regarding the AI products that have been introduced in the area of waste classification recycling. [43] developed an enhanced YOLO model to automate the detection of garbage, with a focus on addressing river pollution caused by urbanization. Their method attained a remarkable 89% mean average accuracy (mAP) in identifying floating trash, such as plastic bottles and aluminum cans, even under difficult circumstances such as fluctuating illumination and different backdrops. [7] investigated the efficacy of YOLOv8, the most recent version of YOLO object identification models, in automating trash sorting. The objective was to improve the effectiveness and safety of waste treatment procedures. Their research showed the superior performance of YOLOv8 compared to other algorithms in precisely detecting and classifying rubbish. This indicates the potential of YOLOv8 as a beneficial tool for enhancing waste management techniques in the face of urbanization and industrialization concerns. [26] developed a thorough strategy using sophisticated machine learning methods to tackle the evaluation of water quality and the identification of undersea garbage. Their approach used the YOLOv8 model to identify garbage, a rule-based classifier to evaluate water quality, and XGBoost to predict water potability. The results demonstrated high accuracy and recall rates for trash detection and accurate sample categorization. This versatile approach shows potential for achieving substantial improvements in environmental preservation and protecting aquatic ecosystems from the detrimental effects of water pollution. Yudin et al., [41] expands the idea of weakly supervised learning, which relies on image-level labels for training data, to identify garbage in recycling operations. The system utilizes the Mask R-CNN architecture, a widely used model for instance segmentation, to segregate trash in pictures. The hierarchical waste identification algorithm demonstrated a robust performance in precisely distinguishing trash items in recycling plant photos, achieving an average Intersection over Union (IoU) score of 0.65 in waste segmentation tasks. The model exhibited a balanced performance in properly recognizing trash items while limiting false positives, as seen by its accuracy and recall values of 0.78 and 0.72, respectively.

3 Methods and materials

This section initiates with an extensive examination of the Warp dataset, including its preprocessing stages and the procedures utilized for classification and detection in the context of an industrial waste sorting facility. The subsequent sections provide detailed descriptions of the deep machine learning approaches utilized in this research, providing additional information regarding their intricacies and methodologies. Furthermore, the quantitative metrics and qualitative method utilized to evaluate the performance of each experiment and interpret the results, respectively, are described. Figure 1 illustrates the comprehensive workflow of our proposed methodology for the classification and detection of Warp datasets.

3.1 Dataset description

Automated trash recycling systems need the capability to automatically detect and classify recyclable items as they move down the conveyor belt. Within our context, the products we are referring to fall into many distinct main groups, including different plastic and glass bottles, cardboard, detergents, canisters, and cans. Efficient recycling of the first three categories requires understanding their color and function since recycling techniques vary. The WaRP dataset [41] consists of annotated photographs acquired from a waste recycling facility that specializes in the sorting of industrial garbage. The recycling program includes a total of 28 different types of recyclable waste. These are divided into specific categories: 17 types of plastic bottles (each identified by a class name starting with "bottle-"), three variations of glass bottles (prefixed with "glass-"), two types of cardboard, four categories of detergents, and containers and cans. The suffix "-full" denotes bottles that are filled with air and not compressed. A key feature of this dataset is that objects have the ability to overlap, undergo significant deformation, or be under unfavorable lighting circumstances. We have utilized the WaRP-D and WaRP-C sections of the dataset for the purpose of training and evaluating the accuracy of detection (WaRP-D) and classification (WaRP-C) tasks. The sample images in **Supplementary Fig. 1** display examples of occurrences from five super class categories of the WaRP Dataset.

3.2 Data Pre-processing

Efficient data preparation is crucial for improving images before using them to train machine learning models [10, 34]. Common methods include adjusting the size of images to preserve uniformity, standardizing pixel values to guarantee constant scaling, and converting images to grayscale to save processing requirements. Data augmentation is used to increase the variety or range of data; while, filtering is utilized to reduce or eliminate unwanted or irrelevant information or noise. The initial step may use histogram equalization to enhance contrast; while, cropping or padding is employed to adjust photos to fixed-size inputs. Data standardization entails calculating the arithmetic mean and standard deviation. Moreover, integrating picture classification or detection capabilities might enhance the preprocessing processes.

3.2.1 Train test split

The classification methodology used included the utilization of Trash pictures provided from the Warp Dataset. WaRP-C consists of picture regions extracted from the WaRP-D dataset, each labeled with corresponding class labels. This section has a total of 7,505 images for the purpose of training, along with an additional 1,550 images designated for testing. The images vary in dimensions, with widths ranging from 40 to 703 pixels and heights ranging

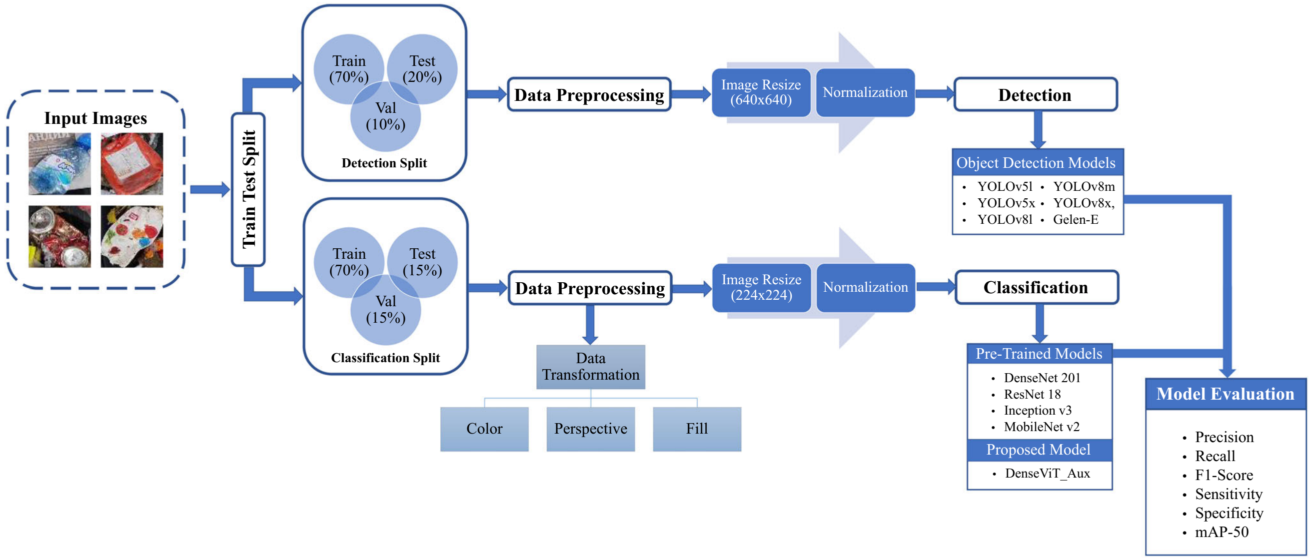


Fig. 1 Architecture and framework of the proposed approach

from 35 to 668 pixels. The dataset is imbalanced due to the actual circumstances of an industrial business. The dataset was divided into two distinct subsets: the training set, and the testing set. We used 20% of our training data as validation data to get the best training model based on the highest performance on the validation data. **Supplementary Table 1** shows the data distribution of 5 superclass across Train, Test and validation set.

3.2.2 Data augmentation

In order to improve the robustness and diversity of our training dataset, we implemented a range of image augmentation methods utilizing the TensorFlow and Keras libraries. In order to achieve this, the augmentation process utilized transformations including image infill, color variation, and perspective adjustments to produce up to 1500 augmented images per class which is displayed in **Supplementary Table 2**. To enhance the generalizability of the model, color variations consisted of arbitrary adjustments to hue, saturation, brightness, and contrast, which simulated various illumination conditions and color variations. Random horizontal and vertical inversions were incorporated into perspective alterations so that the model could identify objects irrespective of their orientation. As a substitute for prospective domain-specific methods to manage occlusions and incomplete objects, image infill was utilized. In order to ensure uniformity, every image was resized to 224 $\times$ 224 pixels. The 'ImageDataGenerator' class was employed to implement these modifications. To enhance organization and traceability, augmented images were prefixed with the name of the applied transformation.

3.2.3 Image resizing and normalization

With the objective of improving the efficacy and durability of our classification model, we executed an encompassing image preprocessing pipeline comprising resizing and data normalization. For consistency and to align with the input specifications of our model, every image in our dataset was resized to 224 $\times$ 224 pixels for classification and 640 $\times$ 640 for detection. Ensuring a consistent aspect ratio and facilitating efficient batch processing throughout the training process are both dependent on this resizing step. Normalization is a widely used image processing approach in the field of computer vision that aims to standardize the pixel values of images in a dataset. The mean and standard deviation (SD) are computed from each image. Every picture is transformed into the RGB format, and the pixel values are adjusted to ensure they fall within the range of 0 to 1. Initially, the mean and variance are calculated. The standard deviation is obtained by taking the square root of the variance. These are used to standardize the picture. Image normalization modifies the levels of pixel intensity. In order to alleviate the computational load, the pixel values were normalized from a range of 0 to 255 to a range of 0 to 1.

3.3 Proposed model architecture for classification

Our proposed model 2S_DenseViT (2 Stream Network with DenseNet and MaxViT) is an innovative dual-stream neural network structure specifically developed for classification purposes. It combines the useful characteristics of DenseNet-201 and Multi-axis Vision Transformer model

(MaxViT) [37] as feature extractors. The MaxViT model introduces a multi-axis attention approach that combines local and global interactions efficiently. It outperforms parallel models while using fewer parameters and computations. The proposed MaxViT block combines MBConv [30], block attention, and grid attention to perform both local and global spatial interactions effectively in a single block. By hierarchically stacking repeated blocks composed of multi-axis self-attention (Max-SA) and feed-forward networks (FFN), it obtained superior performance on various vision tasks. On the other hand, the DenseNet-201 model consists of multiple dense blocks, each comprising several layers where each layer is connected to every other layer within the block. The architecture of our proposed 2S_DenseViT model utilizes pre-trained DenseNet-201 and MaxViT models, initializing them with their respective weights to improve the ability to extract features from input images. The basic components of the '2S_DenseViT ' model consist of obtaining feature maps from the DenseNet-201 and MaxViT models. The DenseNet-201 model uses its feature layers and adds a MaxPool2d layer to decrease the spatial dimensions. Similarly, MaxViT utilizes its feature extraction layers in its model. The purpose of these extracted feature blocks is to gather a wide range of varied and detailed representations from the input data. To further analyze and manipulate these characteristics, two more classifiers are included. After extracting the features from DenseNet-201 model we used pooling operation with AdaptiveAvgPool2d, followed by a flattening operation and a linear layer that produces log probabilities in a lower dimensional space. Similarly, we performed the same operations after extracting features from the MaxViT model. The results from both auxiliary classifiers are joined together to create a unified feature representation. The feature vector, which contains information from both the DenseNet-201 and MaxVit models, is then processed into a final classifier. The final classification block consists of a dropout layer, which helps prevent overfitting, a flattening operation, and a linear layer that translates the combined feature vector to the number of target classes. The ultimate result is achieved by using a LogSoftmax Figure 2layer, which provides that the output probabilities are standardized and appropriate for classification purposes. We have employed the tiny version of the MaxViT model to reduce the training time and latency.

3.4 Object detection models

The YOLO series has significantly transformed the field of object identification by introducing innovative ideas in computer vision, such as the ability to analyze full pictures in a single run via a CNN. YOLOv5l is a modified version of the You Only Look Once (YOLO) object detection model created by Ultralytics [45]. Within the framework of YOLOv5, the letter "l" serves as an indicator of the model's dimensions, specifically denoting the big version. YOLOv5l achieves a harmonious combination of rapidity and precision, making it well-suited for a range of real-time object identification tasks. It attains this equilibrium by possessing a greater quantity of parameters in contrast to the smaller iterations of the model. YOLOv5l benefits from an enhanced parameter count, enabling it to identify objects with more precision. This makes it very valuable in situations where accuracy is of utmost importance, such as surveillance systems or autonomous cars. The YOLOv8 series, including YOLOv8l, YOLOv8m, and YOLOv8x, are names that have been suggested but have not been officially released or publicly recognized as models at the time of my latest update (T. [39]). In accordance with the established naming tradition of earlier YOLO iterations, it is feasible to assume that these theoretical models would have varying dimensions and intricacies. The letters "l," "m," and "x" are most likely abbreviations for large, medium, and extra-large, respectively. These abbreviations indicate differences in the size of the model and the number of parameters it has. Each version has the ability to provide a distinct balance between speed and precision, tailored to meet various application needs. YOLOv9 implements the Generalized ELAN (GELAN) framework, which aims to enhance parameters, computational complexity, accuracy, and inference speed [2]. GELAN improves the flexibility and efficiency of YOLOv9 by enabling users to choose suitable compute blocks for various inference devices. The GELAN is an innovative design that integrates the ideas of Cross-Stage Partial Network (CSPNet) and Efficient Layer Aggregation Network (ELAN) to optimize gradient route planning. The main focus is on creating a design that is lightweight, ensuring quick inference, and achieving high accuracy. GELAN enhances ELAN's layer aggregation capability by enabling the use of several processing blocks, hence assuring a high degree of flexibility. The design strives to provide effective feature aggregation while also retaining a high level of performance in terms of both speed and accuracy. The general architecture of GELAN (**Supplementary Fig. 2**) incorporates CSPNet's cross-stage partial connections with ELAN's efficient layer aggregation to provide successful gradient propagation and feature aggregation.

3.5 Experimental setup for classification

The current research has been carried out using the PyTorch framework for deep learning. The input images provided to the model consist of three distinct channels, specifically labeled as red, green, and blue (RGB). The

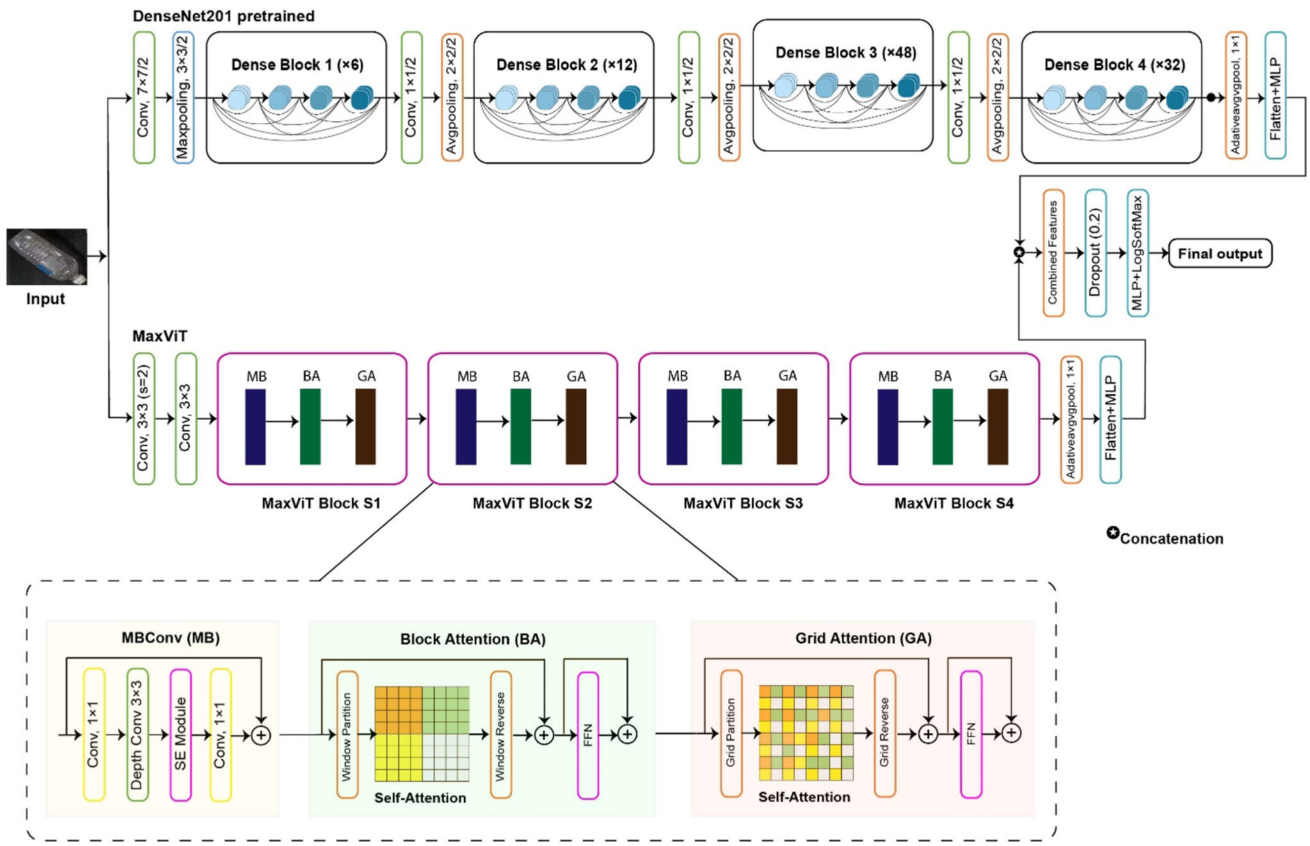


Fig. 2 Proposed Model Architecture of our MaxViT model

submitted photographs were scaled to a dimension of 224 by 224 pixels. The training group comprises eight persons, and the 2S_DenseViT model is trained for a total of fifty epochs; while, the Densenet-201 and Resnet18 models are trained for thirty epochs each. The Adam optimizer is used with a learning rate of 0.00001. The parameters used during the whole training procedure are shown in **Supplementary Table 3.**

The proposed 2S_DenseViT model, which combines DenseNet-201 and MaxViT architectures, has been thoroughly assessed for computational efficiency to determine its suitability for real-time trash classification systems. The model comprises 51,483,180 parameters, predominantly concentrated in the DenseNet backbone and MaxViT transformer blocks. Significantly, elements encompass a considerable share of the parameters, enhancing the model's ability to learn intricate features from input data. The model necessitates 10.06 GFLOPs per forward pass, largely because to the sophisticated vision transformer layers (grid attention and window attention) employed in the MaxViT architecture, together with the hierarchical feature extraction executed by DenseNet-201. The mean inference duration for the 2S_DenseViT model is 261.36 ms per image on a normal GPU, as ascertained from 100 iterations, underscoring its appropriateness for near-real-time applications, particularly in waste management systems where precision is critical. The model demonstrates a maximum memory utilization of 243.80 MB during inference, signifying an efficient memory footprint that permits deployment on systems with moderate GPU memory capacity while executing complicated classification tasks. These results demonstrate that the 2S_DenseViT model, despite its substantial parameter size and transformer construction, is a feasible choice for real-time waste categorization applications. Subsequent optimizations, including model reduction or quantization, could further improve computational efficiency while maintaining accuracy.

3.6 Experimental setup for detection

The validation method entails using the YOLOv9 model with GELAN-E architecture to identify objects in a specific test dataset outlined in the annotation file. The experiment is carried out with a batch size of 8 and an image size of 640 pixels (**Supplementary Table 4**). The confidence threshold is defined as 0.2; while, the IoU threshold is set at 0.5. These thresholds enable a maximum of 300 detections. The procedure does not use single-class mode or data augmentation, and it is not configured to be verbose. The findings are stored in text format along with confidence evaluations, but not in hybrid or JSON forms.

The GELAN-E model has remarkable performance with a batch size of 8, with an average inference time of 12.7 ms per image. This efficiency is enhanced by a very modest parameter count of 16.168 million, highlighting the model's capacity to provide swift forecasts while preserving reasonable complexity. The properties of the GELAN-E model render it an appropriate choice for applications necessitating rapid processing times, especially in real-time scenarios.

4 Result and discussion

This section presents the classification results for both the pre-trained model and our novel model, DenseVit_AUX. Additionally, it includes the detection results acquired from various object detection models, namely YOLOv5l, YOLOv5x, YOLOv8l, YOLOv8m, and YOLOv8x. The models were trained using pre-trained weights on a dataset that included a wide range of waste items. Their effectiveness was evaluated using performance metrics such as Accuracy, precision, recall, mean Average Precision at IoU threshold 0.5 (mAP50), and mean Average Precision across multiple IoU thresholds from 0.5 to 0.95 (mAP50-95).

4.1 Result of classification

Table 1 provides a comprehensive comparison of performance data for eight object identification models: MobileNetV2, InceptionV4, ResNet18, EfficientNet-v2, ConvNext, DenseNet-201, MaxViT, and 2S_DenseViT. The models were assessed using a dataset that consisted of 28 distinct categories of different bottles, canisters, cans, cardboards, and detergents. The metrics include accuracy, precision, sensitivity, F1-score, and specificity, offering a thorough assessment of the performance of each model in successfully categorizing the objects.

The 2S_DenseViT model has the most excellent overall performance, with an accuracy of 83.11%, precision of 83.33%, and F1 score of 83.05. The result demonstrates its exceptional proficiency in precisely identifying and categorizing items, with a negligible occurrence of incorrect positive and negative results across the five categories. DenseNet-201 has excellent performance, notably in terms of accuracy (81.21%) and F1 Score of (81.01%), which demonstrates its ability to accurately detect true positives while successfully excluding false positives. We have also found very good performance with MaxViT model an accuracy of 82.19%, precision of 82.03%, and F1 score of 82.07%. These results highlight its reliability as a suitable option for object identification tasks. We have also achieved promising classification performance using EfficientNet-v2, InceptionV4, Table 2 and ConvNext model. The constant performance across all measures demonstrates the robustness of the system in managing a wide range of object kinds. Compared to the other models, ResNet18 achieved a lower accuracy of 79.14%, precision of 79.42%, and F1 score of 79.38. All of the models achieved high specificity which means the models show low false positive rates, and they are good at identifying negative classes.

The confusion matrix of our proposed 2S_DenseViT model is shown in Fig. 3. These matrices represent the performance of each model in categorizing 28 different waste materials, such as bottles, canisters, cans, cardboard, and detergent. The confusion matrix displays the rates of true positive, false positive, and false negative, which are essential for comprehending the accuracy and resilience of the models in differentiating between various item kinds in an overflowing trash environment.

The confusion matrix for the 2S_DenseViT model has high accuracy in recognizing different bottles. Our performance across various bottle classes has been quite remarkable, with sensitivity ranging between 71 and 92% for 19 out of the 20 classes. Our proposed model did not get reliable sensitivity for bottle multicolor class (46%) due to high similarity in 20 bottle classes. We have got the best performance in cans (97%) and canisters (93%) since these two super classes do not have sub-categories like bottles and detergent.

Table 1 Comparative analysis of performance metrics for object detection models

Models	Accuracy	Precision	Sensitivity	F1 Score	Specificity
MobileNetV2	79.65	78.53	79.65	79.38	99.01
InceptionV4	80.05	79.87	80.05	79.76	98.83
Resnet18	79.14	79.42	79.14	79.06	99.04
ConvNext	80.21	80.28	80.21	80.15	98.97
EffiicientNet-v2	80.98	81.11	80.98	80.88	98.98
MaxViT-T	82.19	82.03	82.19	82.07	98.95
Densenet201	81.21	80.98	81.21	81.01	91.9
2S_DenseViT	83.11	83.33	83.11	83.05	98.94

Table 2 Performance Analysis of our proposed 2S_DenseViT model across 28 different categories of waste images

Classes	Acc (%)	Pre (%)	Sen (%)	F1 (%)	Spe (%)
Bottle-blue	83.11	79.41	77.88	78.64	98.55
bottle-blue5l		80.52	86.11	83.22	98.99
bottle-blue5l-full		80.95	70.83	75.55	99.74
bottle-blue-full		79.07	79.07	79.07	99.4
bottle-dark		91.49	90.53	91.01	99.45
bottle-dark-full		81.08	88.24	84.51	99.54
bottle-green		87.67	86.49	87.08	99.39
bottle-green-full		76.32	85.29	80.56	99.41
bottle-milk		76.27	78.95	77.59	99.06
bottle-milk-full		73.08	90.48	80.85	99.54
bottle-multicolor		68.42	46.43	55.32	99.61
bottle-multicolorv-full		75	71.43	73.17	99.67
bottle-oil		87.23	85.42	86.32	99.6
bottle-oil-full		85.71	75	80	99.94
bottle-transp		82.55	82.91	82.73	96.89
bottle-transp-full		77.42	78.26	77.84	98.56
bottle-yogurt		78.26	85.71	81.82	99.34
canister		96.55	93.33	94.91	99.93
cans		95.96	96.94	96.45	99.72
detergent-box		100	76.47	86.67	100
detergent-color		69.23	83.72	75.79	98.94
detergent-transparent		81.82	65.85	72.97	99.6
detergent-white		71.43	81.4	76.09	99.07
glass-dark		95.65	88	91.67	99.93
glass-green		95.83	92	93.88	99.93
glass-transp		93.33	77.78	84.85	99.87
juice-cardboard		83.08	79.41	81.2	99.26
milk-cardboard		86.6	89.36	87.96	99.11
Weighted Average		83.33	83.11	83.05	98.94

The 2S_DenseViT model demonstrated consistent performance across various detergent categories, with the highest sensitivity observed in the detergent color class at 83.72% and the lowest in detergent transparent at 65.32%. For the two cardboard waste material categories, we have found 79.41% sensitivity in the juice-cardboard class and 89.36% sensitivity in the milk juiceboard class. From the confusion matrix, we can see the incorrect predictions within the bottle and detergent classes which show the challenge posed by their numerous sub-categories. Many of these mispredictions occur within their respective types, highlighting the complexity of accurately classifying such diverse classes.

4.2 Result of object detection

Table 3 presents a comparison of the performance metrics of several object detection models, namely YOLOv5l, YOLOv5x, YOLOv8l, YOLOv8m, YOLOv8x, and GELAN-E. The metrics being evaluated include precision, recall, mAP50, and mAP50-95.

Based on the results of all the models, GELAN-E stands out as it has the greatest accuracy (0.65), recall (0.56), mAP50 (0.63), and mAP50-95 (0.52), showcasing its better overall performance. This suggests that GELAN-E has superior performance in terms of both accuracies, accurately recognizing things, and recall, detecting a larger number of relevant objects, in comparison with its counterparts. GELAN-E's performance in terms of mAP50 and mAP50-95 highlights its strong and accurate performance across diverse IoU thresholds, establishing it as the most dependable model for a range of detection tasks. The YOLO models, while they usually exhibit excellent reliability, are inferior to GELAN-E in terms of performance. YOLOv5l and YOLOv5x exhibit comparable accuracy (0.57 and 0.55, respectively) and mAP50 (0.48 and 0.50, respectively). However, their recall (0.42 and 0.44) and mAP50-95 (0.38 and 0.40) suggest that they are less

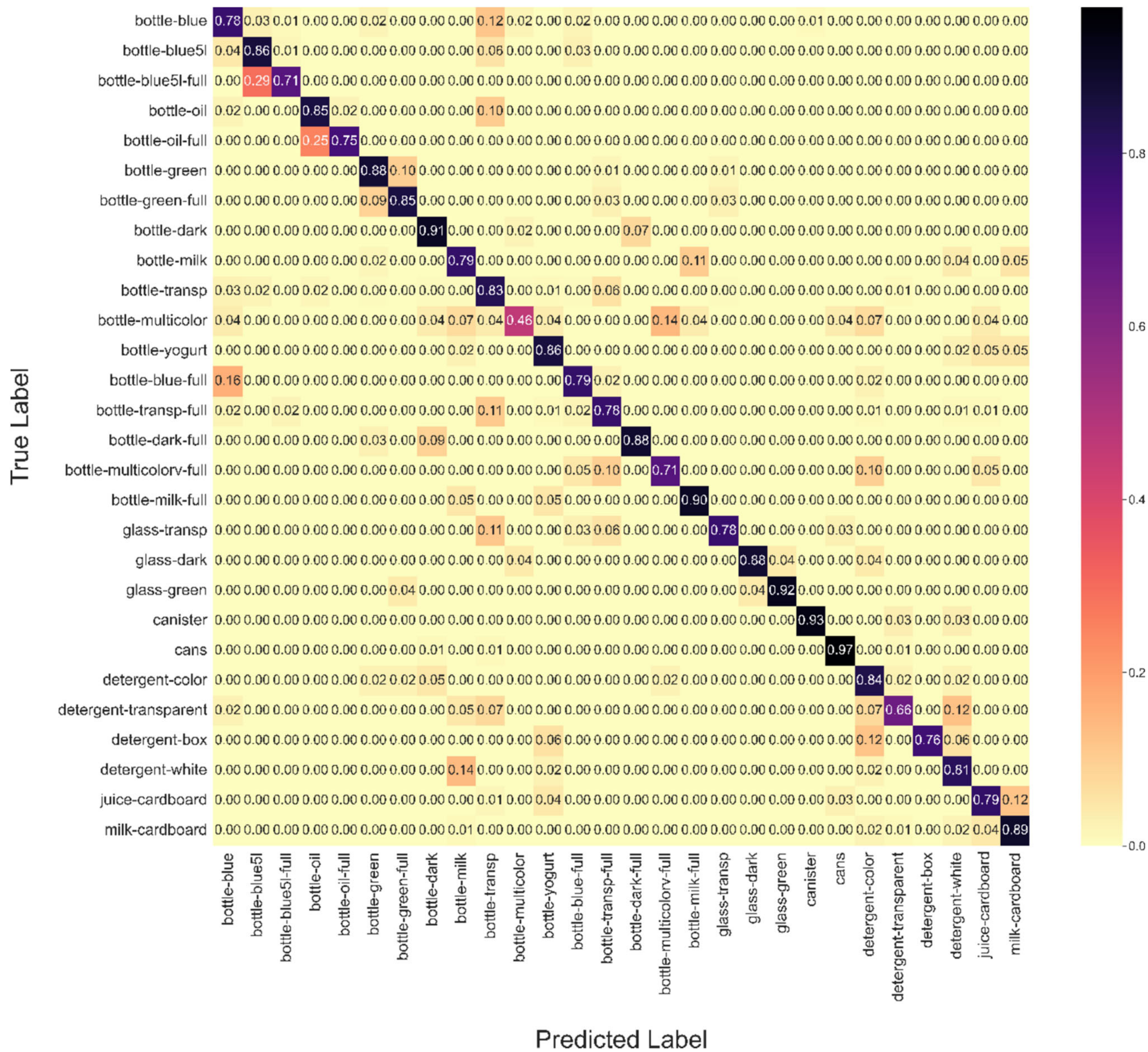


Fig. 3 Confusion matrix of our classification result with 2S_DenseViT Model

Table 3 Comparative analysis of performance metrics for various object detection models

Models	Precision	Recall	mAP50	mAP50-95
YOLOv5l	0.57	0.42	0.48	0.38
YOLOv5x	0.55	0.44	0.50	0.40
YOLOv8l	0.49	0.46	0.49	0.39
YOLOv8m	0.50	0.39	0.44	0.34
YOLOv8x	0.52	0.42	0.48	0.39
GELAN-E	0.65	0.56	0.63	0.52

reliable in recognizing each of the relevant items compared to GELAN-E. The YOLOv8 series, which includes YOLOv8l, YOLOv8m, and YOLOv8x, has somewhat lower metrics overall. YOLOv8l performs well in terms of balance, but it still falls short of GELAN-E's superior efficiency and accuracy. As a result, whereas the YOLO models show decent performance in many measures, GELAN-E outperforms them greatly, especially in terms of accuracy and recall. GELAN-E is a very effective model for precise and thorough object detection. It is capable of retaining exceptional performance across various detection thresholds. Table 4 demonstrates the performance metrics of a GELAN-E model on a dataset's different classes.

Table 4 Evaluation of the GELAN-E Model's performance on different object classes

Class	Precision	Recall	mAP50	mAP50-95
all	0.655	0.569	0.637	0.522
bottle-blue	0.699	0.558	0.635	0.494
bottle-green	0.727	0.646	0.737	0.618
bottle-dark	0.698	0.747	0.792	0.655
bottle-milk	0.487	0.386	0.472	0.382
bottle-transp	0.624	0.556	0.624	0.497
bottle-multicolor	0.753	0.286	0.482	0.411
bottle-yogurt	0.545	0.333	0.391	0.332
bottle-oil	0.557	0.5	0.604	0.506
cans	0.69	0.429	0.542	0.407
juice-cardboard	0.36	0.338	0.346	0.277
milk-cardboard	0.525	0.411	0.478	0.393
detergent-color	0.466	0.419	0.476	0.387
detergent-transparent	0.422	0.268	0.295	0.245
detergent-box	0.505	0.706	0.801	0.645
canister	0.563	0.433	0.533	0.472
bottle-blue-full	0.647	0.81	0.78	0.579
bottle-transp-full	0.613	0.761	0.737	0.606
bottle-dark-full	0.767	0.873	0.862	0.761
bottle-green-full	0.827	0.844	0.859	0.748
bottle-multicolorv-full	0.711	0.762	0.816	0.678
bottle-milk-full	0.786	0.875	0.844	0.718
bottle-oil-full	1	0.217	0.498	0.448
detergent-white	0.457	0.442	0.459	0.367
bottle-blue5l	0.661	0.75	0.783	0.64
bottle-blue5l-full	0.79	0.785	0.838	0.73
glass-transp	0.813	0.482	0.657	0.48
glass-dark	0.812	0.691	0.769	0.585
glass-green	0.839	0.625	0.728	0.565

Precision (P), recall (R), mAP50, and mAP0.5 to 0.95 are the measurements.

The model achieved a precision of 0.655, recall of 0.569, mAP50 of 0.637, and mAP50-95 of 0.522 across all classes. The class "bottle-dark-full" had the greatest precision (0.767), recall (0.873), mAP50 (0.862), and mAP50-95 (0.761) among all the individual classes, which indicates outstanding detection ability. On the other hand, bottle-yogurt and detergent-transparent had lower performance. Bottle-yogurt had a precision of 0.545, recall of 0.333, mAP50 of 0.391, and mAP50-95 of 0.332. Detergent-transparent had a precision of 0.422, recall of 0.268, mAP50 of 0.295, and mAP50-95 of 0.245. Some classes, like bottle-oil-full, had impeccable accuracy (1.000), but had a poor recall (0.217), indicating a strong certainty in detecting occurrences when they occur, but also missing a considerable percentage of them. On the other hand, the bottle-multicolor-full and bottle-milk-full classes had excellent overall performance, with accuracy and recall values above 0.7, and mAP50-95 values also exceeding 0.7. This demonstrates a harmonious and accurate performance in detecting and correctly identifying occurrences within these categories. The variations in performance across several categories highlight the model's capabilities and areas that need improvement. Classes that perform well tend to have distinct characteristics and a consistent visual presentation; while, classes that perform weakly may be impeded by visual complexity or obstruction, which may impact the accuracy and consistency of detection.

Supplementary Fig. 3 illustrates the correlation between precision and confidence for 28 different classes. The x-axis represents the confidence level of predictions produced by the model, ranging from 0.0 to 1.0. Higher values indicate increased confidence in the model's predictions. The y-axis, on the other hand, measures precision, which ranges from 0.0 to 1.0, with higher precision indicating fewer false positives. Similarly, the recall-confidence curve in Supplementary Fig. 3 shows that higher confidence thresholds lower recall rates, resulting in fewer true positives identified by the model. Lower confidence levels boost recollection, permitting more true positives. On the other hand, Supplementary Fig. 3 displays the precision–recall curve for all waste types after model training with the dataset. The abscissa axis represents recall rate; while, the ordinate axis represents precision rate. The curve represents precision–recall trade-offs.

The 2S_DenseViT and GELAN-E models exhibit remarkable efficacy in trash classification and detection; however, their computational requirements are a significant factor for large-scale, real-time applications. In practical deployment circumstances, computational overhead can be alleviated using techniques such as model pruning, quantization, and knowledge distillation, which are frequently employed to decrease model size and enhance inference speed. Subsequent efforts will concentrate on applying these optimization strategies to guarantee that the models maintain scalability and efficiency while preserving accuracy.

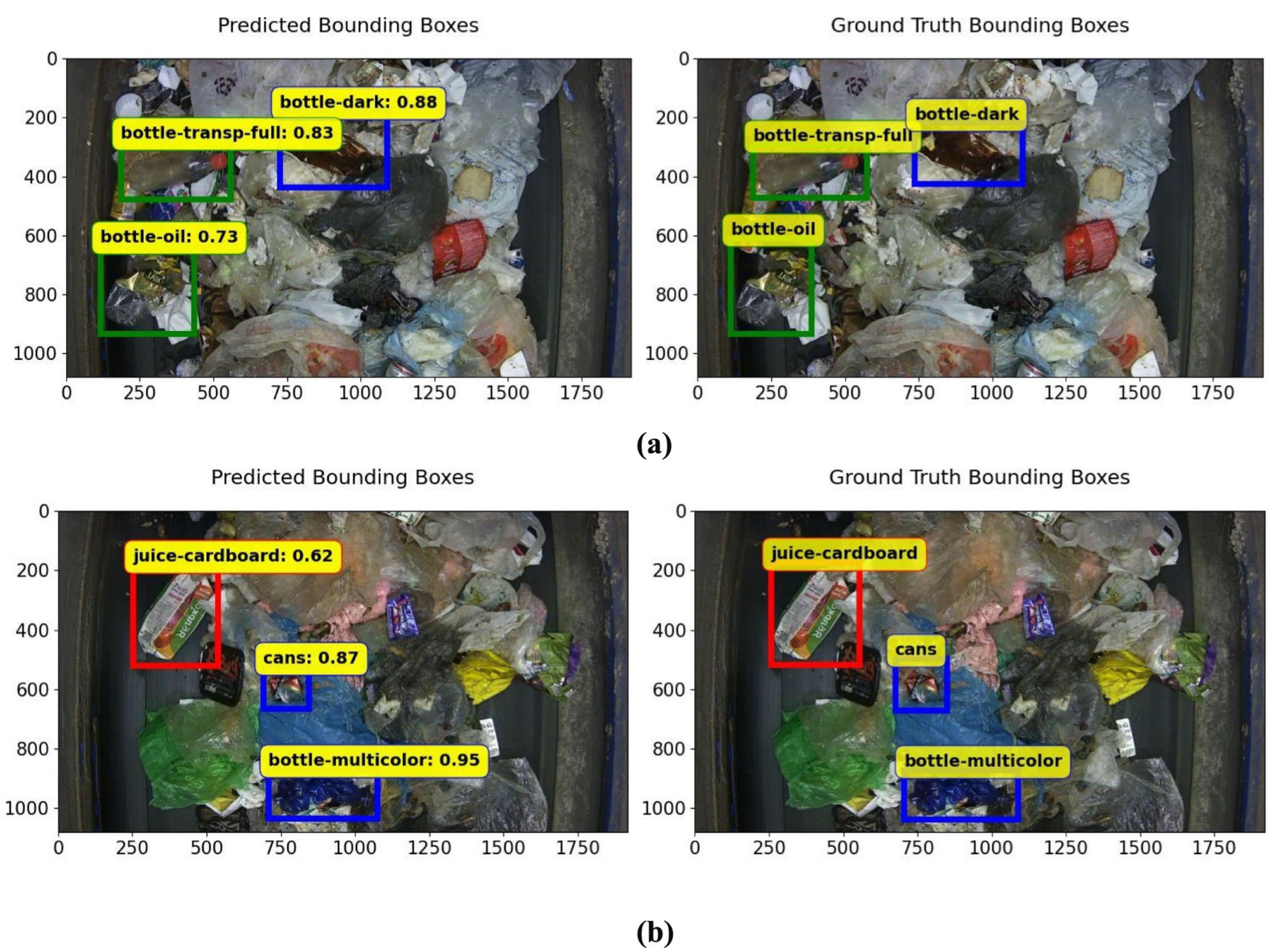
4.3 Waste recycling plant prediction

The diagram displays the object's prediction achieved by GELAN-E, showcasing the model's sophisticated capacity to precisely detect and locate things in a complex garbage environment. The merging of CSPNet and ELAN concepts in GELAN-E guarantees effective feature aggregation and gradient propagation, leading to accurate bounding boxes that closely match the ground truth annotations. The accuracy is demonstrated clearly by the high confidence ratings and the accurate recognition of different bottle types in challenging backdrops.

Figure 4 illustrates the comparison between the predicted and actual bounding boxes using GELAN-E. The picture on the left in the top row displays the predicted bounding boxes of the model. It shows a three bounding box identified as "bottle-dark" with a high confidence score of 0.88, "bottle-transp-full" with a confidence score of 0.83, and "bottle-oil" with a confidence score of 0.73. The expected areas roughly correspond to the actual bounding box in the right picture, which are additionally labeled "bottle-dark," "bottle-transp-full," and "bottle-oil." This indicates that the model has accurately aligned and identified the object. The picture on the left in the bottom row also depicts three predicted bounding boxes labeled "bottle-multicolor," "cans," and "juice-cardboard" with a confidence score of 0.95, 0.87, and 0.62, respectively. The prediction is compared to the ground truth bounding boxes in the right picture, which are likewise labeled as "bottle-multicolor," "cans," and "juice-cardboard." The projected bounding box has a tight correspondence with the ground truth in terms of both location and labeling, though having a slightly reduced confidence score.

4.4 Qualitative assessment using computer-aided modeling (CAM)

This section provides a qualitative assessment of our 2S_DenseViT model, intended for the categorization of waste photos. We used Score-CAM-based Class Activation Mapping (CAM) to evaluate the model's prediction skills and its proficiency in categorizing various types of garbage. This approach allows us to identify the areas of the input photos that the model emphasizes during categorization choices. The CAM heatmaps produced for different trash categorization situations provide a comprehensive insight into the model's focal points and decision-making mechanisms. Figure 5 presents nine test pictures obtained from various folds of the test dataset, together with their corresponding original images and Score-CAM heatmaps. This research enables a rigorous evaluation of the 2S_DenseViT



(a)

(b)

Fig. 4 Comparison between the predicted and actual bounding boxes for object detection in a clustered waste environment using GELAN-E

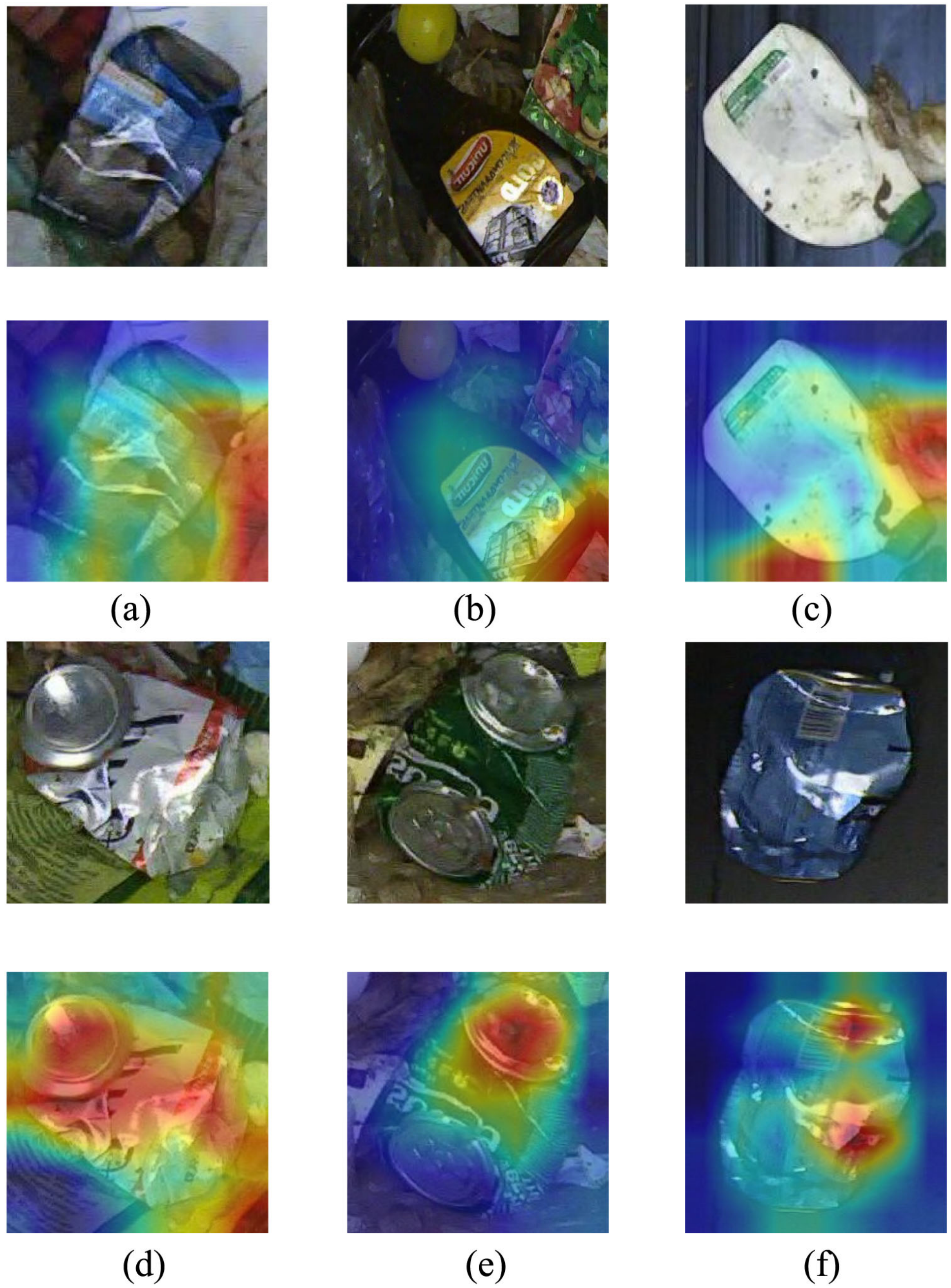


Fig. 5 Class Activation Mapping (CAM) of garbage images and original images based on score-CAM

model's efficacy in differentiating various waste categories.

4.5 Performance comparison of GCDN-Net with existing studies

Table 5 represents the performance comparison of our proposed method with existing research studies on various waste classification and detection tasks using different datasets. For classification task, different deep learning networks (improved ResNet, Vision Transformer, and RWC-Net) were proposed to get reliable performance in waste identification task in earlier datasets with limited classes. Also, various deep learning networks, including Mask R-CNN, YOLOv5, ResNet50, and RWC-Net, have been proposed previously to achieve reliable performance in waste detection tasks across different datasets.

Table 5 Performance comparison of our proposed method with the research studies in WaRP dataset and other studies

	Dataset	Images	Classes	Task	CNN Model	Performance
[44]	TrashNet	2527	6	Classification	Improved ResNet	F1 Score: 95.87%
Proença & Simões, 2020	TACO	1500	28	Detection	Mask R-CNN	mAP: 0.194
Sukel et al., 2023	GIGO	9352	5	Multi-label Classification	Vision Transformer	F1 Score 80% and mAP 0.47
Meng & Chu, 2020	AquaTrash	369	4	Detection	ResNet50	mAP 0.8148
Hossen et al., 2024	TrashNet	2527	6	Classification	RWC-Net	95.01% ACC
[41]	WaRP	10,456 (CL)	28	Classification +	ConvNext + YoloV5	81.81% ACC (CL) + mAP 0.58
Our Methods		+ 2974 (DE)		Detection	2S_DenseViT + GELAN-E	83.11% ACC (CL) + 0.63 (DE)

The WaRP dataset was published in 2023 becoming the new benchmark dataset for both waste classification and detection task with 28 classes of waste materials. Our proposed classification (CL) model and object detection (DE) model (2S_DenseViT + GELAN-E) achieved a classification accuracy of 83.11%, and mAP of 0.63 significantly outperforming this study [42].

During our test, we discovered that the 2S_DenseViT model performed inadequately in several trash categories, specifically the bottle-multicolor class. This performance gap is due to the visual complexity and unpredictability inherent in this class, which may impair the model's capacity to reliably discriminate these pictures. To solve this issue, future research will investigate sophisticated object detecting approaches. Specifically, including attention processes or using more advanced object detection models may improve the model's ability to effectively distinguish visually complicated waste categories. By examining these techniques, we want to increase our model's resilience and accuracy in recognizing a wider variety of waste sorts.

5 Conclusion

The integration of machine learning and deep learning technologies in waste classification and detection is a significant advancement in environmental sustainability. These technologies have several advantages such as high efficiency, large capacity, and low cost compared to the traditional methods of waste disposal. Deep learning models help in the real-time sorting and classification of wastes which in turn enhances recycling efficiency, reducing pollution, and mitigating environmental impact. In our research, we employed 28 distinct categories of waste materials for both classification and detection tasks. Unlike other datasets such as TrashNet, our dataset consists of images without any colored background, offering a more realistic environment for machine learning and deep learning-based investigation. We have developed 2S_DenseViT, a novel dual-stream deep learning model that achieved an overall accuracy of 83.11% in classifying 28 categories of waste materials. Our waste detection model, GELAN-E achieved a notable good mAP of 0.63. Both of our classification and detection performance outperformed the existing method in the WaRP dataset. The research effectively illustrates the usefulness of deep learning in garbage sorting; despite that, numerous drawbacks are there. This encompasses difficulties in precisely classifying waste kinds and inconsistencies in performance across varying environmental circumstances. Future enhancements may concentrate on expanding dataset variety and optimizing the model to manage more intricate waste streams. Future research will focus on enhancing the computational efficiency of the proposed models through techniques including model pruning, quantization, and knowledge distillation. The optimizations will seek to decrease the computational demands of the models, thereby enhancing their applicability for real-time use in extensive waste management systems. Moreover, subsequent investigations could examine the incorporation of alternative data modalities, such as thermal imaging, or the development of autonomous sorting algorithms, thereby strengthening the system's scalability and efficiency. Our future research endeavors will focus on enhancing the classification accuracy specifically within the 20 waste categories related to bottle types. We are also planning on collecting more waste materials of different categories to develop a new large benchmark dataset. This will facilitate a more rigorous comparison of deep learning models and their performance.

Authors contribution Conceptualization contributed by FRS, MSBI, MEHC; data curation contributed by FRS, MSBI, MN; investigation contributed by FRS, MSBI, MN; funding acquisition, supervision and project management contributed by MoN, MH, AKAK, AK, AA, MEM, SBAK, MEHC; software, resources and validation contributed by MEHC; writing—original draft preparation contributed by FRS, MSBI, MN, MEHC. All authors are involved in methodology development, and writing—reviewing and editing. FRS and MSBI equally contributed to the study, drafting and reviewing.

Funding Open Access funding provided by the Qatar National Library. This work was made possible by Qatar National Research Fund # UREP29-205–2-056. The statements made herein are solely the responsibility of the authors. The open-access publication cost is covered by the Qatar National Library.

Data availability The pre-processed dataset can be made available upon a reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00521-024-10855-2.

References

1. Abdallah M, Talib MA, Feroz S, Nasir Q, Abdalla H, Mahfood B (2020) Artificial intelligence applications in solid waste management: A systematic research review. Waste Manage 109:231–246
2. Adil M, Ullah R, Noor S, Gohar N (2022) Effect of number of neurons and layers in an artificial neural network for generalized concrete mix design. Neural Comput Appl 34(11):8355–8363
3. Ahmad K, Khan K, Al-Fuqaha A (2020) Intelligent fusion of deep features for improved waste classification. IEEE Access 8:96495–96504
4. Ahmed MIB, Alotaibi RB, Al-Qahtani RA, Al-Qahtani RS, Al-Hetela SS, Al-Matar KA, Al-Saqer NK, Rahman A, Saraireh L, Youldash M (2023) Deep learning approach to recyclable products classification: Towards sustainable waste management. Sustainability 15(14):11138
5. Alsabei A, Alsayed A, Alzahrani M, Al-Shareef S (2021) Waste classification by fine-tuning pre-trained CNN and GAN. Inter J Com Sci & Network Security 21(8):65–70
6. Bashkirova, D., Abdelfattah, M., Zhu, Z., Akl, J., Alladkani, F., Hu, P., Ablavsky, V., Calli, B., Bargal, S. A., & Saenko, K. (2022). Zerowaste dataset: Towards deformable object segmentation in cluttered scenes. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21147–21157.
7. Bawankule R, Gaikwad V, Kulkarni I, Kulkarni S, Jadhav A, Ranjan N (2023) Visual detection of waste using YOLOv8. Inter Con Sus Com Smart Systems (ICSCSS) 2023:869–873
8. Bui TD, Tsai FM, Tseng M-L, Ali MH (2020) Identifying sustainable solid waste management barriers in practice using the fuzzy Delphi method. Resour Conserv Recycl 154:104625
9. Bui T-D, Tseng M-L (2022) Understanding the barriers to sustainable solid waste management in society 5.0 under uncertainties: a novelty of socials and technical perspectives on performance driving. Envir Sci Poll Rese 29(11):16265–16293
10. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, Gallivanone F, Cozzi A, D'Amico NC, Sardanelli F (2021) AI applications to medical images: From machine learning to deep learning. Physica Med 83:9–24
11. Chu Y, Huang C, Xie X, Tan B, Kamal S, Xiong X (2018) Multilayer hybrid deep-learning method for waste classification and recycling. Computational Intelligence and Neuroscience 2018:1–9. https://doi.org/10.1155/2018/5060857
12. Fang B, Yu J, Chen Z, Osman AI, Farghali M, Ihara I, Hamza EH, Rooney DW, Yap P-S (2023) Artificial intelligence for waste management in smart cities: a review. Environ Chem Lett 21(4):1959–1989
13. Iyamu HO, Anda M, Ho G (2020) A review of municipal solid waste management in the BRIC and high-income countries: A thematic framework for low-income countries. Habitat Int 95:102097
14. Kabir E, Kaur R, Lee J, Kim K-H, Kwon EE (2020) Prospects of biopolymer technology as an alternative option for non-degradable plastics and sustainable management of plastic wastes. J Clean Prod 258:120536
15. Kabirifar K, Mojtahedi M, Wang C, Tam VWY (2020) Construction and demolition waste management contributing factors coupled with reduce, reuse, and recycle strategies for effective waste management: A review. J Clean Prod 263:121265
16. Khan S, Anjum R, Raza ST, Bazai NA, Ihtisham M (2022) Technologies for municipal solid waste management: Current status, challenges, and future perspectives. Chemosphere 288:132403
17. Kiyokawa T, Takamatsu J, Koyanaka S (2022) Challenges for future robotic sorters of mixed industrial waste: a survey. IEEE Trans Autom Sci Eng 21(1):1023–1040
18. Longo E, Sahin FA, Redondi AEC, Bolzan P, Bianchini M, Maffei S (2021) A 5g-enabled smart waste management system for university campus. Sensors 21(24):8278

19. Majchrowska S, Mikołajczyk A, Ferlin M, Klawikowska Z, Plantykow MA, Kwasigroch A, Majek K (2022) Deep learning-based waste detection in natural and urban environments. Waste Manage 138:274–284
20. Malik M, Sharma S, Uddin M, Chen C-L, Wu C-M, Soni P, Chaudhary S (2022) Waste classification for sustainable development using image recognition with deep learning neural network models. Sustainability 14(12):7222
21. Marin I, Mladenović S, Gotovac S, Zaharija G (2021) Deep-feature-based approach to marine debris classification. Appl Sci 11(12):5644
22. Ogrezeanu, I. A., Suciu, C., & Itu, L. M. (2024) Automated Waste Sorting: A Comprehensive Approach Using Deep Learning for Detection and Classification. *2024 32nd Mediterranean Conference on Control and Automation, MED* https://doi.org/10.1109/MED61351.2024.10566174
23. Ozkaya, U., & Seyfi, L. (2019). Fine-tuning models comparisons on garbage classification for recyclability. *ArXiv Preprint.*
24. Pan Y, Zhang G, Zhang L (2020) A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. Autom Constr 119:103357
25. Pang, H., & Huang, C. (2023). A Two-Stage Deep Learning Framework for Enhanced Waste Detection and Classification. *Proceedings - 22nd IEEE International Conference on Machine Learning and Applications, ICMLA 2023*, 2014–2021. https://doi.org/10.1109/ICMLA58977.2023.00304
26. Parasar, D., Vadalia, S. R., Chavan, S. S., Bhere, K. R., Nabi, F., & Patel, A. Z. (2024). *Waste Detection and Water Quality Assessment in Aquatic Environments: A Comprehensive Study Using YoloV8 and XGBoost*.
27. Qu, D. (2021). Application of Artificial Intelligence in Waste Classification Management at University. *International Conference on Intelligent Vision and Computing*, 330–343.
28. Rahman MW, Islam R, Hasan A, Bithi NI, Hasan MM, Rahman MM (2022) Intelligent waste management system using deep learning with IoT. J King Saud Univ-Com Info Sci 34(5):2072–2087
29. Riyadi S, Andriyani AD, Masyhur AM (2024) Classification of recyclable waste using deep learning: a comparison of yolo models. Revue d'Intelligence Artificielle 38(4):1089–1096. https://doi.org/10.18280/ria.380404
30. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520.
31. Satav AG, Kubade S, Amrutkar C, Arya G, Pawar A (2023) A state-of-the-art review on robotics in waste sorting: scope and challenges. Inter J Interac Design Manu (IJIDeM) 17(6):2789–2806
32. Shi C, Tan C, Wang T, Wang L (2021) A waste classification method based on a multilayer hybrid convolution neural network. Appl Sci 11(18):8572
33. Siddiqi A, Haraguchi M, Narayanamurti V (2020) Urban waste to energy recovery assessment simulations for developing countries. World Dev 131:104949
34. Singla A, Gupta M (2023) Investigating Deep learning models for NFT classification: A Review. Scie J Meta Bloc Techn 1(1):91–98
35. Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*.
36. Tsai F-M, Bui TD, Tseng M-L, Lim MK, Wu K-J, Mashud AHM (2021) Assessing a hierarchical sustainable solid waste management structure with qualitative information: Policy and regulations drive social impacts and stakeholder participation. Resour Conserv Recycl 168:105285
37. Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., & Li, Y. (2022). MaxViT: Multi-axis Vision Transformer. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-031-20053-3_27
38. Wu F, Yang Z, Mo X, Wu Z, Tang W, Duan J, Zou X (2023) Detection and counting of banana bunches by integrating deep learning and classic image-processing algorithms. Comput Electron Agric 209:107827
39. Wu T, Dong Y (2023) YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition. Appl Sci 13(24):12977
40. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P (2021) SegFormer: Simple and efficient design for semantic segmentation with transformers. Adv Neural Inf Process Syst 34:12077–12090
41. Yudin D, Zakharenko N, Smetanin A, Filonov R, Kichik M, Kuznetsov V, Larichev D, Gudov E, Budennyy S, Panov A (2024) Hierarchical waste detection with weakly supervised segmentation in images from recycling plants. Eng Appl Arti Intell. https://doi.org/10.1016/j.engappai.2023.107542
42. Yudin D, Zakharenko N, Smetanin A, Filonov R, Kichik M, Kuznetsov V, Larichev D, Gudov E, Budennyy S, Panov A (2024) Hierarchical waste detection with weakly supervised segmentation in images from recycling plants. Engineering Applications of Artificial Intelligence. https://doi.org/10.1016/j.engappai.2023.107542
43. Zailan NA, Azizan MM, Hasikin K, Mohd Khairuddin AS, Khairuddin U (2022) An automated solid waste detection using the optimized YOLO model for riverine management. Front Public Health 10:907280
44. Zhang Q, Zhang X, Mu X, Wang Z, Tian R, Wang X, Liu X (2021) Recyclable waste image recognition based on deep learning. Resour Conserv Recycl 171:105636. https://doi.org/10.1016/J.RESCONREC.2021.105636

45. Zhou Yan (2023) IYOLO-NL: An improved you only look once and none left object detector for real-time face mask detection. Heliyon. https://doi.org/10.1016/j.heliyon.2023.e19064

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Faizul Rakib Sayem[1] · Md. Sakib Bin Islam[2] · Mansura Naznine[3] · Mohammad Nashbat[4] · Mazhar Hasan-Zia[4] · Ali K Ansaruddin Kunju[4] · Amith Khandakar[5] · Azad Ashraf[4] · Molla Ehsanul Majid[6] · Saad Bin Abul Kashem[7] · Muhammad E. H. Chowdhury[5]

✉ Muhammad E. H. Chowdhury
mchowdhury@qu.edu.qa

Faizul Rakib Sayem
faizulrakib-2017214832@eee.du.ac.bd

Md. Sakib Bin Islam
201926034@student.mist.ac.bd

Mansura Naznine
1703008@student.ruet.ac.bd

Mohammad Nashbat
mohammad.nashbat@UDST.edu.qa

Mazhar Hasan-Zia
mazhar.hasanzia@udst.edu.qa

Ali K Ansaruddin Kunju
aliyarukunju.kunju@udst.edu.qa

Amith Khandakar
amitk@qu.edu.qa

Azad Ashraf
azad.ashraf@udst.edu.qa

Molla Ehsanul Majid
mmajid@qf.org.qa

Saad Bin Abul Kashem
saad.kashem@afg-aberdeen.edu.qa

1 Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka, Bangladesh

2 Department of Biomedical Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh

3 Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh

4 Chemical Engineering Department, University of Doha for Science and Technology, Doha, Qatar

5 Department of Electrical Engineering, Qatar University, 2713 Doha, Qatar

6 Computer Applications Department, Academic Bridge Program, Qatar Foundation, Doha, Qatar

7 Department of Computing Science, AFG College With the University of Aberdeen, Doha, Qatar