PLDs-CNN-Ridge-ELM: A parallel lightweight two stage waste classification model guided by SHAP

Mansura Naznine[1], Md. Nahiduzzaman[2], Md. Jawadul Karim[2], Md. Faysal Ahamed[2], Abdus Salam[2,3], Mohamed Arselene Ayari[3], Amith Khandakar[3], Azad Ashraf[4], Mominul Ahsan[5,*] and Julfikar Haider[6]

[1]Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh.

[2]Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh.

[3]Department of Electrical Engineering, Qatar University, Doha 2713, Qatar.

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

[5]Department of Computer Science, University of York, Deramore Lane, Heslington, York YO10 5GH, UK

[6]Department of Engineering, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK

***Corresponding author:**

Dr Julfikar Haider

Department of Engineering,

Manchester Metropolitan University,

Chester Street, Manchester M1 5GD, UK

E-Mail: j.haider@mmu.ac.uk

Abstract

Accurate classification of classifying waste is absolutely crucial for effective managing the vast amount of waste generated by everyday activities in societies. Using technological solutions for garbage classification can lower labor expenses, promote economic recycling, and help protect the environment to achieve UN's sustainability goals across the world. This study presents a novel two-stage waste classification system using the Parallel Lightweight Depth-Wise Separable Convolutional Neural Network (PLDs-CNN) in combination with the Ridge Regression Extreme Learning Machine (Ridge-ELM) classifier using waste images as the input data. The proposed method, which only has nine layers and 1.09 million parameters, effectively distinguishes various types of waste as hazardous, household food, recyclable, and residual. By precisely capturing relevant features from waste images, the PLDs-CNN drastically lowers computational complexity and produces outstanding classification performance. In the first stage (four-class classification), the average values of accuracy, precision, recall, f1, and ROC-AUC are 99%, 97.25±0.02%, 96±0.03%, 96.5±0.01%, and 99.51%, respectively. Likewise, in the second stage, the framework attains remarkable performance ratings for the twelve-class classification: 96.00%, 95±0.033%, 94.33±0.031%, 94.66±0.02%, and 99.59%. Ridge-ELM accomplishes this excellent outcome only with an average testing time of 0.0079 seconds (first stage) and 0.0041 seconds (second stage). The real-time analytical performance is verified by comparing the method with cutting-edge transfer learning methods. Moreover, adding SHAP (Shapley Additive Explanations) improves the model's interpretability by offering essential insights that increase confidence in classifying waste items in practical situations. Furthermore, this study provides a complete hardware architecture that includes a convenient Graphical User Interface (GUI) for real-time waste categorization, a webcam-based conveyor belt sorting mechanism, and a 2-axis pan-tilt system for waste sorting on its own. This comprehensive strategy demonstrates the potential of the research work for creating a cost-effective waste categorization and separation solution to facilitate global recycling and reutilization of waste materials and support circular economy.

Keywords

Waste classification, Parallel Lightweight Depth-Wise Separable Convolutional Neural Network (PLDs-CNN), Ridge Regression Extreme Learning Machines (Ridge-ELM), Hardware Architecture, Graphical User Interface (GUI), and Shapley Additive Explanations (SHAP).

1. Introduction

The rapid growth of global populations and economies has led to a significant rise in resource consumption, resulting in a concerning increase in waste generation [1], [2]. Based on prior studies, it is expected that the worldwide production of solid waste will reach 2.2 billion tons annually by 2025, which will require a budget of $375.5 billion for waste management [3]. 34% of the world's municipal solid garbage is produced by 16% of the population in developed countries [4]. Compared to the global average waste of 0.74 kg per day, these countries generate almost 2.1 kg of waste per person per day [4]. The production of waste is expected to increase rapidly even in low- and middle-income nations such as Africa and some parts of Asia. These countries produce approximately 35% of the world's solid waste [5]. Unfortunately, there is no efficient and automated system for waste disposal to handle this ever-expanding problem [6]. The United Nations Environment Program (UNEP) has identified this as a substantial issue that has negative effects on economic progress, human communities, and public well-being [7]. The inadequate disposal of waste, specifically through landfills and the burning process, presents a substantial danger to urban ecosystems and the welfare of inhabitants [8]. The harmful effects of unlimited waste production, such as the buildup of dangerous compounds and widespread plastic pollution, emphasize the need for scientific solutions in waste management. Recycling and composting are the primary methods of sustainable waste management. However, less than 19% of waste is reused through recycling and composting globally, while roughly 40% of waste ended up in landfills [9]. A thorough knowledge of waste classification is essential for the implementation of an efficient waste management system, as there are numerous distinct types of waste. Therefore, many countries have initiated research on intelligent garbage classification and recycling technologies [10], [11], [12], [13], [14].

The amount of waste produced in Bangladesh increased from 1,100,000 tons in 1970 to 14,778,497 tons in 2012, indicating a 134,300 tons annual increase [9], [15], [16]. According to the recent data, the average amount of solid waste produced by per person in different areas of Bangladesh varies between 0.2 and 0.56 kg [17]. Dhaka, the capital city, generated a daily average of 6,448.37 tons of solid waste in the year 2016-2017 [18]. Although City Corporation collects around 50% of Dhaka's waste, a large amount of waste, approximately 40-60%, remains uncollected and is not properly disposed of. This uncollected waste contains 80% organic material [9]. By 2025, the urban population is expected to reach 78.44 million, and the rate of trash production is estimated to rise to 220 kg per capita per year [19]. The government has initiated the implementation of the National 3R (Reduce, Reuse, and Recycle) Strategy to solve the waste management problem [20]. Another organization “Waste Concern”, a social business enterprise that has emerged to address the issue of municipal garbage accumulation by collaborating with families. UNICEF has also implemented recycling programs and waste management initiatives in collaboration with city corporations and municipalities. Still, there are not enough efforts set up now to improve the standards. Factors like land scarcity and insufficient technical skills have further worsen the problem when dealing with large amounts of garbage. To tackle the issue of huge waste production in Bangladesh, a variety of smart technological approaches can be implemented. Automated sorting systems, sensor-based technologies, and artificial intelligence can improve the accuracy of waste separation, assuring effective recycling.

The advancement of artificial intelligence introduces novel concepts to this domain. Several researchers have successfully employed Convolutional Neural Networks (CNN) for precise waste categorization [21], [22], [23] resulting in a range of notable accomplishments. For waste detection and classification, some studies employed YOLO based models [24], [25], [26] [27], [28] . Some also employed Transfer Learning (TL) based models [29], [30], which have shown high accuracy performance, ranging from 87%-96%. Nevertheless, these TL models possess higher parameters (1.2 - 1.3 million) [31], [32], and the process of determining them necessitates a substantial number of floating-point operations [33]. Also, most of these studies used datasets with limited classes (3-6 classes) and a limited number of images (less than 5000 images). As a result, implementing real-world applications based on these models is a challenging task. Hence, researchers have begun investigating lightweight models based on TL for waste classification, specifically targeting embedded devices or platforms with low resources [34]. Still, there needs to be more adequate research on lightweight garbage classification models. Because the existing lightweight models show lower classification accuracy (83-95%) on large datasets, which can lead to misclassification of waste elements. Also, many of these studies did not explore the potential for real-life implementation of the proposed models.

[29], [30], [31], [32], [33], [34] To overcome the challenges, specifically aiming for a model of small size with low parameter count, yet with high classification accuracy to reduce computing cost, suitable for real-world implementation via embedded system, a two-stage parallel lightweight CNN model has been proposed. The primary contributions of this paper are as follows:

(1) A two-stage parallel lightweight depth-wise separable CNN (PLDs-CNN) has been developed to lower computing requirements during feature extraction. This model can extract vital features while maintaining a low parameter count, thus showing a notable reduction in training and testing time.

(2) A modified Ridge-Extreme Learning Machine (Ridge-ELM) classifier has been implemented to enhance the accuracy of classifying four class and twelve class of waste.

(3) For the first time, the framework's interpretability is highlighted using SHAP (Shapley Additive Explanations) in real-time, which offers valuable insights into the model's decision-making process by identifying the specific areas of an image that the proposed model focused on with significantly higher attention compared to other areas.

(4) This study presents an exhaustive view of the proposed model through integrating hardware execution with a user-friendly Graphical User Interface (GUI) facilitating real-time waste classification for effective recycling. With this approach, the practical usability of the proposed model is verified.

Section 2 of this research presents an in-depth overview of the prior relevant studies conducted in this topic. Section 3 presents the proposed methodology, which includes a comprehensive framework, a description of the dataset, feature extraction methods, and performance metrics. Section 4 provides a detailed presentation of comprehensive classification results, along with an in-depth explanation. Section 4 also presents the interpretability of proposed model using SHAP and hardware-software structure for real life implementation. Conclusions are presented in Section 5.

2. Related Works

So far, researchers have proposed several lightweight and deep learning models for automated classification of wastes. Mao et al. [30] proposed an optimized version of DenseNet121 by applying genetic algorithm to optimize its fully-connected-layer for waste classification. The study was based on the TrashNet dataset, which consists of 2527 images classified into six distinct categories. Employing data augmentation technique, they increased the number of train images for better classification accuracy. The proposed approach achieved a remarkable accuracy of 99.60%, with a training time of 5542 seconds. Nowakowski et al. [35] proposed a deep learning CNN and faster region-based CNN for e-waste analysis and type classifying. They built an e-waste dataset of 210 images with three classes and obtained 90-96.7% accuracy. Khan et al. [36] proposed another waste classification approach, called RWC-EPODL, which utilizes the emperor penguin optimizer with a deep learning model, to generate bioenergy from recyclable garbage. This model employed AX-RetinaNet for object identification and used a stacked auto-encoder (SAE) for classification. The study attained a success rate of 98.96% on Kaggle's garbage categorization dataset, which comprises 750 images categorized into six distinct classes. Lin et al. [37] employed different ResNet architectures derived from transfer learning models to classify waste on the TrashNet dataset. Although they achieved 88.8% accuracy, the model's performance is hindered by the vast number of parameters it possesses. On average, it took about 700 seconds to train for one epoch. Yang et al. [34] introduced a lightweight neural network system called WasNet, with 1.5 million parameters. The performance of the proposed model was demonstrated using three existing datasets. The accuracy achieved on the ImageNet dataset was 64.5%, on the Huawei Garbage Classification dataset it was 82.5%, and on the TrashNet dataset it was 96.10%.

Using modern technologies, some researchers have created intelligent waste categorization tools, which have developed automated and effective garbage classification systems in real life. These solutions are intended to improve garbage management procedures using advanced algorithms to achieve higher classification accuracy and promote a sustainable environment. Chen et al. [31] proposed a garbage classification network (GCNet) based on improved ShuffleNetv2. Employing the parallel mixed attention mechanism (PMAM), incorporating novel activation functions FReLU, and leveraging transfer learning, they enhanced the model's performance and attained an impressive accuracy of 97.9% on their custom dataset. The dataset comprises 4256 photos, which were divided into 14 distinct subcategories. The proposed model can categorize garbage into four separate groups: recyclable, wet, hazardous, and dry garbage. The categorization process required 0.88 seconds, utilizing 1.3 million parameters. Similarly, Fan et al.[38] designed an intelligent garbage bin that separates regular household waste into four categories. The system consists of an automated image classification system, which utilizes a Raspberry Pi, a camera, and three rotating plates. The image classification technique used EfficientNetB2 model in combination with PMAM to obtain high accuracy. Additionally, the researchers proposed a background noise removal (BNR) approach to tackle the impact of environmental factors on garbage recognition. They achieved classification accuracy of 93.38% on the Huawei Cloud Garbage Classification dataset. Feng et al. [32] employed the GECM-EfficientNet model for effective waste classification to create an intelligent waste bin and achieved high accuracy (94.54% and 94.23%) on self-built and TrashNet datasets with 1.23 million parameters. Based on EfficientNet, GECM-EfficientNet used transfer learning, efficient channel attention (ECA) and coordinate attention (CA) modules and streamlining techniques to achieve better accuracy and real-

time performance. The waste bin had a camera and servos for sorting waste into fan-shaped bins. It was operated by a Raspberry Pi 4B.

Jin et al. [39] designed a device utilizing deep learning techniques to facilitate sustainable garbage recycling. Their model utilized MobileNetV2 as the main framework, incorporating one convolutional attention module (CBAM), one principal component analysis (PCA) dimensionality reduction module, and one fully connected classification layer. The proposed approach significantly decreases the time required for garbage identification by 170ms compared to the conventional MobileNetV2 network and effectively categorizes garbage into four distinct categories. Using the Huawei Cloud Garbage dataset, which includes 14,683 pictures, the recommended approach achieved 90.7% accuracy. Similarly, Zhang et al. [40] developed an automated waste sorting device for categorizing domestic waste. The authors presented a two-step trash recognition-retrieval technique adopting the VGG16 model. A dataset including 1040 waste images was created and the data augmentation technique was employed. In order to mitigate the issue of overfitting, they utilized the ten-fold cross-validation technique. The model categorized 13 different forms of waste into four distinct categories. According to the experimental results, the model's average accuracy was 94.71%.

According to the state-of-art models, most researchers employed deep learning and transfer learning models with huge number of parameters. For instance, in [32], they utilized GECM-EfficientNet with 1.23 million parameters, whereas the authors of [31] employed GCNet with 1.3 million parameters. Thus, these models require a significant amount of time for GPU training. The proposed method in [30] needed 5542 seconds for training purpose. Similarly, in [37], authors utilized various ResNet architectures that required 7000 seconds for training.

Undoubtedly, implementing these algorithms in real life is challenging. To develop a practical, cost effective, intelligent waste sorting system, it is necessary to construct a lightweight model with reduced parameters and layers, which will require shorter training time compared to the current models. Again, the literature shows that certain research was able to achieve higher classification accuracy [30], [35]. However, they used a dataset with a small number of classes and images to demonstrate their proposed model. Although it was observed that some dataset in literature with 13, 14 or 18 sub-classes, but these datasets are not publicly accessible [31], [32], [40]. Furthermore, no studies demonstrate the use of real-time explainable AI, such as SHAP or LIME, emphasizing the impact of individual features.

3. Materials and Methods

3.1. Dataset Description

The effectiveness of the learning models depends on the quality of the dataset used. TrashNet, the widely used dataset for garbage classification research, has a limited six-category taxonomy. Jin et al. [39] used the Huawei Cloud Garbage Classification dataset in their research of garbage identification and classification; however, this dataset is no longer available. The dataset utilized in this study was mainly obtained from Kaggle's Garbage Classification dataset. The database comprises 15,150 images representing twelve distinct categories of domestic waste. All the images of the dataset have been thoroughly classified into four groups according to characteristics such as origin, composition, and perceived hazard levels. Afterwards, the images were additionally categorized into twelve sub-classes. Table 1 and Figure 1 provide a comprehensive overview of the dataset and sample images.

Table 1. Overall datasets on both four classes and twelve sub-classes.

Testing phase	Trash Type		Training	Testing	Validation
First Stage: 4 classes	Hazardous Waste		766	94	85
	Household Food Waste		797	99	89
	Recyclable Waste		10439	1289	1160
	Residual Waste		564	70	63
	Total		**12,566**	**1552**	**1397**
Second Stage: 12 sub-classes	Battery	Hazardous Waste	766	94	85
	Expired Food	Household Food Waste	797	99	89
	Brown Glass	Recyclable Waste	491	61	55
	Cardboard		722	89	80
	Clothes		4313	533	479
	Green Glass		509	63	57
	Metal		623	77	69
	Paper		851	105	94
	Plastic		701	86	78
	Shoes		1601	198	178
	White Glass		628	77	70
	Trash	Residual Waste	564	70	63
	Total		**12,566**	**1552**	**1397**

Figure 1: Garbage Classification dataset include 4 classes: (1) Hazardous Waste, (2) Household Food Waste, (3) Recyclable Waste, (4) Residual Waste; 12 sub-classes: (A) Battery, (B) Expired Food, (C) Brown Glass, (D) Cardboard, (E) Clothes, (F) Green Glass, (G) Metal, (H) Paper, (I) Plastic, (J) Shoes, (K) White Glass, (L) Trash.

3.2. Proposed Framework

Figure 2 delineates the main stages of the proposed deep learning framework: the first stage, pre-processing and splitting image dataset, application of deep learning model for classification, and generation of an interpretable model using SHAP. While the second stage classification occurs in the same way and then a hardware implementation of the proposed model is demonstrated. After collecting

and categorizing the dataset, 80% of the images are allocated for training, 10% for testing, and 10% for validation of the deep learning (DL) models. An advanced novel neural network architecture called PLDs-CNN and parallel CNN (PL-CNN) were utilized to extract essential image features. Furthermore, two classifiers, namely pseudo–Extreme Learning Machine (ELM) and Ridge-ELM, were developed to assess class identification performance. The SHAP approach was employed to explain the models' output generation process.

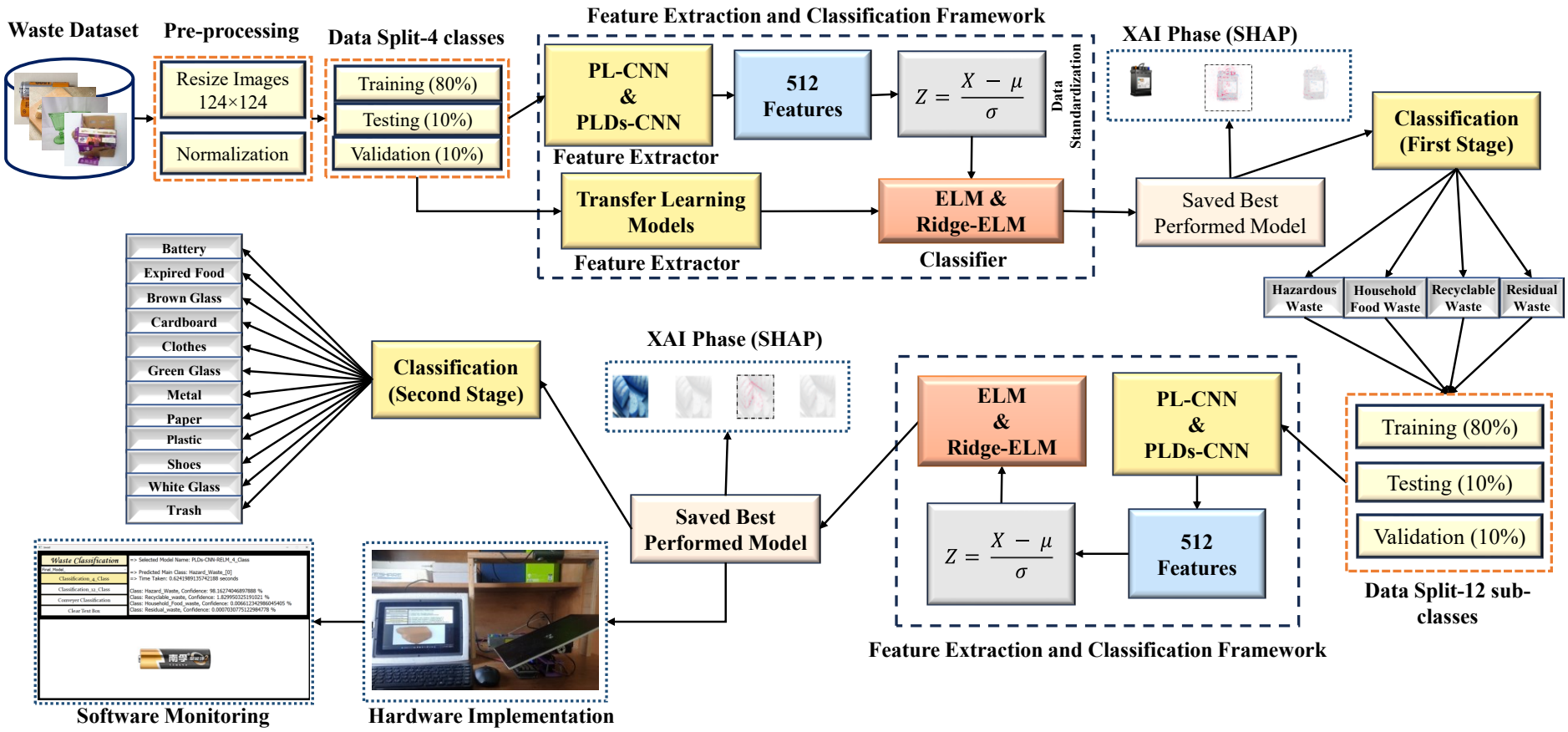


Figure 2: Proposed research framework on two stage multi-class waste image classification.

3.3. Image Pre-processing

The image pre-processing phase involves two essential procedures: normalization [41] and image down sampling. The operations aim to shrink the images to a standardized dimension of 124×124 pixels. These techniques are essential for enhancing model efficiency and extracting features.

3.4. Deep Learning Model

Existing literature indicates that, currently, most of the studies concentrate on employing large models for waste image classification tasks, with less attention given to the development of lightweight models and their practical applications in real-world scenarios. To address these challenges, PLDs-CNN was developed. This customized design incorporates reduction in model parameters, layers, and total size, requiring less computational resources. The subsequent sections contain a detailed explanation of the PLDs-CNN model, along with brief insights into the state-of-the-art transfer learning (TL) models.

3.4.1. Feature Extraction

The primary challenge when building a CNN model is to determine the optimal layer set up. Limited parameters and layers may restrict the model's ability to capture unique features, imposing restrictions on its performance. On the other hand, an excessive number of parameters and layers might cause overfitting, which leads to longer processing times and higher computational requirements. Therefore, it is essential to achieve the optimal balance to ensure effective feature extraction with successful implementation. The main objective of this study was to create a CNN model that can extract key features with the least number of parameters and layers.

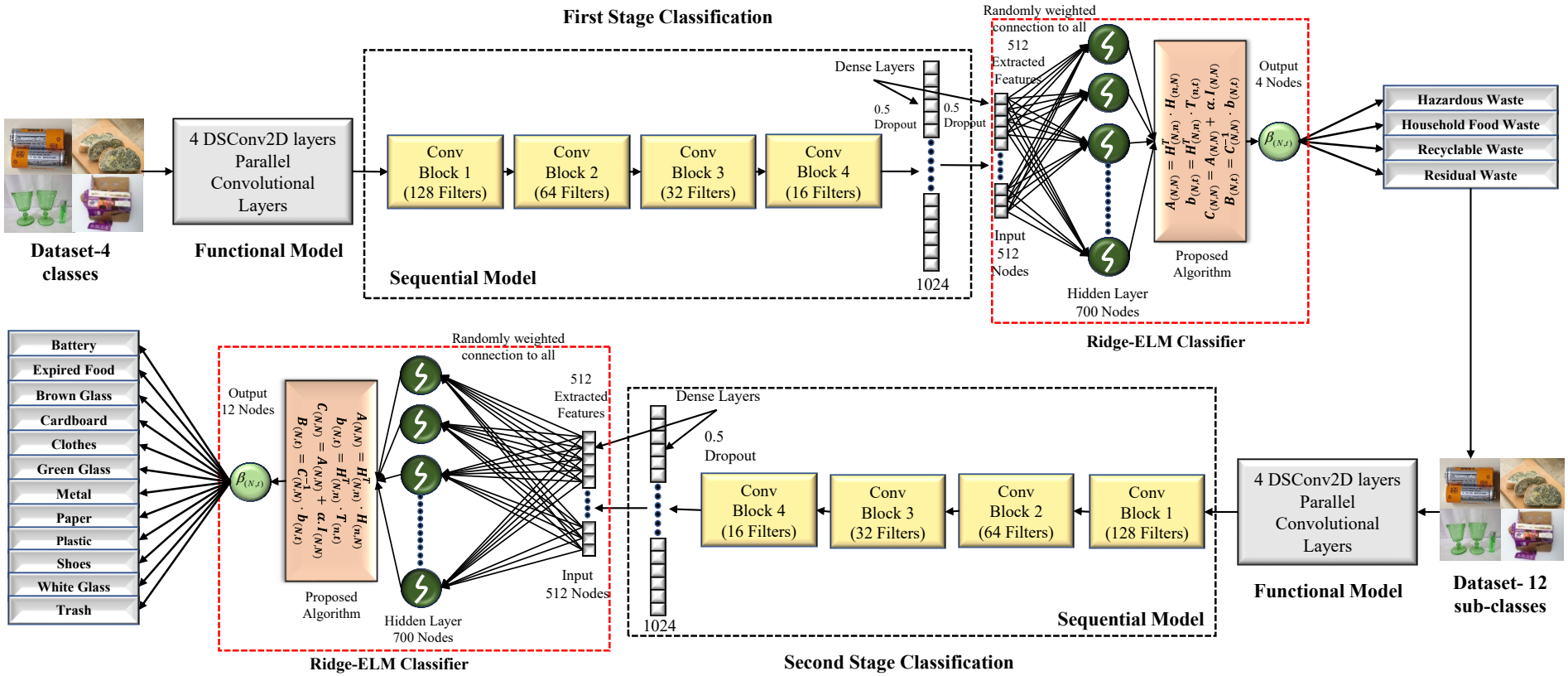


Figure 3: Proposed PLDS-CNN-Ridge-ELM architecture

After taking every aspect into account, a lightweight PLDs-CNN model was created to effectively extract unique features while utilizing the fewest possible resources. The architecture of the PLDs-CNN model is illustrated in Figure 3. The CNN model's accessibility is enhanced by implementing a refined approach. Nine convolution layers (CL) and two fully connected layers (FC) are employed in the model to achieve the most optimal balance. Four parallel CLs instead of a single CL are utilized to extract the discriminating features, as depicted in Figure 4. However, it should be noted that four consecutive CLs will increase the complexity of the model by adding more layers (depth). To address this issue, it was opted to simultaneously run the first four CLs, and their selection was determined through an iterative experimentation process. The convolutional layers (CL) used a total of 256 kernels, with the sizes of the first to fourth kernels set at 9×9, 7×7, 5×5, and 3×3, respectively. Krizhevsky et al. [42] and Nahiduzzaman et al. [43] both support using larger kernel sizes, like 9×9, to guarantee suitable classification performance, which is consistent with the choice of kernel sizes. Since various kernels produce different feature maps, a thorough analysis and integration of a wide range of kernels, from small to large sizes, were conducted to obtain excellent classification performance. Maintaining the same padding size for the first five CLs is essential when extracting important information from the trash image's border element. The feature maps obtained from these simultaneous CLs must be combined and effectively transferred to a subsequent CL without any errors.

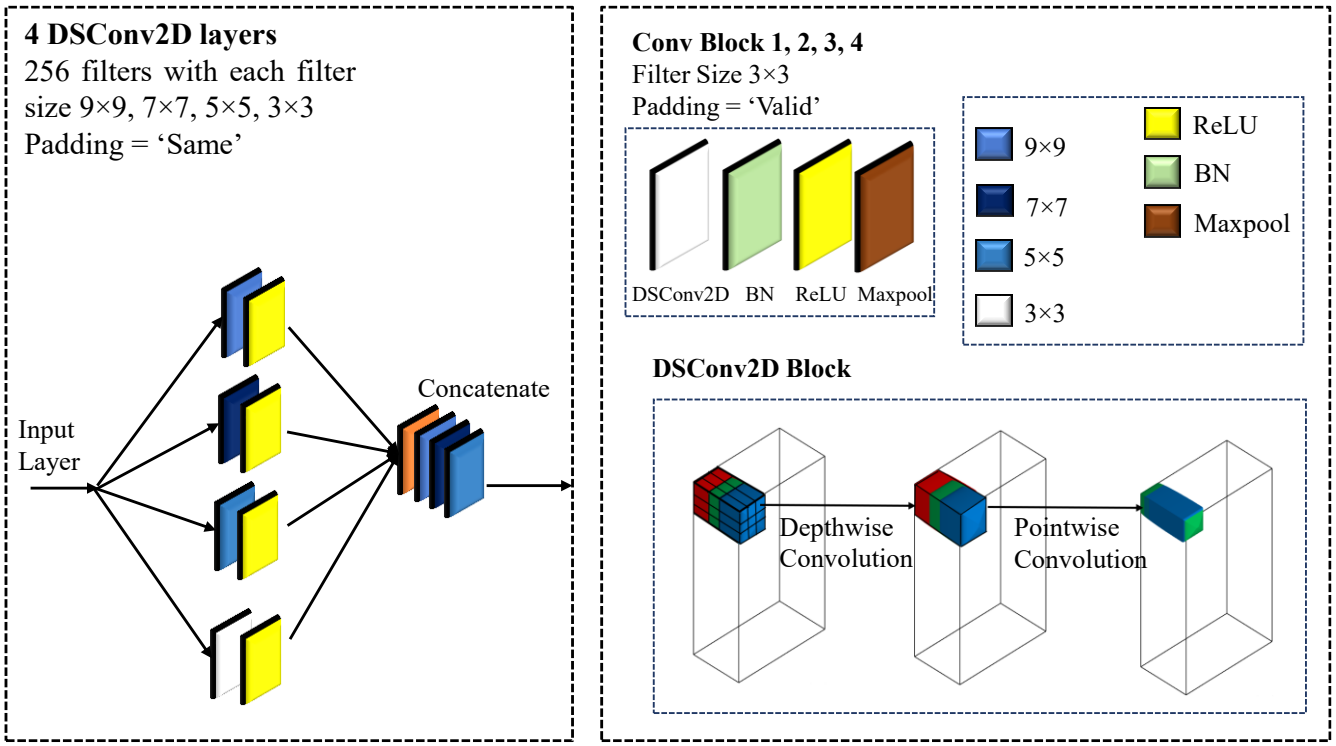


Figure 4: Detail overview of the convolution block.

The CNN architecture was improved by integrating depth-wise separable convolution (DSC), a method that divides the traditional convolution operation into two separate steps: depth-wise and pointwise. A compressed kernel is utilized on a specific portion of DSC to process an infusion feature map, resulting in a new feature map output with the same number of channels. During pointwise convolution, a 1×1 convolutional kernel is applied independently to each channel to create a new feature map with fewer channels. This emphasizes the utmost significance of DSC. This fine-tuning of the parameters leads

directly to a notable decrease in computational complexity. During the final phase, we incorporated three convolutional layers (CLs) and employed N and max pooling with a kernel size of 2×2. The CLs filters had 128, 64, 32, and 16 values, respectively. Each filter was set up with three 3×3 kernels and designed to use valid padding. Batch Normalization (BN) is included to enhance the model's effectiveness. This technique efficiently restores the mean and standard deviation of the inputs for each layer, resulting in improved speed and stability during model execution. All convolutional layers (CLs) employed the Rectified Linear Unit (ReLU) activation function.

Furthermore, to address overfitting and improve the effectiveness of the training process, dropout was employed alongside incorporating two fully connected (FC) layers. A dropout strategy was utilized throughout each training iteration, where 50% of all nodes were randomly deactivated. The objective of this technique was to improve the ability to apply knowledge to new situations and speed up the process of reaching a solution. In the last fully connected (FC) layer, 512 features were extracted. We replaced the SoftMax function with the Ridge-ELM classifier to enhance the classification performance. The model was trained using a loss function derived from the sparse categorical loss function. It was opted to utilize a 32-batch-size ADAM optimizer for this process. After determining the learning rate by trial and error, the model was trained for 200 epochs at 0.001. The overview of the detailed model is displayed in Table 2.

Table 2. Lightweight parallel depth-wise separable convolutional neural network (PLDs-CNN) model summary.

Layer Type	**Output Shape**	**Parameters**
Model Input	(None, 124, 124, 3)	0
Model (Functional)	(None, 124, 124, 1024)	4588
Separable conv2d 4	(None, 122, 122, 128)	140416
Batch normalization	(None, 122, 122, 128)	512
Activation	(None, 122, 122, 128)	0
Max Pooling 2D	(None, 61, 61, 128)	0
Separable conv2d 5	(None, 59, 59, 64)	9408
Batch Normalization 1	(None, 59, 59, 64)	256
Activation 1	(None, 59, 59, 64)	0
Max Pooling 2D 1	(None, 29, 29, 64)	0
Separable conv2d 6	(None, 27, 27, 32)	2656
Batch normalization 2	(None, 27, 27, 32)	128
Activation 2	(None, 27, 27,32)	0
Max Pooling 2D 2	(None, 13, 13, 32)	0
Last conv	(None, 11, 11, 16)	816
Batch normalization 3	(None, 11, 11, 16)	64
Activation 3	(None, 11, 11, 16)	0
Max Pooling 2D 3	(None, 5, 5, 16)	0
Dropout	(None, 5, 5, 16)	0
Flatten	(None, 400)	0
Dense	(None, 1024)	410624
Batch normalization 4	(None, 1024)	4096
Dropout 1	(None, 1024)	0
DenseLastPL	(None, 512)	524800
Total Parameters:	**1,098,364**	
Trainable Parameters:	**1,095,836**	
Non- Trainable Parameters:	**2,528**	

3.4.2. Transfer Learning Models

Transfer learning models such as DenseNet201 [44], EfficientNetB6 [45], InceptionResNetV2 [46], MobileNetV3Small [47], ResNet152V2 [48], VGG16 [49], and Xception [50] have significance to improve the classification of trash images across many categories. Due to their extensive pre-training on big datasets, these models exhibit high effectiveness in extracting significant features from images. These models can efficiently capture intricate patterns and precise information associated with trash images by being fine-tuned on a limited amount of data. The pre-trained models were trained using over 14 million classifications from ImageNet dataset, spanning around 1,000 categories. To produce accurate classification results, we integrated the training of transfer learning models with the proposed Ridge-ELM classifier. A comparative analysis of the novel PLDs-CNN model and TL methods was performed, focusing on classification outcomes and computational resources utilization. This comparison encompasses several aspects like performance, model parameters, layer sizes, as well as training and testing duration. After initializing these models, the final layers were modified by incorporating two fully connected (FC) layers, one with 1024 nodes and the other with 512 nodes. Figure 5 depicts a comprehensive representation of TL with the Ridge-ELM classifier.

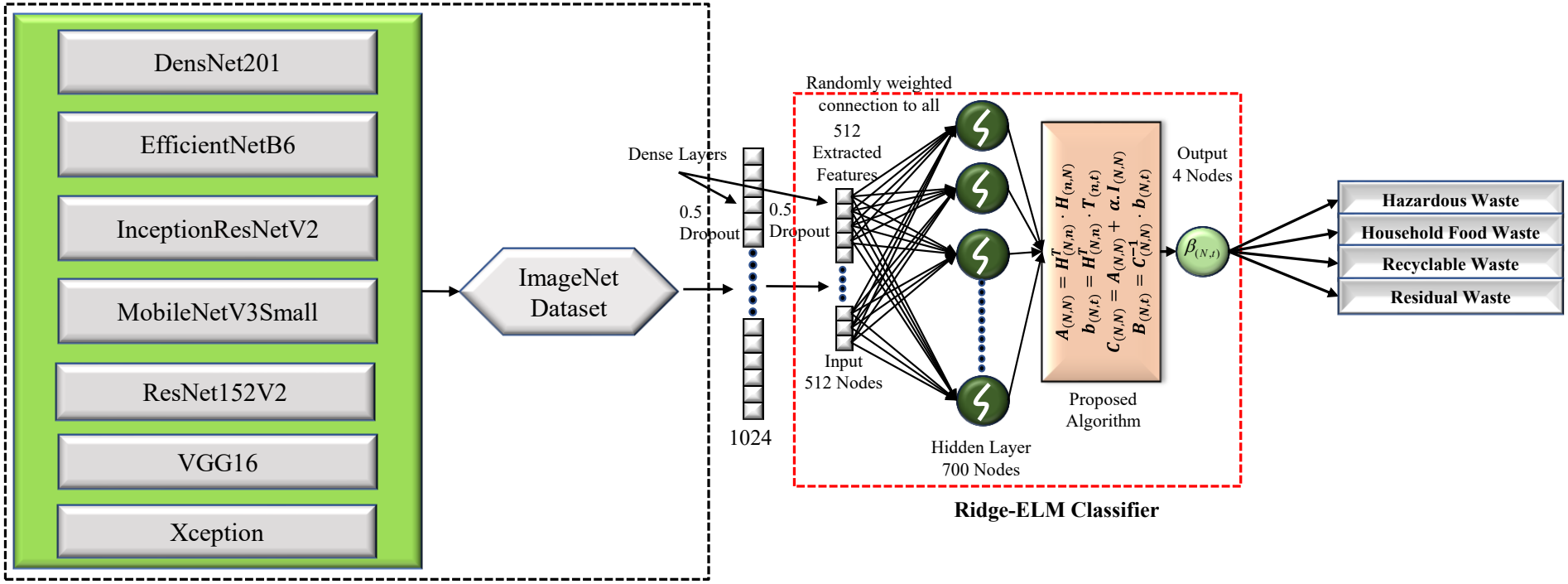


Figure 5: The modified transfer learning architecture to classify waste images.

The DenseNet [44] architecture in CNN is characterized by its distinctive dense connectivity network. Each layer in this architecture is provided with feature maps from the preceding levels, facilitating the seamless transmission of information and establishing densely interconnected routes across the network. The dense interconnectivity facilitates the acquisition and use of complex feature representations, resulting in efficient feature acquisition and enhanced performance. The various iterations of DenseNet, such as DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264, vary in the quantity of layers they possess. As an example, DenseNet-201 consists of 201 layers, whereas DenseNet-169 consists of 169 layers.

EfficientNetB6 was proposed in 2019 by Tan and Le [45]. It belongs to the EfficientNet family. The architecture utilizes compound scaling, which entails proportionally scaling the network's depth, width, and resolution. EfficientNetB6 is trained on the ImageNet dataset and consists of 87 million parameters. It has been employed for several transfer learning applications, including image classification, object detection, and semantic segmentation.

The InceptionResNetV2 [46] model integrates the Inception and ResNet architectures, which efficiently extract features using inception modules and residual connections. Conversely, the MobileNetV3Small [47] belongs to the MobileNet series, which prioritises lightweight operations, making it appropriate for real-time applications on devices with minimal resources. The MobileNetV3Small model has been developed to achieve an ideal trade-off between the size of the model, its speed, and its accuracy. On the other hand, the InceptionResNetV2 model is most suitable for effectively processing delicate visual patterns due to its complex architecture and robust feature learning capabilities.

The Visual Geometry Group (VGG) [49] was a CNN architecture with several convolutional layers and filters. After each CL, an MP layer and a ReLU function aid in feature extraction. The last FC layer uses a SoftMax function with three FC layers. ResNet152V2 is another CNN architecture that was first proposed by He et al. in 2017 [48]. The model's architecture relies on the theory of residual learning, which entails incorporating shortcut connections between layers to enhance the network's learning efficiency. ResNet152V2 includes 60 million parameters.

In 2016, Google developed Xception [50], an architecture that extends the Inception framework. The system employs DSCs, which separately filter each channel of the input feature map, along with pointwise convolutions. This technique greatly decreases computing requirements and memory utilization while upholding accuracy. Xception is commonly employed for diverse computer vision tasks, especially when there are constraints on processing resources, owing to its high efficiency.

3.5. Ridge Extreme Learning Model (Ridge-ELM)

The ELM (Extreme Learning Machine), developed by Huang et al. [51], represents a significant change in feature classification methodology. The employed approach is a feed forward network based on supervised learning which is a pioneering innovation. By applying the strength of neural networks (NN), the ELM eliminates the necessity of backpropagation, resulting in a remarkable thousand-fold improvement in training speed. This innovative approach has completely transformed the field of feature classification.

The recent advancement has provided the model with remarkable abilities in classification and generalization. More precisely, the pseudo ELM has demonstrated exceptional competence in handling large-scale multi-class classification tasks and has outperformed the most recent machine learning models [47], [52], [53], [54], [55], [56]. The significance of ELM resides in its unique and adaptable parameter initialization technique, that requires only one hidden layer between the input and hidden layers. The parameters connecting the hidden and output layers are identified by the utilization of the pseudoinverse approach. However, by substituting the ridge regression methodology with the pseudoinverse method, this work added a new degree of competence. This modification greatly improves the model's capacity to successfully learn and handle features, hence improving its potential for generalization and getting outstanding accuracy in comparison to the ELM. The architecture of the model comprises 512 nodes in the input layer and a powerful ensemble of 700 nodes in the hidden layer. In conclusion, the Ridge-ELM algorithm generated a total of four nodes that played a vital role in accurately categorizing various types of samples from waste images in the output layer. Figure 3 depicts the Ridge-ELM architecture alongside the proposed PLDs-CNN model for multiclass categorization. The explanation of Ridge-ELM algorithm is given in Algorithm 1.

Algorithm 1: Proposed Ridge-ELM algorithm for multi-class classification.

$$X_{(n,m)} = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & \cdots & x_{(1,m)} \\ x_{(2,1)} & x_{(2,2)} & \cdots & x_{(2,m)} \\ x_{(3,1)} & x_{(3,2)} & \cdots & x_{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(n,1)} & x_{(n,2)} & \cdots & x_{(n,m)} \end{bmatrix} \quad Y_{(n,t)} = \begin{bmatrix} y_{(1,1)} & y_{(1,2)} & \cdots & y_{(1,t)} \\ y_{(2,1)} & y_{(2,2)} & \cdots & y_{(2,t)} \\ y_{(3,1)} & y_{(3,2)} & \cdots & y_{(3,t)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{(n,1)} & y_{(n,2)} & \cdots & y_{(n,t)} \end{bmatrix}$$

$$W_{(m,N)} = \begin{bmatrix} w_{(1,1)} & w_{(1,2)} & \cdots & w_{(1,N)} \\ w_{(2,1)} & w_{(2,2)} & \cdots & w_{(2,N)} \\ w_{(3,1)} & w_{(3,2)} & \cdots & w_{(3,N)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(m,1)} & w_{(m,2)} & \cdots & w_{(m,N)} \end{bmatrix} H_{(n,N)} = \begin{bmatrix} h_{(1,1)} & h_{(1,2)} & \cdots & h_{(1,N)} \\ h_{(2,1)} & h_{(2,2)} & \cdots & h_{(2,N)} \\ h_{(3,1)} & h_{(3,2)} & \cdots & h_{(3,N)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{(n,1)} & h_{(n,2)} & \cdots & h_{(n,N)} \end{bmatrix}$$

The input matrix is denoted as X, and the output matrix is represented as Y.

1. $W_{(m,N)}$ is denoted as input weight and $B_{(1,N)}$ is denoted as bias matrices.

$$B_{(1,N)} = \begin{bmatrix} b_{(1,1)} & b_{(1,2)} & \cdots & b_{(1,N)} \end{bmatrix}$$

2. The second step is to find the output $H_{(n,N)}$ of the hidden layer.

$$H_{(n,N)} = G(X_{(n,m)} \cdot W_{(m,N)} + B_{(1,N)})$$

Where, G denotes an activation function in this context.

3. Determine the output weight matrix $\beta_{(N,t)}$ using pseudoinverse.

$$\beta_{(N,t)} = H^{\dagger}_{(N,n)} \cdot T_{(n,t)}$$

In this proposed hybrid Ridge regression, the pseudoinverse has been replaced by these equations:

$$A_{(N,N)} = H^{T}_{(N,n)} \cdot H_{(n,N)}$$

$$b_{(N,t)} = H^{T}_{(N,n)} \cdot T_{(n,t)}$$

$$C_{(N,N)} = A_{(N,N)} + \alpha . I_{(N,N)}$$

$$B_{(N,t)} = C^{-1}_{(N,N)} \cdot b_{(N,t)}$$

Where, α denotes regularization parameters.

4. Generate prediction $\beta_{(N,t)}$.

The model displayed a strong sense of assurance in its ability to produce precise and accurate final predictions. Ridge-ELM, a method that seamlessly integrated ridge regression into the ELM framework, achieved a perfect balance between effective feature learning and regularization. Consequently, the predictive capability of the model was enhanced by its greater ability to generalize and interpret intricate patterns in the data.

3.6. XAI

Explainable Artificial Intelligence (XAI) is important for enhancing the transparency and interpretability of the PLDs-CNN model. SHAP (Shapley Additive Explanations) was employed to tackle the "black box" nature of deep learning models, which usually makes them less understandable. Through the integration of the PLDs-CNN model and SHAP, the automatic garbage classification systems are now capable of making smarter and more efficient decisions when classifying waste items into four main classes and twelve different sub-classes [57]. This technique presents new opportunities for smarter waste management and more efficient categorization of waste materials across multiple categories. It was found that there was a clear pattern in the Shapley values that were employed in the study to measure the significance of individual pixels. The presence of red pixels enhances the accuracy of class identification, while the presence of blue pixels diminishes the probability of proper categorization [58]. The Shapley values were computed using Equation (1).

$$\emptyset_k = \sum_{M \subseteq N \setminus k} \frac{M|!\,(A - |M| - 1)!}{A!} [f_x(M \cup k) - f_x(M)] \quad (1)$$

The variable f_x denotes the impact on the output resulting from the Shapely values of a certain feature, k. Subset M comprises of all features that are included in feature N, excluding feature k. $\frac{M|!(A-|M|-1)!}{A!}$ represents the weighted factor of the subset M permutations. Equation (2) yields the predicted result, which is represented by the symbol $f_x\,(M)$,

$$f_x\,(M) = P[f(x)|x_M] \quad (2)$$

The SHAP method involves the substitution of every initial identifiable (x_k) with a binary value (b'_k) that denotes the presence or absence of x_k, as illustrated in Equation (3).

$$l\,(b') = \emptyset_0 + \sum_{k=1}^{A} \emptyset_k b'_k \quad (3)$$

In the proposed framework $f(x)$, the bias is denoted by $\emptyset_0$, and the feature's contribution is represented by $\emptyset_k b'_k$, where $l\,(b')$ represents the substitute model for the framework. The contribution of feature k to the outcome and the role of $\emptyset_k$ are essential aspects that aid in understanding the underlying mechanisms of the model.

3.7. Classification experiments and performance matrices

The deep learning algorithms and XAI were implemented using Keras, with TensorFlow as the backend, running on PyCharm Community Edition (2021.2.3) software. A computer equipped with an 11th generation Intel(R) Core (TM) i9-11900 CPU operating at a frequency of 2.50GHz, 128GB of RAM, and an NVIDIA GeForce RTX 3090 graphics processing unit with 24 GB of memory was utilized for both training and testing the model. The computer was running a 64-bit version of Windows 10 Pro.

Confusion matrix (CM) was used to measure the PLDs-CNN model's performance. The accuracy, precision, recall, f1-score, and area under the curve (AUC) from the CM were determined using the following formulas.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{4}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{5}$$

$$Recall = \frac{T_P}{T_N + F_P} \tag{6}$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{7}$$

$$AUC = \frac{1}{2}\left(\frac{T_P}{T_P + F_N} + \frac{T_N}{T_N + F_P}\right) \tag{8}$$

The symbols T_P, T_N, F_P, and F_N were used to denote true positives, true negatives, false positives, and false negatives, respectively.

4. Results

This study involved the evaluation of the PL-CNN and PLDs-CNN models across four classes (First stage) to assess their performance and compare the results with other state-of-the-art TL models. Subsequently, twelve-class classification (Second stage) was conducted using the PL-CNN and PLDs-CNN models, as these models demonstrated superior performance in the four-class classification compared to other models.

4.1. First Stage: Results for four class classification

4.1.1. PL-CNN-ELM and PL-CNN-Ridge-ELM

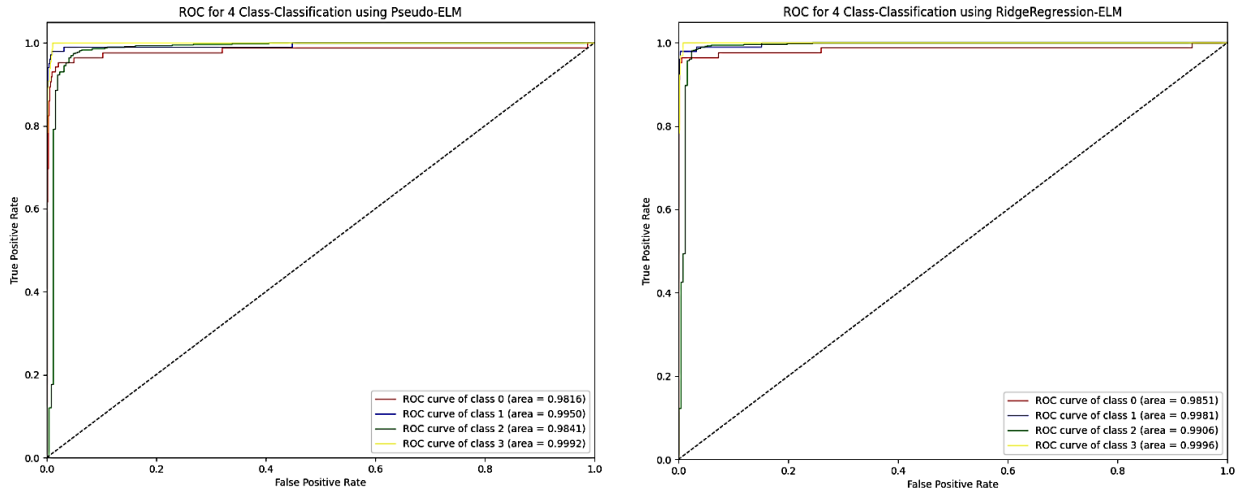
Initially, the PL-CNN model was employed to conduct training using a dataset including 15,150 images that encompassed 4 distinct waste categories (Figure 1 and Table 1). The PL-CNN (without DSC) was evaluated independently using a dataset including 1,552 test pictures. Both ELM and Ridge-ELM were utilized for evaluating class-specific performance of the PL-CNN. The findings of these evaluations are displayed in Table 3. The PL-CNN-ELM achieved an average test precision of 96.75±0.03%, recall of 90.5±0.05%, and f1-score of 93.5±0.03%. The accuracy and area under the curve (AUC) were 97% and 98.99%, respectively. The Ridge-ELM classifier on this model achieved an average precision of 97.5±0.02% (an improvement of 0.75%), recall of 94.5±0.03% (an improvement of 4%), and f1-score of 96±0.02% (an improvement of 2.5%). The mean scores for accuracy and AUC were 98% (with a 1%

improvement) and 99.33% (with a 0.34% improvement), respectively. The utilization of PL-CNN-Ridge-ELM may improve conventional ELM and reduce the misclassification rate. Figure 6 displayed the ROC curves for each class.

Table 3: Four class performances by using PL-CNN-ELM and PL-CNN-Ridge-ELM architecture.

Class Name	PL-CNN-ELM			PL-CNN-Ridge-ELM		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Hazardous Waste (0)	0.99	0.86	0.92	0.99	0.94	0.96
Household Food Waste (1)	0.98	0.89	0.93	0.98	0.93	0.96
Recyclable Waste (2)	0.98	0.99	0.99	0.99	0.99	0.99
Residual Waste (3)	0.92	0.88	0.90	0.94	0.92	0.93
Average (μ) ± SD (σ) (%)	96.75±0.03	90.5±0.05	93.5±0.03	**97.5±0.02**	**94.5±0.03**	**96±0.02**
Accuracy (%)	97.0			**98.0**		
AUC (%)	98.99			**99.33**		

*Bold values indicate the best results.



(A) **(B)**

Figure 6. Class-wise ROCs on (A) PL-CNN-ELM and (E) PL-CNN-Ridge-ELM for four class classification.

4.1.2. PLDs-CNN-ELM and PLDs-CNN-Ridge-ELM (Proposed method)

Table 4 illustrates the notable differences in performance between ELM and Ridge-ELM in the proposed parallel lightweight depth-wise separable CNN (PLDS-CNN) structure. To confirm the models' dependability and accuracy, a benchmark dataset which consists of twelve types of waste material images, was used. The CMs for both classifiers in the first stage classification (Four classes) are displayed in Figures 7A and 7B, providing important insights into the classification result. Ridge-ELM significantly reduced misclassification rates for Hazardous Waste (0), Household Food Waste (1), and Residual Waste (3). For each class, the matrices calculate precision, recall, and f1-scores. Ridge-ELM achieved an average precision of 97.25±0.02% across four categories. This classifier achieved a recall of 96±0.03% and a f1-score of 96.5±0.01%. These values represent improvements of 6.75% and 3.25% respectively compared to ELM. The accuracy score improved from 97% to 99%. The results indicate that Ridge-ELM successfully identifies the several classes from the samples. The Ridge-ELM model achieved an AUC of 99.51%, surpassing the AUC of the ELM model, which was 99.24%. The results presented here offer solid proof for the effectiveness of the proposed Ridge-ELM approach. This

analysis introduces a new and efficient model that utilizes a Ridge-ELM classifier. The model outperforms the traditional methods and other models in terms of performance and accuracy.

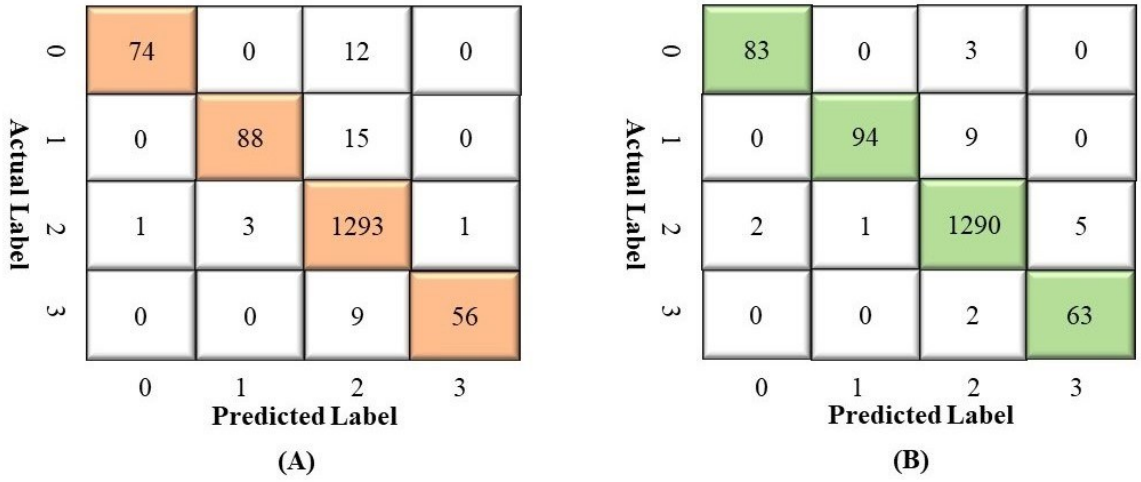


Figure 7. Confusion matrices of PLDs-CNN model with (A) ELM and (B) Ridge-ELM for four class classification.

Figure 8 provides a thorough evaluation of the performance of the PLDs-CNN-ELM and PLDs-CNN-Ridge-ELM models in accurately identifying four different waste categories. The ROC curves for each model demonstrate that both models exhibited outstanding performance, as indicated by the ROC Values above 98.57% for all grades. The average ROC score of 99.61% for the proposed Ridge-ELM combined with the PLDs-CNN model illustrates the model's exceptional accuracy and efficiency in categorizing garbage images. The findings reveal that the PLDs-CNN-Ridge-ELM model exhibits a high level of reliability and effectiveness in its classification performance. It is important to mention that the suggested model successfully achieved a 99.99% accuracy rate in recognizing Residual Waste (3). The study demonstrated that the suggested method effectively recognizes this specific type of garbage with a high level of accuracy. In addition, the ROC curve shows that the proposed model greatly enhanced the detection of Recyclable Waste (2). The increase in the detection rate from 98.57% to 99.39% resulted in an improvement. The proposed PLDs-CNN-Ridge-ELM approach demonstrated superior performance compared to all other methods in class-wise detection. There were not enough images for each group in the Garbage Classification dataset, so the dataset was not balanced. One of the primary goals during the model's development was to efficiently address this data imbalance. An adequate weighting system was implemented to guarantee that each class had an equal influence on the final result. The AUC values ranged from 0.9923 to 0.9999 for PLDs-CNN-Ridge-ELM in all classes, demonstrating that superior AUC outcomes were achieved even when dealing with imbalanced data sets (refer to Figure 8). The model's recall rate for Recyclable Waste (2) was 99%, demonstrating that the suggested framework can consistently differentiate every sort of waste material. The proposed framework has the capacity to greatly diminish hurdles faced by municipal authorities and improve waste management systems through our creative contributions.

Table 4: Four class performances by using PLDs-CNN-ELM and PLDs-CNN-Ridge-ELM architecture.

Class Name	PLDs-CNN-ELM			PLDs-CNN-Ridge-ELM		
	Precision	Recall	F1	Precision	Recall	F1
Hazardous Waste (0)	0.99	0.86	0.92	0.98	0.97	0.97
Household Food Waste (1)	0.97	0.85	0.91	0.99	0.91	0.95

Recyclable Waste (2)	0.97	1.00	0.98	0.99	0.99	0.99
Residual Waste (3)	0.98	0.86	0.92	0.93	0.97	0.95
Average (μ) ± SD (σ) (%)	**97.75±0.009**	89.25±0.07	93.25±0.03	97.25±0.02	**96±0.03**	**96.5±0.01**
Accuracy (%)	97.0			**99.0**		
AUC (%)	99.24			**99.51**		

*Bold values indicate the best results.

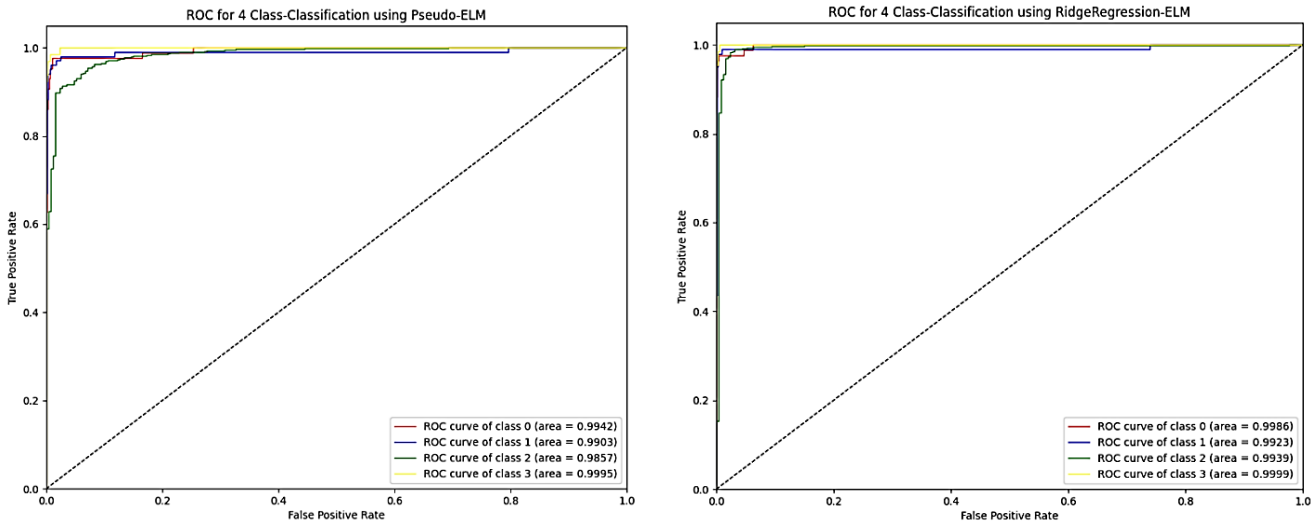


Figure 8. Class-wise ROCs on (A) PLDs-CNN-PELM and (E) PLDs-CNN-RELM for four class classification.

4.1.3. Performance comparison among proposed model with other TL models

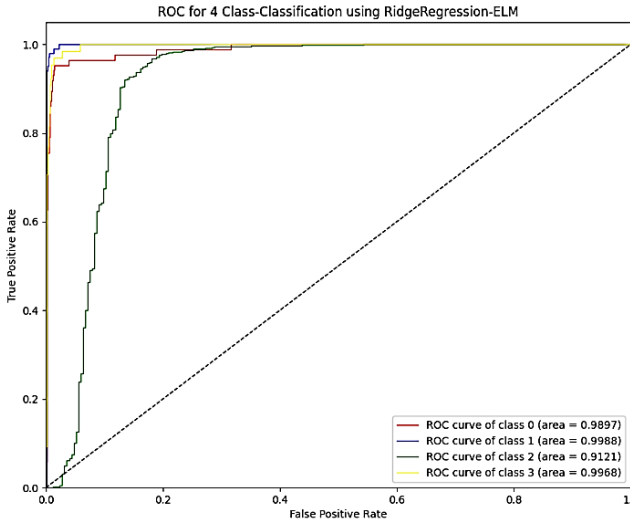
Table 5 compares the performance metrics of PLDs-CNN with other state-of-art TL models using the Ridge-ELM classifier in terms of overall four class classification results. Among the other seven TL models, VGG16 had the best classification results. The mean precision, recall, f1-score, accuracy, and AUC values were 95±0.01%, 90±0.08%, 92.25±0.04%, 97%, and 97.47% respectively. The EfficientNetB6 TL model has the lowest model performance. Among the other models, DenseNet201 achieved the highest mean precision of 98±0.018%. The proposed PLDs-CNN-Ridge-ELM achieved 99% accuracy for four class classification, which is roughly 2% higher than the VGG16. Aside from that, the AUC was 99.51%, which was better than the VGG16 (97.47%). Figure 9 depicts the ROC curve for the TL model for all four classes. Figure 10 shows a bar chart representation of the overall model performance. From the performance we can conclude that PLDs-CNN-Ridge-ELM demonstrates better results than other models.

Table 5. Four class classification performance of TL models and proposed model.

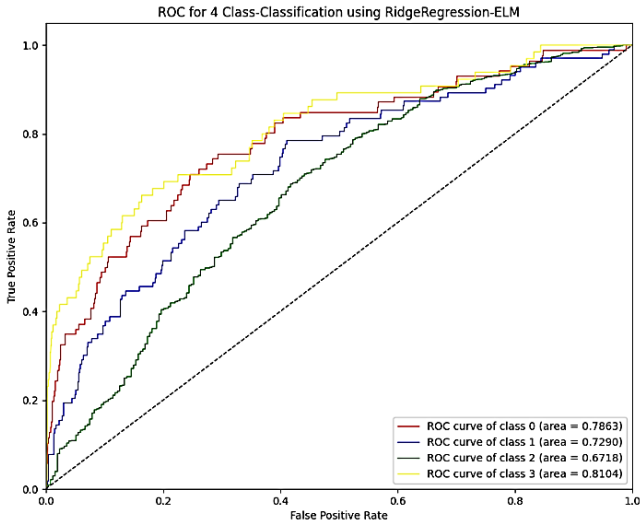
Model Name	**mP (%)**	**mR (%)**	**mF1 (%)**	**Accuracy (%)**	**AUC (%)**
DenseNet201-Ridge-ELM	**98±0.018**	85.25±0.119	90.5±0.066	97	96.85
EfficientNetB6-Ridge-ELM	35±0.42	26.5±0.49	25.5±0.43	84	74.93
InceptionResNetV2-Ridge-ELM	94.75±0.02	86.25±0.08	90.25±0.05	96	97.88
MobileNetV3Small-Ridge-ELM	80±0.11	61.25±0.26	68.25±0.19	89	91.06
ResNet152V2-Ridge-ELM	94.75±0.03	81.5±0.13	87±0.07	95	97.19
VGG16-Ridge-ELM	95±0.01	90±0.08	92.25±0.04	97	97.47
Xception-Ridge-ELM	96±0.02	82.75±0.11	88.5±0.06	96	96.72
PL-CNN-ELM	96.75±0.03	90.5±0.05	93.5±0.03	97	98.99

PLDs-CNN-ELM	97.75±0.009	89.25±0.07	93.25±0.03	97	99.24
PL-CNN-Ridge-ELM	97.75±0.02	94.5±0.03	96±0.02	98	99.33
PLDs-CNN-Ridge-ELM	97.25±0.02	**96±0.03**	**96.5±0.01**	**99.00**	**99.51**

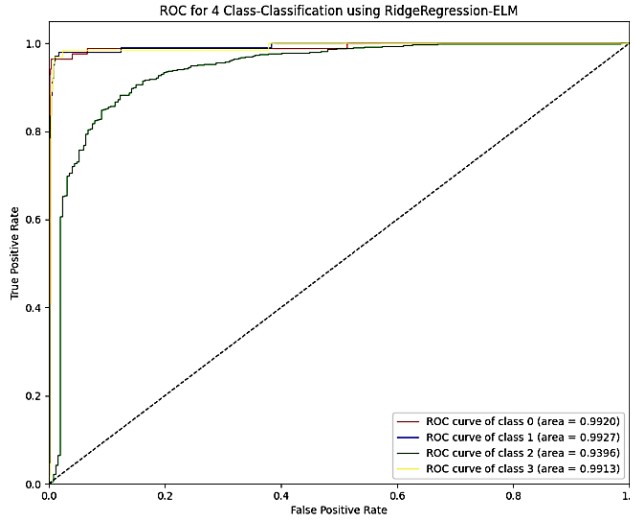
*Bold values indicate the best results.



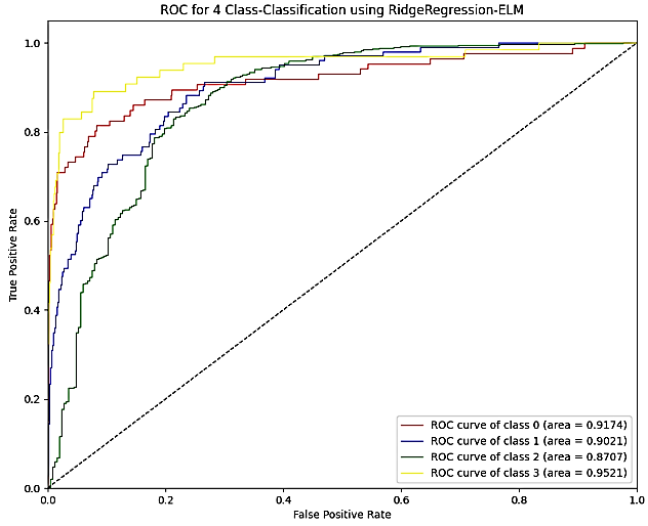
(A)



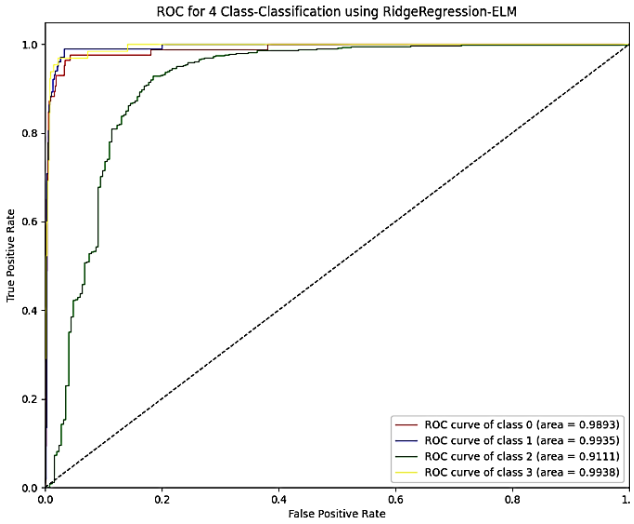
(B)



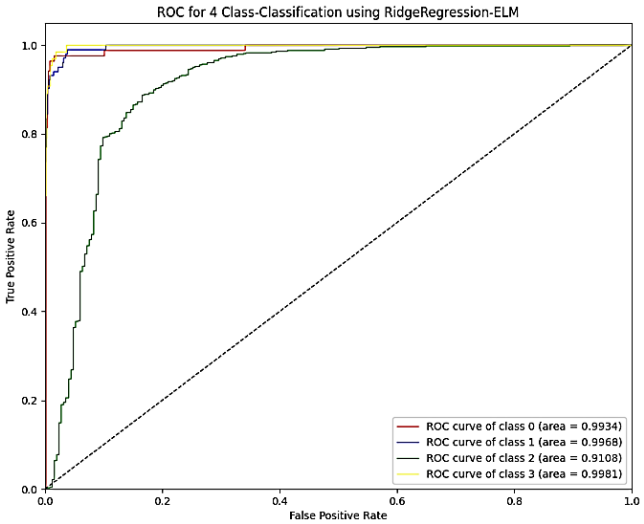
(C)



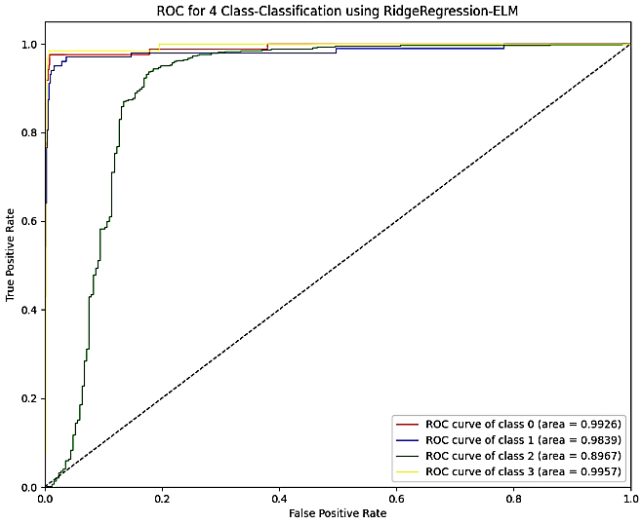
(D)



(E)



(F)



(G)

Figure 9. Class-wise ROCs on (A) DensNET201, (B) EfficientNetB6, (C) InceptionResNetV2, (D) MobileNetV3Small, (E) ResNet152V2, (F) VGG16, and (G) Xception with RELM for four class classification

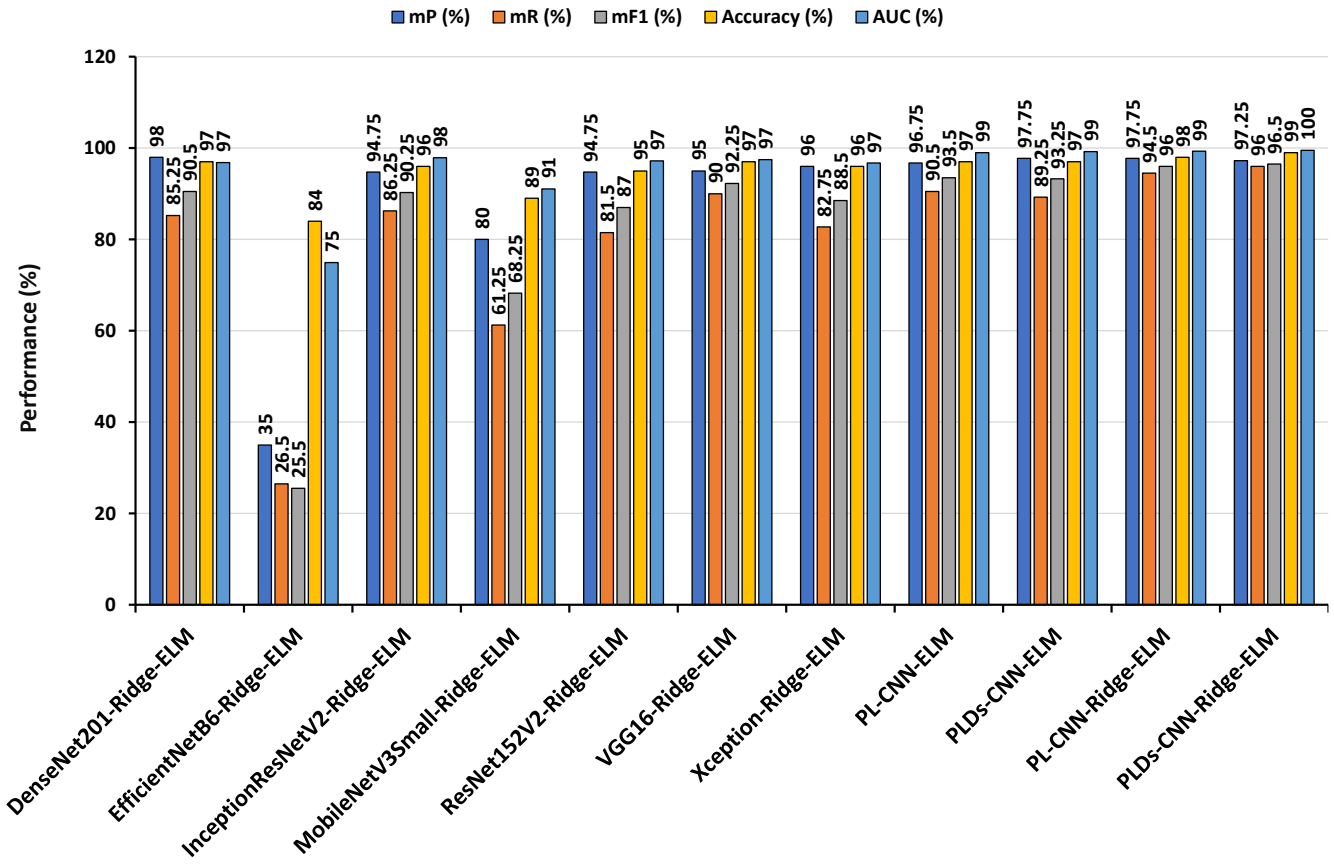


Figure 10. Proposed and TL models performances on (A) DenseNet201, (B) EfficientNetB6, (C) InceptionResNetV2, (D) MobileNetV3Small, (E)ResNet152V2, (F) VGG16, (G) Xception, (H) PL-CNN, and (I) PLDs-CNN with Ridge-ELM for four class classifications.

4.1.4. Computational time and resource comparison

Table 6 compares the computational resources of PLDs-CNN, PL-CNN, and other TL models based on model parameters, layers, sizes, training time, and testing time. Regarding classification accuracy, training and testing time, and model complexity, the PLDs-CNN model performed better than any other model, making it the most efficient model. Figure 11 illustrates the entire amount of effort and resources required. According to the study results, the PLDs-CNN-Ridge-ELM approach is reliable and can efficiently categorize various waste materials.

The ResNet152V2 model possesses the highest number of parameters compared to all other models, with a total of 92.41 million. It consists of 193 layers and has an overall size of 567,614 MB. The InceptionResNetV2 model has the biggest model size of 783,287 MB. It also has 783 layers which is the highest among all models. On the other hand, the PLDs-CNN model is particularly efficient in loading and processing because of its small size of 12.7 MB, 1.09 million parameters, and 9 CL. The proposed model has a total of 1.09 million trainable parameters, which is roughly 84.78 times fewer than the ResNet152V2 model and 2.15 times fewer than the PL-CNN model, which has 2.344 million trainable parameters. In addition, the proposed Ridge-ELM achieved a more appropriate and less consecutive processing time, with 0.1006 seconds for training and 0.0079 seconds for testing. However, because of performance considerations and model complexity, other TL models may have more satisfactory computational times. The PLDs-CNN model demonstrates efficient waste classification while minimizing resource utilization, despite its lower layer count and smaller sizes. The PLDs-CNN-Ridge-ELM model is an ideal method that achieves superior performance while preserving a compact and efficient structure, providing it well-suited for practical garbage management applications.

Table 6. Computational resources (model parameters and size) and time comparison for multiclass classifications.

Criteria	**PL-CNN-Ridge-ELM**	**PLDs-CNN-Ridge-ELM**	**DenseNet 201-Ridge-ELM**	**EfficientNetB6-Ridge-ELM**	**Inception ResNetV2-Ridge-ELM**	**MobileNetV3Small-Ridge-ELM**	**ResNet152V2-Ridge-ELM**	**VGG16-Ridge-ELM**	**Xception-Ridge-ELM**
Total Parameters (Million)	2.344	**1.09**	36.54	79.23	61.15	18.83	92.41	19.95	27.42
Trainable Parameters (Million)	2.341	**1.09**	18.22	38.27	6.81	17.30	34.08	5.24	6.55
Number of Layers	**9**	**9**	598	710	783	246	193	25	36
Size (Megabyte)	27	**12.7**	710280	669596	783287	246204	567614	22116	135470
Training Time (Ridge-ELM) (4-class) (seconds)	0.0657	0.1006	0.0239	**0.01591**	0.03024	0.02217	0.03450	0.01942	0.02722
Testing Time (Ridge-ELM) (4-class) (millisecond)	6	7.9	9.569	15.62	5	9.5	5	**3**	4.7

*Bold values indicate the best results.

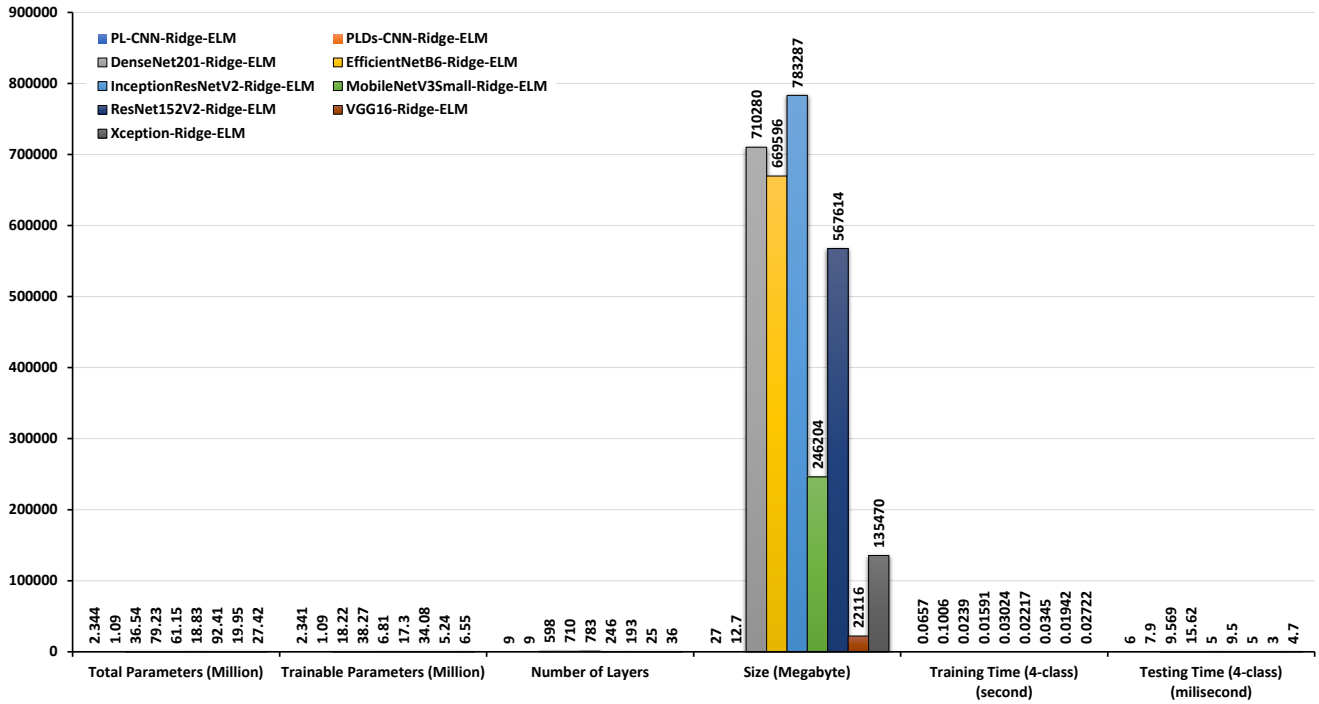


Figure 11. Computational resource (model parameters and size) and time comparisons among the proposed PLDs-CNN-Ridge-ELM and other TL models.

4.2. Second Stage: Results for twelve class classification

4.2.1. PL-CNN-ELM and PL-CNN-Ridge-ELM

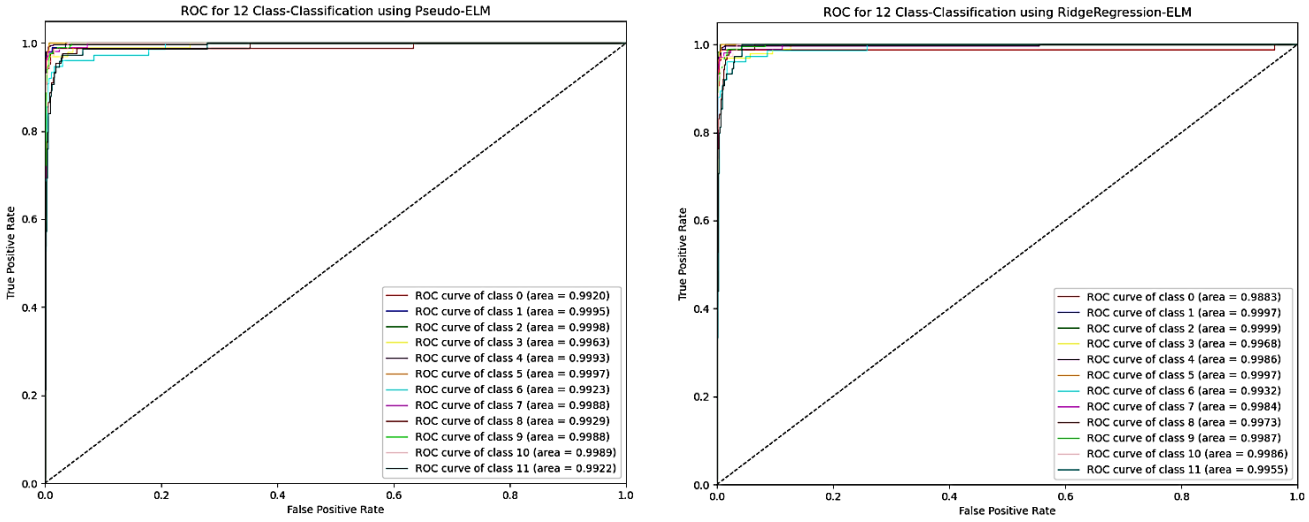
The same multiclass categorization technique was employed for twelve class classification in the second stage. The PL-CNN model performed training, testing, and validation on 12,566, 1,552, and 1,397 images, respectively. The models' classification performance is presented in Table 7. The PL-CNN-ELM model achieved precision, recall, and f1-scores of 94.25±0.04%, 93.75±0.04%, and 94±0.84%, respectively. The model's effectiveness was improved by utilizing Ridge-ELM. The test accuracy for both PL-CNN-ELM and PL-CNN-Ridge-ELM is 96%, indicating that they perform equally well. The PL-CNN-Ridge-ELM model improved the AUC by around 0.03%, increasing it from 99.67% to 99.70%. The precision, recall, and f1-score achieved the highest scores of 94.41±0.04% (with a 0.16% improvement), 93.75±0.04%, and 94.08±0.04% (with a 0.08% improvement), respectively. In Figure 12, the class-wise ROC curve is displayed for both classifiers on the PL-CNN model.

Table 7: Twelve class performances by using PL-CNN-ELM and PL-CNN-Ridge-ELM.

Class Name	PL-CNN-ELM			PL-CNN-Ridge-ELM		
	Precision	Recall	F1	Precision	Recall	F1
Battery (0)	0.94	0.93	0.94	0.97	0.90	0.93
Expired Food (1)	0.99	0.97	0.98	0.96	0.98	0.97
Brown Glass (2)	0.97	0.97	0.97	0.97	0.98	0.98
Cardboard (3)	0.97	0.95	0.96	0.97	0.96	0.96
Clothes (4)	0.98	1.00	0.99	0.98	1.00	0.99
Green Glass (5)	0.97	0.97	0.97	0.98	0.97	0.98
Metal (6)	0.88	0.88	0.88	0.89	0.89	0.89
Paper (7)	0.94	0.96	0.95	0.94	0.95	0.95
Plastic (8)	0.91	0.87	0.89	0.90	0.88	0.89
Shoes (9)	0.96	0.96	0.96	0.96	0.96	0.96

White Glass (10)	0.96	0.95	0.95	0.96	0.93	0.94
Trash (11)	0.84	0.84	0.84	0.85	0.85	0.85
Average (μ) ± SD (σ) (%)	94.25±0.04	93.75±0.04	94±0.84	**94.41±0.04**	**93.75±0.04**	**94.08±0.04**
Accuracy (%)	96.00			**96.00**		
AUC (%)	99.67			**99.70**		

*Bold values indicate the best results.



(A) **(B)**

Figure 12. Class-wise ROCs on (A) PL-CNN-ELM and (E) PL-CNN-Ridge-ELM for twelve class classification.

4.2.2. PLDs-CNN-ELM and PLDs-CNN-Ridge-ELM

Figure 13 illustrates essential information about the performance of both classifiers. Clearly, both classifiers performed very well during the second stage of classification. The precision, recall, and f1-scores demonstrated a high level of balance and surpassed 90% for the maximum category. The results provides solid evidence for the efficacy of the Ridge-ELM approach in successfully classifying multiclass waste images. The proposed method exhibits a substantial performance advantage over conventional models, thus demonstrating its superiority. Upon further examination, a novel and less heavy model is discovered, which incorporates a Ridge-ELM classifier, outperforming existing models in terms of both performance and accuracy. These findings establish a strong foundation for the application of the Ridge-ELM approach in waste image categorization. ELM showed an outstanding average precision of 94.66±0.034% across all 12 categories. ELM also achieved a recall rate of 93.66±0.036% and a f1-score of 94.5±0.033%. on the other hand, Ridge-ELM achieved improvements of 0.34%, 0.67%, and 0.16% in comparison to ELM. The accuracy score experienced a 1.0% improvement, rising from 95% to 96%. Figure 14 and Table 8 thoroughly examines the ability of the PLDs-CNN-ELM and PLDs-CNN-Ridge-ELM models to distinguish between 12 various waste types. The ROC curves for each category exhibit exceptional performance from both models, with ROC values surpassing 98% for the majority of classes. Every class identification record is considered satisfactory on the proposed PLDs-CNN-Ridge-ELM framework. This emphasizes the importance of the proposed system for practical implementation in real-world scenarios.

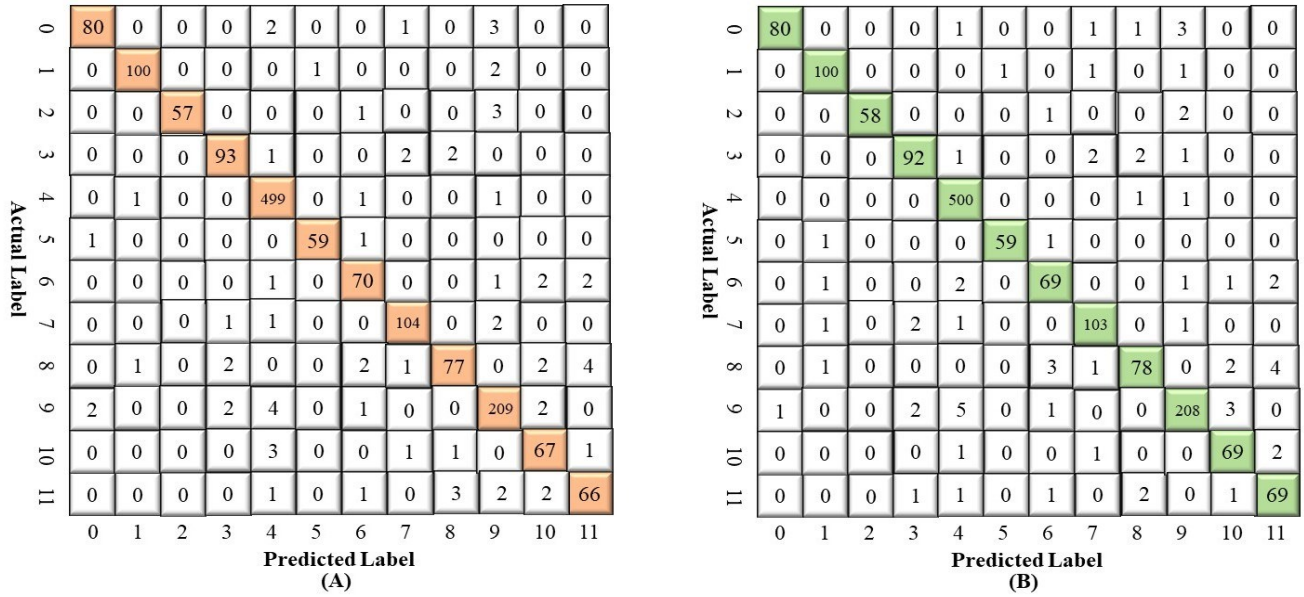


Figure 13. Confusion matrices of PLDs-CNN model with (A) ELM and (B) Ridge-ELM for twelve classes.

Table 8: Twelve class performances by using PLDs-CNN-PELM and PLDs-CNN-RELM architecture.

Class Name	PLDs-CNN-ELM			PLDs-CNN-Ridge-ELM		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Battery (0)	0.96	0.93	0.95	0.99	0.93	0.96
Expired Food (1)	0.98	0.97	0.98	0.96	0.97	0.97
Brown Glass (2)	1.00	0.93	0.97	1.00	0.95	0.97
Cardboard (3)	0.95	0.95	0.95	0.95	0.94	0.94
Clothes (4)	0.97	0.99	0.99	0.98	1.00	0.99
Green Glass (5)	0.98	0.97	0.98	0.98	0.97	0.98
Metal (6)	0.91	0.92	0.92	0.91	0.91	0.91
Paper (7)	0.95	0.96	0.96	0.94	0.95	0.95
Plastic (8)	0.93	0.87	0.90	0.93	0.88	0.90
Shoes (9)	0.94	0.95	0.94	0.95	0.95	0.95
White Glass (10)	0.89	0.92	0.91	0.91	0.95	0.93
Trash (11)	0.90	0.88	0.89	0.90	0.92	0.91
Average (μ) ± SD (σ) (%)	94.66±0.034	93.66±0.036	94.5±0.033	**95±0.033**	**94.33±0.031**	**94.66±0.02**
Accuracy (%)	95.00			**96.00**		
AUC (%)	**99.63**			99.59		

*Bold values indicate the best results.

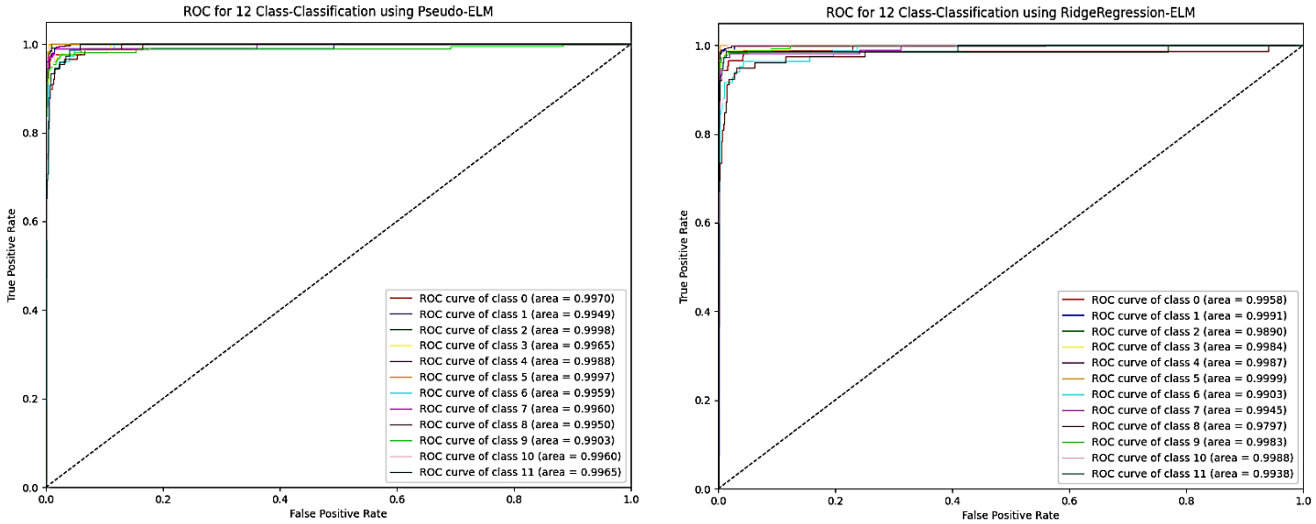


Figure 14. Class-wise ROCs on (A) PLDs-CNN-ELM and (E) PLDs-CNN-Ridge-ELM for twelve class classification.

4.3. Interpretability of PLDs-CNN-Ridge-ELM using SHAP

Through a systematic examination of every conceivable combination of wastage attributes, Shapley values were formulated, giving rise to representations characterized by pixels. A distinct pattern manifested during the investigation; wherein red pixels demonstrated robust efficacy in identifying class distinctions. In the first stage of the testing phase, SHAP results were provided with explanation images for different classifications. These classifications included four classes: hazardous, household, recyclable, and residual waste. The explanation images showed that red pixels indicated a higher probability of being close to the target class. In comparison, blue pixels indicated a higher probability of being distant from the target class. It is important to note that the SHAP explanation graphics have faint grey backgrounds merged with the original input images, as shown in Figure 15. The top row of the SHAP explanation image depicts red pixels, indicating the presence of hazardous waste. Conversely, the absence of blue pixels and fewer red pixels eliminate other class categories. The model identifies the hazardous class result where the red dots are defined with maximum confidence in that class. Where the red dots defined that model found maximum confidence to identify this as hazardous class result. The second row of the SHAP explanation images presented a distinct pattern: red pixels denoted the household food waste class. Conversely, for that sample, the maximum number of blue pixels was observed in recyclable waste, indicating that the confidence level for this category is comparatively lower than others.

The proposed model was tested in the second stage, which included 12 sub-classes, making it more complex to classify. Figure 16 shows how the model extracted the actual class during testing. The proposed model could accurately predict the outcome even when faced with increased data complexity. Visual SHAP explanations were used to confirm our model's results, providing a more comprehensive understanding of different classifications of wastage. These explanations helped to improve the system's understanding of the various forms of wastage.

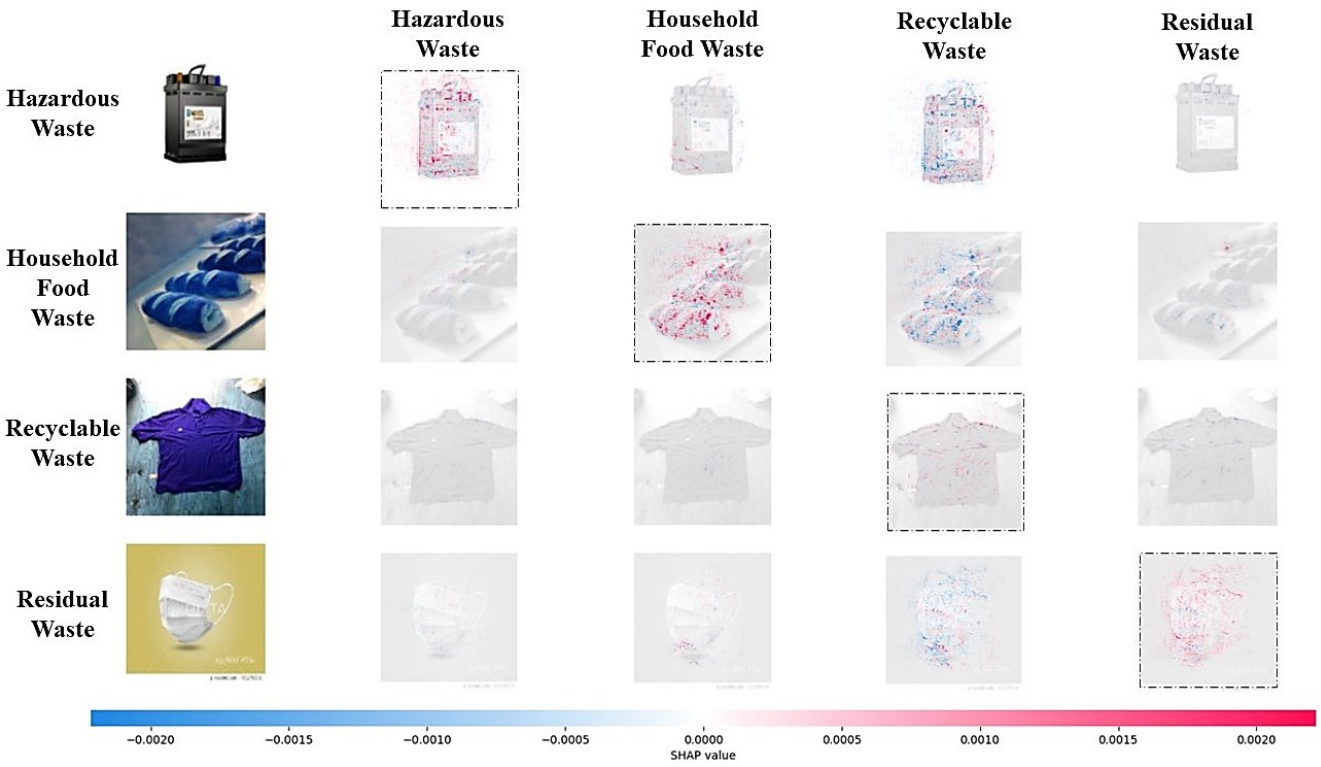


Figure 15. SHAP explanation images for PLDs-CNN-Ridge-ELM for four class classification.

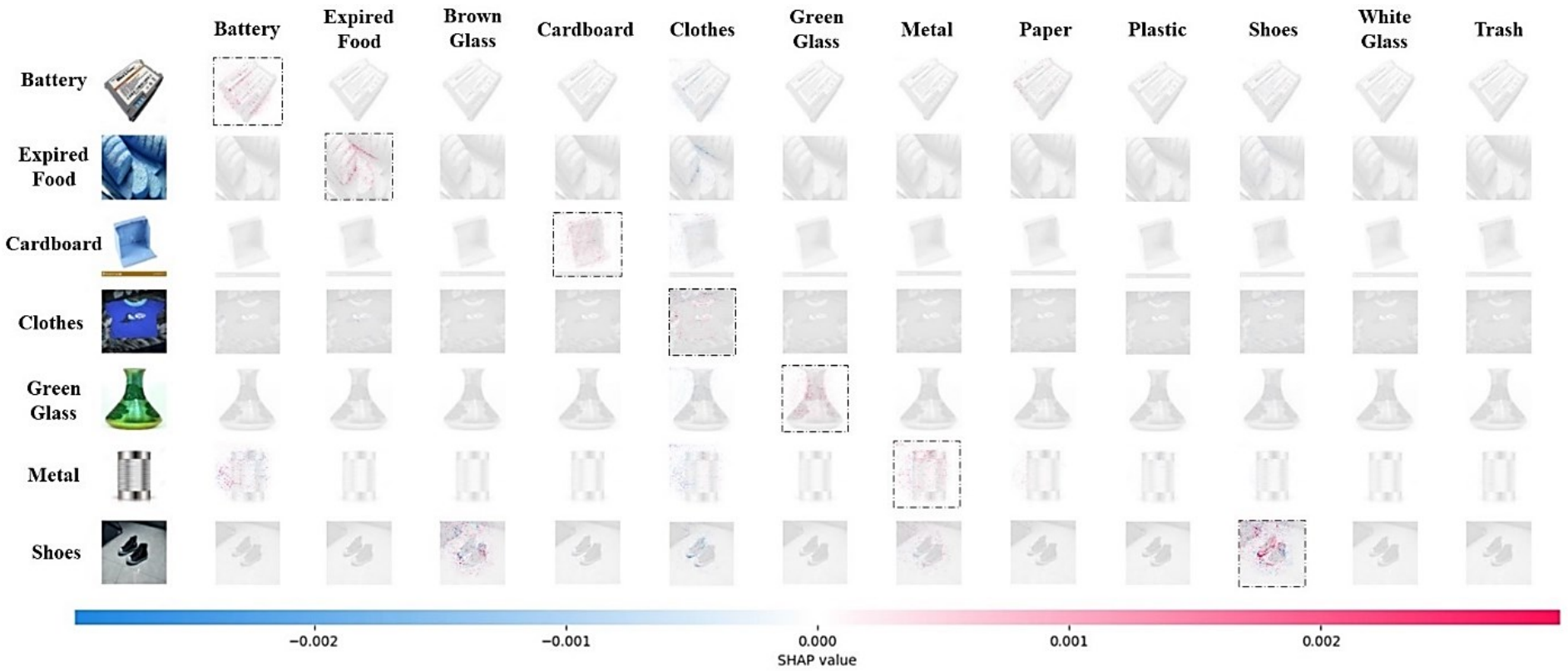


Figure 16. SHAP explanation images for PLDs-CNN-Ridge-ELM for twelve class classification.

4.4. Hardware Structure

4.4.1. Graphical User Interface (GUI)

For better testing and real-time implementation flexibility in the waste management industry, a graphical user interface (GUI) was designed based on PYQT5 used for the QT application framework. The GUI was programmed for three individual tasks, for four-class and twelve-class classifications of the three proposed models and their real-time classifications (Figure 17). The interface is generally designed to support conveyor belt-assisted waste sorting for four classes. For user flexibility, the tasks needed to be performed are made easily accessible by just clicking buttons without any need for loading files for every new runtime.

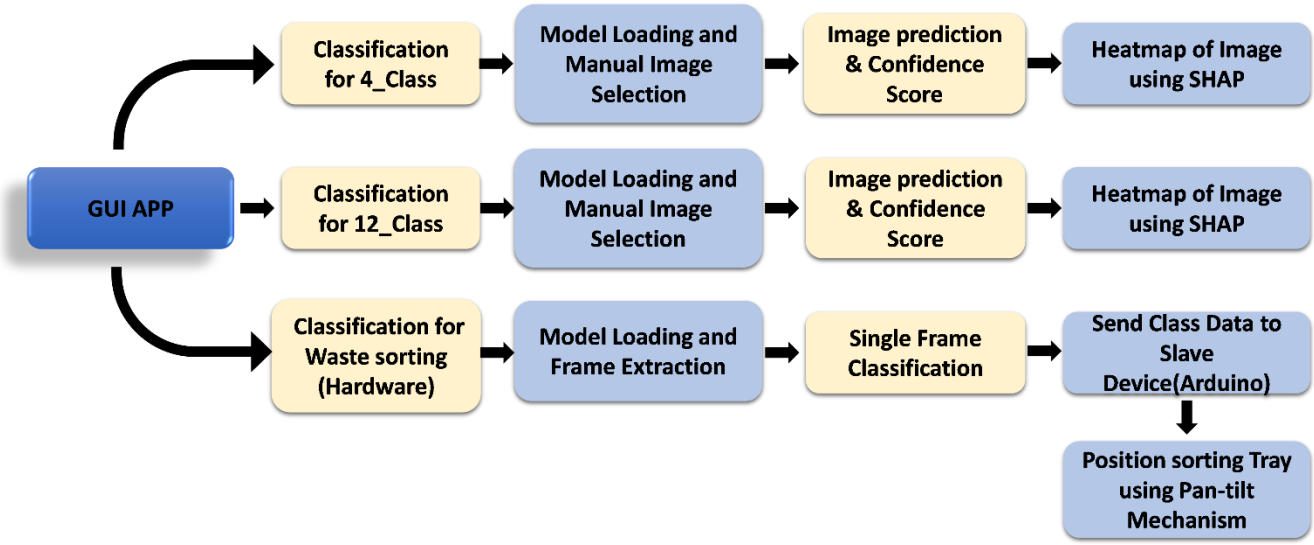
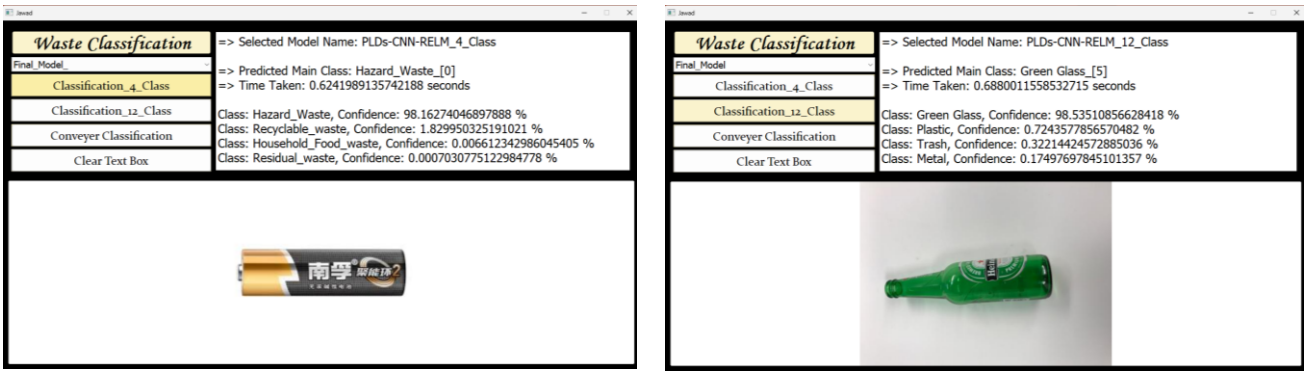


Figure 17. Flowchart of GUI application.

The app can test any of the three proposed models from the dropdown menu. After clicking the "Classification_4_Class" or "Classification_12_Class" button, the app will open a file dialogue for selecting an image for classification. Upon selecting the image, it will undergo several preprocessing tasks like resizing it to 124 × 124 to match the model's requirements. Followed by rescaling the image within 0 to 1 for faster convergence and reduced computational load. Finally, before predicting, the dimension of the numpy image array is increased to align it to the requirements of the model. The time taken for prediction is saved along with some other important classification results like the top 4 class names with corresponding confidence scores in Figure 18.



(A) **(B)**

Figure 18. (a) Classification of hazard items with inference time and confidence of top four class using a four-class model, (b) Classification of recyclable items with inference time and confidence of top four class with a 12-class model.

For understanding the model's decision making process for classifying images, the most influential parts of it are being highlighted using SHAP. This gives a better understanding on which features, the images are being predicted and this can also assist in debugging the model (Figure 19). The process starts by initializing the DeepExplainer using one of the models to identify the garbage object. The explanation process was based on this model, which acquired the ability to correlate certain pixel patterns with distinct classifications. In addition to the model, a background dataset was selected, which was a smaller portion of the training data. This dataset was essential since it embodied the standard input space of the model, functioning as a benchmark for comprehending the degree of novelty or conformity of a new input in relation to what the model encountered throughout its training. The subsequent step involved the calculation of SHAP values for a particular picture, which was the fundamental aspect of this procedure. The SHAP method measures the influence of each pixel on the model's conclusion by systematically altering the visible pixels [59]. This process effectively determined the contribution of each pixel to the final prediction.

(A) **(B)**

Figure 19. Visualization of significant features using SHAP for (a) 4-class model, (b) 12-class model for single image classification

4.4.2. Classification for Conveyer Belt Sorting Mechanism

Followed by accurate waste class prediction by the developed model and app, a real-world smart system can be implemented for sorting the waste automatically. The total system (Figure 20) will require a garbage chute (a) from dropping unsorted garbage items on the conveyer (b) belt. Upon passing over it, the waste will be detected by the motion triggered camera (e), which will use the image to predict the class the waste belongs in the edge device (c). Once the correct waste is identified, the tray (f) will direct it to its suitable bin by moving the pan-tilt mechanism. The system allows loading of a webcam and take multiple image frames of waste carried by a conveyer belt in a recycling plant and conduct classification to direct them to suitable bins for recycling or disposal. Based on the output class from the captured image, the computer sends a command to Arduino using seral communication that makes the tray turn accordingly to any of the four sides where the respective bins are placed. The waste is then sliding to the desired bin due to gravity. The proposed conveyer belt sorting mechanism is just an early concept for a larger and robust waste management system.

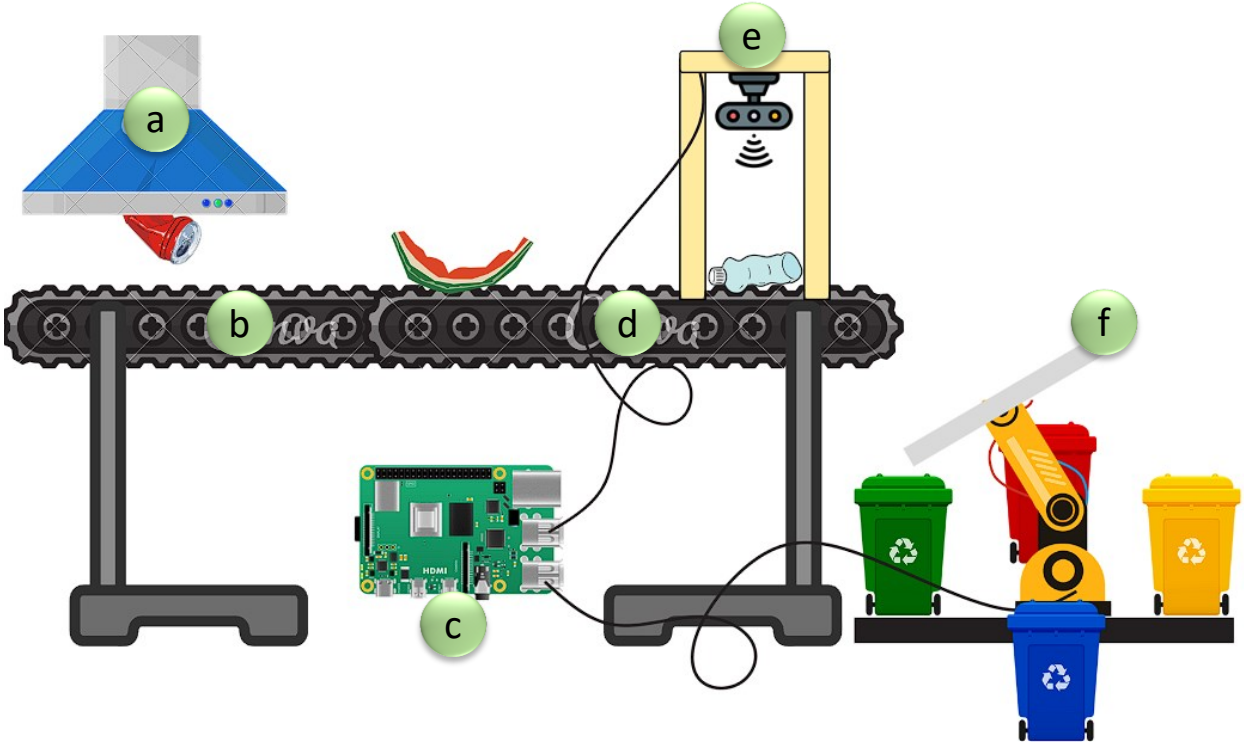


Figure 20. Concept of the total waste sorting mechanism. (a) Garbage Chute, (b) Conveyer Belt, (c) Edge Device, (d) Cable Wire, (e) Motion Sensor with Camera, (f) Pan-tilt joint for Trash sorting.

A circuit diagram of a conceptual system is presented in Figure 21 which is suitable for real-time usage.

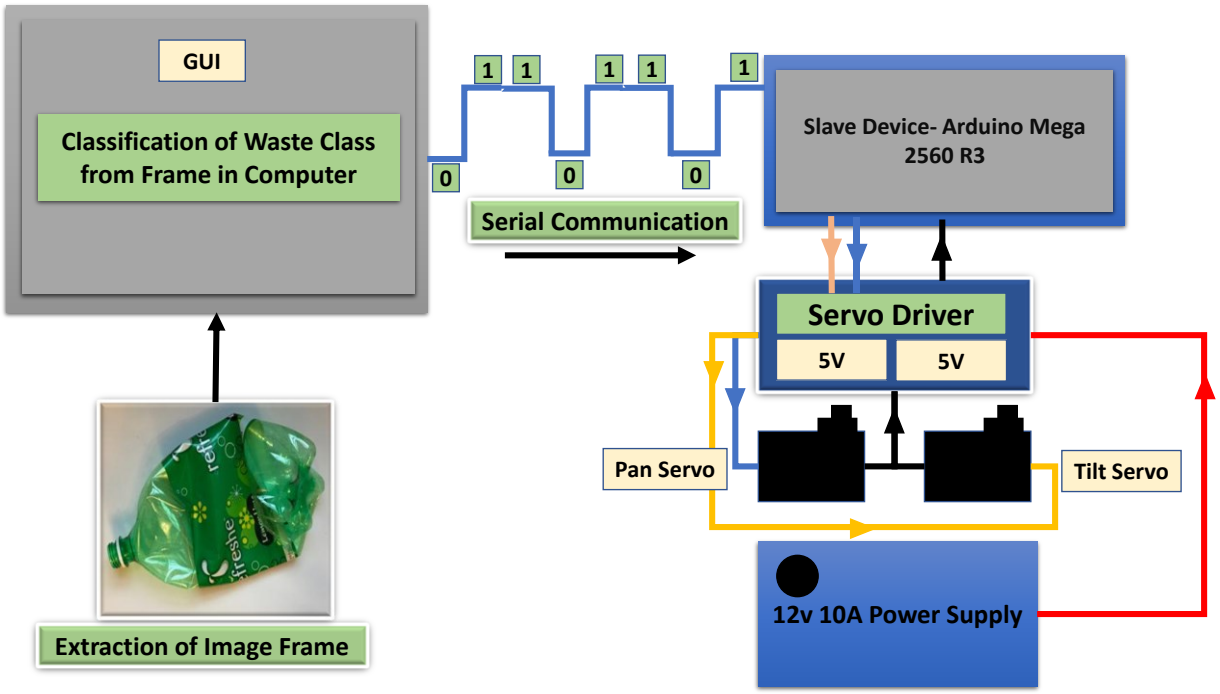


Figure 21: Circuit diagram of the hardware configuration with GUI app.

The preliminary design of the pan-tilt mechanism and its function has been programmed using two servo motors and Arduino as slave device respectively. The 2axis pan-tilt mounted servo assembly in Figure 22 was made using two 5vservos with 3D printed servo brackets. The designs were made on

Autodesk Tinkercad software with high precision allowing the tray to move in any direction. Its design and the coding for tray alignment were conducted for sorting four types of trash. The initial positions are set to 0 degree for pan servo and 90 degree for tilt servo. After predicting waste class for recyclable or residual waste, the tilt servo bracket/mount rotates left (Figure 22(d)) or right (Figure 22(b)) respectively from its initial position, while the pan servo bracket/mount remains fixed for both of them. Similarly for household food waste, tilt servo bracket/mount turns right (Figure 22(a)) and for hazard waste turns left (Figure 22(c)), but pan servo bracket/mount rotates 90 degree anti-clockwise, causing the tilt servo itself to rotate along with the tray. In this manner, depending on the predicted class, the tray is shifted to four different sides for diverting incoming waste.

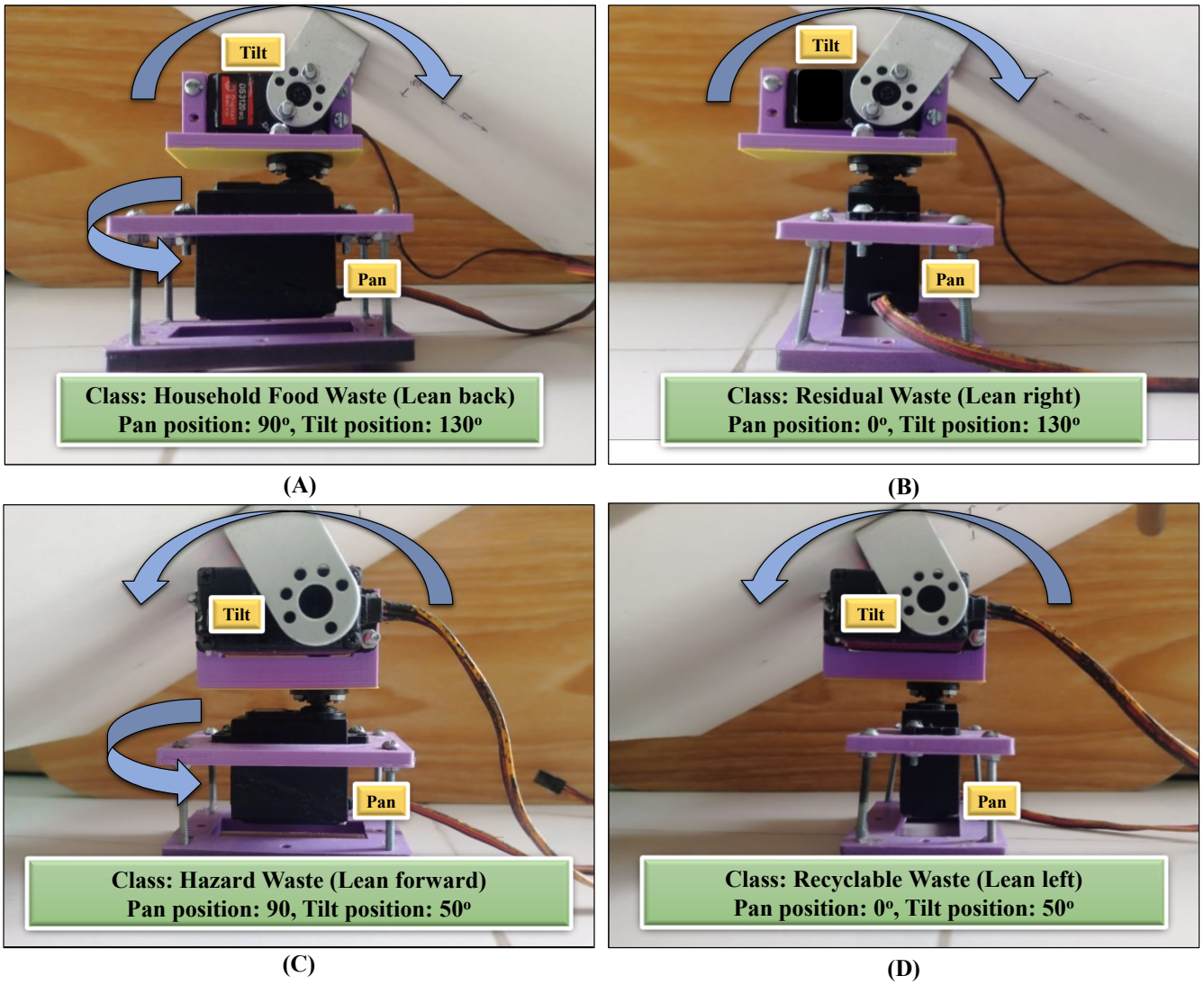


Figure 22. Different servo positions based on class using a pan-tilt mechanism. (A) Servo position for household food waste. (B) Servo position for recyclable waste. (C) Servo position for hazard waste. (D) Servo position for residual waste.

For a practical test on the functioning of the concept, a PowerPoint video of random waste images was accumulated and loaded in our app to create the same condition of wastes passing over a conveyor belt. Upon predicting each frame of waste from the video, the app shows the corresponding class (Figure

23(a) and 23(c)) of the current trash along with its confidence score and immediately sends a command to Arduino via serial communication to position the tray for the correct bin (Figure 23(b) and 23(c)).

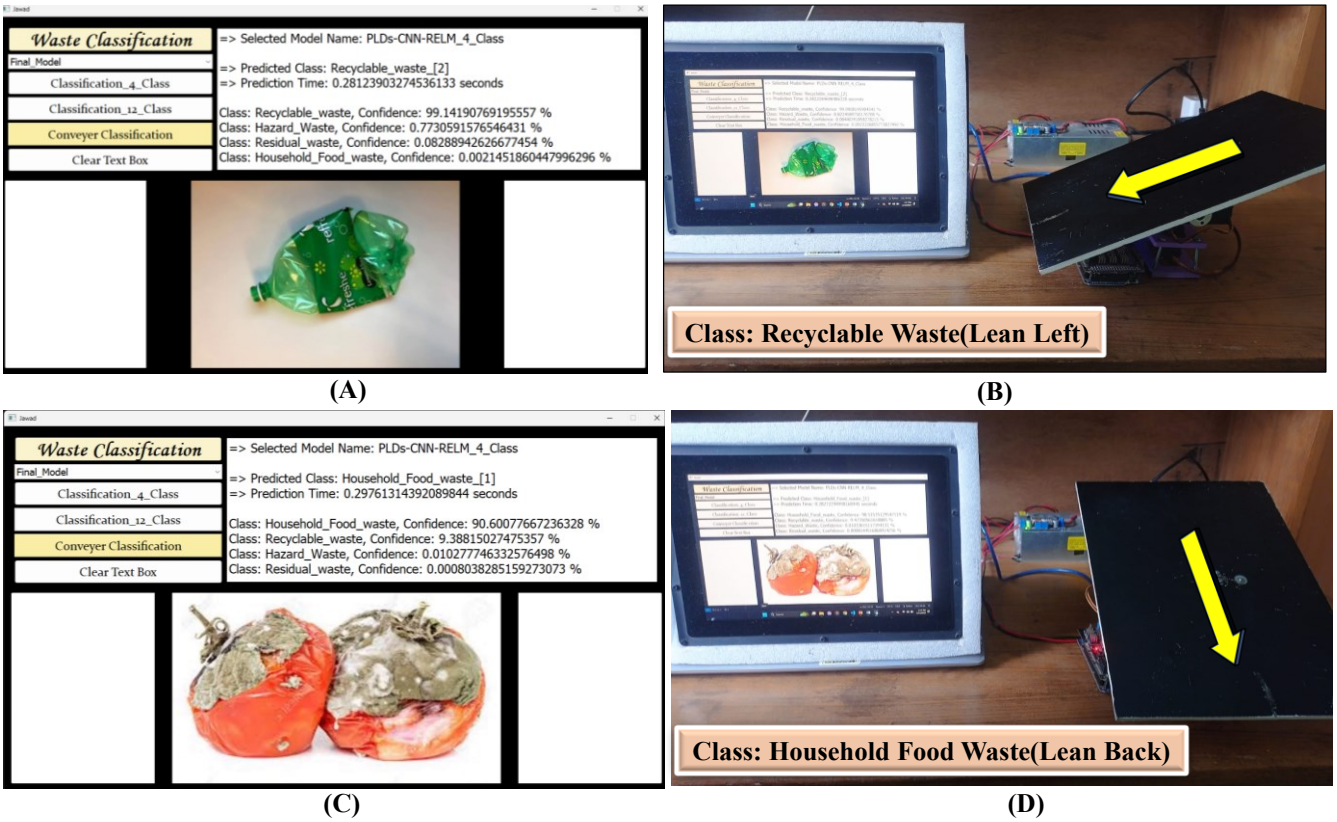


Figure 23. Continuous classification process test for real-time sorting and tray movement for (a), (b) Recyclable Waste, (c)(d) Household Food Waste.

4.5. Discussion and Future Work

Table 9 presents a summary of all state-of-art models alongside the proposed PLDs-CNN-Ridge-ELM model. The overview indicates that in [30], the optimized denseNet121 achieved the highest accuracy of 99.60%. However, it is important to note that they used the TrashNet dataset with six classes, comprising only 2527 images. Meanwhile, the proposed model achieved a comparable accuracy of 99.0% for four-class in the first stage, utilizing a more extensive dataset of 15,150 images. As the model was developed based on DenseNet121, the number of parameters was also much higher than that of the proposed model. ,Moreover the proposed model's testing time was only 0.0079 seconds for four-class classification.

Yang et al. [34] proposed a lightweight model with 1.5 million parameters. However, this model achieved lower classification accuracy (82.5%) for larger datasets like the Huawei garbage classification dataset. On the contrary, the proposed model showed excellent performance with 96% accuracy for twelve class classification on a large dataset with only 1.09 million parameters. Feng et al. [32] also proposed a lightweight model based on EfficientNet. However, the proposed PLDs-CNN-Ridge-ELM successfully outperformed this model in terms of accuracy, model parameters and number of images in the dataset. Additionally, Chen et al. [31] achieved a higher accuracy of 97.9%, exceeding the proposed model's performance. Nevertheless, it is essential to note that they conducted research on a smaller dataset (4256 images) and additionally their model had a higher number of parameters than the proposed model. Moreover, no studies demonstrated a real-time XAI while the proposed model

introduced SHAP. Table 9 also demonstrates that the reduced 4-class classification has 99% accuracy compared to 96% on the granular 12 classes, highlighting specialized sub-categories' problems [60]. Reducing the number of classes simplifies decision boundaries, enhancing the learning process and promoting better generalization. Although closely related groups may share similarities, merging them into broader categories introduces a more intricate decision boundary. In a fixed dataset, a lower number of classes can provide adequate features for effective generalization. Conversely, an increase in classes without a proportional rise in data and feature variability may lead to reduced model accuracy. However, the simplicity of decision boundaries in a smaller number of classes can help mitigate these challenges despite the limited feature set.

The lightweight PLDs-CNN model design is relatively straightforward, consisting of nine convolutional layers and three dense layers. However, to ensure the extraction of discriminant features, the first four convolutional layers are performed in parallel, reducing the total number of convolutional layers from nine to four. As shown in Table 5, the proposed model surpasses the other seven TL models regarding classification performance. The primary objective of this research was to develop a model that will enhance classification outcomes yet minimize the number of parameters and network layers. This was achieved by utilizing DS CLs instead of traditional CLs and adapting them to produce cost-effective devices. In addition, the introduction of SHAP aims to ensure that the proposed model accurately identifies the relevant regions of an image to extract valuable features while overlooking irrelevant parts. This improves the model's interpretability for common people, enabling them to classify waste efficiently and accurately. Moreover, it can assist in developing the ability to differentiate between different types of waste.

Table 9. Summary of the state-of-art models and proposed model.

Ref.	Dataset	Number of Class	Best Model	Testing Accuracy	Best Model's Parameters	Testing Time	Real-time XAI
Mao et al. [30]	TrashNet (2527 images)	6	Optimized DenseNet121	99.60%	7.2 million	---	No
Nowakowski et al. [35]	Custom (210 images)	3	Deep CNN and R-CNN	96.7%	---	---	No
Khan et al. [36]	Kaggle Garbage Categorization Dataset (750 images)	6	RWC-EPODL	98.96%	---	---	No
Lin et al.[37]	TrashNet (2527 images)	6	RWNet	88.8%	58.5 million	---	No
Yang et al. [34]	TrashNet (2527 images), Huawei Garbage Classification Dataset (18079 images)	6 (TrashNet), 4 (Huawei Garbage Classification Dataset)	WasNet	96.10% (For TrashNet), 82.5% (For Huawei Garbage Classification Dataset)	1.5 million	---	No
Chen et al.[31]	Custom (4256 images)	4 (14 sub-classes)	GCNet (Improved ShuffleNetv2)	97.9% (For 14 sub-class)	1.3 million	---	No
Fan et	Huawei Cloud	4	EfficientNetB	93.38%	7.8	6.756 seconds	No

al.[38]	Garbage Classification Dataset (14,802 images)		2 with PMAM		million		
Feng et al. [32]	Custom (7361 images)	4 (18 sub-classes)	GECM-EfficienNet	94.54% (For 18 sub-class)	1.23 million	---	No
Jin et al.[39]	Huawei Garbage Classification Challenge Cup Dataset (14,683 images)	4	Improved MobileNetV2	90.7 %	3.4 million	---	No
Zhang et al.[40]	Custom (1040 images)	4 (13 sub-classes)	RevM	94.71% (For 4 class)	---	---	No
Proposed Model	Kaggle Garbage Classification Dataset (15,150 images)	4 (12 sub-classes)	PLDs-CNN-Ridge-ELM	**99.0%** (For 4 class), 96.0% (For 12 sub-class)	1.09 million	**0.0079 seconds (For 4 class), 0.0041 seconds (For 12 sub-class)**	**Yes**

The study was conducted on existing dataset collected from Kaggle. In the future, this research will be extended by collecting more images of garbage to create a custom dataset and evaluating the model's performance on that dataset. The intension is to compare the performance of custom dataset-trained model with that of conventional datasets. This comparative analysis will provide insights into the generalizability and domain-specific effectiveness of our proposed models. Also, the current designed hardware will be modified further for real time classification in the waste management industry for autonomous waste sorting. The design will be built upon an edge device for better portability and maintenance in real life use. A modified robotic arm mechanism will be applied for quicker and accurate sorting of waste along with tracking and management of waste using IoT. The established approach can also be applied to identify additional categories of waste, such as electronic waste, industrial waste, and medical waste.

5. Conclusions

The presented research offered a novel method for reliably classifying various types of waste by utilizing the PLDs-CNN model and the Ridge-ELM classifier in combination. The suggested PLDs-CNN architecture, consisting of nine layers with only 1.09 million parameters, effectively classifies four and twelve types of waste while minimizing computing load. The testing times of the model were only 0.0079 seconds and 0.0041 seconds for the four-class and twelve-class classifications. The inclusion of the hybrid Ridge-ELM model, which replaces the pseudo ridge regression approach, greatly improves the classification performance. This innovative technique achieves exceptional levels of accuracy, reaching 99% for 4 class classifications in the initial stage and 96% for 12 class classifications in the second stage. The PLDs-CNN-Ridge-ELM model exhibits exceptional accuracy in identifying different types of waste, as evidenced by the precision, recall, and f1-scores of 95±0.033%, 94.33±0.031%, and 94.66±0.02%, respectively, for 12-class classification. Additionally, the model achieves a remarkable AUC score of 99.59%. Moreover, the proposed model's size was only 12.7MB

which makes this approach more accessible and applicable for real-life waste management tools because of its low computational requirements, which facilitate the easy implementation of affordable edge devices. The addition of real-time SHAP significantly benefits consumers as it enables a dependable interpretation of the model's output. The transparency and comprehensibility of SHAP enhance the reliability of the categorization outcomes. In summary, the suggested PLDs-CNN-Ridge-ELM technique greatly improves the accuracy of waste classification and is easy to implement and analyze in real life operation.

Acknowledgement

The authors extend their appreciation to AI tools like ChatGPT, QuillBot, and Grammarly for their invaluable assistance in refining and enhancing the clarity of the writing.

References

[1] D. Abuga and N. S. Raghava, “Real-time smart garbage bin mechanism for solid waste management in smart cities,” *Sustain Cities Soc*, vol. 75, p. 103347, Dec. 2021, doi: 10.1016/j.scs.2021.103347.

[2] K. Özkan, S. Ergin, Ş. Işık, and İ. Işıklı, “A new classification scheme of plastic wastes based upon recycling labels,” *Waste Management*, vol. 35, pp. 29–35, Jan. 2015, doi: 10.1016/j.wasman.2014.09.030.

[3] D. Hoornweg and P. Bhada-Tata, “What a waste: a global review of solid waste management,” *Urban Dev Ser Knowl Pap*, vol. 15, pp. 87–88, Jan. 2012.

[4] M. K. Jaunich, J. W. Levis, J. F. DeCarolis, M. A. Barlaz, and S. R. Ranjithan, “Solid Waste Management Policy Implications on Waste Process Choices and Systemwide Cost and Greenhouse Gas Performance,” *Environ Sci Technol*, vol. 53, no. 4, pp. 1766–1775, Feb. 2019, doi: 10.1021/acs.est.8b04589.

[5] S. Kaza and P. Bhada-Tata, *Decision Maker’s Guides for Solid Waste Management Technologies*. World Bank, Washington, DC, 2018. doi: 10.1596/31694.

[6] Y. Li and X. Zhang, “Intelligent X-ray waste detection and classification via X-ray characteristic enhancement and deep learning,” *J Clean Prod*, vol. 435, p. 140573, Jan. 2024, doi: 10.1016/j.jclepro.2024.140573.

[7] N. L. Harris *et al.*, “Global maps of twenty-first century forest carbon fluxes,” *Nat Clim Chang*, vol. 11, no. 3, pp. 234–240, Mar. 2021, doi: 10.1038/s41558-020-00976-6.

[8] S. Wang, J. Wang, S. Yang, J. Li, and K. Zhou, “From intention to behavior: Comprehending residents’ waste sorting intention and behavior formation process,” *Waste Management*, vol. 113, pp. 41–50, Jul. 2020, doi: 10.1016/j.wasman.2020.05.031.

[9] S. Kaza, L. C. Yao, P. Bhada-Tata, and F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC: World Bank, 2018. doi: 10.1596/978-1-4648-1329-0.

[10] C. G. Cheah, W. Y. Chia, S. F. Lai, K. W. Chew, S. R. Chia, and P. L. Show, “Innovation designs of industry 4.0 based solid waste management: Machinery and digital circular economy,” *Environ Res*, vol. 213, p. 113619, Oct. 2022, doi: 10.1016/j.envres.2022.113619.

[11] X. Lu, X. Pu, and X. Han, “Sustainable smart waste classification and collection system: A bi-objective modeling and optimization approach,” *J Clean Prod*, vol. 276, p. 124183, Dec. 2020, doi: 10.1016/j.jclepro.2020.124183.

[12] K. D. Kang, H. Kang, I. M. S. K. Ilankoon, and C. Y. Chong, “Electronic waste collection systems using Internet of Things (IoT): Household electronic waste management in Malaysia,” *J Clean Prod*, vol. 252, p. 119801, Apr. 2020, doi: 10.1016/j.jclepro.2019.119801.

[13] B. Carrera, V. L. Piñol, J. B. Mata, and K. Kim, “A machine learning based classification models for plastic recycling using different wavelength range spectrums,” *J Clean Prod*, vol. 374, p. 133883, Nov. 2022, doi: 10.1016/j.jclepro.2022.133883.

[14] P. Jiang *et al.*, “Blockchain technology applications in waste management: Overview, challenges and opportunities,” *J Clean Prod*, vol. 421, p. 138466, Oct. 2023, doi: 10.1016/j.jclepro.2023.138466.

[15] Md. Ashikuzzaman and Md. H. Howlader, “Sustainable Solid Waste Management in Bangladesh,” 2020, pp. 35–55. doi: 10.4018/978-1-7998-0198-6.ch002.

[16] S. Shams, J. N. Sahu, S. M. S. Rahman, and A. Ahsan, “Sustainable waste management policy in Bangladesh for reduction of greenhouse gases,” *Sustain Cities Soc*, vol. 33, pp. 18–26, Aug. 2017, doi: 10.1016/j.scs.2017.05.008.

[17] A. Ahsan, M. Alamgir, M. M. El-Sergany, S. Shams, M. K. Rowshon, and N. N. N. Daud, “Assessment of Municipal Solid Waste Management System in a Developing Country,” *Chinese Journal of Engineering*, vol. 2014, pp. 1–11, Mar. 2014, doi: 10.1155/2014/561935.

[18] D. T. Jerin *et al.*, “An overview of progress towards implementation of solid waste management policies in Dhaka, Bangladesh,” *Heliyon*, vol. 8, no. 2, p. e08918, Feb. 2022, doi: 10.1016/j.heliyon.2022.e08918.

[19] Md. A. Habib, M. M. Ahmed, M. Aziz, Mohd. R. A. Beg, and Md. E. Hoque, “Municipal Solid Waste Management and Waste-to-Energy Potential from Rajshahi City Corporation in Bangladesh,” *Applied Sciences*, vol. 11, no. 9, p. 3744, Apr. 2021, doi: 10.3390/app11093744.

[20] A. K. M. Haque and S. Razy, “Practices of 3Rs (Reduce, Reuse and Recycle) Strategy in Urban Solid Waste Management in Rajshahi City Corporation of Bangladesh,” vol. 23, pp. 75–87, Feb. 2021.

[21] Z. Kang, J. Yang, G. Li, and Z. Zhang, “An Automatic Garbage Classification System Based on Deep Learning,” *IEEE Access*, vol. 8, pp. 140019–140029, 2020, doi: 10.1109/ACCESS.2020.3010496.

[22] K. Lin *et al.*, “Toward smarter management and recovery of municipal solid waste: A critical review on deep learning approaches,” *J Clean Prod*, vol. 346, p. 130943, Apr. 2022, doi: 10.1016/j.jclepro.2022.130943.

[23] Y.-L. Zhang, Y.-C. Kim, and G.-W. Cha, “Assessment of deep learning-based image analysis for disaster waste identification,” *J Clean Prod*, vol. 428, p. 139351, Nov. 2023, doi: 10.1016/j.jclepro.2023.139351.

[24] Y. Qiao, Q. Zhang, Y. Qi, T. Wan, L. Yang, and X. Yu, “A Waste Classification model in Low-illumination scenes based on ConvNeXt,” *Resour Conserv Recycl*, vol. 199, p. 107274, Dec. 2023, doi: 10.1016/j.resconrec.2023.107274.

[25] Q. Zhang *et al.*, “A multi-label waste detection model based on transfer learning,” *Resour Conserv Recycl*, vol. 181, p. 106235, Jun. 2022, doi: 10.1016/j.resconrec.2022.106235.

[26] Y. Chen *et al.*, “Classification and recycling of recyclable garbage based on deep learning,” *J Clean Prod*, vol. 414, p. 137558, Aug. 2023, doi: 10.1016/j.jclepro.2023.137558.

[27] W.-L. Mao, W.-C. Chen, H. I. K. Fathurrahman, and Y.-H. Lin, “Deep learning networks for real-time regional domestic waste detection,” *J Clean Prod*, vol. 344, p. 131096, Apr. 2022, doi: 10.1016/j.jclepro.2022.131096.

[28] J. Yang, Y.-P. Xu, P. Chen, J.-Y. Li, D. Liu, and X.-L. Chu, “Combining spectroscopy and machine learning for rapid identification of plastic waste: Recent developments and future prospects,” *J Clean Prod*, vol. 431, p. 139771, Dec. 2023, doi: 10.1016/j.jclepro.2023.139771.

[29] Q. Zhang *et al.*, “Recyclable waste image recognition based on deep learning,” *Resour Conserv Recycl*, vol. 171, p. 105636, Aug. 2021, doi: 10.1016/j.resconrec.2021.105636.

[30] W.-L. Mao, W.-C. Chen, C.-T. Wang, and Y.-H. Lin, “Recycling waste classification using optimized convolutional neural network,” *Resour Conserv Recycl*, vol. 164, p. 105132, Jan. 2021, doi: 10.1016/j.resconrec.2020.105132.

[31] Z. Chen, J. Yang, L. Chen, and H. Jiao, “Garbage classification system based on improved ShuffleNet v2,” *Resour Conserv Recycl*, vol. 178, p. 106090, Mar. 2022, doi: 10.1016/j.resconrec.2021.106090.

[32] Z. Feng, J. Yang, L. Chen, Z. Chen, and L. Li, “An Intelligent Waste-Sorting and Recycling Device Based on Improved EfficientNet,” *Int J Environ Res Public Health*, vol. 19, no. 23, p. 15987, Nov. 2022, doi: 10.3390/ijerph192315987.

[33] S. Gaba, I. Budhiraja, V. Kumar, S. Garg, G. Kaddoum, and M. M. Hassan, “A federated calibration scheme for convolutional neural networks: Models, applications and challenges,” *Comput Commun*, vol. 192, pp. 144–162, Aug. 2022, doi: 10.1016/j.comcom.2022.05.035.

[34] Z. Yang and D. Li, “WasNet: A Neural Network-Based Garbage Collection Management System,” *IEEE Access*, vol. 8, pp. 103984–103993, 2020, doi: 10.1109/ACCESS.2020.2999678.

[35] P. Nowakowski and T. Pamuła, “Application of deep learning object classifier to improve e-waste collection planning,” *Waste Management*, vol. 109, pp. 1–9, May 2020, doi: 10.1016/j.wasman.2020.04.041.

[36] A. I. Khan, A. S. Almalaise Alghamdi, Y. B. Abushark, F. Alsolami, A. Almalawi, and A. Marish Ali, “Recycling waste classification using emperor penguin optimizer with deep learning model for bioenergy production,” *Chemosphere*, vol. 307, p. 136044, Nov. 2022, doi: 10.1016/j.chemosphere.2022.136044.

[37] K. Lin *et al.*, “Applying a deep residual network coupling with transfer learning for recyclable waste sorting,” *Environmental Science and Pollution Research*, vol. 29, no. 60, pp. 91081–91095, Dec. 2022, doi: 10.1007/s11356-022-22167-w.

[38] H. Fan *et al.*, “Raspberry Pi-based design of intelligent household classified garbage bin,” *Internet of Things*, vol. 24, p. 100987, Dec. 2023, doi: 10.1016/j.iot.2023.100987.

[39] S. Jin, Z. Yang, G. Królczykg, X. Liu, P. Gardoni, and Z. Li, “Garbage detection and classification using a new deep learning-based machine vision system as a tool for sustainable waste recycling,” *Waste Management*, vol. 162, pp. 123–130, May 2023, doi: 10.1016/j.wasman.2023.02.014.

[40] S. Zhang, Y. Chen, Z. Yang, and H. Gong, “Computer Vision Based Two-stage Waste Recognition-Retrieval Algorithm for Waste Classification,” *Resour Conserv Recycl*, vol. 169, p. 105543, Jun. 2021, doi: 10.1016/j.resconrec.2021.105543.

[41] Md. Nahiduzzaman *et al.*, “Explainable deep learning model for automatic mulberry leaf disease classification,” *Front Plant Sci*, vol. 14, Sep. 2023, doi: 10.3389/fpls.2023.1175515.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[43] Md. Nahiduzzaman *et al.*, "Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier," *Expert Syst Appl*, vol. 217, p. 119557, May 2023, doi: 10.1016/j.eswa.2023.119557.

[44] C. Zhao, R. Shuai, L. Ma, W. Liu, D. Hu, and M. Wu, "Dermoscopy Image Classification Based on StyleGAN and DenseNet201," *IEEE Access*, vol. 9, pp. 8659–8679, 2021, doi: 10.1109/ACCESS.2021.3049600.

[45] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019.

[46] Y. Bhatia, A. Bajpayee, D. Raghuvanshi, and H. Mittal, "Image Captioning using Google's Inception-resnet-v2 and Recurrent Neural Network," in *2019 Twelfth International Conference on Contemporary Computing (IC3)*, IEEE, Aug. 2019, pp. 1–6. doi: 10.1109/IC3.2019.8844921.

[47] S. Maheta and Manisha, "Deep Learning-Based Cancelable Biometric Recognition Using MobileNetV3Small Model," 2023, pp. 347–356. doi: 10.1007/978-981-99-1203-2_29.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015.

[49] V. Sudha and Dr. T. R. Ganeshbabu, "A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning," *Computers, Materials & Continua*, vol. 66, no. 1, pp. 827–842, 2020, doi: 10.32604/cmc.2020.012008.

[50] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

[51] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[52] Md. Nahiduzzaman, Md. R. Islam, and R. Hassan, "ChestX-Ray6: Prediction of multiple diseases including COVID-19 from chest X-ray images using convolutional neural network," *Expert Syst Appl*, vol. 211, p. 118576, Jan. 2023, doi: 10.1016/j.eswa.2022.118576.

[53] Md. Nahiduzzaman, Md. R. Islam, S. M. R. Islam, Md. O. F. Goni, Md. S. Anower, and K.-S. Kwak, "Hybrid CNN-SVD Based Prominent Feature Extraction and Selection for Grading Diabetic Retinopathy Using Extreme Learning Machine Algorithm," *IEEE Access*, vol. 9, pp. 152261–152274, 2021, doi: 10.1109/ACCESS.2021.3125791.

[54] Md. Nahiduzzaman *et al.*, "A Novel Method for Multivariant Pneumonia Classification Based on Hybrid CNN-PCA Based Feature Extraction Using Extreme Learning Machine With CXR Images," *IEEE Access*, vol. 9, pp. 147512–147526, 2021, doi: 10.1109/ACCESS.2021.3123782.

[55] H. B. Kibria, M. Nahiduzzaman, Md. O. F. Goni, M. Ahsan, and J. Haider, "An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI," *Sensors*, vol. 22, no. 19, p. 7268, Sep. 2022, doi: 10.3390/s22197268.

[56] Md. Nahiduzzaman *et al.*, "Parallel CNN-ELM: A multiclass classification of chest X-ray images to identify seventeen lung diseases including COVID-19," *Expert Syst Appl*, vol. 229, p. 120528, Nov. 2023, doi: 10.1016/j.eswa.2023.120528.

[57] S. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*. 2017.

[58] M. Bhandari, T. B. Shahi, B. Siku, and A. Neupane, “Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI,” *Comput Biol Med*, vol. 150, p. 106156, Nov. 2022, doi: 10.1016/j.compbiomed.2022.106156.

[59] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, “Explainable AI: A Review of Machine Learning Interpretability Methods,” *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.

[60] Z. Li, K. Kamnitsas, and B. Glocker, “Analyzing Overfitting Under Class Imbalance in Neural Networks for Image Segmentation,” *IEEE Trans Med Imaging*, vol. 40, no. 3, pp. 1065–1077, Mar. 2021, doi: 10.1109/TMI.2020.3046692.