Please cite the Published Version

An automated waste classification system using deep learning techniques: Toward efficient waste recycling and environmental sustainability

Md. Nahiduzzaman [a], Md. Faysal Ahamed [a], Mansura Naznine [b], Md. Jawadul Karim [a], Hafsa Binte Kibria [a], Mohamed Arselene Ayari [c], Amith Khandakar [c], Azad Ashraf [d], Mominul Ahsan [e], Julfikar Haider [f,*]

[a] Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[b] Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[c] Department of Electrical Engineering, Qatar University, Doha 2713, Qatar
[d] Chemical Engineering Department, University of Doha for Science and Technology, Doha, Qatar
[e] Department of Computer Science, University of York, Deramore Lane, Heslington, York YO10 5GH, UK
[f] Department of Engineering, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK

ARTICLE INFO

ABSTRACT

The growing negative effects caused by inadequate waste processing have led to the widespread implementation of waste classification systems. One effective approach is to develop an automated classification system that uses advanced waste recognition technology. This method can decrease the amount of manual labor required for waste separation and recycling activities. In the present study, a novel three-stage waste classification system was proposed. It incorporates the parallel lightweight depth-wise separable convolutional neural network (DP-CNN) in conjunction with the ensemble extreme learning machine (En-ELM) classifier. Waste items are first classified into two main categories: biodegradable and non-biodegradable. The dataset is then split into nine distinct categories in the second stage based on the overall waste characteristics. The final stage of the classification process involves a more detailed granularity, as all images are assigned to one of thirty-six specific classes. With an average accuracy, precision, recall, f1, and ROC-AUC values of 96 %, 95.0 ± 0.02 %, 95.0 ± 0.02 %, 95.0 ± 0.02 %, and 98.77 %, respectively, the proposed model demonstrates promising results in the first stage of two-class classification. Advancing to the second stage, the framework showed excellent results in nine-class classification, with performance rates of 91.00 %, 90.0 ± 0.04 %, 89.44 ± 0.06 %, 89.66 ± 0.05 %, and 98.57 %, respectively. Similar to the previous stages, the model continued to perform effectively in the third stage, achieving 85.25 % accuracy, 85.02 % precision, 85.25 % recall, 84.54 % f1-score, and 98.68 % AUC in the thirty-six-class classification. The En-ELM classifier, a fusion of pseudoinverse ELM (PI-ELM) and L1 regularized ELM (L1-RELM), achieved impressive results with an average testing time of 0.00001 s. A novel comprehensive dataset titled TriCascade WasteImage, which combines four smaller preexisting datasets, was used to measure the performance of the DP-CNN-En-ELM model. With only nine layers and 1.09 million parameters, the proposed approach precisely extracts pertinent information from images to classify diverse waste materials. The effectiveness of the model is confirmed by comparing it to advanced transfer learning methods. Various explainable AI (XAI) methods are used to explore the interpretability and decision-making capability of the proposed model. Additionally, this study presents a comprehensive prototype hardware architecture for rapid waste categorization in an augmented environment, enabling autonomous waste sorting in industrial applications.

---

* Corresponding author.

E-mail addresses: nahiduzzaman@ece.ruet.ac.bd (Md. Nahiduzzaman), faysal.ahamed@ece.ruet.ac.bd (Md.F. Ahamed), 1703008@student.ruet.ac.bd (M. Naznine), 1710021@student.ruet.ac.bd (Md.J. Karim), hafsabintekibria@ece.ruet.ac.bd (H.B. Kibria), ArslanA@qu.edu.qa (M.A. Ayari), amitk@qu.edu.qa (A. Khandakar), azad.ashraf@udst.edu.qa (A. Ashraf), mominul.ahsan@york.ac.uk (M. Ahsan), j.haider@mmu.ac.uk (J. Haider).

1. Introduction

Waste generation has become an increasingly pressing issue in the modern world. With the increase in population and consumption patterns, the volume of waste produced has reached staggering levels [1,2]. This accumulation of waste poses significant challenges, ranging from environmental degradation to public health concerns. According to organizations such as the World Bank, global waste generation is expected to increase to 3.8 billion tons by 2050, highlighting the urgent need for effective waste management strategies [3]. According to data from the World Bank, high-income countries often spend between 1 % and 3 % of their GDP on waste management annually. In contrast, middle- and low-income countries tend to have lower waste management budgets, with expenditures ranging from 0.5 % to 1 % of GDP or even less in some cases. Regarding waste generation, while developed countries have advanced waste management systems in place, they also tend to produce more waste per capita due to higher levels of consumption and production. However, middle- and low-income countries, particularly those experiencing rapid urbanization and industrialization, are experiencing significant increases in waste generation. This trend is attributed to population growth, urban expansion, changing consumption patterns, and inadequate infrastructure for waste management.

In Bangladesh, waste generation has significantly increased over the years. In 2012, it reached 1,47,78,497 tons from 11,00,000 tons in 1970, with an annual increase of 1,34,300 tons [4]. Urban areas produced 5200,919 tons per year in 2014 (0.35 kg per capita per day) [4]. In Dhaka, only half of the waste is collected properly, leaving 40–60 % uncollected, mostly organic. With the urban population expected to reach 78.44 million by 2025, waste generation will increase to 220 kg per capita per year. Land scarcity and a lack of expertise and clear policies complicate waste management, impacting health, safety, and the environment. Safe disposal is crucial for city functionality and well-being [4]. Therefore, while developed countries may have higher waste management budgets, the challenge of waste generation and management is increasingly significant in middle- and low-income countries due to rapid urbanization and industrialization. The impact of unchecked waste extends far beyond mere inconvenience. Improper disposal of waste leads to soil, air, and water pollution, which threatens ecosystems and biodiversity. Plastic waste, in particular, has garnered attention due to its persistence in the environment and harmful effects on marine life [5]. Moreover, uncontrolled landfilling contributes to greenhouse gas emissions, exacerbating climate change.

Recognizing the gravity of these issues, the importance of efficient waste management cannot be overstated. Proper waste management not only mitigates environmental harm but also conserves resources and supports sustainable development goals. However, traditional waste management approaches face challenges such as limited resources, inadequate infrastructure, and inefficient sorting methods [6]. To address these limitations, advancements in technology, particularly machine learning algorithms, offer promising solutions. By leveraging the power of artificial intelligence, waste classification models can categorize and recycle waste materials efficiently. This not only optimizes resource utilization but also minimizes environmental impact by diverting waste from landfills.

Over time, numerous researchers have explored the application of artificial intelligence (AI) for waste classification. Some have employed YOLO-based models [7–11], leveraging their real-time object detection capabilities, while others have opted for transfer learning techniques [12–16], capitalizing on pretrained models to enhance classification accuracy. Although transfer learning often yields high accuracy, it has the drawback of high parameter complexity, making it computationally intensive and resource-demanding. Moreover, many existing studies have focused on a limited range of target classes, restricting their applicability to broader waste categorization scenarios and hindering their scalability. Recognizing these challenges, recent efforts have emerged to develop lightweight models for garbage classification

[17–19]. These models prioritize efficiency and scalability, aiming to reduce computational overhead while maintaining satisfactory performance. However, despite advancements in lightweight model development, they still encounter limitations that need to be addressed for more effective waste management solutions. These include issues such as reduced classification accuracy compared to more complex models [20], particularly when dealing with complex waste materials or in environments with varying lighting conditions. Additionally, the interpretability of lightweight models may be limited, posing challenges in understanding the underlying decision-making processes and potentially affecting trust and adoption in practical waste management settings.

The main contributions of this paper include the following:

1. **Largest Waste Classification Dataset**: A comprehensive dataset of 35,264 waste images has been introduced, supporting a three-stage classification system: binary classification, followed by nine specific classes, and thirty-six granular classes. This extensive dataset enables more accurate and robust waste sorting, improving overall recycling processes and supporting better waste management practices in real life.
2. **Lightweight Ensemble Classifier**: The En-ELM, a lightweight ensemble classifier combining PI-ELM and L1-RELM, has been proposed. This classifier is ideal for embedded systems in resource-constrained environments.
3. **Interpretable AI Techniques**: GradCam, guided GradCam, salience mapping, and SHAP techniques have been employed to enhance the transparency and interpretability of the classification process, ensuring understandable decision-making.
4. **Real-Time Classification with Hardware Support**: Real-time waste classification has been implemented in an augmented environment with a prototype hardware architecture, improving efficiency and scalability for industrial applications.
5. **Addressing Infrastructure and Resource Constraints**: The proposed framework is specifically designed for deployment in resource-constrained environments, making it highly applicable in regions with inadequate waste management infrastructure, such as developing countries.
6. **Scalable and Efficient Waste Sorting**: The multi-stage classification approach enhances sorting efficiency and scalability, making the system suitable for large-scale industrial waste management and reducing manual labor.
7. **Environmental Impact**: By improving waste sorting accuracy and real-time processing, the system contributes to reducing environmental pollution, supporting better recycling practices and sustainability goals.

A thorough review of previous studies on the subject is presented in Section 2. In Section 3, the proposed approach is outlined, including the framework, data description, feature extraction techniques, and performance metrics. An in-depth examination of the results at each of the three classification stages is presented in Section 4, along with a thorough explanation. This section also describes the hardware configuration for feasible deployment and presents the interpretability of the proposed model using explainable AI (XAI) techniques. Finally, Section 5 presents the key conclusions.

2. Related works

For waste classification, many works [13–15,21–26] have leveraged transfer learning, primarily due to its effectiveness in utilizing pretrained models and adapting them to specific tasks.

Jin et al. [25] introduced a device utilizing deep learning techniques employing MobileNetV2. With the Huawei Cloud Garbage dataset, the model achieved 90.7 % accuracy in classifying images into four categories. In [13], a novel waste classification method was proposed

utilizing a DenseNet169 model based on transfer learning. The authors introduced the NWNU-TRASH dataset with 2528 images across 5 classes, which was split into 70 % for training and 30 % for testing. The results showed an accuracy of >82 %, surpassing that of other algorithms. However, limitations included a small sample size, uneven waste distribution, and a high number of parameters in the trained model, which hindered the real-world applicability of the model. The limitation of the small sample size was addressed in [15], where the study introduced metaheuristics with deep transfer learning enabled detection and classification methods for industrial waste management (MDTLDC-IWM) models. This model aimed to streamline waste classification in industrial settings by utilizing YOLO-v5 for object detection and a stacked sparse autoencoder (SSAE) for classification. The model achieved high precision (96.84 %) and F scores (96.71 %) on a dataset comprising 2467 images across six classes. Anh H. Vo [14] proposed the DNN-TC model, which likely contains a significant number of parameters due to its deep neural network architecture. The VN-trash dataset was utilized, and DNN-TC was developed based on the ResNext architecture with enhancements. The experimental results showed that DNN-TC outperformed state-of-the-art methods on both the Trashnet and VN-trah datasets, achieving 94 % and 98 % accuracy, respectively. One limitation of this work is that future endeavours entail refining the framework for real-world applications and exploring segmentation techniques for image preprocessing. However, the issue of real-time application has been addressed in other studies. Cheema et al. [26] proposed SWMACM-CA, a real-time waste management and classification system. It integrated IoT and deep learning, achieving over 90 % accuracy in waste classification using waste grid segmentation and the VGG16 DL algorithm. The system minimized latency by training the model on a cloud server and employing a PAN for reliable IoT communication. This approach enhanced waste management efficiency, effectively addressing critical environmental challenges.

All the models discussed here, as presented in previous papers [13–15,26], employed transfer learning, capitalizing on models characterized by a large number of parameters. However, while effective, such approaches may pose limitations in terms of computational efficiency, making them less suitable for resource-constrained devices or real-time applications. Recognizing the importance of lightweight models, many studies have been conducted to address the challenges associated with transfer learning, aiming to enhance efficiency and applicability in diverse settings. Several lightweight machine learning and deep learning models have been explored for waste classification tasks, offering computational efficiency and scalability.

Yang et al. [27] proposed a novel approach to address pollution from unseparated garbage, focusing on a garbage classification system. They developed a lightweight neural network called WasNet, with 1.5 million parameters, which is significantly less than that of mainstream networks. Despite its compact size, WasNet demonstrated impressive performance, achieving accuracies of 64.5 % on the ImageNet dataset, 82.5 % on the Huawei Garbage Classification dataset, and 96.10 % on the TrashNet dataset. This indicates its effectiveness in classifying various types of garbage. The number of parameters was further reduced in [28]. A lightweight network architecture named Focus-RCNet, which employs knowledge distillation to further compress and optimize the model, was introduced. The performance of Focus-RCNet was validated on the TrashNet dataset, which comprises six target classes. With a total of 0.525 million parameters, the model demonstrated advantages such as low computational cost, a small number of parameters, high speed, and high accuracy. This model was well suited for deployment on mobile devices and holds promise for automatic waste classification, reducing the need for human intervention. However, the number of target classes can be further increased, potentially broadening its applicability. Testing on the TrashNet dataset yielded a remarkable accuracy of 92 %. Another study [29] focused on addressing the challenges in managing construction and demolition waste (C&DW), which significantly impacts project costs. They proposed a deep convolution neural network to

automatically identify different C&DW materials from digital images of waste deposited in construction site bins. The experiment achieved a high accuracy of 94 %, indicating the potential for reducing project costs and diverting C&DW from landfills. Recent research indicates that deep learning models outperform traditional techniques in object detection and classification. As urbanization accelerates, smart cities are incorporating Internet of Things (IoT) technologies for automated waste management, aiming to enhance efficiency, flexibility, and sustainability [30–32]. IoT solutions facilitate the immediate monitoring, gathering, and management of garbage. In a study by Hussain et al. [33], IoT-based smart bins were developed that integrate DL and machine learning (ML) models for waste monitoring, collection, management, and forecasting of air pollutants in the environment. Hence, recognizing the growing importance of real-time monitoring and management in waste systems, this work has implemented a lightweight model with a large number of classes, explainable artificial intelligence (AI) techniques, and real-time data exchange to complement these advancements that align closely with real-world requirements.

3. Materials and methods

3.1. Dataset description

This paper introduces a novel dataset named TriCascade WasteImage, which fills the present gap in datasets that include three distinct stages of waste classification. This customized dataset is a combination of four existing datasets: TrashBox [34] from GitHub, the Dead Animals Pollution Image Dataset [35] from Roboflow, and two datasets obtained from Kaggle, namely, the waste_picture [36] dataset and the Garbage Classification [37] dataset. These four datasets are currently the largest and most widely utilized datasets among various researchers. The Tri-Cascade WasteImage dataset plays a pivotal role in advancing waste image detection and classification research [13,34].

The TriCascade WasteImage dataset was initially divided into two categories, biodegradable and nonbiodegradable waste, denoting binary classification. In the second stage, the dataset is divided into nine specific classes based on the general characteristics of waste. Biodegradable wastes are classified into green waste and recyclable waste. On the other hand, non-biodegradable wastes are classified into glass, metal, polymer (petroleum-based), leather and fabric, medical waste, e-waste, and hazardous waste. In the third and final stage, all the images are allocated to one of thirty-six specific classes.

Using a total of 35,264 images, the TriCascade WasteImage dataset is meticulously organized to facilitate comprehensive waste classification research. Table 1 and Fig. 1 provide a comprehensive analysis of the dataset, demonstrating its structure and presenting a few samples as examples. This dataset promises to be a valuable resource for advancing the state of the art in waste image classification and related research domains.

3.2. Proposed framework

A waste classification method was implemented using a deep learning (DL) approach. Fig. 2 illustrates the three fundamental stages of the suggested framework: preprocessing the dataset, employing DL for classification, and creating an interpretable model using XAI. Four different datasets are used to collect images, and different classes from each dataset are combined since there is currently no dataset available for three-stage waste classification. After collecting and evaluating images, the waste images are separated into thirty-six different classes. There is an 80:10:10 split among the sample images used for training, testing, and validation. To ensure normalization, preprocessing was applied to the images. To classify waste objects, a unique and lightweight parallel depth-wise separable convolutional neural network (DP-CNN) is implemented. XAI is utilized for model interpretation.

The decision to implement a three-stage classification model is

Table 1

TriCascade WasteImage dataset overview.

First Stage of Classification (Two classes)	Second Stage of Classification (Nine classes)	Third Stage of Classification (Thirty-six classes)	Class No.	Training	Testing	Validation
Biodegradable Waste	Green Waste	Foods	0	3317	424	373
		Animal Dead Body	1	179	23	18
	Recyclable Waste	Cardboard	2	1647	179	209
		Newspaper	3	925	106	89
		Paper Cups	4	419	52	40
		Paper	5	700	76	73
		Bowls and Dishes (Porcelain)	7	499	56	52
Non-biodegradable Waste	Glass	Brown Glass	6	679	60	74
		Green Glass	8	489	70	70
		White Glass	9	622	88	65
	Metal	Beverage Cans	10	1306	160	136
		Construction Scrap	11	350	38	43
		Metal Containers	12	354	55	28
	Polymer (Petroleum Based)	Plastic Bag	13	791	103	76
		Plastic Bottle	14	588	73	94
		Plastic Container	15	372	62	30
		Plastic Cup	16	325	42	38
		Tetra Pak	17	679	88	75
	Leather and Fabric	Cloths (Synthetic Fabric)	18	4354	516	454
		Shoes	19	1580	210	187
	Medical Waste	Gloves	20	284	36	33
		Masks	21	313	47	40
		Bandage	22	336	39	30
		Medicine and Medicine Strip	23	1069	124	114
		Syringe	24	319	39	47
		Diaper	25	627	87	64
	E-waste	Electric Cable	26	444	58	51
		Electric Chips	27	395	52	45
		Laptop	28	324	30	44
		Small Appliances	29	601	69	70
		Smartphones	30	179	14	26
		Battery	31	1619	198	172
	Hazardous Waste	Thermometer	32	745	85	78
		Cigarette Butt	33	73	9	15
		Pesticide Bottle	34	750	104	86
		Spray Can	35	310	55	35
Total				**28,563**	**3527**	**3174**

driven by the complexities and challenges associated with real-world industrial waste sorting. Waste materials vary significantly in terms of their physical characteristics, composition, and recyclability, which makes it difficult for a single-stage end-to-end model to perform effectively. By dividing the classification process into three distinct stages, our model can handle this diversity more efficiently. The first stage provides a broad categorization into biodegradable and non-biodegradable waste, reducing the complexity for subsequent stages. In the second stage, the waste is classified into nine specific classes, refining the categories based on their general characteristics. The final stage involves a detailed classification into thirty-six distinct classes, allowing for granular sorting of waste. This staged approach ensures that the model progressively narrows down the classification task, improving accuracy and reducing the potential for misclassification.

This model has been specifically designed for deployment in industrial waste management environments. Each stage functions as an independent end-to-end model, which allows for distributed processing across different machines. In practical industrial setups, these stages can be implemented sequentially across multiple devices, with each stage handling a particular aspect of the waste sorting process. For instance, the first machine could handle the binary classification of waste, the second machine could sort the waste into nine categories, and the final machine could complete the process by classifying waste into thirty-six granular classes. This distributed setup not only optimizes resource utilization but also enables parallel processing, thereby increasing the system's overall throughput.

The three-stage model enhances waste sorting efficiency by breaking down the task into more manageable stages. Each stage specializes in different aspects of the classification, which allows for more effective feature extraction and categorization. By progressively refining the classification at each stage, the model reduces the complexity faced by each individual stage. This results in better accuracy compared to a conventional single-stage end-to-end model, particularly in industrial environments where diverse waste materials need to be sorted rapidly. Additionally, the staged approach allows for easier troubleshooting and upgrading of individual stages without affecting the entire system, making it more flexible and scalable in real-world applications.

Furthermore, each stage in the model categorizes waste from different perspectives. For example, in the first stage, the broad distinction between biodegradable and non-biodegradable waste focuses on the environmental impact. In the second stage, the classification is based on the physical characteristics and general categories of waste. Finally, the third stage dives deeper into the composition of waste, ensuring that each item is placed into a specific category that aids in recycling and processing. This method allows for more nuanced and comprehensive waste management, addressing various aspects of waste categorization that a single-stage model might overlook.

3.3. Dataset preprocessing

The classification accuracy is significantly affected by the image preprocessing quality. This study streamlines the image processing phases to facilitate their implementation on embedded systems. As part of the preprocessing stage, the images in the dataset were resized to dimensions of 124 pixels in both width and height to save the storage space required and to optimize the utilization of processing resources. Frequently, a large quantity of intensity values is employed to represent an image. Normalization [38] was performed to reduce the complexity
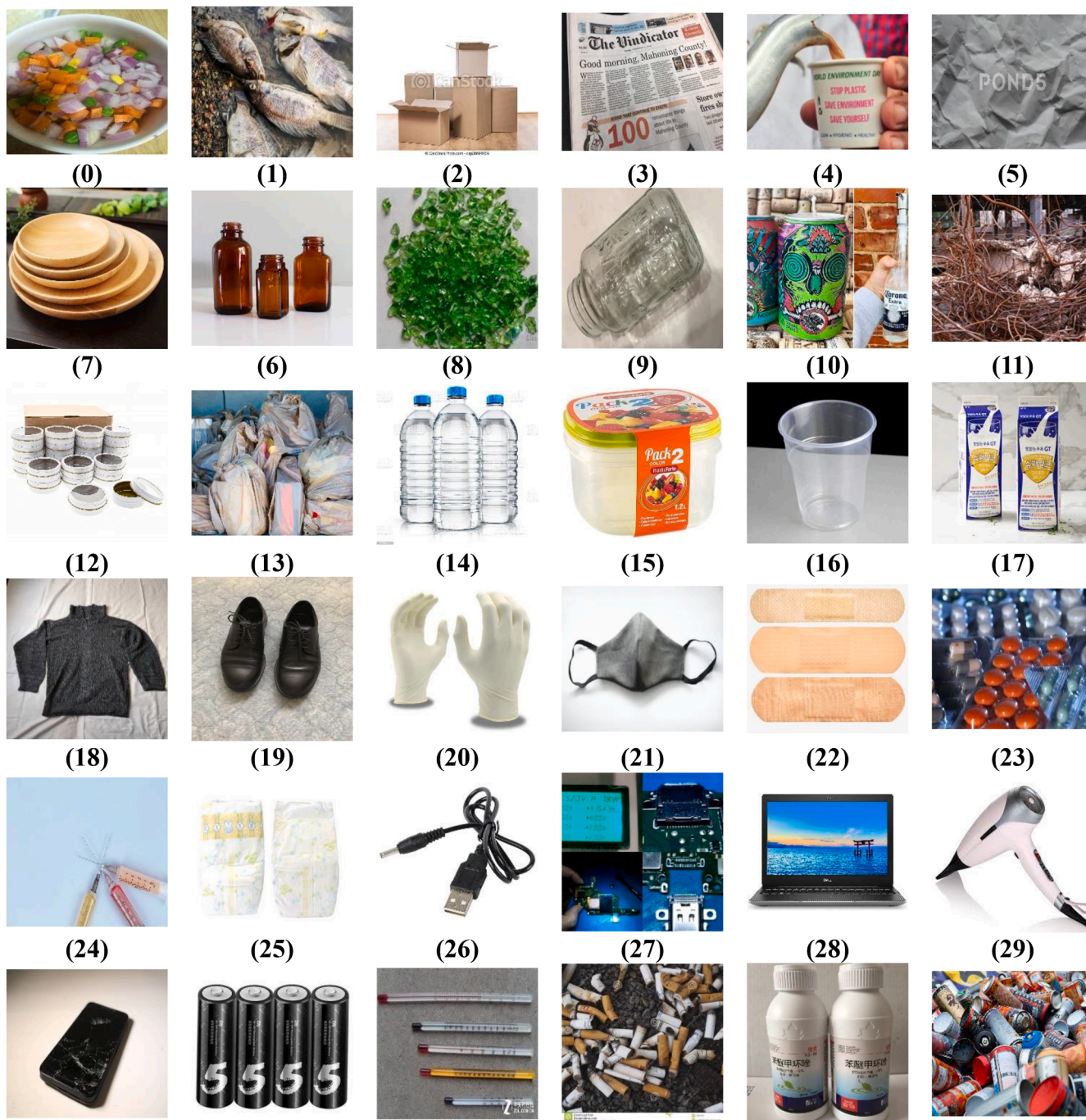
Fig. 1. TriCascade WasteImage dataset include 36 classes and 3 stages: (0) Foods, (1) Animal Dead Body, (2) Cardboard, (3) Newspaper, (4) Paper Cups, (5) Paper, (7) Bowls and Dishes (Porcelain), (6) Brown Glass, (8) Green Glass, (9) White Glass, (10) Beverage Cans, (11) Construction Scrap, (12) Metal Containers, (13) Plastic Bag, (14) Plastic Bottle, (15) Plastic Container, (16) Plastic Cup, (17) Tetra Pak, (18) Cloths (Synthetic Fabric), (19) Shoes, (20) Gloves, (21) Masks, (22) Bandage, (23) Medicine and Medicine Strip, (24) Syringe, (25) Diaper, (26) Electric Cable, (27) Electric Chips, (28) Laptop, (29) Small Appliances, (30) Smartphones, (31) Battery, (32) Thermometer, (33) Cigarette Butt, (34) Pesticide Bottle, (35) Spray Can.

of the images, and the scale was changed from 0 to 255 to 0–1 by dividing the pixel values by 255.

3.4. Deep learning model

Most transfer learning (TL) models possess significantly large parameters, layers, and sizes, leading to a substantial increase in computational requirements. To overcome these challenges, a customized version of a parallel lightweight depth-wise separable convolutional neural network (DP-CNN) was built. This model is characterized by its low complexity, lightweight nature, and low number of parameters, layers, and size, resulting in minimal overhead. The next subsections provide a comprehensive description of the DP-CNN feature extractor, together with concise explanations of the state-of-the-art TL feature extractors used in this study [39].

3.4.1. Parallel lightweight depth-wise separable convolutional neural network (DP-CNN) feature extractor

Efficiently extracting vital features with minimal parameters and network layers is crucial in a CNN model to achieve optimal
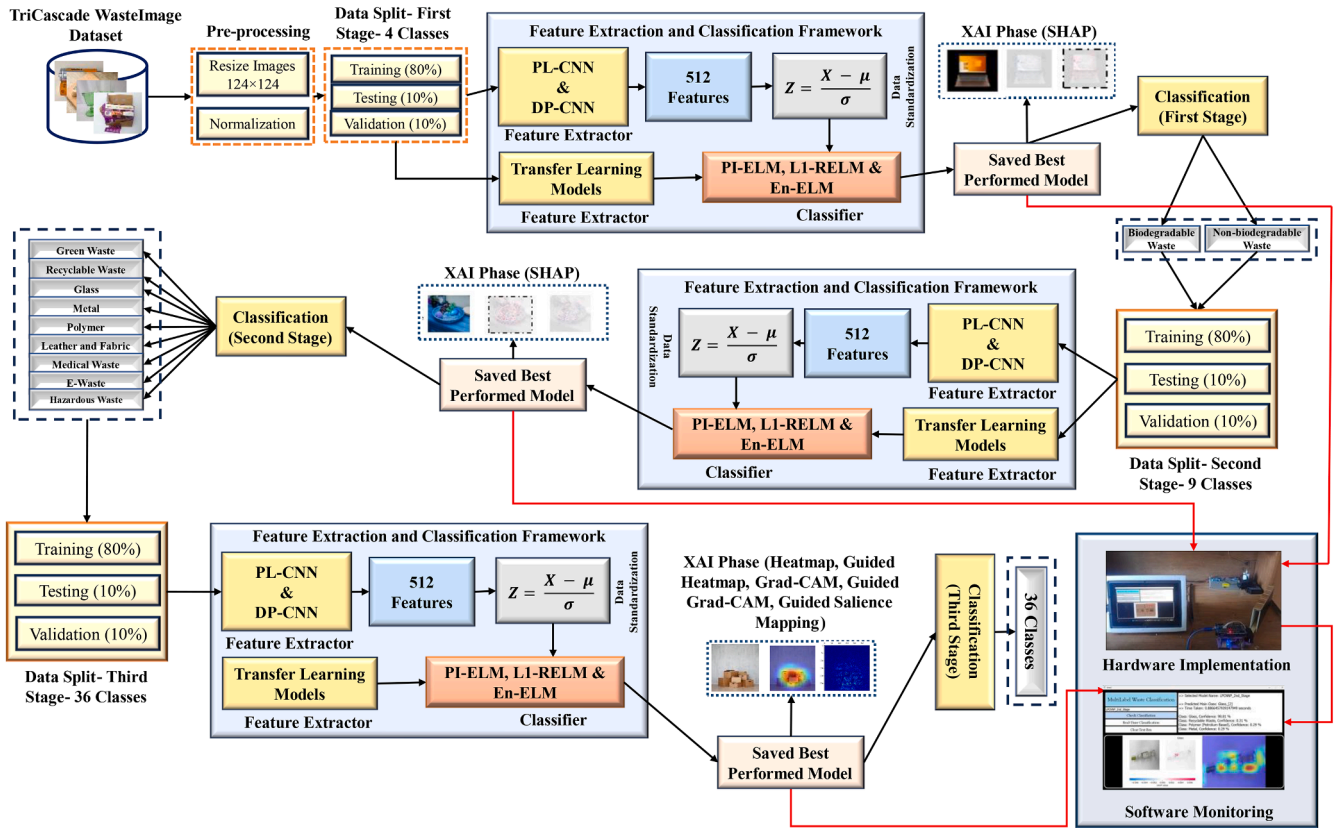
Fig. 2. Proposed model for three-stage multiclass classification of waste images.

performance and practical applicability. Attaining an optimal layer structure is essential, as a lower number of layers and parameters might limit the model's ability to identify special features, resulting in performance constraints. Conversely, an excessive number of layers and parameters could result in overfitting, leading to longer processing times and increased computational demand. Hence, the main goal was to develop a CNN model capable of effectively identifying the most critical features using a minimal number of layers and parameters.

Fig. 3 illustrates the structure of the proposed DP-CNN model, which adeptly handles layer complexity and conforms to parameter
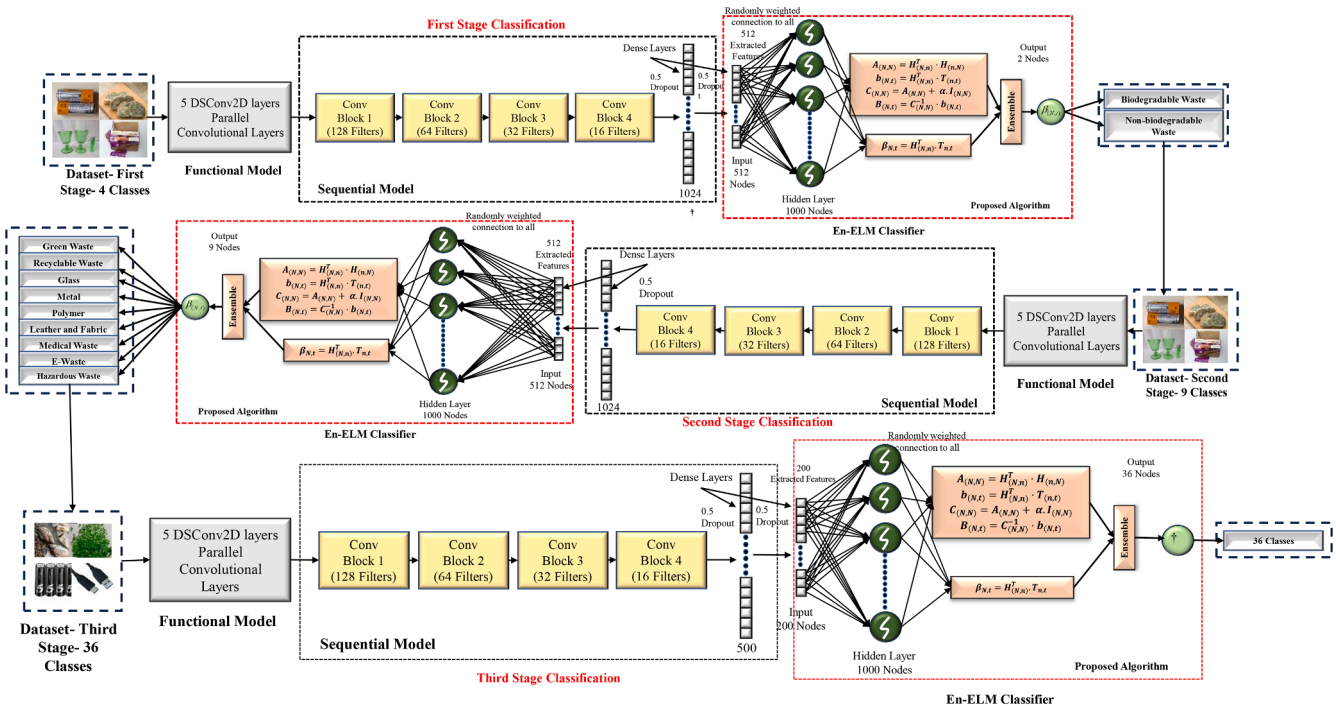
Fig. 3. Proposed DP-CNN-En-ELM architecture.

constraints, ensuring efficient feature extraction within a practical size. The model consists of nine convolution layers (CLs) with two fully connected (FC) layers to obtain an ideal balance. Instead of using only one CL, five parallel CLs were employed (Fig. 4). On the other hand, using five consecutive CLs would increase the model complexity by increasing the depth of the layers. To address this issue, parallel execution of the initial five convolutional layers (CLs) was implemented, with their selection being determined via a trial-and-error approach. There were 256 kernels used in total for each CL; the first, second, third, fourth, and fifth kernel sizes were 11×11, 9×9, 7×7, 5×5, and 3×3, respectively. The kernel size was determined according to the design ideas of Krizhevsky et al. [40], who recommended using large kernel sizes such as 11×11 to achieve the most effective classification results. As various kernels produce distinct feature maps, this study examined and combined kernels of varying sizes, ranging from small to large, to identify noteworthy features and achieve satisfactory classification performance. For the first five CLs, the padding size was maintained at the same value to collect important data from the border element of the waste images. The feature maps produced by the parallel convolutional layers (CLs) were subsequently merged and fed into a subsequent sequential CL.

The CNN framework was enhanced by integrating depth-wise separable convolution (DSC), a technique that splits the standard convolution procedure into two separate phases: depth-wise and pointwise. Initially, depth-wise convolution employs a smaller kernel to operate on a specific segment of an input feature map, generating a new feature map that retains the original number of channels. The output from the depth-wise convolution is then processed through a pointwise convolution. Here, a 1×1 convolutional kernel acts on every channel, resulting in a new feature map with a decreased number of channels. In the conclusion stage, four convolutional layers (CLs) were added, utilizing batch normalization (BN) and max pooling with a 2×2 kernel size. The filters of these CLs were set to 128, 64, 32, and 16. Each filter was supplied with 3×3 kernels and was configured to utilize VALID padding. By effectively recentering and rescaling the inputs to each layer, BN proved advantageous, enhancing the model's execution speed

and stability. Additionally, the rectified linear unit (ReLU) activation function was applied across all CLs. In addition to three FC layers, dropout was employed to mitigate overfitting and accelerate the training process by randomly discarding 50 % of all nodes. The proposed model has a complex architecture consisting of nine CLs. The first five layers act simultaneously, effectively operating as a single layer in the overall system. After implementing this distinctive arrangement, four more CLs are added, resulting in a total of five CLs. The model additionally incorporates two FC layers alongside the CLs. This complex configuration yields a grand total of eight unique layers. For this experiment, two dropouts were utilized following the last two CLs, and an additional two dropouts were implemented after the initial two FC layers. The SoftMax activation function was used in the final FC layer to classify different types of waste materials. After extracting 200 features from the final FC layer, the ensemble ELM classifier was substituted for the SoftMax function to improve the classification performance. The model was trained using a loss function based on sparse categorical loss, with training facilitated by a 32-batch-size ADAM optimizer. A learning rate of 0.001, established through trial and error, was utilized, and the model underwent training for 200 epochs. Table 2 displays a summary of the detailed model.

3.4.2. Transfer learning feature extractors

Recent years have witnessed the successful application of transfer learning (TL) models for a wide range of potential uses. This study employed six TL models: DenseNet201, InceptionResNetV2, MobileNetV3Small, ResNet152V2, VGG16, and Xception. With over 14 million images and 1000 classifications, the ImageNet dataset was utilized as the training resource for all the pretrained models. The training of the TL feature extractors was combined with the proposed ELM classifier to classify waste images (Fig. 5). After initializing these feature extractors, the final layers were modified by incorporating two fully connected (FC) layers with 500 and 200 nodes to enhance the detection of waste images. TL approaches were compared with the unique lightweight DP-CNN model that was suggested regarding classification outcomes and processing resources (sizes, layers, and parameters).
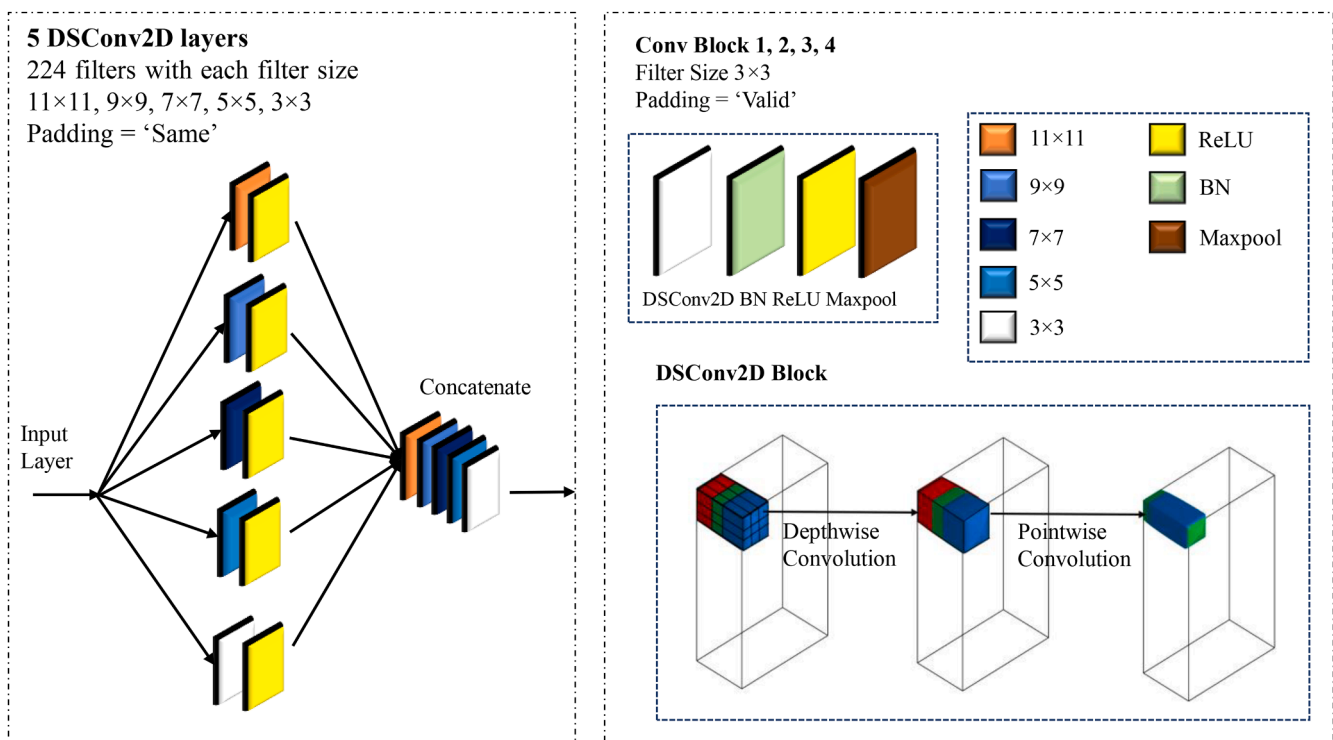

Fig. 4. Detailed summary of the convolution block structure.

Table 2

Parallel lightweight depthwise separable convolutional neural network (DP-CNN) feature extractor summary.

Layer Type	Output Shape	Parameters
Model Input	(None, 124, 124, 3)	0
Model (Functional)	(None, 124, 124, 1280)	5975
Separable conv2d 5	(None, 122, 122, 128)	175,488
Batch normalization	(None, 122, 122, 128)	512
Activation	(None, 122, 122, 128)	0
Max Pooling 2D	(None, 61, 61, 128)	0
Separable conv2d 6	(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 7	(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)	410,624
Batch normalization 4	(None, 1024)	4096
Dropout 1	(None, 1024)	0
DenseLastPL	(None, 512)	524,800
Batch normalization 5	(None, 512)	2048
Dropout 2	(None, 512)	0
Dense 1	(None, 36)	18,468
Total Parameters:	**1155,339**	
Trainable Parameters:	**1151,787**	
Non-Trainable Parameters:	**3552**	

The DenseNet architecture is notable in the field of TL for feature extraction because of its dense connectivity. This means that each layer is interconnected, allowing for a smooth flow of information and enabling the capture of complicated features. As a result, the performance of the model improved. The InceptionResNetV2 model combines the advantages of the ResNet and Inception frameworks by utilizing inception modules and residual connections to achieve enhanced feature extraction. MobileNetV3Small is specifically designed for real-time applications on devices that have limited resources. It focuses on finding the best balance between size, speed, and accuracy. The ResNet152V2 model improves learning efficiency by utilizing advanced residual learning methods, incorporating 60 million parameters to facilitate subtle feature learning. The VGG architecture utilizes deep convolutional layers, followed by Max Pooling and ReLU functions, which ultimately leads to a fully connected layer equipped with SoftMax. This demonstrates a strong and effective method for extracting features. Google's Xception model is an improvement on Inception. It uses depthwise separable convolutions to achieve efficient and accurate performance in computer vision tasks, especially when there are computing limitations. This represents a notable advancement in TL feature extractors.

3.5. Ensemble extreme learning model (En-ELM)

A notable paradigm shift in feature classification was made by Huang et al. [41] with the introduction of the ELM. This approach utilizes a feedforward network supported by supervised learning and has been recognized as a major advancement. Through the use of neural network capabilities, the ELM eliminates the need for backpropagation, leading to an incredible increase in training speed of up to a thousand times. The field of feature classification has undergone a significant transformation because of this innovative method.

Recent advancements have significantly enhanced the model's classification and generalization performance, particularly the pseudoinverse ELM (PI-ELM), which excels in large-scale multiclass classification [42–46]. The PI-ELM is significant because it uses only one hidden layer and provides an innovative and flexible approach for initializing parameters between the input and hidden layers. The pseudoinverse method is used to find the parameters that connect the hidden and output layers. By replacing the pseudoinverse approach with ridge regression methodology and L1-regularized parameters, the degree of sophistication increased. This improvement significantly boosts the model's ability to effectively capture and manage features, thereby enhancing its generalization capabilities and achieving unmatched accuracy relative to the PI-ELM. The model consists of 200 nodes in the input layer and 500 nodes in the hidden layer.

This research advances innovation by introducing a novel ensemble method that combines the strengths of both the PI-ELM and L1-RELM. This cooperative strategy utilizes the distinct strengths of PI-ELM's expertise in large-scale multiclass classification and L1-RELM's improved feature learning and regularization abilities. The suggested ensemble technique tries to use the individual capabilities of both ELMs, creating a synergistic combination that shows promise for advancing image classification research. The Ensemble ELM (En-ELM) classifier is shown in Algorithm 1.

3.6. XAI

In the context of deep learning, explainable artificial intelligence (XAI) is the capacity to understand and characterize the decision-making process of a deep neural network [47]. This is particularly important for DL models since they might be confusing and difficult to comprehend.

3.6.1. Shapley additive explanations (SHAP)

In this study, SHAP was applied to eliminate the "black box" aspect of DL models. This allowed for a more thorough evaluation and explanation of the DP-CNN model's output. SHAP computes the average marginal contributions of each feature value to evaluate the relevance of the model features. To help with categorization interpretation, each pixel in a predicted image has a score that indicates its role. By considering every
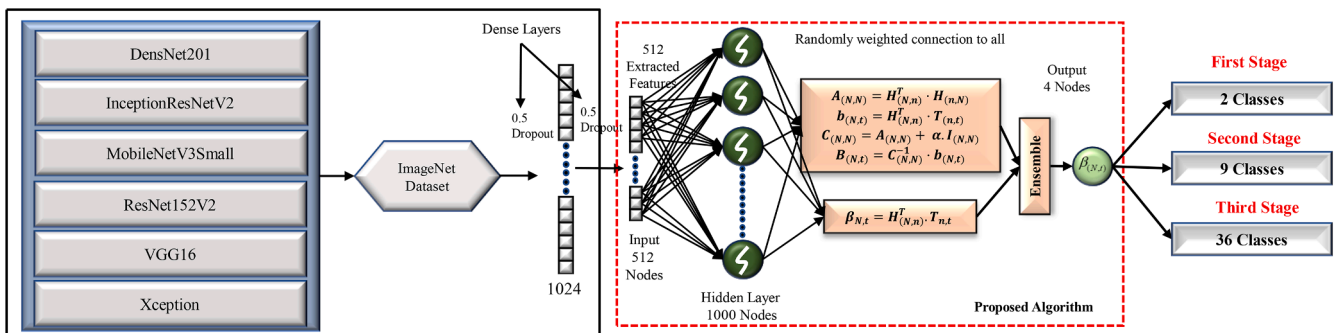

Fig. 5. The modified transfer learning architecture for classifying waste images.

Algorithm 1

Proposed En-ELM classifier for multiclass classification.

$$= \begin{bmatrix} in_{(1,1)} & in_{(1,2)} & \cdots & in_{(1,m)} \\ in_{(2,1)} & in_{(2,2)} & \cdots & in_{(2,m)} \\ in_{(3,1)} & in_{(3,2)} & \cdots & in_{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ in_{(n,1)} & in_{(n,2)} & \cdots & in_{(n,m)} \end{bmatrix} \quad Op_{(n,t)} = \begin{bmatrix} op_{(1,1)} & op_{(1,2)} & \cdots & op_{(1,t)} \\ op_{(2,1)} & op_{(2,2)} & \cdots & op_{(2,t)} \\ op_{(3,1)} & op_{(3,2)} & \cdots & op_{(3,t)} \\ \vdots & \vdots & \ddots & \vdots \\ op_{(n,1)} & op_{(n,2)} & \cdots & op_{(n,t)} \end{bmatrix}$$

$$Wg_{(m,N)} = \begin{bmatrix} wg_{(1,1)} & wg_{(1,2)} & \cdots & wg_{(1,N)} \\ wg_{(2,1)} & wg_{(2,2)} & \cdots & wg_{(2,N)} \\ wg_{(3,1)} & wg_{(3,2)} & \cdots & wg_{(3,N)} \\ \vdots & \vdots & \ddots & \vdots \\ wg_{(m,1)} & wg_{(m,2)} & \cdots & wg_{(m,N)} \end{bmatrix} \quad Hl_{(n,N)} = \begin{bmatrix} hl_{(1,1)} & hl_{(1,2)} & \cdots & hl_{(1,N)} \\ hl_{(2,1)} & hl_{(2,2)} & \cdots & hl_{(2,N)} \\ hl_{(3,1)} & hl_{(3,2)} & \cdots & hl_{(3,N)} \\ \vdots & \vdots & \ddots & \vdots \\ hl_{(n,1)} & hl_{(n,2)} & \cdots & hl_{(n,N)} \end{bmatrix}$$

The input matrix is denoted as 'In', and the output matrix is indicated by 'Op'.

1. The input weight is signified by $Wg_{(m,N)}$ and the bias matrices by $Bs_{(1,N)}$.

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

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

$$Hl_{(n,N)} = G\left(In_{(n,m)} \cdot Wg_{(m,N)} + Bs_{(1,N)}\right)$$

Here, 'G' is an activation function.

3. Calculate the output weight matrix $\beta_{(N,t)}$ through the use of pseudo inverse method.

$$\beta_{(N,t)} = Hl^t_{(N,n)} \times T_{(n,t)}$$

In L1 regularized approach, the pseudo inverse is substituted with the following equations:

$$A_{(N,N)} = Hl^T_{(N,n)} \cdot Hl_{(n,N)}$$

$$b_{(N,t)} = Hl^T_{(N,n)} \cdot T_{(n,t)}$$

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

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

Where, α denotes L1 regularization parameters.

4. Determine the final output using ensemble technique (En-ELM) that incorporates both PI-ELM and L1-RELM:

$$En_{(N,t)} = \frac{\beta_{(N,t)} + Bs_{i(N,t)}}{2}$$

$$En_{(N,t)} = \frac{H^t_{(N,n)} \cdot T_{(n,t)} + C^{-1}_{(N,N)} \cdot b_{(N,t)}}{2}$$

5. Generate prediction $En_{(N,t)}$.

---

potential combination of characteristics linked to waste images, the Shapley value can be determined. After the resulting Shapley values are pixelated, the number of red pixels increases, and the number of blue pixels decreases the likelihood of correctly identifying the class [47,48]. The Shapley value was calculated by applying Eq. (1).

$$\Theta_j = \sum_{T \subseteq O \setminus j} \frac{|T|!(B - |T| - 1)!}{B!} \left[g_y(T \cup j) - g_y(T)\right] \qquad (1)$$

Here, g_y represents the change in output inclusion attributed to Shapley values for a particular feature, denoted as j. The subset T comprises all features from the O features, excluding feature j. $\frac{|T|!(B-|T|-1)!}{B!}$ calculates the weighted factor of the subset T permutations. The anticipated outcome, represented as $g_y(T)$, is obtained from Eq. (2).

$$g_y(T) = P[g(y)|y_T] \qquad (2)$$

The SHAP technique entails the replacement of each initial feature (x_j) with a binary indicator (d'_j) that signifies the presence or absence of x_j, as shown in Eq. (3).

$$m(d') = \Theta_0 + \sum_{j=1}^{B} \Theta_j d'_j \qquad (3)$$

Here, Θ_0 represents the bias, $\Theta_j d'_j$ represents the feature contribution, and $m(d')$ represents the substitute model for the framework. Understanding the role of Θ_j and the contribution of feature j to the result are key elements in learning the model's underlying functionality.

3.6.2. Gradient-weighted class activation mapping (Grad-CAM)

Model visualization is essential for analyzing and debugging model performance. Tools such as CAM [49], Grad-CAM [50], and GradCAM++ [51] provide valuable insights into deep learning model predictions by emphasizing significant areas within an image. Grad-CAM produces heatmaps and saliency maps that highlight the specific regions that the model focuses on while generating predictions. These visualizations improve the clarity of the model and build

confidence among the users by revealing the decision-making processes of the model.

Heatmap Visualization: A heatmap is created to show the areas of the original image that most significantly influenced the final classification decision. This heatmap is produced by calculating the gradient of the final CL's output class score in relation to the feature maps [49].

$$H(x) = \Sigma_i \frac{\partial S_c}{\partial F_i} \qquad (4)$$

Here, $H(x)$ represents the heatmap for a given input image x, S_c represents the class score, and F_i corresponds to the i th feature map.

Guided Heatmap Visualization: The heatmap created in the previous phase was improved through guided heatmap visualization. Gradients are determined by guided backpropagation and then multiplied by the ReLU activation of the associated feature map [52].

Grad-CAM Visualization: Grad-CAM merges class discrimination with location. It creates a heatmap after computing the weights for each feature map based on the gradient of the class score compared to the feature maps [50].

$$\gamma^e_m = \frac{1}{x} \Sigma_o \Sigma_p \frac{\partial y^E}{\partial C^m_{op}} \qquad (5)$$

$$L^e_{GradCAM} = ReLU \left(\Sigma_m \gamma^e_m \cdot C^m\right) \qquad (6)$$

Here, y^E indicates the score for class e with respect to feature map C^m, γ^e_m indicates the calculated weight for every neuron, and $\frac{1}{x} \Sigma_o \Sigma_p$ defines a global average pooling across dimensions o and p.

Guided Grad-CAM Visualization: Guided Grad-CAM visualization is an advanced interpretability technique that combines the principles of guided backpropagation and Grad-CAM. This method refines the heatmap generated by Grad-CAM by incorporating guided backpropagation gradients and emphasizing the most influential features in the final classification decision [49].

$$Guided\ Grad - CAM^e(u,v) = ReLU\left(\Sigma_m \gamma^m_e \cdot C^m(u,v)\right) .^* G^e(u,v) \qquad (7)$$

Here, *e* stands for the target class, (u, v) represents the coordinates of a pixel in the input image, γ_e^m indicates important weights for each feature map, C^m is the activation of feature map *m* at pixel (u, v), and $G^e(u, v)$ indicates the refined gradient map obtained through guided backpropagation for class *e*.

Guided Salience Mapping: Saliency mapping evaluates a class's spatial support. Neural networks help with interpretability by providing an image that highlights the region of interest. Using backpropagation, the saliency map is produced. This allows for a better understanding of the decisions made by the model by identifying pixels that have little impact on the score and computing the derivative of the class score with respect to the image.

First, we calculate the distance of each pixel to the rest of the pixels in the same frame:

$$SALS(I_k) = \sum_{i=1}^{N} |I_k - I_i| \tag{8}$$

I_i is the value of the pixel *i*. The following equation is the expanded form of this equation.

$$SALS(I_k) = |I_k - I_1| + |I_k - I_2| + ... + |I_k - I_N| \tag{9}$$

Where, *N* represents all the pixels in the present frame. After that, we can further refine our formula. We combine the values that have the same *I*.

$$SALS(I_k) = \Sigma F_n \times |I_k - I_N| \tag{10}$$

Where, F_n is the frequency of I_N. The value of *n* belongs to [0,255].

3.7. Performance evaluation metrics and implementation

The implementation of DL algorithms and XAI was carried out using the Keras library, with TensorFlow as its backend, on PyCharm Community Edition (version 2021.2.3). The hardware configuration included an 11th generation Intel(R) Core (TM) i9–11,900 CPU at 2.50 GHz, 128 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory, supporting both the training and testing phases. The system was run on a 64-bit Windows 10 Pro operating system. The DP-CNN model's performance was assessed using a confusion matrix (CM), from which metrics such as accuracy, precision, recall, F1-score, and the area under the curve (AUC) were derived using standard formulas.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TN + FP} \tag{13}$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{Precision + Recall} \tag{14}$$

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{15}$$

The symbols *TP*, *TN*, *FP*, and *FN* were used to denote true positives, true negatives, false positives, and false negatives, respectively.

The cross-entropy formula is a comparison between the model's predicted probability distribution and the integer-based actual class label. To reduce the amount of cross-entropy loss, the difference between the actual and predicted labels is measured. In DL applications, such as image classification, the sparse categorical cross-entropy loss is commonly employed, especially when several classes are involved.

$$L_{ce} = -\sum_{i=1}^{n} y_r \times \log(y_p) \tag{16}$$

where *n* denotes the class number, the truth label is defined as y_r, and y_p is the probability.

4. Results and discussion

This research focused on assessing the proposed models (the PL-CNN and DP-CNN feature extractors in conjunction with the PI-ELM, L1-RELM, and En-ELM classifiers) across various classes in three different stages to evaluate their performances and compare the results with those of other state-of-the-art TL models. First, binary classification was performed to distinguish between biodegradable and non-biodegradable wastes. In the next stage, biodegradable and non-biodegradable waste materials were further categorized into nine different generic classes. In the final stage, thirty-six types of waste were classified using the proposed models.

4.1. First stage: binary classification results

4.1.1. PL-CNN-PI-ELM, PL-CNN-L1-RELM, and PL-CNN-En-ELM

The PL-CNN model was initially utilized and trained using a dataset consisting of 35,264 images that represent two separate waste categories, as shown in Table 1 and Fig. 1. The PL-CNN (excluding DSC) was subjected to individual testing using a dataset consisting of 3527 images. The performance of the PL-CNN feature extractor was evaluated using the PI-ELM, L1-RELM, and En-ELM classifiers for each class. The results of these assessments are summarized in Table 3.

The PL-CNN-PI-ELM exhibited an average test precision of 92.5 ± 0.03 %, a recall of 92.0 ± 0.07 %, and an F1-score of 92.5 ± 0.04 %. The accuracy and AUC were reported as 94.0 % and 97.2 %, respectively. Employing the L1-RELM classifier with the feature extractor resulted in average precision, recall, and F1-score values of 92.5 ± 0.03 %, 91.5 ± 0.06 %, and 92.0 ± 0.05 %, respectively. The average accuracy and AUC were 94.0 % and 97.19 %, respectively. Conversely, the hybrid PL-CNN-En-ELM demonstrated an average precision of 92.5 ± 0.03 %, a recall of 92.0 ± 0.07 %, and an F1-score of 92.0 ± 0.05 %. The accuracy and AUC were reported as 94.0 % and 97.27 %, respectively. All three classifiers exhibited identical accuracy results in the PL-CNN method, but En-ELM surpassed the other two classifiers in terms of precision, recall and AUC. The class-wise ROC curves are visualized in Fig. 6.

4.1.2. DP-CNN-PI-ELM, DP-CNN-L1-RELM, and DP-CNN-En-ELM (Proposed method)

The DP-CNN approach was assessed using the PI-ELM, L1-RELM, and En-ELM classifiers. The evaluation results are presented in Table 4, which demonstrates a significant performance evaluation for the two waste categories. For biodegradable waste (Class 0), all classifiers exhibited precision, recall, and F1-Score values of 0.93, indicating that they have the capacity to detect this type of waste accurately. Similarly, regarding non-biodegradable waste (Class 1), every classifier demonstrated precision, recall, and F1-score values of 0.97, confirming their expertise in differentiating non-biodegradable materials. The performance of all classifiers consistently achieved an average precision, recall, and F1-score of 95.0 ± 0.02 %. The classifiers demonstrated a substantial accuracy of 96 %, highlighting their dependability in producing precise predictions. Moreover, the AUC values varied between 98.68 % and 98.78 %, indicating that the models were able to maintain a balance between true positive and false positive rates. Thus, when combined with the PI-ELM, L1-RELM, or Ensemble-ELM classifiers, the DP-CNN approach consistently and efficiently performs waste classification, validating its ability to handle various categories. The confusion matrices for all three classifiers are displayed in Fig. 7, which provides essential information about the classification results. Fig. 7C shows that compared with PI-ELM, En-ELM decreases the misclassification rate for non-biodegradable (class 1) waste. The classwise ROC curves are visualized in Fig. 8.

Table 3

First stage: Binary classification performances of the PL-CNN-PI-ELM, PL-CNN-L1-RELM, and PL-CNN-En-ELM models.

Class Name	PL-CNN-PI-ELM			PL-CNN-L1-RELM			PL-CNN-En-ELM		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Biodegradable (0)	0.9	0.87	0.89	0.9	0.87	0.88	0.9	0.87	0.88
Nonbiodegradable (1)	0.95	0.97	0.96	0.95	0.96	0.96	0.95	0.97	0.96
Average (μ) ± SD (σ) (%)	**92.5 ± 0.03**	**92.0 ± 0.07**	**92.5 ± 0.04**	**92.5 ± 0.03**	91.5 ± 0.06	92.0 ± 0.05	**92.5 ± 0.03**	**92.0 ± 0.07**	92.0 ± 0.05
Accuracy (%)	**94.0**			**94.0**			**94.0**		
AUC (%)	97.2			97.19			**97.27**		

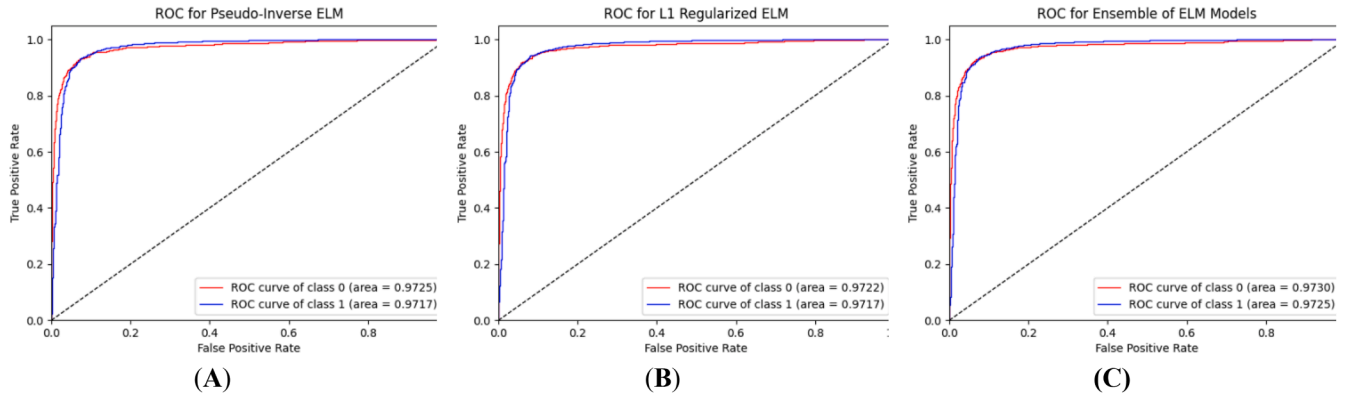Note: Bold values indicate the best results.

Fig. 6. Classwise ROCs of (A) PL-CNN-Pseudo-Inverse-ELM, (B) PL-CNN-L1-Regularized-ELM, and (C) PL-CNN-Ensemble-ELM models for first-stage binary classification.

Table 4

First stage: Binary classification performances of the DP-CNN-PI-ELM, DP-CNN-L1-RELM, and DP-CNN-En-ELM architectures.

Class Name	DP-CNN-PI-ELM			DP-CNN-L1-RELM			DP-CNN-En-ELM		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Biodegradable (0)	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
Nonbiodegradable (1)	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Average (μ) ± SD (σ) (%)	95.0 ± 0.02	95.0 ± 0.02	95.0 ± 0.02	95.0 ± 0.02	95.0 ± 0.02	95.0 ± 0.02	95.0 ± 0.02	95.0 ± 0.02	95.0 ± 0.02
Accuracy (%)	96			96			96		
AUC (%)	98.68			**98.78**			98.77		

Note: Bold values indicate the best results.
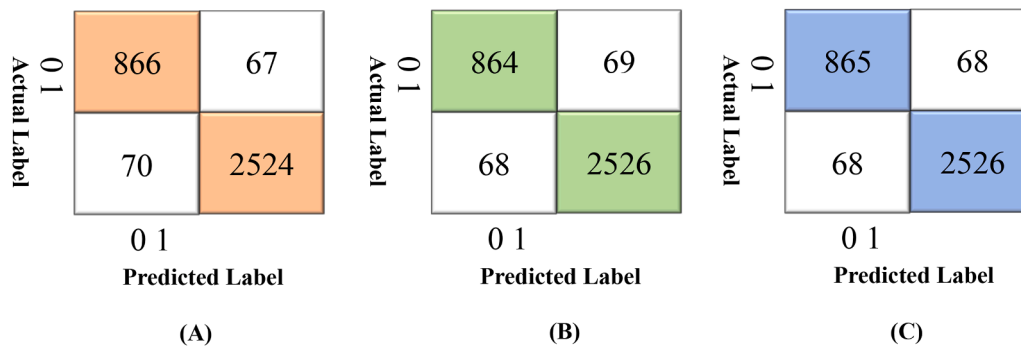
Fig. 7. Confusion matrices of DP-CNN model with (A) DP-CNN-PI-ELM, (B) DP-CNN-L1-RELM, and (C) DP-CNN-En-ELM for first-stage binary classification.

4.1.3. Performance comparison among the proposed model and other TL models

Table 5 displays the performance metric comparisons of the PL-CNN and DP-CNN with other TL models using the En-ELM classifier in terms of binary class classification results at the first stage. Among the six transfer learning models, VGG16 and Xception achieved the highest classification performances. The average precision, recall, F1-score, accuracy, and AUC of the VGG16 model were 96.0 ± 0 %, 93.5 ±

0.07 %, 94.5 ± 0.03 %, 96 %, and 95.93 %, respectively. Similarly, Xception achieved 96.5 ± 0.02 % precision, 92.0 ± 0.09 % recall, 94.0 ± 0.04 % f1-score, 96 % accuracy, and 96.01 % AUC. The MobileNetV3Small-TL model demonstrated the poorest performance among all the models. The DenseNet201 model achieved the highest AUC of 97.7 % compared to the other TL models. The table also demonstrates that the proposed DP-CNN-En-ELM achieved an accuracy of 96.0 % for binary class classification, similar to VGG16 and Xception,

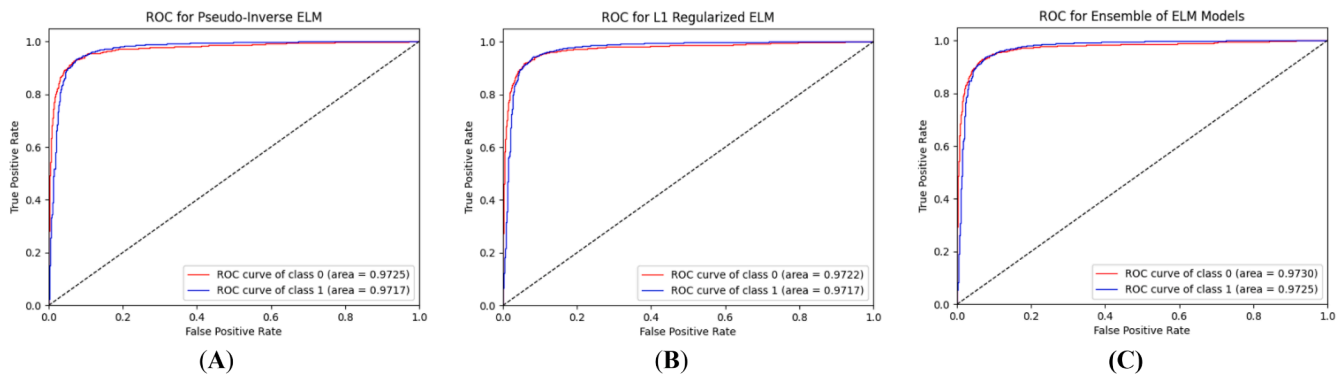**Fig. 8.** Classwise ROCs of (A) DP-CNN-PI-ELM, (B) DP-CNN-L1-RELM, and (C) DP-CNN-En-ELM for first-stage binary classification.

Table 5
Binary classification performance of the TL models integrating the En-ELM and the proposed models.

Model Name	mP (%)	mR (%)	mF1 (%)	Accuracy (%)	AUC (%)
DenseNet201-En-ELM	95.5 ± 0.02	90.5 ± 0.12	92.5 ± 0.04	95	97.7
InceptionResNetV2-En-ELM	94.5 ± 0.02	89.0 ± 0.14	91.0 ± 0.07	93	95
MobileNetV3Small-En-ELM	71.0 ± 0.12	63.5 ± 0.40	65.5 ± 0.28	77	78.23
ResNet152V2- En-ELM	93.5 ± 0.02	88.0 ± 0.15	90.0 ± 0.07	93	94.53
VGG16- En-ELM	96.0 ± 0	93.5 ± 0.07	94.5 ± 0.03	96	95.93
Xception- En-ELM	**96.5 ± 0.02**	92.0 ± 0.09	94.0 ± 0.04	96	96.01
PL-CNN- En-ELM	92.5 ± 0.03	92.0 ± 0.07	92.0 ± 0.05	94.0	97.27
DP-CNN- En-ELM	95.0 ± 0.02	**95.0 ± 0.02**	**95.0 ± 0.02**	**96.0**	**98.77**

Note: Bold values indicate the best results.

and 2.0 % higher than that of the PL-CNN-En-ELM. In addition, the AUC achieved a greater value of 98.77 % compared to DenseNet201. The ROC curves for all the TL models and the PL-CNN and DP-CNN models are displayed in Figs. 9 and 10 presents a bar chart that visually represents the overall results of the models.

4.2. Second stage: results for nine-class classification

4.2.1. PL-CNN-PI-ELM, PL-CNN-L1-RELM, and PL-CNN-En-ELM

In the second stage, the same classification methodology was employed, broadening the process by incorporating the classification of waste images into nine different classes. Table 6 displays an in-depth evaluation of the performance of three classifiers: PL-CNN-PI-ELM, PL-CNN-L1-RELM, and PL-CNN-En-ELM. The performances of all three classifiers are excellent for Green Waste (Class 0), with precision, recall, and F1-score values of 0.87, 0.94, and 0.9, respectively, for the En-ELM classifier. The classifiers demonstrate their robustness in managing various waste types, as evidenced by the consistent patterns detected in other classes. Leather and Fabric (Class 5) have exceptional precision, recall, and F1 scores for all classifiers, highlighting their effectiveness in recognizing this particular type of waste.

The precision, recall, and f1-scores of the PL-CNN-PI-ELM model were 94.25±0.04 %, 93.75±0.04 %, and 94±0.84 %, respectively. The PL-CNN-L1-RELM achieved the highest precision of 81.88 ± 0.06 %. Moreover, PL-CNN-En-ELM had the highest recall and F1-score, with values of 81.44 ± 0.11 (0.68 % greater than those of L1-RELM and 0.95 % greater than those of PI-ELM) and 81.66 ± 0.07 (0.53 % greater than those of L1-RELM and 0.95 % greater than those of PI-ELM),

respectively. The accuracy is 83.0 % across all classifiers, indicating the classifiers' efficacy in accurately predicting outcomes. The AUC values ranged from 96.88 % to 97.19 %, where En-ELM successfully achieved the highest AUC value of 97.19 %. However, the En-ELM consistently outperforms the other methods in terms of the recall, F1-score, and AUC across all the waste classes. Notably, it excels in categorizing waste materials composed of leather and fabric (Class 5). The ROC curves for all the models are displayed in Fig. 11.

4.2.2. DP-CNN-PI-ELM, DP-CNN-L1-RELM, and DP-CNN-En-ELM (Proposed method)

Fig. 12 and Table 7 provide valuable insights into the performance of three classifiers utilizing the DP-CNN model. All classifiers demonstrated exceptional effectiveness during the second round of classification, with precision, recall, and F1-scores consistently exceeding 89 %. The En-ELM approach stands out as particularly effective in accurately classifying waste images across diverse classes, showing a significant performance advantage over alternative models. The integration of PI-ELM and L1-RELM into an En-ELM classifier introduces a novel model that outperforms others. These outcomes provide a strong basis for the utilization of the En-ELM classifier in the classification of waste images. Across all 9 categories, DP-CNN-En-ELM achieved the highest average precision of 90.0 ± 0.04 % (0.13 % greater than that of PI-ELM) and an F1 score of 89.66 ± 0.05 % (0.12 % greater than that of L1-RELM). DP-CNN-PI-ELM achieved the highest recall of 89.55 ± 0.05 %. All three classifiers achieved a similar accuracy of 91 %. The En-ELM has the maximum AUC of 98.57 %, exceeding the AUC of 98.42 % for the PI-ELM and 98.46 % for the L1-RELM. The confusion matrices in Fig. 13 prove that the En-ELM decreases the misclassification rate for green waste (class 0), recyclable waste (class 1), polymer (petroleum-based) waste (class 4), leather and fabric (class 5), and medical waste (class 6).

4.2.3. Performance comparison among the proposed model and other TL models

Table 8 displays the performance metric comparisons between the DP-CNN and other TL models using the En-ELM classifier. The comparisons are based on a total of 9 multiclass classification results in the second stage. DenseNet201 outperformed the other five transfer learning models in classifying all nine categories. The average precision, recall, F1-score, accuracy, and AUC were 90.77 ± 0.04 %, 88.22 ± 0.08 %, 89.11.92 ± 0.04 %, 90.0 %, and 96.05 %, respectively. The MobileNetV3Small-TL model exhibited the poorest performance among all the models. The table also demonstrates that the DP-CNN-En-ELM model achieved an accuracy of 91.0 % for the classification task with nine classes, which is approximately 1.0 % greater than the accuracy attained by the DenseNet201 model and 8.0 % greater than that of the PL-CNN model. In addition, the AUC achieved a greater value of 98.57 % compared to DenseNet201, which achieved an AUC of 96.05 %. The average precision, recall, and f1-score of the DP-CNN model were 90.0 ±
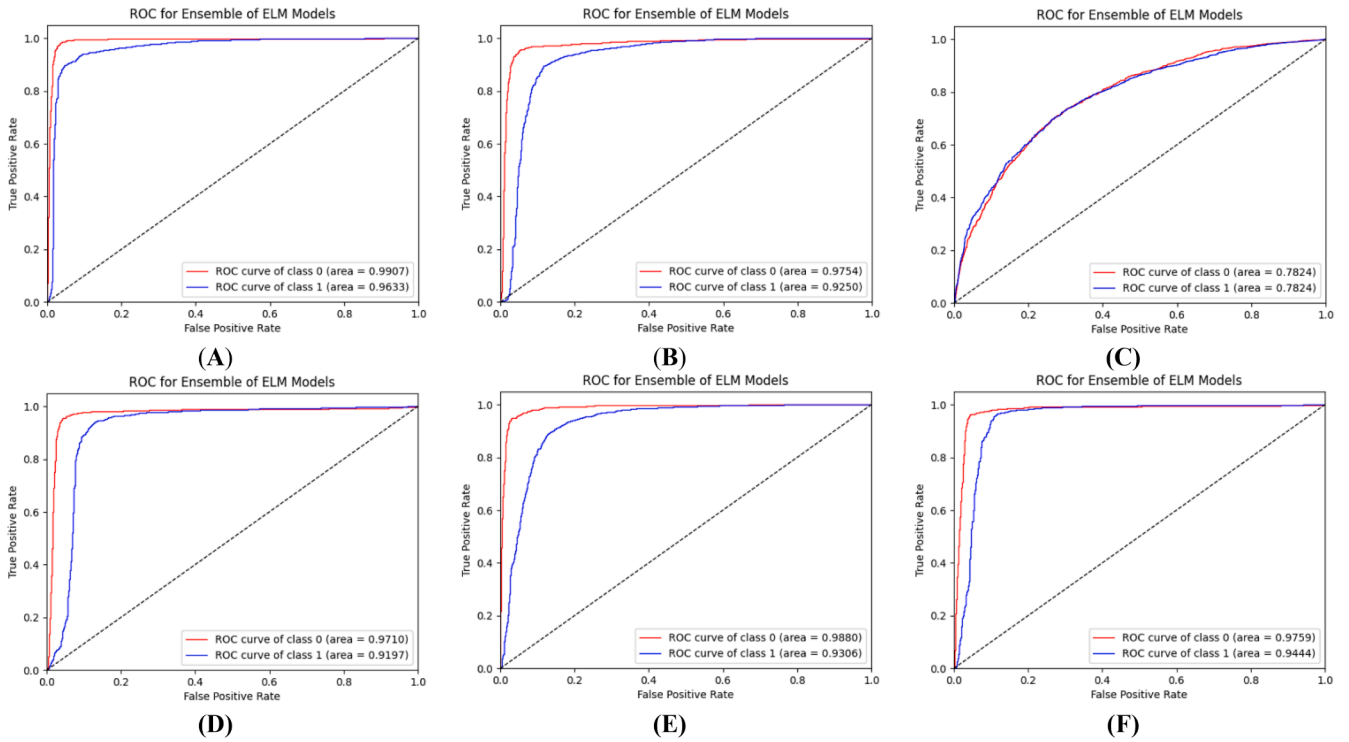
Fig. 9. Classwise ROCs on (A) DensNet201, (B) InceptionResNetV2, (C) MobileNetV3Small, (D) ResNet152V2, (E) VGG16, and (F) Xception for En-ELM for first-stage binary classification.
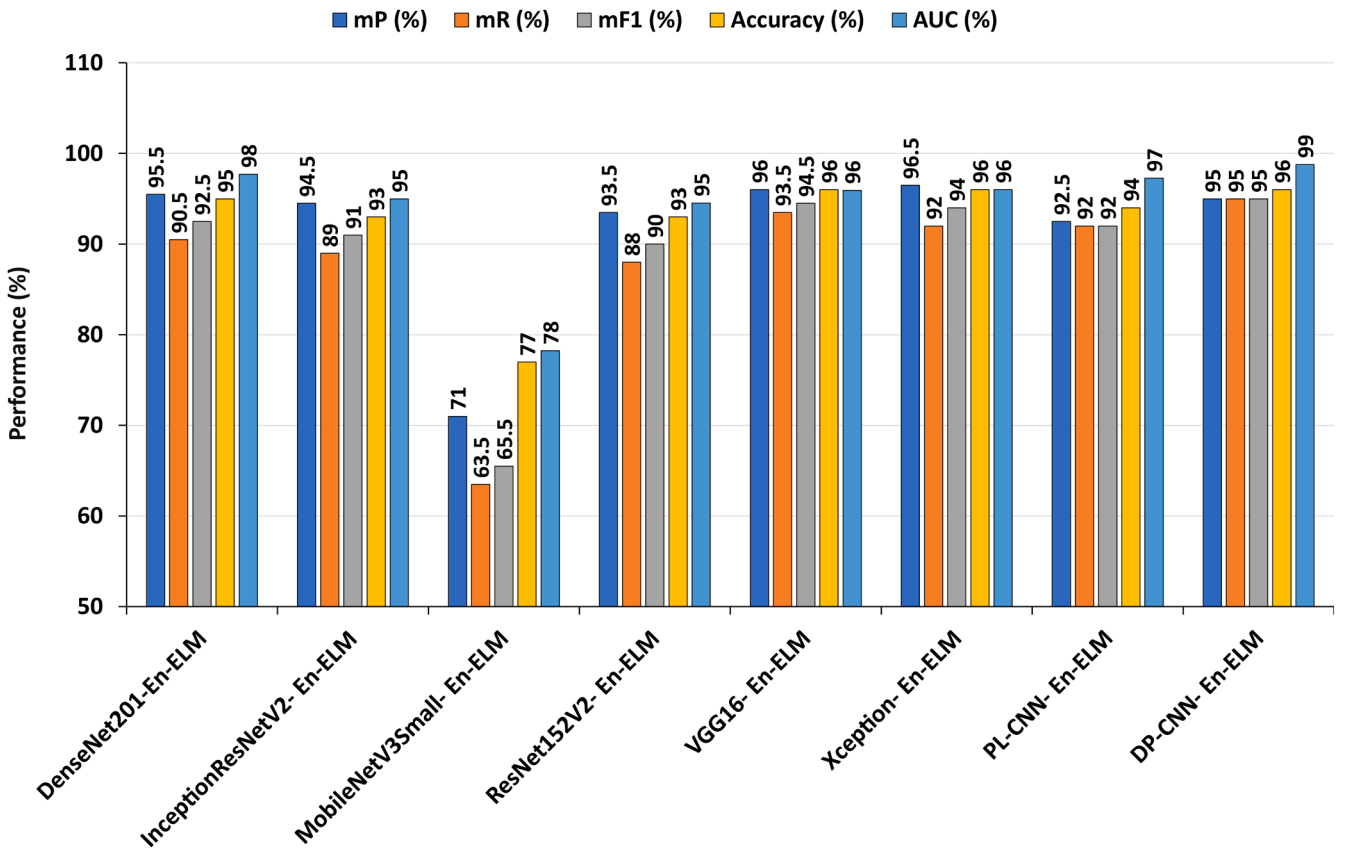

Fig. 10. Performance of the proposed and TL models for two-class classification.

Table 6
Second Stage: Nine-class performances of the PL-CNN-PI-ELM, PL-CNN-L1-RELM, and PL-CNN-En-ELM.

Class Name	PL-CNN-PI-ELM			PL-CNN-L1-RELM			PL-CNN-En-ELM		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Green Waste (0)	0.87	0.94	0.9	0.87	0.95	0.91	0.87	0.94	0.9
Recyclable Waste (1)	0.83	0.82	0.82	0.83	0.81	0.82	0.83	0.82	0.82
Glass (2)	0.87	0.9	0.88	0.87	0.91	0.89	0.88	0.93	0.9
Metal (3)	0.76	0.73	0.75	0.75	0.74	0.74	0.76	0.76	0.76
Polymer (Petroleum Based) (4)	0.72	0.67	0.69	0.72	0.68	0.7	0.71	0.68	0.7
Leather and Fabric (5)	0.91	0.95	0.93	0.91	0.95	0.93	0.91	0.96	0.93
Medical Waste (6)	0.76	0.72	0.74	0.77	0.72	0.75	0.78	0.72	0.75
E-waste (7)	0.78	0.85	0.81	0.78	0.84	0.81	0.80	0.84	0.82
Hazardous Waste (8)	0.86	0.68	0.76	0.87	0.68	0.76	0.88	0.68	0.77
Average (μ) ± SD (σ) (%)	**81.77 ± 0.06**	**80.66 ± 0.11**	**80.88 ± 0.08**	**81.88 ± 0.06**	**80.88 ± 0.11**	**81.22 ± 0.08**	**82.44 ± 0.06**	**81.44 ± 0.11**	**81.66 ± 0.07**
Accuracy (%)	**83.0**			**83.0**			**83.0**		
AUC (%)	96.88			97.11			**97.19**		

Note: Bold values indicate the best results.
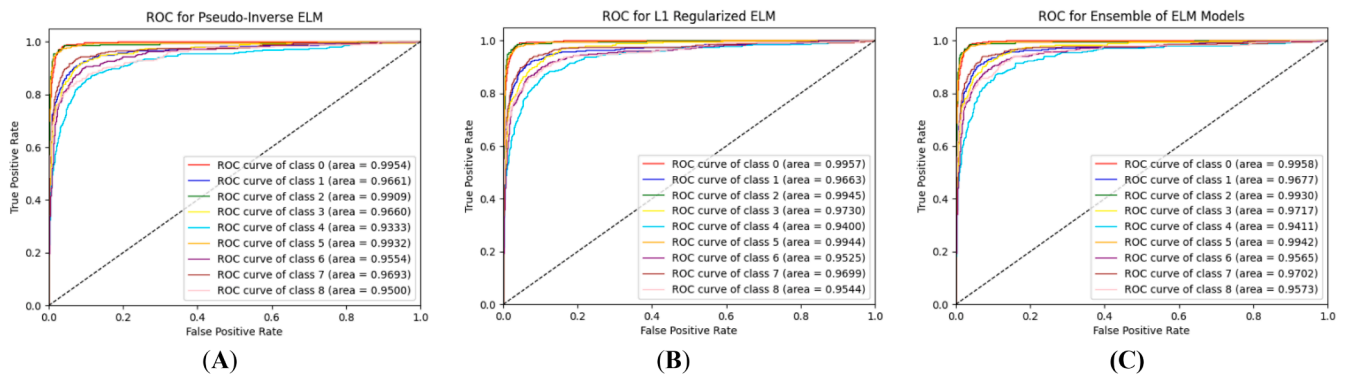
Fig. 11. Classwise ROCs of (A) PL-CNN-PI-ELM, (B) PL-CNN-L1-RELM, and (C) PL-CNN-En-ELM for second-stage nine-class classification.
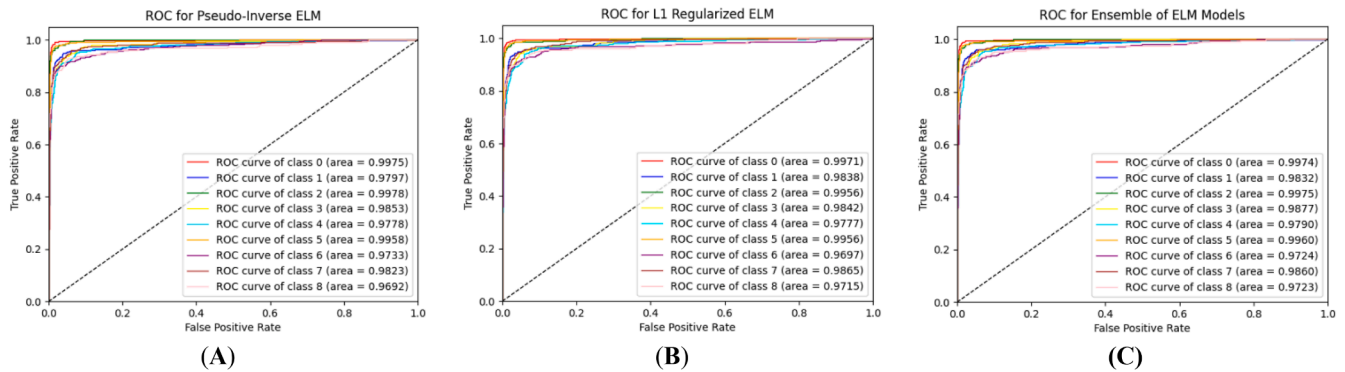
Fig. 12. Classwise ROCs of (A) DP-CNN-PI-ELM, (B) DP-CNN-L1-RELM, and (C) DP-CNN-En-ELM for second-stage nine-class classification.

0.04 %, 89.44 ± 0.06 (1.36 % gain), and 89.66 ± 0.05 (0.61 % gain), respectively. Fig. 14 displays the ROC curve of the TL model for all nine classes. Fig. 15 presents a bar chart that visually represents the overall performance of the models.

4.3. Third stage: results for thirty-six class classifications using the proposed model and TL approaches

Table 9 provides a comprehensive overview of the performance in the third-stage classification for different TL models, PL-CNN, and DP-CNN models employing various feature extractors and classifiers. Each row corresponds to a specific model, while the columns showcase the performance metric values. The classifiers utilized include PI-ELM, L1-

RELM, and En-ELM. After evaluation, a consistent trend emerges across different feature extractors, indicating that the En-ELM classifier tends to yield superior results in comparison to the PI-ELM and L1-RELM classifiers. Notably, the En-ELM classifier consistently exhibits strong precision, recall, F1 score, accuracy, and AUC metrics when combined with the proposed DP-CNN feature extractor. This suggests that these combinations offer robust and reliable classification performance.

For the PI-ELM classifier, DenseNet201 and VGG16 are the most effective TL models. DenseNet201 attains the highest precision (83.51 %) and F1 score (78.83 %). On the other hand, VGG16 achieved the highest recall (80.72 %), accuracy (80.72 %), and AUC (97.75 %). However, the proposed PL-CNN-PI-ELM outperforms both models. The proposed model achieved a precision of 84.43 % (1.08 % greater than

Table 7

Second-stage nine-class performances of the DP-CNN-PI-ELM, DP-CNN-L1-RELM, and DP-CNN-En-ELM architectures.

Class Name	DP-CNN-PI-ELM			DP-CNN-L1-RELM			DP-CNN-En-ELM		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Green Waste (0)	0.95	0.98	0.96	0.95	0.98	0.96	0.95	0.98	0.96
Recyclable Waste (1)	0.92	0.91	0.92	0.92	0.91	0.92	0.93	0.91	0.92
Glass (2)	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
Metal (3)	0.84	0.86	0.85	0.83	0.85	0.84	0.83	0.85	0.84
Polymer (Petroleum Based) (4)	0.85	0.85	0.85	0.86	0.83	0.84	0.85	0.84	0.85
Leather and Fabric (5)	0.96	0.97	0.96	0.96	0.97	0.96	0.96	0.97	0.97
Medical Waste (6)	0.88	0.86	0.87	0.87	0.86	0.87	0.88	0.87	0.87
E-waste (7)	0.88	0.9	0.89	0.88	0.9	0.89	0.88	0.9	0.89
Hazardous Waste (8)	0.88	0.8	0.84	0.9	0.81	0.85	0.89	0.8	0.84
Average (μ) ± SD (σ) (%)	89.88 ± 0.04	**89.55 ± 0.05**	89.66 ± 0.04	**90.0 ± 0.04**	89.33 ± 0.06	89.55 ± 0.04	**90.0 ± 0.04**	89.44 ± 0.06	**89.66 ± 0.05**
Accuracy (%)	91			91			91		
AUC (%)	98.42			98.46			**98.57**		

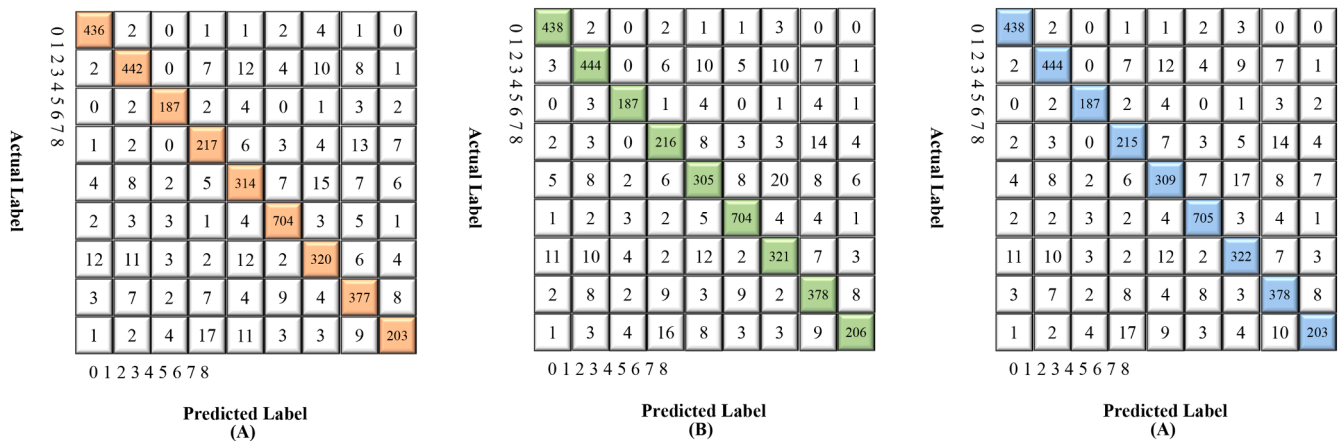Note: Bold values indicate the best results.

Fig. 13. Confusion matrices of the DP-CNN model with (A) PI-ELM and (B) L1-RELM (C) En-ELM for nine classes.

Table 8

Nine-class classification performances of the TL models that integrate Ensemble-ELM and the proposed models.

Model Name	mP (%)	mR (%)	mF1 (%)	Accuracy (%)	AUC (%)
DenseNet201-En-ELM	**90.77 ± 0.04**	88.22 ± 0.08	89.11 ± 0.04	90.0	96.05
InceptionResNetV2-En-ELM	87.33 ± 0.04	84.33 ± 0.15	85.0 ± 0.10	87.0	94.90
MobileNetV3Small-En-ELM	44.66 ± 0.10	42.88 ± 0.20	43.22 ± 0.15	49.0	83.57
ResNet152V2-En-ELM	90.55 ± 0.06	87.11 ± 0.08	88.22 ± 0.03	89.0	94.31
VGG16-En-ELM	90.0 ± 0.03	87.88 ± 0.07	88.66 ± 0.03	90.0	96.60
Xception-En-ELM	89.77 ± 0.05	87.55 ± 0.08	88.22 ± 0.04	90.0	98.44
PL-CNN-En-ELM	82.44 ± 0.06	81.44 ± 0.11	81.66 ± 0.07	83.0	97.19
DP-CNN-En-ELM	90.0 ± 0.04	**89.44 ± 0.06**	**89.66 ± 0.05**	**91.0**	**98.57**

Note: Bold values indicate the best results.

DenseNet201), a recall of 84.77 % (4.77 % greater than VGG16), an F1 score of 84.00 % (6.15 % greater than DenseNet201), an accuracy of 84.77 % (4.77 % greater than VGG16), and an AUC of 98.51 % (0.77 % greater than VGG16).

In the context of the L1-RELM classifier, DenseNet201 achieved the highest precision (82.91 %), Xception achieved the highest recall (80.40 %), F1 score (78.22 %), and accuracy (80.40 %), while VGG16 achieved the highest AUC (97.75 %) among the other TL models. However, the PL-CNN model achieved notable improvements over these other models by achieving precision, recall, F1 score, accuracy, and AUC values of 83.19 %, 83.52 %, 82.60 %, 83.52 %, and 98.59 %, respectively. However, the proposed DP-CNN-L1-RELM model outperforms all the other models. This model achieved a precision of 84.55 % (exceeding that of the PL-CNN by 1.60 %), a recall of 84.80 % (a notable 1.50 % improvement), an F1 score of 84.03 % (showing a 1.70 % gain), an accuracy of 84.80 % (a substantial 1.50 % gain over that of the PL-CNN), and an AUC of 98.45 %.

For the En-RELM classifier, DenseNet201 achieved the highest precision (86.16 %) among the TL models. However, the PL-CNN model demonstrates notable enhancements, achieving precision, recall, F1 score, accuracy, and AUC values of 83.49 %, 83.75 %, 82.85 %, 83.75 %, and 98.70 %, respectively. However, the proposed DP-CNN-L1-RELM model outperforms all the other models. This model had a precision of 85.02 %, a recall of 85.25 % (a 1.75 % greater than that of the PL-CNN), an F1 score of 84.54 % (a 2 % greater than that of the PL-CNN), an accuracy of 85.25 % (a substantial 1.75 % greater than that of the PL-CNN), and an AUC of 98.68 %. While individual model performances may vary, the En-ELM has emerged as a promising classifier, demonstrating its effectiveness in diverse scenarios. In summary, the DP-CNN-En-ELM model stands out as the best choice because of its superior performance and effectiveness in handling the complexities of waste classification problems. Fig. 16 shows the ROC curves of all classifiers across all the models. Fig. 17 presents a bar chart that visually represents the overall performance of the En-ELM classifier.
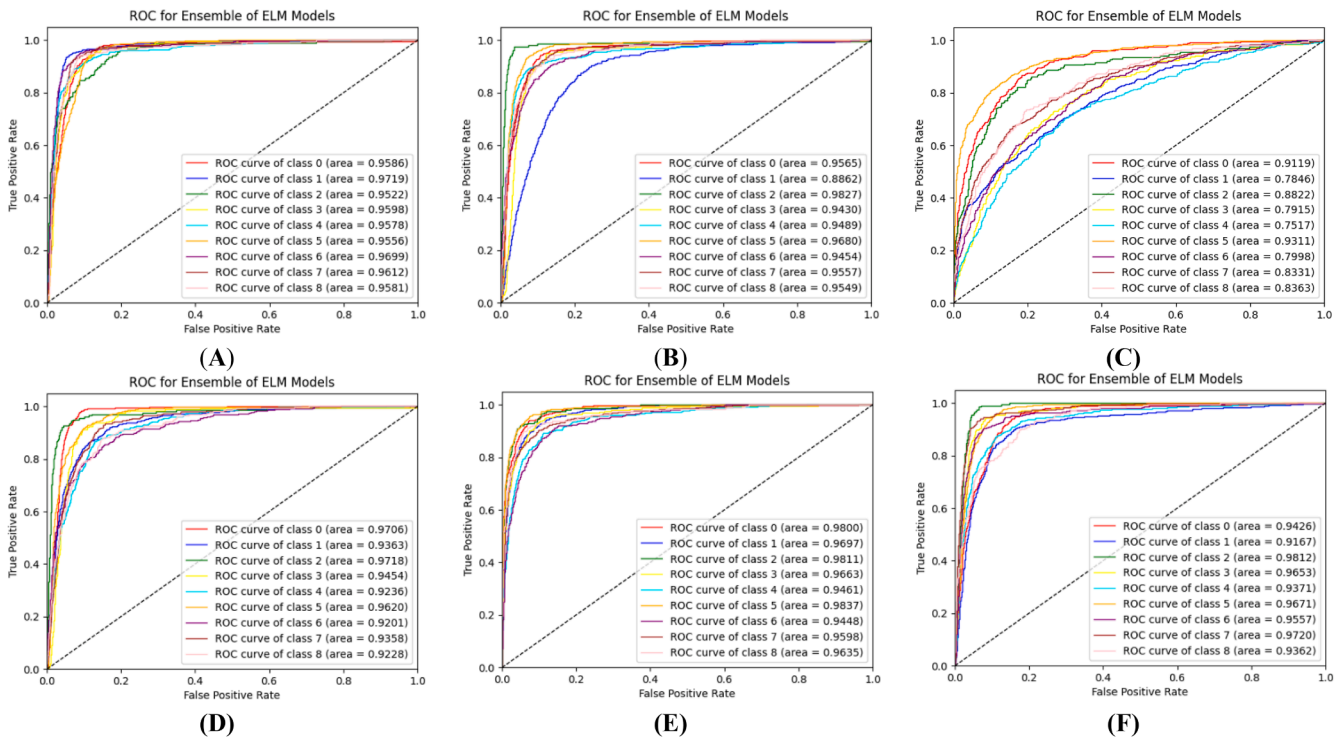
Fig. 14. Classwise ROCs on (A) DensNet201, (B) InceptionResNetV2, (C) MobileNetV3Small, (D) ResNet152V2, (E) VGG16, and (F) Xception for second stage-nine class classification.
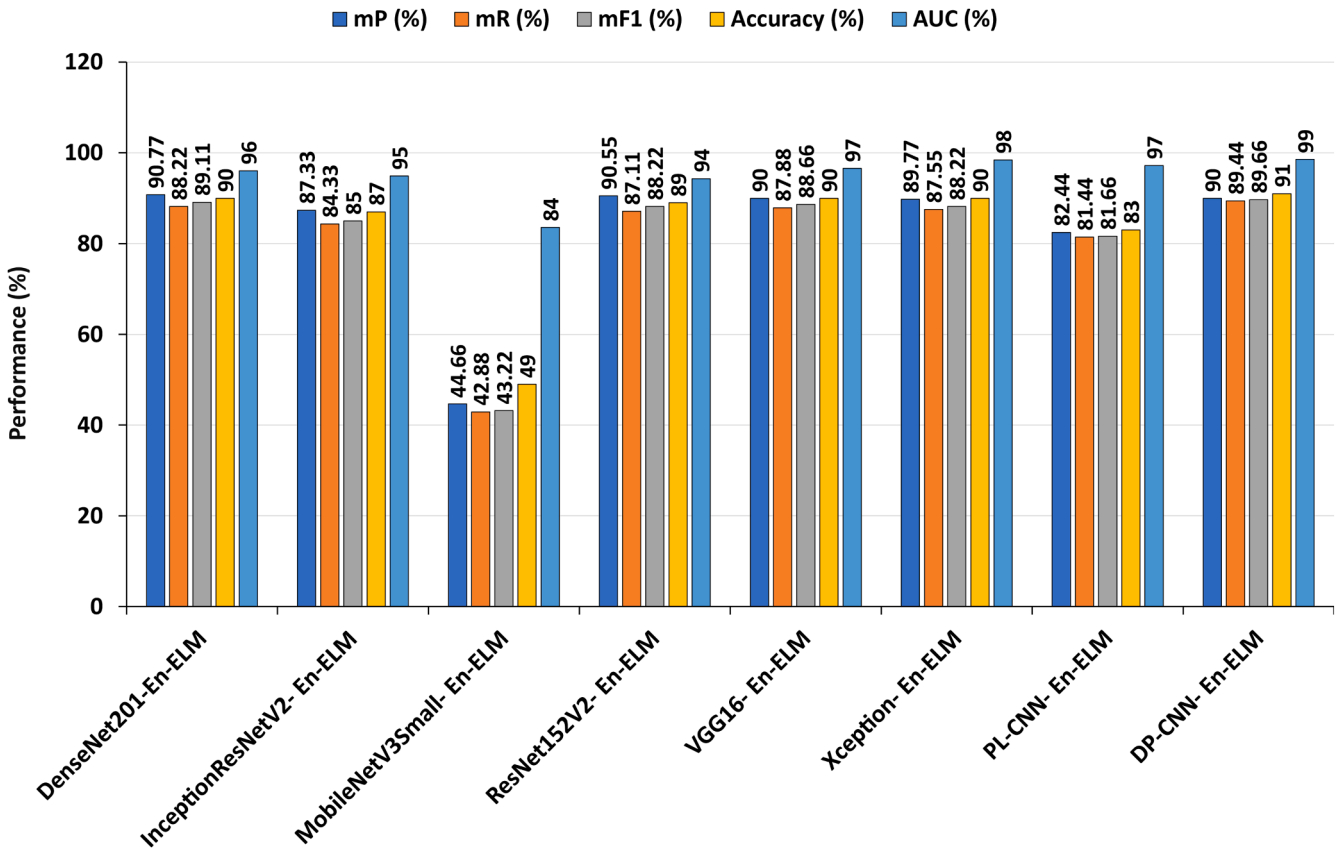

Fig. 15. Performance of the proposed and TL models for nine-class classification.

Table 9
Third Stage 36 class performances achieved by using the PL-CNN, DP-CNN, and TL architectures.

Feature Extractor	Classifier	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)	AUC (%)
DenseNet201	PI-ELM	83.51	80.49	78.83	80.49	97.37
	L1-RELM	82.91	78.84	77.44	78.84	97.12
	En-ELM	**86.16**	83.52	82.55	83.52	98.04
InceptionResNetV2	PI-ELM	73.58	72.12	68.93	68.93	96.32
	L1-RELM	73.96	71.44	67.37	67.37	96.04
	En-ELM	77.98	76.24	73.24	73.24	97.23
MobileNetV3Small	PI-ELM	42.75	45.90	40.41	45.90	89.92
	L1-RELM	42.55	45.70	40.28	45.70	89.95
	En-ELM	44.26	45.98	40.60	45.98	90.71
ResNet152V2	PI-ELM	83.12	79.10	77.49	79.10	97.21
	L1-RELM	82.83	79.64	77.87	79.64	97.27
	En-ELM	85.69	82.42	80.89	82.42	97.99
VGG16	PI-ELM	81.48	80.72	78.76	80.72	97.75
	L1-RELM	81.30	79.75	77.36	79.75	97.75
	En-ELM	84.37	83.75	82.22	83.75	98.46
Xception	PI-ELM	83.26	81.14	79.31	81.14	97.43
	L1-RELM	82.80	80.40	78.22	80.40	97.64
	En-ELM	85.25	83.49	81.93	83.49	98.18
PL-CNN	PI-ELM	83.24	83.49	82.53	83.49	98.53
	L1-RELM	83.19	83.52	82.60	83.52	98.59
	En-ELM	83.49	83.75	82.85	83.75	**98.70**
DP-CNN	PI-ELM	84.43	84.77	84.00	84.77	98.51
	L1-RELM	84.55	84.80	84.03	84.80	98.45
	En-ELM	85.02	**85.25**	**84.54**	**85.25**	98.68

Note: Bold values indicate the best results.

4.4. Computational time and resource comparison

Table 10 presents a comparison of computational resources among DP-CNN, PL-CNN, and other transfer learning (TL) feature extractors, considering aspects such as parameters, layers, sizes, and testing duration. In terms of the classification accuracy, testing time, and complexity, the DP-CNN outperformed all others, indicating that it is the most effective. Fig. 18 displays the total resources needed. The findings of the research indicate that the DP-CNN-En-ELM strategy is dependable and can be used to effectively classify different types of waste materials.

The ResNet152V2 model possesses the highest parameter count among all models, totaling 91.60 million, comprising 567 layers, and occupying a space of 567,614 MB. In comparison, the InceptionResNetV2 model boasts the largest size of 783,287 MB, accompanied by 783 layers, the most extensive layer count across all models. In contrast, the DP-CNN model exhibits efficiency in loading and processing, featuring a compact size of 12.7 MB, 1.09 million trainable parameters, and 9 CLs. Notably, the proposed model comprises 1.10 million total parameters, which is approximately 83.27 times less than that of ResNet152V2 and 2.45 times less than that of PL-CNN, which encompasses 2.698 million parameters. Additionally, the proposed En-ELM model achieves more favorable and less consecutive processing times—specifically, 0.00001 s for testing. When performance and model complexity are considered, other TL models may have better computational speeds. Despite having fewer layers and smaller sizes, the DP-CNN is highly effective at rapidly sorting waste while using as few resources as possible. For real-world waste management tools, the DP-CNN-En-ELM model stands out as the best option because it offers better performance while maintaining a compact yet efficient structure.

4.5. Comparative analysis of SOTA end-to-end models and proposed multi-stage model

A comparative analysis between the proposed multi-stage model and several state-of-the-art (SOTA) end-to-end models is presented in Table 11 to assess the effectiveness of the proposed DP-CNN-En-ELM approach, which incorporates a multi-stage framework designed to enhance classification accuracy across a broad range of categories. Specifically, this framework achieves an accuracy of 96 % in the first stage with 2 classes, 91 % in the second stage with 9 classes, and 85.25 % in the third stage with 36 classes.

In contrast, several notable end-to-end models exhibit varying performance levels. For instance, Optimized DenseNet169 [13] achieves 82.8 % accuracy with 5 classes, while MDTLDC-IWM [15] attains 99.26 % accuracy with 6 classes. DNN-TC [14] records 98 % accuracy with 3 classes, and Focus-RCNet-KD [28] reaches 92 % accuracy across 6 classes. Optimized DenseNet121 [53] demonstrates exceptional performance with 99.60 % accuracy in a 6-class scenario. Additionally, EfficientNetB2 with PMAM [23] secures 93.38 % accuracy with 4 classes, and Improved MobileNetV2 [25] achieves 90.7 % with 4 classes.

The three-stage architecture of the proposed model is specifically designed to address limitations observed in end-to-end systems, particularly when dealing with a large number of classes and diverse classification tasks. By decomposing the classification process into more manageable stages, each optimized for different levels of classification granularity, the model adapts to a broader spectrum of classification challenges. This hierarchical approach not only improves classification performance but also provides more detailed insights at each stage, which end-to-end systems often struggle to deliver effectively.

Thus, while end-to-end models demonstrate strong performance in specific contexts, the multi-stage model offers a valuable alternative by leveraging its staged design to handle complex classification problems with greater accuracy and granularity. This approach underscores the effectiveness and advantages of a multi-stage methodology over a singular end-to-end solution.

4.6. Interpretability and visualization of DP-CNN-En-ELM using SHAP and Grad-CAM

Shapley values were developed through a methodical analysis of every potential combination of waste features, resulting in representations that are defined by pixels. During the experiment, a clear pattern emerged: red pixels in the explanation images suggested that there was a greater chance of being in the target class. Blue pixels, on the other hand, represented a greater probability of being outside of that class. It is significant to note that, as shown in Fig. 19, the SHAP explanation visuals merge subtle gray backgrounds into the original input images.

Fig. 19-(A) shows two classes of SHAPs: biodegradable and non-biodegradable waste. The upper row of the SHAP explanation image displays red pixels, signifying the detection of non-biodegradable waste.
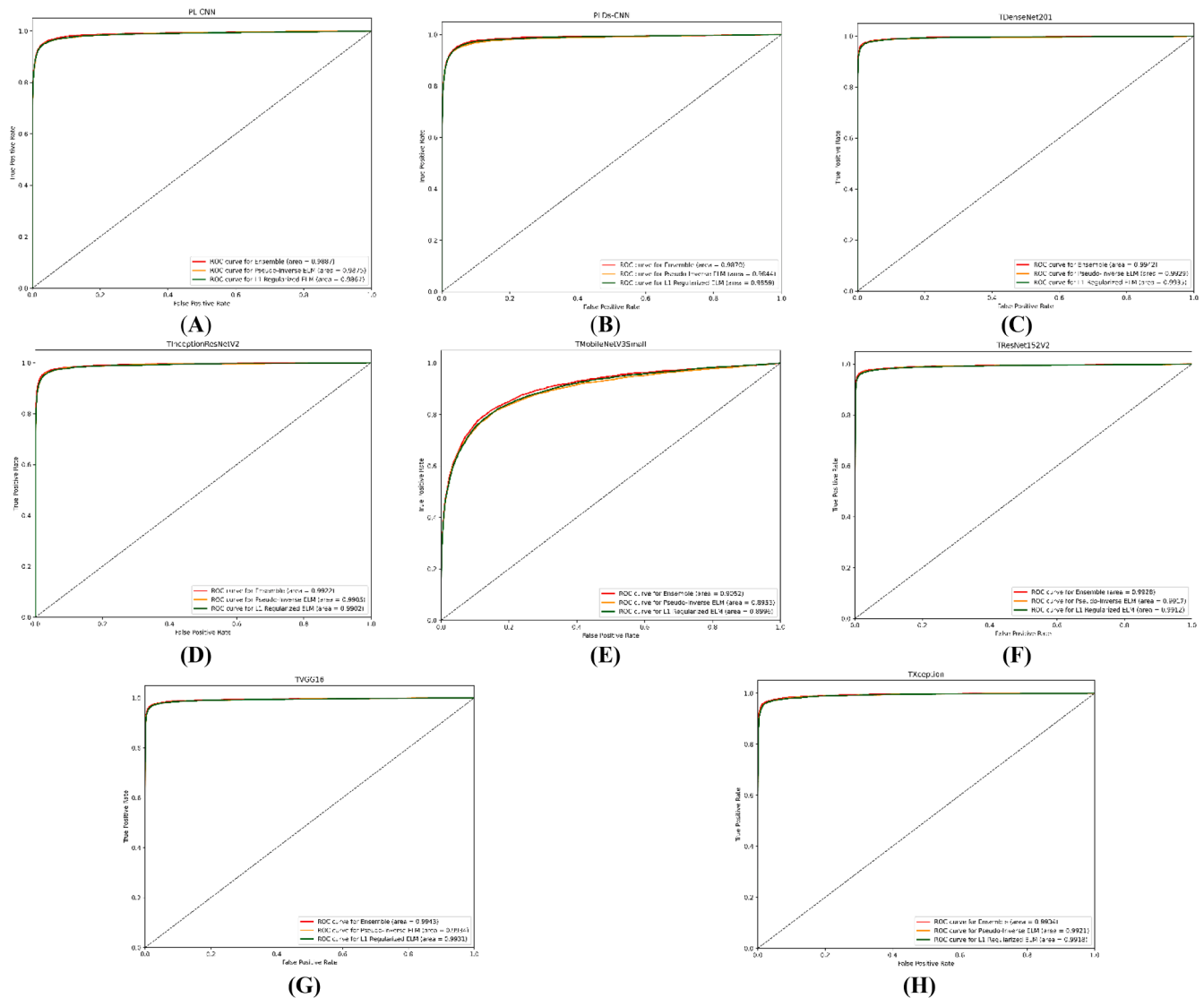
Fig. 16. ROC-AUCs on (A) PL-CNN, (B) DP-CNN, (C) DenseNet201-PI-ELM, (D) InceptionResNetV2, (E) MobileNetV3Small, (F) ResNet152V2, (G) VGG16, (H) Xception with PI-ELM, L1-RELM and En-ELM for third stage-36 class classification.

Conversely, the presence of blue pixels eliminates the possibility of biodegradable waste. The fifth row in the SHAP visuals shows a specific trend: red pixels represent the class of biodegradable waste. In contrast, for the same sample, a higher count of blue pixels was noted in the non-biodegradable waste category, suggesting that the confidence level for this category was relatively lower than that for the other categories. The second stage of the proposed model's testing, which is presented in Fig. 19-(B), involves 9 classes, making the classification even more difficult. The first row reveals that the first explanation image has more red pixels, indicating that the material is green waste. On the other hand, the absence of blue pixels and fewer red pixels effectively eliminated other class categories. Similarly, the second and third rows demonstrate that the sample images are from recyclable waste and polymer, respectively, as shown by the red pixels. A surplus of red pixels in the SHAP explanation image accurately identifies the sample image as belonging to the e-waste class in the fifth row. Despite the presence of more intricate data, the proposed framework was still able to predict the result with higher accuracy.

Model decision-making processes are made transparent through visualizations, which improve model transparency and user trust. Fig. 20 depicts a selection of five images from thirty-six waste categories. The images were randomly selected from the original dataset to provide a fair evaluation of the model's performance. The heatmap, grad-CAM, and guided grad-CAM highlight the model's ability to focus on the most relevant regions of an image and accurately classify various waste types. The waste image heatmap and guided heatmap visualization are shown in the third and fourth columns, respectively. The yellow and green areas highlight the relevant regions of the sample images that are important to the decision-making process. Grad-CAM and guided Grad-CAM depict exactly where the model focuses its attention on an image. The fifth and sixth columns show that the model mainly concentrates on the center of the object in most scenarios. When dealing with larger waste items that constitute an extensive portion of the image, the model focuses on several regions inside the waste area. The seventh column presents guided salience mapping, an interpretability method that combines the ideas of salience mapping and guided backpropagation. Brighter areas in the saliency map depict the pixels that contribute most to the model's classification. The first row of the visualization proves that the proposed model focuses on the right features, as all brighter pixels highlight the cardboard area instead of the irrelevant background. Similarly, in the second and third rows, brighter pixels indicate animal dead bodies and battery areas. In summary, by highlighting key features, areas, or neurons, various visualization techniques substantially improve the interpretability of DL models.
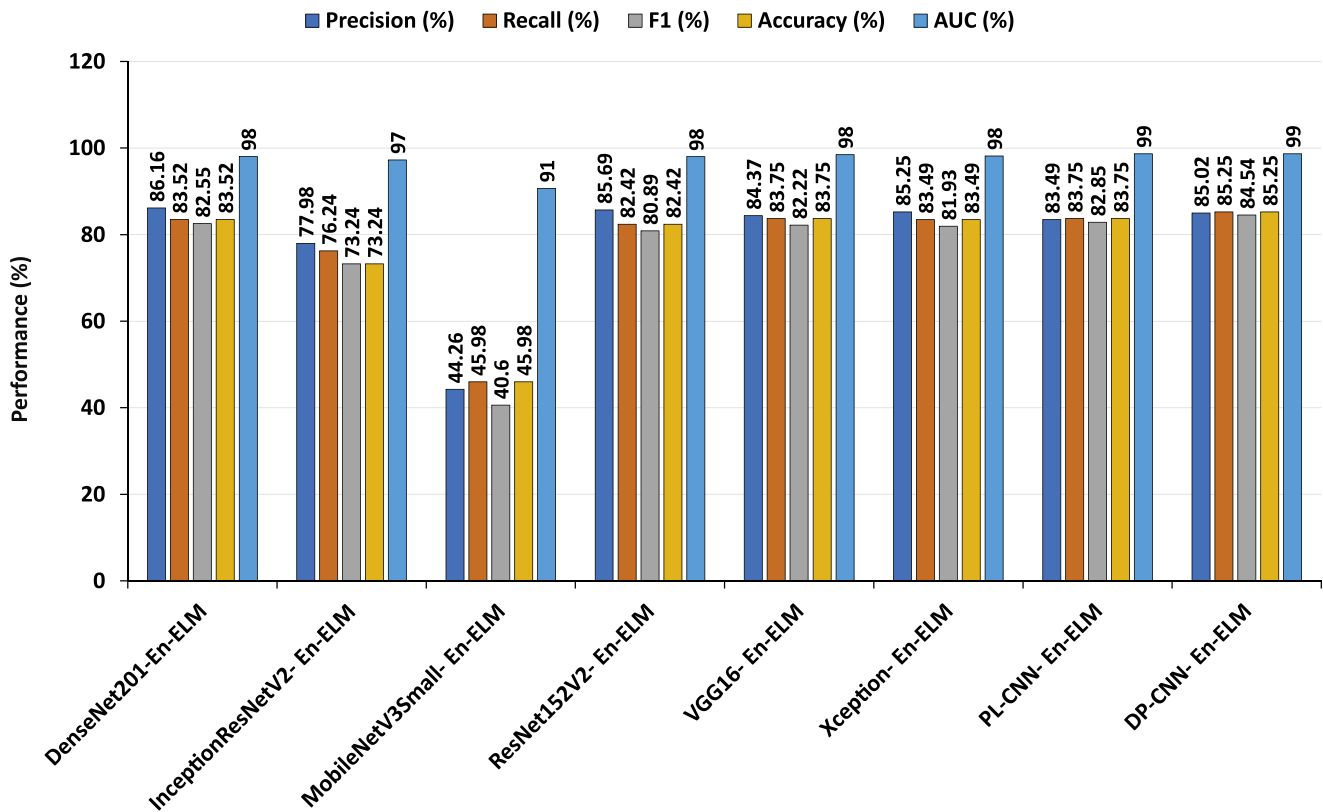
Fig. 17. Performance of the proposed and TL models for thirty-six class classifications.

Table 10
Computational resources (model parameters and size) and time comparison for multiclass classifications.

Criteria		PL-CNN	DP-CNN	DenseNet201	InceptionResNetV2	MobileNetV3Small	ResNet152V2	VGG16	Xception
Total Parameters (Million)		2.698	**1.10**	36.10	60.98	18.41	91.60	19.82	54.13
Trainable Parameters (Million)		2.695	1.09	17.78	6.6	16.88	33.26	5.1	33.26
Number of Layers		9	**9**	710	783	246	567	22	135
Size (Megabyte)		27	**12.7**	710,280	783,287	246,204	567,614	22,116	135,470
1st Stage	PI-ELM	0.0312	0.0349	0.0040	0.00009	0.0334	**0.00006**	0.0156	0.00009
Testing Time (Seconds)	L1-RELM	0.0312	0.0469	0.0087	0.0156	0.0249	**0.00004**	0.00008	0.0027
	En-ELM	0.00005	**0.00001**	0.00006	0.00005	0.00009	0.00003	0.00005	0.00008
2nd Stage Testing Time (Seconds)	PI-ELM	0.0312	0.0312	0.0021	0.00008	0.0156	**0.00006**	0.00008	0.0035
	L1-RELM	0.0312	0.0313	0.00009	0.00009	0.0312	**0.00004**	0.0123	0.0069
	En-ELM	0.00009	**0.00001**	0.00008	0.00006	0.00005	0.00003	0.00009	0.00007
3rd Stage	PI-ELM	0.0130	0.00450	0.0050	0.0101	0.0145	0.0050	0.0303	**0.0028**
Testing Time (Seconds)	L1-RELM	0.0131	**0.00452**	0.0281	0.0101	0.0127	0.0052	0.00458	0.0066
	En-ELM	0.00006	**0.00001**	0.00008	0.00009	0.00006	0.00005	0.0050	0.00009

Note: Bold values indicate the best results.

4.7. Hardware structure

4.7.1. Graphical user interface (GUI)

A hardware-based approach for the utilization of the proposed models in real-time applications is discussed, where a graphical user interface (GUI) app was developed to classify all three stages of the models accurately. The development of the app was carried out using the PYQT5 framework, which provides two individual tasks that are designed to run on desktop platforms. The first task allows for single image classification with proper heatmap visualization; in this case, both SHAP and Grad-CAM were used for visualizing the regions of the provided image that contributed most to the classification. This option was provided not for real-time application but for evaluating each mode's performance and whether it is capable of correctly localizing the crucial points for new sets of images. In addition, it also provides corresponding classes, frame rates, and confidence scores for better

clarification. The second task involved the development of a hardware system with an accurate sorting mechanism for use in industrial waste sorting plants. As manual management of trash is a very unhygienic and challenging task, an automated approach was proposed that used the developed models in addition to a hardware device for effective waste sorting. The flowchart in Fig. 21 provides a more detailed method for GUI task interfaces.

Visualization techniques such as SHAP and Grad-CAM are essential for developing interpretability and transparency in classification tasks. These approaches aid in solving the nontransparent characteristics of deep learning models by isolating the significant features and emphasizing areas that have an impact on prediction. Additionally, they contribute to a clearer understanding of how the model arrives at its decisions. The nature of waste materials is not the same and always changes based on their structure and appearance; thus, the functionality of the developed models is limited. To address this dynamic nature,
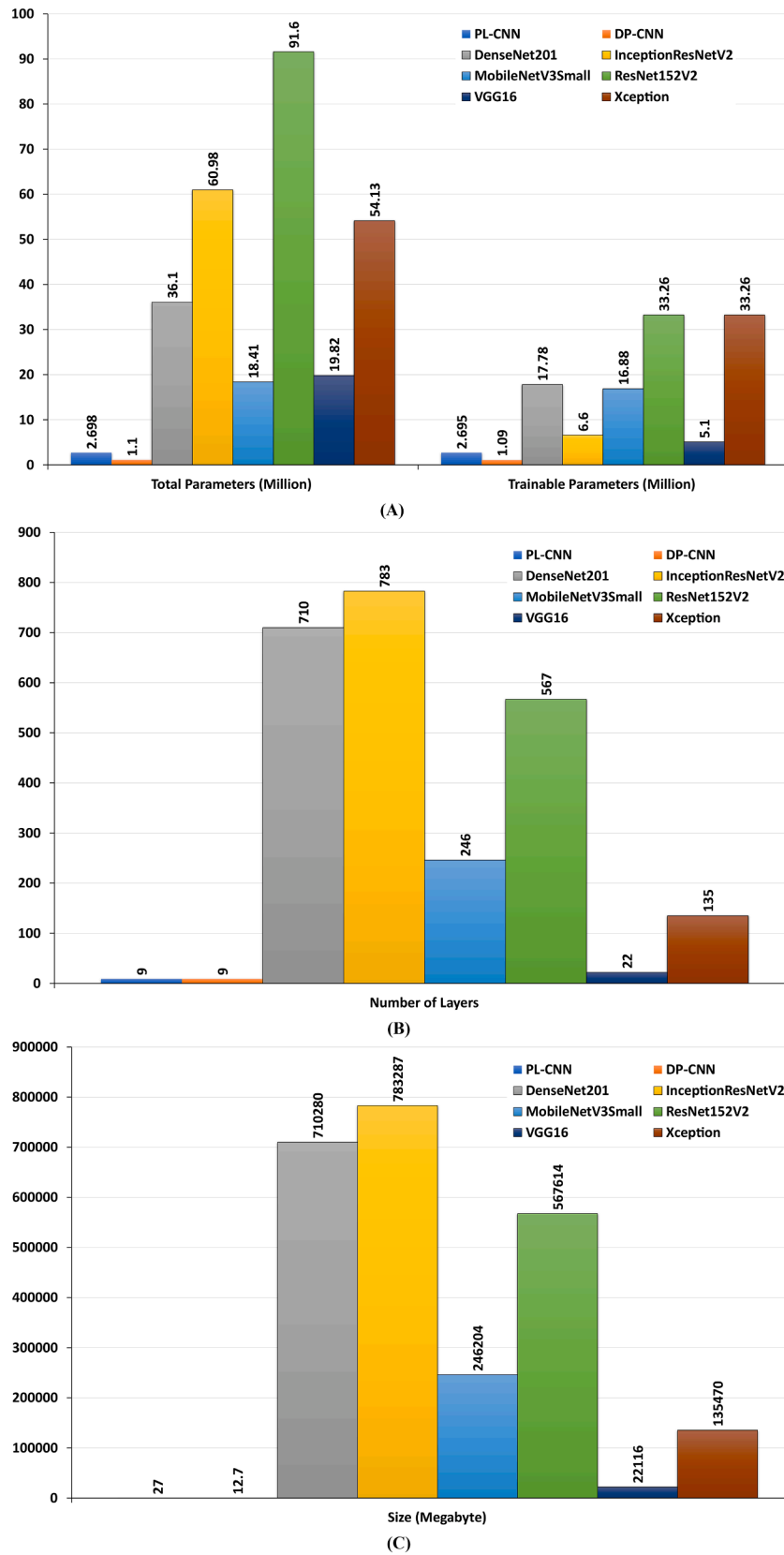
Fig. 18. Computational resources (A) Total parameters and trainable parameters, (B) Number of layers, (C) Size comparisons among the proposed DP-CNN, PL-CNN, and other TL models.

Table 11

Comparative analysis of SOTA End-to-End models vs. proposed multi-stage model.

Model	Dataset	Number of Classes	Accuracy (%)
Optimized DenseNet169 [13]	NWNU-TRASH	5	82.8
MDTLDC-IWM [15]	TR/TS data	6	99.26
DNN-TC [14]	VN-trash	3	98
SWMACM-CA [26]	TrashNet	6	90
Focus-RCNet-KD [28]	TrashNet	6	92
Optimized DenseNet121 [53]	TrashNet	6	99.60
EfficientNetB2 with PMAM [23]	Huawei Cloud Garbage Classification Dataset	4	93.38
Improved MobileNetV2 [25]	Huawei Garbage Classification Challenge Cup Dataset	4	90.7
DP-CNN-En-ELM (Proposed)	TriCascade WasteImage	First Stage: 2 Classes, Second Stage: 9 Classes, Third Stage: 36 Classes	96 (For 2 Classes), 91 % (For 9 Classes), 85.25 % (For 36 Classes)

ongoing model adaptation is crucial. The ability of the models to continuously learn and evolve with new waste materials ensures their relevance and accuracy over time. The first task focuses on overcoming this problem, where its sole objective is to validate whether the models used are capable of classifying different wastes through accurate prediction. This validation process becomes an iterative loop, incorporating new data and refining the models to ensure adaptability to emerging waste patterns. This option uses the developed models to locate and identify the pixel areas of an unseen waste image and verifies whether the models are making accurate predictions by chance. Through continuous validation and testing, the models are fine-tuned to enhance their precision and reliability in diverse scenarios. Fig. 22 shows that the marked regions are on the waste and its edges, proving that the models are capable of identifying a glass bottle based on accurate classification. The interpretability provided by visualization techniques ensures that these identified regions align with human intuition, further reinforcing the trustworthiness of the model's predictions. In addition, model prediction confidence for other classes is also shown to clarify similar resemblances. This comprehensive analysis of prediction confidence across various waste classes allows for a nuanced understanding of the model's strengths and potential areas of improvement. Using this approach, other images can be analyzed for evaluation of the model's performance and improvement according to a more diverse classification performance. The iterative nature of this process, coupled with visualization techniques, ensures a robust and adaptable waste classification system that can keep up with the evolving nature of waste materials.

4.7.2. Hardware development for a servo-controlled diverter for waste sorting

The hardware part was designed based on the principle of the model and its output for real-time applications. To enhance real-time capabilities, the hardware comprises a servo-controlled diverter (Fig. 23A) that can divert waste into two separate areas while passing through a conveyor belt. A simulation was conducted using the FlexSim simulation software version 2019 developed by Eamonn Lavery and Anthony Johnson for a clear understanding of how the proposed model can be used for industrial waste sorting (video link: https://www.youtube.com/watch?v=cd7Xb1hB6CM). The simulation process is shown in

Fig. 23, where each waste is automatically checked by a camera (Fig. 23B) upon passing through the conveyor. This simulation process provides a virtual testing ground for the hardware, ensuring its functionality and effectiveness in a controlled environment. Frames taken by the camera are processed using the models used for this research. The integration of real-world hardware with simulation results allows for a comprehensive evaluation of the model's performance under various conditions. Based on the predicted class, the servo diverter (Fig. 23A) diverts each waste to its correct path for sorting and recycling purposes. This automated sorting mechanism not only improves efficiency but also reduces the margin of error in waste classification. Using such a method is easy and customizable because it requires no complex or sophisticated technology and relies entirely on image processing-based algorithms. This simplicity in implementation makes the system accessible and cost-effective, paving the way for practical applications in waste management systems.

The proposed idea encompasses a larger system where multiple pairs of diverters and a camera mount can be installed at each dividing junction of the conveyor belt, allowing seamless waste sorting. In addition, the utilization of a wireless sensor network (WSN) instead of a conventional wiring system is far more convenient due to its cost, energy efficiency, and real-time monitoring. Such an approach is heavily customizable, which is beneficial, especially for the waste sorting industry where frequent modification is needed. The circuit design and hardware mechanism are simple and inexpensive but effective for this development. To ensure cost-effectiveness, the simplicity of the circuit design is complemented by its efficiency in waste sorting tasks.

According to Fig. 24, the camera module continuously sends the captured frames to the central computer, where using one of the developed models, the class of waste is identified. This seamless integration between the camera module and the central computer streamlines the data flow for efficient waste classification. Based on the identification, a binary value, either 0 or 1, is sent to the hardware circuit by serial communication, which determines the diverter's position for directing the waste path. The binary communication protocol provides a straightforward yet reliable method for real-time control of waste diverters. The data transfer is set to 9600 bits per second, and the buffer size of the Arduino is 128 bytes, which allows for storing commands in case of overflow of data. These parameters are carefully chosen to optimize data transmission and ensure robustness under varying conditions.

The prototype was constructed with a single 15-watt servo motor, an Arduino Mega 2560, a servo power supply, and a 34 cm long diverter. To ensure a compact and efficient assembly, several 3D-printed parts were used to hold the components in place, leveraging the design capabilities of Tinkercad software. The concept was to keep the design and circuit components as optimal as possible for easier installation and portability. This emphasis on simplicity aligns with the goal of user-friendly installation and convenient portability for practical applications. At the boot stage, the diverter is programmed to lock itself in the middle at 90°. This default position ensures a standardized starting point for the diverter, providing consistency in its initial configuration. As soon as the GUI starts classifying, the binary decision is sent to the Arduino for hardware processing according to initial tests performed using the first-stage model, which can classify two classes: biodegradable and non-biodegradable. If the detected class is biodegradable, Arduino receives command '1′, which immediately directs the servo to rotate to 125° from its initial state, executing a precise movement that strategically blocks the right passage depicted in Fig. 25 (A, B). In the opposite scenario, for the other class, the servo rotates to 55°, causing the lever to block the left passage, as shown in Fig. 25 (C, D). This dynamic response of the hardware ensures effective waste sorting based on real-time classification results. This test was performed with a PowerPoint video of random trash images that was used in the GUI app to illustrate the same scenario of a trash passing through the conveyor belt. The integration of real-world simulation scenarios with the GUI app provides a practical
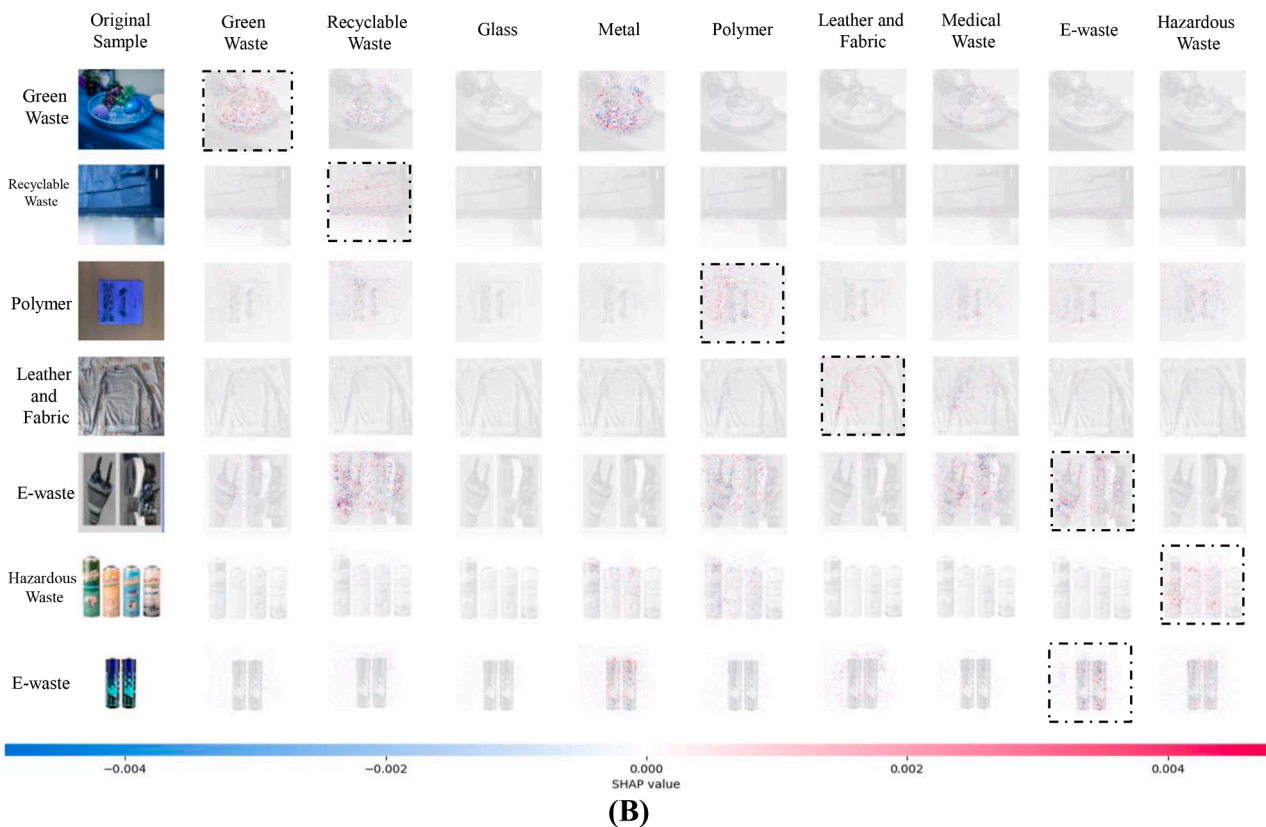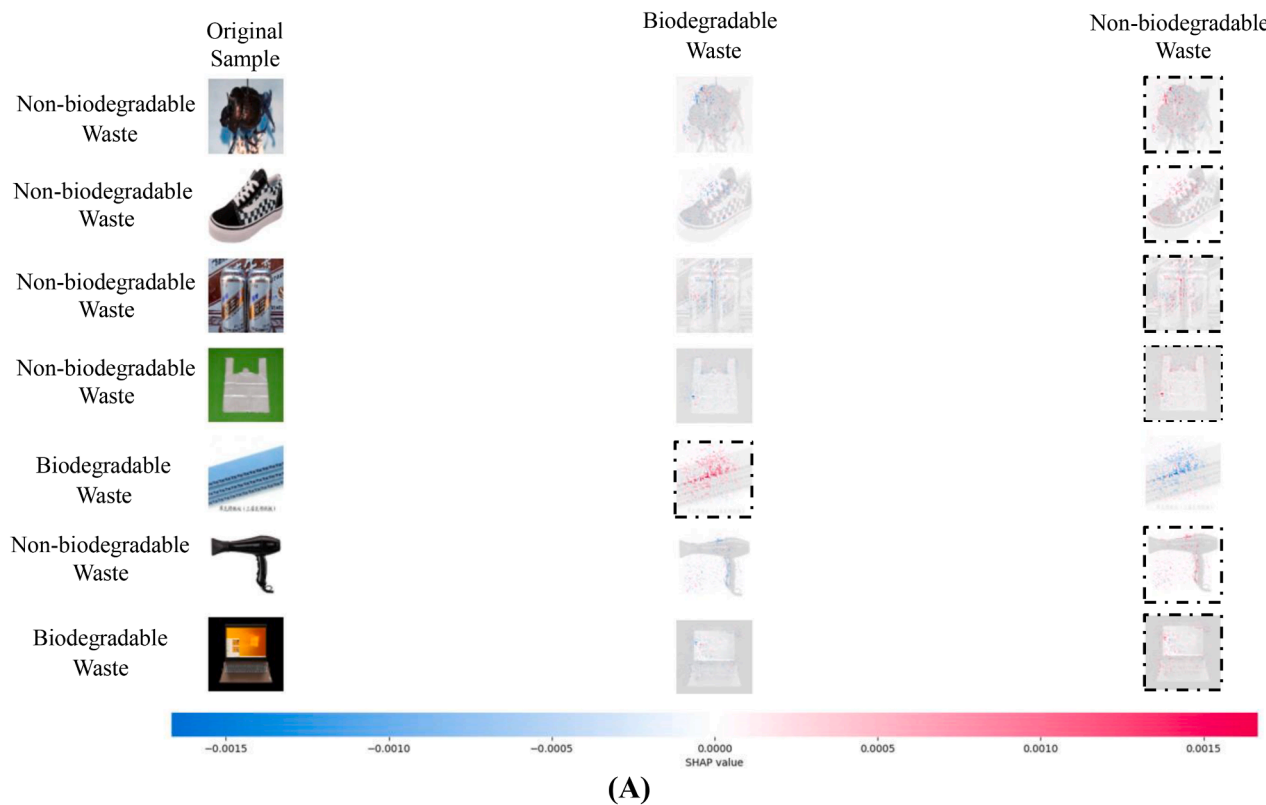
(A)

(B)

Fig. 19. SHAP explanation images for DP-CNN-En-ELM for (A) two-class classification and (B) nine-class classification.

demonstration of the responsiveness of hardware mechanisms to varying waste types. Using it, the hardware mechanism was switched on so that it would react to the trash changes in every frame determined via

the classification using the proposed model. Fig. 23 shows that detecting a biodegradable waste causes the diverter to move right, dynamically responding to the classification outcome and facilitating the correct path
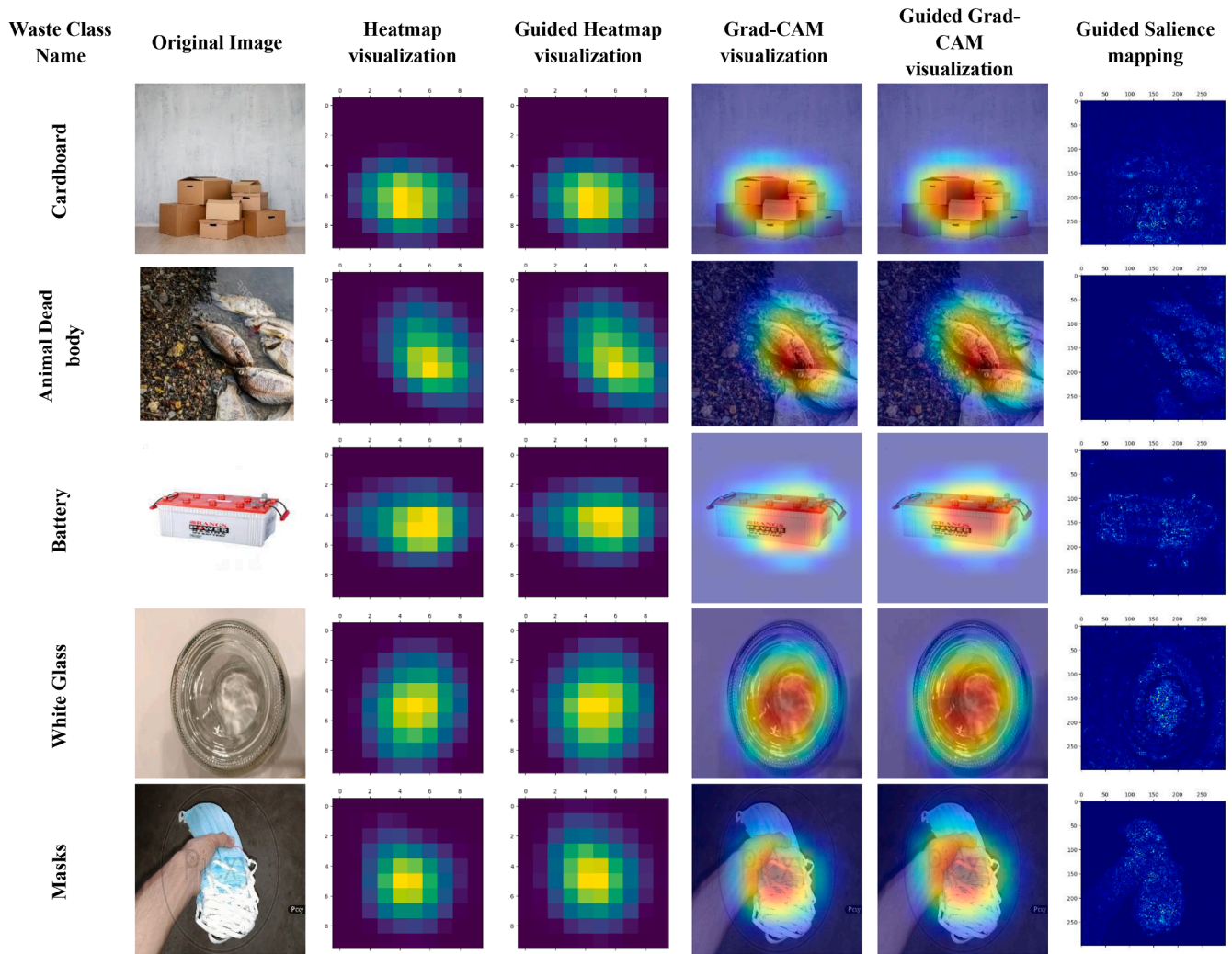
Fig. 20. Model Visualization by Heatmap, Guided Heatmap, Grad-CAM, Guided Grad-CAM, and Guided Salience Mapping.
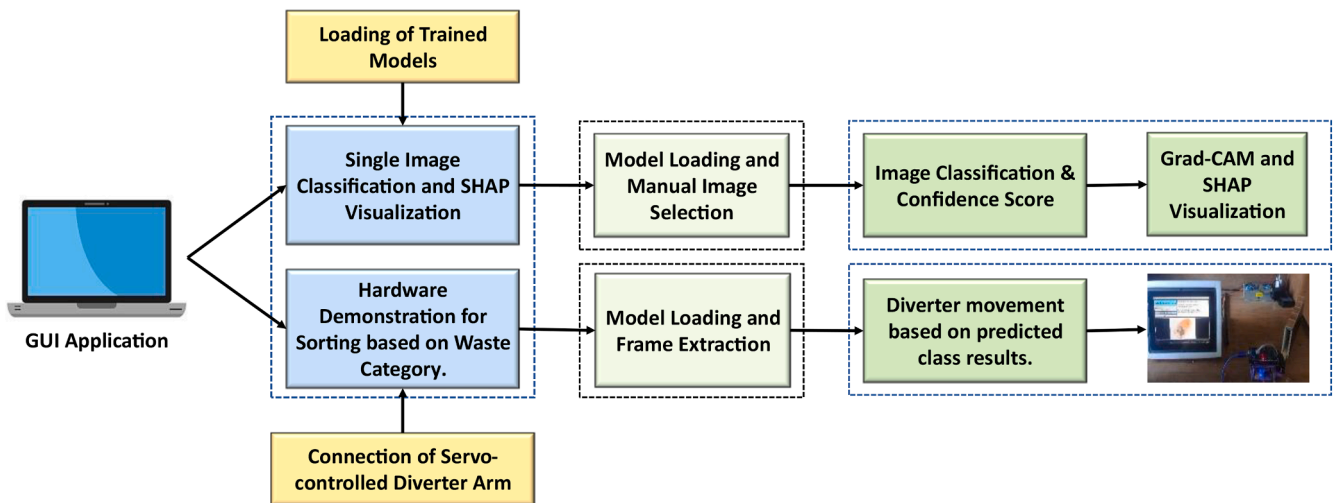

Fig. 21. Flowchart of the working steps in the GUI application.

for biodegradable waste, as depicted by the blue box of the simulation and the blue arrow. For non-biodegradable waste, the diverter shifts left, allowing the items to move right, as depicted by the green box indicated by the green arrow. This interactive behavior of the hardware aligns

with the efficient sorting of waste materials based on their identified class, providing a tangible representation of the system's functionality.

The proposed model was also tested with real-world data shown in Fig. 26 where multiple pieces of vegetable considered biodegradable
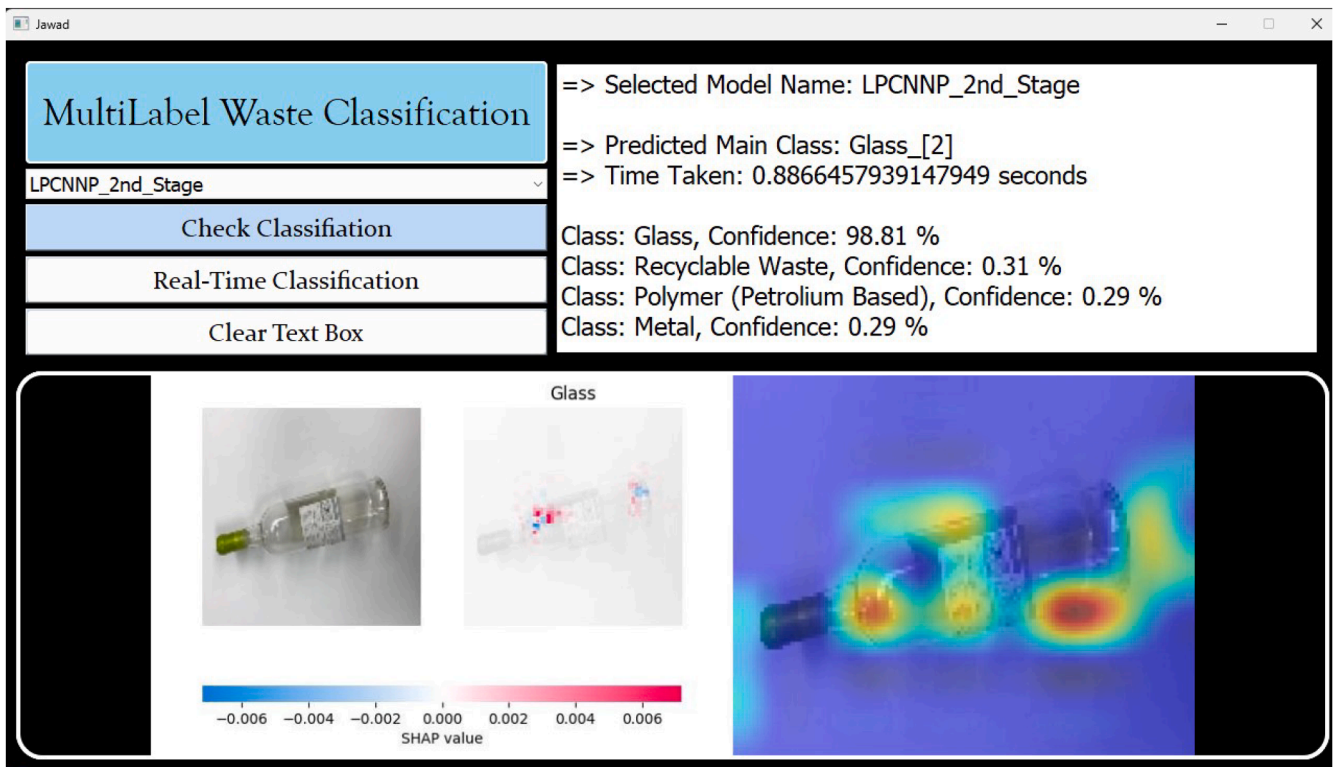
Fig. 22. Visualization of significant features using SHAP and Grad-CAM for model performance evaluation.
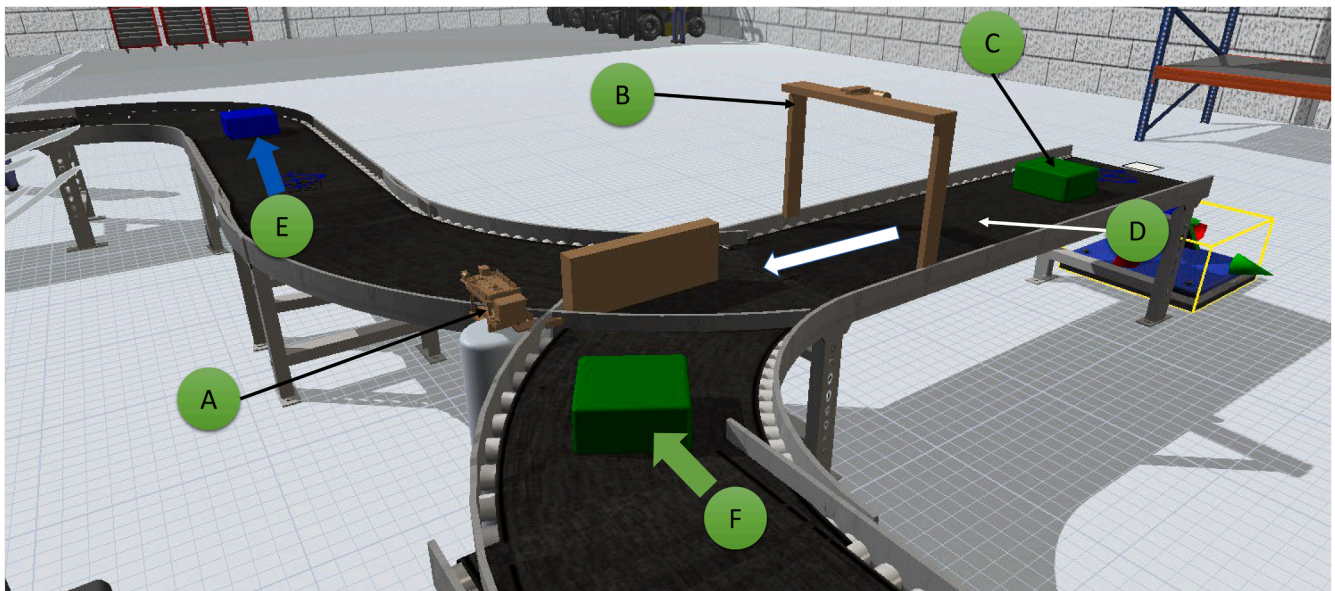

Fig. 23. Simulation diagram of waste sorting using a servo-based diverter and conveyor system approach. (A) Servo-based diverter. (B) Camera mount for waste classification. (C) Passing unidentified waste through the conveyor. (D) Conveyor platform (E) Identified biodegradable waste and (F) Identified non-biodegradable waste.

trash and a plastic container as non-biodegradable trash were used. A webcam was used to extract the image from which the GUI app collected the live image frames on which prediction was done. The model could correctly distinguish between both of the trash with a high confidence score. During the test, it was found that natural lighting had effects on the confidence score as they showed more accurate results for images under bright light.

4.8. Discussion

Although several studies have investigated waste classification, more work is needed to encompass a wide range of waste types. This discussion provides a comprehensive analysis of the experimental outcomes obtained by the proposed method, which successfully classifies 36 types of waste products. The present study consists of three primary steps: merging the dataset, extracting features, and performing classification
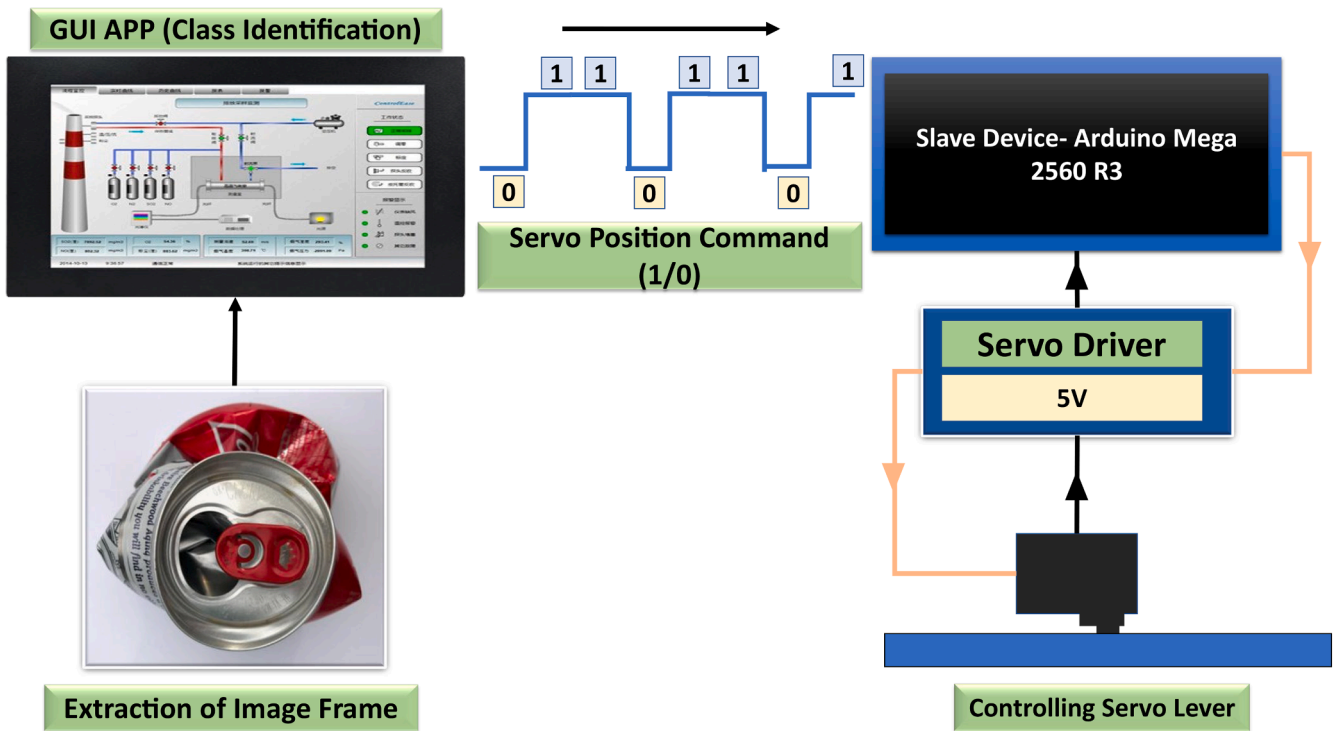
GUI APP (Class Identification)

Servo Position Command (1/0)

Extraction of Image Frame

Slave Device- Arduino Mega 2560 R3

Servo Driver

5V

Controlling Servo Lever

Fig. 24. Hardware circuit configuration with the GUI app.

(B) Diverter to Left (125 Degrees)

(C)

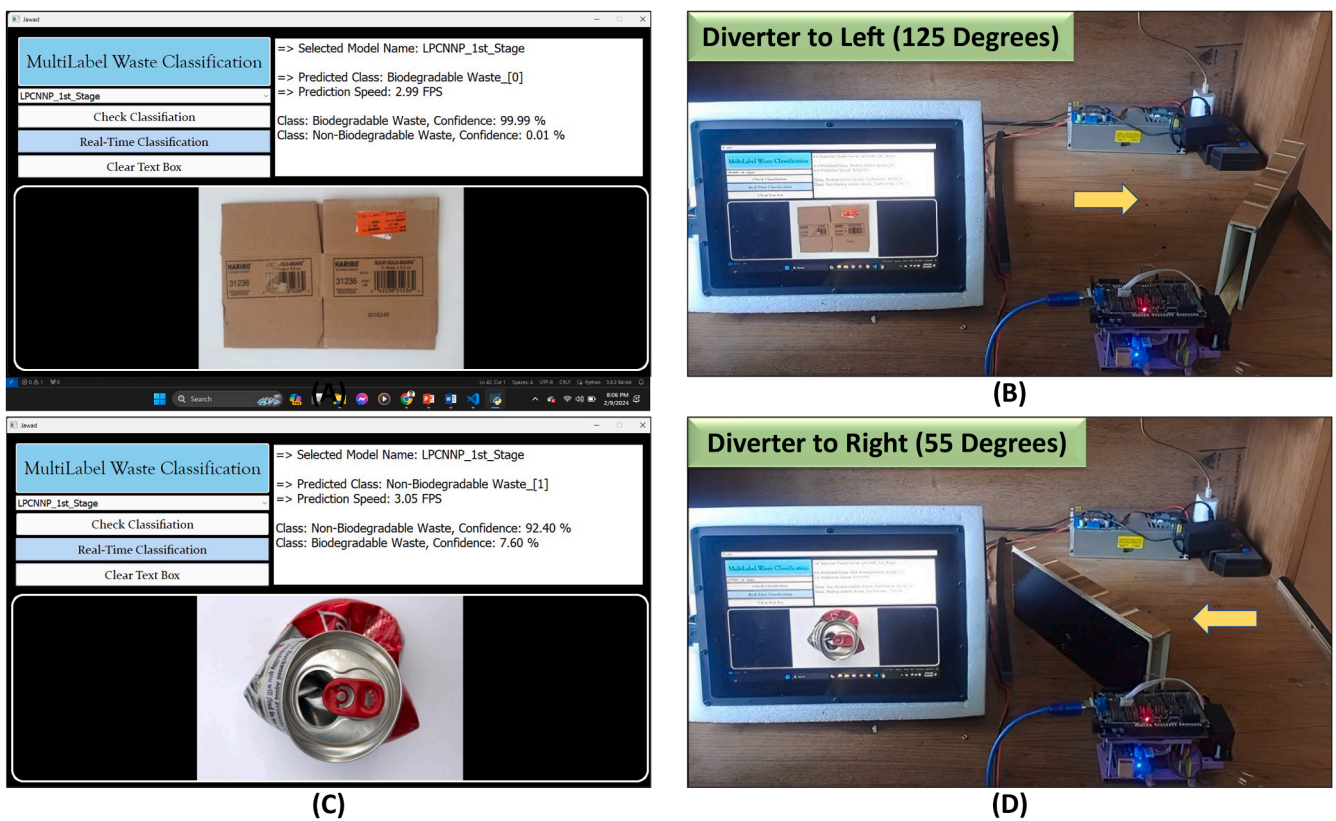**(D)** Diverter to Right (55 Degrees)

Fig. 25. Classification process test for real-time sorting and diverter movement for (A) (B) biodegradable waste and (C) (D) nonbiodegradable waste.

with interpretability. Merging the datasets results in larger and standardized input data, which improves the reliability of the proposed model in real-world scenarios.

Table 12 provides an overview of all state-of-the-art models in comparison with the proposed DP-CNN-En-ELM model. In summary, in [53], the optimized denseNet121 model peaked accuracy of 99.60 %. They utilized the TrashNet dataset, which comprises six classes and a total of 2527 pictures. Neelakandan et al. [15] attained nearly the same
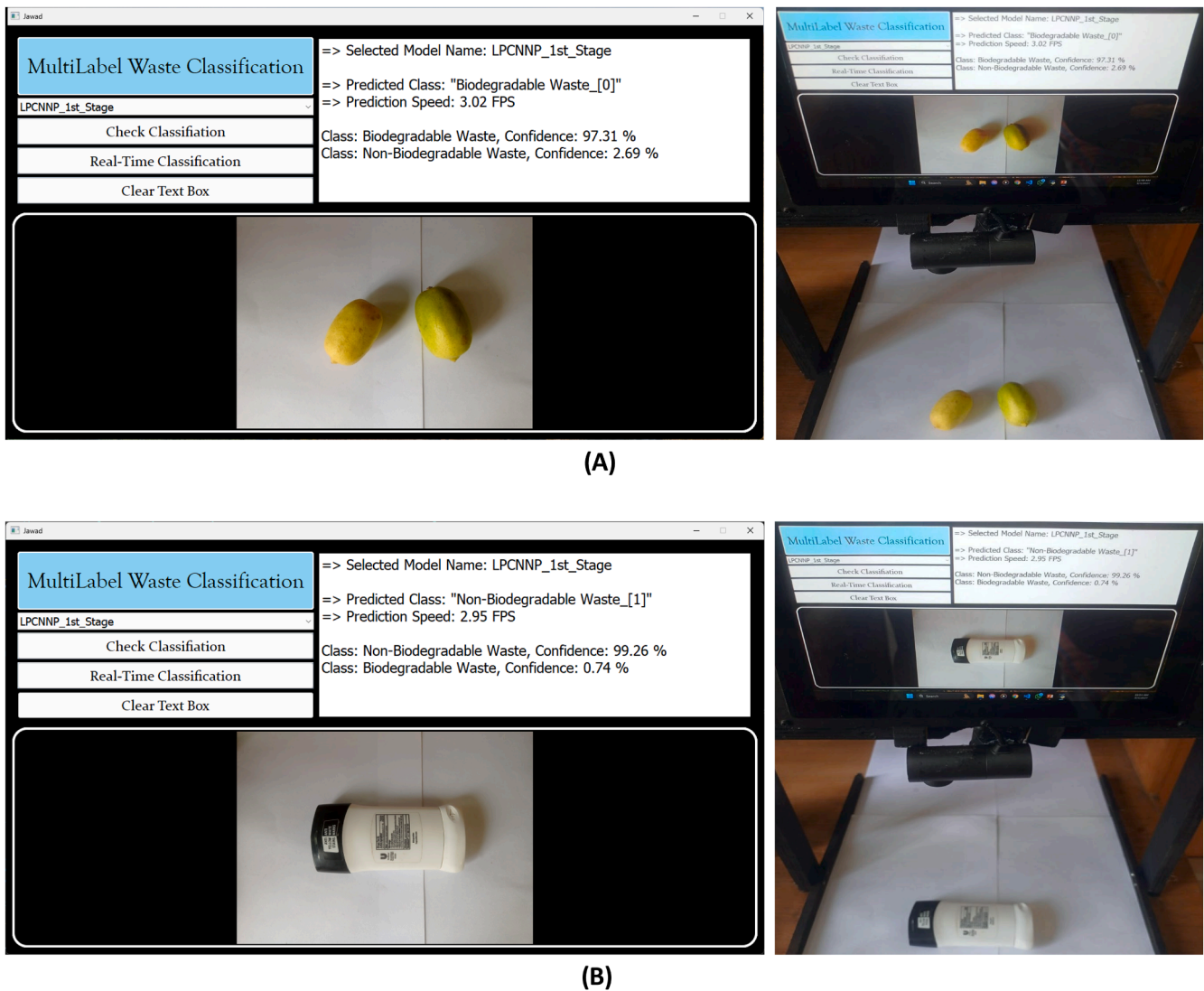
Fig. 26. Classification process test for real-time image for (A) biodegradable waste and (B) non-biodegradable waste.

accuracy of 99.26 % on the TR/TS dataset of 2467 images. Davis et al. [29] analyzed a CNN model on a customized dataset of 1758 images from 8 different classes. They attained a testing accuracy of 94 %. On the other hand, using an extensive dataset of 35,264 images, the proposed model achieves a comparable accuracy of 96.0 % for two classes in the first stage and 91.0 % for nine classes in the second stage. Furthermore, the DP-CNN model contains only 1.09 million parameters, whereas most of the existing models have enormous numbers of parameters (in [21] 58.5 million and in [26] 138 million). Although Zheng et al. [28] have shown that their Focus-RCNet-KD model can achieve a 92 % accuracy rate with just 0.53 million parameters, they also evaluate the model on the smaller TrashNet dataset.

A compact lightweight model featuring 1.5 million parameters was proposed by Yang et al. [30]. However, on larger datasets such as the Huawei garbage classification dataset, this model has shown a lower classification accuracy of 82.5 %. Another lightweight model based on EfficientNet was proposed by Feng et al. [24] for two-stage waste classification. In the second stage of classification, they used 1.23 million parameters to achieve a 94.54 % testing accuracy for 18 classes. However, the suggested DP-CNN-En-ELM model outperformed these models [23–25,27] based on model parameters and dataset size. Additionally, it should be noted that there is an absence of research focusing on three-stage classification. The proposed three-stage waste classification

model achieved an impressive accuracy of 85.25 % in the third stage for classifying waste materials into 36 distinct categories. Moreover, there is not much research illustrating real-time XAI, but the suggested model includes several XAI techniques, such as heatmap, guided heatmap, Grad-CAM, guided Grad-CAM, guided salience mapping, and SHAP methods. Additionally, incorporating hardware increases the acceptability of the model in the real world.

Despite the impressive performance of the proposed method, which employs a lightweight model with 1.09 million parameters, 9 layers, and a size of 12.7 megabytes, it still has certain limitations. As evident from the data presented in Table 12, the model's classification accuracy decreased as the number of classes increased. In the third stage, the model achieved the lowest accuracy of 85.25 %. This is because in the third stage, the dataset has 36 different classes. Expanding the number of classes hinders the learning process by complicating decision boundaries [54]. Having fewer classes in a specific dataset can offer adequate features for successful generalization. On the other hand, expanding the number of classes without correspondingly increasing the dataset size might result in a notable decrease in model efficacy. Additionally, the combined TriCascade WasteImage dataset encompasses various waste image categories, which is another cause of lower classification accuracy.

The proposed three-stage waste classification framework effectively

Table 12
Summary of the state-of-the-art models and the proposed model.

Ref.	Dataset	Number of Images	Number of Classes	Best Model	Testing Accuracy (%)	Best Model's Parameters (million)	Testing Time (seconds)	Real-time XAI
[13]	NWNU-TRASH	2528	5	Optimized DenseNet169	82.8	—	22.56	No
[15]	TR/TS data	2467	6	MDTLDC-IWM	99.26	—	—	No
[14]	VN-trash	5904	3	DNN-TC	98	—	—	No
[29]	Custom	1758	8	CNN	94	—-	—-·	No
[26]	TrashNet	2527	6	SWMACM-CA	90	138	—	No
[28]	TrashNet	2527	6	Focus-RCNet-KD	92	0.53	—	No
[21]	TrashNet	2527	6	RWNet	88.8	58.5	—	No
[53]	TrashNet	2527	6	Optimized DenseNet121	99.60	7.2	—	No
[27]	TrashNet and Huawei Garbage Classification Dataset	2527 and 18,079	6 and 4	WasNet	96.10 (For TrashNet), 82.5 (For Huawei Garbage Classification Dataset)	1.5	—	No
[23]	Huawei Cloud Garbage Classification Dataset	14,802	4	EfficientNetB2 with PMAM	93.38	7.8	6.756	No
[24]	Custom	7361	First Stage: 4 Classes Second Stage: 18 Classes	GECM-EfficientNet	94.54 (For 18 Classes)	1.23	—	No
[25]	Huawei Garbage Classification Challenge Cup Dataset	14,683	4	Improved MobileNetV2	90.7	3.4	—	No
Proposed Model	**TriCascade WasteImage**	**35,264**	**First Stage: 2 Classes, Second Stage: 9 Classes, Third Stage: 36 Classes**	**DP-CNN-En-ELM**	**96 (For 2 Classes), 91 % (For 9 Classes), 85.25 % (For 36 Classes)**	**1.09**	**0.00001**	**Yes**

addresses core issues prevalent in traditional waste management systems, such as limited resources, inadequate infrastructure, and inefficient sorting methods. By employing a lightweight DL architecture, specifically the DP-CNN (Depth-wise Separable Convolutional Neural Network) combined with the En-ELM (Ensemble Extreme Learning Machine) classifier, our method significantly reduces computational complexity, making it suitable for deployment in resource-constrained environments like Bangladesh, where infrastructure for efficient waste management is lacking. Establishing waste sorting centers that utilize this technology can help streamline the sorting process, reducing manual labor and improving sorting efficiency, which ultimately leads to lower environmental pollution by ensuring proper recycling and disposal. However, to fully address these issues, the method needs to be scaled up commercially and implemented in large-scale industrial applications, particularly in urban areas experiencing rapid population growth. Furthermore, public awareness and support for waste segregation at the source, such as separating biodegradable and non-biodegradable waste at home, are crucial for maximizing the efficiency of these sorting centers. Investments in developing countries, such as Bangladesh, are essential to building the necessary infrastructure, facilitating the widespread adoption of such technologies, and mitigating the environmental challenges posed by improper waste management. By combining technological innovation with public engagement and infrastructure investment, our framework offers a comprehensive solution that can substantially reduce environmental pollution while supporting global sustainability goals.

In future research, the authors will improve the model's efficiency by balancing the dataset. The researchers plan to gather and create a more evenly distributed dataset to enhance classification results. To enhance the applicability and contribution of our work to the community, future research will include the collection of real-world datasets directly from operational waste management facilities. This approach will provide a more accurate reflection of the challenges and variations encountered in practical applications. Collecting real datasets will also allow us to validate and fine-tune our model under actual operating conditions,

improving its robustness and reliability. It has been planned to collaborate with waste management companies and municipal waste sorting centers to gather diverse and representative samples from real-world scenarios. This will not only enhance the dataset's relevance but also contribute to developing more practical and effective waste classification solutions. Incorporating real-world data will strengthen the contributions of our research and provide valuable waste management solutions for the community.

Additionally, emphasis will be given to developing hardware using different communication protocols between the waste classification system and the conveyor-solving mechanism to ensure smooth integration. The use of the IoT system here can improve the synchronization mechanism to guarantee accurate coordination between the diverter and the conveyor, avoiding any delays or mistakes. In addition, the sorting mechanism could be modified to a rotating diverter mechanism that will allow for multiple class sorting of waste with fewer space and hardware requirements. Such integration with other automation systems will provide a comprehensive strategy to optimize space usage and ensure long-term cost efficiency.

5. Conclusions

This study proposed an innovative three-stage waste classification model combining the parallel lightweight depthwise separable convolutional neural network (DP-CNN) model with the ensemble extreme learning machine (En-ELM) classifier. The proposed DP-CNN architecture, which comprises nine layers and 1.09 million parameters, efficiently categorizes 36 types of wastes while reducing computational overhead. The model's testing time was only 0.00001. Combining the pseudoinverse ELM (PI-ELM) and L1-regularized ELM (L1-RELM) to create the En-ELM classifier significantly enhances the classification performance. This combined approach achieves impressive accuracy levels, with a 96.0 % accuracy rate for two-class classification during the initial phase, a 91.0 % accuracy for nine-class classification during the second stage, and an 85.25 % accuracy for thirty-six-class classification

during the final stage. The DP-CNN-En-ELM model has outstanding accuracy in distinguishing various forms of waste, with precision, recall, and F1-scores of 85.02 %, 85.25 %, and 84.54 %, respectively, for 36-class classification. The model also achieved an impressive AUC of 98.68 %. The design of the model is optimized by simultaneously running the first five CLs to improve feature extraction. The model also succeeds in terms of classification performance and computational requirements compared to advanced transfer learning models. The effectiveness of the model is demonstrated through the utilization of a substantial dataset comprising 35,264 images of diverse types. Furthermore, the size of the proposed model was only 12.7 MB, which increases the model's usability and affordability for waste management tools in daily life due to its low computational demands. Moreover, its integration with real-time XAI and hardware structures fosters transparency and reliability, ensuring informed decision-making and advancing the paradigm of sustainable waste management.

CRediT authorship contribution statement

Md. Nahiduzzaman: Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Md. Faysal Ahamed:** Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Mansura Naznine:** Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Md. Jawadul Karim:** Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Hafsa Binte Kibria:** Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Mohamed Arselene Ayari:** Writing – review & editing, Validation, Supervision, Formal analysis. **Amith Khandakar:** Writing – review & editing, Validation, Supervision, Formal analysis. **Azad Ashraf:** Writing – review & editing, Validation, Supervision, Formal analysis. **Mominul Ahsan:** Writing – review & editing, Validation, Supervision, Formal analysis. **Julfikar Haider:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

The customized dataset used for classification can be found in the following link. https://www.kaggle.com/datasets/salam035/waste-datasets.

Conceptual system operation video is available in the following link https://www.youtube.com/watch?v=cd7Xb1hB6CM

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.knosys.2025.113028.

References

[1] M. Mazzanti, R. Zoboli, Waste generation, waste disposal and policy effectiveness, Resour. Conserv. Recycl. 52 (2008) 1221–1234, https://doi.org/10.1016/j.resconrec.2008.07.003.

[2] R. Sha'Ato, S.Y. Aboho, F.O. Oketunde, I.S. Eneji, G. Unazi, S. Agwa, Survey of solid waste generation and composition in a rapidly growing urban area in Central Nigeria, Waste Manag. 27 (2007) 352–358, https://doi.org/10.1016/j.wasman.2006.02.008.

[3] T. Kumari, A.S. Raghubanshi, Waste management practices in the developing nations: challenges and opportunities. Waste Management and Resource Recycling in the Developing World, Elsevier, 2023, pp. 773–797, https://doi.org/10.1016/B978-0-323-90463-6.00017-8.

[4] D.T. Jerin, H.H. Sara, M.A. Radia, P.S. Hema, S. Hasan, S.A. Urme, C. Audia, Md. T. Hasan, Z. Quayyum, An overview of progress towards implementation of solid waste management policies in Dhaka, Bangladesh, Heliyon 8 (2022) e08918, https://doi.org/10.1016/j.heliyon.2022.e08918.

[5] S. Rhein, M. Schmid, Consumers' awareness of plastic packaging: more than just environmental concerns, Resour. Conserv. Recycl. 162 (2020) 105063, https://doi.org/10.1016/j.resconrec.2020.105063.

[6] R.E. Marshall, K. Farahbakhsh, Systems approaches to integrated solid waste management in developing countries, Waste Manag. 33 (2013) 988–1003, https://doi.org/10.1016/j.wasman.2012.12.023.

[7] A. Ye, B. Pang, Y. Jin, J. Cui, A YOLO-based Neural Network with VAE for intelligent garbage detection and classification, in: 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence, New York, NY, USA, ACM, 2020, pp. 1–7, https://doi.org/10.1145/3446132.3446400.

[8] A.B. Wahyutama, M. Hwang, YOLO-based object detection for separate collection of recyclables and capacity monitoring of trash bins, Electronics. (Basel) 11 (2022) 1323, https://doi.org/10.3390/electronics11091323.

[9] W.-L. Mao, W.-C. Chen, H.I.K. Fathurrahman, Y.-H. Lin, Deep learning networks for real-time regional domestic waste detection, J. Clean. Prod. 344 (2022) 131096, https://doi.org/10.1016/j.jclepro.2022.131096.

[10] N.A. Zailan, M.M. Azizan, K. Hasikin, A.S. Mohd Khairuddin, U. Khairuddin, An automated solid waste detection using the optimized YOLO model for riverine management, Front. Public Health 10 (2022), https://doi.org/10.3389/fpubh.2022.907280.

[11] Z. Lun, Y. Pan, S. Wang, Z. Abbas, M.S. Islam, S. Yin, Skip-YOLO: domestic garbage detection using deep learning method in complex multi-scenes, Int. J. Comput. Intell. Syst. 16 (2023) 139, https://doi.org/10.1007/s44196-023-00314-6.

[12] G. Huang, J. He, Z. Xu, G. Huang, A combination model based on transfer learning for waste classification, Concurr. Comput. 32 (2020), https://doi.org/10.1002/cpe.5751.

[13] Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, X. Liu, Waste image classification based on transfer learning and convolutional neural network, Waste Manag. 135 (2021) 150–157, https://doi.org/10.1016/j.wasman.2021.08.038.

[14] A.H. Vo, L. Hoang Son, M.T. Vo, T. Le, A novel framework for trash classification using deep transfer learning, IEEE Access 7 (2019) 178631–178639, https://doi.org/10.1109/ACCESS.2019.2959033.

[15] S. Neelakandan, M. Prakash, B.T. Geetha, A.K. Nanda, A.M. Metwally, M. Santhamoorthy, M.S. Gupta, Metaheuristics with Deep Transfer Learning Enabled Detection and classification model for industrial waste management, Chemosphere 308 (2022) 136046, https://doi.org/10.1016/j.chemosphere.2022.136046.

[16] H. Panwar, P.K. Gupta, M.K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, S. Sharma, I.H. Sarker, AquaVision: automating the detection of waste in water bodies using deep transfer learning, Case Stud. Chem. Environ. Eng. 2 (2020) 100026, https://doi.org/10.1016/j.cscee.2020.100026.

[17] X. Cai, F. Shuang, X. Sun, Y. Duan, G. Cheng, Towards lightweight neural networks for garbage object detection, Sensors 22 (2022) 7455, https://doi.org/10.3390/s22197455.

[18] Z. Chen, J. Yang, L. Chen, H. Jiao, Garbage classification system based on improved ShuffleNet v2, Resour. Conserv. Recycl. 178 (2022) 106090, https://doi.org/10.1016/j.resconrec.2021.106090.

[19] M. Fan, K. Zuo, J. Wang, J. Zhu, A lightweight multiscale convolutional neural network for garbage sorting, Syst. Soft Comput. 5 (2023) 200059, https://doi.org/10.1016/j.sasc.2023.200059.

[20] C. Shi, C. Tan, T. Wang, L. Wang, A waste classification method based on a multilayer hybrid convolution neural network, Appl. Sci. 11 (2021) 8572, https://doi.org/10.3390/app11188572.

[21] K. Lin, Y. Zhao, X. Gao, M. Zhang, C. Zhao, L. Peng, Q. Zhang, T. Zhou, Applying a deep residual network coupling with transfer learning for recyclable waste sorting, Environ. Sci. Pollut. Res. 29 (2022) 91081–91095, https://doi.org/10.1007/s11356-022-22167-w.

[22] Q. Zhang, X. Zhang, X. Mu, Z. Wang, R. Tian, X. Wang, X. Liu, Recyclable waste image recognition based on deep learning, Resour. Conserv. Recycl. 171 (2021) 105636, https://doi.org/10.1016/j.resconrec.2021.105636.

[23] H. Fan, Q. Dong, N. Guo, J. Xue, R. Zhang, H. Wang, M. Shi, Raspberry Pi-based design of intelligent household classified garbage bin, Internet of Things 24 (2023) 100987, https://doi.org/10.1016/j.iot.2023.100987.

[24] Z. Feng, J. Yang, L. Chen, Z. Chen, L. Li, An intelligent waste-sorting and recycling device based on improved efficientNet, Int. J. Environ. Res. Public Health 19 (2022) 15987, https://doi.org/10.3390/ijerph192315987.

[25] S. Jin, Z. Yang, G. Królczykg, X. Liu, P. Gardoni, Z. Li, Garbage detection and classification using a new deep learning-based machine vision system as a tool for

sustainable waste recycling, Waste Manag. 162 (2023) 123–130, https://doi.org/10.1016/j.wasman.2023.02.014.

[26] S.M. Cheema, A. Hannan, I.M. Pires, Smart waste management and classification systems using cutting edge approach, Sustainability 14 (2022) 10226, https://doi.org/10.3390/su141610226.

[27] Z. Yang, D. Li, WasNet: a Neural Network-Based Garbage Collection Management System, IEEE Access 8 (2020) 103984–103993, https://doi.org/10.1109/ACCESS.2020.2999678.

[28] D. Zheng, R. Wang, Y. Duan, P.C.-I. Pang, T. Tan, Focus-RCNet: a lightweight recyclable waste classification algorithm based on focus and knowledge distillation, Vis. Comput. Ind. Biomed. Art. 6 (2023) 19, https://doi.org/10.1186/s42492-023-00146-3.

[29] P. Davis, F. Aziz, M.T. Newaz, W. Sher, L. Simon, The classification of construction waste material using a deep convolutional neural network, Autom. Constr. 122 (2021) 103481, https://doi.org/10.1016/j.autcon.2020.103481.

[30] Md.W. Rahman, R. Islam, A. Hasan, N.I. Bithi, Md.M. Hasan, M.M. Rahman, Intelligent waste management system using deep learning with IoT, J. King Saud Univ. Comput. Inf. Sci. 34 (2022) 2072–2087, https://doi.org/10.1016/j.jksuci.2020.08.016.

[31] F.S. Alsubaei, F.N. Al-Wesabi, A.M. Hilal, Deep learning-based small object detection and classification model for garbage waste management in smart cities and IoT environment, Appl. Sci. 12 (2022) 2281, https://doi.org/10.3390/app12052281.

[32] G.K. Ijemaru, L.M. Ang, K.P. Seng, Transformation from IoT to IoV for waste management in smart cities, J. Netw. Comput. Appl. 204 (2022) 103393, https://doi.org/10.1016/j.jnca.2022.103393.

[33] A. Hussain, U. Draz, T. Ali, S. Tariq, M. Irfan, A. Glowacz, J.A. Antonino Daviu, S. Yasin, S. Rahman, Waste Management and prediction of air pollutants using iot and machine learning approach, Energies (Basel) 13 (2020) 3930, https://doi.org/10.3390/en13153930.

[34] N.V. Kumsetty, A. Bhat Nekkare, TrashBox: trash detection and classification using quantum transfer learning, in: 2022 31st Conference of Open Innovations Association (FRUCT), IEEE, 2022, pp. 125–130, https://doi.org/10.23919/FRUCT54823.2022.9770922.

[35] Hamdi Ali, Dead animals pollution, (2024). https://universe.roboflow.com/hamdi-ali/dead-animals-pollution (accessed February 9, 2024).

[36] 且听风吟, wastepictures, (2019). https://www.kaggle.com/datasets/wangziang/waste-pictures (accessed February 9, 2024).

[37] Mostafa Mohamed, Garbage classification (12 classes), (2021). https://www.kaggle.com/datasets/mostafaabla/garbage-classification?fbclid=IwAR3OjKZepc8ML8AmwJek_Iwb2JGL7VtCeRDhsIn1GALeLfi5CUptBVQVwmY (accessed February 9, 2024).

[38] Md. Nahiduzzaman, M.E.H. Chowdhury, A. Salam, E. Nahid, F. Ahmed, N. Al-Emadi, M.A. Ayari, A. Khandakar, J. Haider, Explainable deep learning model for automatic mulberry leaf disease classification, Front. Plant Sci. 14 (2023), https://doi.org/10.3389/fpls.2023.1175515.

[39] Md.R. Islam, Md. Nahiduzzaman, Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach, Expert. Syst. Appl. 195 (2022) 116554, https://doi.org/10.1016/j.eswa.2022.116554.

[40] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90, https://doi.org/10.1145/3065386.

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

[42] Md. Nahiduzzaman, Md.R. Islam, R. Hassan, ChestX-Ray6: prediction of multiple diseases including COVID-19 from chest X-ray images using convolutional neural network, Expert. Syst. Appl. 211 (2023) 118576, https://doi.org/10.1016/j.eswa.2022.118576.

[43] Md. Nahiduzzaman, Md.R. Islam, S.M.R. Islam, Md.O.F. Goni, Md.S. Anower, K.-S. Kwak, Hybrid CNN-SVD based prominent feature extraction and selection for grading diabetic retinopathy using extreme learning machine algorithm, IEEE Access. 9 (2021) 152261–152274, https://doi.org/10.1109/ACCESS.2021.3125791.

[44] Md. Nahiduzzaman, Md.O.F. Goni, Md.S. Anower, Md.R. Islam, M. Ahsan, J. Haider, S. Gurusamy, R. Hassan, Md.R. Islam, A novel method for multivariant Pneumonia classification based on hybrid CNN-PCA based feature extraction using extreme learning machine with CXR Images, IEEE Access. 9 (2021) 147512–147526, https://doi.org/10.1109/ACCESS.2021.3123782.

[45] H.B. Kibria, M. Nahiduzzaman, Md.O.F. Goni, M. Ahsan, J. Haider, An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI, Sensors 22 (2022) 7268, https://doi.org/10.3390/s22197268.

[46] Md. Nahiduzzaman, Md.O.F. Goni, R. Hassan, Md.R. Islam, M.K. Syfullah, S.M. Shahriar, Md.S. Anower, M. Ahsan, J. Haider, M. Kowalski, Parallel CNN-ELM: a multiclass classification of chest X-ray images to identify seventeen lung diseases including COVID-19, Expert. Syst. Appl. 229 (2023) 120528, https://doi.org/10.1016/j.eswa.2023.120528.

[47] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, 2017.

[48] M. Bhandari, T.B. Shahi, B. Siku, A. Neupane, Explanatory classification of CXR images into COVID-19, pneumonia and tuberculosis using deep learning and XAI, Comput. Biol. Med. 150 (2022) 106156, https://doi.org/10.1016/j.compbiomed.2022.106156.

[49] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, (2015).

[50] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 618–626, https://doi.org/10.1109/ICCV.2017.74.

[51] A. Chattopadhay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018, pp. 839–847, https://doi.org/10.1109/WACV.2018.00097.

[52] J.T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, Striving for simplicity: the all convolutional Net, (2014).

[53] W.-L. Mao, W.-C. Chen, C.-T. Wang, Y.-H. Lin, Recycling waste classification using optimized convolutional neural network, Resour. Conserv. Recycl. 164 (2021) 105132, https://doi.org/10.1016/j.resconrec.2020.105132.

[54] Z. Li, K. Kamnitsas, B. Glocker, Analyzing overfitting under class imbalance in neural networks for image segmentation, IEEE Trans. Med. ImAging 40 (2021) 1065–1077, https://doi.org/10.1109/TMI.2020.3046692.