Lightweight Deep Learning Algorithm for Sorting Medical Waste Embedded System

Y. S. Gan[1], Yuan-Zhi Liu[2], Bo-Cheng Tseng[2], Gen-Bing Liong[3], Sze-Teng Liong[2]*

[1] School of Architecture, Feng Chia University, Taiwan
[2] Department of Electronic Engineering, Feng Chia University, Taiwan
[3] Faculty of Computer Science & Information Technology, University of Malaya, Malaysia
ysgan@fcu.edu.tw, ximen89211@gmail.com, cs4150252@gmail.com, genbing67@gmail.com, stliong@fcu.edu.tw

Abstract

The recent COVID-19 pandemic has led to an increased number of hospitalized patients, and in some countries, the overwhelming problem has even caused the collapse of the healthcare system. Thus, the associated waste generated and inadequate waste disposal may have deleterious impacts on public health and the surrounding environment. This paper aims to design and implement an automatic medical waste identification and sorting mechanism that can distinguish the items disposed of via a series of image processing procedures. Specifically, the system can classify, sort, and calculate the disposed items based on their visual appearance. To establish this, both software and hardware systems have been devised. Particularly, the vision system for object detection and classification is constructed based on several popular pre-trained convolutional neural networks. Upon verifying the type of waste (i.e., general infection, dangerous infection, and general garbage), an automatic sorting mechanism will be triggered to dispose of the object in the corresponding garbage bin. The experimentation has been validated on a total of self-collected 2025 images from 3 categories, and the best accuracy attained is 99.34% when adopting GoogLeNet as the backbone architecture.

Keywords: Deep learning, Medical waste, Classification, Sorting, Transfer learning

1 Introduction

In recent years, the topic of waste classification has garnered extensive discussion. However, with the advancement of technology and the growth of civilization, the volume of waste generated has increased exponentially. According to statistics from the Ministry of Health and Welfare and the Environmental Protection Department in 2017, the average output of waste produced by medical institutions in Taiwan is about 120,000 metric tons per year [1]. However, the outbreak of novel coronavirus (COVID-19) infections has placed tremendous pressure on existing waste management systems, leading to an increase in waste accumulation in nearly all healthcare units and services. It should be noted that improper disposal of medical waste and poor management of healthcare waste, including waste classification, waste minimization, containerization, color coding, and labeling, potentially expose serious hazards of secondary disease transmission [2]. Additionally, it has been highlighted by [3] that more than 85% of medical waste can be recycled to reduce waste output.

Currently, artificial intelligence has matured into a well-established technology, and its applications have been rapidly expanding across various fields, with expectations for continued growth. However, there are still numerous scenarios that heavily rely on human efforts, such as cleaning and sorting tasks. Particularly, a significant amount of manpower is allocated to manage infectious and general commercial waste, resulting not only in increased personnel costs but also in the potential for personnel negligence. Therefore, the objective of this paper is to address these challenges by incorporating image recognition technology into the waste classification and sorting system. In essence, the designed mechanism presents a user-friendly system where garbage can be placed on the platform for automatic sorting. This approach significantly reduces the workload of cleaners and enhances overall quality of life.

This study focuses on establishing a real-time automated system for classifying medical waste using appearance features extracted through a deep learning technique known as the convolutional neural network (CNN). Notably, there is currently no existing comprehensive dataset exclusively containing medical waste images. To address this, we initially curate a dataset comprising three distinct classes: general infection, dangerous infection, and general garbage. In order to evaluate our approach, we leverage the transfer learning strategy, employing popular CNN architectures for feature extraction and waste classification. In addition to the CNN approach, the capabilities of state-of-the-art deep learning models are explored herein, such as transformer-based models, which include layers like patch embedding and attention mechanisms. This exploration is a significant aspect of our research, allowing us to investigate the potential advantages and limitations of such models in the context of medical waste classification. To demonstrate the practical applicability of our system, we design and implement a hardware platform capable of sorting the identified waste into appropriate bins. This platform integrates various components such as a low-cost webcam, ultrasonic sensor,

*Corresponding Author: Sze-Teng Liong; E-mail: stliong@fcu.edu.tw
DOI: https://doi.org/10.70003/160792642024092505004

micro-controller board, servo motor, and slider.

Furthermore, the practical applications of this study encompass automating medical waste sorting in healthcare facilities and hospitals, streamlining waste management centers' processes, enhancing public health and safety through proper disposal, optimizing resource allocation, and acting as an educational tool for waste classification awareness. Thus, promoting positively impact healthcare, waste management, environmental preservation, and public health on a broader scale.

The primary objectives accomplished in this research are summarized as follows:

1. Creation of a medical waste dataset encompassing 2025 images distributed across three categories.
2. Development of an automated medical waste recognition system utilizing state-of-the-art pre-trained CNNs.
3. Implementation of a hardware platform to validate the real-time waste sorting process.
4. Rigorous experimental validation of the recognition system's effectiveness and the mechanism's robustness, supported by both quantitative and qualitative results.

The rest of the paper is structured as follows: Section 2 reviews publicly available datasets and discusses recent research progress in recognition models and hardware mechanisms. Subsequently, Section 3 offers an overview of the proposed framework and elaborates on each step in detail. In Section 4, experimental results are reported and analyzed. Finally, Section 5 outlines the conclusion of this work and highlights potential future directions.

2 Related Works

2.1 Automatic Trash Classification

So far, numerous studies have addressed automatic trash classification. One particularly notable work in this domain is presented by [4], who introduced the TrashNet dataset. This dataset encompasses six waste classes: cardboard (403 samples), glass (501 samples), metal (410 samples), paper (594 samples), plastic (482 samples), and trash (137 samples), totaling 2,527 images. All images were captured against a clean background, utilizing objects placed in front of a white poster board. The image capture process employed various mobile devices, including Apple iPhone 7 Plus, Apple iPhone 5S, and Apple iPhone SE, with a standardized resolution of 512 x 384 pixels. Utilizing a scale-invariant feature transform (SIFT) [14] feature descriptor and a support vector machine (SVM) [15] classifier, a classification accuracy of 63% was achieved [4]. This relatively unsatisfactory result could be attributed to an imbalanced distribution within the waste classes. Notably, the trash class attained a precision of only 0.2 due to its limited representation (occupying approximately 5% of the dataset), often leading to misclassification with the paper class, which is the majority class.

Since then, several previous research works [5-8] have attempted to enhance classification accuracy by introducing deep learning approaches based on artificial neural networks for automatic feature extraction. Specifically, their methodologies employ a transfer learning strategy, leveraging feature knowledge gained from a large-scale dataset, such as ImageNet [16], which contains over a million images spanning 1000 object categories. Parameters, encompassing weights and biases, in pre-trained models are fine-tuned by exposing them to the TrashNet images. This process not only accelerates learning but also reduces the necessary input data, substantially elevating classification performance. A summary of outcomes is presented in Table 1, which juxtaposes the models employed in various studies and their evaluations on the TrashNet dataset. Notably, DenseNet-121 [17] and ResNet-50 [18] emerge as common architectures within this research domain.

On the other hand, Proença and Simões [10] curated a collection of trash images named Trash Annotations in Context (TACO). To provide a succinct overview, this dataset comprises 1500 images, encompassing a total of 4784 annotations. The annotated labels span a wide array of 60 categories, with notable examples of major classes being plastic bags, cigarettes, bottles, and cans. All images are meticulously annotated through pixel-wise segmentation, wherein the label with the largest area spans up to 2048 pixels, while the minimum area of a label is less than 16 pixels. To establish a benchmark result using the TACO dataset, Proença and Simões employed the Mask R-CNN [19] model, leveraging the ResNet-50 architecture as its backbone.

Table 1. Comparison of related works that evaluated on different datasets

No.	Dataset	# Images	# Labels	Model	Classification accuracy (%)	Segmentation (mAP)
1	TrashNet [4]	2527	6	DenseNet-121+GA [5]	99.6	-
2				ResNet-50 [6]	95.35	-
3				DenseNet-121 [7]	95	-
4				DenseNet-121 [8]	95	-
5				SIFT+SVM [4]	63	-
6	VN-trash [9]	5904	3	DNN-TC [9]	98	-
7	TACO [10]	1500	60	Mask R-CNN [10]	-	0.1590
8				DeepLabv3-ML [11]	-	0.9607
9	AquaTrash [12]	369	4	Faster R-CNN [12]	-	0.8148
10	MJU-Waste [11]	2475	N/A	DeepLabv3-ML [11]	-	0.9714
11	ScrapNet [13]	8135	6	ScrapNet [13]	-	0.8234

In continuation, [12] expands upon this work by creating a novel dataset named AquaTrash, primarily derived from the TACO dataset. Succinctly, this new dataset comprises a total of 369 images, primarily focusing on waste found in aquatic environments. AquaTrash includes annotations for four categories: glass, metal, paper, and plastic. An object detection approach, specifically Faster R-CNN [20], is employed to accurately localize waste within bounding boxes. The outcome is an achieved segmentation average precision of 0.81, establishing a strong baseline for waste detection within this context.

More recently, [11] compiled a set of real-life images depicting waste items found within a university campus, known as the MJU dataset. Precisely, the image capture setting is situated in a laboratory environment, with each waste item held by a person. Notably different from conventional waste datasets captured using 2D cameras, this dataset utilizes the Microsoft Kinect RGBD device. This approach allows each image to contain both intensity and depth information at various levels of spatial granularity. In total, the dataset comprises 2475 images. To establish a baseline segmentation performance, the images are divided into training, validation, and test sets in a ratio of 6:1:3. Additionally, the deep learning architecture DeepLabv3 [21], incorporating the proposed multi-level model, achieves an impressive segmentation mean intersection over union result of 93.79%.

Subsequently, Masand et al. [13] introduced a composite dataset to address the disparity among various waste categories featured in distinct publicly available datasets. To achieve this, they curated the ScrapNet dataset by selecting images from multiple sources. These sources encompassed datasets such as TrashNet, Openrecycle [22], TACO, and waste classification data [23]. The amalgamated dataset comprises a total of 8135 images, representing six categories: plastic (2014 samples), metal (950 samples), glass (1055 samples), paper (1043 samples), compost (913 samples), and trash (1602 samples). Employing the EfficientDet [24] network architecture for the object detection task yielded an mAP score of 0.8234.

2.2 Medical Waste Classification

Regarding medical waste identification, Bian et al. [25] employed a Single Shot Multibox Detector (SSD) [26] with the MobileNet backbone architecture for waste segmentation. Their tested garbage comprised materials like hemostatic forceps, gloves, infusion bags, and syringes. Although they achieved a high classification accuracy of 98.5% with a dataset of 2825 images, the data distribution across the training, validation, and test sets remains unclear. Another medical waste detection study was conducted by Chen et al. [27], where a camera above the waste container recorded and classified the total trash amount. However, these wastes need to be properly separated for appropriate disposal. The classification of medical waste includes categories like infectious, hazardous, radioactive, and general.

In a similar vein, Mythili and Anbarasi [28] assembled a trash image dataset encompassing 200 images across five categories of biomedical waste: infectious waste, chemical waste, sharp waste, pharmaceutical waste, and pathological waste. They applied a 1:1 train/test split to validate their proposed encoder-decoder network architecture's capability. A two-stage waste recognition process was undertaken, beginning with a segmentation step followed by classification. They evaluated their approach using accuracy and Root Mean Square Error (RMSE) metrics for classification and segmentation performance assessment, respectively. This resulted in an 84% accuracy for the 5-class classification task. However, the dataset's scale is relatively small and susceptible to overfitting, especially when utilizing a neural network approach to extract discriminant features. Furthermore, ambiguities arose in their segmentation experimentation due to a lack of clarity in protocol settings and configurations within the study.

In their pursuit of fostering an environmentally conscious and sustainable ecosystem, [29] significantly enrich the ongoing discourse on effective waste management practices. Their research presents a focused exploration into the complex domain of medical waste management systems and their alignment with circular economy principles. By investigating the integration of these principles, the authors underscore the potential for a paradigm shift in conventional practices within the medical waste sector. Noteworthy is their emphasis on the deployment of optimization techniques and multi-criteria decision-making (MCDM) methods, revealing a thorough investigation of these strategies' applicability to optimize circular economy-driven medical waste management systems. This meticulous examination attests to the authors' profound insights.

Simultaneously, the same research team underscores their dedication to refining waste management practices through their earlier work [30]. Here, they introduce an inventive and robust approach to address the intricate challenges inherent in urban waste management systems, particularly in the presence of uncertainty. The study candidly acknowledges the unpredictable nature of waste management scenarios and introduces a pioneering perspective by proposing a comprehensive framework that integrates location-allocation-inventory optimization. This approach underscores the researchers' commitment to a holistic solution that adeptly considers the intricate interplay of diverse factors within waste management systems.

2.3 Waste Classification with Hardware Implementation

In addition to software development with high accuracy, numerous studies focus on implementing robots for waste detection, sorting, and picking. For instance, Bai et al. [31] designed a mobile robot that utilizes ResNet-34 to distinguish between garbage and non-garbage categories. The garbage class includes objects like bottles, cans, cartons, plastic bags, and waste papers. This mobile robot integrates a localization module (Inertial Measurement Unit (IMU) and odometer), controller, manipulator, garbage container, and multiple sensors. Consequently, it autonomously cleans designated areas of grass by removing predefined garbage. Conversely, a floating robot was developed by Kong et al. [32] to retrieve plastic waste from water surfaces. They employed a YOLOv3 [33] architecture to identify the target object. The system was designed to navigate toward the object and perform grasping operations.

2.4 Summary

Notably, most of the existing works apply CNN architectures to image classification due to the recent advancements in deep learning. In order to compete in the ImageNet recognition challenge [34], several architectures were proposed, including AlexNet [35], VGG16 [36], GoogLeNet [37], ResNet-18 [18], and ResNet-50 [18]. In 2012, AlexNet was the pioneering deep network that made significant progress in the challenge. Subsequently, VGG16 was developed, utilizing multiple small-size kernels to create a deeper network with lower costs in learning object discriminant features. Concurrently, GoogLeNet, inspired by the human visual system, introduced the inception module, which increased both depth and width while maintaining computational efficiency in network learning. However, deep architectures posed challenges such as vanishing gradients. To address this, ResNet was introduced, incorporating residual blocks and skip connections that add the identity of the previous layer to preserve gradients. Due to its effectiveness, the authors further designed deeper architectures, namely ResNet-18 and ResNet-50.

3 Proposed Method

There are two primary components to achieving this automatic medical waste sorting mechanism: software configuration and hardware setup. Designing a robust medical waste classification system involves five key steps: data collection, data augmentation, CNN model training and testing, and graphical user interface (GUI) visualization. On the other hand, the hardware system encompasses four steps: image acquisition, object classification, movement of the garbage bin, and waste disposal. An overall flowchart of the proposed method is portrayed in Figure 1, and the specifics of each step will be detailed in the following subsections.

3.1 Software Configuration

3.1.1 Data Collection

The image data consists of three categories: general garbage (e.g., rag, medicine cup, medical package, marker pen, straw, chopsticks, and rag), general infection (e.g., cotton, cotton swab, gauze, gloves, masks, and tongue depressor), and dangerous infection (e.g., syringe, hypodermic needle, and blade). The details of each category and the corresponding statistics are provided in Table 2. In total, 2025 images have been collected, with a data distribution of 810 for general garbage, 810 for general infection, and 405 for dangerous infection. All the images were captured using the KTnet camera model, with a resolution of 1280 x 720 pixels. Sample data examples are illustrated in Figure 2. Notably, the images were taken from different angles, with varying object quantities and distances, to enhance the variability of the collected data and ensure the recognition system's viability in a wide range of operating conditions.

Table 2. The statistics of the experimental data

No.	Category	Object	Total
1.	General infection	Cotton	135
		Cotton sawb	135
		Gauze	135
		Gloves	135
		Masks	135
		Tongue depressor	135
2.	General garbage	Medical cup	135
		Medical package	135
		Marker pen	135
		Straw	135
		Chopsticks	135
		Rag	135
3.	Dangerous infection	Syringe	135
		Hypodermic needle	135
		Blade	135
			2025

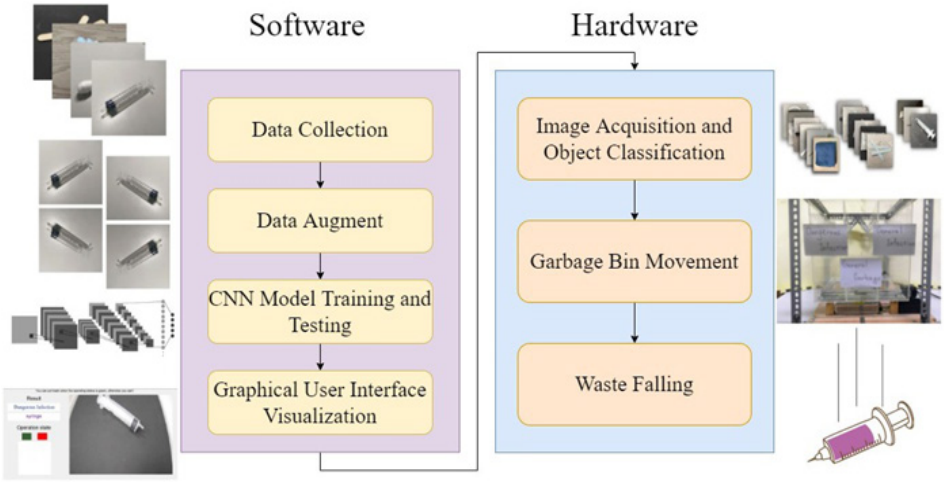


Figure 1. The proposed framework for the medical waste classification and sorting mechanism that composes hardware and software parts

Figure 2. Samples of the self-collected medical waste dataset

3.1.2 Data Augmentation

The data augmentation process is performed to address the issue of data insufficiency. Additionally, it mitigates the overfitting phenomenon in the CNN model training process during the subsequent stage, thereby improving its generalization ability. Specifically, artificial image data is created by adding slight modifications to existing data. Random transformations such as reflections, translations, scaling, and rotations are implemented. Specifically, the augmentation options include both horizontal and vertical reflections, horizontal and vertical translations ranging from [-3, 3] pixels, a scaling range of [0.5, 2], and a rotation angle range of [-15°, 15°]. The detailed range of the augmentation parameters is provided in Table 3.

Table 3. Parameters for data augmentation

Augmenter	Parameter
Horizontal reflection	True
Vertical reflection	True
Scaling	[0.5, 2]
Rotation	[-15°, 15°]
Horizontal translation	[-3, 3]
Vertical translation	[-3, 3]

3.1.3 CNN Model Training and Testing

The CNN models used in this study rely on a transfer learning strategy to offset the limitations of a small dataset, thereby facilitating faster convergence during training. The pre-trained CNN models under consideration include AlexNet [35], GoogLeNet [37], ResNet-18 [18], ResNet-50 [18], and VGG-16 [36]. These models are fine-tuned by adding a new fully-connected layer with an output size of 3 for the 3-class classification task and 15 for the 15-class classification task. The updated models are trained using the Adam optimizer and the following loss function:

$$L(\bar{y}, y) = -\Sigma^{K} y_{(k)} log\bar{y}^{(k)} \quad (1)$$

where $y^{(k)}$ is 0 or 1 indicating correct classification on class k. All models have been previously trained on the ImageNet dataset [16], which contains approximately 1.2 million images across 1000 categories. A summary of the properties of these pre-trained networks and their respective training times is provided in Table 4. It is observed that the training time for all networks is approximately 1.5 hours. These networks vary in architectural design, with layer depths ranging from 8 to 50 and model sizes ranging from 27MB to 515MB.

Table 4. The properties of the pre-trained network and the corresponding training time required for the medical waste dataset

Network	Depth	Model size	Training time
AlexNet	8	227 MB	4840*s*
GoogLeNet	22	27 MB	5353*s*
VGG-16	16	515 MB	5877*s*
ResNet-18	18	44 MB	4964*s*
ResNet-50	50	96 MB	5620*s*

To ensure a fair evaluation of the experiment, a 3-fold cross-validation strategy is employed to randomly divide the data into three equally sized subsamples. Two experiments are conducted for medical waste recognition: a 3-class classification task and a 15-class classification task. The former is performed on a composite dataset, while the latter considers each object item as an individual class. The train/test data distribution for each category in both tasks is detailed in Table 5. The model training is configured with a learning rate of 0.0001 and an epoch size ranging from 10 to 200, in intervals of 10. Early stopping at the 200th epoch is implemented because the training accuracy and loss values plateau, indicating that further training would not result in improvements. The training progress for the GoogLeNet

architecture is illustrated in Figure 3. Importantly, the training options and experimental protocols for both the 3-class and 15-class classification tasks are identical, as summarized in Table 6.

Table 5. The train/ test split for the 3-class and 15-class classification tasks

3-class classification				15-class classification			
Object	Train	Test	Total	Object	Train	Test	Total
General infection	567	243	810	Cotton	95	40	135
				Cotton swab	95	40	135
				Gauze	95	40	135
				Gloves	95	40	135
				Masks	95	40	135
				Tongue depressor	95	40	135
General garbage	567	243	810	Medicine cup	95	40	135
				Medical package	95	40	135
				Marker pen	95	40	135
				Straw	95	40	135
				Chopsticks	95	40	135
				Rag	95	40	135
Dangerous infection	284	121	405	Syringe	95	40	135
				Hypodermic needle	95	40	135
				Blade	95	40	135
Total	1418	607	2025	Total	1425	600	2025

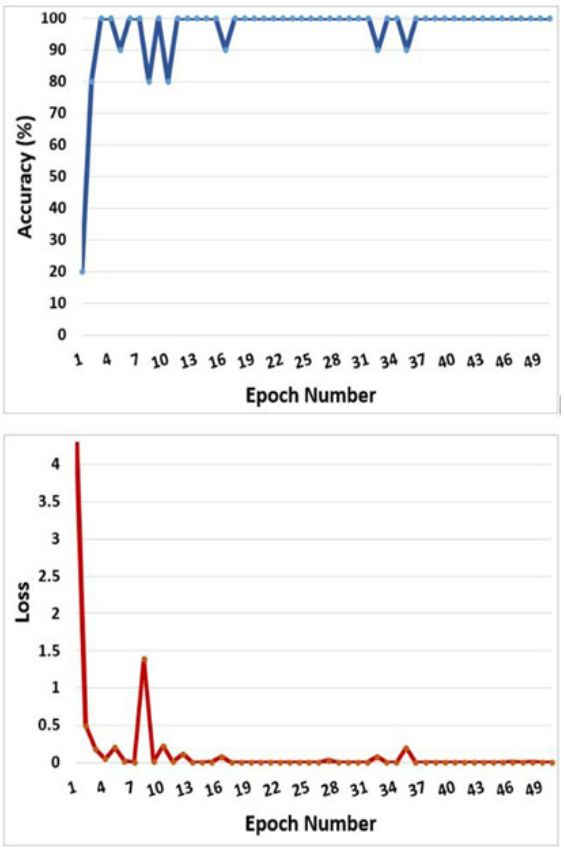


Figure 3. The training accuracy and loss for GoogLeNet architecture

Table 6. Training options for both the 3-class and 15-class classification tasks

Option	Parameter
Solver	adam
Initial learn rate	0.0001
Mini-batch size	16
Epoch	[10, 200]
Shuffle	once
Gradient decay factor	0.9
Gradient threshold method	l2 norm

3.1.4 Graphical User Interface Visualization

Upon completing the training of the CNN models, a GUI is designed for intuitive user interaction. A screenshot of the GUI is shown in Figure 4. The large section of the window on the right displays the real-time scene captured via the webcam, while the left section provides the name of the detected object. This GUI displays results for both the 3-class and 15-class classification tasks. The operational state is indicated: a green light signifies the system is ready for operation, and a red light implies the system is busy, either executing trash bin movement or awaiting waste disposal.

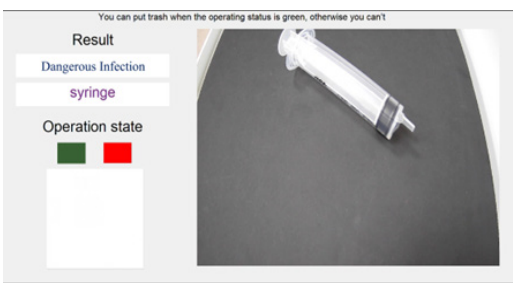


Figure 4. The GUI for intuitive interaction

3.2 Hardware Setup

The automatic medical waste classification mechanism consists of six hardware components: an ultrasonic sensor, an Arduino, a sliding rail, a webcam, a servo motor, and a garbage bin. The overall dimensions of the system are 36 cm in width, height, and length, as depicted in Figure 5. The functions and detailed operations of the system are elaborated in the following subsections, which cover the processes of image acquisition, object classification, garbage bin movement, and waste falling.

Figure 5. The hardware setup of the automatic medical waste classification mechanism

3.2.1 Image Acquisition and Object Classification

Users can place trash on a platform measuring 30 cm x 30 cm when the GUI displays a green operational state. The platform is constructed from black paperboard and has a payload capacity of 1 kg. A webcam is mounted atop the system at a -45° angle, perpendicular to the horizontal plane, to capture images of the object. An ultrasonic sensor is employed to automatically detect the presence of a randomly placed object. The sensor's detection range is set to 30 cm, sufficient to cover the entire platform. Although an infrared sensor was considered during the design phase, it was found to perform less effectively at longer distances and to be susceptible to environmental factors such as dust, sunlight, and smoke. Therefore, we opted for ultrasonic sensors for object detection.

Upon object detection, the webcam activates to capture an image, which is then sent to the computer for further analysis. At this point, the operational state displayed on the GUI will turn red. The webcam operates at a spatial resolution of 1280 x 720 pixels. The captured image is resized to 224 x 224 pixels before being fed into the CNN model for object recognition. The object detection process via the ultrasonic sensor and the object classification using the GoogLeNet architecture each take about 1 second, ensuring real-time processing.

3.2.2 Garbage Bin Movement

After identifying the object type, the sliding rail moves either to the left or right, depending on the category. Specifically, the garbage bin has dimensions of 36 x 36 x 36cm^3, and it is partitioned into three equal compartments to accommodate the three categories of medical waste. Each compartment has a volume capacity of approximately 16 liters. The sliding rail, designed with dimensions of 36 x 36 x 36 cm^3, facilitates bin movement. To enhance the stability of this movement, two vertical supporting beams are positioned beneath the bin, flanking the sliding rail.

The garbage bin starts in a centered position. Maximum movement limits of 12 cm to both the left and right are established, corresponding to our three classifications: general garbage, general infection, and dangerous infection. The bin's movement speed is set at 30 mm/s, allowing it to move from the center to the left or right in a maximum time of 4 seconds.

3.2.3 Waste Falling

After the garbage bin moves to the desired position, the servo motor is activated to allow the medical waste on the platform to drop. The servo motor is programmed to rotate 90° clockwise, facilitating the waste's descent due to gravitational force. The entire waste-disposal process takes less than 1 second. Subsequently, the hardware system resets, and the operational state displayed on the GUI returns to a green light, signaling readiness for the next round of medical waste identification.

4 Experiment Setting and Results

This section discusses the metrics and corresponding experimental settings used to evaluate the proposed method. Additionally, in-depth analyses, both numerical and

statistical, are presented to provide insightful perspectives by revealing patterns or trends in the extracted feature data. Furthermore, the results obtained are interpreted visually through graphical representations, which offer valuable information and facilitate the investigation of subsequent explanatory mathematical models. Finally, the limitations of this study are outlined to suggest areas for further improvement in the proposed methodology.

4.1 Experiment Settings

The experiments are conducted on an Intel(R)Core(TM) i7-10750H 2.60GHz CPU with an NVIDIAGTX 1660Ti GPU. Image classification is implemented using MATLAB version R2020b. To assess the effectiveness and robustness of the proposed medical waste classification system, the performance metrics used are accuracy and F1-score, which can be mathematically expressed as follows:

$$Accuracy := \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

and Precision × Recall

$$F_1 - score := 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{3}$$

where

$$Recall : \frac{TP}{TP + FN}, \tag{4}$$

and

$$Precision := \frac{TP}{TP + FP}. \tag{5}$$

where:

- TP is the true positive, indicating the model correctly distinguishes the class of the object.
- TN is the true negative, indicating the model correctly predicts that the object does not belong to the class.
- FN is the false negative, indicating the model does not predict the class of the object correctly.
- FP is the false positive, indicating the model is incorrectly predicts the negative class as positive.

On the other hand, the Pearson correlation coefficient is used here to measure the linear relationships between two feature vectors, aiming to observe the associations within interclass or intraclass groups. Specifically, the Pearson coefficient is approximated using a least-squares fit. A coefficient value of 1 represents a perfect positive relationship, -1 signifies a perfect negative relationship, and 0 indicates the absence of any relationship between the two vectors. Mathematically, the Pearson correlation coefficient can be formulated as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{6}$$

where n refers to the sample size, x_i and y_i indicate the values of the first and second vectors in the sample, respectively. $\overline{x}$ and $\overline{y}$ represent the mean of the values of the first and second vectors in the sample, respectively.

4.2 Result and Discussion

This section primarily discusses the experimental results for the software component, focusing on the performance analysis of CNN model training and evaluation. Specifically, two separate classification tasks are conducted to validate the suitability of the collected database: a 3-class and a 15-class classification. Table 7 and Table 8 present the accuracies and F1-scores obtained using the transfer learning strategy with five well-known CNN architectures (i.e., AlexNet, GoogLeNet, VGG-16, ResNet-18, and ResNet-50).

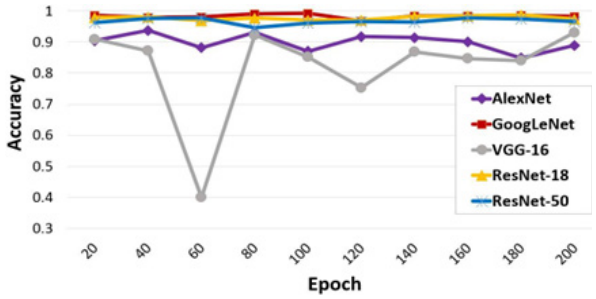
Overall, the average results achieved in both tasks are promising. The best accuracies are 99.34% and 98.67% when using GoogLeNet as the backbone architecture for the 3-class and 15-class classification tasks, respectively. Other CNN architectures, such as AlexNet, ResNet-18, and ResNet-50, also achieve an average accuracy of 90% or higher, indicating the robustness of the applied networks. Additionally, the small standard deviation (i.e., less than 3%) indicates high consistency in the classification tasks. To offer a clearer visualization of the performance achieved, Figure 6 summarizes the trend of classification results across different epoch sizes for various backbone architectures. It can be observed that the accuracy for all the networks, except VGG-16, plateaus after around 20 epochs and rarely increases thereafter. This demonstrates the effectiveness of the transfer learning strategy through fine-tuning the model's parameters, thereby significantly reducing training time. However, the results from AlexNet diverge noticeably from those of other networks like GoogLeNet, ResNet-18, and ResNet-50, showing an average accuracy of around 91% in both classification tasks. This discrepancy in accuracy is partly due to the model's shallow structure; its lower number of parameters may not be well-suited for this complex feature-learning task. In contrast, the performance of GoogLeNet, ResNet-18, and ResNet-50 is very similar across all epoch sizes, achieving the highest accuracies of over 98%. Therefore, GoogLeNet is chosen as the primary model for in-depth analysis in subsequent discussions. On the other hand, VGG-16 experiences a sharp drop in accuracy during the training process, especially at epoch=60 as shown in Figure 6(a). This fluctuation could be attributed to its large number of trainable parameters (i.e., 515MB, as referred to in Table 4), making it less ideal for model generalization on this medical waste dataset.

Table 7. The performance results in terms of accuracy and F1-score for the 3-class classification task (i.e., general infection, general garbage, and dangerous infection) when adopting different pre-trained networks

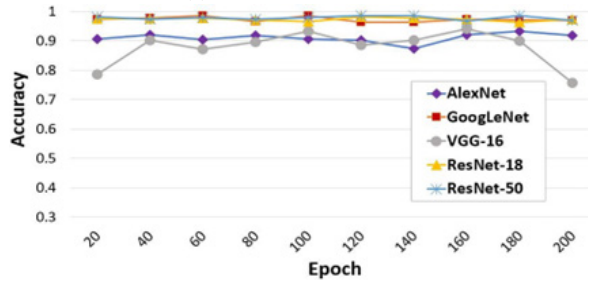
Epoch	AlexNet		GoogLeNet		VGG-16		ResNet-18		ResNet-50	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
10	0.9176	0.9166	0.9852	0.9830	0.7661	0.7364	0.9687	0.9698	0.9786	0.9774
20	0.9044	0.9066	0.9852	0.9821	0.9094	0.912	0.9786	0.9769	0.9621	0.9578
30	0.9357	0.9369	0.9753	0.9760	0.9209	0.927	0.9736	0.9709	0.9572	0.9551
40	0.9374	0.9387	0.9786	0.9781	0.8715	0.8612	0.9802	0.9782	0.9753	0.9720
50	0.9357	0.9337	0.9852	0.9834	**0.9506**	**0.9451**	0.9835	0.9822	**0.9852**	**0.9823**
60	0.8814	0.8837	0.9802	0.9787	0.4003	0.5718	0.9687	0.9672	0.9769	0.9753
70	**0.9374**	**0.9395**	0.9802	0.9807	0.9094	0.8997	0.9736	0.9715	0.9605	0.9589
80	0.9308	0.9341	0.9901	0.9897	0.9226	0.9243	0.9769	0.9741	0.9456	0.9413
90	0.9308	0.9300	0.9802	0.9787	0.9012	0.9002	0.9786	0.9787	0.9703	0.9681
100	0.8699	0.8675	**0.9934**	**0.9925**	0.8517	0.8241	0.9703	0.9658	0.9605	0.9557
110	0.9110	0.9122	0.9835	0.9802	0.4003	0.5718	0.9835	0.9822	0.9621	0.9617
120	0.9176	0.9119	0.9671	0.9611	0.7529	0.7529	0.9703	0.9686	0.9654	0.9627
130	0.9044	0.9076	0.9852	0.9842	0.9193	0.912	0.9802	0.9767	0.9769	0.9746
140	0.9143	0.9166	0.9835	0.9814	0.8682	0.8424	0.9819	0.9809	0.9638	0.9607
150	0.8962	0.8860	0.9802	0.9763	0.9226	0.9175	0.9720	0.9673	0.9671	0.9665
160	0.9012	0.8994	0.9835	0.9835	0.8468	0.8527	0.9835	0.9821	0.9769	0.9760
170	0.9308	0.9312	0.9703	0.9673	0.8847	0.8864	0.9835	0.9829	0.9341	0.9249
180	0.8484	0.8597	0.9852	0.9842	0.8402	0.8082	**0.9868**	**0.9856**	0.9736	0.9714
190	0.9176	0.9154	0.9802	0.9808	0.8962	0.9007	0.9786	0.9801	0.9769	0.9767
200	0.8896	0.8840	0.9819	0.9795	0.9308	0.9333	0.9736	0.9726	0.9654	0.9624
Max	0.9374	0.9395	0.9934	0.9925	0.9506	0.9451	0.9868	0.9856	0.9852	0.9823
Avg	0.9106	0.9106	0.9817	0.9801	0.8333	0.8440	0.9773	0.9757	0.9667	0.9641
σ	0.0244	0.0240	0.0060	0.0068	0.1567	0.1094	0.0056	0.0061	0.0121	0.0135

Table 8. The performance results in terms of accuracy and F1-score for the 15-class classification task when adopting different pre-trainednetworks

Epoch	AlexNet		GoogLeNet		VGG-16		ResNet-18		ResNet-50	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
10	0.8917	0.8933	0.9850	0.9850	0.8667	0.8684	0.9750	0.9751	0.9783	0.9783
20	0.9067	0.9057	0.9733	0.9734	0.7850	0.7869	0.9767	0.9767	0.9817	0.9817
30	0.9350	0.9358	0.9750	0.9750	0.8967	0.8991	0.9767	0.9764	0.9783	0.9783
40	0.9200	0.9198	0.9783	0.9785	0.9033	0.9025	0.9767	0.9767	0.9733	0.9732
50	0.8850	0.8859	0.9800	0.9800	0.9133	0.9124	0.9733	0.9732	0.9783	0.9784
60	0.9050	0.9060	0.9850	0.9851	0.8717	0.8695	0.9783	0.9782	0.9800	0.9799
70	**0.9383**	**0.9389**	0.9600	0.9604	0.8867	0.8860	0.9783	0.9784	0.9783	0.9784
80	0.9183	0.9183	0.9650	0.9648	0.8967	0.8945	0.9717	0.9720	0.9733	0.9730
90	0.9233	0.9236	0.9783	0.9782	0.8317	0.8312	0.9817	0.9815	0.9700	0.9698
100	0.9067	0.9089	**0.9867**	**0.9866**	0.9333	0.9334	0.9650	0.9646	0.9800	0.9798
110	0.8983	0.9033	0.9767	0.9767	0.8517	0.8581	0.9750	0.9748	0.9583	0.9587
120	0.9033	0.9034	0.9633	0.9633	0.8867	0.8847	0.9817	0.9814	**0.9867**	**0.9866**
130	0.9100	0.9094	0.9800	0.9800	0.7983	0.7922	0.9767	0.9767	0.9767	0.9765
140	0.8733	0.8721	0.9633	0.9635	0.9017	0.9006	0.9783	0.9784	0.9850	0.9850
150	0.9100	0.9103	0.9617	0.9620	0.9183	0.9172	**0.9817**	**0.9817**	0.9783	0.9779
160	0.9200	0.9203	0.9733	0.9736	**0.9417**	**0.9423**	0.9733	0.9733	0.9683	0.9683
170	0.8683	0.8672	0.9733	0.9733	0.8400	0.8303	0.9700	0.9703	0.9833	0.9833
180	0.9333	0.9337	0.9717	0.9714	0.9000	0.8994	0.9633	0.9630	0.9850	0.9850
190	0.9317	0.9311	0.9767	0.9768	0.8800	0.8764	0.9767	0.9765	0.9767	0.9766
200	0.9183	0.9191	0.9700	0.9699	0.7567	0.7584	0.9733	0.9735	0.9700	0.9699
Max	0.9383	0.9389	0.9867	0.9866	0.9417	0.9423	0.9817	0.9817	0.9867	0.9866
Avg	0.9098	0.9103	0.9738	0.9739	0.8730	0.8722	0.9752	0.9751	0.9770	0.9769
σ	0.0195	0.0196	0.0080	0.0079	0.0493	0.0495	0.0049	0.0050	0.0067	0.0067



(a) 3-class classification



(b) 15-class classification

Figure 6. Accuracy results for the 3-class and 15-class classification tasks across different epoch sizes when adopting five different famous pre-trained CNN architectures

In addition to the accuracy and F1-score previously reported, the confusion matrices for both classification tasks offer further, more granular insights, as detailed in Table 9 and Table 10. A total of 607 and 600 testing images were used in the 3-class and 15-class classification tasks, respectively. Specifically, in the 3-class classification task, the CNN model perfectly classifies the "general infection" category, while making minor errors in classifying images from the "general garbage" and "dangerous infection" categories—misclassifying 3 and 1 images, respectively. Conversely, Table 10 reveals that among the 15 categories, only a few test images from 5 classes (i.e., "cotton swab", "gloves", "marker pen", chopsticks", and "blades") exhibit minor miscategorization errors. Despite the high similarity in appearance between different objects, which leads to potential misclassifications such as mistaking "chopsticks" for "straw" or "tongue depressor", as shown in Figure 7 the resulting classification rates for all categories exceed 90%. This implies the robustness of the applied CNN model.

Succinctly, this study not only introduces a novel framework for automatic medical waste recognition, but also incorporates a real-time hardware platform for classification and sorting. The model's strength lies in its robustness analysis, evidenced by comprehensive assessments such as intraclass and interclass correlations, ensuring its capability to handle complex relationships between waste categories. Achieving a remarkable 99% classification accuracy on a self-collected database of around 2000 images underscores its proficiency. However, the hardware dependency could limit its adaptability in certain contexts. Compared to similar methods, this model uniquely combines algorithmic advancement and hardware implementation, offering both theoretical soundness and practical applicability, setting it apart from purely algorithm-based approaches.

Table 9. The confusion matrix of recognition result (%) for the 3-class classification task when employing GoogLeNet in the transfer learningstrategy

		Predicted class		
		General infection	General garbage	Dangerous infection
Target class	General infection	**100**	0	0
	General garbage	0	**98.77**	1.23
	Dangerous infection	0	0.41	**99.59**

Table 10. The confusion matrix of recognition result (%) for the 15-class classification task when employing GoogLeNet in the transfer learning strategy

		Predicted Class														
		Cotton	Cotton swab	Gauze	Gloves	Masks	Tongue depressor	Medicine cup	Medical package	Marker pen	Straw	Chopsticks	Rag	Syringe	Hypodermic needle	Blade
Target Class	Cotton	**100**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Cotton swab	0	**97.5**	0	0	0	0	2.5	0	0	0	0	0	0	0	0
	Gauze	0	0	**100**	0	0	0	0	0	0	0	0	0	0	0	0
	Gloves	0	2.5	0	**97.5**	0	0	0	0	0	0	0	0	0	0	0
	Masks	0	0	0	0	**100**	0	0	0	0	0	0	0	0	0	0
	Tongue depressor	0	0	0	0	0	**100**	0	0	0	0	0	0	0	0	0
	Medicine cup	0	0	0	0	0	0	**100**	0	0	0	0	0	0	0	0
	Medical package	0	0	0	0	0	0	0	**100**	0	0	0	0	0	0	0
	Marker pen	0	0	0	0	0	0	0	0	**97.5**	0	2.5	0	0	0	0
	Straw	0	0	0	0	0	0	0	0	0	**100**	0	0	0	0	0
	Chopsticks	0	0	0	0	0	5	0	0	0	2.5	**92.5**	0	0	0	0
	Rag	0	0	0	0	0	0	0	0	0	0	0	**100**	0	0	0
	Syringe	0	0	0	0	0	0	0	0	0	0	0	0	**100**	0	0
	Hypodermic needle	0	0	0	0	0	0	0	0	0	0	0	0	0	**100**	0
	Blade	0	0	0	0	0	0	0	0	0	0	0	2.5	0	0	**97.5**

Compared to similar methods in the literature, this proposed framework stands out due to its multifaceted approach. Unlike some existing approaches that solely focus on algorithmic improvements, this model integrates both algorithmic advancements and hardware implementation. This dual focus ensures that the model's performance is not only theoretically sound but also practically applicable in real-world scenarios. Moreover, the meticulous analysis of intraclass and interclass correlations sets this work apart. It goes beyond traditional evaluation metrics to provide a more nuanced understanding of the model's classification abilities. This level of insight into the model's performance is often lacking in comparable methods. This high classification accuracy (99%), along with the comprehensive evaluation, provides strong evidence of the model's effectiveness and contributes to its superiority over existing methods. In summary, the proposed model combines innovation, practical implementation, rigorous analysis, and superior performance. This multifaceted approach positions it as a noteworthy advancement in the field of medical waste recognition.

4.3 Statistical Analysis and Data Visualization

A statistical analysis is conducted to examine the feature similarity extracted by the CNN model for images within both intraclass and interclass groups, as presented in Table 11 and Table 12, respectively. This metric can also serve as an alternative way to assess and numerically describe the relationships between feature representations within the same and different groups. Specifically, a high correlation coefficient close to 1 suggests high feature similarity between two feature representations, while a value close to 0 indicates that the extracted features are dissimilar. Table 11 includes five randomly selected test images from each category (i.e., masks, hypodermic needles, cotton swabs, and chopsticks) for statistical evaluation. Overall, features extracted from the masks and hypodermic needle categories exhibit very high correlation coefficients (i.e., >80%), while the cotton swab and chopsticks categories show relatively lower values (i.e., 60%). This discrepancy could explain the correct and incorrect classifications observed in the 15-class classification task (as referred to in Table 10). In terms of interclass relationships, the average correlation coefficient value in Table 11 is approximately 15%, suggesting that features from different groups have minimal correlation. This implies that the CNN architecture is capable of extracting distinct and qualitatively important features from images across different categories.

The graphical illustration in Figure 8 displays the t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis used to visualize the high-dimensional data learned by the pre-trained GoogLeNet. Specifically, a random image from each of the 3 or 15 categories is selected, and the activations from the last pooling layer are extracted. Each video, represented by a 1024-D feature set, is then transformed into a 3D map based on the probability distribution between pairs of instances. The distinct clusters, with 3 and 15 well-separated and spaced groups, facilitate the correct identification of image categories in both classification tasks.

Last but not least, Table 13 presents the Grad-CAM activations that highlight the features learned by the CNN model. The Grad-CAM model outputs a visual explanation heatmap with colors ranging from red to blue. In this heatmap, the red region indicates high activation, while blue denotes low activation. It is observed that the proposed method can achieve promising recognition results, especially for images with noisy backgrounds (e.g., medical packages and chopsticks samples) and for tilted or deformed objects (e.g., medicine cups and masks samples). This demonstrates that the localization maps of meaningful regions within the images are clearly highlighted, attesting to the strong generalization capabilities of the adopted model. Figure 9 showcases examples of correctly classified test images along with their corresponding confidence levels, most of which exceed 95%.

Table 11. Correlation coefficient measurement for the images from intraclass groups

Masks

	1	2	3	4	5
1	1	0.85	0.78	0.85	0.80
2	0.85	1	0.82	0.85	0.80
3	0.78	0.82	1	0.78	0.76
4	0.85	0.85	0.78	1	0.93
5	0.80	0.80	0.76	0.93	1

Cotton swab

	1	2	3	4	5
1	1	0.82	0.61	0.57	0.46
2	0.82	1	0.53	0.53	0.47
3	0.61	0.53	1	0.73	0.70
4	0.57	0.53	0.73	1	0.72
5	0.46	0.47	0.70	0.72	1

Hypodermic needle

	1	2	3	4	5
1	1	0.80	0.93	0.90	0.86
2	0.80	1	0.77	0.87	0.87
3	0.93	0.77	1	0.84	0.85
4	0.90	0.87	0.84	1	0.95
5	0.86	0.87	0.85	0.95	1

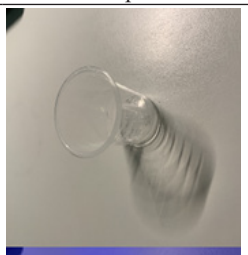
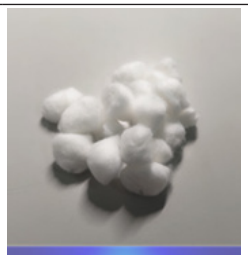
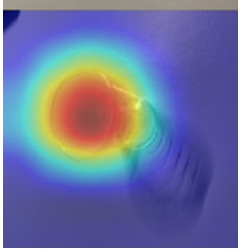
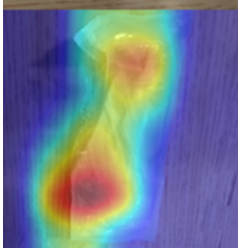
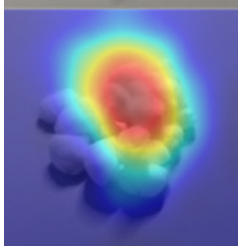
Chopsticks

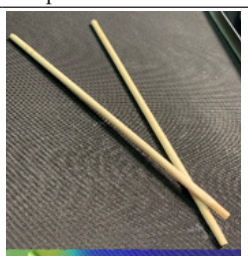
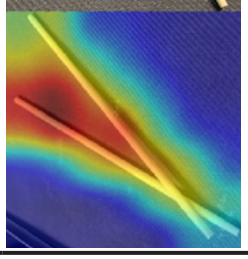
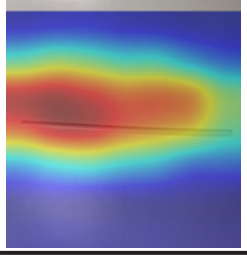
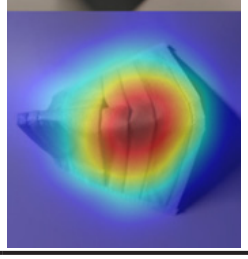
	1	2	3	4	5
1	1	0.89	0.43	0.35	0.34
2	0.89	1	0.51	0.43	0.40
3	0.43	0.51	1	0.78	0.80
4	0.35	0.43	0.78	1	0.98
5	0.34	0.40	0.80	0.98	1

Table 12. Correlation coefficient measurement for the images from interclass groups

	Cotton	Cotton swab	Gauze	Gloves	Masks	Tongue depressor	Medicine cup	Medical package	Marker pen	Straw	Chopsticks	Rag	Syringe	Hypodermic needle	Blade
Cotton	1	0.18	0.20	0.06	0.17	0.02	0.29	0.23	0.14	-0.01	0.14	-0.03	0.04	0.22	0.18
Cotton swab	0.18	1	0.03	0.14	0.02	0.25	0.20	0.08	0.25	0.06	0.25	0.43	0.00	0.18	0.18
Gauze	0.20	0.03	1	0.30	0.30	0.20	0.19	0.10	0.12	-0.02	0.12	0.01	0.39	0.03	0.19
Gloves	0.06	0.14	0.30	1	0.20	0.20	0.21	0.06	0.07	0.04	0.07	0.03	0.13	0.02	0.02
Masks	0.17	0.02	0.30	0.20	1	0.15	0.24	0.27	0.19	0.03	0.19	0.04	0.14	0.10	0.20
Tongue depressor	0.02	0.25	0.20	0.20	0.15	1	0.07	0.03	0.21	0.08	0.21	0.12	0.09	0.07	0.18
Medicine cup	0.29	0.20	0.19	0.21	0.24	0.07	1	0.23	0.15	0.03	0.15	0.01	0.06	0.21	0.11
Medical package	0.23	0.08	0.10	0.06	0.27	0.03	0.23	1	0.07	0.09	0.07	-0.01	0.01	0.16	0.07
Marker pen	0.14	0.25	0.12	0.07	0.19	0.21	0.15	0.07	1	0.09	1	0.20	0.05	0.18	0.41
Straw	-0.01	0.06	-0.02	0.04	0.03	0.08	0.03	0.09	0.09	1	0.09	0.21	0.10	0.34	0.16
Chopsticks	0.14	0.25	0.12	0.07	0.19	0.21	0.15	0.07	1	0.09	1	0.20	0.05	0.18	0.41
Rag	-0.03	0.43	0.01	0.03	0.04	0.12	0.01	-0.01	0.20	0.21	0.20	1	0.10	0.15	0.13
Syringe	0.04	0.00	0.39	0.13	0.14	0.09	0.06	0.01	0.05	0.10	0.05	0.10	1	0.20	0.10
Hypodermic needle	0.22	0.18	0.03	0.02	0.10	0.07	0.21	0.16	0.18	0.34	0.18	0.15	0.20	1	0.37
Blade	0.18	0.18	0.19	0.02	0.20	0.18	0.11	0.07	0.41	0.16	0.41	0.13	0.10	0.37	1

Table 13. The six types of waste images and their Grad-CAM images

Class	Medicine cup	Medical package	Cotton
Testing image			
Grad-CAM image			

Class	Chopsticks	Straw	Mask
Testing image			
Grad-CAM image			

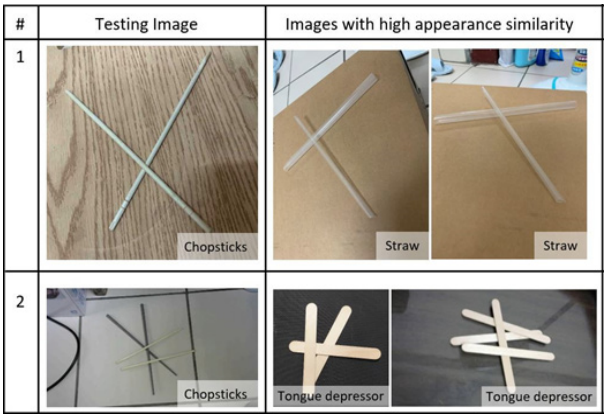
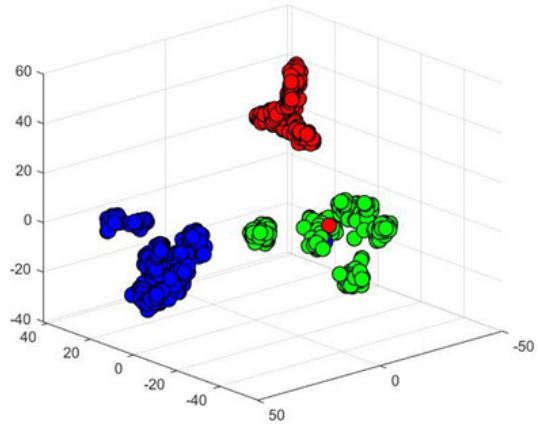
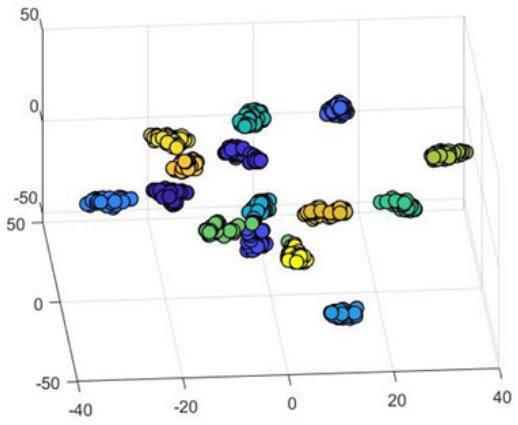


Figure 7. The example of the misclassified testing images with the images with high appearance similarity



(a) 3-class classification



(b) 15- class classification

Figure 8. T-SNE map of the activation layer extracted from deep learning architecture

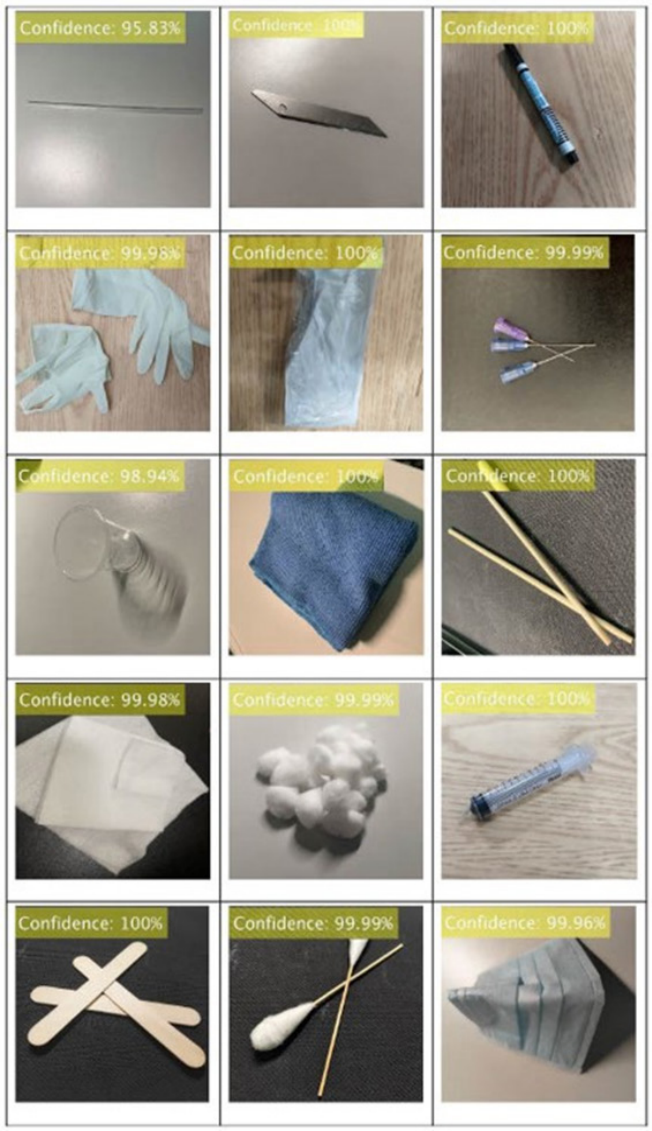


Figure 9. Confidence level of the testing images

4.4 Ablation Studies

To investigate the impact of hyperparameter selection on recognition performance, an ablation study is conducted focusing on the types of optimizers and the values of the learning rates. Three types of optimizers are evaluated: Adaptive Moment Estimation (Adam), Stochastic Gradient Descent with Momentum (SGDM), and Root Mean Square Propagation (RMSProp). The experiments are carried out using AlexNet and GoogLeNet architectures for the 3-class classification task, with the results illustrated in Figure 10 and Figure 11, respectively. It is observed that GoogLeNet outperforms AlexNet overall, corroborating the previously discussed findings that GoogLeNet consistently yields promising recognition rates. Regarding the choice of optimizer, the Adam solver outperforms the other two in both cases, as seen in Figure 10 and Figure 11. Therefore, Adam is selected for all subsequent experiments.

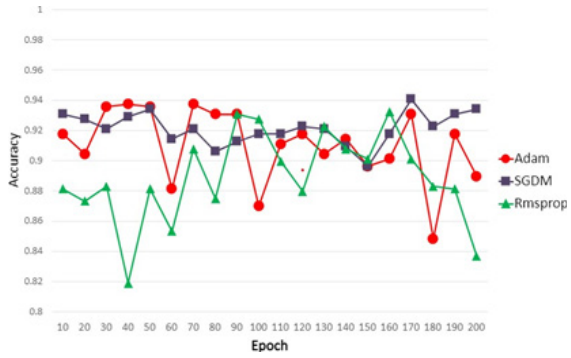


Figure 10. The recognition accuracy when employing different types of optimizers in training the AlexNet architecture for the 3-class classification task

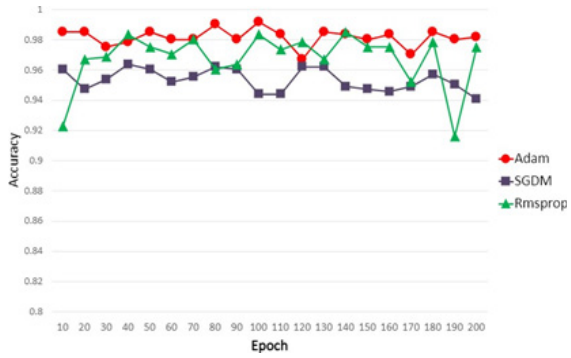


Figure 11. The recognition accuracy when employing different types of optimizers in training the GoogLeNet architecture for the 3-class classification task

Conversely, the learning rate is varied to identify the optimal value. Figure 12 displays the accuracy rates obtained when using three different learning rates: 0.00001, 0.0001, 0.001. This experiment employs GoogLeNet as the backbone architecture and uses the Adam solver as the optimizer. It is evident that a learning rate of 0.0001 results in a stable recognition rate with an accuracy of at least 95%. On a related note, the model training is configured such that the epoch size is set to [10, 200] with an interval of 10.

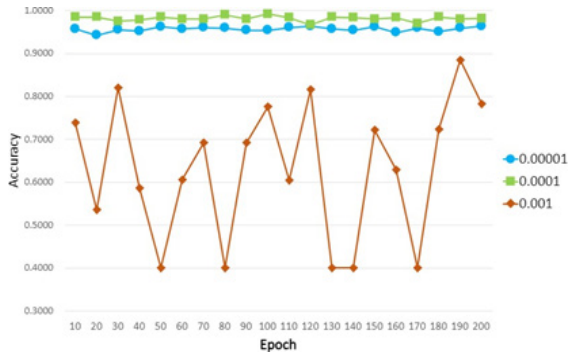


Figure 12. The recognition accuracy when employing different values of learn rate in training the GoogLeNet architecture for the 3-class classification task

The reason for the early stopping at the 200 epoch is because further training does not lead to any improvement. Concretely, the justification of selecting the values or options of the parameters are summarized below:

1. **Solver:** The performance results when employing different types of optimization solvers (i.e., adaptive moment estimation (Adam), stochastic gradient descent with momentum (SGDM), and Root Mean Square Propagation (RMSProp)) are demonstrated in Figure 10 and Figure 11. The recognition result when adopting Adam outperformed the two other optimization algorithms.
2. **Learn rate:** An ablation study is performed to evaluate the impact of adopting different learning rates (i.e., 0.00001, 0.0001, 0.001) when training the GoogLeNet architecture. Detailed performance yielded is demonstrated in Figure 12. As a result, we opt for the learning rate of 0.0001 in all the experiments herein, as it exhibits the best recognition rate among them.
3. **Epoch size:** For the learn rate settings, Figure 3 portrays the learning progress during the model training using GoogLeNet architecture. It can be observed that the network converges at the beginning of the few epochs and the value of training accuracy and training loss start to remain stagnant. In addition, the recognition result yielded after conducting the 3-fold cross-validation strategy reported in Table 7 and Table 8 in the manuscript evidenced the adequacy of the parameter option, as there is no overfitting phenomenon occurring.
4. **Mini-batch size:** This parameter value is often tuned to an aspect of the computational architecture on which the implementation is being executed. It does not affect accuracy, but it affects the training speed and memory usage.
5. **Shuffle:** The data is shuffled once randomly before training the model to reduce overfitting and variance. As such, the weights are more generalized and do converge faster, and produce better results.
6. **Gradient decay factor:** The decay rate of the gradient moving average for the Adam solver is specified as a value that is less than 1. As suggested by the MATLAB toolbox [38], the default value of 0.9 works well for most tasks.

On a related note, the 15-class classification aims to identify items that may be incorrectly sorted into the wrong garbage bin during the 3-class classification. This analysis sheds light on items that are prone to misclassification. For instance, the confusion matrix in Table 10 reveals that 2% of cotton swab samples are misclassified as medicine cups, suggesting that infectious waste could end up in the general garbage bin. Similarly, the table indicates that 2.5% of blade samples are misclassified as rags, meaning the blades could also be sorted into the general garbage bin. Such misclassification scenarios pose potential risks to both human health and the environment if waste management is not executed properly. To mitigate overfitting, K-fold cross-validation is employed as a widely recognized and effective measure against performance ambiguity. The number of parameters, network size, FLOPs, and MACs for each architecture are detailed in Table 14, providing insights into the computational burden associated with each architecture.

Table 14. The number of parameters and the size of each network, along with their FLOPs and MACs

	Network	# Parameter (Millions)	Size (MB)	FLOPs (billions)	MACs (billions)
1	AlexNet	61	227	1.11	7.36
2	GoogLeNet	7	27	1.46	1.39
3	VGG-16	138	138	15.5	30.8
4	ResNet-18	11.7	44	1.82	3.67
5	ResNet-50	25.6	96	3.67	7.58

4.5 Evaluation on Transformer-based Model

The rising prominence of deep learning models has indeed been remarkable in recent years, especially with the evolution of various kinds of models, most notably the transformer-based model. This model, with its core components like patch embedding and attention layers, is meticulously crafted to excel at capturing discriminant features in data. Figure 13 presents the performance results in terms of accuracy when training and testing the transformer-based model for a 3-class classification task. It is noteworthy that the accuracy of the model generally improved as the epoch size increased. Specifically, there was a consistent upward trend in accuracy from 100 to 400 epochs. Subsequently, the accuracy appeared to plateau, exhibiting slight fluctuations from epoch size 400 to 800. However, beyond epoch=800, a significant decrease in accuracy was observed. Overall, this transformer model achieved its maximum accuracy of 75% at epoch=500. Table 15 provides a detailed performance result, including other metrics such as F1-score, recall, and precision. These metrics generally followed a similar trend to accuracy.

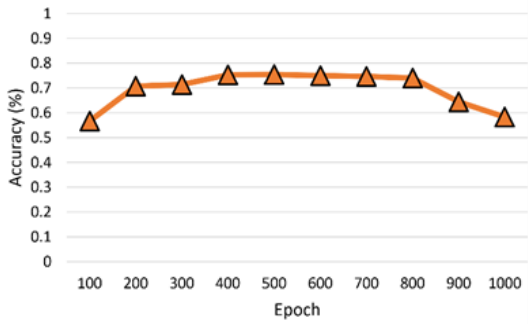


Figure 13. The performance result in terms of accuracy when training and testing using a transformer-based model for the 3-class classification task

The results presented in Table 15 and the accompanying discussion indicate that the transformer-based model may not be the optimal choice when evaluating it on a dataset comprising a relatively small number of samples, specifically 2025 images across three classes. Succinctly, the unsuitability of the transformer-based model can be attributed to several factors. Firstly, transformers thrive on large datasets to generalize effectively, but the limited data here may lead to overfitting or insufficient pattern capture. Secondly, the inherent complexity of transformer models can be excessive for the task, potentially causing overfitting and computational inefficiency. Additionally, fine-tuning on a small dataset becomes challenging, hampering the model's ability to adapt. Furthermore, transformers emphasize capturing long-range dependencies and relationships, which may not be prevalent in a small dataset, leading the model to focus on noise or irrelevant features. Moreover, to train a transformer-based network is generally highly computationally expensive compared to CNN, making them impractical for these applications.

Table 15. The performance results in terms of the metrics accuracy, F1-score, recall, and precision when training and testing using a transformer-based model for the 3-class classification task

	Accuracy	F1-score	Recall	Precision
100	0.5667	0.5409	0.5742	0.6115
200	0.7068	0.6749	0.6728	0.6878
300	0.7133	0.6989	0.7129	0.6967
400	0.7529	0.7297	0.7306	0.7380
500	0.7545	0.7436	0.7679	0.7585
600	0.7496	0.7389	0.7583	0.7419
700	0.7463	0.7356	0.7556	0.7539
800	0.7397	0.7280	0.7335	0.7447
900	0.6442	0.6294	0.6691	0.6817
1000	0.5832	0.5524	0.6045	0.6450

4.6 Limitation

Although the above-reported experimental results offer insightful and highly feasible evaluations of the proposed pipeline, several limitations warrant mention. Since this study performs two types of classification tasks, namely 3-class and 15-class, the dataset used is relatively small, consisting of only 2000 images. A larger dataset would be preferable to account for variations in input data, especially concerning objects with distinct colors and perspective angles. Another significant limitation is that the proposed model can only identify a single object in each image, leading to potential misclassification errors when multiple objects of different classes appear in the same image. To address this, detection or segmentation algorithms could be implemented to efficiently localize objects and identify their respective bounding boxes or boundaries.

Regarding the type of medical waste, this study focuses solely on solid waste, as it is recommended that biological liquid waste be poured down the drain and into the sewage system. However, the hardware platform is limited to handling small-scale (i.e., < 15 x 15 cm) and lightweight (i.e., < 1kg) waste. Despite these limitations, this automatic medical waste classification mechanism serves as a prototype that can easily be scaled up to manage multiple types of waste, such as regular waste, biohazardous waste, sharps waste, pharmaceutical waste, and hazardous pharmaceutical waste [39].

5 Conclusion

In summary, this work introduces a novel framework for the automatic recognition of medical waste. The proposed system includes a hardware platform for real-time classification and sorting. Extensive analyses and investigations confirm the system's reliability and robustness. The examination of complete confusion matrix entries and

correlation coefficient measurements for both intraclass and interclass groups indicates that appropriate experimental procedures and network models have been employed. Overall, the framework effectively performs both 3-class and 15-class classification tasks, achieving a remarkable classification rate of 99% on a self-collected database comprising approximately 2000 images. The experimental results provide both numerical and qualitative evidence, confirming the validity of the collected database.

While 85% of medical waste is non-hazardous and general, the growing volume of daily waste generated can be more efficiently managed both on-site and off-site through advanced computer vision technologies. As for future work, the proposed automatic medical waste classification and sorting system can be further improved by adding more categories, making it more practical for real-world implementation. For example, the system could include the five healthcare waste classes: regular waste, biohazardous waste, sharps waste, pharmaceutical waste, and hazardous pharmaceutical waste. Additionally, the hardware system could be enhanced to reduce computational time and the time required for garbage bin movement, necessitating more robust hardware components to improve the platform's mechanics.

Acknowledgement

This work was funded by the Ministry of Science and Technology (MOST) (Grant Number: MOST 111-2221-E-035-059-MY3 and MOST 113-2221-E-035-024-).

References

[1] Medical Waste Management, https://medwaste.epa.gov.ww/, 2023.

[2] Hong Kong Environmental Protection Agency, https://www.epd.gov.hk/epd/clinicalwaste/tc/largeproducer_duty_segregation.html, 2023.

[3] Taiwan Watch Institute, https://www.taiwanwatch.org.tw/node/653, 2023.

[4] M. Yang, G. Thung, *Classification of Trash for Recyclability Status*, CS229 Project Report, 2016.

[5] W.-L. Mao, W.-C. Chen, C.-T. Wang, Y.-H. Lin, Recycling Waste Classification using Optimized Convolutional Neural Network, *Resources, Conservation and Recycling*, Vol. 164, Article No. 105132, January, 2021.

[6] S. Meng, W.-T. Chu, A Study of Garbage Classification with Convolutional Neural Networks, *2020 Indo–Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*, Rajpura, India, 2020, pp. 152–157.

[7] C. Bircanoğlu, M. Atay, F. Be¸Ser, O¨ . Gen¸C, M. A. Kızrak, Recyclenet: Intelligent Waste Sorting using Deep Neural Networks, *2018 Innovations in Intelligent Systems and Applications (INISTA)*, Thessaloniki, Greece, 2018, pp. 1–7.

[8] R. A. Aral, Ş. R. Keskin, M. Kaya, M. HacïOmerõglu, Classification Of Trashnet Dataset based on Deep Learning Models, *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 2058–2062.

[9] A. H. Vo, L. H. Son, M. T. Vo, T. Le, A Novel Framework for Trash Classification using Deep Transfer Learning, *IEEE Access*, Vol. 7, pp. 178631–178639, December, 2019.

[10] P. F. Proen¸Ca, P. Simões, *Taco: Trash Annotations in Context for Litter Detection*, Arxiv, Preprint Arxiv: 2003.06975, March, 2020. https://arxiv.org/abs/2003.06975

[11] T. Wang, Y. Cai, L. Liang, D. Ye, A Multi-Level Approach to Waste Object Segmentation, *Sensors*, Vol. 20, No. 14, Article No. 3816, July, 2020.

[12] H. Panwar, P. 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 Studies in Chemical and Environmental Engineering*, Vol. 2, Article No. 100026, September, 2020.

[13] A. Masand, S. Chauhan, M. Jangid, R. Kumar, S. Roy, Scrapnet: An Efficient Approach to Trash Classification, IEEE Access, Vol. 9, 130947–130958, September, 2021.

[14] A. E. Abdel-Hakim, A. A. Farag, Csift: A Sift Descriptor with Color Invariant Characteristics, *2006 IEEE Computer Society Conference On Computer Vision and Pattern Recognition (CVPR)*, New York, NY, USA, 2006, pp. 1978–1983.

[15] L. Wang, *Support Vector Machines: Theory and Applications*, Vol. 177, Springer Science & Business Media, 2005.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, F. F. Li, Imagenet: A Large-Scale Hierarchical Image Database, *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009, pp. 248–255.

[17] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, 2017, pp. 4700–4708.

[18] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.

[19] W. Abdulla, *Mask R-CNN for Object Detection and Instance Segmentation on Keras and Tensorflow*, Github Repository, 2017.

[20] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks, *Transactions on Pattern Analysis & Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149, June, 2017.

[21] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, *Rethinking Atrous Convolution for Semantic Image Segmentation*, Arxiv, Preprint Arxiv: 1706.05587, December, 2017. https://arxiv.org/abs/1706.05587

[22] Openrecycle, GitHub, Inc., https://github.com/openrecycle/dataset, 2019.

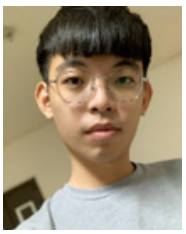
[23] Wasteclass, Kaggle, https://www.kaggle.com/techsash/

waste-classification-data, 2019.
[24] M. Tan, Q. Le, Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks, *36th International Conference on Machine Learning (PMLR)*, Long Beach, California, 2019, pp. 6105–6114.
[25] X. Bian, Y. Chen, S. Wang, F. Cheng, H. Cao, Medical Waste Classification System based on Opencv And Ssd-Mobilenet for 5G, *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Nanjing, China, 2021, pp. 1–6.
[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, SSD: Single Shot Multibox Detector, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), *European Conference on Computer Vision, Vol. 9905*, Springer, Cham, 2016, pp. 21–37.
[27] J. Chen, J. Mao, C. Thiel, Y. Wang, Iwaste: Video-Based Medical Waste Detection And Classification, *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, 2020, pp. 5794–5797.
[28] T. Mythili, A. Anbarasi, Enhanced Segmentation Network with Deep Learning for Biomedical Waste Classification, *Indian Journal of Science and Technology*, Vol. 14, No. 2, pp. 141–153, January, 2021.
[29] E. B. Tirkolaee, A. Goli, S. Mirjalili, Circular Economy Application in Designing Sustainable Medical Waste Management Systems, *Environmental Science and Pollution Research*, Vol. 29, No. 53, pp. 79667–79668, November, 2022.
[30] E. B. Tirkolaee, I. Mahdavi, M. M. S. Esfahani, G.-W. Weber, A Robust Green Location-Allocation-Inventory Problem to Design an Urban Waste Management System Under Uncertainty, *Waste Management*, Vol. 102, pp. 340–350, February, 2020.
[31] J. Bai, S. Lian, Z. Liu, K. Wang, D. Liu, Deep Learning based Robot for Automatically Picking Up Garbage on the Grass, *IEEE Transactions on Consumer Electronics*, Vol. 64, No. 3, pp. 382–389, August, 2018.
[32] S. Kong, M. Tian, C. Qiu, Z. Wu, J. Yu, Iwscr: An Intelligent Water Surface Cleaner Robot for Collecting Floating Garbage, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 51, No. 10, pp. 6358-6368, October, 2021.
[33] J. Redmon, A. Farhadi, *Yolov3: An Incremental Improvement*, Arxiv, Preprint Arxiv: 1804.02767, April, 2018. https://arxiv.org/abs/1804.02767
[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, F. F. Li, Imagenet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision*, Vol. 115, No. 3, pp. 211–252, December, 2015.
[35] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet Classification with Deep Convolutional Neural Networks, *Communications of the ACM*, Vol. 60, No. 6, pp. 84–90, June, 2017.
[36] K. Simonyan, A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, Arxiv, Preprint Arxiv: 1409.1556, September, 2014. https://arxiv.org/abs/1409.1556
[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, *IEEE Conference On Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1–9.
[38] Options for Training Deep Learning Neural Network, https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html. 2016.
[39] Medical Waste Disposal and Management Services, MedPro Disposal, https://www.medprodisposal.com/medical-waste-disposal/, 2023.

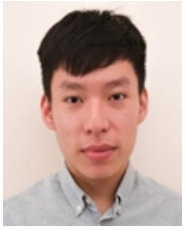
Biographies

Y. S. Gan is an assistant professor at the School of Architectures at Feng Chia University. He was a researcher at the National Taipei University of Nursing and Health Sciences Taiwan and an assistant professor at Xiamen University Malaysia. He received his B.E. and Ph.D. Degrees from Universiti Teknologi Malaysia. His research interests include DNA computing, artificial intelligence, machine learning, and computer vision.

Yuan-Zhi Liu is currently pursuing at the Department of Electronic Engineering at Feng Chia University. He is assisting Sze-Teng Liong and Y.S. Gan's research in the field of image processing study and mainly focuses on the industrial application of neural networks to improve accuracy. His main research interests include machine learning, pattern recognition, and artificial intelligence.



Bo-Cheng Tseng is currently pursuing at the Department of Electronic Engineering at Feng Chia University. He is assisting Sze-Teng Liong and Y.S. Gan's research in the field of computer vision and pattern recognition with applications in multimedia analysis. His research interests include artificial intelligence, statistical machine learning, and data mining.



Gen-Bing Liong is a PhD student at the Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, Malaysia. He received his B.CS degree from Multimedia University (MMU) Malaysia in 2021. His research interests include machine learning, image processing, and computer vision. Currently, he is actively working on facial micro-expression analysis, image understanding, and image restoration.

Sze-Teng Liong is an associate professor at the Department of Electronic Engineering at Feng Chia University. She received her B.E. degree from Multimedia University in 2014 and a Ph.D. Degree in the University of Malaya in 2017. Her research topics include machine learning, pattern recognition, image processing, and computer vision. She is currently leading the Computational Analytics & Cognitive Vision Lab at Feng Chia University.