



A deep learning approach based hardware solution to categorise garbage in environment

Tanya Gupta¹ · Rakshit Joshi¹ · Devarshi Mukhopadhyay¹ · Kartik Sachdeva¹ · Nikita Jain¹ · Deepali Virmani² · Laura Garcia-Hernandez³

Received: 10 January 2021 / Accepted: 10 July 2021 / Published online: 19 November 2021
© The Author(s) 2021

Abstract

Garbage detection and disposal have become one of the major hassles in urban planning. Due to population influx in urban areas, the rate of garbage generation has increased exponentially along with garbage diversity. In this paper, we propose a hardware solution for garbage segregation at the base level based on deep learning architecture. The proposed deep-learning-based hardware solution *SmartBin* can segregate the garbage into biodegradable and non-biodegradable using Image classification through a Convolutional Neural Network System Architecture using a Real-time embedded system. Garbage detection via image classification aims for quick and efficient categorization of garbage present in the bin. However, this is an arduous task as garbage can be of any dimension, object, color, texture, unlike object detection of a particular entity where images of objects of that entity do share some similar characteristics and traits. The objective of this work is to compare the performance of various pre-trained Convolution Neural Network namely AlexNet, ResNet, VGG-16, and InceptionNet for garbage classification and test their working along with hardware components (PiCam, raspberry pi, infrared sensors, etc.) used for garbage detection in the bin. The InceptionNet Neural Network showed the best performance measure for the proposed model with an accuracy of 98.15% and a loss of 0.10 for the training set while it was 96.23% and 0.13 for the validation set.

Keywords Inception net · CNN · Infrared sensors · Garbage segregation · Image categorization · Image classification

-
- ✉ Nikita Jain
nikita.jain@bharatividyaapeeth.edu
 - Tanya Gupta
23.tanyagupta@gmail.com
 - Rakshit Joshi
rakshitjoshi97@gmail.com
 - Devarshi Mukhopadhyay
debom97@gmail.com
 - Kartik Sachdeva
ks8346@gmail.com
 - Deepali Virmani
deepalivirmani@gmail.com; deepali.virmani@vips.edu
 - Laura Garcia-Hernandez
ir1gahel@uco.es

¹ Department of Computer & Science Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi, India

² Vivekananda Institute of Professional Studies - Technical Campus, College of Engineering, New Delhi, India

³ Area of Project Engineering, University of Córdoba, 14071 Córdoba, Spain

Introduction

Waste management in urban areas is an essential aspect of urban governance. With increasing waste generation daily and public apathy towards the situation, garbage is pooling in pockets of urban areas leading to the general decline in overall public hygiene. Efficient waste management is exceedingly essential to society. Currently, employed methods rely on manual separation, classification, and transportation, which require a lot of manual labor and money, hence making it a cumbersome and expensive process [1]. Several techniques of automatic waste detection and sorting have been proposed, such as automated sorting [2–5]. Hence, a massive amount of potential lies in the domain of waste management techniques and improving them.

Autonomous Garbage Detection is one such new technique [6] coupled with IoT [7]. In this technique, integration with other data systems such as Geographic Information Systems [8] can provide a system where overflow, as well as pooling of garbage, could be detected without the requirement of explicit human intervention. At present, various

models closely resemble autonomous garbage detection systems with the help of several techniques. Image processing via Convolutional Neural Network for classification, detection and categorization of images [9–12], image classification via Gray level co-occurrence matrix [13] and Gray level Aura Matrix [14], garbage detection using Support Vector Machines [15, 16], detection of pools of garbage on roadside via Video Content Analysis using VGGNet [17] and keyframe extraction, Volumetric estimation of garbage in the bin for preventing overflowing using poisson surface reconstruction [11], smart garbage bins for garbage level detection with RFID/IR/ultrasonic sensors [18], detection of unauthorized pooling of garbage using real-world surveillance [19], garbage detection and collection by robotic arms using Artificial Intelligence algorithms and raspberry pi [20], application based on Convolutional Neural Network (CNN) to track and report pools of garbage [21], Multilayer Hybrid deep learning method to extract image features using CNN and multilayer perceptrons (MLP) to consolidate image features to classify wastes as recyclable or the others [11], identifying the waste container using radio frequency identification [22, 23].

The proposed hardware solution *Smart Bin* is the first step towards garbage segregation into biodegradable and non-biodegradable objects at the base level. This model uses four pre-trained Convolutional Neural Networks, namely AlexNet, ResNet, VGG-16, and InceptionNet, out of which InceptionNet performs best and hence is the base model due to its better performance with an accuracy of 96.23–98.15% and loss varies between 0.10–0.13. The sample image is captured using the pi-cam and is transferred to the raspberry pi module, which in turn feeds it to the InceptionNet as input for classification. After classification, the servo motor rotates the separator disk either in a clockwise or anticlockwise direction. This solution provides a low-cost solution for base-level garbage detection and segregation. Unlike the models implemented so far for garbage classification, the proposed model also applies the hardware for waste segregation. The deep learning-based hardware solution aims to make it easier for municipal authorities to collect and segregate waste without much additional cost.

Section [Introduction](#) provides a brief introduction to the problem statement and the work done in that field. Section [Related work](#), describes the related work and the existing state-of-art in the form of models and technology development. Section [Dataset](#) presents the dataset description and the sources of data collection for the proposed work. In Sect. [System design](#), a system design that gives a basic idea of the working of the proposed solution along with cost estimation, is provided. The methodology describes the methods followed by the system, and the flow of the system is given in Sect. [Methodology](#). Section [Hardware component analysis](#) presents a hardware component analysis stating the description of hardware components used along with system

images. Section [Results and discussion](#) gives the results, and discussion shows the results obtained by the practical implementation of the system and discusses the limitations of the system in the real-world environment, along with a brief comparison of the current state of the art and Sect. [Conclusion](#): gives the conclusion derived from the research.

Related work

The field of image classification is widely used for garbage detection and categorization. Image classification is a process via which an image can be classified based on its visual content. Hiremath et al. [24] proposed a convolutional neural network model intending to segregate waste into three categories, namely compostable waste, recyclable waste, and reject waste, i.e., which is not recyclable or reusable. Ying-hao chu et al. [11] proposes a method using MHS (Multilayer hybrid deep learning method) that uses CNN to extract features of images and MLP to combine these features extracted from images through sensors, finally sorting the waste into two categories namely recyclable and others. The method achieved higher than 90% overall accuracy. Mindy Yang et al. [15] suggest a model that uses Support vector machines (SVM) with scale-invariant Feature Transformation (SIFT) highlights and a convolutional neural system (CNN). The tests demonstrated that the SVM performed superior to CNN because of the absence of several pictures. The model orders object present in the image into six classes to be specific glass, metal, paper, cardboard, plastic and refuse with an exactness of 63% with SVM and 22% with CNN. Alvaro Salamander et al. [25] proposes a framework that can break down pictures and control an automated arm and a transport line to arrange and isolate waste, using a canny algorithm for border detection and gaussian blur to filter noise which can visually classify types of waste effectively. Gaurav Mittal et al. [21] in SpotGarbage, which is an app to detect garbage using deep learning, proposes a smartphone application for the detection of pools of waste in images clicked by users that are geotagged. It uses CNN on images for the detection of garbage pools; the application achieves a mean accuracy of 87.69%. Mandar Satvilkar [12] in image-based trash classification using machine learning algorithms for recyclability status proposes a model that classifies the garbage image into five categories using CNN, the model uses SVM for extracting SIFT features. Miguel Valente et al. [23] have used computer vision to propose a model that detects different types of garbage containers using two different approaches. One is feature detectors/descriptors, which uses a vector of locally aggregated descriptors and CNN using You Only Look Once (YOLO), which gave an accuracy of 90%.

In the field of semantic segmentation, an image is divided into segments where all pixels in a segment have some com-

mon characteristics [26]. Cenk Bircanoglu et al. [9] compared various deep convolutional networks like Deep residual networks, Inception Res-Net, MobileNet, DenseNet, etc. [27]. Ying Liu et al. [10] proposed an improved YOLOv2 model, where the authors tweaked the parameters and used optimization and acceleration algorithms to strike a balance between real-time performance application and precision of target box clustering. Ying Wang et al. [6] proposes an autonomous system of garbage detection using the Faster R-CNN open-source framework. Here instead of VGG as the essential convolutional layers, ResNet Network was used. ShubenduSngt et al. [28] proposed a mobile application that takes complaints from the user in the form of images of garbage and classifies them into spurious and non-spurious claims. It analyses and classifies the image with the help of SVM and CNN library (TFlearn deep learning library in Tensorflow) and detects the presence or absence of garbage in the image with an accuracy of 85% and informs the concerned authority in case of non-spurious complaints. Ross Girchick et al. [29] proposed accurate object detection using R-CNN, which consists of three modules, the first generates category-independent region proposals, the second is a CNN that extracts features. The third is a set of specific linear SVMs. Alex Krizhevsky et al. [30] trained an extensive deep convolutional neural network consisting of 5 CNN and three fully connected layers. They wrote a highly optimized GPU version and trained their model using the ImageNet dataset. They adopted two optimization techniques, i.e., Data Augmentation and Dropout, to reduce overfitting; hence they achieved a minimum error rate of 15.3%, which was the best error rate in the competition.

The volumetric analysis provides the volume of an object or a pile of objects. In [31] proposes a way of processing the image structure based on image segmentation, 3D reconstruction, and volume estimation. The work uses sliding window segmentation based on deep neural networks, 3D reconstruction uses SIFT (Scale Invariant Feature Transform) features, and volume estimation is obtained using Poisson surface reconstruction. The output of the system is the volume of the garbage detected and analyzed.

Video Content Analysis (VCA) is the strategy for consequently examining video streams to recognize and decide worldly and spatial occasions with the goal that one gets valuable data about the video content. VCA can be utilized for constant observation of waste removal at the side of the road or flooding of dustbins. In [17] presents a model wherein video observation content is dissected, utilizing object arrangement. Items showing up in the video are recognized and ordered using a convolutional neural system model. A log record that contains the classes of unique items and the hour of appearance is produced for later search. Kimin Yun et al. [19] in Vision-based garbage dumping action detection for a real-world surveillance platform proposes a method for

detecting dumping of garbage in an uncouth and unauthorized way, in the real world. For detection and tracking of garbage, a background subtraction algorithm is used, while for the detection of humans, human joint estimation is used. A polling-based module detects the dumping action by humans.

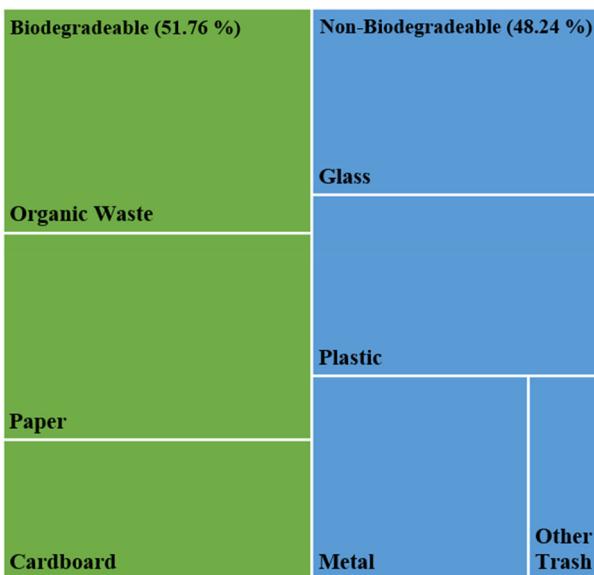
IoT implements hardware using the outputs from the different software implementation, which involves the calculation and semi-intelligent part [32]. Jinqiang Bai et al. [33] proposes a design of a robot that can detect and classify the garbage on the grass and picks it up with an average computational time of ground segmentation and garbage recognition of about 10.3 ms and 8.1 ms respectively, using CNN with SegNet. Sandeep et al. [18] suggested an IoT-based smart bin that has ultrasonic sensors embedded in it for the detection of the level of garbage and a microcontroller that will process the information provided by the sensor and send it further to the respective authorities. If waste reaches a certain level in the bin, the buzzer indicates overflow. The system provides a graphical view of the level of garbage in the container. Chen Zhihong et al. [34] proposes a conveyor belt system that deploys a camera array and robotic arms to detect garbage and estimate their poses using deep neural networks (RPN with VGG-16 model) with a computing time of about 220 ms. Siddhant Bansal et al. [20] in Automatic Garbage detection and collection proposes a model of a robotic arm that is controlled by a microcontroller to pick up the garbage and put it in a garbage bin. The system uses real-time video analysis, which can achieve four frames a second on Raspberry Pi, which uses AI algorithms to detect and calculate the distance of garbage by using a camera. The model is capable of detecting waste with 90% accuracy in real-time. Shah et al. [35] proposed a model for maximum value recovery from the trash bin while minimizing the transportation cost. Here the proposed model, the smart city, consists of several discreet sectors where various low capacity trucks and high capacity trucks handle the waste collection. The stochastic optimization model based on chance constraint programming is proposed to optimize waste collection operations' planning. For collecting the knowledge about bin status, RFID tags, actuators, capacity sensors, and a static GPS is used. Md. Abdulla Al Mamun et al. [36] suggests a bin that works in three-level architecture, i.e., smart bin, gateway, control station. The bin monitors the real-time parameters of the bin status, and an accelerometer is used to monitor the lid opening. It uses the ZigBee to transmit data to the gateway from the bin., and GPRS to transfer the data from the gateway to the control station. Six different sensors are used to monitor bin status that is driven by the hall effect of ultrasound load cell temperature-humidity sensors. Patric marques et al. [37] suggest an IOT based infrastructure to segregate waste in organic and recyclable waste in both indoor and outdoor scenarios. The results present a good response time; the architecture can handle 3902 garbage bins at the same time.

Table 1 Dataset description

Class label	Biodegradable			Non-biodegradable				Total images		
	Sub-label	Organic waste	Cardboard	Paper	Metal	Glass	Plastic			
No. of images		1935	1209	1782	1230	1503	1446	411	9516	
Percentage (%)		20.33		12.7		18.7	12.9	15.7	15.19	4.3

Dataset Tree Map

■ Biodegradeable (51.76 %) ■ Non-Biodegradeable (48.24 %)

**Fig. 1** Tree map representation of dataset

Dataset

The data used is obtained from three different sources; firstly, from the trashnet, this repository [38] contains images divided into five categories glass, paper, cardboard, plastic, metal, and others. Second, waste classification data [39] which includes two categories of images: organic and recyclable objects. Third, Drinking waste classification [40] which contains images of 4 types of drinking waste glass, aluminum cans, PET, and HDPE. All of these images are high-resolution images taken against a white background with almost no noise from other images of interfering objects.

Table 1 shows the division of the dataset compiled into seven classes, namely organic waste, cardboard, and paper for bio-degradable material constituting 51.76% of the dataset and metal, glass, plastic and other trash for non-bio-degradable representing 48.24% of the dataset material.

Figure 1 gives a treemap representation of the dataset used. Figure 2 shows the samples of dataset images.

System design

The proposed solution is a *Smart Bin*, which uses image classification for dumping the waste into either bio-degradable or non-biodegradable sections of the bin. It has a separating disk supported with an axle on the motor, which then rotates to up the detected segment. This model is currently based on the segregation of one entity at a time. Its implementation will help us understand and tackle the problem of multiple entity segregation.

Table 2 gives the approximate cost estimation of the proposed solution in the testing environment as well as for street-level implementation.

Two implementations can be performed based on the available hardware. They can also be found on the types of cameras used, which are Infrared Cameras and standard low-light sensitivity cameras. The difference between the two is that the latter won't be able to function properly at night, whereas IR camera modules will be accurate in the dark too. The use of these different modules is subject to testing, and the system will be able to determine the differences after implementation.

The proposed and implemented system solution uses Image Classification by CNN. CNN is the ace performer in the present-day image classification. Therefore, it will help in the efficient and accurate classification of trash. The data set will be self-acquired depending on the type of camera which will be used. Infrared cameras will need a set of infrared output images, and a standard camera will need generic images.

Methodology

The system for the waste segregation bin consisting of two functional components: Image Classification and Real-time Embedded System. The bin uses both of the modules to achieve the required operational results. The bin is also structurally divided into two sections, the Detection Section and the Segregated waste section. The system uses a sensor that signals the camera to take a picture, which is then processed by the image classification system, which sends the trash to the detected section type. The detected section type is divided on the x-axis into two halves, each for the

Fig. 2 Dataset samples of
(a) Organic waste, **(b)** Metal,
(c) Cardboard, **(d)** Paper,
(e) Glass, **(f)** Plastic and
(g) Other trash

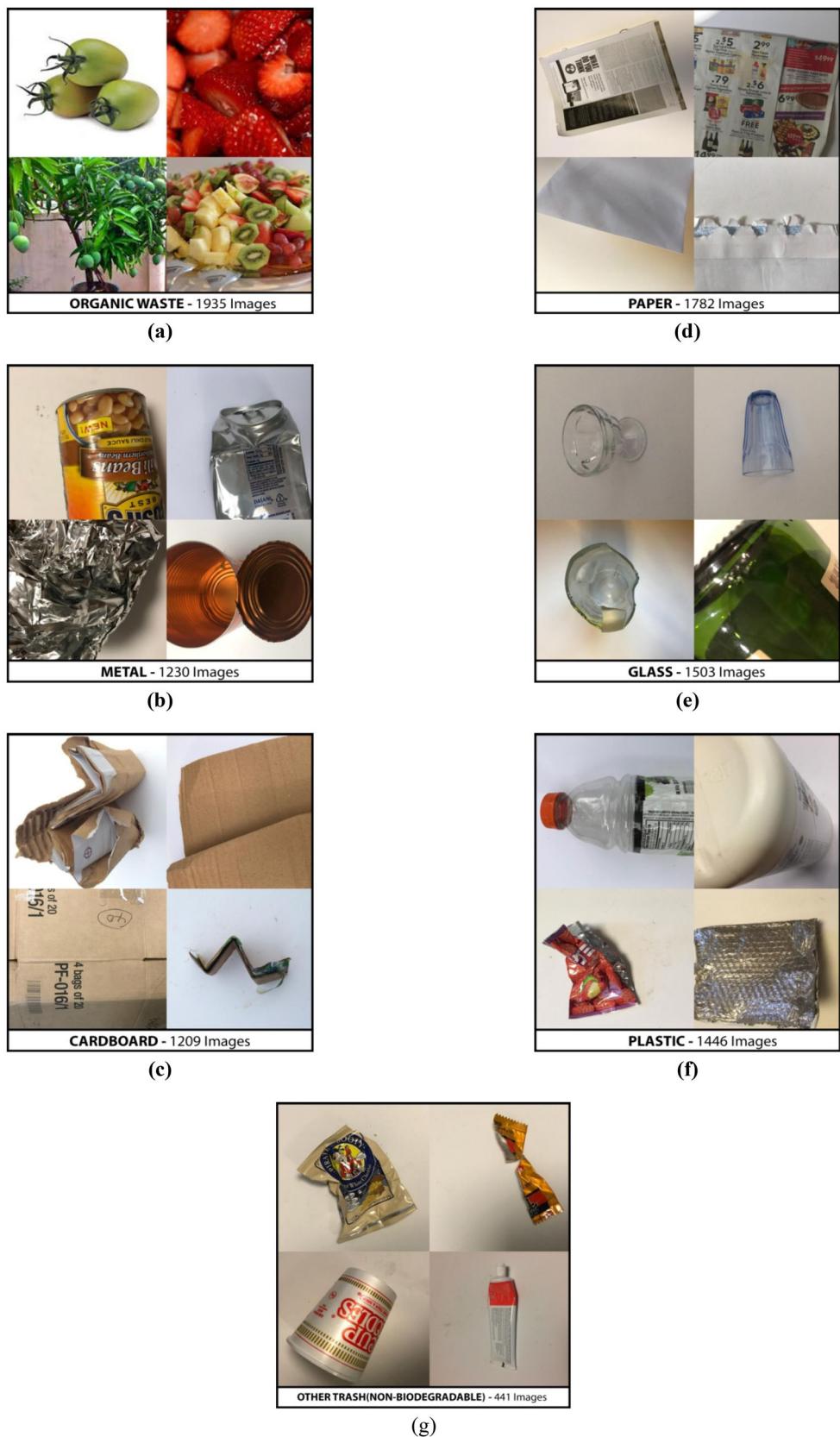


Table 2 Cost Estimation of the proposed model

Components I (for testing purposes)				
S.no	Component	Type	Quantity	Approx. Cost
1	Camera	Simple low light sensitive camera module	2	₹600
2	Bin	Plastic	1	₹300
3	Motor	Servo motor	1	₹150
4	Circuit chip	Arduino uno (less power implementation) supported via PC	1	₹400
Total cost				₹1450
Proposed components II (for street level implementation)				
1	Camera	Infrared sensor & camera module	2	₹650
2	Bin	Stainless steel + required components	1	₹600
3	Motor	DC Motor	1	₹50
4	Circuit chip	Raspberry Pi 3B	1	₹2750
Total cost				₹4050

two major class labels, i.e., Biodegradable waste and Non-Biodegradable waste.

Figure 3 explains the control flow of the proposed system. In building this system, various components are used, such as a sensor, a motor, and a push-button. A 5 MP Pi camera module is used for capturing the image of the trash thrown on the bin's separator disk.

On the press of a button, the IR sensor starts up and transmits IR waves. If the transmitting signal strikes any obstacle on its way, which is trash, it reverts to the receiver and triggers the camera module when the IR sensor gives a high signal at the output. The camera switches to take a picture after 5 s. These 5 s hence given ensures that the trash item is still and the camera sensors are adjusted to the lighting. The time interval also depends on the quality of the camera being used and can be changed accordingly. Since it's the cheapest 5 Megapixel Camera, the time for obtaining the best results is set as 5 s. After the camera successfully takes a picture, the program triggers the Image classification software module.

This system is tested using various CNN-based architectures for efficient classification: AlexNet, ResNet, VGG16, and InceptionNet. The first bit of leeway of CNN contrasted with its antecedents is that it consequently identifies the significant features with no human supervision. CNN is additionally computationally productive. It utilizes unique convolution and pooling tasks and performs parameter sharing. Hence CNN models are empowered to run on any gadget, making them all around alluring. Since InceptionNet shows the best performance for the system proposed & implemented, the *Smart Bin* Utilizes Google's deep convolutional network Inception Net V3 to classify the garbage object based on its image. Inception Net V3 is a deep CNN trained for the ImageNet Visual Recognition challenge, i.e., it is trained on millions of images and can differentiate between a thousand different classes. The 5 Megapixel camera, when

triggered, captures an image and redirects the image path to the classification program, which then runs a series of computations. Based on the output of the neural network, the images are classified into six categories. These six categories include metal, cardboard, paper, plastic, organic waste (food, etc.), and glass. These categories are then divided into two major categories biodegradable and non-biodegradable waste. The program returns a signal of high (1) for biodegradable and low (0) for non-biodegradable.

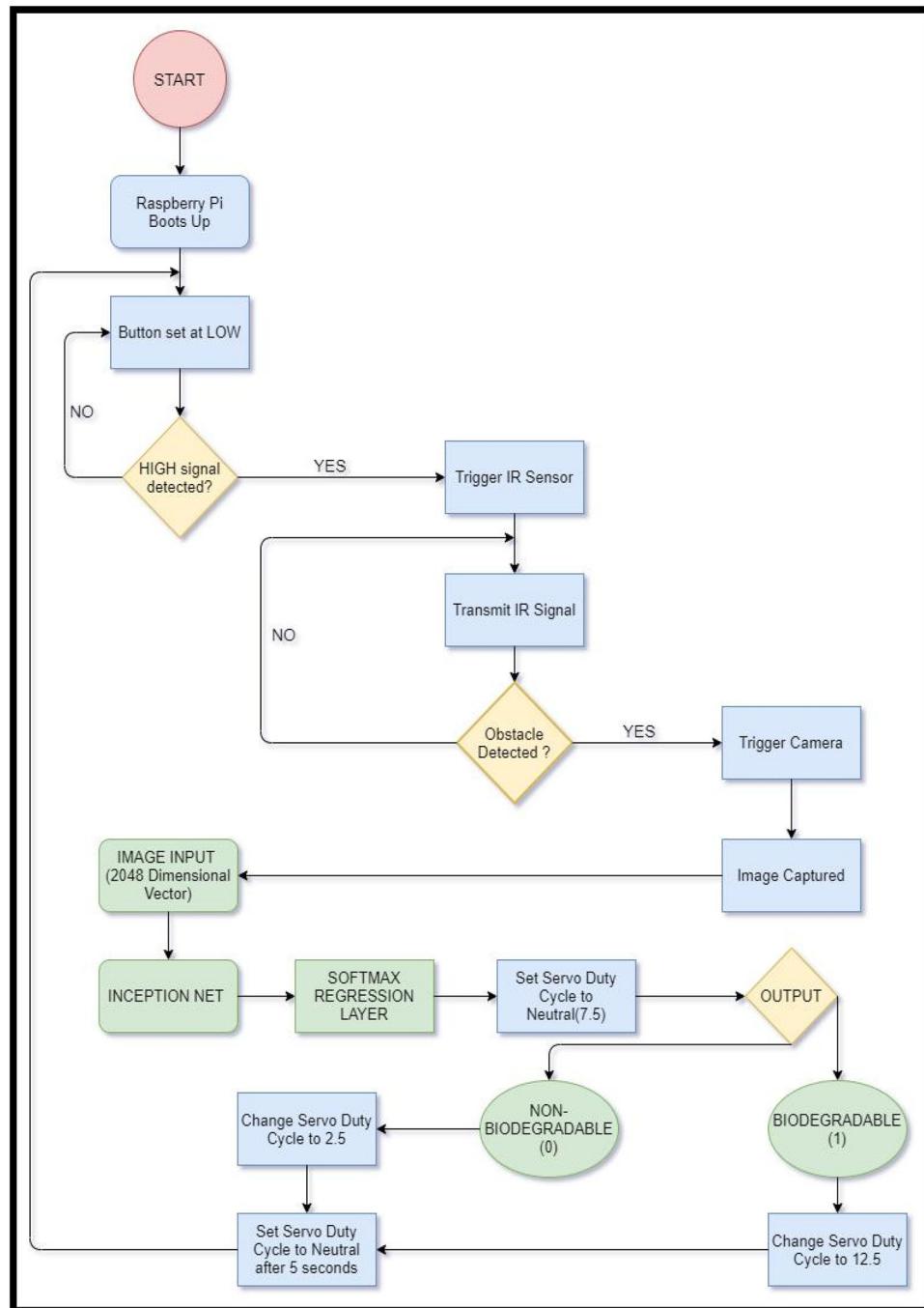
If trash is biodegradable, then the duty cycle of the servo motor, connected to the separator disk responsible for rotating the disk to sort out the waste into one of the two sections, is set to 12.5. Consequently, the motor rotates from the neutral state of duty cycle 7.5, which is 90 degrees, to 180 degrees, and in the biodegradable section. If trash is non-biodegradable, then the duty cycle of the motor is set to 2.5, which is 0 degrees, and the motor goes from 90 to 0 degrees to the non-biodegradable side. After 5 s, the motor rotates back to a neutral 7.5 duty cycle at 90 degrees. This is to ensure that the trash falls in the required section without failure.

Figure 4 represents hardware interrupts, such as a switch, which can be used to stop the driver script and end the infinite loop.

AlexNet

.AlexNet: the winner of the 2012 ImageNet LSVRC competition had a significant impact on the field of application of deep learning to computer vision. AlexNet consists of 8 layers, i.e., five convolutional layers and three fully connected layers. AlexNet introduced many of the fundamental approaches used as a standard today. They are—ReLU non-linearity instead of the standard tanh function, training on multiple GPUs, Overlapping pooling, Data Augmentation

Fig. 3 Solution system flow diagram



Methods (flipping and rotation), and dropout [41]. The block representation of AlexNet architecture is shown in Fig. 5.

a previous layer until the adjacent layer learns its weights. Figure 6 gives the resNet architecture [42].

ResNet

Introduction of Resnet in 2015 was one of the groundbreaking works in computer vision. ResNet50 consisted of 50 layers with skip connections after every 2nd layer to reduce the vanishing gradient problem by reusing activations from

VGG-16

VGG-16 is a CNN architecture introduced in 2014. The number ‘16’ stands for the number of layers in the network. This network utilizes 3×3 convolution of stride one layers stacked up, followed by max-pool layers of 2×2 of stride 2. This sub-structure is consistent throughout the net-

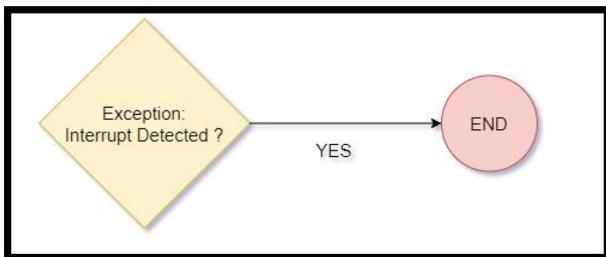


Fig. 4 Interrupt flow

work, followed by two fully-connected layers (each having 4096 nodes) and a softmax layer for classifying the output. Although a relatively simple network structure with fewer numbers of hyperparameters, the network is quite large and has an estimated 138 million parameters. Figure 7 shows the block diagram architecture of VGG 16 [43].

Inception net

Inception Net V3 is a 42-layer deep network with an approximate 7 million parameters. Inception Net is an architecture that utilizes a $1 \times 1 \times f$ convolutional block on a $N \times N \times N$ to produce a $N \times N \times f$ block. This block is further convolved by a $P \times P \times M$ block to create an $N-P+1$ output block finally.

$$f_{i,j,k_n}^n = \max(w_{k_n}^n f_{i,j}^{n-1} + b_{k_n}, 0) \quad (1)$$

Equation 1 is equivalent to the convolution layer with a 1×1 convolution kernel. Here (i, j) pixel index, $w_{k_n}^n$ is the numerical value of 1×1 block and b_{k_n} is the bias.

The motivation behind this is to reduce the computational cost incurred when a $N \times N \times N$ block is convolved by $P \times P \times M$ blocks to produce the $N-P+1$ block. Furthermore, the output from the parallel operation of multiple filter blocks is stacked up and concatenated to form the output block from one inception module. This not only helps in building a more in-depth and more extensive network but significantly

reduces the computational cost associated with building a deep convolutional neural network. Furthermore, this network utilizes one ‘dropout’ layer to fine-tune the network operations as well as using ‘label smoothening’ to prevent overfitting. Figure 8 represents the block diagram of Inception Net V3 architecture.

Activation functions-Rectified Linear units:—is a mathematical function defined as $y = \max(0, x)$. The reason for its common usage as an activation function is simple mathematical computation converges faster due to its linear nature. It is sparsely activated because, for all negative values, it outputs zero. Figure 9 gives an expanded view of the stem block already given in Fig. 8.

Inception Module A:

Two 3×3 convolutions replace one 5×5 convolution. This reduces the no. of parameters from 25 to 18, i.e., by a factor of 28%. Figure 10 gives an expanded view of the Inception A module in Fig. 8.

Grid Size reduction:

Another grid size reduction method is proposed instead of a simple max pooling. Two parallel stride blocks are used whose outputs are concatenated, i.e., **320 feature maps** are done by **convolution with stride 2**. **320 feature maps** are obtained by **max pooling**. And these two sets of feature maps are **concatenated as 640 feature maps** and go to the next level of inception module. Figures 11 and 12 give the expanded view of reduction A module and reduction B module in Fig. 8, respectively.

Inception module B:

One 3×1 convolution followed by one 1×3 convolution replaces one 3×3 convolution. Hence there is a reduction in the no. of parameters from 9 to 6, i.e., by a factor of 33%.

Figure 13 gives an expanded view of the Inception B module in reference to Fig. 8.

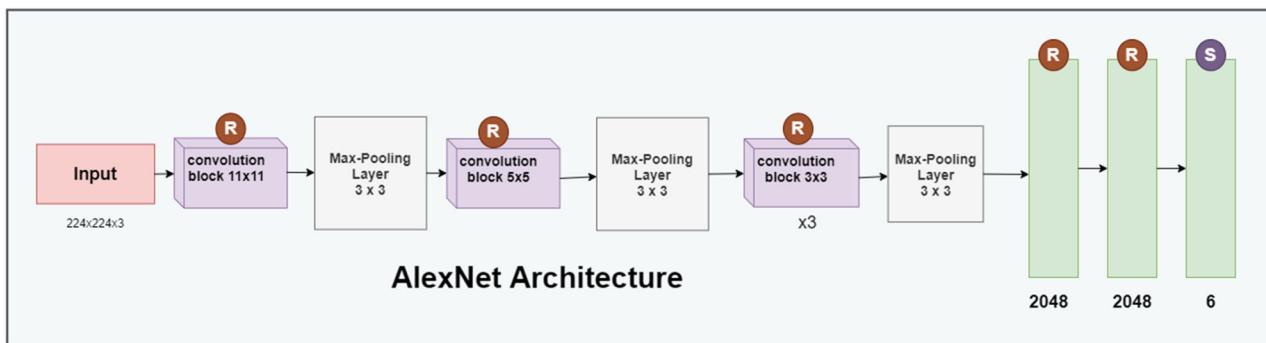
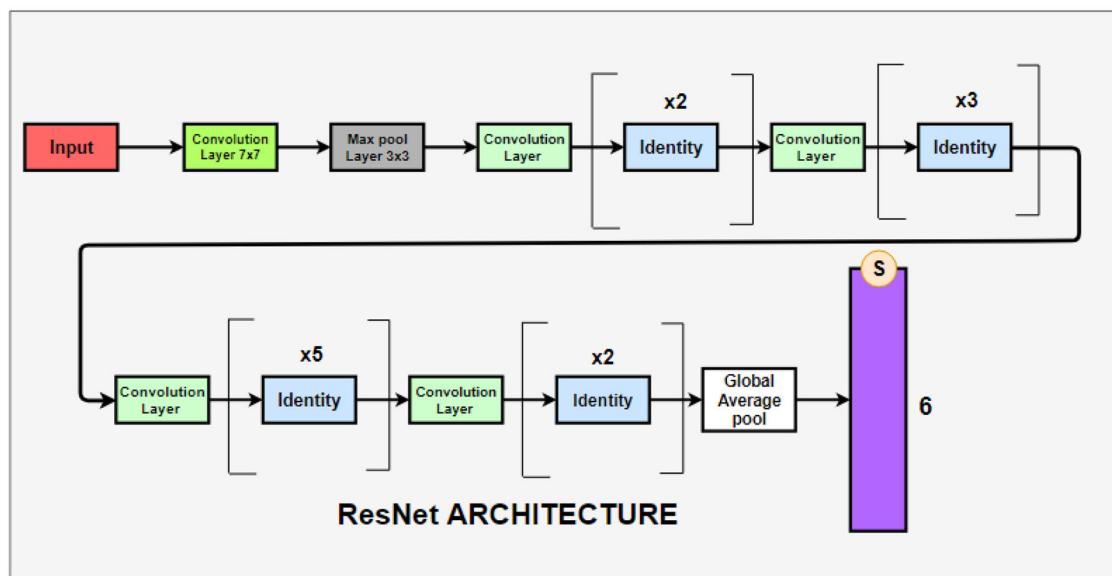
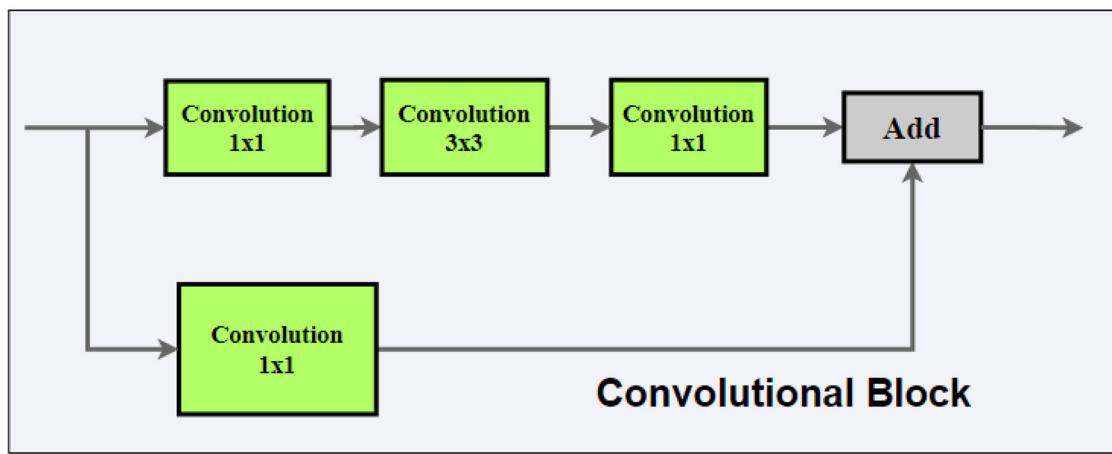


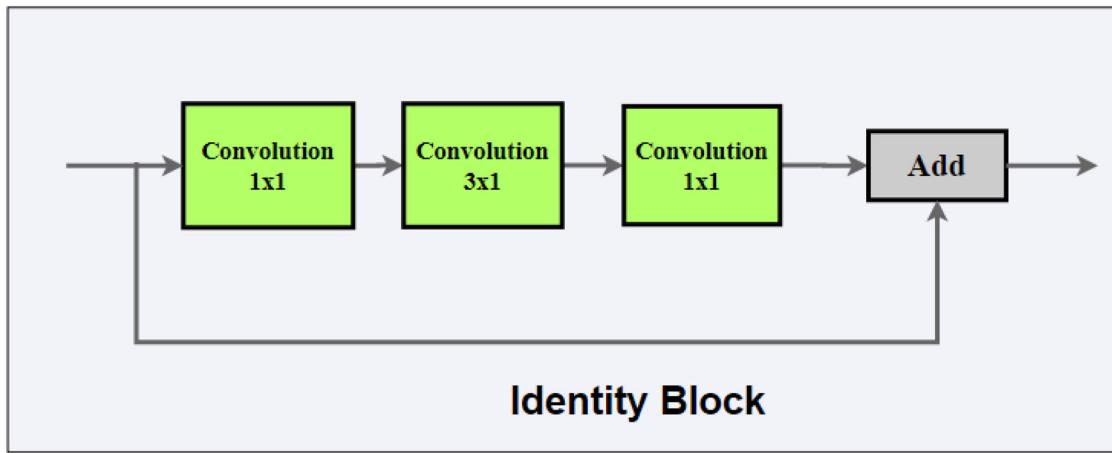
Fig. 5 AlexNet architecture block diagram



(a)



(b)



(c)

Fig. 6 (a) ResNet architecture block diagram, (b) Convolutional block and (c) Identity block

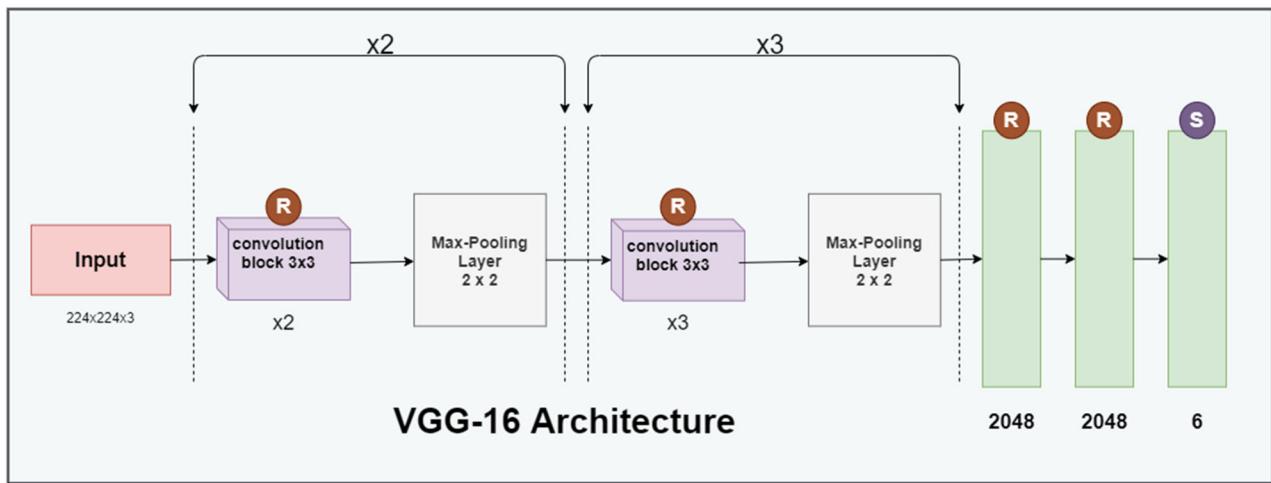


Fig. 7 VGG-16 architecture block diagram

Fig. 8 Inception net architecture block diagram

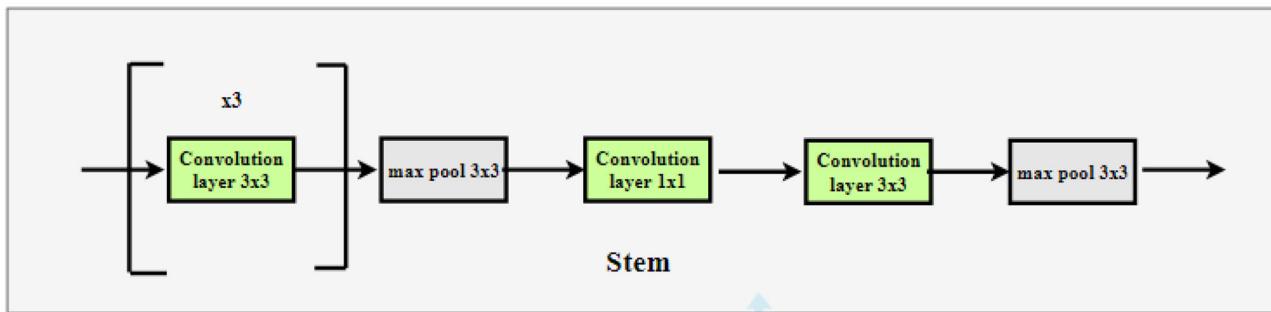
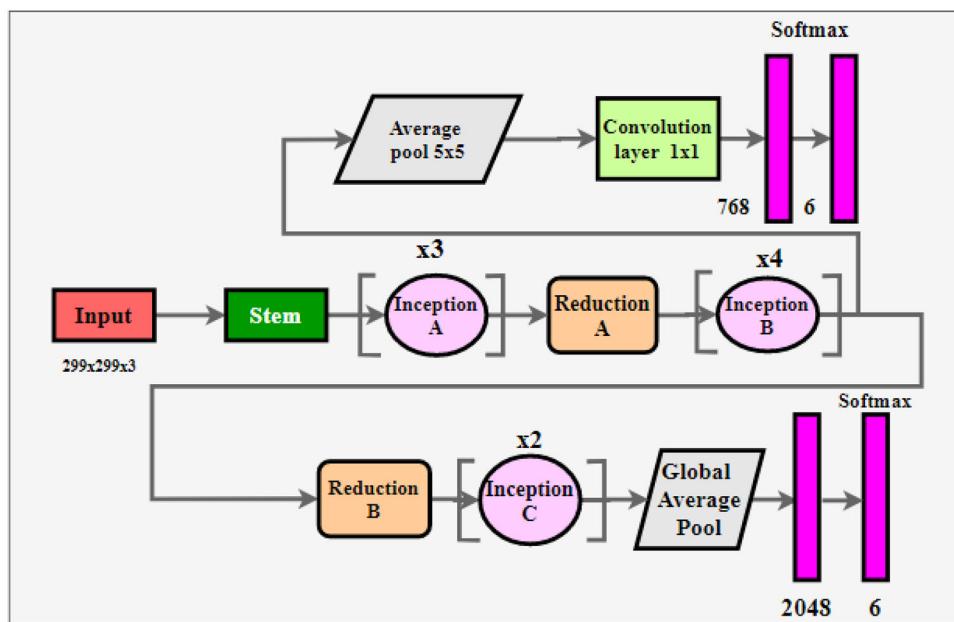


Fig. 9 Stem block diagram

Fig. 10 Inception A block diagram

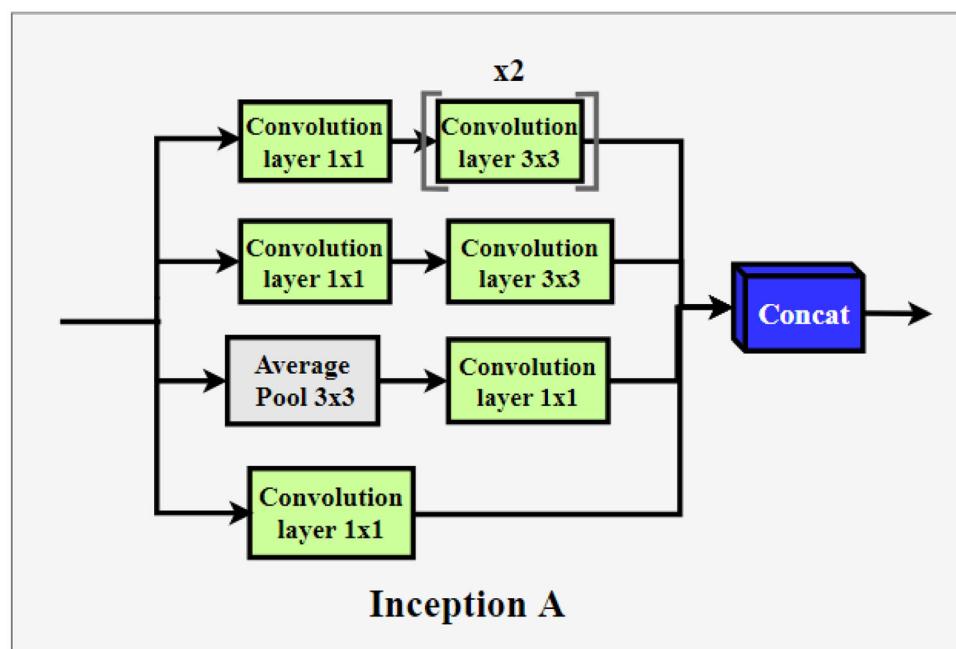


Fig. 11 Reduction A block diagram

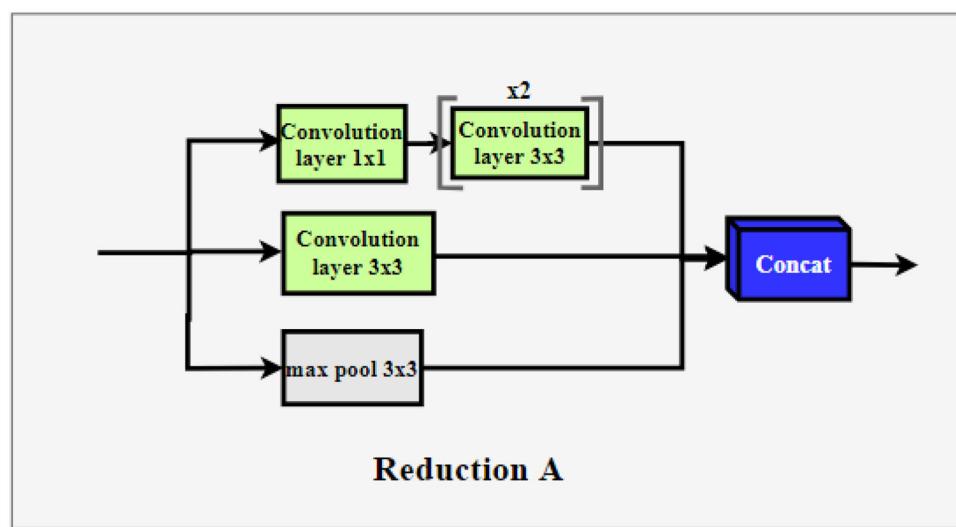


Fig. 12 Reduction B block diagram

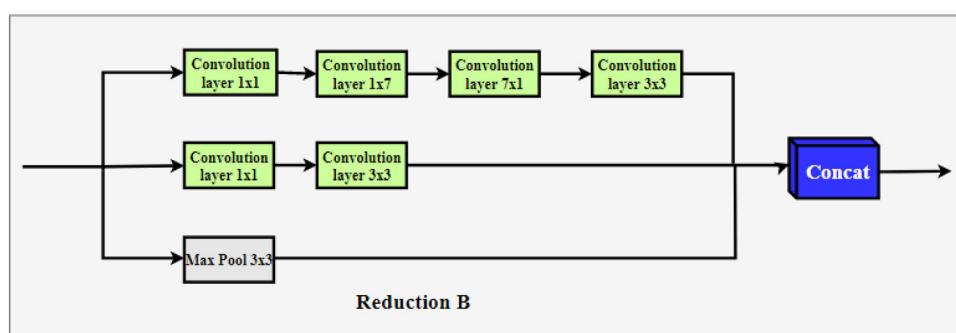


Fig. 13 Inception B block diagram

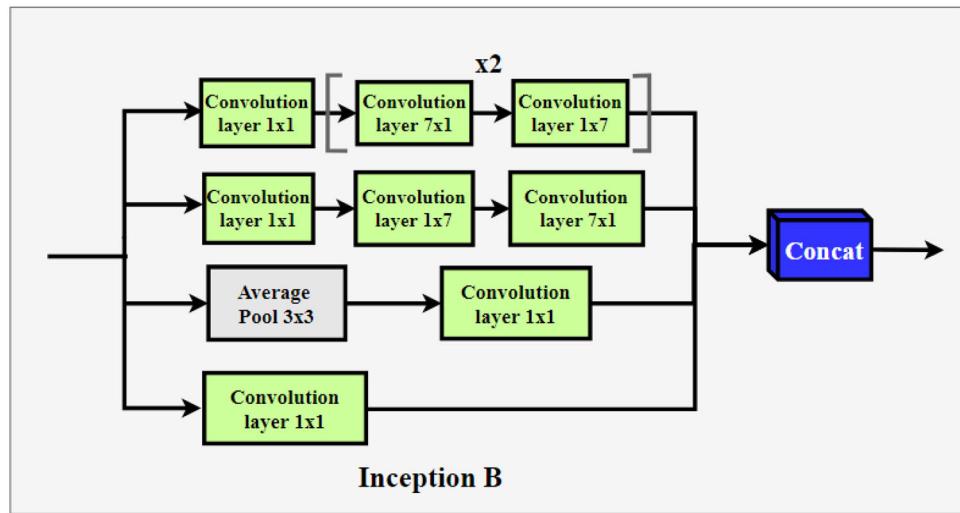


Fig. 14 Inception C block diagram

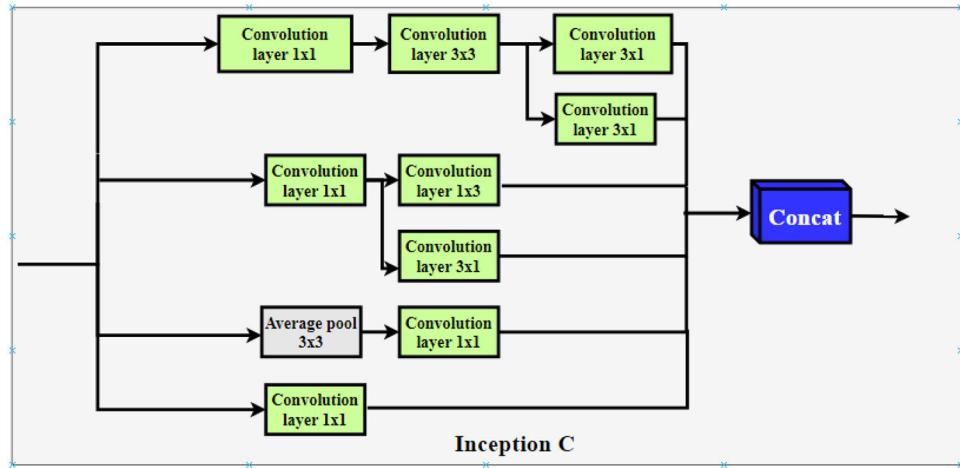
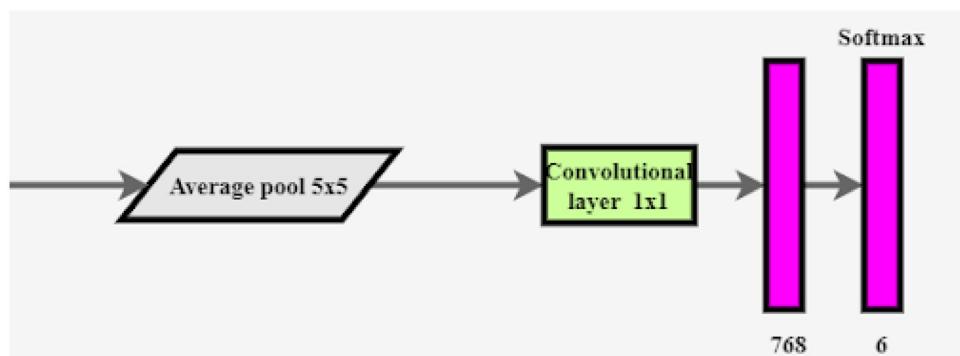


Fig. 15 Auxiliary classifier block diagram



Inception module C

It utilizes asymmetric factorization (replacing a $N \times N$ convolution with $1 \times N$ followed by $N \times 1$ convolution block) to promote high dimensional representation i.e. each sub-block of convolution is factorized into further sub-blocks and then whose output blocks are concatenated to give the output of the module as shown in the above Fig. 14.

Auxiliary classifier:

One auxiliary classifier is used on top of the last 17×17 layers as the function of the auxiliary classifier is more of a regularizer. Figure 15 gives an expanded view of the Auxiliary classifier.

RMS prop optimizer:

It is an optimization algorithm for parameter updates. It is a gradient descent based adaptive learning method in which the

oscillations in the vertical directions are damped to increase the learning rate and hence take more significant steps in the horizontal direction, thus reaching the minimum faster.

$$v_t = \rho v_{t-1} + (1 - \rho) * g_t^2 \quad (2)$$

$$\Delta\omega_t = -\frac{\varphi}{\sqrt{v_t + \epsilon}} * g_t \quad (3)$$

$$\omega_{t+1} = \omega_t + \Delta\omega_t \quad (4)$$

Here, φ is the initial learning rate, v_t is the exponential average of square gradients and g_t is the gradient at a given time along ω .

Label smoothening:

It is a mechanism to regularize the classifier layer by estimating the marginalized effect of label- dropout during training. Label smoothening is a variation on one hot encoding where the negative labels have a value slightly higher than 0, say 0.1, and the positive labels have a value somewhat lower than 1, say 0.9. The idea behind layer smoothening is that it aims to tune the layer parameters of the penultimate layer in such a manner that it outputs a value close to the value of the correct class. At the same time, the probability values of the incorrect class are equally distant.

$$y_k^{LS} = y_k(1 - \alpha) + \frac{\alpha}{k} \quad (5)$$

where k is the number of classes, α is the hyperparameter that determines the amount of smoothening, y_k is one hot encoded label vector and y_k^{LS} as the class value.

Softmax classifier: It is an activation function, similar to a sigmoid function, i.e., it outputs a value depicting a probability of a data instance belonging to a class. It is a multi-class activation function, as it outputs a matrix of $n \times 1$ where each cell holds a value depicting the probability of that data instance belonging to that class.

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=0}^k e^{x_j}} \quad (6)$$

where x_i, x_j are the vector outputs from the previous layer.

Transfer learning:

Transfer learning is utilized when instead of developing a deep learning model from scratch, the weights of a pre-trained model is tweaked by retraining the model through a particular dataset with a predetermined number of classes. The output is a class value, as determined by the dataset class values. Transfer learning cuts down on the need to have high-end processing hardware as well as a significant amount of time needed to develop a model by utilizing pre-built architectures like Inception Net, VGG-16, ResNet, etc. to function according to our requirements.

Implementation

Pre-processing:

A portion of the sample images was subject to distortions like crops, scales, or flips. These introduce variation in the training process for an object type, thus mirroring a real-world situation where the image of the object might be in a different orientation compared to that present in the dataset. However, when training with distortions, cached bottlenecks values can't be used, and it has to be recalculated again for each image. Thus, only random images from each category were chosen to be trained with distortion so as not to slow down the training process.

Raspberry Pi automatic execution: driver script

An automatic launch is set up using an endless loop and using the command “`sudo python /home/pi/garbageseg.py &`.” This command allows raspberry pi to execute this script as soon as the system is booted on being powered up and run without human intervention. This script runs in an endless loop and is the driver script for the smart bin. The primary function of the driver script is to control all the hardware functions used to trigger various connected devices and to execute the used classification program. This script also loads up all the libraries required to perform different tasks at boot up time. A hardware interrupt or power cut can stop the indefinite loop in the script.

Button control: “USE ME” and the background functions

Hardware component button “USE ME” opens up the lid of the bin on being pressed. But it has another background function, which is to allow the Driver Script to startup the IR sensor and PiCam. The following function helps to save power and prolong device life. After the required peripherals are used, they have shut down automatically until the button initiates the whole process again. The button control is at the start of the loop, which runs indefinitely.

Sensor control mechanism and the PiCamera

The peripheral in use, here, is the IR Sensor. IR Sensor's main objective is to keep emitting rays of frequency lying in the region of the Near-Infrared, which is 700 nm to 1400 nm. As they are not visible to the human eye, the camera can capture images without an issue even if the sensor is on. The rays emitted by the transmitter, intercept the obstruction and are reflected in the receiver by the sensor module. The following obstruction triggers and starts the camera to a high signal. The camera captures the image and stores it. The driver script shuts down the camera and sensor as soon as one input is acquired.

Algorithm 1 explains how the Raspberry Pi Interfaces with the peripherals using the driver script and sends control over to Algorithm 2 to classify images into two class labels and return control over to Algorithm 1.

Algorithm 1: Pseudo-code of Raspberry Pi Interface

```

1 Initialize GPIO;
2 GPIO.setup(switch, GPIO.IN, pull_up_down=GPIO.PUD_UP);
3 while switch is HIGH do
4     GPIO.setup(but2, GPIO.IN, pull_up_down=GPIO.PUD_UP);
5     #Enable input and pull up resistors of button
6     try:
7         i=0;
8         while i is 0 do
9             input_state = GPIO.input(but);
#reads signal from button
11            if input_state is FALSE then
12                GPIO.setup(sensor,GPIO.IN);
#Enable IR sensor
14                if GPIO.input(sensor) then
15                    picam captures an image after 5 seconds;
16                    run Algorithm 2;
17                    out=binary output;
#algorithm 2 returns 1 or 0
19                    GPIO.setup(servo, GPIO.OUT);
20                    Servo set to neutral duty cycle = 7.5 and 50Hz frequency;
21                    if out is 1 then
22                        servo set to 10.5;
end
23                    else
24                        servo set to 4.5;
end
25
26
27
28
29
30
31 except HardwareInterrupt:
32     GPIO.cleanup();
33
34 end

```

The driver script enables the classification program and waits for the output. **Algorithm 2** explains how Inception Net does the classification process before redirecting control to Algorithm 1.

Algorithm 2: Pseudo Code of Inception Net Implementation

```

1 for num in iterations do
2     for mini_batch in batches do
3         distort_images=random(mini_batch) // random images from the mini batch
4          $\nabla f(w_k)$  =inception(distort_images) //change of predicted value w.r.t correct value
5         #update parameter
6          $w_{k+1} = \beta w_k + \frac{\epsilon}{\sqrt{s_{k+1}^2}} \nabla f(w_k)$ 
7     end
8     //compute moving averages of mean and variance over the mini batches for normalization
9      $\hat{\mu} = \alpha \hat{\mu}_{t-1} + (1 - \alpha) \mu_t$ 
10     $\hat{\sigma}_t^2 = \alpha \hat{\sigma}_{t-1}^2 + (1 - \alpha) \sigma_t^2$ 
11 end

```

Table 3 Image processing parameters across the inception net V3

Type	Patch size/stride	Parameters	Activations	Input	Output
Input layer		0	268,203	299 × 299 × 3	299 × 299 × 3
Convolution	3 × 3/2	896	710,432	299 × 299 × 3	149 × 149 × 32
Convolution	3 × 3/1	9,248	691,488	149 × 149 × 32	147 × 147 × 32
Padded convolution	3 × 3/1	18,496	1,382,976	147 × 147 × 32	147 × 147 × 64
Pool	3 × 3/2	36,928	341,056	147 × 147 × 64	73 × 73 × 64
Conv	3 × 3/1	46,160	403,280	73 × 73 × 64	71 × 71 × 80
Conv	3 × 3/2	138,624	235,200	71 × 71 × 80	35 × 35 × 192
Conv	3 × 3/1	497,952	352,800	35 × 35 × 192	35 × 35 × 288
5 × inception module A	Figure 10	5,141,564	1,463,104	35 × 35 × 288	17 × 17 × 768
4 × inception module B	Figure 13	9,342,567	303,872	17 × 17 × 768	8 × 8 × 1280
2 × inception module C	Figure 14	7,125,321	262,144	8 × 8 × 1280	8 × 8 × 2048
Pool	8 × 8	–	2048	8 × 8 × 2048	1 × 1 × 2048
Linear	Logits	–	6	1 × 1 × 2048	1 × 1 × 6
Softmax	Classifier	–	–	1 × 1 × 6	Class

INCEPTION v3 MODEL

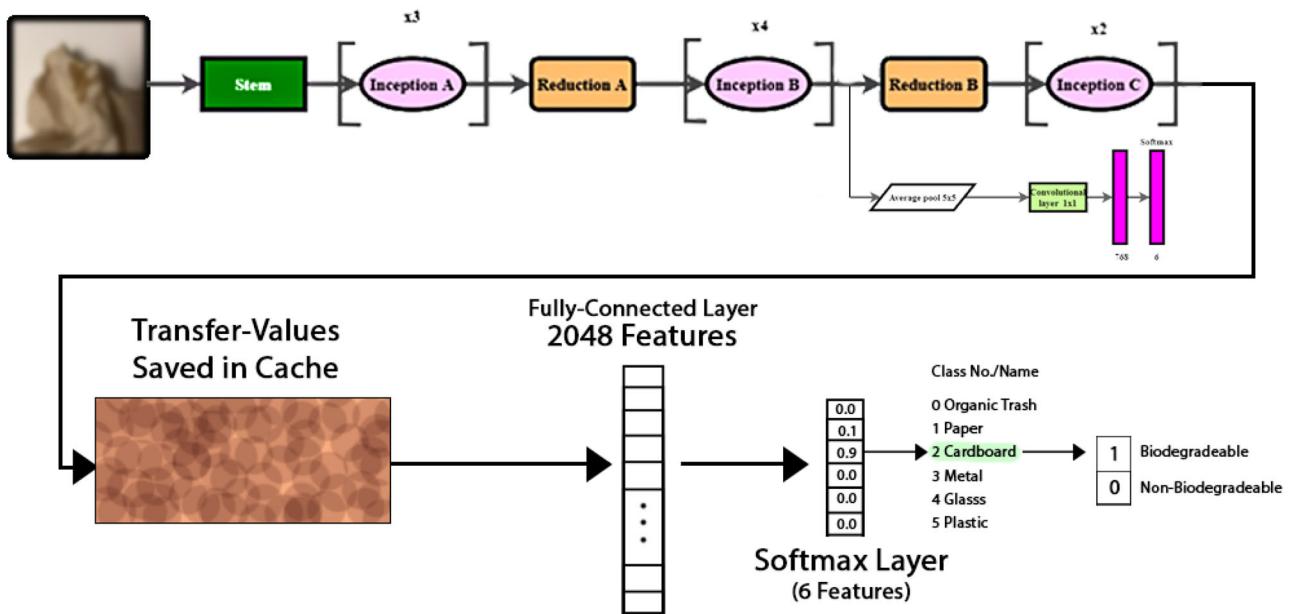
**Fig. 16** Inception net V3 system model

Table 3 represents the parameters for the original InceptionV3 architecture except for the softmax layer, which categorized the output into six classes instead of a thousand. In actuality, we only retrain the softmax layer or build another layer on top of the softmax layer to categorize the object according to the determined classes set by us. Hence from the perspective of our retrained model, the number of trainable parameters is 4,719,105, while the non-trainable settings are 19,802,784.

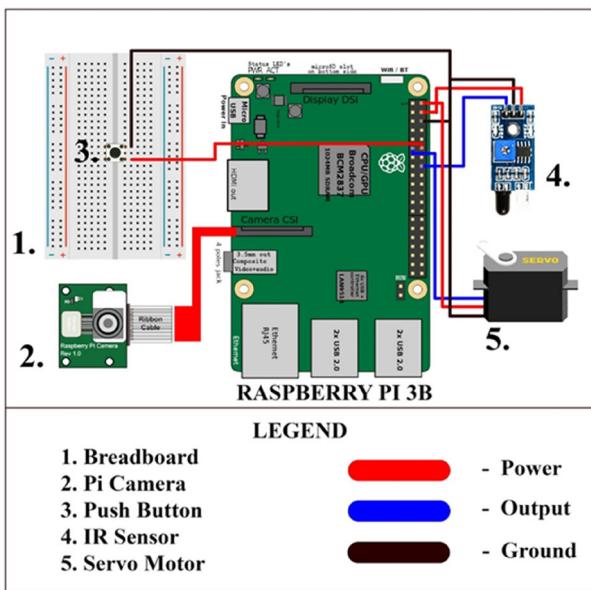
Figure 16 describes the proposed system, the softmax layers of each model are retrained to produce a 6-class output based upon which a decision is taken as to whether the object is biodegradable or non-biodegradable.

Separating mechanism

The driver script initializes the servo motor and sets the duty cycle of the motor to a neutral position, which is 7.5. The driver script receives a binary output of the classification program. The duty cycle is changed to 10.5 for output 1.

Table 4 Raspberry pi 3B features

Features	Description
Processor	1.2 GHz Broadcom BCM2837 64bit
RAM	1 GB
Connectivity	BCM43438 wireless LAN, Bluetooth Low Energy
I/O	40 pin GPIO, 4 USB Ports, HDMI, CSI Camera Port

**Fig. 17** Circuit diagram

After a 5-s delay, the cycle is changed back to 7.5, helping the garbage to fall into the biodegradable section. And if the output is 0, the duty cycle is changed to 4.5, allowing the garbage to fall in the non-biodegradable part (Table 4).

Hardware component analysis

Figure 17 depicts the circuit diagram of the proposed system.

Raspberry Pi 3B

The Raspberry Pi is a small single-board PC that was created in the United Kingdom by the Raspberry Pi Foundation to advance the education of essential software engineering in schools and developing nations. The Raspberry Pi 3B is the first model in the third generation of the series and involves a great stride in the performance as compared to the previous generations.

Infrared (IR) sensor

Infrared innovation tends to a wide assortment of remote applications. The primary zones are sensors and remote con-

trols. In the electromagnetic spectral range, the infrared range is separated into three types: near-infrared, mid-infrared, and far-infrared. The near-infrared region has a range of 700 nm to 1400 nm. This frequency is higher than the microwave spectrum and less than the visible light spectrum. For optical detection and trans-receiving, optics utilizes the near-infrared spectrum as the light is not as complex as Radio Frequency. IR information transmission for short-range applications. The module uses an IR sensor to transmit near IR waves, which, when obstructed, reflects the signals which are received by the transmitter and helps light up the sensing LED.

Pi camera

Raspberry Pi has a CSI Connector which allows the Pi camera to be directly connected. The camera has a 15-pin MIPI Camera Serial Interface, which allows high data transfer rates. The main feature of this camera is that it is really cheap and readily available to anyone. It is fully compatible with different models of Raspberry Pi. It has a 5 MegaPixelOmnivision 5647 Camera Module, which allows still picture resolution of up to 2592×1944 pixels and video resolution of 1080p at 30fps, 720p at 60fps and 640×480 p at 60/90. It is small in size ($20 \times 25 \times 9$ mm), which allows it to be easily carryable and fits in almost all Raspberry cases and weighs a mere 3 g.

Servo motor

A servo motor is an electric motor. It uses servomechanism for control. A motor associated with servomechanism of a DC is known as DC servo motor. A linear or rotary actuator that gives accurate control of linear and angular movement, acceleration, and velocity is known as a servo motor. A motor coupled to its sensor for feedback of the position helms a sophisticated control system that is dedicated and specially designed for the servo.

Figure 18 shows the implemented hardware system where Fig. 18a–d shows all the components of the smart bin and Fig. 18e shows how IR sensor detects Trash and Fig. 18f shows how the system responds to the output of the classification algorithm. Figure 18e shows a cardboard piece being detected by an IR sensor, which is then segregated in the biodegradable section in Fig. 18f.

Results and discussion

Figure 19 gives a graphical view of accuracy vs. epochs and loss vs. epoch for each pre-trained CNN. Each graph depicts the change in the accuracy of the model. Also, the loss of the model is decreasing as the model is trained (no. of epochs). These are used as the parameters based upon

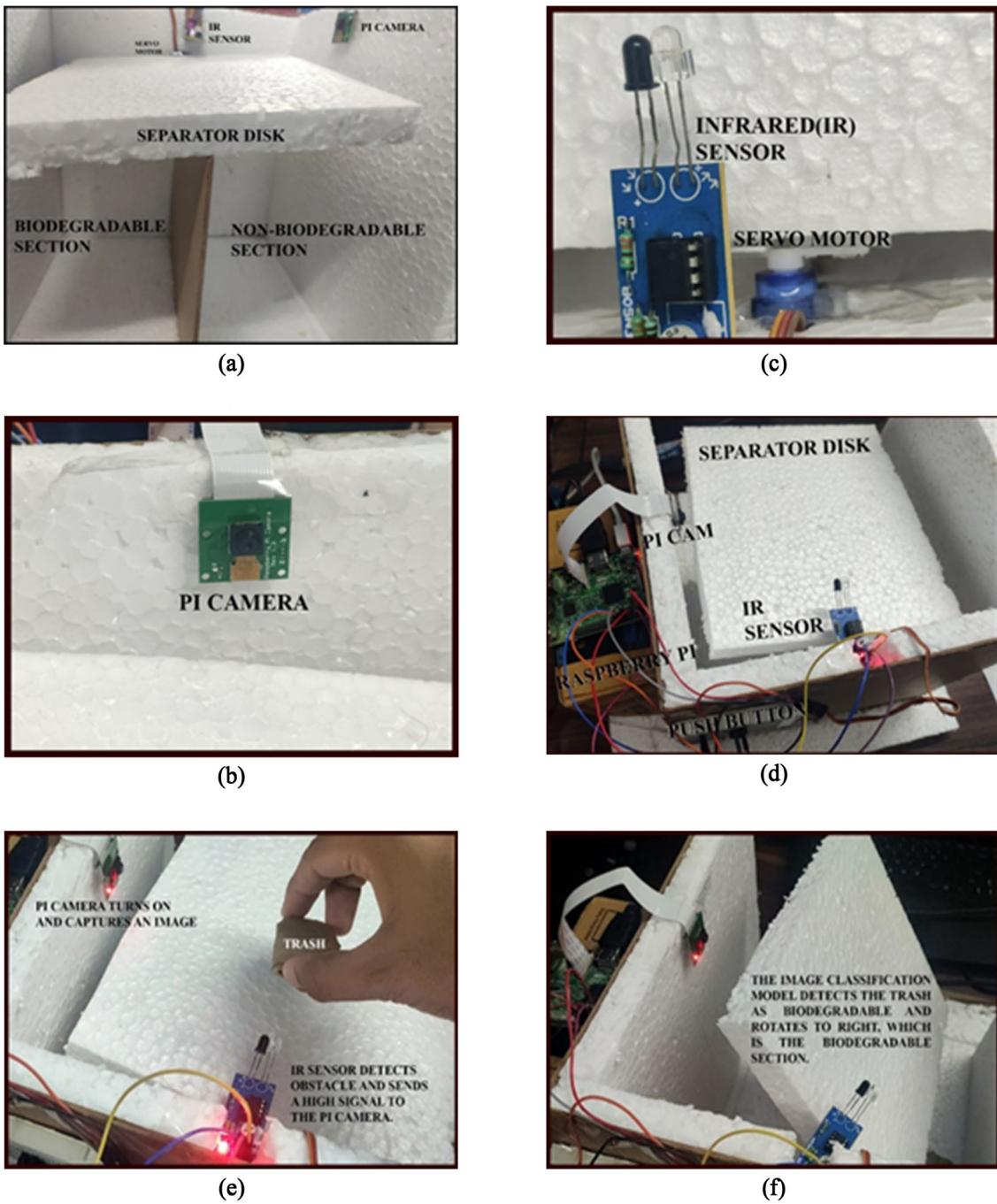


Fig. 18 Implemented hardware system (a–d) Hardware components, (e) IR sensor and (f) System response to the output of classification algorithm

which a model was evaluated. As seen in the below graphs of AlexNet and InceptionV3, the accuracy for the validation set trends towards the accuracy observed in the training set, as well as the loss minimizes and tends towards the values observed in the training set, indicating that robust and efficient parameter learning aimed at low variance and low bias for both AlexNet and InceptionV3.

$$\text{Accuracy} : \frac{\text{True Positive} + \text{False Negative}}{\text{Total number of samples}} \quad (7)$$

$$\text{Loss} : -(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (8)$$

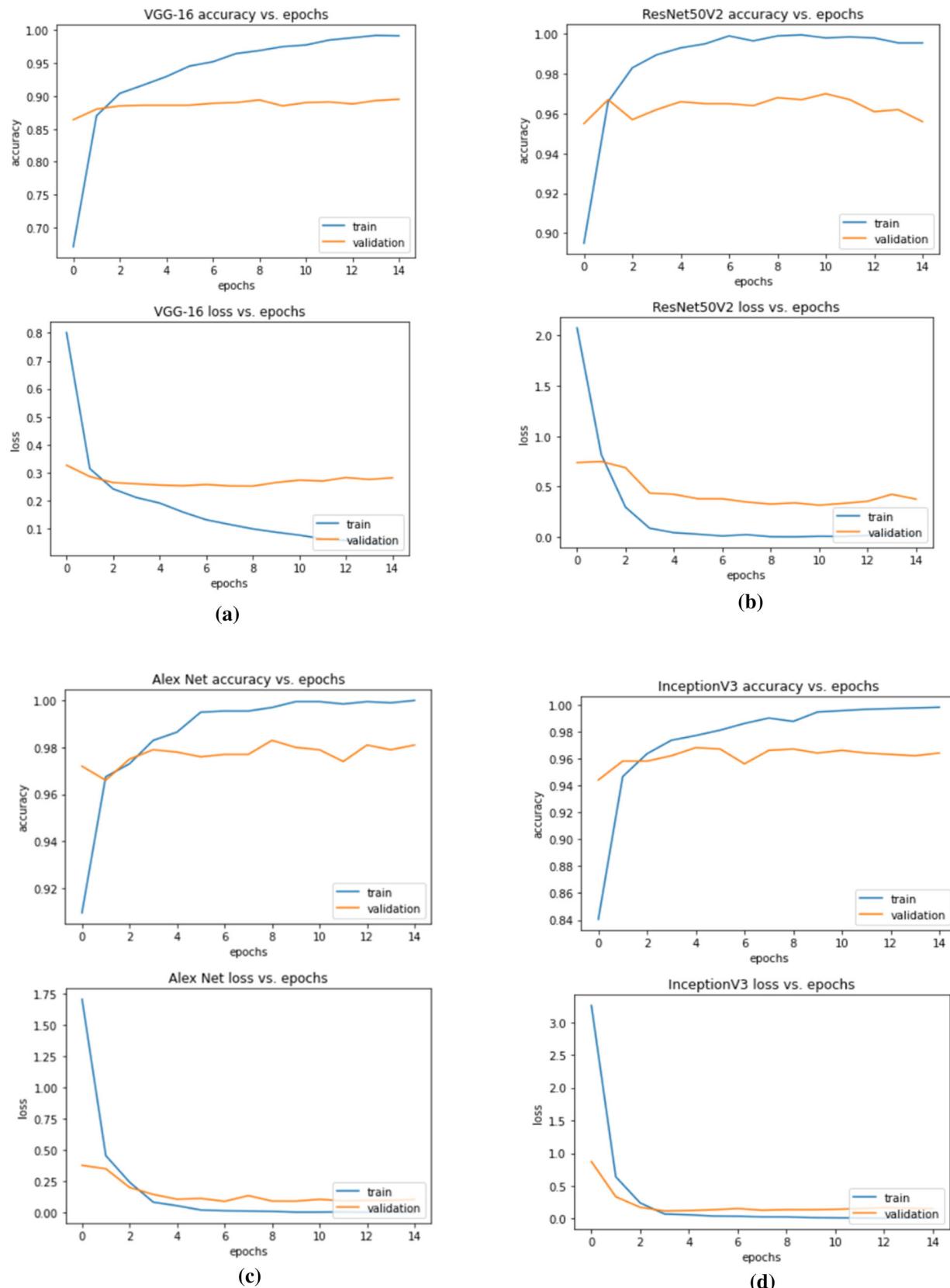


Fig. 19 Output graphs accuracy vs epoch and loss vs epoch graphs of (a) VGG-16, (b) ResNet50, (c) AlexNet and (d) Inception Net V3

Table 5 Results of different models used

Model	Accuracy-training set	Accuracy-validation set	Loss-training set	Loss-validation set
AlexNet	98.27%	97.95%	0.10	0.22
ResNet	97.94%	97.21%	0.09	0.48
InceptionNet V3	98.15%	96.23%	0.10	0.13
VGG16	98.76%	87.52%	0.09	0.32

Table 6 Comparison of current state-of-art

Existing work	Description	Algorithms	Accuracy
Spotgarbage: smartphone app to detect garbage using deep learning (2016) [21]	It proposes a smartphone application that detects pools of garbage in images clicked by users that are geo-tagged	CNN	87.69%
Image-Based Trash Classification using Machine Learning Algorithms for Recyclability Status (2018) [12]	It proposes a model that classifies the garbage image into five categories, namely paper, glass, plastic, cardboard, and metal	CNN	89.81%
Detection of Waste Containers Using Computer Vision (2019) [23]	The model classifies garbage containers	VLAD, CNN	90%
Multilayer hybrid deep-learning method for waste classification and recycling (2018) [11]	The model proposed segregates the garbage into two recyclable categories and others using a camera and other sensors	MHS	91.6–98.2%
Automatic Garbage detection and collection (2019) [20]	The system uses real-time video analysis to detect garbage and calculate the distance of garbage from the robotic arm that picks up the garbage	CNN, Raspberry Pi	90%
Automatic Garbage Volume Estimation Using Sift Features Through Deep Neural Networks And Poisson Surface Reconstruction. (2018) [31]	The image is segmented, reconstructed, and then the volume of garbage in the bin is calculated	SIFT, Poisson surface reconstruction	NA
In this paper	The proposed hardware simulated by proposed software classifies waste into two categories, Biodegradable and Nonbiodegradable, and uses hardware solutions to segregate them into different partitions of the same bin	CNN (Inception Net), Raspberry Pi	96.23–98.15%

As seen in Table 5, the accuracy calculated on the training set of all the models is around 98%. Still, the accuracy calculated on the validation set of VGG-16 is significantly lower compared to the other models.

As can be seen in Fig. 20 AlexNet and Resnet both perform highly well on the validation set compared to InceptionNet V3 but are outperformed by InceptionNet V3 when calculating the ‘Loss’ on the Validation set, where their average ‘Loss’ values are significantly higher than that of their training counterpart.

Furthermore as shown in the graph in Fig. 20 the prediction time in case of InceptionNet V3 is the lowest, hence

for deployment on a raspberry pi model with a RAM of 1 GB(900 MHz), InceptionNet V3 would be the optimum choice compared to the other four models as low prediction time model is required in the proposed architecture of the smart bin.

The proposed system worked well when a high-resolution image of the object held against a white background was available. When the image of the object was dull/distorted (held against a darker background) or a noisy image of the object (where the object was broken and mixed with other objects), it was able to classify the object correctly. However, on inspection of the values obtained from the softmax layer,

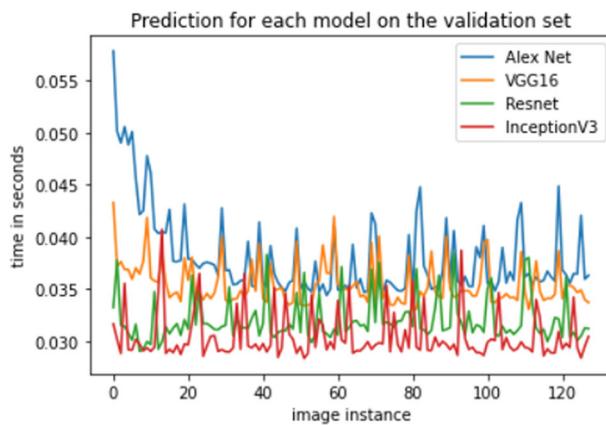


Fig. 20 Prediction time vs image instances graph for each model

it was found that the values were much more distorted and distributed among different classes rather than being majorly concentrated in one class, as observed previously for high-resolution, low noise images.

The test objects were classified under two labels that are biodegradable and nonbiodegradable. The first test was done with a biodegradable cardboard piece crushed to form a ball. The PiCam captured the image shown in Fig. 21 (a). The early detection was not accurate, and there could have been the following reasons: The cardboard piece comprised a lesser area of the image, whereas other background detailing was more evident. Thermocol and background noise could have led to the wrong classification. The image was also shot at a lower ISO level, and the shutter speed of the camera was for less duration. Hence, the quality of the image captured was not as good and detailed. The object in question was blurred as it was close to the camera, and the camera could not focus on it quickly. This led to the decision to give the camera sleep time before actually capturing the picture. This was done to get a better result and get the camera to focus on the object which needs to be classified.

To improve the accuracy of the model, a hardware model was modified, and a plain sheet of paper was introduced opposite to the camera to deal with background noise. Low-quality images could also be a hindrance. Hence, ISO speed was increased to 640 from 240, and the shutter speed was set to 1/15, which is a slight change to the exposure value but effective in distinguishing features in a detailed manner. The images captured were of resolution 1920×1080 pixels with a focal aperture as $f/2.9$. These settings improved the accuracy of the model immensely. Testing was done on different objects found in a household, which comprises the major waste content located at the source. Objects like a keychain in Fig. 21b, metallic pen in Fig. 21c, a newspaper in Fig. 21d, a plastic wrapper in Fig. 21e, a potato in Fig. 21f, a glass jar in Fig. 21g and a comb in Fig. 21h were successfully clas-

sified into their respective categories and segregated by the smart bin.

Many tests were run on different household objects, which were successfully classified. However, there was a unique case; an earbud couldn't be classified as Non-Biodegradable by the smart bin. Tests were run at different angles, and earbuds were classified as Biodegradable, as shown in Fig. 21i. This could only be overcome by training the inception net model with more images and preferably more instances of different household objects.

Hence, there is a vast potential in the *Smart Bin* model as it can be trained on more images, which can help it make more accurate decisions for complex cases.

A brief comparison of current state-of-art and the proposed model

Table 6 gives a comparison of the current state-of-art with the proposed model. The model based on MHS [11] classifies the waste into two classes, recyclable and other trash, yet there is no provision for physical segregation of garbage. On the other hand, the model proposed in this paper classifies the garbage into two categories, biodegradable and nonbiodegradable, and also segregates them physically using a smart bin. The accuracy of the proposed model varies between 96.23–98.15%, which is even better than the former model mentioned.

Conclusion

Garbage detection and disposal are some of the significant issues faced by India today. Heaps of garbage are pooled on the roadside resulting in an unhealthy environment for people around that area. Many efforts have been made to resolve this issue and automate the procedure of garbage detection and segregation. Multiple applications and software are developed for the detection of pooled garbage on the roadside with the help of image classification (images being taken by the general public). Softwares for garbage segregation (segregating biodegradable from non-biodegradable) are also developed. To alert authorities about overflowing bins, many hardware implementations are also there, which will analyze the volume of garbage in the bin or the level of garbage in the bin. So far, there are various technologies that are being used for automating the task of garbage detection or segregation like CNN, SVM, YOLOv2, etc. Although the techniques implemented so far are instrumental, yet there are few shortcomings in the systems (whether it is the efficiency of the system or the need for a controlled environment). There are various types of smart bin modules which are being proposed worldwide, such as volumetric estimation, detect filling levels, automate trash can lids with the help of IR sensors, etc. A single module can be made to accomplish all such functions with proper resource optimization. The speed at which



Fig. 21 Output of the system algorithm

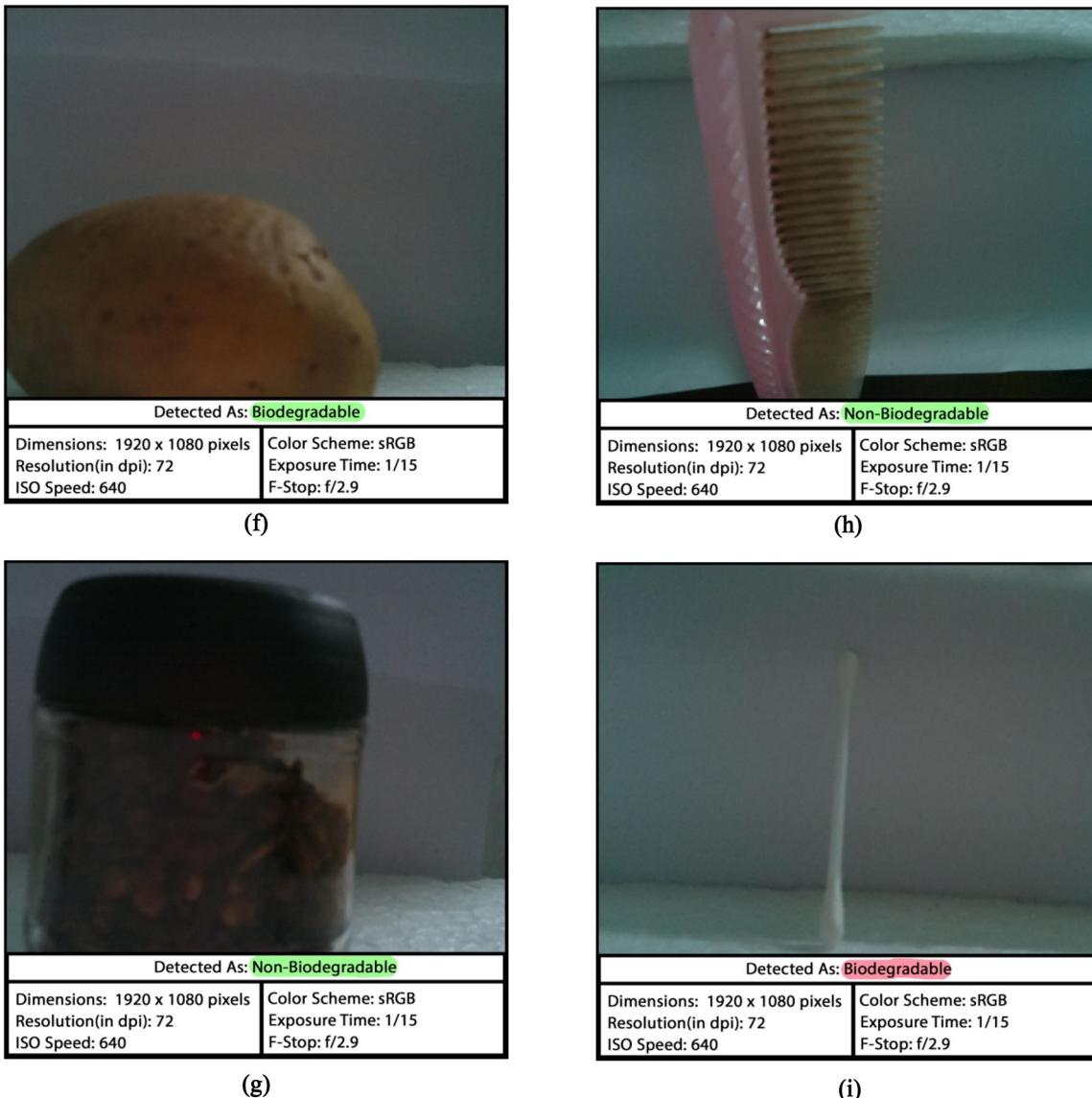


Fig. 21 continued

the detection takes place and waste is segregated can be improved. For testing purposes, the system was built with adequate delays to maintain accuracy; hence, time gaps can be reduced by reducing such delays.

Funding This work is self-funded.

Data availability <https://github.com/TanyaGupta23/Smartbin>.

Code availability <https://github.com/TanyaGupta23/Smartbin>.

Declarations

Conflicts of interest The authors declare that they have no conflicts of interest to report regarding the present study

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aziz F, Arof H, Mokhtar N, Mubin M, Talip MSA (2015) Rotation invariant bin detection and solid waste level classification. Elsevier Meas 65:19–28
- Bhatia A (2019) Fighting India's garbage crisis: what are the challenges India faces in managing its waste. *New Delhi Television Limited*. Retrieved from <https://swachhindia.ndtv.com/>
- Swaminathan M (2018) How can India's waste problem see a systemic change? *EPW Engage*. Retrieved from <https://www.epw.in/engage/article/>
- Chahine K, Ghazal B (2021) Automatic sorting of solid wastes using sensor fusion. Int J Eng Technol 9(6):4408–4414
- Gundupalli SP, Hait S, Thakur A (2017) Multi-material classification of dry recyclables from municipal solid waste based on thermal imaging. Waste Manag 70:13–21
- Wang Y, Zhang X (2018) Autonomous garbage detection for intelligent urban management. In MATEC Web of Conferences (Vol. 232, p. 01056). EDP Sciences
- (2019) SAIL launches smart garbage bins made of steel. *The Hindu*. Retrieved from <https://www.thehindubusinessline.com/companies/>
- Gutierrez JM, Jensen M, Henius M, Riaz T (2015) Smart waste collection system based on location intelligence. Proc Comput Sci 61:120–127
- Bircanoglu C, Atay M, Beşer F, Genç Ö, Kızırkı MA (2018) RecycleNet: intelligent waste sorting using deep neural networks. In 2018 Innovations in Intelligent Systems and Applications (INISTA) (pp. 1–7). IEEE
- Liu Y, Ge Z, Lv G, Wang S (2018) Research on automatic garbage detection system based on deep learning and narrowband internet of things. In Journal of Physics: Conference Series (Vol. 1069, No. 1, p. 012032). IOP Publishing.
- Chu Y, Huang C, Xie X, Tan B, Kamal S, Xiong X (2018) Multilayer hybrid deep-learning method for waste classification and recycling. Comput Intell Neurosci
- Satvikar M (2018) Image-based trash classification using machine learning algorithms for recyclability status (Doctoral dissertation, Dublin, National College of Ireland).
- Arebey M, Hannan MA, Basri H, Begum RA (2012) Bin level detection using gray level co-occurrence matrix in solid waste collection. In Proceedings of the World Congress on Engineering and Computer Science, Papers (Vol. 2).
- Hannan MA, Arebey M, Begum RA, Basri H, Al Mamun MA (2016) Content-based image retrieval system for solid waste bin level detection and performance evaluation. Waste Manag 50:10–19
- Yang M, Thung G (2016) Classification of trash for recyclability status. CS229 Project Report, 2016.
- Wang Z, Peng B, Huang Y, Sun G (2019) Classification for plastic bottles recycling based on image recognition. Waste Manag 88:170–181
- Aljarrah I, Mohammad D (2018) Video content analysis using convolutional neural networks. In 2018 9th International Conference on Information and Communication Systems (ICICS) (pp. 122–126). IEEE
- Sandeep M, Chaware SD, Akshay J, Namrata B, Rohini K (2017) Smart garbage monitoring system using internet of things (IoT). Int J Innov Res Electr Electron Instrum Control Eng 5(1):201
- Yun K, Kwon Y, Oh S, Moon J, Park J (2019) Vision-based garbage dumping action detection for real-world surveillance platform. ETRI J 41(4):494–505
- Bansal S, Patel S, Shah I, Patel P, Makwana P, Thakker D (2019) AGDC: automatic garbage detection and collection. arXiv preprint arXiv: 1908.05849.
- Mittal G, Yagnik KB, Garg M, Krishnan NC (2016) Spotgarbage: smartphone app to detect garbage using deep learning. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 940–945).
- Gao Q, Lim S, Jia X (2018) Hyperspectral image classification using convolutional neural networks and multiple feature learning. Remote Sens 10(2):299
- Valente M, Silva H, Caldeira JM, Soares VN, Gaspar PD (2019) Detection of waste containers using computer vision. Appl Syst Innov 2(1):11
- Hiremath T, Rajarajeswari S (2019) A survey on existing convolutional neural networks and waste management techniques and an approach to solve waste classification problem using neural networks
- Salmador A, Cid JP, Novelle IR (2008) Intelligent garbage classifier. IJIMAI 1(1):31–36
- Zulpe N, Pawar V (2012) GLCM textural features for brain tumor classification. Int J Comput Sci Issues (IJCSI) 9(3):354
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv: 1409.1556*.
- Singh S, Mehta KS, Bhattacharya N, Prasad J, Lakshmi SK, Subramaniam KV, Sitaram D (2017) Identifying uncollected garbage in urban areas using crowdsourcing and machine learning. In 2017 IEEE Region 10 Symposium (TENSYMP) (pp. 1–5). IEEE
- Girshick R, Donahue J, Darrell T, Malik J (2015) Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans Pattern Anal Mach Intell 38(1):142–158
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097–1105).
- Nilopherjan N, Piriyadharisini G, Rajmohan R, Sandhya SG (2018) Automatic garbage volume estimation using sift features through deep neural networks and poisson surface reconstruction. Int J Pure Appl Math 119(14):1101–1107
- Romero A, Gatta C, Camps-Valls G (2015) Unsupervised deep feature extraction for remote sensing image classification. IEEE Trans Geosci Remote Sens 54(3):1349–1362
- Bai J, Lian S, Liu Z, Wang K, Liu D (2018) Deep learning based robot for automatically picking up garbage on the grass. IEEE Trans Consum Electron 64(3):382–389
- Zhihong C, Hebin Z, Yanbo W, Binyan L, Yu L (2017) A vision-based robotic grasping system using deep learning for garbage sorting. In 2017 36th Chinese Control Conference (CCC) (pp. 11223–11226). IEEE
- Shah PJ, Anagnostopoulos T, Zaslavsky A, Behdad S (2018) A stochastic optimization framework for planning of waste collection and value recovery operations in smart and sustainable cities. 78, Waste management.
- Al Mamun MA, Hannan MA, Hussain A, Basri H (2016) Theoretical model and implementation of a real-time intelligent bin status monitoring system using rule-based decision algorithms. 48, Expert Systems with Applications.
- Marques P, Manfroi D, Deitos E, Cegoni J, Castilhos R, Rochol J, Kunst R (2019) An IoT-based smart cities infrastructure architecture applied to a waste management scenario. 87, Ad Hoc Networks.
- Github | “Trashnet waste repository”. <https://github.com/garythung/trashnet>. Accessed on 24 May 2020
- Sekar S. Waste classification data. Kaggle, 16 June 2019, www.kaggle.com/techsash/waste-classification-data. Accessed on 24 May 2020
- Serezhkin A. Drinking waste classification. Kaggle, 10 May, 2020, www.kaggle.com/arkadiyhacks/drinking-waste-classification. Accessed on 24 May 2020

41. Rawat W, Wang Z (2017) Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput* 29(9):2352–2449
42. Wu Z, Shen C, Van Den Hengel A (2019) Wider or deeper: revisiting the resnet model for visual recognition. *Pattern Recogn* 90:119–133
43. Sengupta A, Ye Y, Wang R, Liu C, Roy K (2019) Going deeper in spiking neural networks: Vgg and residual architectures. *Front Neurosci* 13:2

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