

## RESEARCH ARTICLE

# Garbage Content Estimation Using Internet of Things and Machine Learning

EUNICE LIKOTIKO<sup>1</sup>, YUKI MATSUDA<sup>1</sup>, (Member, IEEE),  
AND KEIICHI YASUMOTO<sup>1</sup>, (Member, IEEE)

Nara Institute of Science and Technology, Nara 630-0192, Japan

Corresponding author: Eunice Likotiko (likotiko.eunice\_david.ld4@is.naist.jp)

This work was supported by the Japan Science and Technology Agency (JST), Precursory Research for Embryonic Science and Technology (PRESTO), under Grant JPMJPR2039.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Review Board at the Nara Institute of Science and Technology under Approval No. 2019-1-9-2.

**ABSTRACT** Much garbage is produced daily in homes due to living activities, including cooking and eating. The garbage must be adequately managed for human well-being and environmental protection. Although the existing IoT-based smart garbage systems have gained high garbage classification accuracy, they still have a problem that they provide a small number of garbage categories, not enough for reasonable practices of household garbage separation. This study presents a new smart garbage bin system, SGBS, embedded with multiple sensors to solve the problem. We deployed temperature, humidity, and gas sensors to know the condition and identify the garbage content disposed of. Then, we introduce a new garbage content estimation method by training a machine learning model using daily collected fuse sensor readings combined with detailed household garbage contents annotations to perform garbage classification tasks. For evaluation, we deployed the designed SGBS in five households over one month. As a result, we confirmed that the leave-one-house cross-validation results showed an accuracy of 91% in 5 kitchen waste contents, also, 89% in 5 paper/softbox contents, and 85% in the 8 garbage categories for the classification tasks.

**INDEX TERMS** IoT-based smart garbage system, garbage content estimation, machine learning algorithms.

## I. INTRODUCTION

Much garbage is produced daily in homes due to living activities, including cooking and eating. Therefore, garbage must be adequately managed for human well-being and environmental protection. In the standard municipal garbage management system, households are responsible for sorting and managing garbage produced in their home. However, it is hard to depend solely on public awareness to provide the correct garbage management at the source. Therefore, an automation tool that can reflect the home's daily life and understand households' routine behaviour of garbage disposal would be necessary to influence behaviour change on garbage disposal and increase home monitoring for the case of elderly anomaly detection and healthy living. Furthermore,

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi<sup>1</sup>.

it would improve garbage management services through proper garbage separation practices for the well-being of people and the environment.

It is reported that the world generates 2.01 billion tonnes of municipal solid waste annually, with at least 33% of that not managed environmentally safely [1]. In fact, daily waste generated per person ranges widely, from 0.11 to 4.54 kilograms [2]. Furthermore, only 17% of electronic garbage is collected and recycled [3]. Moreover, 32% of plastic packages still need to be managed, which leads to severe implications for ecological balance and human well-being. But, again, garbage separation by the person who disposes of garbage has been widely accepted as ethical behaviour and best practice for reducing, reusing, and recycling [4]. Several existing IoT-based smart garbage systems and the classification methods using computer vision and artificial intelligence have been developed to improve household garbage

management [5], [6], [7], [8]. However, the existing systems have the following problems: first, they can not learn the amount of garbage disposed of each time; second, they provide a small number of garbage categories, not enough for reasonable practices of household garbage separation; and third, they can not understand the routine behaviour of garbage disposal by households.

In our previous study [9], we addressed the first problem by proposing a smart garbage bin system with ToF and weight sensors and the ARIMA model based garbage growth prediction method. In this paper, we focus on solving the second and the third problems, we propose a newly designed and developed smart garbage bin system (SGBS) embedded with multiple sensors to identify the garbage contents disposed of. The SGBS architecture comprised two subsystems.

The first subsystem is the smart garbage bin (SGB), embedded with DHT22 (temperature and humidity) and MQ135 gas sensors to know the conditions and identify the disposed garbage content since garbage contents have different shapes and moisture. Therefore, the type of garbage content affects the humidity and air quality found in the smart bin. Also, the SGB is embedded with ToF (time of flight) and load cell sensors to detect the new garbage content disposed of each time. Then, data are updated and stored in the cloud via a Wi-Fi gateway.

The second subsystem is a garbage annotation mobile application (GAA). The GAA interface consists of 8 garbage categories and 25 garbage content identities, providing an easy way for household users to annotate garbage content they dispose of daily using a handy smartphone.

We conducted experiments where the SGBS was deployed in five houses of heterogeneous characteristics to examine the impact. As a result, the household user daily uses the installed smart garbage bin system and annotates their garbage contents, which they dispose of in smart garbage bins. Therefore, information about identified garbage and produced amounts were continuously monitored and collected in the garbage log for each household. To perform garbage classification tasks, we introduce a new garbage content estimation method by training a machine learning model using daily collected fuse sensor readings combined with detailed household garbage contents annotations. As a result, we confirmed that the leave-one-house-out cross-validation results showed an accuracy of 91% in 5 kitchen waste contents, also, 89% in 5 paper/softbox contents, and 85% in the 8 garbage categories for the classification tasks. In summary, the contributions of this work are:

- 1) Identification of garbage content and understanding household garbage disposal behaviour for influencing family's behaviour change in the garbage disposal and increase home monitoring.
- 2) The provision of more satisfactory garbage content categories for the reasonable practice of separating garbage in the household.
- 3) Providing and discussing a new garbage content estimation model based on daily garbage contents disposed

of in households, built with data-efficiency machines learning classifiers with satisfactory relative accuracy.

The remainder of this paper is as follows: Section II provides an overview of related work from the recent work on garbage classification using the image and Deep learning models also Municipal garbage separation rules. Section III describes the materials and tools used in the study, including systems design and development details. Section IV presents the experiment, data collection and pre-processing data procedures. Section V introduces the garbage content estimation model and the step by steps process of building the model using a machine learning algorithms. Finally, Section VI discusses results from the classification tasks and compares our approach with literature works, whereas, in Section VII, we conclude our paper.

## II. RELATED WORK

This section gives an overview of related work from two different perspectives. First, we provide an overview of the separation and disposal of garbage with an emphasis on municipals in Japan, where this study was conducted. Secondly, we discuss recent work on garbage classification from images using deep learning to recall existing approaches to assess it. Thirdly we briefly discuss our preliminary study.

### A. SEPARATION AND DISPOSAL OF GARBAGE IN JAPAN

Garbage separation has been a major challenge across developing countries than in developed countries where there are various collection systems for house-separated garbage, such as in Sweden and Germany [10], China [11], and Japan [12]. While in other developed countries, garbage separation is often classified into three categories: recyclable, household, and vegetation garbage. In Japan, the garbage separation and disposal system is different and complex. The rules for the separation and disposal of garbage depend on the particular local municipality, whereby each city in Japan provides a well-documented pamphlet explaining the garbage disposal rules. In general, garbage is divided into four categories: Burnable garbage (Kitchen waste, paper scraps, clothing, etc.), non-burnable garbage (Metal, glass, ceramics and pottery, etc.), recyclable (Plastic bottles, container jars, cans, newspapers, etc.), and oversized (Large furniture, etc.) [12]. Therefore, each municipality uses such a general garbage division to classify garbage for their residents. Table 1 provides an overview of the division of burnable garbage content in four cities in Japan: Kashihara [13], Ikoma [14], Nara [15] and Kyoto [16]. Apart from garbage descriptions from the municipal pamphlets, residents use designated plastic garbage bags of up to 45 litres to dispose of garbage. Moreover, garbage collection for each category of garbage is set by the municipal for instance, Mondays and Thursdays in Ikoma city [14] are used for the collection of burnable garbage only. The above facts show that families in Japan play a hand role in their municipal rules for garbage separation and disposal systems. However, the failure of households

**TABLE 1. Overview of burnable garbage separation in Japan.**

| City Name           | Burnable garbage separation   |
|---------------------|---|
| Kashihara city [13] | 1:Kitchen scraps<br>2:Small plastic e.g DVDs/CDs,toys<br>3:Waste paper (tissue, mixed papers)<br>4:Weeds, twigs, and leave (30 cm)<br>5:Containers e.g for mayonnaise, oil<br>6:Polystyrene<br>7:Shoes,Bag, Cloth |
| Ikoma city [14]     | 1:Kitchen waste<br>2:Plastic/unclean products e.g storage containers<br>3:Paper/textile e.g Tissues,Milk box, cloth<br>4:Others   |
| Nara city [15]      | 1:Kitchen garbage<br>2:Styrofoam e.g noodle cup<br>3:Cassette tape/videotape (less than five)<br>4:Waste Woodchips e.g chopsticks, pencils<br>5:Waste Paper   |
| Kyoto city [16]     | 1:Raw garbage<br>2:Non-recyclable paper<br>3:Broken glasses<br>4:Small plastics   |

to sort the garbage renders the whole system useless [7]. Therefore, automation tools are necessary to monitor daily family garbage disposal and improve garbage separation and management.

## B. GARBAGE CLASSIFICATION FROM IMAGES WITH DEEP LEARNING MODELS

A possible solution to overcome the existing challenges in household garbage separation and management is to adopt sustainable automation tools to improve garbage separation. Presently, several works have been devoted to the automation and detection of garbage from images, which has now become a popular choice to replace manual garbage separation while taking advantage of the rapid advances in computer vision and artificial intelligence. Various standard CNN architectures have been recently proposed to perform image classification tasks with high accuracies, such as VGGNet [17], AlexNet [18], ResNet [19] and DenseNet [20].

Nnamoko et al. [5] investigated the problem of manual household garbage separation into two categories, namely, organic and recyclable. Experiments presented in this paper were conducted with Sekar's waste classification image dataset available in the Kaggle library [21]. Later, a bespoke 5-layer CNN architecture was used to perform image classification tasks. In this work, the training was conducted on two datasets, smaller model ( $80 \times 45$  pixels) and a larger model ( $225 \times 264$  pixels), for performance comparison, thus obtaining similar cross-validation accuracy of 79%. Likewise, Mookkaiah et al. [22] proposed a model to identify and classify two types of garbage, biodegradable and non-biodegradable. First, the images were collected in the respective garbage bin by Raspberry Pi Camera Module v2. Then garbage classification task was done by CNN architecture. However, separating garbage into two categories is insufficient for logical household garbage separation.

Besides, there is still a shortage of publicly available garbage image datasets and an information gap in their experimental procedures.

Furthermore, Wang et al. [7] revealed garbage sorting and classification at the source, the beginning of garbage collection while utilizing the combined method of IoT and CNN. The study used experimental data available in the Trashnet [23] dataset, merged with other datasets thus, resulted in nine categories of garbage (Kitchen waste, other waste, hazardous waste, plastic, glass, paper or cardboard, metal, fabric and other recyclable waste). In addition, the study developed an intelligent bin embedded with ultrasonic sensors, MQ9, and MQ135 gas sensors to monitor the garbage's running state in the bin. Finally, the CNN model was deployed in mobile phones and cloud computing servers for garbage classification. The system required citizens to take pictures of garbage using their mobile phones and send them to a cloud server to run the deep-learning algorithm to recognize categories. Despite the high-performance accuracies of 92.44% and 92.00% achieved by Xception and MobileNetV3 models on classifying nine types of garbages, the author presented more generalizable garbage categories that need to be improved for proper household garbage separation.

Besides, a distributed architecture for smart recycling using machine learning was realized by Ziouzos et al. [6] as a solution for garbage classification in collection facilities to solve the problem of non-segregated garbage, which exists more in developing and developed countries. The Trashnet [23] dataset was used for training the models by utilizing computation offloading to the cloud. The CNN architecture classified the garbage materials into five categories: paper, glass, plastic, metal, carton, and trash. Similarly, Sami et al. [24] used the Trashnet [23] dataset to automate the garbage classification problem into six classes: glass, paper, metal, cardboard, and trash using a Support Vector Machine, Random Forest, Decision tree, and CNN to find the optimal algorithm that best fits garbage classification solution. However, the available public garbage image datasets need more classes of garbage categories for proper garbage classification. Therefore, the garbage categories presented in both studies [6], [24] are not practical for household garbage separation and for improving the garbage management systems.

Despite the high accuracies achieved by the existing solutions on garbage classification through the automation and detection of garbage from images by the deep learning models, they still have problems: (Problem 1) They can not learn the amount of garbage disposed of each time; (Problem 2) They provide a small number of garbage categories, not enough for reasonable practices of household garbage separation; (Problem 3) They can not understand households' routine behaviour of garbage disposal. Therefore, to the best of our knowledge, an automation tool that can learn and identify the daily garbage content disposed of in homes and perform classification tasks, as investigated throughout this work, has yet to be considered.

### C. PRELIMINARY STUDY

To solve Problem 1, we conducted a preliminary study to learn the amount of garbage disposed of each time and predict growth behaviour at a single house [9]. In this study, we designed and developed the initial smart garbage bin prototype embedded with ToF (time of flight) and load cell sensors to track the amount of garbage during disposal. Using a Wi-Fi gateway, data were sent to a cloud platform. For evaluation, we deployed the smart garbage bin in a student laboratory over one month. An autoregressive integrated moving average (ARIMA) model was applied, providing an average mean absolute error (MAE) of 5.17 cm and a standard deviation (SD) of 0.33 cm, thus was considered satisfactory accuracy for the garbage growth prediction. Therefore, our prediction model was suitable for predicting future garbage growth behaviour, enhancing flexibility in the garbage collection schedule and the frequency of changing garbage bags in the smart bin.

However, Problem 2 and Problem 3 in Section II-B remain open. Therefore, in this paper, we try to address these problems.

### III. MATERIALS AND TOOLS

This section presents the details of the system requirements necessary for designing and developing a smart garbage bin system (SGBS), tools and the procedure for selecting important garbage categories for developing garbage annotation application design.

#### A. SYSTEM REQUIREMENTS

In this subsection, we describe the system requirements for the proposed system. Based on the discussions in Section I and Section II, we find the following two requirements for a smart garbage bin system:

- 1) The smart garbage bin system should automatically collect sensor data without any additional activities by users.
- 2) The smart garbage bin system should estimate detailed garbage categories and garbage content identities corresponding to each disposal behaviour.

To address requirement (1), we designed and developed a smart garbage bin system which is always connected to the internet, uploads all sensor data to the cloud to store them. To address requirement (2), we built a new machine learning model for estimating garbage categories and garbage content identities with high accuracy.

#### B. ARCHITECTURE DESIGN

Fig. 1 demonstrates a designed and developed SGBS architecture to revolutionize the existing household garbage management system by tracking daily household garbage disposal information and identifying the type of garbage contents disposed of at the source. The smart garbage bin system architecture consists of two subsystems: the smart garbage bin (SGB), embedded with distance and weight sensors to

TABLE 2. Sensor used in development of smart garbage bin.

| Sensor              | Purpose  |
|---------------------|--|
| 1. AE-VL53L0X(ToF)  | Measure fill level of garbage in a bin                   |
| 2. DHT22            | Measure temperature and humidity in a bin                |
| 3. HX711-load cell  | Measure increase of weight of garbage in a bin           |
| 4. MQ135 gas sensor | Measure CO <sub>2</sub> ,NH <sub>3</sub> ,Smoke in a bin |

detect the timestamp of newly disposed of garbage content during garbage disposal. On the other hand, the smart garbage bin (SGB) is embedded with temperature, humidity, and gas sensors to identify and distinguish disposed of garbage contents. Secondly, SGBS architecture comprises the garbage annotation mobile application (GAA) with a smooth interface that allows users to annotate their daily disposal of garbage content during garbage disposal. The two subsystems (SGB and GAA) later create a daily garbage log data for each house. Moreover, the designed architecture comprises the analysis part that uses machine learning algorithms to classify garbage contents found in the house logs. The outcome of the analysis produces a garbage content estimator for each home which helps identify and classify garbage content at the source.

#### C. SMART GARBAGE BIN

Fig. 2 shows the overview of a designed and developed smart garbage bin system (SGBS). Considering the significant roles of the proposed SGBS architecture described in Section III-B, a set of lightweight, low-cost, high-precision IoT sensors were chosen and embedded in the smart garbage bin (SGB). The selected devices have different hardware configurations and purposes. In our SGB prototype, we used a DHT22 (temperature and humidity) and MQ135 gas sensors to monitor the moisture and air quality of the disposed garbage content in the smart garbage bin. Furthermore, we used a ToF (time of flight) and HX711-load cell to track the garbage filling level and weight at each time of disposal. Using a Wi-Fi gateway, the smart garbage bin system is always connected to the internet, uploads all sensor data to the cloud, and stores them. In addition, the Secure Digital non-volatile flash memory card format (SD), connected to an I2C real-time clock with 32.768 kHz frequency (DS3231 RTC) module data are also collected and stored in the SD-created file in one-minute intervals daily. On the other hand, the SGB comprises the 2 × 16 character LCD Module with a blue backlight, which uses an I2C interface to communicate with the host Arduino Mega 2560 microcontroller Rev3. Therefore, the LCD module displays the garbage's current filling level and temperature data of the smart bin. The proposed smart garbage bin prototype allows easy tracking of garbage amount information at the source. Table 2 provides the purpose of the chosen sensors used to develop the smart garbage bin.

#### D. GARBAGE ANNOTATION APPLICATION

To provide a smooth and easy way for households to annotate garbage content they dispose of daily. We further present

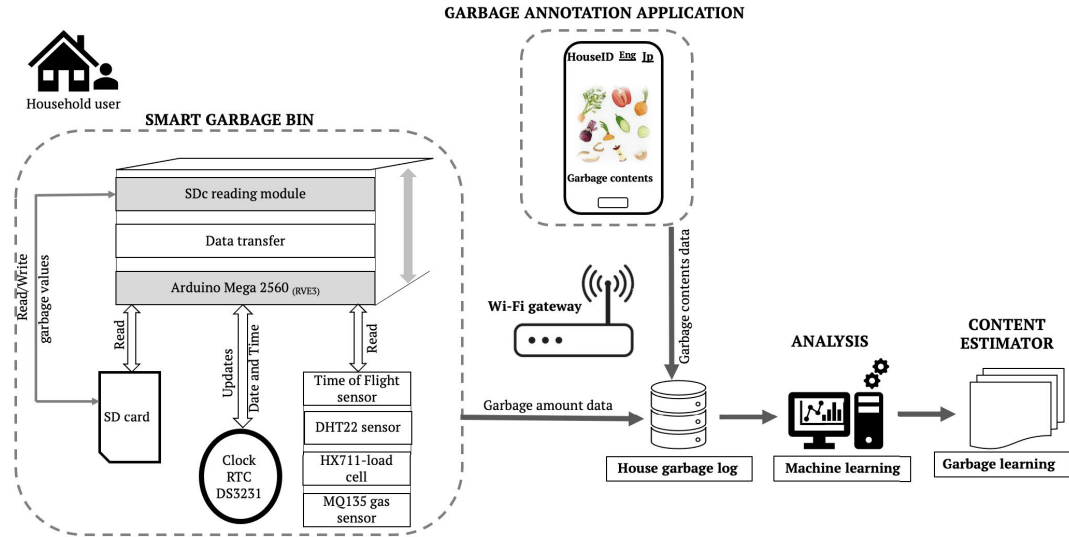


FIGURE 1. Smart garbage bin system architecture design.

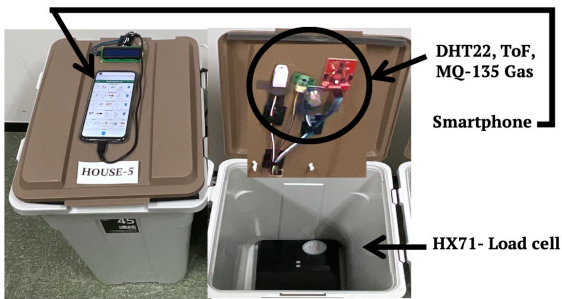


FIGURE 2. Smart garbage bin system overview.

a garbage annotation mobile application (GAA). The GAA designed and installed in a handy smartphone made a significant value consideration to household users by allowing annotation in a more efficient and tailored way through a smooth interface. The selection of the garbage categories in our proposed study is based on the rules for separating and disposing of burnable garbage as provided in four random selected municipal’s pamphlets in Japan that explain the garbage disposal rules described in Section II-A, including the city of Kashihara [13], Ikoma [14], Nara [15], Kyoto [16]. Additionally, we conducted a short survey with fifteen (15) students living in the city of Ikoma and Nara for one week. The survey participants were asked to annotate their daily burnable garbage disposal on paper. The annotation included the name of the garbage contents and the frequency of disposing of such garbage. Thus, by analyzing the survey results and the rules for disposing of the garbage from municipal pamphlets, we established important categories of burnable garbage with specific content identities for the mobile annotations application. The garbage annotations application interface comprises the garbage categories and a menu with two languages, English and Japanese, giving users flexibility to

switch between the languages. Also, the interface consists of house numbers as an identification for the experimental data collection.

Fig. 3 demonstrates the garbage annotation application interface whereby vertically depicts 8 garbage categories (i.e., Kitchen waste, Meal garbage, Paper/softbox, Fabric/textile, Plastic, Dust, Plant, and All others) and horizontally depicts 25 garbage contents identities (i.e., Food garbage, Edible food, Sink basin, Kitchen waste bag, Unclean cup, Unclean container, Unclean packages, Waste wood, Tissues, Mixed Papers, Milk/Juice box, Masks, Clothes, Shoe, bag, Rubber products, Disposable diapers, Plastic product, Toys, CD, Cigarette ashes/stick, Vacuum cleaner, Plant and Others) belonging to each category. The garbage annotation application provides a guide knowledge that allows individual households to smoothly select the type of garbage content each time they dispose of garbage in the SGB from a handy smartphone fixed outside on top of the SGB cover. Then, data about the garbage category and its specific identity content are sent to the cloud data server using a Wi-Fi network.

IV. DEPLOYMENT AND DATA COLLECTION EXPERIMENT

Herein we present the experimental setup and data collection, including datasets, the data preprocessing steps undertaken to build the garbage contents estimation model, and the methods adopted to address the study aims. This study was approved by the Ethical Review Committee for Research Involving Human Subjects at the Nara Institute of Science and Technology (Approval No.: 2020-I-16).

A. EXPERIMENT AND PARTICIPANT INFORMATION

We conducted the evaluation experiment from June to August 2022 in five households of heterogeneous characteristics in the city of Nara, Ikoma, and Kyoto in Japan for 3-5 weeks. We considered family size, type of family, age group, number

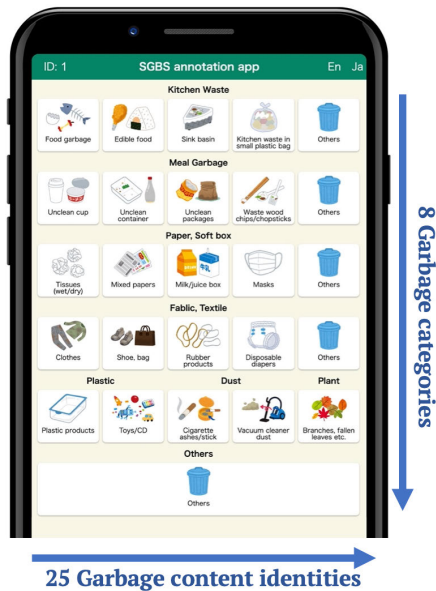


FIGURE 3. Garbage annotation application interface.

TABLE 3. Information of participants.

| House ID | Family size | Family type   | Child/infant | Age group | City  |
|----------|-------------|---------------|--------------|-----------|-------|
| 1        | 2           | Couple        | 0            | 50-55     | Kyoto |
| 2        | 4           | Couple        | 2            | 7-50      | Nara  |
| 3        | 2           | Couple        | 0            | 28-29     | Ikoma |
| 4        | 2           | Single-shared | 0            | 22-23     | Ikoma |
| 5        | 3           | Couple        | 1            | 27-29     | Ikoma |

of children, and city as the criteria for selecting participants for the experiment. Table 3 outlines the participant’s information. All participants were well informed about the experiment and provided their own consent to participate in the experiment. In addition, smart garbage bins were distributed and installed in each house. Fig. 2 shows the overview of the deployed SGBS.

**B. DATASETS**

The experiment resulted in five garbage logs data from the five households. The garbage log consists of data from the SGB (i.e., timestamp, filling level, weight, temperature, humidity, and air quality), collected every one-minute interval. Also, data from the GAA (i.e., timestamp, garbage categories, and content identities) collected only when a user disposes of and annotates the garbage in a smart garbage bin. The frequency of garbage disposal and annotation of garbage contents differ in each household due to household characteristics. Table 4 details the full annotations of garbage contents found in houses 1 to 5 by the household users during the experiment. Therefore, we define the following rules to merge the multiple sensor data from the smart garbage bin (as features) and garbage content annotations by the households (as labels) to create a single dataset of each house.

TABLE 4. Garbage annotation frequency found in house 1 to 5.

| Category ID | Category name  | House 1 | House 2 | House 3 | House 4 | House 5 |
|-------------|----------------|---------|---------|---------|---------|---------|
| 1           | Kitchen waste  | 78      | 37      | 104     | 23      | 152     |
| 2           | Meal garbage   | 66      | 63      | 125     | 11      | 135     |
| 3           | Paper/softbox  | 374     | 200     | 183     | 61      | 53      |
| 4           | Fabric/textile | 21      | 16      | 21      | 8       | 77      |
| 5           | Plastic        | 5       | 4       | 14      | 1       | 9       |
| 6           | Dust           | 50      | 11      | 3       | 1       | 6       |
| 7           | Plant          | 19      | 9       | 13      | 0       | 1       |
| 8           | All other      | 74      | 24      | 78      | 18      | 17      |
|             | Total          | 687     | 364     | 541     | 121     | 449     |

We considered a time stamp of 10-minute intervals from the disposal time recorded by the annotation application to calculate features for the particular label. The features include maximum, minimum, and rate of change of the garbage filling level, weight, temperature, humidity, and air quality. At the same time, the label consists of 8 garbage categories and 25 garbage identities. Thus, we obtained the total original datasets of each house for both garbage categories and content identities. Below are the rules used to merge the collected data;

- 1) Every 10 minutes, if a new garbage label is input, and then calculate new features for the label.
- 2) If at the same time or in less than 10 minutes, another new label is input, then use the previously calculated features for the new label (Overlap features).

**C. CLASS IMBALANCE**

A lower frequency of disposing of a particular type of garbage content than the others experienced in all houses leads to a minority of such garbage content. Therefore, the minority class labels affect the model-building process, i.e., a model that always chooses the majority class regardless of the corresponding feature. To solve this, we utilize the resampling technique to enhance the classifier model’s size and quality and avoid biases class during training. There are two main approaches for random resampling: Oversampling, which duplicates the minority class, and Undersampling, which deletes the majority class. In our case, due to the low number of annotations in garbage category 4 (Fabric/textile), garbage category 5 (Plastic), garbage category 6 (Dust), and garbage category 7 (Plant) experience in all five houses (see Table 4), we applied the Oversampling technique to increase the minority class using the imbalanced-learn sci-kit-learn library. Table 5 and Table 6 show the total number of datasets of garbage categories and content identities before and after resampling.

**V. GARBAGE CONTENT ESTIMATION MODEL**

This study aims to identify garbage contents disposed of and perform the garbage classification from garbage contents disposed of daily in the household by adopting IoT and data-efficient machine learning algorithms. Therefore we present a garbage content estimation model to classify

**TABLE 5. Re-sampling and cross-validation split for the 8 garbage categories.**

| House ID | Before re-sampling |            | After re-sampling |            |
|----------|--------------------|------------|-------------------|------------|
|          | Original dataset   | 25% splits | New dataset       | 25% splits |
| 1        | 682                | 170-171    | 2618              | 654-655    |
| 2        | 360                | 90         | 1400              | 350        |
| 3        | 538                | 134-135    | 1281              | 320-321    |
| 4        | 121                | 30-31      | 305               | 76-77      |
| 5        | 449                | 112-113    | 1064              | 266        |

**TABLE 6. Re-sampling and cross-validation split for the 25 garbage content identities.**

| House ID | Before re-sampling |            | After re-sampling |            |
|----------|--------------------|------------|-------------------|------------|
|          | Original dataset   | 25% splits | New dataset       | 25% splits |
| 1        | 687                | 170-172    | 1930              | 482-483    |
| 2        | 364                | 91         | 570               | 142-143    |
| 3        | 541                | 134-136    | 1295              | 323-324    |
| 4        | 121                | 30-31      | 360               | 90         |
| 5        | 450                | 112-113    | 1165              | 291-292    |

8 categories of garbage and a total of 25 garbage contents identities relating to a particular category, as demonstrated in Fig. 3 of the garbage annotation application. The subsequent section details the process of building classification models.

### A. MODEL BUILDING

Fig. 4 demonstrates model building steps and order of operations. Below we give a details explanation of the importance of each model-building step. We performed the classification tasks from daily collected fuse sensor readings combined with detailed household garbage contents annotations intending to find the class (i.e., 8 garbage categories: *Kitchen waste, Meal garbage, Paper/softbox, Fabric/textile, Plastic, Dust, Plant, and All others*) and (i.e., 25 garbage content identities: *Food garbage, Edible food, Sink basin, Kitchen waste bag, Unclean cup, Unclean container, Unclean packages, Waste wood, Tissues, Mixed Papers, Milk/Juice box, Masks, Clothes, Shoe, bag, Rubber products, Disposable diapers, Plastic product, Toys, CD, Cigarette ashes/stick, Vacuum cleaner, Plant and Others*) to which a new unseen observation belongs. During the model-building steps in Fig. 4, we only consider utilizing data-efficient methods, namely: Random forest, Naive Bayes, Extreme Gradient Boosting (Xgboost), and Decision tree algorithms to build the garbage content estimation model, for the reasons such as the comparison of the machine learning classifiers, the small number of available datasets, the popularity of the classifier and data preprocessing to avoid minority class labels. We eventually defined the order of operations applied to the selected classifiers during the model-building steps.

More precisely, we train and test by splitting the dataset of each house into four (4) chunks of 25% equal size dataset as shown in the Table 5 and Table 6 for garbage categories and content identities. To avoid overfitting as much as possible,

first, we utilize repeated k-fold cross-validation to evaluate the machine learning models in steps 1 and step 2 (see Fig. 4). Then, we averaged the results with 4-fold cross-validations to compute the final validation score for each investigated model configuration. Therefore, the model created in step 1 used the original (unbalanced) datasets, i.e., before resampling (see Table 5). While the model developed in step 2 used the balanced class dataset, i.e., after resampling (see Table 6), as discussed in Section IV-C. Thus, for performance comparison of balanced and unbalanced datasets, our model-building process output two models, an unbalanced model and a balanced model (see Fig. 4).

Afterwards, for better comparison reasons of the cross-validation methods applied to the classifiers, and, in order to increase the training set, in step 3 (see Fig. 4), we changed the cross-validation method to leave one house out cross-validation method where we repeatedly trained our models with total balanced datasets from the four houses and testing the model with the remaining one house. Thus, we obtained the Leave one house out model.

Furthermore, we built the overall result models in step 4 (see Fig. 4) of the classification tasks for both class garbage categories and content identities for each house to investigate the overall performance of the classifiers. We first made the overall result model on all 8 garbage categories, i.e. *Kitchen waste, Meal garbage, Paper/softbox, Fabric/textile, Plastic, Dust, Plant, and All others* found in House 1, House 2, House 3, House 4 and House 5. Nonetheless, because each garbage category comprises 5 to 2 specific garbage content identities (see Fig. 3), in total, there are 25 different garbage content identities belonging to the eight categories expected to be annotated by the users daily using the garbage annotations application. Therefore because of the majority number of garbage content identities and differences in frequency behaviour of garbage disposal and annotation exhibited from each house (see Table 4). In this study, we first selected the five garbage content identities from the Kitchen waste (category 1) as it has had a higher frequency of annotation in house 3, house 4 and house 5. Also, we chose the five garbage content identities from the paper/softbox (category 3) as it has had a higher frequency of annotation in house 1 and house 2 to learn the performance of the classifiers on garbage content identities. Therefore, to this point of the study, we created three overall result models for garbage content estimation, namely;

- 1) Overall result model for general garbage categories
- 2) Overall result model for kitchen waste contents identities
- 3) Overall result model for paper, softbox contents identities

### B. PERFORMANCE EVALUATION

Our model evaluation performance is based on accuracy, which is the percentage of correct comparison classifications. Moreover, we evaluate the performance of our models

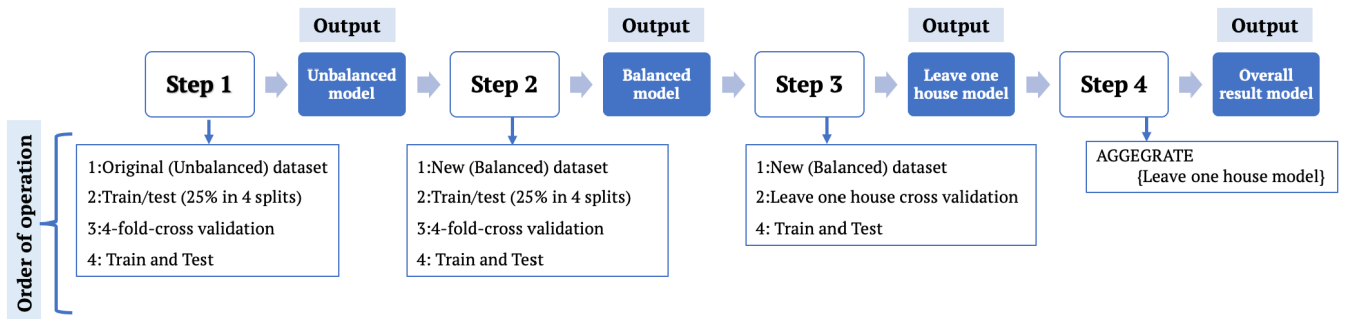


FIGURE 4. Model building steps and order of operations.

TABLE 7. 4-fold cross-validation performance accuracy for the 8 garbage categories.

| House ID | Accuracy (%) of unbalanced model |             |         |               | Accuracy (%) of balanced model |             |         |               |
|----------|----------------------------------|-------------|---------|---------------|--------------------------------|-------------|---------|---------------|
|          | Random forest                    | Naive bayes | Xgboost | Decision tree | Random forest                  | Naive bayes | Xgboost | Decision tree |
| 1        | <b>90</b>                        | 89          | 85      | 67            | 65                             | 71          | 68      | 67            |
| 2        | 85                               | 85          | 74      | 71            | 73                             | 72          | 73      | 63            |
| 3        | <b>86</b>                        | 72          | 85      | 82            | 86                             | 79          | 79      | 72            |
| 4        | 80                               | 79          | 73      | 77            | 75                             | 76          | 78      | 63            |
| 5        | 87                               | 86          | 83      | 81            | 80                             | 78          | 62      | 69            |

TABLE 8. Leave one house cross-validation performance accuracy for the 8 garbage categories.

| House ID | Accuracy (%) of Leave one house model |             |         |               |
|----------|---------------------------------------|-------------|---------|---------------|
|          | Random forest                         | Naive bayes | Xgboost | Decision tree |
| 1        | 83                                    | 83          | 79      | 57            |
| 2        | 84                                    | 84          | 84      | 69            |
| 3        | <b>88</b>                             | 81          | 81      | 80            |
| 4        | 80                                    | 80          | 78      | 72            |
| 5        | 84                                    | 78          | 78      | 69            |

using other metrics, such as Confusion matrices, Precision, Recall and F1-score. We will especially give the most informative metrics for the overall result models because they aggregated the garbage class label results from all houses belonging to the same classification and averaged the result into a single metric measurement. Furthermore, the model parameters tuning was applied on all classifiers, Random forest, Naive Bayes, Extreme Gradient Boosting (Xgboost), and Decision tree. As a result, the accuracy slightly increased by increasing the number of parameters such as estimators, criterion, and random state for each model separately. Therefore, we independently investigated the model performance on all experimental datasets found in House 1, House 2, House 3, House 4, and House 5 on garbage categories and garbage content identities classification tasks. The percentage performance accuracy results using 4-fold cross-validation and leave-one-house-out cross-validation as applied to the four machine learning classifiers for the 8 garbage categories and 25 garbage identities are summarized in Table 7, Table 8, Table 9, and Table 10.

### C. RESULTS

Throughout this subsection, we describe results obtained from the classification tasks as detailed in Section V-B. Specifically, we look into and compare the performance accuracy from the unbalanced, balanced, leave one house out, and overall result models using the four machine learning classifiers.

#### 1) UNBALANCED MODEL

We see from the results of the unbalanced model (see Table 7) and (see Table 9) using the 4-fold cross-validations that Random forest performs slightly better than other classifiers (Naive Bayes, Xgboost, and Decision tree), for classification tasks of both garbage categories and garbage content identities. For garbage categories, the highest accuracy was 90% obtained in house 1, and the 67% lowest accuracy resulted from the Decision tree in the same house. Also, 93% for garbage content identities was the highest accuracy found in house 1 by Random forest, and the lowest accuracy was 80% by the Decision tree found in house 4.

#### 2) BALANCED MODEL

Afterwards, we compared the four classifiers with the same 4-fold cross-validations method in all five houses on a balanced dataset with the approaches discussed in Section IV-C to deal with the unequal class balance. The results can be seen in Table 7 and Table 9. We observed that the performance accuracy slightly decreased compared with the unbalanced model performance. Yet, Random forest manifested the highest accuracy and thus outperformed the rest of the classifiers. For the garbage categories, the Random forest exhibited 86%



**TABLE 9.** 4-fold cross-validation performance accuracy for the 25 garbage content identities.

| House ID | Accuracy (%) of unbalanced model |             |         |               | Accuracy (%) of balanced model |             |         |               |
|----------|----------------------------------|-------------|---------|---------------|--------------------------------|-------------|---------|---------------|
|          | Random forest                    | Naive bayes | Xgboost | Decision tree | Random forest                  | Naive bayes | Xgboost | Decision tree |
| 1        | <b>93</b>                        | 89          | 86      | 86            | <b>88</b>                      | 79          | 63      | 72            |
| 2        | 89                               | 87          | 84      | 85            | <b>88</b>                      | 81          | 68      | 76            |
| 3        | 87                               | 87          | 87      | 82            | 83                             | 85          | 68      | 76            |
| 4        | 83                               | 84          | 80      | 80            | 79                             | 73          | 71      | 72            |
| 5        | 86                               | 85          | 82      | 83            | 75                             | 78          | 65      | 62            |

**TABLE 10.** Leave one house cross-validation performance accuracy for the 25 garbage content identities.

| House ID | Accuracy (%) of Leave one house model |             |         |               |
|----------|---------------------------------------|-------------|---------|---------------|
|          | Random forest                         | Naive bayes | Xgboost | Decision tree |
| 1        | <b>91</b>                             | 89          | 88      | 74            |
| 2        | 90                                    | 87          | 84      | 78            |
| 3        | 89                                    | 89          | 89      | 84            |
| 4        | 86                                    | 86          | 86      | 81            |
| 5        | 88                                    | 85          | 85      | 65            |

in house 3, and 63% by the Decision tree in house 2 was the lowest accuracy. While for garbage content identities, the accuracy was 88% by Random forest from house 1 and house 2, and the most insufficient accuracy was 62% by a decision tree in house 5.

3) LEAVE ONE HOUSE OUT MODEL

In the next step, we compare the results of the repeated 4-fold cross-validation in step 2 to the Leave one house out (LoH) cross-validation approaches in step 3 (see Fig. 4). In order to investigate the classification performance in all five houses. Therefore, we applied the LoH on the balanced class datasets using the four classifiers in step 3. However, we maintained the same order of operation as in step 2. With this approach, the sum of four houses increases the size of the training set during repeated testing with only one house dataset. The results for Random forest, Naive Bayes, XGBoost, and Decision tree in the case of the garbage categories and garbage content identities for all four classifier sets are shown in Table 8 and Table 10. We see an apparent accuracy increase in each house compared to the balanced model of 4-fold cross-validation in Table 7 and Table 9. For the garbage categories, the Random forest revealed the highest accuracy of 88% in house 3, while the decision tree showed the lowest accuracy of 57% in house 1. In addition, garbage content identities in the leave one house out model achieved the highest accuracy of 91% and 90% by Random forest in house 1 and house 2, respectively. On the other hand, the Decision tree exhibited unsatisfactory performance, 65% in house 5. Moreover, Random forest again steadily outperformed the rest of the classifiers.

4) OVERALL RESULT MODEL

To realize the performance of the three overall result models described in Section V-A above Overall result model

of garbage categories, (2) Overall result model of kitchen waste contents identities and (3) Overall result model of Paper/softbox contents identities. The performance accuracy results for the three models are shown in Table 11. Moreover, we compared the Recall, Precision, and F1-score for the overall result models as they can better judge the performance by showing the metric measurements of each class label.

For the garbage categories overall result model (see Table 11), Random forest achieved the highest accuracy of 85%, followed by Naive Bayes at 82% and Xgboost at 80%, while the decision tree lags with the least accuracy of 64%. Table 12 summarises the metric accuracies of the 8 garbage categories overall result model with Recall, Precision, and F1-score using the Random forest classifier.

Further, for the overall result model of kitchen waste contents identities (see Table 11) (*i.e., food garbage, edible food, sink basin, kitchen waste bag, and others*). The Random forest has steadily revealed the best classification accuracy of 91%, while the accuracies of the rest of the models are: 88% Naive Bayes, 84% Xgboost and 76% Decision tree. Likewise, the overall result model of the paper/softbox contents identities (see Table 11) (*i.e., tissues, mixed papers, milk/juice box, masks, and others*) are 85% Naive Bayes, 83% Xgboost and 71% Decision tree were outperformed by the Random forest at 89%. The summary of the Recall, Precision, and F1-score for the overall result models of the 5 kitchen waste and the 5 paper/softbox content identities are shown in Table 13 and Table 14, using the Random forest as it has been portrayed as the best classifier.

The aggregated confusion matrix plots using the Random forest of each overall result model are shown in Fig. 6, where the columns represent the actual values (Truth) of the target class label. The rows represent the predicted values (Predicted) of the target variable class label. The number of validation samples that were correctly classified are demonstrated in the diagonal cells, and that were incorrectly classified are demonstrated in the off-diagonal cells.

In addition, to investigate the impact of the collected multiple sensor readings on the garbage content estimation model, we applied the features importance method using the Random forest classifier as our chosen classifier for the garbage content estimation model. The results in Fig. 5 show that air quality, humidity, temperature, and fill level values are more relevant features for identifying garbage content in the smart bin. Therefore, the

TABLE 11. Accuracy performance of the three overall result models.

| Overall result model of: | Accuracy(%) of Overall result models |             |         |               |
|--------------------------|--------------------------------------|-------------|---------|---------------|
|                          | Random forest                        | Naive bayes | Xgboost | Decision tree |
| Kitchen waste            | 91                                   | 88          | 84      | 76            |
| Paper/softbox            | 89                                   | 85          | 83      | 71            |
| Garbage category         | 85                                   | 82          | 80      | 64            |

TABLE 12. Summary of 8 garbage categories overall result model.

| Name           | Precision | Recall | F1-Score |
|----------------|-----------|--------|----------|
| Kitchen waste  | 0.84      | 0.85   | 0.85     |
| Meal garbage   | 0.79      | 0.81   | 0.80     |
| Paper/softbox  | 0.69      | 0.89   | 0.78     |
| Fabric/textile | 0.96      | 0.88   | 0.92     |
| Plastic        | 0.99      | 0.88   | 0.93     |
| Dust           | 0.90      | 0.88   | 0.89     |
| Plant          | 0.95      | 0.78   | 0.86     |
| All other      | 0.87      | 0.84   | 0.86     |

TABLE 13. Summary of 5 Kitchen waste contents identities overall result model.

| Name         | Precision | Recall | F1-Score |
|--------------|-----------|--------|----------|
| Food garbage | 0.82      | 0.90   | 0.86     |
| Edible food  | 0.93      | 0.91   | 0.92     |
| Sink basin   | 0.92      | 0.86   | 0.89     |
| Kitchen bag  | 0.94      | 0.94   | 0.94     |
| Others       | 0.97      | 0.96   | 0.96     |

TABLE 14. Summary of 5 Paper/softbox contents identities overall result model.

| Name           | Precision | Recall | F1-Score |
|----------------|-----------|--------|----------|
| Tissues        | 0.81      | 0.92   | 0.86     |
| Mixed paper    | 0.91      | 0.91   | 0.91     |
| Milk juice box | 0.91      | 0.86   | 0.89     |
| Masks          | 0.89      | 0.84   | 0.86     |
| Others         | 0.95      | 0.92   | 0.93     |

identified garbage content disposed of daily and annotation procedures contributes to the garbage classification tasks. Furthermore, the cross-validation approaches provided satisfactory results, especially for the leave-one-house-out cross-validation, which performed better than the 4-fold cross-validation.

## VI. DISCUSSION

Throughout this section, we discuss our findings and possible implications. Due to the sufficient classification outcomes, we chose the Random forest algorithm as the best classifier. We also decided on the overall result models as the final model for our garbage content estimation tasks. Generally,

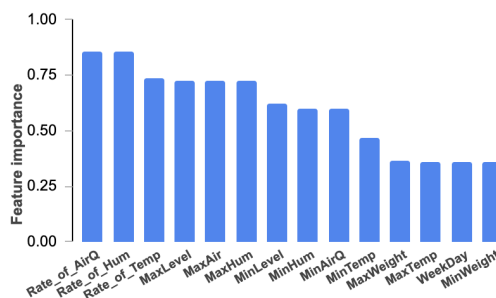


FIGURE 5. Features importance on multiple sensor readings.

the highest accuracy is between 85% and 91%, and the lowest is 64%, which is satisfactory for garbage content classification tasks. However, the lowest amount of annotation on certain class (imbalance) labels makes the classification task difficult. We start the detailed discussion by comparing garbage annotations from each house and then classification tasks by the machine learning algorithms, followed by the usefulness of the garbage content estimation model. Finally, we look at the comparison of our approach to the literature.

### A. COMPARISON OF HOUSEHOLD GARBAGE DISPOSAL ANNOTATION AND CLASSIFICATION

In general, we observed different behaviour of garbage disposal in all five houses, which is due to the heterogeneity behaviour in each family, such as living style, size of the family, type of the family, number of children/infants, age group, and city. In this case, the study observed differences in the routine frequency of garbage disposal and the type of garbage content disposed among the houses. Therefore, using the smooth garbage annotation interface (see Fig. 3) that allowed household users to annotate garbage contents during disposal, the study found that certain garbage contents were important in some houses, i.e., daily disposed and annotated, compared to others. Table 4 shows the annotation frequency of garbage category disposal among houses, as briefly detailed below.

- **House 1:** as shown in Table 3, this house consists of a married couple in Kyoto prefecture. In this house, garbage category 3 (Paper/softbox) was the most important category compared to other categories annotated 374 times during the experiment (see Table 4). In comparison, garbage category 5, which consisted of plastic contents, appeared as the least important annotated only

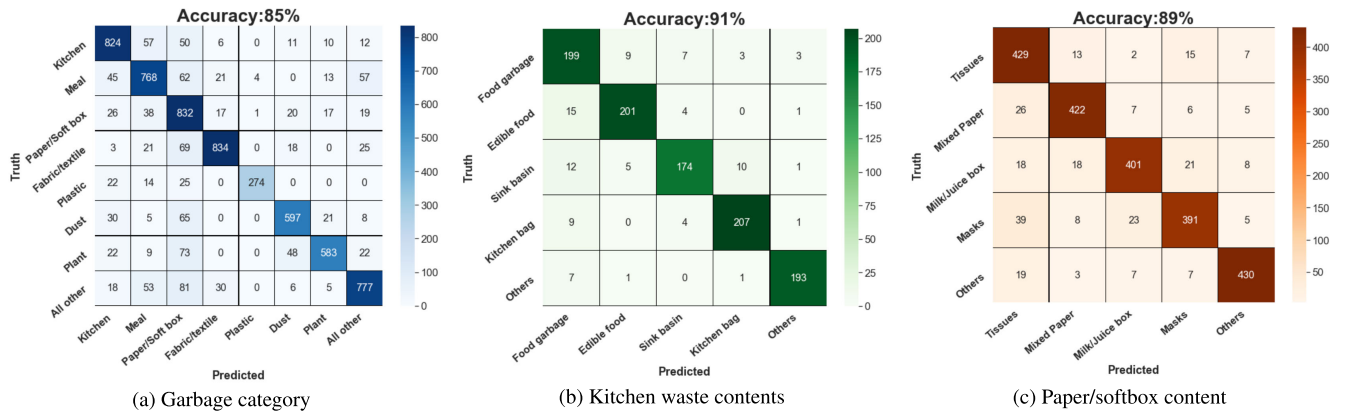


FIGURE 6. Confusion matrices of the three overall result models.

5 times. In addition, other categories had almost a similar frequency of annotation, such as Kitchen waste (78), Meal garbage (66), All others (74), and Dust (50). On the other hand, fabric/textile had 21 annotations, while the plant had 19 annotations.

- **House 2:** consists of a married couple with two children living in Nara city (see Table 3). Like in house 1 (see Table 4), garbage category 3 (Paper/softbox) was the most important category in this house, annotated 200 during the experiment, and Category 5 (Plastic) was the least annotated, only 4 times. Compared with other categories, Kitchen waste had 37 annotations, Meal garbage 63, All others 24, Fabric/textile had 16, dust 11, and Plant 9. House 2 had fewer annotations than house 1.
- **House 3:** as shown in Table 3, this house comprises a young married couple in Ikoma city. Even though garbage category 3 (Paper/softbox) is steady as the most important and Plastic as the minor category observed in houses 1 and house 2, in this house, the study observed a slight difference in annotation frequency exhibited among Kitchen waste, Meal garbage, and Paper/softbox categories. The result in Table 4 shows that the annotations frequency kept, such as Paper/softbox (183), was the most important, followed by Meal garbage (125), and Kitchen waste (104) was the third in the garbage category importance ranking.
- **House 4:** While Houses 1, 2, 3, and 5 comprise married couples, house 4 consists of two singles living in a shared house in Ikoma city (see Table 3). The study observed less annotations frequency in this house than in other houses. However, similar to houses 1, 2, and 3, garbage category 3 (Paper/softbox) had the highest annotation frequency and ranked as the most important, while the plastic was minor. Therefore, the annotation frequency in Table 4 is as follows: Paper/softbox had 61 annotations, followed by Kitchen waste (23) and Meal garbage (11), which similarly ranks with house 3. In addition, not only Plastic was the minor but also dust which was

annotated only once each. Moreover, category 7 (Dust) was not annotated in this house.

- **House 5:** This house comprises a young married couple with an infant in Ikoma city (see Table 3). Contrary to all other houses, the study observed a fewer annotation frequency of garbage category 3 (Paper/softbox), which prevailed in houses 1, 2, 3, and 4 as the most important garbage category (see Table 4). Instead, kitchen waste was the most important category in this house, with 152 annotations, followed by Meal garbage (135) and Fabric/textile (77) third in the ranks. The high annotation frequency of category 4 (Fabric/textile) was due to the disposal frequency of disposable diapers (the fourth garbage content in the Fabric/textile category 4 see Fig. 3) thus increasing the number of fabric/textile. On the other hand, Plant category 7 was annotated only once and therefore appeared as a minor category, similar to house 3. Plastic had 9 annotations, and dust had 6 annotations.

Eventually, daily disposed garbage contents and detailed garbage annotation frequency by households impacted the classification tasks in each house. For instance, in Random forests, the chosen classifier for this study (see Table 7) and (see Table 9), the accuracies for classification tasks of both garbage category and content identities in house 1 were higher than in house 4, which had fewer annotations frequencies. Moreover, the study found that the Decision tree was the insufficient classifier model compared to Random forest, Naive Bayes, Xgboost applied on the datasets in all five houses. Over and above that, the leave-one-house-out cross-validation method showed better performance compared to the 4-fold cross-validation approach despite its computational cost (see Table 8 and Table 10). Therefore, in the overall result models, we aggregated the classification result of the same class label into one metric performance using the leave-one-house-out approach, which has manifested better performance than 4-fold cross-validation on the balanced model. The following section compares our approaches with the literature.

## B. COMPARISON WITH LITERATURE

As discussed in the Section II, similar approaches in other domains/applications were investigated. As detailed below, we compare our strategies and experimental setups with those more similar to ours.

- **Suitable practice for house garbage separation**  
Our study has considered the identification of daily disposed of garbage content and provided a satisfactory garbage category suitable for burnable garbage separation practice for most families in Japan. However, Nnamoko et al. [5] and Mookkaiah et al. [22] investigated only two kinds of garbage, i.e., Organic and recyclable, which is not enough for rational garbage separation in houses. Likewise, apart from increasing the number of classes as demonstrated by Ziouzios et al. [6] and Sami et al. [24], to find respective garbage categories such as (*kitchen waste, other waste, hazardous waste, plastic, glass, paper or cardboard, metal, fabric, and other recyclable waste*). Yet these studies provided a small number and more generalizable garbage categories, which is not the best practice for proper house garbage separation and can not fully solve the problem of profound implications for ecological balance and threat to global sustainability, development, and human well-being.
- **Use of daily garbage contents and experiment transparency**  
Our study proposed to perform garbage content estimations from the daily collected fuse sensor readings and household annotations with transparency on experiments and thus can be reproducible in the field. On the contrary, the studies by [6], [22], and [24] used publicly available garbage image datasets to improve classification tasks with less transparency information on their experimental setup. However, the publicly available image datasets are associated with problems such as resizing, resolutions, and inappropriate colour presentation, thus lowering the quality of the classification task.
- **Use of efficient data models**  
Our study applied more data-efficient methods, namely Random forest, Naive Bayes, Xgboost, and Decision tree, for the classification tasks. On the contrary, most of the previous works applied the existing standard models for the classification tasks, such as VGGNet [17], AlexNet [18], ResNet [19], and DenseNet [20]. A common issue associated with image classification using the existing standard model is high computational cost which often results in high development time and prediction model size because they are often pre-trained for more than one purpose [5]. In addition, CNN-based models are difficult to run on embedded systems suitable for garbage bins, and their architecture requires large amounts of data for training which is yet to be available.

## C. STUDY LIMITATIONS

- **Few numbers of annotation**  
Our study provided sufficient burnable garbage identification to guide house users during garbage disposal through the mobile application interface. Yet, few annotations were recorded on some garbage categories because of the difference in garbage disposal behaviour exhibited in each house. For instance, the low number of plastic, dust, and plant categories in houses 1, 2, 3, and 4 (see Table 4), therefore, were removed during model building as they were affecting the performance accuracy. For this reason, more garbage annotation is required for additional training data to ensure a robust garbage estimation in application scenarios.
- **Learn correct annotation**  
Even though the study identified the frequency of annotations for each category in every house, households need to learn and remember to correctly annotate garbage in a category and contents, which can further improve the garbage classification tasks.

## VII. CONCLUSION

This study presented a new smart garbage bin system (SGBS) embedded with multiple sensors to identify the disposed garbage content categories by households. First, we designed and developed a smart garbage bin system (SGBS) architecture comprised of the smart garbage bin (SGB) equipped with temperature, humidity, gas, ToF, and load cell sensors. Then, we developed the garbage annotation mobile application (GAA) consisting of a smooth interface of 8 garbage categories and 25 content identities to allow users to annotate garbage contents during garbage disposal. Finally, we introduce a new garbage content estimation method by training a machine learning model using daily collected fuse sensor readings combined with detailed household garbage contents annotations to perform garbage classification tasks. We deployed the designed SGBS in five households over one month and applied the leave-one-house-out cross-validation to the model trained and tested with the collected data. As a result, our proposed method achieved an accuracy of 91% in 5 kitchen waste contents, 89% in 5 paper/softbox contents, and 85% in 8 garbage categories for the classification tasks. Moreover, our results show that air quality, humidity, temperature, and fill level values are more relevant features in the garbage content estimation model.

The proposed SGBS contributes to household garbage identification and classification to ensure that valuable materials are recycled and utilized.

Our future work includes extending our design to an event-based detection system to understand household garbage disposal behaviour. Also, expansion of the experiment to more families and experiment with other types of garbage, such as non-burnable garbage (Metal, glass, ceramics, pottery, etc.) and recyclable (Plastic bottles, container jars, cans, newspapers).

## REFERENCES

- [1] *What a Waste 2.0*. Accessed: Nov. 5, 2022. [Online]. Available: <https://datatopics.worldbank.org/what-a-waste/>
- [2] Ellen MacArthur Foundation. *The New Plastics Economy: Rethinking the Future of Plastics and Catalysing Action*. Accessed: Nov. 5, 2022. [Online]. Available: <https://ellenmacarthurfoundation.org/the-new-plastics-economy-rethinking-the-future-of-plastics-and-catalysing>
- [3] *The Global E-Waste Monitor 2020: Quantities, Flows and the Circular Economy Potential*. Accessed: Nov. 5, 2022. [Online]. Available: <https://collections.unu.edu/view/UNU:7737>
- [4] B. Wang, M. Farooque, R. Y. Zhong, A. Zhang, and Y. Liu, "Internet of Things (IoT)-enabled accountability in source separation of household waste for a circular economy in China," *J. Cleaner Prod.*, vol. 300, Jun. 2021, Art. no. 126773.
- [5] N. Nnamoko, J. Barrowclough, and J. Procter, "Solid waste image classification using deep convolutional neural network," *Infrastructures*, vol. 7, no. 4, p. 47, Mar. 2022.
- [6] D. Ziouzos, D. Tsiktiris, N. Baras, and M. Dasygenis, "A distributed architecture for smart recycling using machine learning," *Future Internet*, vol. 12, no. 9, p. 141, Aug. 2020.
- [7] C. Wang, J. Qin, C. Qu, X. Ran, C. Liu, and B. Chen, "A smart municipal waste management system based on deep-learning and Internet of Things," *Waste Manage.*, vol. 135, pp. 20–29, Nov. 2021.
- [8] M. Cubillos, "Multi-site household waste generation forecasting using a deep learning approach," *Waste Manage.*, vol. 115, pp. 8–14, Sep. 2020.
- [9] E. Likotiko, S. Misaki, Y. Matsuda, and K. Yasumoto, "SGBS: A novel smart garbage bin system for understanding household garbage disposal behaviour," in *Proc. 13th Int. Conf. Mobile Comput. Ubiquitous Netw. (ICMU)*, Nov. 2021, pp. 1–8.
- [10] *Managing Municipal Solid Waste 32 European Countries EEA Report no 2/2013ISSN 1725-9177*. Accessed: Nov. 5, 2022. [Online]. Available: <https://www.eea.europa.eu/publications/managing-municipal-solid-waste>
- [11] W. Guo, B. Xi, C. Huang, J. Li, Z. Tang, W. Li, C. Ma, and W. Wu, "Solid waste management in China: Policy and driving factors in 2004–2019," *Resour., Conservation Recycling*, vol. 173, Oct. 2021, Art. no. 105727.
- [12] *Japanese Waste Management and Recycling Industry*. Accessed: Nov. 5, 2022. [Online]. Available: <https://www.env.go.jp/en/index.html>
- [13] *Separation and Disposal of Garbage in Kashihara Separation and Disposal of Garbage in Kashihara*. Accessed: Nov. 5, 2022. [Online]. Available: <https://www.city.kashihara.nara.jp/documents/5c34c0f2f1a7f00f31b18cc1>
- [14] *Garbage Collection Schedule in Ikoma City*. Accessed: Nov. 5, 2022. [Online]. Available: <https://www.city.ikoma.lg.jp/cmsfiles/contents/0000005/5895/gomical.pdf>
- [15] *How to Divide and Take Out Garbage (for Household Garbage)*. Accessed: Nov. 5, 2022. [Online]. Available: <https://www.city.nara.lg.jp/uploaded/attachment/34246.pdf>
- [16] *Combustible Garbage*. Accessed: Nov. 5, 2022. [Online]. Available: <http://kyoto-kogomi.net/wp-content/uploads/2016/04/gg-eng.pdf>
- [17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 7, Jun. 2015, pp. 770–778.
- [20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [21] *Waste Classification Data*. Accessed: Nov. 5, 2022. [Online]. Available: <https://www.kaggle.com/datasets/techsash/waste-classification-data>
- [22] S. S. Mookkaiah, G. Thangavelu, R. Hebbar, N. Haldar, and H. Singh, "Design and development of smart Internet of Things-based solid waste management system using computer vision," *Environ. Sci. Pollut. Res.*, vol. 29, pp. 64871–64885, Apr. 2022.
- [23] *Trashnet*. Accessed: Nov. 5, 2022. [Online]. Available: <https://github.com/garythung/trashnet>
- [24] K. N. Sami, Z. M. A. Amin, and R. Hassan, "Waste management using machine learning and deep learning algorithms," *Int. J. Perceptive Cognit. Comput.*, vol. 6, no. 2, pp. 97–106, Dec. 2020.



**EUNICE LIKOTIKO** received the B.E. degree from Ardhi University, the United Republic of Tanzania, in 2012, and the M.E. degree from The Nelson Mandela African Institute of Science and Technology (NM-AIST), the United Republic of Tanzania, in 2017. She is currently pursuing the Ph.D. degree with the Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST), Nara, Japan. Her research interests include ubiquitous computing, smart home, and sensing technology.



**YUKI MATSUDA** (Member, IEEE) was born in 1993. He received the B.E. degree in advanced course of mechanical and electronic system engineering from the National Institute of Technology, Akashi College, Japan, in 2015, and the M.E. and Ph.D. degrees from the Graduate School of Information Science, Nara Institute of Science and Technology, Japan, in 2016 and 2019, respectively. Since 2019, he has been an Assistant Professor with the Ubiquitous Computing Systems Laboratory, Graduate School of Science and Technology, Nara Institute of Science and Technology. Since 2020, he has been a Researcher at Japan Science and Technology Agency, PRESTO. His current research interests include participatory sensing, location-based information systems, wearable computing, and affective computing. He is a member of IPSJ, IEICE, JSAI, and ACM.



**KEIICHI YASUMOTO** (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in information and computer sciences from Osaka University, Osaka, Japan, in 1991, 1993, and 1996, respectively. He is currently a Professor with the Graduate School of Science and Technology, Nara Institute of Science and Technology. His research interests include distributed systems, mobile computing, and ubiquitous computing. He is a member of ACM, IPSJ, SICE, and IEICE.

...