

# A probability approach in cans identification

Yulia Resti<sup>1,\*</sup>, Amrifan S.Mohruni<sup>2</sup>, Firmansyah Burlian<sup>2</sup>, Irsyadi Yani<sup>2</sup>, and Ali Amran<sup>1</sup>

<sup>1</sup>Department of Mathematics, Faculty of Mathematics and Natural Science Universitas Sriwijaya, 30162 Inderalaya, Indonesia

<sup>2</sup>Department of Mechanical Engineering Faculty of Engineering Universitas Sriwijaya, 30162 Inderalaya, Indonesia

**Abstract.** The objective of this study is to identify can waste into three types based on the images by using a probability approach of trinomial distribution in term regression. Predictor variables considered are the color intensity of red, green, and blue of the images taken at the top, down, and side pose successively. From an independence test between each of the predictor variable and can waste type noted that only the color intensity of red which the image taken at top pose that does not correspond to the can waste types. Based on the Nagelkerke value is found that the variance of the predictor variable data in identifying the can waste type is able to explain the variance of the types of 59.1 percent. The final model show that the significant predictor variables are the colors intensity of green and blue which the image taken at the top pose, the color intensity of red which the image taken at down pose, and the color intensity of red, green and blue which the image taken at side pose successively. The model can identify cans waste into three types based on the images correctly by 73.13%.

## 1 Introduction

Aluminum cans are the most widely recycled solid waste in the world. A lot of energy is taken to make them, throwing away one aluminum can wastes as much energy as pouring out half a can of gasoline, and if aluminum cans are thrown in a landfill, it takes between 100-500 years for just one aluminum can to decompose [1]. In recent years, solid waste management, especially recycling based on automatic sorting system is a major challenge for large cities in developing countries [2].

A cans is a metal container designed to hold a fixed portion of liquid such as carbonated soft drink, fruit juice, herbal tea and the like. Cans are made of aluminium (75% international manufacture) or tin-plated steel (25% international manufacture). Worldwide production for all beverage cans is approximately 52 billion units. An aluminium can is a container for packaging, made primarily of aluminum. It is commonly used for foods and beverages but also for products such as oil, chemicals, and other liquids. A common 12-ounce-size can weigh 15 grams when empty.

The identification process of cans waste is needed in an automatic sorting system of cans recycling. Several approaches have been applied to identify cans waste either the parametric such as discriminant analysis and non-parametric approach such as back propagation algorithm. The application of discriminant analysis to identify cans waste can be found in [4], while the implementation of back propagation algorithm can be found in [3].

The identification system research using automatic or semi-automatic techniques was started in the 1960s and intensified in the last two decades. With the increased

availability of powerful computer, it has received significant attention from the research community. There are currently automatic systems that perform well when the object images are captured under uniform and controlled conditions. However, these technologies have several limitations due to the fact that they are difficult to develop, the quality of the image captured, and lighting problem.

The identification system is generally based on the selection of the color space to detect the color identification methods. Object identification is a feature based approach in which face geometry is taken which includes shape and other features like size, edge, color etc. The algorithm requires 2-D images whose threshold values of intensities are taken into consideration in the measurement of the number of the pixels to get the entire object feature area.

Trinomial logistic regression is also the parametric approach which often considered in many analysis because it does not assume normality, linearity, or homoscedasticity, whereas discriminant analysis requires these assumptions. [10] applied trinomial logistic regression to classify the health level of BUMN is based on the financial ratios, [9] classify the severity of traffic accident victims based on the characteristics of victims, types of vehicles, the characteristics and time of collision, while [8] applied trinomial logistic regression in another name as polynomial logistic regression to classify the risk of motor insurance for the third party. Application of trinomial logistic regression in other fields can be found in [11], [12], and [13]. In this paper, the trinomial logistic regression is applied to identify the cans waste based on the some images into three types.

\* Corresponding author: [yulia\\_resti@mipa.unsri.ac.id](mailto:yulia_resti@mipa.unsri.ac.id)

## 2 Material and methods

### 2.1 Data collection and description

We collected three types of cans waste; 100% recyclable (type 1), not 100% recyclable (type 2), and unrecyclable (type 3). The proportion of each of cans waste types is presented in Table 1.

Correct identifying and sorting of aluminium cans from non-degradable materials in solid waste requires many facets of intelligent sorting system. First, for the item to identify itself, it should minimize the number of information carried or remembered on the automatic sorting system. For example, it must be able to distinguish between aluminium and non-aluminium cans and add a new object for automatic sorting in the future. This is to ensure that the best and optimal methods can be chosen in order to produce significant and efficient results.

**Table 1.** Proportion of cans waste for each type.

Type	% populations
1	50.00%
2	31.25%
3	18.75%

We have collected 3 pictures for each cans. In order to store the feature values, a database with two tables was created using Excel. One of the tables was used to store raw sample data images and another was used to store the template data.

An edge detection algorithm based on the Sobel function was applied to extract the contour of the object after eliminate noise of gray image. Figure 4.4 provides the edge of beverage cans after filtering the global edge.

Each of them is placed on a conveyor belt, which connect to a webcam. The images of them are taken using the webcam for three phases; top, down, and side poses. We save it in JPEG format with a resolution of 320 x 240 pixels, and then extract the color intensity of red, green and blue of each pose using a Matlab comment. Of this population, 50% is taken as a database or for learning, and the rest for testing.

The image capture contained image acquisition and image pre-processing steps. The feature extraction contained cropping and extraction steps, where the cropping consisted of template matching and window feature. On the other hand, the extraction comprised edge and color extraction, thus similarity measurement is to measure the similarity between the measured features and the values stored in the database.



**Fig 1.** object on the conveyor was detected by a webcam

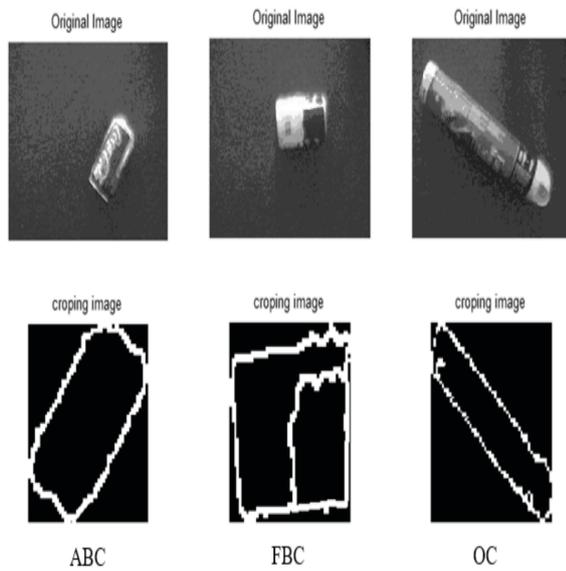
Some factors, such as lighting conditions, speed of the conveyor belt, camera gain, contrast and brightness, height of the camera above the object, may strongly influence the quality and sharpness of the final image; experiments were carried on in order to find an optimum configuration.



**Fig 2** The standard pose of the cans

In feature extraction phase also considered both color and gray scale images. For color image each of the three color components RGB are considered separately. For gray scale image, standard grayscale transformation is obtained from the original RGB image. Processing with the gray-level of one image is called point processing. Once the raw image has been provided by the webcam and grabbing routine, the edges are extracted and the shape of the cans is identified. This procedure has been designed to also work in conditions of bad lighting, high

image noise although the experimental condition may not require all the following image processing steps. Processing with the gray-level of one image is called point processing. The manipulation of image is meant for extracting pixel value and pixel position of each pixel of the image. The image is loaded in a form or picture box using load Picture function. Then, using the point method, pixel values of the pixels of an image are obtained. Next, the red, green and blue values for the pixel values are separated. After the necessary operations, the new pixel values are placed in the appropriate positions in a form or in a picture box using pset method and RGB function. All these steps will result in image manipulation.



**Fig 3.** Cropping image

The identification process was handled by the system, after each can was automatically detected by the webcam. The features of the captured object were then matched with the data for the cans waste in the database. For cans waste identification, the rule based classification was developed using the energy distribution with different quantization level.

The main process in develop the database system consisted of image pre-processing, feature extraction and classification. Precision is defined as the ratio of genuine records in the template database retrieved by the identification system and the total number of templates retrieved. Recall is defined as the ratio of the genuine records in the template database retrieved by the identification system and the total number of genuine records in the template database.

The identification process is undertaken by comparing the value of the image with the values in the database to find the closest values. To determine the accuracy of the identification system, the can images are stored in the database with a standard pose.

## 2.2 Trinomial logistic regression

Consider an image data of cans waste with  $n$  observations ( $i = 1, 2, \dots, n$ ), three cans types ( $j = 1, 2, 3$ ), the number of sample in  $j$ -th type is denoted  $n_{ij}$  and  $n = \sum_{j=1}^3 n_{ij}$ . Let  $N_{ij}$  be the random variable for  $i$ -th can waste with probability into  $j$ -th type,  $\pi_j(\mathbf{x}_i)$ ,  $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ip})^T$  be the predictor variables for  $i$ -th can waste. If  $N_{ij}$  follow trinomial logistic regression model, the probability mass function is

$$P(N_{ij} = n_{ij} | \mathbf{x}_i) = \frac{n!}{\prod_{j=1}^3 n_{ij}!} \prod_{j=1}^3 \pi_j(\mathbf{x}_i)^{n_{ij}} \quad (1)$$

This probability mass function in equation (1) implied that probability of  $n_{i1}, n_{i2}, n_{i3}$  fall into can type 1, 2, 3 respectively given a set of predictor variables  $\mathbf{x}_i$ ,

where  $\sum_{j=1}^3 \pi_j(\mathbf{x}_i) = 1$ . The trinomial logistic regression

model was formed with stated  $P(N_{i1} = 1 | \mathbf{x}_i) = \pi_1(\mathbf{x}_i)$ ,  $P(N_{i2} = 1 | \mathbf{x}_i) = \pi_2(\mathbf{x}_i)$  and  $P(N_{i3} = 1 | \mathbf{x}_i) = \pi_3(\mathbf{x}_i)$  as the following logistic functions,

$$\ln \frac{\pi_j(\mathbf{x}_i)}{\pi_3(\mathbf{x}_i)} = \beta_j^T \mathbf{x}_i \quad (2)$$

where we chose the cans waste type 3 as reference category in this research. Using the logit link function, we have

$$\pi_1(\mathbf{x}_i) = \frac{g_1}{1 + g_1 + g_2} \quad (3)$$

$$\pi_2(\mathbf{x}_i) = \frac{g_2}{1 + g_1 + g_2} \quad (4)$$

and

$$\pi_3(\mathbf{x}_i) = \frac{1}{1 + g_1 + g_2} \quad (5)$$

where

$$g_1 = \exp(\beta_1^T \mathbf{x}_{i1}) = \sum_{k=0}^p \exp(\beta_{1k} x_{ik}) \quad (6)$$

$$g_2 = \exp(\beta_2^T \mathbf{x}_{i2}) = \sum_{k=0}^p \exp(\beta_{2k} x_{i2k}) \quad (7)$$

The regression parameters,  $\beta_k$ , can be estimated from (1) using maximum likelihood estimation, where the log likelihood function is (Agresti, 2013 in Thorsen, 2014),

$$\begin{aligned} \text{Ln}(\beta_{ij}) &= \sum_{i=1}^n \sum_{j=1}^3 n_{ij} \ln \pi_j(\mathbf{x}_i) \\ &= \sum_{i=1}^n \sum_{j=1}^3 n_{ij} \ln \left( \frac{\exp(g_j(\mathbf{x}_i))}{1 + \sum_{j=1}^2 \exp(g_j(\mathbf{x}_i))} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^3 n_{ij} g_j(\mathbf{x}_i) - n_{ij} \ln \left( 1 + \sum_{j=1}^2 \exp(g_j(\mathbf{x}_i)) \right) \\ &= \sum_{i=1}^n n_{ij} \left( \sum_{j=1}^2 \sum_{k=1}^p \beta_{j0} + \beta_{jk} x_{ijk} \right) \\ &\quad - n_{ij} \ln \left( 1 + \sum_{j=1}^2 \exp \left( \sum_{k=1}^p \beta_{j0} + \beta_{jk} x_{ijk} \right) \right) \end{aligned}$$

### 3 Results and discussion

From the results of extracting color intensity of red, green, and blue of the images of cans waste types, we obtain the value of intensity of red, green, and blue is quite varied. They both mean and standard deviation of the values range of color intensity of each cans waste type can be seen in Table 2.

**Table 2.** Mean and standard deviation of values range of color intensity.

Type	Mean	Standard Dev
1	33.78	12.41
2	25.78	2.33
3	9.11	1.45

The cans waste for type 1 has the biggest mean and standard deviation of values range, whereas for type 3 has the smallest mean.

For identification cans waste based on the colors intensity of red, green and blue, we employed the probability approach; trinomial distribution in term regression as described in Sec.2, where predictor variables  $X_1, X_2, X_3$  denote color intensity of red, green,

and blue which the image is taken at top,  $X_4, X_5, X_6$  denote color intensity of red, green, and blue which the image is taken at down, and  $X_7, X_8, X_9$  denote color intensity of red, green, and blue which the image is taken at side successively.

We employed an independence test between each of the predictor variable and cans waste type at first. Based on the chi-square test ( $\alpha = 5\%$ ) is revealed that only the color intensity of red which the image is taken at top pose ( $X_1$ ) that does not correspond with the can waste type. Then the variable is not involved in the trinomial regression modeling to identify the type. By using  $G$ -statistic, the significance of predictor variables simultaneously in the trinomial model denotes that at least one predictor variable of eight predictor variables that influence in identifying the type cans waste ( $\chi^2 = 235.74 > \chi^2_{(0.05;14)} = 23.69$ ). The significance of predictor variables partially based on the Wald test ( $\alpha = 5\%$ ) notes that some variables have p-value  $> \alpha$ . By using the backwardstepwise procedure, we have the significant variables as presented in Table 3.

**Table 3.**The significant variables.

Predictor Variable	p-value
$X_2$	0,00
$X_3$	0,02
$X_4$	0,03
$X_7$	0,02
$X_8$	0,00
$X_9$	0,01

The final trinomial regression model parameters obtained given in Table 4, where type 3 as the reference category. The variance of the predictor variables data in identifying the cans waste type are able to explain the variance of the types by 59.1 percent (based on Nagelkerke value), while the rest is explained by the other predictor variables than this model.

From Table 4, we known that for cans waste type1, the higher the color intensity of the green of the image taken at top pose ( $X_2$ ), the tendency of an image to be identified as type 1 is 1.31 times more than in type 3. The higher the color intensity of the blue of the image taken at top pose ( $X_3$ ), the tendency of the image to be identified as type 1 is 1.06 times more than type 3, etc, while for type 2, the higher the color intensity of the green of the image taken at top pose ( $X_2$ ), the tendency of the image to be identified as type 2 is 1.75 times more than in type 3. The higher the color intensity of the color of the image taken at top pose ( $X_3$ ), the tendency of the

image to be identified as type 2 is 0.68 times more than type 3, etc.

The color intensity of red of the image taken at side pose ( $X_7$ ) has the highest tendency to be identified as type 1 whereas the color intensity of blue of the image taken at side pose ( $X_9$ ) has the lowest tendency. The color intensity of blue of the image taken at side pose ( $X_9$ ) has the highest tendency to be identified as type 2 whereas the color intensity of blue of the image taken at top pose ( $X_3$ ) has the lowest tendency.

**Table 4.** The final model of trinomial regression.

Can waste type		$\beta$	Wald	exp ( $\beta$ )	
1	intercept	-1,67	2,59	-	
	$X_2$	0,27	2,83	1,31	
	$X_3$	0,06	0,13	1,06	
	$X_4$	-0,50	7,67	0,61	
	$X_7$	0,79	12,63	2,21	
	$X_8$	0,62	8,26	1,87	
	$X_9$	-0,56	6,46	0,57	
	2	intercept	-6,49	26,12	-
		$X_2$	0,56	7,15	1,75
$X_3$		-0,39	3,74	0,68	
$X_4$		-0,32	2,14	0,73	
$X_7$		0,23	0,80	1,25	
$X_8$		0,27	1,20	1,30	
$X_9$		1,64	35,68	5,15	

The correct result percentage of testing to identify the can waste type is presented in Table 5.

**Table 5.** Percentage of identification result.

Identification to Cans Waste Type	True Cans Waste Type		
	1	2	3
1	43.13	6.88	9.38
2	6.56	22.81	2.19
3	0.31	1.56	7.19
% correct	43.13	22.81	7.19
Total	73.13		

The detection algorithm was simplified for cans classification. When a can is moved on the belt conveyor through the detection area, it will automatically be identifying.

The higher the level correctly of identification cans waste will denote the higher the level accuracy which will indicate the better the sorting system performance. We found that the approach can identify correctly more than 70%. It is enough for a good sorting system.

This method is a useful procedure for analysing and comparing three types of cans. These results are a definite improvement on all other published results so far which quote a sorting accuracy of only.

## 4 Conclusion

Employing the probability approach of trinomial distribution in term regression, the predictor variables that affect the identification of cans waste type based on the images are the color intensity of green and blue which the image is taken at top pose, the color intensity of red which the image is taken at down pose, and the color intensity of red, green and blue which the image is taken at side pose successively.

The model can identify cans waste into three types based on the images correctly by 73.13%. The level correctly which more than 70% is enough for a good sorting system. This method is a useful procedure for analysing and comparing three types of cans. but the other approaches can be tried to improve performance sorting system.

It is estimated that more investigations on the modern system are necessary. Although knowing how each different component works is a step in the right track, complete combination will give a better sense of the identification system's overall cycle time, failure rates, and problem areas.

This research was funded by Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia (Penelitian Unggulan Perguruan Tinggi, Anggaran DIPA Direktorat Penelitian dan Pengabdian kepada Masyarakat, 2016). The author are grateful to the Universitas Sriwijaya. The author are also many thanks to reviewers for providing valuable comments that help to improve this manuscript

## References

- [1] <http://triaa.com/environment-society/recycling/tri-arrows-recycling/just-kids>, accessed on Nov 28, (2015).
- [2] S. Cointreau, Seminar on Solid Waste Primers and Lessons Learned from World Bank Projects, (2005)
- [3] Irsyadi Y, Edgar S, M. A. Hannan, Djuraidah AW, Hassan B, in Proc. 3rd Int. Conf. on Soft Computing Technology in Civil, Structural and Environmental Engineering, pp.1-11, (2013).
- [4] Yulia R, in Proc. 1st Int. Conf. on Education, Technology and Science, pp. 414-417, (2014).
- [5] Yulia R, J. of Physics: Conference Series, doi:10.1088/1742-6596/622/1/012052, pp. 1-6, (2015).

- [6] Agresti, A., New York. John & Son, (2002).
- [7] Hosmer, D.W. & Lemeshow, S., New York. John & Son, (2000).
- [8] Yulia R, Prosiding Seminar Nasional Sains dan Teknologi 2015, Bandung, Indonesia, (2015).
- [9] Ambarwati, A.N. , H. Kuswantodan I. Zain, Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, Fakultas MIPA, Universitas Negeri Yogyakarta, (2012).
- [10] Buana, I.S., Mahendrawathidan N. Irawan, Seminar Nasional Informatika 2010, UPN Veteran, Yogyakarta, ISSN: 1979-2328, (2010).
- [11] Abdalla El-Habil, Pakistan Journal of Statistics and Operation Research. Vol.VIII No.2, pp.271-291, (2012).
- [12] Mahdi Khodadadzadeh, Jun Li, Antonio Plaza, and José M. Bioucas-Dias, IEEE Geoscience and Remote Sensing Letters, Vol. 11, No. 12, pp. 2105-2109, (2014).
- [13] Erik Thors'en, Mathematical Statistics Stockholm University Bachelor Thesis, (2014).