Science and Technology Indonesia
e-ISSN:2580-4391 *p*-ISSN:2580-4405
Vol. 6, No. 4, October 2021

Research Paper

Prediction of Plastic-Type for Sorting System using Fisher Discriminant Analysis

Irsyadi Yani[1], Yulia Resti[2]*, Firmansyah Burlian[1], Ansyori[3]

[1]*Department of Mechanical Engineering, Faculty of Engineering, Sriwijaya University, Palembang, 30662, Indonesia*
[2]*Department of Mathematics, Faculty of Mathematics and Natural Science, Sriwijaya University, Palembang, 30662, Indonesia*
[3]*Department of Electrical Engineering, Faculty of Engineering, Sriwijaya University, Palembang, 30662, Indonesia*
***Corresponding author**: yulia_resti@mipa.unsri.ac.id

Abstract

Recycling is a more environmentally friendly method of managing and reducing plastic waste that can significantly reduce land degradation, pollution, and greenhouse gas emissions. According to its composition, an essential first step in the recycling process is sorting out plastic waste. However, inadequate sorting of plastic types can result in cross-contamination and increasing industrial operating costs. A low-cost automated plastic sorting system can be developed by using digital image data in the red, green, and blue (RGB) color space as the dataset and predicting the type using learning datasets. The purpose of this paper is to demonstrate how to use Fisher Discriminant Analysis (FDA) to predict the plastic type from a digital image of the RGB model and then evaluate the performance using cross-validation. This work has four main steps: collecting plastic digital image data, forming statistical tests, predicting plastic types, and evaluating prediction performance. FDA is quite effective for predicting the type of plastic. Performance measures the accuracy of 87.11 %, the recall-micro of 91.67 %, the recall-micro of 80.97 %, the specificity-micro of 90.33 %, and the specificity-macro of 90.38 %, respectively. The micro is determined by the number of decisions made for each object. In comparison, the macro is calculated based on the average decision made by each class.

Keywords

Fisher Discriminant Analysis, Plastic-Type, Prediction

Received: 11 July 2021, Accepted: 4 October 2021
https://doi.org/10.26554/sti.2021.6.4.313-318

1. INTRODUCTION

Although plastic is the most widely used inorganic material globally, particularly in countries experiencing rapid economic growth (Srigul et al., 2016), plastic can be harmful to the environment due to its hundreds-year decomposition time (Shuai et al., 2020). Recycling is a viable option for managing and reducing plastic waste instead of landfills and incineration (Chow et al., 2016). This step can significantly reduce land degradation, pollution, and greenhouse gas emissions while also saving up to 95 % of the energy used in the plastic manufacturing process (Siddique et al., 2008). Sorting plastic waste according to its material composition is the initial step in the recycling process. This stage is critical because the improper classification of plastic types can result in cross-contamination, which increases industrial operating costs (Pivnenko et al., 2016). In addition, this process frequently encounters difficulties when attempting to differentiate between different types of plastic (Ruj et al., 2015). The plastic types Polyethylene Terephthalate (PET/PETE), High-Density Polyethylene (HDPE), and Polypropylene are widely used in the community and have the potential to become waste (PP).

Due to the ineffectiveness and inefficiency of the manual method, automatic plastic sorting is a viable solution to this problem. A low-cost automatic plastic sorting system can be developed by utilizing machine learning and a digital image with the RGB color model as a dataset. Machine learning-derived predicted plastic-type values have a purpose in the sorting process. The artificial neural network backpropagation (ANNB) method also is implemented to predict plastic-type based on digital images (Khona'ah et al., 2015; Yani et al., 2020). The ANNB algorithm is a widely used and popular prediction/classification algorithm. However, the minimum accuracy of the classification method is 85 % (Aronoff, 1985). Additionally, the performance of the method is solely based on its accuracy. Therefore, numerous metrics must be used to evaluate the effectiveness of methods (Gorunescu, 2011).

One of the prediction methods in machine learning is Fisher Discriminant Analysis. This method is a powerful tool for developing a statistical prediction algorithm (Raudys and Young, 2004). It has proven very successful in a variety of tasks, including recognizing, assessment of risk, identification, diagnosis, or classifying (Vranckx et al., 2021; Chumachenko et al., 2021; Bari and Fattah, 2020; Wang et al., 2018). This method has several models, such as Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Fisher Discriminant Analysis (FDA). The first two models require a Gaussian multivariate assumption. Only the LDA and FDA assume that the covariance matrix is homogeneous. When the covariance matrix is not homogeneous, the more appropriate model is QDA. LDA is more appropriate than QDA for small sample sizes in learning data and vice versa for enormous sample sizes (James et al., 2013). The LDA can also be more appropriate than QDA when the data dimension is small (Wahl and Kronmal, 1977).

This article proposes using FDA to predict the three plastic types used in sorting systems, with five metrics for method performance: accuracy, the micro and macro proportion of plastic types correctly predicted (recall-micro and recall-macro), and the micro and macro proportion of plastic types correctly predicted (specificity-micro and specificity-macro) (Dinesh and Dash, 2016; Sokolova and Lapalme, 2009).

2. EXPERIMENTAL SECTION

2.1 Materials

The statistics summary of image data collected related to the five normalized predictor variables is noted in Table 1.

Table 1. Summary Statistic of Variable

Statistic	Predictor Variable				
	Red X_1	Green X_2	Blue X_3	Entropy X_4	Variance X_5
Minimum	0.33	0.35	0.33	0.00	0
1^{st} Quartile	0.61	0.63	0.67	0.01	0.01
Median	0.7	0.75	0.78	0.02	0.02
Mean	0.76	0.79	0.8	0.01	0.05
3^{rd} Quartile	0.98	0.99	0.98	0.02	0.12
Maximum	1	1	1	0.03	0.13

2.2 Methods

Figure 1 presents the main stages of this research. Each stage has a minimal one step. The first need to get images of plastic is to build the acquisition system. This system has two key components: a web camera that takes digital images and a computer that processes the images into the RGB format.

There are 450 different plastic data collected by capturing the images in three different random poses. Plastic waste comes from three types; PET, HDPE, and PP. The obtained images are processed into RGB color format, where each color component has a value of 8 bits so that each color component has a scale of 28 = 256 or a pixel value range of 0 to 255. The resolution of the image stored in the database is 560 × 420 pixels. The image is cropped to 34 × 34 pixels with cropping coordinates [280 180 33 33]. Figure 2 presents the three types of the cropped plastic waste digital image.

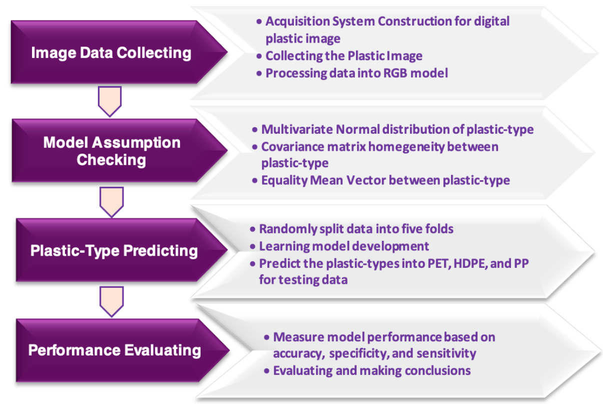

Figure 1. Research Methodology

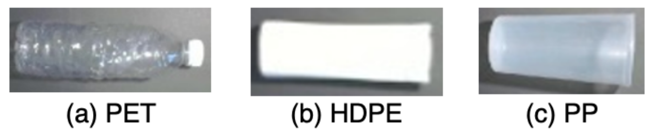

Figure 2. Digital Image of Plastic-Type

The second step is to check the discriminant analysis assumptions. This work proposed the Discriminant Analysis s related to the plastic-types prediction method. The Doornik-Hansen (Adefisoye et al., 2016), the Fligner-Killeen (Stevens, 2012), and the Pillai Trace (Carey, 1998) tests to check multivariate Gaussian distribution, covariance matrix homogeneity, and mean vector equality assumptions related to the prediction method assumptions. The tests are written in (1)–(3),

$$Doornik - Hansen = \left(Z(\sqrt{\theta_1}) + +Z_2^2\right) \tag{1}$$

$$FK = \frac{\sum_{j=1}^{k} n_j \left(\bar{x}_j^z - \bar{x}^2\right)^2}{S^2} \tag{2}$$

$$PT = trace\left(B(B+W)^{-1}\right) \tag{3}$$

For the Doonik-Hansen test Adefisoye et al. (2016), $Z(\sqrt{\theta_1})$ and z_2 are defined respectively as,

$$Z(\sqrt{\theta_1}) = \frac{ln(G/c) + \sqrt{(G/c)^2 + 1}}{\sqrt{ln(\omega)}} \tag{4}$$

$$z^2 = \left(\left(\frac{\xi}{2\varphi}\right)^{\frac{1}{3}} - 1 + \frac{1}{9\varphi}\right)(9\varphi)^{\frac{1}{2}} \tag{5}$$

where $G, c, \omega^2, \xi, and \varphi$ are written successively as,

$$G = \sqrt{\theta_1}\sqrt{\frac{(n+1)(n+3)}{6(n-1)}} \tag{6}$$

$$c = \sqrt{\frac{2}{(\omega^2-1)}} \tag{7}$$

$$\omega^2 = -1 + \sqrt{2\beta_2 - 1} \tag{8}$$

$$\xi = (b_2 - 1 - b_2)2k \tag{9}$$

$$\varphi = \frac{(n+5)(n+7)((n-2)(n^2+27n-70)+b_1(n-7)(n^2+2n-5))}{6(n-3)(n+1)(n^2+15n-4)} \tag{10}$$

For m_2 and m_3 are the second and third central moments, respectively,

$$\theta_1 = \frac{m_3^2}{m_2^3} \tag{11}$$

$$\beta_2 = \frac{3(n^2+27n-70)(n-3)(n+1)}{(n-2)(n+5)(n+7)(n+9)} \tag{12}$$

$$k = \frac{(n+7)(n+7)(n^3+37n^2+11n-313)(n-3)(n+1)}{12(n-3)(n+1)(n^{2+15n-4})} \tag{13}$$

The Fligner-Killeen test are defined Stevens (2012) successively as,

$$S = \frac{1}{\Sigma_{j=1}^k n_j}\Sigma_{j=1}^k n_j S_j \tag{14}$$

For the Pillai Trace test B and W are formulated Carey (1998) as,

$$B = \sum_{j=1}^{k} n_j(\bar{X}_j - \bar{X})(\bar{X}_j - \bar{X})^T \tag{15}$$

$$W = \sum_{j=1}^{k} n_j \sum_{i=1}^{k} n_j(x_{ij} - \bar{X})(x_{ij} - \bar{X})^T \tag{16}$$

The third step is to implement the discriminant analysis to build learning models and predict the plastic types. The stages in this step are randomly split data, learning model development, and predict the plastic types into PET, HDPE, and PP for testing data. The data were randomized into five-folds, four folds to build a learning model, and the remaining one-fold to predictive data (Lantz, 2019; Alpaydin, 2016). The model analysis that is implemented is one of three models: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), or Fisher Discriminant Analysis (FDA). The model selection is based on the results of statistical testing assumptions. LDA or QDA can be implemented when Gaussian assumptions are fulfilled. LDA considers that all groups have the same covariance matrix, whereas QDA is calculated based on the covariance matrix of each group (Hastie et al., 2009). The sample size is critical when deciding whether to use LDA or QDA (Wahl and Kronmal, 1977). Generally, LDA is more appropriate than QDA for small sample sizes in learning data and vice versa for enormous sample sizes (James et al., 2013). However, if this assumption is not met, it is more appropriate to implement the FDA.

The plastic image with $X = (X_1, X_2, X_3, X_4, X_5)^T$ is classified as the j -th plastic-type if the discriminant function $\hat{d}_j(x)$ is the largest. The $\hat{d}_j(x)$ for both models, LDA and QDA, respectively (James et al., 2013).

$$\delta(x) = ln\pi_j + X^T\Sigma^{-1}\mu_j - \frac{1}{2}\mu_j^T\Sigma_j\mu_j \tag{17}$$

$$\delta(x) = ln\pi_j - \frac{1}{2}ln|\Sigma_j| - \frac{1}{2}X^T\Sigma^{-1}X + X^T\Sigma_j^1\mu_j - \frac{1}{2}\mu_j^T\Sigma_j\mu_j \tag{18}$$

with covariance matrix respectively, Σ and $\Sigma_j, \forall_j$.

In FDA, X is classified as the j -th plastic-type if the linear combination, $Y_j = V^T X$, is maximum where,

$$V = S_W^{-1}(\mu_1 - \mu_2) \tag{19}$$

$$S_W = \sum_{j=1}^{2} S_j \tag{20}$$

$$S_j = \sum_{x_i \epsilon j^{th} g} (X_i - \mu_j)(X_i - \mu_j)^T \tag{21}$$

$$\mu_j = \frac{1}{n_j}\sum_{x_i \epsilon j^{th} g} X_i \tag{22}$$

The final step is to evaluate the performance of the discriminant analysis. Scalar values are used to represent classification performance in various metrics such as accuracy, recall-micro (μ), recall-macro (M), specificity-micro (μ), and specificity-macro (M). The TP_j, FP_j, TN_j, and FN_j values are determined for each plastic type, $j = 1, 2, 3$. The micro proportion is calculated based on the number of decisions per object, while the macro proportion is calculated based on the average decision per class. The performance measurements refer to Table 2 for the first plastic type. The performance measure for other plastic types is determined similarly (Dinesh and Dash, 2016; Sokolova and Lapalme, 2009).

Table 2. Confusion Matrix for Plastic-Type, $j = 1$

		Actual		
	j	1	2	3
Prediction	1	True-Positive (TP)	False-Negative (FN)	False-Negative (FN)
	2	False-Positive (FP)	True-Negative (TN)	True-Negative (TN)
	3	False-Positive (FP)	True-Negative (TN)	True-Negative (TN)

$$Accuracy = \frac{\Sigma_{j=1}^{3} \frac{TP_j+TN_j}{TP_j+FP_jFN_j+TN}}{3} \tag{23}$$

$$Recall_\mu = \frac{\Sigma_{j=1}^{3} TP_j}{\Sigma_{j=1}^{3}(TP_j + FN_j)} \tag{24}$$

$$Recall_M = \frac{\Sigma_{j=1}^{3} \frac{TP_j}{(TP_j+FN_j)}}{3} \tag{25}$$

$$Specificity_\mu = \frac{\Sigma_{j=1}^{3} TN_j}{\Sigma_{j=1}^{3}(FP_j + TN_j)} \tag{26}$$

$$Specificity_M = \frac{\Sigma_{j=1}^{3} \frac{TN_j}{(FP_j+TN_j)}}{3} \tag{27}$$

3. RESULTS AND DISCUSSION

Tables 3–4 summarize the results of the assumption tests for discriminant analysis for all of the learning data. This work used the Doornik-Hansen and the Fligner-Killeen tests to assess multivariate Gaussian distributions of explanatory variables and homogeneity of covariance matrices between types of plastic waste, respectively.

Table 3 demonstrates that not all plastic types in all learning data have a multivariate Gaussian distribution at the 5 % significance level. Only the first, second, and fifth folds datasets,

Table 3. Multivariate Gaussian Test

Doornik-Hansen Test			Learning Data				
			1	2	3	4	5
Type of plastic	PET	statistic	140.96	143.88	123.75	104.97	117.42
		p-value	0	0	0	0	0
	HDPE	statistic	397.25	456.53	351.74	369.93	300.34
		p-value	0	0	0	0	0.09
	PP	statistic	29.05	27.37	25.77	36.07	37.22
		p-value	0	0	0	0	0

and even then, only HDPE plastic-type data have a multivariate Gaussian distribution. The assumption of a multivariate Gaussian distribution is required for the majority of multivariate analyses. However, it is challenging to locate data with a multivariate Gaussian distribution over all real-world groups (Hallin and Paindaveine, 2009).

Table 4. Homogeneity of Covariance Matrices Test

Fligner-Killeen Test	Learning Data				
	1	2	3	4	5
Chi-sq	4.35	0.58	0.95	1.35	1.7
p-value	0.11	0.74	0.62	0.51	0.43

The next assumption test in discriminant analysis is the homogeneity of the covariance matrix. This independent variable test is carried out when the Gaussian multivariate assumption is not met. Currently, the assumption of an equal mean vector is not necessary. Related to the homogeneity test as described in Table 4, the result shows that all learning data have a homogenous covariance matrix with a significance level of 5 %.

FDA is used to make predictions based on the findings test of the Gaussian multivariate and the covariance matrix homogeneity assumptions.

Table 5. Performance of Plastic Waste Classification using FDA

Performance Measurement	Testing Data					Average	Variance
	1	2	3	4	5		
Accuracy	87.41	85.19	85.93	88.89	88.15	87.11	2.37
Recall miu	81.11	77.78	78.89	79.07	83.22	91.67	5.29
RecallM	81.03	78.86	79.07	83.22	82.68	80.97	4.04
Specificity miu	90.56	88.89	89.44	91.67	91.11	90.33	1.32
SpecificityM	90.74	88.87	89.53	91.49	91.26	90.38	1.30

This work has an accuracy of 87.11 %, recall-micro (μ) and recall-macro (M) at 91.67 % and 80.97 % respectively, specificity-micro (μ) and specificity-macro (M) at 90.33 % and 90.38 % respectively. This information shows that the FDA method is quite good at predicting plastic type since, according to Aronoff (1985), the minimum accuracy of the classification method is 85 %. Other than that, the specificity that calculates the truth in all plastic-types other than the selected types against all other types has the higher standard deviation (about 2 %), and the recall calculates the correctness model of statistical learning in predicting that the plastic-type has the lowest standard deviation (about 1 %). Thus, this work's result is better than Khona'ah et al. (2015), who implemented the ANNB algorithm to predict the plastic types with an accuracy of 86.67 %. Although the difference in prediction accuracy does not reach 1 %, this work has proposed different validation techniques and more performance measures than Khona'ah et al. (2015) to show that the prediction results have low variance. Therefore, better prediction performance for plastic types than our proposed method can be obtained by implementing classification methods that do not require the assumption of a multivariate Gaussian distribution and homogeneity of the covariance matrix. These methods include k-NN, decision tree, or Support Vector Machine.

4. CONCLUSIONS

Plastic recycling is a more environmentally friendly method of managing and reducing plastic waste that can significantly reduce land degradation, pollution, and greenhouse gas emissions. This stage is crucial because inaccurate sorting of plastic types can cross-contamination and increase industrial operating costs. This paper evaluates the performance of the Fisher Discriminant Analysis model to predict the plastic type using digital images. This model successfully predicts the plastic-type. Performance measures the accuracy of 87.11 %, the micro and macro proportion of plastic-type with correctly predicted (recall) was 91.67 % and 80.97 %, respectively. In contrast, the micro and macro proportion of the plastic-type into other types predicted correctly (specificity) was 90.33 % and 90.38 %, respectively. However, superior prediction performance for plastic types can be obtained using classification methods that do not require the assumption of a multivariate Gaussian distribution and homogeneity of the covariance matrix, for the examples k-NN, decision tree, or Support Vector Machine.

5. ACKNOWLEDGEMENT

The research/publication of this article was funded by DIPA of Public Service Agency of Sriwijaya University 2021. SP DIPA-023.17.2.677515 /2021, On November 23, 2020. In accordance with the Rector's Decree Number: 0010/ UN9/ SK.LP2M.PT/2021, On April 28, 2021.

REFERENCES

Adefisoye, J., B. Golam Kibria, and F. George (2016). Performances of several univariate tests of normality: An empirical study. *J. Biom. Biostat*, **7**; 1–8

Alpaydin, E. (2016). *Machine learning: the new AI*. MIT press

Aronoff, S. (1985). The minimum accuracy value as an index of classification accuracy. *Photogrammetric Engineering and Remote Sensing*, **51**(1); 99–111

Bari, M. F. and S. A. Fattah (2020). Epileptic seizure detection in EEG signals using normalized IMFs in CEEMDAN domain and quadratic discriminant classifier. *Biomedical Signal Processing and Control*, **58**; 101833

Carey, G. (1998). Multivariate analysis of variance (MANOVA) II: practical guide to ANOVA and MANOVA for SAS. *Retrieved September*, **1**; 2009

Chow, C.-f., W.-M. W. So, and T.-Y. Cheung (2016). Research and development of a new waste collection bin to facilitate education in plastic recycling. *Applied Environmental Education & Communication*, **15**(1); 45–57

Chumachenko, K., J. Raitoharju, A. Iosifidis, and M. Gabbouj (2021). Speed-up and multi-view extensions to subclass discriminant analysis. *Pattern Recognition*, **111**; 107660

Dinesh, S. and T. Dash (2016). Reliable evaluation of neural network for multiclass classification of real-world data. *arXiv preprint arXiv:1612.00671*

Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques*, volume 12. Springer Science & Business Media

Hallin, M. and D. Paindaveine (2009). Optimal tests for homogeneity of covariance, scale, and shape. *Journal of Multivariate Analysis*, **100**(3); 422–444

Hastie, T., R. Tibshirani, and J. Friedman (2009). The elements of statistical learnin. *Cited on*; 33

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An introduction to statistical learning*, volume 112. Springer

Khona'ah, B., D. Rosiliani, and I. Yani (2015). Identification and Clasification of Plastic Color Images Based on The RGB Method. . *Journal of Multidisciplinary Engineering Science and Technology*, **6**(6); 10170–10174

Lantz, B. (2019). *Machine learning with R: expert techniques for predictive modeling*. Packt publishing ltd

Pivnenko, K., M. Eriksen, J. Martín-Fernández, E. Eriksson, and T. Astrup (2016). Recycling of plastic waste: Presence of phthalates in plastics from households and industry. *Waste Management*, **54**; 44–52

Raudys, Š. and D. M. Young (2004). Results in statistical discriminant analysis: A review of the former Soviet Union literature. *Journal of Multivariate Analysis*, **89**(1); 1–35

Ruj, B., V. Pandey, P. Jash, and V. Srivastava (2015). Sorting of plastic waste for effective recycling. *International Journal of Applied Science and Engineering Research*, **4**(4); 564–571

Shuai, C., Y. Cheng, W. Yang, P. Feng, Y. Yang, C. He, F. Qi, and S. Peng (2020). Magnetically actuated bone scaffold: Microstructure, cell response and osteogenesis. *Composites Part B: Engineering*, **192**; 107986

Siddique, R., J. Khatib, and I. Kaur (2008). Use of recycled plastic in concrete: A review. *Waste Management*, **28**(10); 1835–1852

Sokolova, M. and G. Lapalme (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, **45**(4); 427–437

Srigul, W., P. Inrawong, and M. Kupimai (2016). Plastic classification base on correlation of RGB color. In *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, pages 1–5

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge

Vranckx, I., J. Raymaekers, B. De Ketelaere, P. J. Rousseeuw, and M. Hubert (2021). Real-time discriminant analysis in the presence of label and measurement noise. *Chemometrics and Intelligent Laboratory Systems*, **208**; 104197

Wahl, P. W. and R. A. Kronmal (1977). Discriminant functions when covariances are unequal and sample sizes are moderate. *Biometrics*; 479–484

Wang, X., X. Li, R. Ma, Y. Li, W. Wang, H. Huang, C. Xu, and Y. An (2018). Quadratic discriminant analysis model for assessing the risk of cadmium pollution for paddy fields in a county in China. *Environmental Pollution*, **236**; 366–372

Yani, I., D. Rosiliani, B. Khona'ah, and F. Almahdini (2020). Identification and plastic type and classification of PET, HDPE, and PP using RGB method. In *IOP Conference Series: Materials Science and Engineering*, volume 857. IOP Publishing, page 012015