



## Data Article

Post-harvested *Musa acuminata* Banana Tiers DatasetEduardo Jr Piedad<sup>a,\*</sup>, June Anne Caladcad<sup>b</sup><sup>a</sup> Computer Software Division, Advanced Science and Technology, Quezon City 1101, Philippines<sup>b</sup> Department of Industrial Engineering, University of San Carlos, Cebu City 6000, Philippines

## ARTICLE INFO

## Article history:

Received 21 November 2022

Revised 15 December 2022

Accepted 20 December 2022

Available online 25 December 2022

Dataset link: [Tier-based Dataset: Musa-Acuminata Banana Fruit Species \(Original data\)](#)

## Keywords:

Banana

*Musa acuminata*

RGB image

LAB color

## ABSTRACT

Post-harvested *Musa acuminata* banana species from a local banana plantation in the Philippines are the subject of this article. All banana tier samples used were pre-classified into four classes by a local expert. These four classifications are extra class, class I, class II, and reject. There are six images captured per banana tier sample from the six different views. Each captured image underwent a three-step image transformation to finely extract the RGB numerical values while the size measurement feature was gathered through manual measurement. The dataset presented in this article provides a brief differentiation of the different classes of banana tiers for commercial use through image processing. This dataset can be useful in establishing an advanced intelligent system in a non-invasive approach through machine and deep learning techniques.

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)DOI of original article: [10.1016/j.postharvbio.2018.06.004](https://doi.org/10.1016/j.postharvbio.2018.06.004)

\* Corresponding author.

E-mail address: [eduardojr.piedad@asti.dost.gov.ph](mailto:eduardojr.piedad@asti.dost.gov.ph) (E.J. Piedad).<https://doi.org/10.1016/j.dib.2022.108856>2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

Subject	Agriculture Engineering, Computer Science
Specific subject area	Postharvest Classification of Bananas, Image Processing, RGB (red, green, and blue) Color Values, Banana Classes
Type of data	Table Image Figure
How the data were acquired	Bananas were gathered from a local banana plantation in the Philippines. A local expert of the Department of Agriculture, Cebu, Philippines, pre-classified the harvested banana samples into their respective classes. A hardware was designed to capture the images of each sample. It is composed of six A4Tech PK-910H Web cameras and LED bar lights to provide sufficient illumination constant in all samples. The developed hardware also provided a clean white background, making it easier for the samples to undergo pre-processing. For the fruit's dimension, a straightforward manual measurement was performed.
Data format	Raw Analyzed Filtered
Description of data collection	Bananas are classified based on tiers into four classes: "extra class", class I", "class II", and "reject". A total of 194 banana tiers were chosen randomly, in which there are 65, 49, 30, and 50 samples per class, respectively. Each banana tier sample was captured with white background in six different views: front, back, left, right, top, and bottom views (Figure 1). There was a total of 1,164 images gathered, which then underwent feature extraction. The features extracted for each image were the RGB values and the size dimension. For banana tiers with an odd set of top fingers, the length of the middle finger was used while two middle fingers for an even set of top fingers.
Data source location	All banana samples were acquired from: <ul style="list-style-type: none"> <li>· City/Town/Region: Carmen</li> <li>· Country: Philippines</li> </ul> <p>The capturing of images was conducted in:</p> <ul style="list-style-type: none"> <li>· Institution: University of San Jose – Recoletos</li> <li>· City/Town/Region: Cebu</li> <li>· Country: Philippines</li> </ul>
Data accessibility	Repository name: Mendeley Data Data identification number: 10.17632/zk3tkxndjw.2 Direct URL to data: <a href="https://data.mendeley.com/datasets/zk3tkxndjw/2">https://data.mendeley.com/datasets/zk3tkxndjw/2</a>
Related research article	E.J. Piedad, J.I. Lazada, G.J. Pojas, L.V.V. Ferrer, Postharvest classification of banana ( <i>Musa acuminata</i> ) using tier-based machine learning. <i>Postharvest Biology and Technology</i> . 145 (2018) 93-100. <a href="https://doi.org/10.1016/j.postharvbio.2018.06.004">10.1016/j.postharvbio.2018.06.004</a>

## Value of the Data

- The data extracted are the features of banana tiers – red, green, and blue color values, and the size dimensions of the top middle finger and their corresponding classification.
- The data would be useful to the agricultural sectors and banana plantations for improved postharvest classification of bananas.
- If utilized in machine learning and deep learning, this dataset will help supplement existing data compilations to establish an automated classification system for large banana plantations and banana exporting companies.

## 1. Objective

One of the most important horticultural products in the Philippines, being a top export and dollar earner, is the banana fruit [1]. Classification of commodities is one essential processes in

the agriculture industry; however, this is typically done manually, which is unreliable due to individual differences in perception [2,3]. Moreover, the ability of manual inspection to deliver accurate grading diminishes with the increase in the number of factors that have to be considered [4,5]. High value crops in the Philippines revealed postharvest losses for bananas ranging from 3% up to 30% due to the farm practice of manual classification [6].

Bananas tiers are traditionally classified by local farmers through a manual process. The color and finger sizes of the banana tiers are used to sort them into their classes. The classification system was developed following this manual sorting process. Images of each sample were captured at different angles, which served as the initial data. These sample underwent feature extraction to generate its RGB values and size dimensions as the primary dataset of this article.

This dataset also served as a pre-requisite in developing a classification system for bananas using tier-based machine learning in the study of Piedad et al. [7]. Furthermore, the dataset was also divided for the training and testing phases, particularly, into a 70-30% proportion, respectively.

## 2. Data Description

The data associated with this work contains a folder (Banana image dataset) of the captured images of banana tier in its six angles and a .csv file containing the extracted features for each sample. There are a total of 194 banana tier samples with six images from their side views: front left, right, back, top, and bottom; and thus, there is a total of 1,164 banana images. Each sample are separated in a folder labeled with a number that corresponds to its sampleID. The `banana_feature.csv` contains the following information:

- sampleID (pertaining to the labeled folders in the 'banana image dataset' folder)
- the extracted features in RGB (red, green, blue) color values
- image side view types
- banana class
- size dimension in (mm)

Note that the finger size dimensions were taken manually by measuring the length of the top middle finger of a banana tier in millimeter (mm). The specific finger measured was assumed to have a regular size in comparison to the other fingers in a banana tier.

## 3. Experimental Design, Materials and Methods

### 3.1. Banana sample preparation

The banana samples were all acquired from a local banana plantation in Cebu, Philippines. The harvested bananas are pre-classified by the local experts of the Department of Agriculture, Cebu, Philippines into four different classes: "extra class", "class I", "class II", and "reject". Shown in Table 1 are the specifications of bananas per class taken from Philippine National Standards for fresh fruit – banana. Bananas are sorted either by tiers or by clusters, which is commonly done in actual postharvest practice [7,8]. There was a total of 194 banana tier samples gathered where there are 65, 49, 30, and 50 samples for extra class, class I, class II, and reject, respectively.

A straightforward approach was used in extracting the size of the bananas. The top middle fingers are manually measured (in mm). For the even-numbered top fingers set, the average of the top two middle fingers is calculated. The size measurement is illustrated in Fig. 1.

**Table 1**

Banana classes specification.

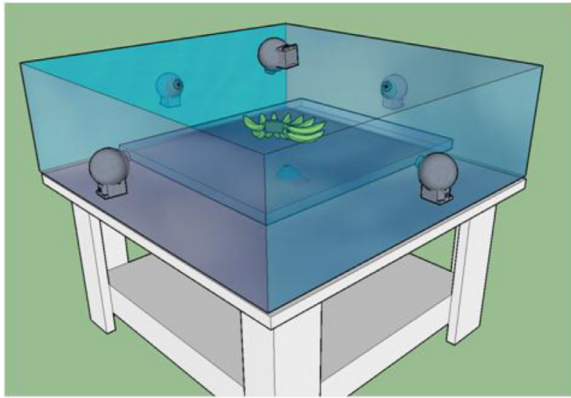
Classification	Description
Extra class	<ul style="list-style-type: none"> <li>- superior quality</li> <li>- must have all the typical characteristics and color of the variety and/or commercial type</li> <li>- must be mature, clean, well-formed, well-trimmed and free of decay, split fingers, loose fingers, bruises, blemishes and discoloration caused by diseases, insects, molds, latex burn, and mechanical damage</li> <li>- very slight superficial defects are allowed, provided that these do not affect the general appearance, the quality, the keeping quality and presentation in the package</li> </ul>
Class I	<ul style="list-style-type: none"> <li>- must be of good quality</li> <li>- must have all the typical characteristics and color of the variety and/or commercial type</li> <li>- must be mature, clean, well-formed, well-trimmed and free of decay, split fingers, loose fingers, bruises, and discoloration caused by diseases, insects, molds, latex burns, and mechanical damage</li> <li>- slight defects of the fingers may be allowed, provided these do not affect the general appearance, the quality, the keeping quality and presentation in the package</li> <li>- slight defects in shape and color</li> <li>- slight defects on the skin due to rubbing and other defects such as sunburns and blemishes not exceeding 5% of the total surface</li> <li>- defects must not, in any case, affect the flesh of the fruit</li> </ul>
Class II	<ul style="list-style-type: none"> <li>- includes banana which do not qualify for inclusion in higher classes but satisfy the minimum requirements</li> <li>- defects may be allowed, provided the banana retain their essential characteristics as regards the quality, the keeping quality and presentation in the package</li> <li>- defects in shape and color</li> <li>- defects on the skin due to rubbing and other defects such as sunburn and blemishes not exceeding 10% of the total surface area</li> <li>- defects must not, in any case, affect the flesh of the fruit</li> </ul>



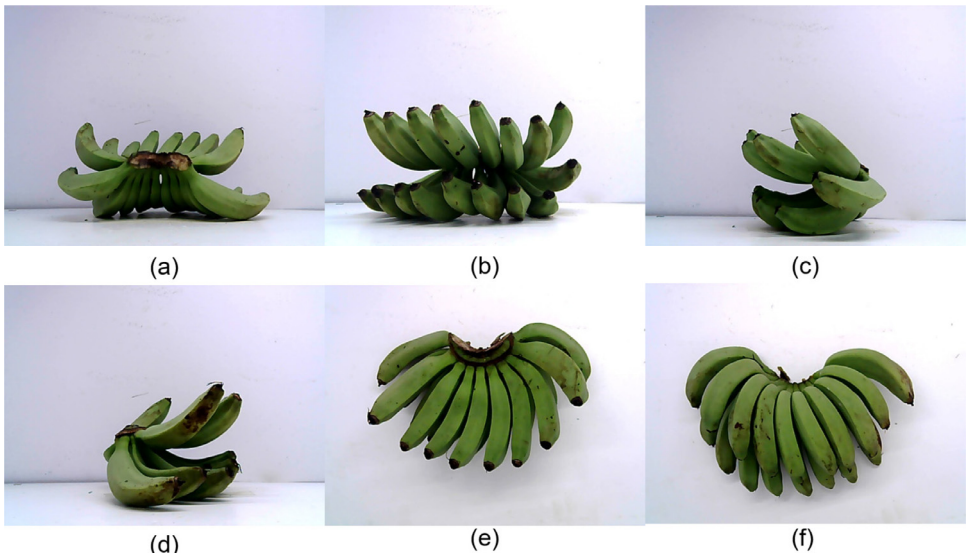
**Fig. 1.** Banana tier size measurement  $d$  based on the (a) top middle finger, and for an even set of top fingers, the average size of the (b) top two middle fingers is taken.

### 3.2. Hardware specifications

An A4Tech PK-910H web camera and LED bar light illumination were used to capture the banana hands in six different views. Shown in Fig. 2 is the prototype design when the images of each banana sample were captured and in Fig. 3 is a sample of the captured banana tier in six different views. An isometric view of the banana hand was obtained by choosing the viewing direction such that the angles between the projections of the  $x$ ,  $y$ , and  $z$  axes are all the same, or



**Fig. 2.** Hardware design for image acquisition.



**Fig. 3.** A banana tier taken in (a) front, (b) back, (c) left, (d) right, (e) top, and (f) bottom side views.

120°. All images are taken on a uniform-colored background. Furthermore, each sample's images were captured at a sufficiently close distance but exact measurement of camera distance was not observed. This is because the impact of the measurement error due to close-distance variations is statistically insignificant [9,10].

The captured images of each sample were processed to extract their features. The extracted features are the red, green, and blue color values and the size dimensions. The extracted feature served as the independent variable while the banana classes are the dependent variable.

### 3.3. Pre-processing

The RGB color space is the most used computer display [11]. The image segmentation operations used and evaluated was background subtraction to separate the subject from its background and  $L^*a^*b$  color space algorithm to extract the actual numerical value of the hand. These

features were transcribed in Emgu libraries for the mechanics of the database. A sample of the background subtractions is shown in the study of Piedad et al.

## Ethics Statements

Not applicable.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

Tier-based Dataset: [Musa-Acuminata Banana Fruit Species \(Original data\)](#) (Mendeley Data).

## CRediT Author Statement

**Eduardo Jr Piedad:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing; **June Anne Caladcad:** Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

## Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- [1] Philippine Statistics Authority, Agricultural foreign trade statistics of the Philippines. 2017 <https://psa.gov.ph/content/agricultural-foreign-trade-statistics-philippines-2015>. Accessed October 2022.
- [2] A.P. Gawande, S.S. Dhande, Implementation of fruit grading system by image processing and data classifier - a review, *Int. J. Eng. Res. Generic Sci.* 2 (2014) 411–413.
- [3] Y. Zhang, P. Phillips, S. Wang, G. Ji, J. Yang, J. Wu, Fruit classification by biogeography-based optimization and feed-forward neural network, *Expert Syst* 33 (2016) 239–253, doi:10.1111/exsy.12146.
- [4] K. Dimililer, E. Olaniyi, Intelligent sorting system based on computer vision for banana industry, *Int. J. Sci. Eng. Res.* 6 (2015) 332–337.
- [5] D. Lorente, N. Aleixos, J. Gomez-Sanchis, S. Cubero, O.L. Garcia-Navarrete, J. Blasco, Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment, *Food Bioprocess Technol.* 5 (2012) 1121–1142, doi:10.1007/s11947-011-0725-1.
- [6] P. Nuevo, M. Maunahan, L. Amatorio, Postharvest handling systems of banana in the Philippines. 2001 <https://agris.fao.org/agris-search/search.do?recordID=PH2003001697>. Accessed October 2022.
- [7] E.J. Piedad, J.I. Larada, G.J. Pojas, L.V.V. Ferrer, Postharvest classification of banana (*Musa acuminata*) using tier-based machine learning, *Postharvest Biol. Technol.* 145 (2018) 93–100, doi:10.1016/j.postharvbio.2018.06.004.
- [8] T.-T. Le, C.-Y. Lin, E.J. Piedad, Deep learning for noninvasive classification of clustered horticultural crops – a case for banana fruit tiers, *Postharvest Biol. Technol.* 156 (2019) 110922, doi:10.1016/j.postharvbio.2019.05.023.
- [9] E.J. Piedad, R. Villeta, Displacement and illumination levels effect on short-distance measurement errors of using a camera, *Recoletos Multidiscipl. Res. J.* 4 (2016), doi:10.32871/rmrj1604.01.06.
- [10] H. Choi, S.-w. Choe, J. Ryu, Optical design of a novel collimator system with a variable virtual-object distance for an inspection instrument of mobile phone camera optics, *Appl. Sci.* (2021) 11, doi:10.3390/app11083350.
- [11] M. Loesdau, S. Chabrier, A. Gabillon, Hue and saturation in the RGB color space, *Image Signal Process.* (2021) 203–212, doi:10.1007/978-3-319-07998-1\_23.