International Research Journal of Agricultural Science and Soil Science (ISSN: 2251-0044) Vol. 13(1) pp. 1-5, April, 2025
Available online https://www.interesjournals.org/agricultural-science-soil-science.html

Research Article

Weight Estimation of Mangoes (*Mangifera indica*) using Machine Learning

Joaquin AC* and Avila RP

Department of Agriculture, Philippine Center for Postharvest Development and Mechnization (DA-PHilMech), Science City of Munoz, Nueva Ecija, Philippines

*Corresponding Author's E-mail: arlyn1111@gmail.com

Received: 23-Dec-2024, Manuscript No. IRJAS- 26-156120; **Editor assigned:** 25-Dec-2024, PreQC No. IRJAS- 26-156120 (PQ); **Reviewed:** 08-Jan-2025, QC No. IRJAS-26-156120; **Revised:** 16-Apr-2025, Manuscript No. IRJAS-26-156120 (R); **Published:** 23-Apr-2025, DOI: 10.14303/2251-0044.2025.037

Abstract

Sorting and classification of fruits according to weight is traditionally done by hand. But human factors often cause error and inaccuracies. Automatically performing this process with machines is important in terms of speeding up the process, reducing costs, and minimizing errors. This research aimed to develop a machine-learning model to improve weight estimation of Carabao mangoes (*Mangifera indica*). To accomplish this, a conveyor belt system and web camera were employed to collect videos for training and validation purposes. Image processing techniques were utilized to extract and calculate pixel count, length, and width of the captured mangoes. The dataset comprised randomly selected carabao mangoes, each with pre-determined weights obtained through digital weighing scales. The developed machine learning Multiple Linear Regression (MLR) model employed a total of 106 training samples and achieved an accuracy of 92.44%, with an additional 106 validation samples attaining 97.32% accuracy.

Keywords: Fruit sorting, Machine learning, Mangoes, Weight estimation

INTRODUCTION

The Philippines has emerged as a significant player in the global mango export market, ranking seventh worldwide. Among various fruit crops, mango holds the distinction of being the third most exported fruit after banana and pineapple. Weight is one of the crucial quality parameters in the export of mangoes, alongside factors such as color, shape, and bruise damage. Precise grading and sorting are essential to meet the stringent requirements for mango export, but the conventional visual inspection methods employed are labor-intensive and prone to human errors.

To address these challenges, researchers have explored diverse techniques that incorporate digital image processing to estimate mango weight (Teoh CC et al., 2007; Thong ND et al., 2019). In that same study, a mechanical system was developed; employing image processing and artificial intelligence to classify mangoes based on color, volume, shape, defects, and subsequently estimate weight and volume using a load cell sensor. Same study of Menguito JJ et al., (2018), proposed an alternative method for determining mango weight through natural frequency analysis, utilizing two program algorithms, namely: Frequency Detector (FD) and Frequency and Weight Detector (FWD). Additionally, another study utilized machine learning algorithms and image processing to estimate the weight of Chokanan mangoes based on pixel count (Teoh CC et al., 2007). Machine learning techniques have also been applied to other crops, such as the work of Basak JK et al., (2022) where models were developed for measuring fruit weight using a non-destructive method based on pixel count from strawberry images and machine learning algorithms. However, there remains a dearth of literature on estimating mango fruit weight using machine learning algorithms that leverage pixel counts and estimated mango dimensions.

Hence, in this study, proposed a mechanical system with rollers, integrated with a vision system, to capture comprehensive video footage of the mango fruit's physical appearance and measure its length, width, and corresponding pixel counts as area. Multiple Linear Regression (MLR) served as the decision-making algorithm for mango classification in this study. Specifically, this paper focuses on the algorithm aspect, particularly the classification of mangoes based on size (weight).

The findings of this study are significant as they offer a robust and efficient approach to estimate the weight of carabao mangoes, contributing to the improvement of the mango supply chain and facilitating better grading and sorting processes for mango exportation. The integration of machine learning techniques and precise digital measurements not only enhances accuracy but also streamlines the estimation process, potentially benefiting the mango industry and global trade of this popular tropical fruit.

MATERIALS AND METHODS

Sample collection

A total of 212 random samples of Philippine Carabao mangoes were acquired from Pangasinan and Manila where most of mango exporters do their business. Mangoes were pre-weighed using a digital balance scale, whereas length and width (mm) were measured using a digital caliper. The samples were separated into training and validation sets of 113 samples each. The training set was utilized to develop the Multiple Linear Regression (MLR) model, where the mangoes' length and width served as input features for weight prediction.

In the subsequent phase, the validation set was used to assess the accuracy and performance of the developed MLR model. The effectiveness and precision of the MLR model were evaluated by comparing the predicted weights from the model with the actual measured weights of the mangoes.

Video capturing and image processing

The collected samples were fed into the imaging chamber system Figure 1, consisting of a 1080p web camera and lighting system for capturing mango videos. A total of 33,900 frame images were extracted from several videos of rotating mangoes, and images underwent pre-processing to convert the RGB color space into HSV color space to separate the background from the real mango images. Elimination of background and leaving only the targeted object is part of the pre-processing (KC K et al., 2021; Sann SS et al., 2021). A grayscale image was then constructed. After pre-processing, the estimated volume of the mango is determined by the sum of the volume of 10 transverse cuts along the length of the mango. Analysis was derived from computing the semi-axes a and b at different points along the length of the mango Figure 2 and relationship between the values of the semi-axes called Depth Factor (DF). This method was adopted from the works of Bermudez AM et al., (2013).

Figure 1. Sample image of mango inside the imaging chamber system.

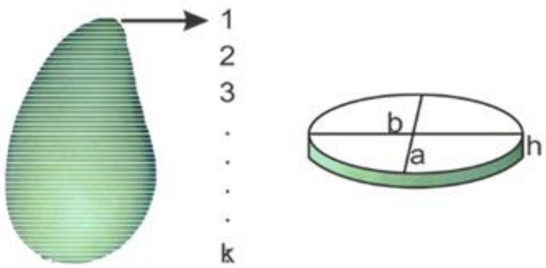

Figure 2. Transverse-cut method of mango weight estimation.

Statistical analysis

Regression models have proven to be highly valuable in numerous applications for predicting certain variables based on other predictor variables (Navidi W, 2011). Among these models, multiple linear regression stands out as a widely used statistical technique that enables the prediction of a response variable using multiple explanatory variables. It represents an extension of the simple linear regression model, which involves only a single explanatory variable (Trunfio TA et al., 2022) This study employs a multiple linear regression analysis to explore the intricate relationship between mango length, width, pixel counts as area, and their corresponding actual weight.

By developing a linear model within the regression analysis, the study aims to predict mango weight, the dependent variable, utilizing length, width, and the counted mango pixels as independent variables. The comprehensive analysis leverages these morphological attributes to facilitate an accurate estimation of the weight of carabao mangoes.

RESULTS AND DISCUSSION

Figure 3 shows sample images obtained from the imaging chamber system, and the sample threshold image is shown in Figure 4, respectively. The thresholding was used to convert the original images into a binary image, where pixels are represented as either black or white, which aims to isolate the mango from the background.

The objective of the regression analysis was to determine mango weight based on four predictor variables: Length (L1), length pixels (P1), width (W2), and width pixels (P2) as the regression equation [1] derived from the analysis.

$$Weight = 103.3 - 1.040\ L1 + 0.000452\ P1 - 2.252\ W2 + 0.002034\ P2 \qquad [1]$$

Figure 3. Images captured from the imaging chamber system.

Figure 4. Sample threshold image of mango.

The model summary shown in Table 1 provided various statistics to assess the performance of the regression model. The coefficient of determination (R-squared) is 93.04%, indicating that approximately 93.04% of the variance in the mango weight can be explained by the predictor variables. The adjusted R-squared [R-sq(adj)] accounts for the number of predictor variables in the model, and it is 92.76%. The R-squared value for prediction (R-sq(pred)) is 92.44%, representing the model's predictive ability.

The Analysis of Variance (ANOVA) on Table 2 showed the variance explained by the regression model and the residual error variance. The F-value for the regression model is 337.56, with a p-value of 0.000, indicating that the regression model is statistically significant, and at least one predictor variable is significantly related to the mango weight.

Overall, the regression analysis demonstrates a strong relationship between the predictor variables (mango length, pixel counts, and width) and the actual weight of carabao mangoes. The model exhibits good predictive performance and provides valuable insights into the factors influencing mango weight.

Table 1. Model summary of the generated MLR model.

S	R-sq	R-sq(adj)	R-sq(pred)
10.8212	93.04%	92.76%	92.44%

Table 2. Analysis of variance of the generated MLR model.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	158111	39527.7	337.56	0

L1	1	974	974.1	8.32	0.005
P1	1	462	462.2	3.95	0.05
W2	1	1392	1392.3	11.89	0.001
P2	1	7209	7208.8	61.56	0
Error	101	11827	117.1		
Total	105	169938			

The performance of the developed linear model was evaluated using a validation set consisting of 106 mangoes, and the results are presented in Table 3. The mean absolute percentage error for estimating mango weight using the multiple linear regression model was found to be 2.68% with an accuracy of 97.32%, which corresponds to approximately 2.68 g error for every 100 grams of mango weight. These results demonstrate that the image processing and analysis techniques, along with the multiple linear regression model, enable the accurate estimation of mango weight with a high level of accuracy. This validates the effectiveness and reliability of the developed machine learning approach for mango weight estimation, which can be valuable for ensuring the quality and standardization of mango exports in the agricultural industry.

Table 3. Validation testing of the developed linear model.

		Weight reading, g		Accuracy	
Number of samples	**Weigth range, g**	**Actual**	**Predicted**	**% Error**	**Percentage**
30	200-249	229.4	232	3.87	96.13
40	250-299	275.9	276.9	3.09	96.91
35	300-349	323.1	319.8	2.62	97.38
1	>350	350	346	1.14	98.86
			Means	2.68	97.32

CONCLUSION

The developed machine learning Multiple Linear Regression (MLR) model achieved a high accuracy of 92.44% in the training phase, using 106 training samples, and 97.32% accuracy in the validation phase, using an additional 106 validation samples. The model's regression equation provided insight into the relationship between mango weight and the predictor variables (length, width, and pixel counts). The findings from the study revealed that four predictors had a strong relationship with mango weight.

The results also showed that the developed MLR model can accurately estimate the weight of carabao mangoes using image processing and machine learning techniques, and that it provides a non-destructive and efficient method for classifying carabao mangoes based on their weight, which is required to meet the strict global market requirements.

The integration of machine learning techniques and precise digital measurements can streamline the estimation process, benefiting the mango industry and contributing to the improvement of the mango supply chain. By enhancing accuracy and efficiency, this research holds significant implications for the quality control and standardization of mango exports, thus reinforcing the Philippines' position as a key player in the global mango market.

ACKNOWEDGEMENTS

Sincere appreciation and wholehearted gratitude to the mango growers, farmers and cooperatives in the provinces of Pangasinan and Nueva Ecija for allowing us to have access to their areas and willingly provided the needed nformation. Also to the mango exporters in Taguig, Metro Manila, who accommodated the project team in the conduct of the experiments and availability of samples; and to all the individual and institutions who provided significant data for the completion of the study.

REFERENCES

1. Basak JK, Paudel B, Kim NE, Deb NC, Kaushalya Madhavi BG, et al. (2022). Non-destructive estimation of fruit weight of strawberry using machine learning models. Agronomy. 12: 2487.
2. Bermúdez AM, Padilla D, Torres G (2013). Image analysis for automatic feature estimation of the *Mangifera indica* fruit. Ing Desarro. 31: 84-104.
3. KC K, Yin Z, Li D, Wu Z (2021). Impacts of background removal on convolutional neural networks for plant disease classification *in situ*. Agriculture. 11: 827.
4. Menguito JJ, Suministrado DC, Zubia O, Keylord M, Quilloy E (2018). Weight determination of mango (*Mangifera indica* L.) by frequency analysis. Agric Eng Int CIGR J. 20: 186-194.
5. Navidi W (2011). Statistics for engineers and scientists. McGraw-Hill.
6. Sann SS, Win SS, Thant ZM (2021). An analysis of various image pre-processing techniques in butterfly image. Int J Adv Res Ideas Innov Technol. 6.

7. Teoh CC, Syaifudin ARM (2007). Image processing and analysis techniques for estimating weight of Chokanan mangoes. J Trop Agric Food Sci. 35: 183-190.
8. Thong ND, Thinh NT, Cong HT (2019). Mango sorting mechanical system uses machine vision and artificial intelligence. Int J Eng Technol. 11: 321-327.
9. Trunfio TA, Scala A, Giglio C (2022). Multiple regression model to analyze the total LOS for patients undergoing laparoscopic appendectomy. BMC Med Inform Decis Mak. 22: 141.