Oil Palm Fresh Fruit Bunches Maturity Prediction by Using Optical Spectrometer

Adilijiang Tuerxun¹*, Abdul Rashid Mohamed Shariff¹, Rimfiel Janius¹, Zulkifly Abbas², and Ghafour Amouzad Mahdiraji³

¹Department of Biological & Agriculture, Faculty of Engineering, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
²Department of Physics, Faculty of Science, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
³School of Engineering, Taylor’s University, Jalan Taylors, 47500 Subang Jaya, Selangor, Malaysia

*E-mail: mrahdel@gmail.com

Abstract
The accuracy in grading the maturity level of oil palm fruits is one of the important factors for the milling industry since it can directly affect the oil palm yield. Numerous studies have investigated for oil palm maturity classification using different approaches. However, there is a lack of clear guidance on which algorithm or technique to be applied for the oil palm maturity classification under different situations. In this research, several machine learning algorithms in the Weka data mining tool were applied for building classifier models and tested on samples of fresh fruit bunch (FFB). The maturity grade determined from the extracted sample features under the spectrometer is compared to the results of the manual standard grader of the milling industry. The validated result shows that simple lazy KStar algorithm performs better with 63% classifier model performance and the weighted average of receiver operating characteristics (ROC) curve area of 83.2%, which established the classifier model accuracy and enhances the model generalization when applied to the new set of test data for validation. The results of these studies make a contribution to the ongoing research in evaluating new techniques for oil palm maturity determination.

Keywords: Data Mining, Maturity Classification, Oil Palm, Spectroscopy

1. Introduction
Oil palm fruit is one of the major oil sources in our daily life. It provides vital food and has been concerned as a healthy product [1]. According to Oil World from Malaysian Palm Oil Board (MPOB), the world production of oil palm is 29% out of 203.91Mn T [2, 3]. Well, the determination of the maturity level of oil palm FFB contributes to the productivities of milling industries [4]. Efficient and cost-effective techniques are required for oil palm maturity detection in most of the private and government industries since the oil palm extraction rate (OER) is directly affected by the correct ripeness prediction [5, 6]. Expert graders rely mostly on mesocarp color and the total number of empty sockets to determine the ripeness of oil palm FFB in the classical method. The FFB is considered ripe if the mesocarp is orange in color. Mesocarp that is yellowish or yellow is considered as unripe or underripe.
Furthermore, other techniques consider bunches as ripe when 10% to 50% of the fruitlets have been detached from the bunch, while overripe when 50% to 90% of fruitlets are detached from bunch, and under-ripe if only 1 to 9 fruitlets have been detached from the bunch [7]. Fruits and vegetables have previously been categorized based on physical characteristics [1]. Optical sensors have been used for fruit quality detection in various horticultural crops [8-10]. However, those methods have some disadvantages such as high time consumption as well as high labor cost. There are some other researches based on the classification and categorization of fruits and vegetables using its physical appearance [11]. However, the physical characteristics of oil palm FFB are unique and hard to measure the exact quality using this technique, because there is variation in the size, shape, and weight [12, 13]. Also, other attributes can differ based on the type of fruits, the age of the tree, and other factors that cannot be easily measured [14]. Several methods applied as proposed by different authors in earlier research work. Methods involving fluorescence technology has been applied for the maturity level of the oil palm FFB. In order to compensate for all these shortcomings, a proposed artificial intelligence (AI), machine learning algorithms in Weka is adopted [15]. In this paper, our objective is to develop effective and accurate classification models for and predicting maturity level oil palm FFB.

2. Materials and Methods
A total of 106 oil palm FFB samples of different maturity levels were prepared for this experimental work, each of the samples belongs to one of the four ripeness categories which are unripe, under-ripe, ripe, and overripe. The detailed procedures adopted in the execution of this research is illustrated in the flowchart in Figure 1.

![Flowchart of the study.](image)

2.1. Sample Preparation
The FFB used in this study were provided by FELDA Agricultural Services Sdn Bhd, located at Jerantut, Pahang, Malaysia. All the samples were harvested from the oil palm of nigrescens type. Several data features were collected for this experimental research in compliance with the standard of the Malaysian Palm Oil Board (MPOB). The maturity level of oil palm FFB can be classified into six categories which are: black, under-ripe, ripe, overripe, empty, and rotten with the standard which was established by the
Malaysian Palm Oil Board (MPOB). This research focused on four categories (classes) which are: unripe, under-ripe, over-ripe, and ripe. These categories are determined by an expert grader using visual inspection in the oil palm milling industry, using standard established by MPOB. Figure 2 displays three of these samples: for under-ripe, ripe, and over-ripe respectively [6]. The samples were kept in a stable position when the measurements were taken. Each of the samples were scanned with the OPRID (Oil Palm Ripeness Detector), which is equipped with in-built transmission light and receptors to receive the reflected light.

Table 1. Grading standard used by MPOB.

| Grading Method | Total Number of Empty Fruitlets Sockets/Det | Mesocarp Colour |
|----------------|---------------------------------------------|-----------------|
|                |                                             | Yellow | Yellowish/Orange | Orange |
| Number of loose fruit sockets on the bunch | 0 | Unripe | Unripe | Ripe |
|                | 0–10 | Unripe | Under Ripe | Ripe |
|                | >10  | Unripe | Ripe | Ripe |
| Number of loose fruits on the ground | Ripe | 10%–50% of fruits detached from a bunch |
|                | Overripe | 50%–90% of fruits detached from a bunch |
|                | Under-ripe | 1–9 fruits detached from a bunch |

Figure 2. Oil palm FFB; (a) Under-ripe, (b) Ripe, (c) Overripe.

2.2. Instrument setup and data collections

The OPRiD is the portable prototype spectrometer, which can measure reflected energy from the surface of an object (FFB) with the use of ultraviolet (UV), visible light, and Infra-red (IR). This instrument contains 8 different wavelengths of light. In this research, four sensors with eight-band wavelength measurement capabilities were installed to detect the reflected light emitted from the 8 different LED modules. The usage of these 8 band sensor has several advantages, such as its portability, affordability, similarity shaped in external cover size for Oil Palm FFBs, as well as easy handling. The OPRiD spectrometer was used to scan each sample. The sensors detect the reflected wavelength bands at 365 nm, 460 nm, 523 nm, 590 nm, 623 nm, 660 nm, 735 nm, and 850 nm. In contrast, conventional systems employed use manual inspection for the oil palm ripeness based on the physical observations by a human grader without optical tools. These proposed sensors detect the amount of energy reflected from the outer surface of oil palm FFBs across various bands of the spectrum. The general system setup diagram for the experiment is shown in Figure 3.
2.3. Data pre-processing and model buildings
The data collected from the spectrometer sensors readings were saved in an excel file format in the memory of a device. The data file is converted to attribute relation file format (ARFF) or comma-separated value (CSV) as the recognized standard file formats by Weka 3.8 data mining tool. In order to manage the data precisely, 32 attributes were selected from the readings taken from the four sensors of the spectrometer from the reflected 8 band light spectrum from the 106 oil palm FFB samples. The randomized filter in the Weka data mining tool was applied for even random distribution of all class values across all folds within the training dataset as a pre-processing method [15, 16]. However, 10% of the data set after stratification is selected to be used as testing data in the build classifier model, to validate the generalization of the classifier model on a new set of instances. The remaining 90% is applied as training data for the classifier algorithm model development [17]. Moreover, five wavelengths with best attributes were selected (AmberD2, FredD3, RedD4, AmberD4, and DredD4) from the whole collected data with the application of subset attributes evaluator (wrapper method) to evaluate the effect on building model performance accuracy. The comparative study of the algorithm performance based on selected attributes for the study is achieved by considering the percentage, a weighted average of the receiver operating characteristics (ROC) area, and the percentage performance of the built model with applied classifier algorithm under two scenarios as illustrated in Table 2. Hence, the selected attribute performance reduces for 50% of the applied algorithm on the built model with an increase in accuracy when considering the percentage weighted-average ROC area, under five selected attributes.

3. Results and Discussion
The results obtained from the built model with the application of all the 32 selected attributes shown in Table 2. The Lazy KStar classifier algorithm was found to be the best and most accurate amongst all tested models across all Weka classifiers. The built classifier model detail result is displayed in Tables 3 and 4 accordingly. The multilayer perceptron has the highest performance value of 60% with an ROC area of 79.3% as compared with another classifier model applied. Contrarily, 80% performance accuracy is achieved from all 32 attributes applied, with an increase in the weighted average of the ROC area. The Lazy KStar algorithm has the best performance of 63%, percentage weighted-average ROC area of 83.2% as displayed in Table 2.
Table 2. Comparative analysis of algorithm on different attributes values.

| Classifier Algorithm       | All Attributes | Selected Attributes |
|----------------------------|----------------|---------------------|
|                            | Performance (%) | ROC Weighted Average (%) | Performance (%) | ROC Weighted Average (%) |
| Naïve base                 | 54.74          | 74.9                | 51.58           | 77.1                   |
| Logistic                   | 46.32          | 65.1                | 51.58           | 78.4                   |
| Multilayer Perceptron      | 54.74          | 76.7                | 60              | 79.3                   |
| Simple Logistic            | 56.84          | 80.7                | 54.74           | 76.3                   |
| Supporting Vector Machine (SMO) | 52.58      | 73.3                | 51.58           | 76.5                   |
| Lazy IBK                   | 60             | 73.5                | 54.74           | 69.4                   |
| Lazy KStar                 | 63             | 83.2                | 55.79           | 80.2                   |
| Cost sensitive             | 58.95          | 78.9                | 57.89           | 79.1                   |
| LMT                        | 56.84          | 80.7                | 54.74           | 76.3                   |
| Random Forest              | 58.95          | 82.8                | 54.74           | 80.5                   |

Table 3. Detail accuracy by class result.

|                  | Precision | Recall | F-Measure | MCC | ROC Area | Class   |
|------------------|-----------|--------|-----------|-----|----------|---------|
| 0.520            | 0.591     | 0.553  | 0.409     | 0.790 | Under Ripe |
| 0.650            | 0.591     | 0.619  | 0.512     | 0.819 | Over Ripe  |
| 0.826            | 0.679     | 0.745  | 0.659     | 0.894 | Unripe     |
| 0.556            | 0.652     | 0.600  | 0.461     | 0.809 | Ripe      |

Table 4 displays the confusion matrix obtained from the applied classifier-built model, to show the level of right prediction accuracy for the four class levels by the classifier model. The validation performance on the test data also established the accuracy and generalization of the built classifier model.

Table 4. Confusion matrix.

| Classified as | a | b | c | d | Classes |
|---------------|---|---|---|---|---------|
| a             | 13| 2 | 3 | 4 | a=Under-Ripe |
| b             | 2 | 13| 1 | 6 | b=Over-Ripe |
| c             | 6 | 1 | 19| 2 | c=Unripe   |
| d             | 4 | 4 | 0 | 15| d=Ripe    |

Furthermore, considering the displayed result of the confusion matrix as displayed in Table 4, the lazy Kstar was able to classify each of the four classes effectively considering the value of the true positive element of the diagonal matrix with reference to true negative values. Moreover, the ROC curve plots generated from each of the predicted classes are displayed in Figure 4 to 7. The result validated that a simple lazy KStar algorithm performs better with a 63% classifier model performance. The effective area under the curve covers a reasonable percentage for each of the classes which determines the classified model accuracy over the performance values as also displayed in Table 2.
Figure 4. The ROC curve plot for the classified Under-Ripe classified class with AUC of 79 percent.

Figure 5. The ROC curve plot for the classified Over-Ripe classified class with AUC of 81.9 percent.

Figure 6. The ROC curve plot for the classified Unripe classified class with AUC of 89.4 percent.
Figures 4 to 7 demonstrate that the result for the receiver operating characteristics (ROC) curve area for the underripe, overripe, unripe, and ripe fruit maturity categories are 79%, 81.9%, 89.4%, and 80.9% respectively. It can be seen from the figure that the model performance on the Unripe category presents the highest accuracy with the percentage at 89.4 among all four different maturity levels categories of oil palm fresh fruit bunches.

4. Conclusion
The proposed classified model in Weka 3.8 has successfully provided an accurate classification that could generate the efficient prediction of the oil palm FFB ripeness predictive model using the application of light reflection attributes measured from the portable prototype spectrometer. More in-depth studies on ripeness prediction using spectral data are obtained. Several machine learning algorithms were tested for selecting the significant spectrum band. The proposed classification model has been generated for the efficient prediction for oil palm FFB ripeness using the light reflection attributes measured from the spectrometer. The result shows that the simple lazy KStar algorithm performs overall accuracy with 63% and the weighted average of receiver operating characteristics (ROC) curve with 83.2 percent.

Acknowledgments
We acknowledge the LRGS Grant number 5526101 from the Ministry of Higher Education (MOHE) that helps in support on this research. We record our appreciation and thanks to the FELDA Agricultural Services Sdn Bhd, and Universiti Putra Malaysia for all support provided as well as the anonymous reviewers for their comments on improving this manuscript which is highly appreciated.

References
[1] Bensaeed O, Shariff A, Mahmud A, Shafri H, Alfatni M, editors. Oil palm fruit grading using a hyperspectral device and machine learning algorithm. IOP conference series: Earth and environmental science; 2014: IOP Publishing.
[2] Kushairi A, Nambiappan B, editors. Malaysia’s Palm Oil Supply and Demand for 2017 and Outlook for 2018. Conference: Palm Oil Internet Seminar (POINTERS); 2018.
[3] Kushairi A, Singh R, Ong-Abdullah M. The oil palm industry in Malaysia: thriving with transformative technologies. J Oil Palm Res. 2017;29(4):431-9.
[4] Harun NH, Misron N, Sidek RM, Aris I, Ahmad D, Wakiwaka H, et al. Investigations on a novel inductive concept frequency technique for the grading of oil palm fresh fruit bunches. Sensors. 2013;13(2):2254-66.
[5] Hazir MHM, Shariff ARM, Amiruddin MD. Determination of oil palm fresh fruit bunch ripeness—Based on flavonoids and anthocyanin content. Industrial Crops and products. 2012;36(1):466-75.

[6] Saeed OMB, Sankaran S, Shariff ARM, Shafri HZM, Ehsani R, Alfatni MS, et al. Classification of oil palm fresh fruit bunches based on their maturity using portable four-band sensor system. Computers and Electronics in Agriculture. 2012;82:55-60.

[7] Hazir MHM, Shariff ARM, Amiruddin MD, Ramli AR, Saripan MI. Oil palm bunch ripeness classification using fluorescence technique. Journal of food engineering. 2012;113(4):534-40.

[8] Alfatni M, Shariff A, Abdullah M, Marhaban M, Shafie S, Bamiruddin M, et al., editors. Oil palm fresh fruit bunch ripeness classification based on rule-based expert system of ROI image processing technique results. IOP Conference Series: Earth and Environmental Science; 2014: IOP Publishing.

[9] Alfatni MSM, Shariff AM, Shafri HM, Saaed OB, Eshanta OM. Oil palm fruit bunch grading system using red, green and blue digital number. Journal of Applied Sciences. 2008;8(8):1444-52.

[10] May Z, Amaran M. Automated oil palm fruit grading system using artificial intelligence. Int J Eng Sci. 2011;11(21):30-5.

[11] Yeow Y, Abbas Z, Khalid K. Application of microwave moisture sensor for determination of oil palm fruit ripeness. Measurement Science Review. 2010;10(1):7-14.

[12] Giovenzana V, Beghi R, Civelli R, Guidetti R. Optical techniques for rapid quality monitoring along minimally processed fruit and vegetable chain. Trends in Food Science & Technology. 2015;46(2):331-8.

[13] Shaarani SM, Cardenas-Blanco A, Amin MG, Soon N, Hall LD. Monitoring development and ripeness of oil palm fruit (Elaeis guineensis) by MRI and bulk NMR. International Journal of Agriculture and Biology. 2010;12(1):101-5.

[14] Tan YA, Low KW, Lee CK, Low KS. Imaging technique for quantification of oil palm fruit ripeness and oil content. European journal of lipid science and technology. 2010;112(8):838-43.

[15] Khan M, Quadri S. Evaluating various learning techniques for efficiency. International Journal of Engineering and Advanced Technology (IJEAT). 2012;2(2):326-31.

[16] Scott IM, Vermeer CP, Liakata M, Corol DI, Ward JL, Lin W, et al. Enhancement of plant metabolite fingerprinting by machine learning. Plant Physiology. 2010;153(4):1506-20.

[17] Aginako N, Lozano J, Quartulli M, Sierra B, Olaizola IG, editors. Identification of plant species on large botanical image datasets. EMR@ ICMR; 2014.