Advance Models for Camera-Vision Based Oil Content Prediction of Intact Oil Palm Fruits (*Elaeis Guineensis* Jacq) on Trees with Nondestructive Evaluation

Dinah Cherie[#], Sam Herodian [*], Tineke Mandang[*], Usman Ahmad[*]

[#] *Department of Agricultural Engineering, Faculty of Agric. Tech., Andalas University, Padang, 25163, West Sumatera, Indonesia.*
E-mail: dinah.cherie@gmail.com

* *Department of Agricultural Engineering and Bio-systems, Faculty of Agric. Tech., Bogor Agricultural University, Bogor 16680, West Java, Indonesia.*
E-mail: s_herodian@yahoo.com

***Abstract*— In this study, the correlation between oil palm fresh fruits bunch (FFB) appearance and its oil content (OC) was explored. The FFB samples were recorded from various distance (2, 7, 10, and 15 m) with different lighting spectrums (Ultraviolet: 280-380nm, visible: 400-700nm, and Infrared: 720-1100nm) and light intensities (600watt and 1000watt lamps). The FFB images were segmented and its color features were subsequently extracted to be used as input variables for modeling the OC of the sample. Twenty models were developed, one for every experiment-setup arrangement, using the MLP-ANN methods. Statistical engineering software was used to create the models. The number of FFB samples in this study was limited. Consequently, only five out of 20 developed models were selected and consider as valid to predict the FFB's OC. The five models were namely UV-15m, UV-10m, UV-2m, Vis1-7m, and IR1-10m. The coefficient of correlation for each model when validated was 0.962, 0.897, 0.993, 0.993, and 0.401 respectively. For bias and error, these selected models have root-mean-square error (RMSE) of 2.385, 3.707, 2.868, 3.795, and 6.374. The best of five selected models was developed for predicting FFB's OC in Vis1-7m arrangement, i.e. recording FFB images under visible light with 600watt lamps from 7 m of distance.**

***Keywords*— FFB, Oil Content, Nondestructive Evaluation, Machine Vision, MLP-ANN Methods**

I. INTRODUCTION

Oil palm crops (*Elaeis Guineensis* Jacq.) is among the principal commodities for Indonesian industry that support the country economy [1]. Its absorb millions of labors in all of its processes chains [2], from primary products to all the derivatives. This sector drives the economic growth of this nation and contributes as one of foreign currency generator for the country [3]. Furthermore, Indonesia exports of oil palm products and its derivatives, make up 77.13 % of the country agricultural sub-sector values in 2013 [1]. However, compared to the closest competitor, i.e. Malaysia and Nigeria [4], the productivity of oil palm in Indonesia is still by far left behind, due to mishandling the crops.

Currently, Indonesia only able to produce 1000 kg of crude palm oil (CPO) per hectare, compared to 1400kg by the Malaysian [5]. The CPO obtained by processing the oil palm fresh fruits bunches (FFBs) in mills. The yield and quality of the CPO will depend on the FFBs ripeness being processed. For example, processing raw or unripe FFBs will result in low CPO yield [6], while milling the over-ripe FFBs will produce low quality CPO [7]. Therefore, the FFBs should be processed when they reached the optimum ripeness [8].

Unlike the climacteric fruits [9], where the ripening process still continue after the fruits harvested [10], the physiological development [11] of oil palm FFB ceased when it was harvested, and the oil accumulation in the fruits mesocarp and kernel will subsequently stopped [12]. Harvesting will also trigger degradation process of the oil in fruits due to the transformation of oil into free fatty acids (FFA) [13]. Therefore, the FFBs should be transported into processing mills directly after harvest.

Based on this basis, increasing the productivity of oil palm in Indonesia can be achieved, among other, by harvesting the FFBs at the correct ripeness stages, where oil in fruits reached its maximum amount [14]. Moreover, harvesting the FFBs in optimum ripeness will ensure the CPO produced by mills has the prime quality, which correspondingly to the selling price of this product [15].

In order to get FFB at ideal ripeness upon harvest, the labors that performed the harvest should correctly identify the FFB on trees. This proved to be arduous and intricate tasks [16]. First, most oil palm plantations in Indonesia employed unskilled laborers with little or no education in order to pay wages at minimum amount [2]. Secondly, the incentive will be added to the labors wages using commissions system, based on the number, or the total weight of FFBs, harvested every month by the worker [3]. Therefore, in current practice, labors will try to harvest as many FFBs as possible with no regard to the ripeness stages of the FFBs harvested. This premature harvesting practice results in low yield of CPO per hectare since the oil in fruits has not yet reached the optimum level [17]. Consequently, there is an urgent need to find method for correctly estimated the ripeness of FFB prior to harvest. Moreover, this method should have the ability to perform assessment from a long-range in non- destructive manner.

Development in detection technology has enabled the assessment of intact fruits on trees as well as differentiations between ripe and non-ripe fruits [6-8]. The appearance of FFB, mainly the fruits color can be used to determine its ripeness, based on chromaticity analysis [17]. There are some correlations between the fruits hue and its oil content, as reported by Ismail et al. [18]. Using a quadratic poly-fit model, correlation between fruits color (hue) and its oil content can be established with coefficient of correlation of 95.41% [18]. Similar method was carried out by Razali et al. [15], using a statistical F-test and Annova. The FFBs sample's oil content were correlated with its color, measured in RGB color channel and HSI values from recorded image [19]. The results indicated that the value of hue of FFB image has a significant correlation level with its oil content. Using a quadratic poly-fit modeling, the correlation of hue from image and FFB oil content can be established with a R2 of 0.884. Compared with the report from Ismail et al. [18], this result was less accurate, indicating that the selection of statistical analysis to develop the model plays a significant influence to the accuracy of the model. The third study, reported by Ismail and Hudzari [20], used the hue color of FFB images to predict the timing of the harvest. A triangulation method was used to predict the harvest time of the FFB under consideration, by comparing its hue data with the calibration database. The developed method can predict the FFB harvest time with a success rate of 92.9%. The same method was validated by Razali et al., [15], with correlation of 92.39%. Nonetheless, all these previous studies have many drawbacks, such as limited assessment distance [21], and the lack of incoming light spectrum segmentation by means of photo-selective-transitive filters to highlight the optical properties of FFB studied [22].

Based on these, this study will established a non-destructive assessment technique for determining the oil content of intact FFB on trees, based on camera vision and optical auxiliary devices. Photo-selective-transitive filters will be used in this study to select the best optical properties of FFB (i.e. fruits color) for modeling its oil content. The technique will substitute the observation made by the harvester labor, with higher accuracy and consistency. In addition, using the optical auxiliary devices, this technique has the capability to observe the FFB under consideration from a longer distance (15m), compared to previous studies.

II. Materials and Methods

This study was carried out on May 2013, in PT. Nirwana Alam Lestari, a subsidiary oil palm plantation of PT. Astra Agro Lestari, Tbk. The study was done in the Nanga Bulik, Pangkalan Bun, Central Kalimantan, Indonesia (2 ° 05 ' N and 111 ° 15 'E). Geographically, the location was 20-50 meter above sea level, a hilly area with gradient of less than 25%. The local rainfall was between 2000-2500 mm / year, with daily air temperature of 23-32° C, and humidity of 81 - 92%. The FFBs samples were from cv. Marihat of 8 years old trees.

To select a FFB, first, the bunch was assessed by a panel of three observers. The FFB considered as ripe based on the observation of its color and detached fruitlets. If the panel agrees that the FFB under consideration has reached optimum ripeness, then the bunch selected as a sample. The camera was then positioned on the observer's spots, and its lens was directed toward the selected FFB sample on tree to record its image. The distance between the sample and camera lens was measured using a digital laser ranging (DLR130K, Bosch, Germany) with accuracy of 1 mm. the sunlight intensity was measured using a digital light intensity meter. Based on the light intensity, the camera was set accordingly, including the shutter speed, diaphragm opening, ISO number, and white balance. The lens used on camera has 30 times optical magnification capability, and when recording the FFB sample, its focal length was set so the camera field of view cover anterior section of the FFB with an area of 125 x 125 mm. The camera (EOS 60D, Canon, Japan) sensor then records the FFB image using 16 million pixels resolution. The FFB images were recorded in three replications. Subsequently, the FFB was harvested and brought to the laboratory to perform indoor photogrammetry, and chemical analysis as described in Cherie et al. [23].

The means of RGB and HSI data from FFB images were extracted using image processing software developed using the native Win64 advanced programming interface (API). The color data were then plotted into a graph and correlated with the sample's oil contents, as measured in laboratory. The value of color variable which strongly correlated with the changes of sample's oil content was selected for harvest determination of the FFB. The corresponding color variable will be used as the threshold value to decide if the FFB should be harvested [24].

For modeling the oil content of FFB's sample, the extracted color data from FFB's images were analyzed using statistical engineering software (SPSS 20.0, IBM, USA), by employing multi-linear-perceptron artificial-neural-networks analysis (MLP-ANN) [21]. The MLP-ANN method was selected based on its capability to predict correlations between complex and variety data where its variables have abstracts covariate. In addition, the ANN analysis has an excellent capability and flexibility for data processing, and user friendly features [25]. The ANN analysis using the MLP was selected in this study because it offer options for generating model that can predict oil content of FFB based on different types of input variables and multi-layer of coefficients [26]. The oil content model was then validated by comparing its results with laboratory analysis.

The selected input variables for the MLP-ANN modeling in this study were the color features extracted from the FFB's Images. The inputs were comprised of 15 variables, namely means of R, G, B, and H, S, I; the index values of R, G, and B; and the color ratio of R to G, R to B, G to B, G to R, B to R and B to G. The FFB's oil contents measured in laboratory analyses were used as dependent variables, or target output. The samples data were divided into two groups; the first group consists of 70% of data were used for training and model calibration, and the rest 30% of remaining data were used to validate the model. The MLP was set to produce 10 hidden-layers which produced best correlation model. The hyperbolic-tangent activation function was set during model calibration, in order let the software to automatically calculate the bias and weighted the input variables in the algorithm, in order to produced oil content prediction (OCP) model with the best coefficient of correlation (R2 close to 1). The hidden layers produced by the software contain unobservable network nodes (units). Each hidden unit is a function of the weighted sum of the inputs. The function is the activation function, and the values of the weights are determined by the estimation algorithm. The Hyperbolic-tangent function has the form of [25]:

$$Y_{(c)} = \tanh(c) = \frac{(e^{c}-e^{-c})}{(e^{c}+e^{-c})} \tag{1}$$

Where c is the hidden unit and $Y_{(c)}$ is the weighted sums of units in a layer.

The resulting model, developed by activating the identity function, was fed-back to the output target (oil contents) with an adjusted normalized correction value of 0.02. The identity function has the form of [25]:

$$Y_{(c)} = c \tag{2}$$

The input variables in this analysis were adjusted and normalized. They were subtracted by the minimum value and divided by their range using the equation of [25]:

$$x_{n_a} = \left[2 * \frac{(x-x_{min})}{(x_{max}-x_{min})}\right] - 1 \tag{3}$$

Where x_{n_a} is the adjusted normalized value of the input variable, x is the value of input variable, x_{min} is the minimum variable value, and x_{max} is the maximum variable value.

The input variables used for calibrating the model were transformed to the range of (–1, 1) using Eq. 3, to fit the Hyperbolic-tangent algorithm.

After adjusted normalized, the input variable values fall between -1 and 1. This is the required rescaling method for scale-dependent variables if the output layer uses the hyperbolic tangent activation function [25]. The correction option specifies a small number of ε that is applied as a correction to the rescaling formula; this correction ensures that all rescaled dependent variable values will be within the range of the activation function. In particular, the values -1 and 1, the which occur in the uncorrected formula when x takes its minimum and maximum value, define the limits of the range of the hyperbolic tangent function but are not within that range [25]. The corrected formula is [25]:

$$x_{n_{a_c}} = \left\{2 * \left[\frac{(x-(x_{min}-\varepsilon))}{((x_{max}+\varepsilon)-(x_{min}-\varepsilon))}\right]\right\} - 1 \tag{4}$$

Where $x_{n_{a_c}}$ the corrected is adjusted normalized value of the input variable; and ε is the correction value of 0.02.

It takes real-valued arguments and returns them unchanged. When automatic architecture selection is used, this is the activation function for units in the output layer if there are any scale-dependent variables.

When calibrating the oil content model, the MLP-ANN algorithm used batch training type, where it used information from all records in the training dataset. Batch training is often preferred because it directly minimizes the total error; however, batch training may need to update the weights many times until one of the stopping rules is met and hence may need many data passes. It is most useful for "smaller" datasets. Updating the synaptic weights can be performed only after passing all training data records.

Optimization algorithm is the method used to estimate the synaptic weights in this study by applying the scaled conjugate gradient, which can be performed only in batch training types. The optimization algorithm allowed the software to fine-tune the estimation model.

The scaled conjugate gradient algorithm can be used by specify the initial value of the lambda parameter to 0.0000005, and the sigma parameter to 0.00005. To avoid the local minimum loop, simulated annealing is used to weight vectors which randomly generated by the software. With the finding of the global minimum, during application of the optimization algorithm, the weight initialization and automatic architecture selection can be performed by specifying the number for the interval center and the interval offset greater than 0.

The synaptic weights results display the coefficient estimates that show the relationship between the units in a given layer to the units in the following layer. The synaptic weights are based on the training sample even though the active dataset is partitioned into training and validation data.

The model summary displays a summary of the neural network results by partition and overall, including the error, percentage of incorrect predictions, the stopping rule used to stop training, and the training time. The independent variable importance analysis performs a sensitivity analysis, which computes the importance of each predictor in determining the neural network

Stopping rules determine when to stop training the neural network if there is no decrease in error after the specified number of steps fulfilled.

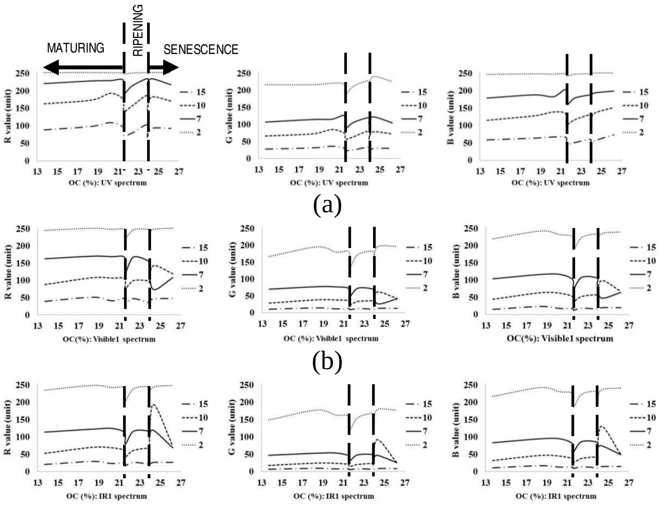
III. Results and Discussion

From the experiment, the results of measurements of FFBs samples in laboratory, it was found that the amount of oil contained in the fruit mesocarp compared to bunch overall weight, ranged from 13 to 26%. The amount of oil in the fruits mesocarp can be explained by its color [26], which correlated to the progress in the physiological developments of the fruit. In this study, the FFB with lower color intensity generally have less oil content (Fig. 1) and vice versa [27]. Differences of colors intensity in the recorded image of FFB were influenced by the light spectrum passes the camera lens filters. When FFB image recorded using photo-selective filter for Ultraviolet (UV) light, and passes lights with wavelength range between 280 and 360nm into the camera sensor, the RGB colors in the FFB images have almost uniform value for sample with oil content 13-18%. This trend was observed for FFB images recorded from 2, 7, 10, and 15 m. Similarly, when the images were recorded using photo-selective filter for visible (400-700nm) and IR (720-1100nm) lights, with illumination intensity of 6000 lux (Vis1 and IR1), the RGB colors of FFB with oil content between 13-18% were relatively equal. However, when images recording performed in higher light intensity (10000lux), the results showed nonlinear correlation between images color and sample oil contents, both for visible and infrared (Vis2 and IR2) filters. Nonetheless, the correlation cannot be explained by a linear regression, since the colors change irregularly. Based on its physiological development, the FFB samples in this group were still in the maturing phase (Fig. 1), were accumulation of oil and pigments in mesocarp progressed slowly.

However, significant changes in colors were observed when the sample oil content reached 21.6%. The changes of colors can be seen in FFB images features, recorded in all treatments except Vis2 and IR2 (Fig. 1). The changed strongly suggest that the physiological development of the FFB has entered the second stages [28], the ripening phase. In this phase, the oil and pigments accumulation in the fruits mesocarp occurred progressively, and the rate of colors change as recorded in the image were preponderant compared to FFB in the earlier stage The progressive accumulation of oil and pigments in the fruits mesocarp produced a more saturated color on the FFB appearance, and when its images was recorded, the RGB colors in images were lower, due to higher lights absorption by the fruits skin. At this point, the value of R, G, and B of the FFB images reach the minimum values in most treatments, except in Vis2 and IR 2, as well as the FFB images recorded in Vis1 treatment from 7m. This unique feature of FFB when its oil content reached 21.6% can be used as a reference that the FFB may be harvest immediately. The similar results had been reported by Cherie et al. [23].

As the oil in the fruits mesocarps progressed and reach maximum level, the FFB entered the third stages of its development, and the senescence of fruits commence. In this stage (Fig. 1), the oil in the mesocarp started to decay and disintegrate into the free fatty acid (FFA). FFB in this condition is considered as overripe, and undesirable for processing, since excessive FFA in its oil may deteriorate the whole quality of CPO in the processing line. The degrading process of the FFB will continue until most of the outer fruits in the bunch detached, and the FFB start producing rotten odor. In this study, this phase observed when the oil content in the FFB samples accumulated to 23.9%, indicated by the shift of RGB color trend in FFB images towards it oil content. Compared to the previous physiological stage, the trend of color change in FFB images oriented to the opposite direction.

Although the second stages of FFB development (ripening) can be observed through its color alteration, estimating the oil content correctly of FFB cannot be performed solely based on this information. The correlation between the color features in FFB images and its oil content were so poor their relationship cannot be explained through simple regression analysis. Furthermore, the color features of images from different FFB samples may have identical values, although their oil contents and development stages were different. Therefore the abstract relationships between the FFB oil contents and its color features were modeled through a neural network modeling analysis, namely the multi-linear-perceptron artificial-neural-network. The models developed based on the FFB images recorded by camera-vision with different lighting setups and recording distances. The color features of the images were extracted and transform to generate 15 input variables for training the models.



(c)

Fig. 1 The graphs explained the relationships between RGB values from FFB images and their mesocarp oil contents. The images of the FFB were recorded from 2, 7, 10, and 15 m (different line styles in the graphs) using a camera-vision with three different photo-selective filters, namely (a) UV:280-380nm, (b) Vis1:400-700nm and (c) IR1:720-1100nm under 6000lux lights.

To develop the model, 70% of data from the samples were used in model calibrations, where 15 colors features from the images were extracted and used as input variables. The developed models can predict the oil content of intact FFBs on trees with high classification accuracy. The MLP-ANN methods used for developing the models can explained the complex and abstracts relationships between the input variables (image's color features) and the oil content of the FFB samples. The output model analysis by the statistical engineering software provides several information of the model. The network structure displays the summary information about the ANN analysis. The information described including the dependent variables, number of input and output units, number of hidden layers and units, and activation functions. A diagram produced by the software displays the complete network diagram. However, since the number of covariates and factor levels used were substantial, the diagram becomes more difficult to interpret and was not display in the manuscript. The Synaptic weights of the model explained the coefficient estimates, indicating the relationship between the units in a given layer to the units in the following layer. The Network Performance of the model indicated an acceptable accuracy, both in model calibration and validation. In summary, the FFB's OCP model developed by means of MLP-ANN produced decent error or incorrect predictions, and the average overall relative error (relative to the mean model) is acceptable. The bias between the actual and prediction value of oil content explained in the residual prediction chart, where the residual-by-predicted-value chart for each scale-dependent variable was described.

The performance of OCP model upon calibration and validation explained in Table I and II respectively. For every experiment setup, one model was developed, and in total there were 20 models developed for predicting the FFB oil content, based on its recorded image. Upon the calibration, OCP models for all experiment setup performed comprehensively, and all coefficients of correlation (R2) were above 0.99. The best model obtained when the FFB images were recorded under UV and Vis2 lights, at 15 and 10 meter respectively. Both models have RMSE of 0.05 and 0.047 respectively. Others models performed similarly, where RMSE were significantly below 0.5.

TABLE I
THE CALIBRATION RESULTS FOR FFB OCP MODELS DEVELOPED WITH MLP-ANN USING 15 COLOR FEATURES FROM RECORDED FFB IMAGES. THE RESULTS INDICATED COEFFICIENT OF CORRELATION AND RMSE OF MODELS FROM ALL EXPERIMENT SETUPS.

Lighting Setup	Recording distance	Coefficient of Correlation (R^2)	Model RMSE
UV	15	1	0.051
	10	0.999	0.192
	7	0.999	0.075
	2	0.999	0.165
Vis1 600W	15	0.999	0.074
	10	0.999	0.192
	7	0.999	0.068
	2	0.999	0.142
Vis2 1000W	15	0.999	0.179
	10	1	0.047
	7	0.999	0.062
	2	0.999	0.070
IR1 600W	15	0.999	0.069
	10	1	0.071
	7	0.999	0.067
	2	0.999	0.077
IR2 1000W	15	1	0.133
	10	0.999	0.066
	7	0.999	0.149
	2	0.999	0.446

In order to test the model consistency, the rest 30% of the samples were validated using the developed models. The validation results will show which model is best-fit for in-field operations. The results showed that of 20 models developed using the MLP-ANN method; several models obtain high coefficient of correlation (R2), above 0.95. However, this parameter cannot be solely used as the main indicator to select the model. Other parameter, such as the RMSE, and residual should be considered to avoid selection of over-fit model which may produce false-positive or true-negative results. Considering RMSE and residual prediction value may lead to better choice of models that best-predict the FFB oil content.

Based on Table II, there are five models which produce low RMSE and high R2. The first model is for FFB's OCP with UV lighting and 15 m image recording distance. This model developed based on the extracted color features of the recorded FFB, and included all the 15 color features as input variables. To develop the model, the engineering statistical software employed 10 hidden layers to explain abstract correlations and variance among these variables, as well as their influence to the model results. The coefficients of input variables in every hidden layer, as explained in Table II, indicated how strong the corresponding variables influenced oil contents prediction model in this experiment setup. The model also incorporate biases as additional input variable, in order to enable accuracy and consistency enhancement. The biases improved coefficient of correlation of the model, and at the same time, minimize the RMSE from the validation results.

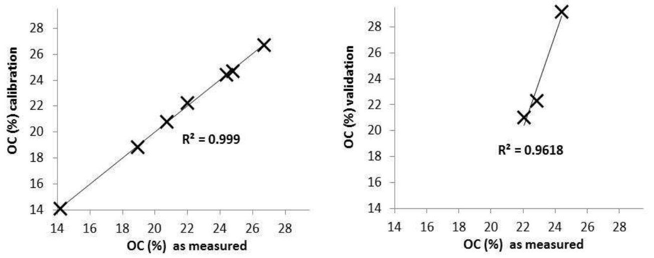
TABLE II
THE VALIDATION RESULTS FOR FFB OCP MODELS BASED ON 30% SAMPLING DATA EXCLUDED IN MODEL CALIBRATIONS. THE RESULTS INDICATED COEFFICIENT OF CORRELATION AND RMSE OF MODELS FROM EACH EXPERIMENT SETUP

Lighting Setup	Recording distance	Coefficient of Correlation (R^2)	Model RMSE
UV	15	0.962	2.385
	10	0.897	3.707
	7	0.582	2.603
	2	0.993	2.868
Vis1 600W	15	0.722	8.301
	10	0.931	15.510
	7	0.993	3.795
	2	0.270	9.760
Vis2 1000W	15	0.201	2.302
	10	0.401	9.289
	7	0.944	4.135
	2	0.871	7.211
IR1 600W	15	0.546	13.840
	10	0.940	6.374
	7	0.995	4.046
	2	0.799	18.252
IR2 1000W	15	0.688	2.391
	10	0.392	6.262
	7	0.891	6.632
	2	0.501	26.107

TABLE III
PARAMETERS ESTIMATED TO MODEL FFB OIL CONTENTS USING MLP-ANN METHOD BASED ON THE FFB RESPOND WHEN ITS IMAGES WERE RECORDED FROM 15 M UNDER UV LIGHT SPECTRUM

		Predicted										
		Hidden Layers										Output Layer
Predictor		H(1)	H(2)	H(3)	H(4)	H(5)	H(6)	H(7)	H(8)	H(9)	H(10)	OC
Input Layer Coefficient Factors	(Bias)	.402	-.417	-.467	.076	-.486	.349	.147	.253	-.456	.724	
	R	.446	-.084	-.194	-.213	.559	.168	-.030	.096	.398	.364	
	G	.407	-.181	.408	.134	-.097	-.351	-.150	.359	-.347	.573	
	B	-.087	.165	-.472	.346	-.392	.058	.369	.493	-.175	-.115	
	H	.056	-.220	-.480	.368	.443	-.014	.387	.530	.365	-.172	
	S	.217	.543	.043	-.389	.646	.285	.428	-.487	-.097	.276	
	I	.102	.562	.175	-.158	.778	-.237	-.060	-.423	-.256	-.055	
	RI	-.505	.418	-.188	-.101	.418	-.236	.233	.415	.240	.537	
	GI	.263	-.409	.641	-.163	-.082	.299	.546	-.422	-.066	.425	
	BI	-.448	.369	-.238	-.016	-.068	-.021	.344	.170	-.287	-.042	
	RG	-.158	.364	-.237	.115	-.154	.004	-.279	-.053	-.242	.541	
	RB	.102	.124	.510	-.155	-.056	.196	-.225	.401	-.275	.572	
	GB	-.422	-.121	.370	.176	-.260	-.279	-.079	-.392	-.318	-.304	
	GR	.361	.076	-.428	-.245	-.007	.014	.142	-.142	.451	-.198	
	BR	.440	.102	-.150	-.313	-.176	.070	-.208	-.297	.455	.040	
	BG	.247	-.014	-.262	-.334	.147	.020	-.174	-.352	-.235	-.549	
Hidden Layer Coefficient Factors	(Bias)											.132
	H(1)											-.186
	H(2)											-.686
	H(3)											.572
	H(4)											.205
	H(5)											-.769
	H(6)											.131
	H(7)											.112
	H(8)											.333
	H(9)											.366
	H(10)											-.850

The model performance upon calibration and validation explained in Fig. 2a and 2b respectively. Upon validation, the model predicted the FFB oil content of the sample above its actual value, as measured by laboratory analysis. This result indicated a false-positive offset of the model and showed that calibration of the model produce an over-fit model. This difference of model performance between the calibration and validation commonly occurred when using different type of samples data. This phenomenon possibly caused when this prediction model was built with limited number of samples. However, since the differences of RMSE and coefficient of correlation upon model calibration and validation was narrow (P<0.05), the models deviation is still within the tolerance limit, and therefore, this developed model can be accepted since it considered producing high accuracy and consistency.



(a) (b)

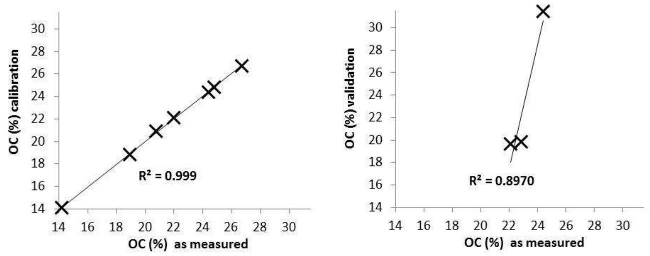
Fig. 2 The FFB sample's OCP based on model calibration (a) and prediction (b) displayed in scatterplot graphs. The graphs represent model performance for predicting FFB oil content based on its images when recorded from 15m under UV light spectrum.

TABLE II
PARAMETERS ESTIMATED TO MODEL FFB OIL CONTENTS USING MLP-ANN METHOD BASED ON THE FFB RESPOND WHEN ITS IMAGES WERE RECORDED FROM 10 M UNDER UV LIGHT SPECTRUM

Predictor		Predicted										
		Hidden Layer										Output Layer
		H(1)	H(2)	H(3)	H(4)	H(5)	H(6)	H(7)	H(8)	H(9)	H(10)	OC
Input Layer Coefficient Factors	(Bias)	-.382	-.516	.076	-.376	.051	-.102	-.315	-.102	.263	.481	
	R	-.218	.321	.238	-.089	.556	-.513	-.275	.292	.532	.186	
	G	.104	-.291	-.301	-.377	.357	.210	-.463	.244	.151	-.477	
	B	.518	-.455	-.392	-.339	.146	-.062	-.182	-.243	-.309	-.648	
	H	-.420	.336	.193	-.437	.479	.396	-.404	.028	-.189	-.251	
	S	-.163	.250	-.289	-.055	-.667	-.311	-.114	.214	-.072	.231	
	I	-.433	.202	.431	.046	.700	-.168	.380	.145	-.496	.034	
	RI	.075	.021	-.502	.356	.459	-.100	-.071	.105	.424	.369	
	GI	.392	.283	.124	.565	-.156	-.418	-.344	.272	.282	-.575	
	BI	.258	-.293	-.197	.186	.156	.147	-.029	-.162	.232	.145	
	RG	.386	.422	.447	.291	-.306	-.302	-.277	-.007	-.250	-.123	
	RB	-.083	.260	-.078	.297	-.168	-.279	-.364	-.144	.247	-.317	
	GB	-.338	-.422	-.516	.438	.409	-.303	.330	-.349	-.202	.086	
	GR	-.430	.080	.483	-.311	-.274	-.339	-.327	-.342	-.064	-.596	
	BR	.285	.206	-.003	.169	-.500	.447	.335	.076	-.159	.308	
	BG	-.506	-.174	-.059	-.288	.349	.115	.011	.366	.056	.428	
Hidden Layer Coefficient Factors	(Bias)											.473
	H(1)											-.280
	H(2)											-.225
	H(3)											.137
	H(4)											-.526
	H(5)											-.718
	H(6)											.240
	H(7)											.427
	H(8)											.090
	H(9)											-.192
	H(10)											-.663

The second selected model is for FFB's OCP with UV lighting and 10 m image recording distance. This model employed all extracted color features from FFB images, and included these features as predictor input. The model developed by the engineering statistical software has 10 hidden layers to explain correlations, variance, and influence of variables to the established model. Variables coefficients in hidden layers were explained in table 4. Model incorporated biases, and its accuracy and consistency enhanced. Low RMSE was observed in model when it was validated using second set of data, excluded when developing the model.

The model performance upon calibration and validation explained in Fig. 3a and 3b respectively. Upon validation, this model also predicted the FFB oil content of the sample above its actual value, thus an over-fit was indicated in this model, and as the results, false-positive and offset was observed in the validation results. Nonetheless, model RMSE and coefficient of correlation were still fall within the tolerance limit (P<0.05), thus this is considered as valid model. Model validations produce R2 of 0.897 with RMSE of 3.707



(a) (b)

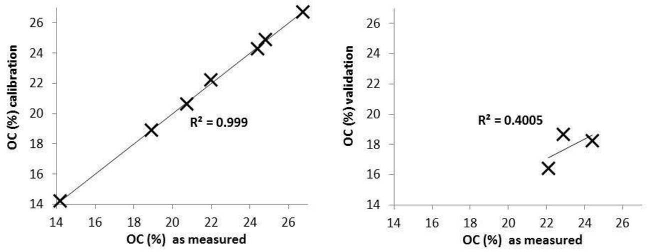
Fig. 3. The FFB sample's OCP based on model calibration (a) and prediction (b) displayed in scatterplot graphs. The graphs represent model performance for predicting FFB oil content based on its images when recorded from 10m under UV light spectrum.

TABLE V
PARAMETERS ESTIMATED TO MODEL FFB OIL CONTENTS USING MLP-ANN METHOD BASED ON THE FFB RESPOND WHEN ITS IMAGES WERE RECORDED FROM 10 M UNDER LOW POWER (600WATT) IR LIGHT

Predictor		Predicted										
		Hidden Layer										Output Layer
		H(1)	H(2)	H(3)	H(4)	H(5)	H(6)	H(7)	H(8)	H(9)	H(10)	OC
Input Layer Coefficient Factors	(Bias)	.263	-.060	.498	.070	.052	-.353	-.012	.083	.592	-.386	
	R	-.499	.571	-.398	.118	-.152	-.540	-.106	-.122	-.942	-.880	
	G	-.250	.411	-.125	-.630	.480	-.229	.483	-.133	-.604	.005	
	B	.062	.544	-.482	-.108	-.079	-.676	-.130	-.439	-.244	-.656	
	H	-.103	.741	-.814	.174	.085	.208	-.219	-.221	-.164	-.510	
	S	.145	.687	-.654	.411	-.193	-.003	.217	-.110	-.632	-.260	
	I	.103	.520	.021	-.245	-.191	.139	.071	.232	-.205	-.650	
	RI	.184	-.033	.109	.618	.020	.392	.448	.311	.375	-.562	
	GI	.148	-.646	.512	-.366	-.579	.574	.431	-.338	.597	.662	
	BI	-.398	.005	.030	-.491	-.177	-.056	.330	.231	-.420	-.481	
	RG	-.134	-.004	-.539	-.297	.586	.100	.267	.205	-.383	-.458	
	RB	-.104	.339	-.320	.037	.813	-.562	-.152	.194	.134	-.462	
	GB	.305	.035	.213	-.231	-.073	.135	.423	.004	.445	-.044	
	GR	-.304	-.266	-.083	.133	.069	.577	.043	-.290	.100	-.050	
	BR	-.289	.145	-.135	-.481	.036	.128	.487	-.055	-.517	-.061	
	BG	.228	.388	-.553	.438	.320	.105	-.410	-.213	.067	-.235	
Hidden Layer Coefficient Factors	(Bias)											.302
	H(1)											-.201
	H(2)											1.010
	H(3)											-1.443
	H(4)											.143
	H(5)											-.113
	H(6)											-.539
	H(7)											-.030
	H(8)											.079
	H(9)											-1.529
	H(10)											-1.243

The third selected model successfully developed to predict FFB's OC using lower intensity IR lighting (600 watt) with 10 m recording distance. The model used 10 hidden layers and 15 color features from its image as input variables. These variables extracted from the image using image processing software, and the model itself was developed using MLP-ANN methods, as built-in part of the engineering statistical software. Parameters in the model, i.e. coefficients of hidden layers, were explained in table 5.

The model performance upon calibration and validation explained in Fig. 4a and 4b respectively. In the validation, model predicted the FFB oil content below its actual value. This model show an over-fit behavior and true-negative prediction, as well as offset was observed in the validation results. Although model's coefficient of correlation (R2) only produce low relationship (0.401) between actual oil content of the FFB and its prediction results, the RMSE of the model (6.374) was still fall within acceptance limit ($P<0.05$), thus this model considered as valid.



(a) (b)

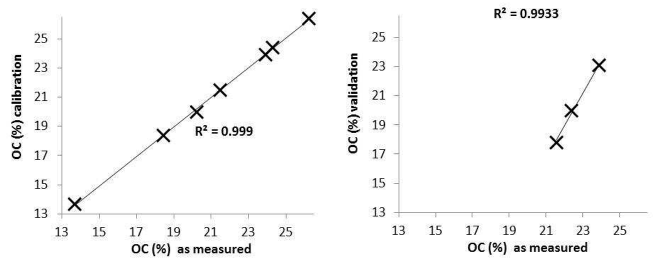
Fig. 4. The FFB sample's OCP based on model calibration (a) and prediction (b) displayed in scatterplot graphs. The graphs represent model performance for predicting FFB oil content based on its images when recorded from 10m under 600watt IR light.

The Fourth selected model successfully developed to predict FFB's OC using lower intensity of visible light (600 watt) with 7 m recording distance. Similar to the previous model, this model also used 10 hidden layers and 15 color features from its image as input variables. The model developed using MLP-ANN methods based on the color features extracted from its image by image processing software. Model's parameters including coefficients of hidden layers, were explained in table 6.

The developed model produced high coefficient of correlation upon calibration and validation. The model's R2 was 0.999 and 0.993 for calibration and validation, respectively. The relationship between prediction and actual values of FFB oil content were described in Fig. 5a and 5b. Validation results of the model produced lower prediction results of FFB oil content, compared to its actual value. The model indicated an over-fit behavior with true-negative and offset in the results. Nevertheless, this model RMSE was considered as minimal (3.795), while its prediction correlation was almost linear (R2=0.993). Both values were well within the tolerance limit ($P<0.05$), thus this model considered as valid.

TABLE VI
PARAMETERS ESTIMATED TO MODEL FFB OIL CONTENTS USING MLP-ANN METHOD BASED ON THE FFB RESPOND WHEN ITS IMAGES WERE RECORDED FROM 7 M UNDER LOW POWER (600WATT) VISIBLE LIGHT

		Predicted										
		Hidden Layer										Output Layer
Predictor		**H(1)**	**H(2)**	**H(3)**	**H(4)**	**H(5)**	**H(6)**	**H(7)**	**H(8)**	**H(9)**	**H(10)**	**OC**
Input Layer Coefficient Factors	(Bias)	-.470	-.489	.027	.244	.383	.076	-.040	.333	.147	-.271	
	R	.303	-.162	.434	.468	.471	-.057	-.037	.087	-.424	.190	
	G	-.152	-.153	-.117	.512	.422	.536	-.466	.419	-.188	.353	
	B	-.192	-.210	-.072	-.328	.353	-.365	-.433	-.302	.176	-.240	
	H	-.413	.012	.073	-.468	-.066	.680	-.173	.634	.003	-.007	
	S	-.061	-.102	.186	.230	.452	.171	-.576	-.378	.126	-.228	
	I	.273	-.387	-.110	.276	-.331	.433	-.205	.307	-.228	-.316	
	RI	.195	.377	.252	-.435	-.158	.393	-.025	.611	.263	.045	
	GI	.306	.307	.369	.378	.415	.454	-.266	-.287	.448	-.477	
	BI	-.443	-.803	.136	.428	.432	.354	.128	.142	.282	.111	
	RG	-.276	-.034	.270	.306	.465	.154	-.128	-.024	.323	.173	
	RB	-.170	.484	-.426	-.495	.166	.239	-.066	.187	-.020	-.095	
	GB	-.505	.417	-.047	-.079	-.346	.380	.035	-.147	.277	-.487	
	GR	-.309	.002	.043	.038	-.352	.287	.355	.237	-.320	-.067	
	BR	-.060	-.085	-.427	.507	-.493	-.491	-.453	-.500	.019	.108	
	BG	.417	.225	.056	-.247	.007	.219	-.483	.211	.058	.406	
Hidden Layer Coefficient Factors	(Bias)											.513
	H(1)											-.411
	H(2)											.710
	H(3)											-.159
	H(4)											-.546
	H(5)											-.007
	H(6)											-.652
	H(7)											.353
	H(8)											-.727
	H(9)											.041
	H(10)											-.086



(a) (b)

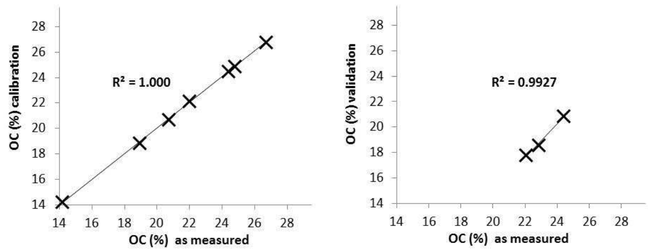
Fig. 5. The FFB sample's OCP based on model calibration (a) and prediction (b) displayed in scatterplot graphs. The graphs represent model performance for predicting FFB oil content based on its images when recorded from 7m under 600watt visible light.

The fifth selected model which predicts FFB's OC based on its image recorded under UV light from 2 m was successfully developed. Similar to other models, this model used MLP-ANN methods with 10 hidden layers based on its 15 color features extracted from its image. The color features were used as input variables upon model development and training, and parameters of the model were explained in table 7.

The coefficient of correlation (R2) of this model during calibration and validation were 1.0 and 0.993, respectively. The results were displayed in Fig. 6a and 6b. This model predicts FFB's oil content below its actual values. Therefore, this model suggests an over-fit trend, with true-negative and offset in the results. However, when compared with other developed models, the results showed that the R2 of the model was among the best from five selected models. Model R2 was 0.993 and its RMSE was 2.868.

TABLE VII
PARAMETERS ESTIMATED TO MODEL FFB OIL CONTENTS USING MLP-ANN METHOD BASED ON THE FFB RESPOND WHEN ITS IMAGES WERE RECORDED FROM 2 M USING UV LIGHT SPECTRUM

Predictor		Predicted										
		Hidden Layer										Output Layer
		H(1)	H(2)	H(3)	H(4)	H(5)	H(6)	H(7)	H(8)	H(9)	H(10)	OC
Input Layer Coefficient Factors	(Bias)	.312	-.111	-.345	-.014	.150	.207	-.620	.752	-.325	.489	
	R	-.430	.305	.259	.050	-.300	.222	-.122	-.441	.289	-.394	
	G	-.326	-.271	.287	.398	.135	.376	-.117	-.090	.553	.402	
	B	.070	-.446	.015	-.111	.521	-.002	-.155	.244	.053	-.302	
	H	-.302	-.285	.198	.340	-.194	-.346	-.441	.307	.235	.502	
	S	.352	.324	.033	-.203	.401	-.497	-.027	.324	.442	.462	
	I	.259	-.458	.373	-.410	-.607	-.307	-.302	-.490	-.430	-.100	
	RI	.151	.076	.171	-.184	.434	-.309	-.189	-.516	.300	-.532	
	GI	-.208	.203	-.326	.382	.644	.201	.401	.229	.404	.102	
	BI	.352	-.408	-.430	.322	-.479	-.308	.235	.369	-.540	.423	
	RG	.474	.336	-.196	.182	-.526	.515	-.145	-.275	-.275	-.295	
	RB	.160	-.366	.458	.326	.057	.353	-.388	-.499	.005	-.169	
	GB	-.109	.166	-.314	-.255	-.112	.376	.120	-.318	.196	.200	
	GR	-.279	-.329	.111	-.549	-.167	-.304	-.031	.435	-.007	.065	
	BR	-.025	.337	.073	-.505	-.023	-.081	.349	-.276	-.044	.271	
	BG	-.092	.180	-.298	.132	-.073	-.535	-.249	-.213	.212	-.181	
Hidden Layer Coefficient Factors	(Bias)											.134
	H(1)											-.019
	H(2)											-.207
	H(3)											.117
	H(4)											.645
	H(5)											.278
	H(6)											-.548
	H(7)											-.772
	H(8)											.522
	H(9)											.436
	H(10)											.237



(a) (b)

Fig. 6. The FFB sample's OCP based on model calibration (a) and prediction (b) displayed in scatterplot graphs. The graphs represent model performance for predicting FFB oil content based on its images when recorded from 2m under UV light.

IV. CONCLUSIONS

From 20 experiments setup to develop OCP models for oil palm FFB, only five models were considered as valid, based on its coefficients of correlations, RMSE and validation results. The experiment setup was set to determine which image recording condition produce best model accuracy to predict intact FFB on trees, using nondestructive evaluation by means of machine-vision assessment. The models were developed from color features of FFB's recorded images, where it was used as input variables prediction. Statistical engineering software was used to develop the models by employing MLP-ANN algorithm, with 10 hidden layers. Five best models were obtained from FFB imaging under UV (recorded from 15, 10, and 2m), Visible (recorded at 7m), and IR lights (recorded at 10 m). Biases were observed in the models, due to limited samples used in this experiment. However, in the general, the five selected models performance were accepted, since their RMSE and coefficient of correlation during calibration and validation was still within the tolerance limit ($P<0.05$).

ACKNOWLEDGMENT

The authors express their noble gratitude to Dr. Eng. Muhammad Makky for his contribution in this study. The Authors would extend their gratitude to PT. Astra Agro Lestari, DP2M Dikti, Bogor Agricultural University, Andalas University, and DGHE DIKTI for providing funding for this research and doctoral scholarship to the first author.

REFERENCES

[1] PPHP. 2014. Agricultural commodity of the Republic of Indonesia: export – import statistics 2011 – 2013. Journal of statistics export-import of the agricultural commodity. ISSN 2337 – 9578.

[2] BPS. 2015. Official statistics report: Mei 2015. Data and information center for employment. The development agency for research and information. Ministry of labor of the republic of Indonesia

[3] Rival A and Levang P. 2014. Palms of controversies: Oil palm and development challenges. Bogor, Indonesia: CIFOR.

[4] RSPO. 2013. Principles and Criteria for the Production of Sustainable Palm Oil. Endorsed by the RSPO Executive Board and Accepted at the Extraordinary General Assembly by RSPO Members. pp.1-71

[5] Bruntse-Dahl, R. 2011. Malaysia and Indonesia bolster defense of palm oil industry to West. The Guardian, 20 May.

[6] Thoriq A., Herodian S., Makky M., Sutejo A., and Cherie D. 2012. Application of Non-destructive spectroscopy technique for determination of Oil Palm Fresh Fruits Bunch (FFB) maturity. The role of agricultural engineering to support food and energy security with environmental insight: Proc. of ISAE national seminar, ed. Hendrawan Y., Al Riza D.F., Dewi S.R., Sugiarto Y., Ubaidillah, and Fatchurrahman D. Brawijaya University, Malang, 30 Nov.-2 Dec. Indonesia. pp.178-185.

[7] Cherie D., Herodian S., Makky M., Mandang T., Ahmad U., and Thoriq A. 2012. Application of Photogrametric for oil palm (Elaeis guineensis Jacq) Fresh Fruits Bunch (FFB) ripeness. The role of agricultural engineering to support food and energy security with environmental insight: Proc. of ISAE national seminar, ed. Hendrawan Y., Al Riza D.F., Dewi S.R., Sugiarto Y., Ubaidillah, and Fatchurrahman D. Brawijaya University, Malang, 30 Nov.-2 Dec. Indonesia. pp. 166-177.

[8] Makky M., Herodian S., and Subrata I.D.M., 2004. Design and Technical test of visual sensing system for palm oil harvesting robot. Proc. Int. Seminar on Advanced Agric. Eng. Farm Work Operation, August 25-26, Bogor, Indonesia. pp.582-592.

[9] Makky M., Paschalidis K.A., Dima K., and Mangganaris A. 2014. Harnessing Untapped Bio-Ethylene Sources from Tomatoes Climacteric Effluent. Proc. International Conference on Agricultural, Environmental and Biological Sciences (AEBS-2014) April 24-25, Phuket, Thailand. pp. 27-32.

[10] Makky M., Paschalidis K.A., Dima K., and Mangganaris A. 2014c. A New Rapid Gas Chromatographic Method for Ethylene, Respirational, and Senescent Gaseous Production of Climacteric Fruits Stored in Prolonged Low Temperature. Proc. International Conference on Agricultural, Environmental and Biological Sciences (AEBS-2014) April 24-25, Phuket, Thailand. pp. 21-26.

[11] Makky M., Paschalidis K.A., Dima K., and Mangganaris A., 2015. Tomato Fruits (Solanaceae Lycopersicon esculentum Mill.) Feedback Mechanism in The Presence of Exogenous Ethylene under Prolonged Chilling Temperature Storage. Journal of Nutrition & Pharmacy Research, 1(1):4-12. International Institute of science and industry research, Australia.

[12] Makky M., Herodian S., Cherie D., Ahmad U., and Mandang T. 2012. Spectroscopy and Photogrammetric techniques for assessing physicochemical properties of oil palm (Elaeis guineensis Jacq) Fresh Fruits Bunch (FFB). Proc. Research Dissemination Seminar, International Convention Center (IICC) – Bogor Agricultural University,, Bogor, 10-11 December, Indonesia.

[13] Makky M., Soni P., and Salokhe V.M. 2012. Machine Vision Application in Indonesian Oil Palm Industry. The Asian Forum of 2012 CSAM (Chinese Society for Agricultural Machinery) International Academic Annual Meeting. "Innovation, Win-win, Development". October 27th -30th, Hangzhou, China.

[14] Makky M. and Soni P., 2013. Towards sustainable green production: exploring automated grading for oil palm fresh fruit bunches (FFB) using machine vision and spectral analysis. Int. J. Advanced Sci. Eng. Information Technol., 3(1), 1-7.
[15] Razali M. H., Ismail W. I. W., Ramli A. R., Sulaiman N., Harun M. H. B. 2011. Technique on simulation for real time oil palm fruits maturity prediction. African Journal of Agricultural Research Vol. 6(7), pp. 1823-1830, 4 April, 2011. ISSN 1991-637X.
[16] Makky M. and Soni P., 2013. Development of an automatic grading machine for oil palm fresh fruit bunches (FFBs) based on machine vision. Computers Electronics Agric., 93, 129-139.
[17] Ismail W. I. W., Bardaie M. Z., Hamid A.M.A. 2000. Optical properties for mechanical harvesting of oil palm FFB. J. Oil Palm Research Vol. 12 No. 2, December 2000, p. 38-45.
[18] Ismail W.I.W., Razali M.H., Ramli A.R., Sulaiman M.N., dan Harun M.H.B. 2009. Development of imaging application for oil palm fruit maturity prediction. Engineering e-Transaction (ISSN 1823-6379). Vol. 4, No. 2, pp 56-63
[19] Hudzari M.R., Ishak W. I.W., Abd. Rahman R., Md. Nasir S., dan Mohd. Haniff H. 2009. Development of image based modeling for determination of oil content and days estimation for harvesting of FFB. International J. Food Engineering, Vol. 5, Iss. 1, Article 1
[20] Ismail W.I.W. dan Hudzari R.M. 2010. Outdoor color recognition system for oil palm fresh fruit bunches (FFB). International Journal of Machine Intelligence, ISSN: 0975–2927 (2), Issue 1, 01-10.
[21] Makky M. and Soni P., 2014. In situ quality assessment of intact oil palm fresh fruit bunches using rapid portable non-contact and nondestructive approach. J. Food Eng., 120, 248-259.
[22] Makky M., Soni P., and Salokhe V.M., 2014. Automatic nondestructive quality inspection system for oil palm fruits. Int. Agrophys., 28, 319-329.
[23] Cherie D., Herodian S., Ahmad U., Mandang T., and Makky M. 2015. Optical Characteristics of Oil Palm Fresh Fruits Bunch (FFB) Under Three Spectrum Regions Influence for Harvest Decision. IJASEIT 5(3): 104-112.
[24] Hudzari R.M., Ishak W.I.W., Noorman M.M. 2010. Parameter acceptance of software development for oil palm fruit maturity prediction. J. Software Engineering, 2010. ISSN. 1819-4311.
[25] IBM. 2012. IBM SPSS Statistics Standard. The Risk of Using Spreadsheets for Statistical Analysis. USA
[26] Ishak W.I.W., dan Hudzari R.M. 2010. Image based modeling for oil palm fruit maturity prediction. J. Food, Agriculture & Environment Vol.8 (2): 469 – 476.
[27] S. Zuraka, “Grading of FFB palm oil based on global colour,” BSc. Thesis, Faculty of Information Technology & Quantitative Sciences, Universiti Teknologi MARA, Oct 2007.
[28] Furferi R., Governi L, and Volpe Y. 2010. ANN-based method for olive Ripening Index automatic prediction. Journal of Food Engineering 101 (2010) 318–328.