Coloured Edge Maps for Oil Palm Ripeness Classification

Wei Yuen Teh
weiyuen@prioritydynamics.com
Ian K.T. Tan
ian.tan1@monash.edu

Priority Dynamics
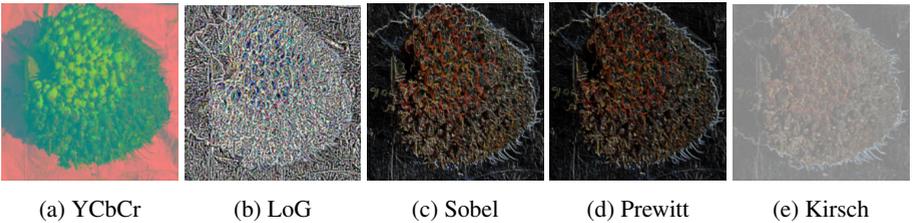Subang Jaya, Malaysia
Monash University
Malaysia

(a) YCbCr   (b) LoG   (c) Sobel   (d) Prewitt   (e) Kirsch

Figure 1: Sample of each input feature

Abstract

The task of grading oil palm bunches by ripeness poses a number of significant challenges for computer vision. The small difference in hue between ripe and unripe bunches means that colour-based models are susceptible to errors when presented with images shot in novel lighting conditions. In this paper, we investigate the effectiveness and performance characteristics of coloured edge maps when used as an input feature to a Convolutional Neural Network (CNN) by comparing the Laplacian of Gaussian, Sobel, Prewitt, and Kirsch edge extraction techniques. We show that under normal lighting conditions, coloured edge maps are able to match the performance of fully-coloured images. More notably, they significantly outperform fully-coloured images when variance to lighting is applied. When images are darkened or brightened, classification accuracy for fully-coloured images drops by 19.89% vs only 4.97% on average for the coloured edge map methods tested. This is of major benefit in commercial applications, where images are often captured by a multitude of devices under different lighting conditions, leading to potentially unreliable performance when fully-coloured images are used. The code used for this paper will be made available at https://github.com/weiyuen.

1 Introduction

Palm oil, which is extracted from the fruits of the oil palm tree, has a variety of common uses, including in the production of consumer edibles, pharmaceuticals, and cosmetics, and increasingly as a biofuel. Demand has been steadily rising over the past decade, with global consumption increasing from 41 metric tons in 2007 to 77 metric tons in 2020 [2].

Conventionally, oil palm fresh fruit bunches (FFBs) are visually inspected to determine ripeness. For the nigrescens variety of oil palm (which makes up the majority of oil palm production), unripe kernels are black, and they turn dark purple just as they are about to ripen [14]. This is not a significant visual difference, which makes the task of classifying FFB ripeness challenging for anyone but experts. Harvesting is typically done at fixed intervals, which results in harvests consisting of a mix of FFBs at different levels of ripeness. As such, the sub-optimal timing of a harvest can result in significantly reduced yields, with Mohanaraj and Donough [6] showing that a 5-day delay resulted in a 5.8% reduction in oil yield. This poses a problem as harvesters are compensated by yield, biasing them towards classifying their harvests in a way that maximizes yield. This results in the need for an independent expert in classification to be present at every harvest to help grade the FFBs for the harvesters. However, these experts are typically affiliated with downstream mill owners, and as such their results are often contested by the plantation owners upstream. Thus, the ability to automate the classification process could significantly streamline the harvesting process and reduce costs for the industry.

The visual similarity that makes manual inspection challenging also poses problems for conventional computer vision techniques. Primarily, the small difference in hue between ripe and unripe FFB leads to models that are susceptible to changes in camera hardware and lighting conditions. Kernels that appear black (unripe) under low light can often appear dark purple under brighter light (ripe). A shadow cast over an image or a camera that has a processing pipeline biased towards oversaturating images (both rather common occurences) can thus significantly change the inferences of conventional colour-based models.

Another often overlooked feature that correlates with oil palm ripeness is the change in textural appearance. Tan et al. [12] have shown that as the FFB ripens, changes in the number of leaves and empty fruitlet sockets results in a change to edge features. In their work, grayscale edge maps were used. Since colour also plays an important role in FFB classification, we have instead set out to explore the effectiveness and performance characteristics of coloured edge maps, which preserve residual colour information around edges. We investigate if the reduction in lighting information relative to fully-coloured images is able to bias a CNN to rely less on lighting, and thus improve generalization.

2   Related Work

2.1   Oil Palm Ripeness Classification

Significant work has been done in this field, with both conventional (RGB) and non-conventional imaging techniques employed successfully. A summary of past results is presented in Table 1.

Among the conventional imaging techniques, Fadilah et al. [3] achieved an accuracy of 86.67% by feeding extracted hue values into a 3-layer MLP (multilayer-perceptron). However, the images in their dataset were captured in a semi-controlled environment, where the camera was placed 35cm away from each palm fruit bunch and shot under controlled lighting conditions. Their images also had to undergo manual segmentation to isolate parts of the FFBs.

In 2016, Shabdin et al. [11] used hue, saturation, and intensity values as inputs to an artificial neural network and obtained an accuracy of 70%. Their work only involved two classes (underripe and ripe) and images were captured in a controlled environment.

Table 1: Related work involving oil palm ripeness classification

Year	Paper	Image Type	Method	Classes	Accuracy
2012	[3]	RGB	Hue values & 3-layer MLP	4	86.67%
2016	[11]	RGB	HSI as inputs to ANN	2	70.00%
2019	[10]	RGB	Extracted colour components from individual fruitlets & SVM	3	92.50%
2020	[15]	RGB	Linear segmentation of mean hue	2	85.00%
2021	[12]	RGB	Mean RGB ratio, red-colour ratio, Canny edge ratio	3	64.67%
2012	[4]	Spectral	Blue-Red Fluorescence Ratio as input to Decision Tree	3	89.70%
2012	[9]	Spectral	Four-band sensor & quadratic discriminant analysis	3	85.00%

Anindita et al. [10] then achieved 92.5% accuracy in 2019 using extracted colour features and SVMs. However, unlike previous work, the model required input images of individual fruitlets instead of full FFBs, which poses a problem as the fruitlets in an FFB can ripen and change in hue at different rates, making it a less reliable indicator of overall FFB ripeness.

In 2020, Wong et al. [15] used linear segmentation to classify bunches by mean hue value, achieving an accuracy of 85%. Unlike the previously discussed methods, their dataset consisted of images shot outdoors under uncontrolled lighting conditions. However, predictions were only made across two classes (ripe/unripe).

Most recently, Tan et al. performed classification by extracting mean RGB values as well as colour ratios. Additionally, they were the first to explicitly use edges as a supplementary feature to colour data through the use of Canny edge detection. This resulted in an accuracy of 64.7% over 3 classes. Like [15], lighting in the dataset was also uncontrolled.

Among the non-conventional imaging techniques, Hazir et al. [4] showed that the Blue-to-Red Fluorescence Ratio obtained by using UV light as an excitation light source could be input into a Decision Tree to obtain 89.7% accuracy in 2012. That same year, Saeed et al. [9] used a portable four-band sensor system and quadratic discriminant analysis to achieve 85% accuracy.

The drawback of these non-conventional imaging approaches is the requirement of specialized equipment and/or the need to conduct testing in controlled environments. Among the conventional imaging techniques, high accuracies have been achieved either through semi-controlled/controlled data gathering conditions [3, 10, 11], or by only considering two classes for evaluation [15]. Despite this, we believe our work can most closely be compared to that of [15] and [12] as these have been the only work to use conventional RGB images captured in uncontrolled conditions.

2.2 Coloured Edge Maps

Extensive work has been done on extracting edges from coloured images. However, these approaches have mostly been focused on producing grayscale edge maps by combining the extracted edges from each (R, G, B) colour channel [7]. Relatively little work has been done on edge extraction techniques that produce coloured edge maps, with the work done by Bora

[□] being one of the few exceptions. In brief, their approach involves first converting an RGB image to the HSV colourspace before performing edge detection on the V channel. The V channel is then replaced with the edge map and the image is converted back to RGB. While our implementation in this paper differs from that of [□], as we will discuss in the following section, the visual characteristics of the resulting coloured edge maps are similar.

3   Input Features

As discussed in Section 2.2, edge maps are most commonly used in their grayscale form. However, colour is an important component of oil palm ripeness classification, leading us to investigate the effectiveness of coloured edge maps. Here, we extract coloured edge maps by using coloured images as an input to classical edge detection techniques, where the method is applied to each colour channel of the image. Thus, our implementation is similar to that discussed in [□], sans the recombination of colour channels at the end (which would produce grayscale edge maps).

To investigate the effectiveness of various extraction techniques for coloured edge maps, the following four methods were used: Laplacian of Gaussian, Sobel, Prewitt, and Kirsch. They were chosen as the grayscale versions of these algorithms are amongst the most commonly used edge extraction techniques today. As a benchmark for comparison, fully-coloured images in the YCbCr colourspace were used. The YCbCr colourspace was chosen in particular as it was shown to be the best performing colourspace for the classification of oil palm ripeness by Sabri et al. [□]. A sample of each input feature is shown in Figure 1, and the conversion process for each feature is discussed in detail below.

3.1   YCbCr

The YCbCr colourspace consists of a luminance channel "Y", which denotes light intensity, as well as two chrominance channels "Cb" and "Cr", which denote the blue-difference and red-difference chroma components respectively. The *cvtcolour* function from the OpenCV library was used for our experiments.

3.2   Laplacian of Gaussian

As its name implies, the Laplacian of Gaussian (LoG) operator involves taking the Laplacian of an image that has been smoothed by a Gaussian filter (to reduce sensitivity to noise). The Laplacian operation computes the second spatial derivative of an image, highlighting regions where this derivative changes rapidly, as is common in edges. In our work, we use the filter f (eq. 1) as a discrete approximation of the Laplacian operator. A Gaussian kernel size of 5x5 was used for smoothing.

$$ f = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{1} $$

3.3   Sobel

The Sobel operator computes the first derivatives of an image in the horizontal and vertical axes by convolving the input image $\boldsymbol{A}$ with two kernels $\boldsymbol{G}_x$ and $\boldsymbol{G}_y$ (equation 2). The gradient
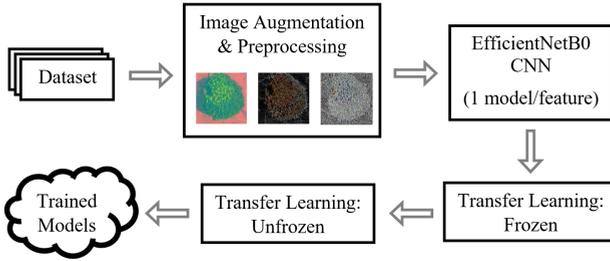
Figure 2: CNN training process

magnitude can then be calculated using Pythagoras' Theorem.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * A \quad , \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \tag{2}$$

3.4 Prewitt

The Prewitt operator utilizes different G_x and G_y kernels when compared to the Sobel operator, where the 2 and -2 values are replaced by 1 and -1 respectively, thus giving equal emphasis to both corner and edge pixels. The two processes are otherwise identical.

3.5 Kirsch

Kirsch edge detection computes first order derivatives and is typically implemented with 8 convolution kernels, formed by rotating a single kernel across the 8 compass directions as shown below. The complete edge map is then obtained by taking the maximum magnitude of each point across all 8 directional edge maps.

$$G_1 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} * A \quad \dots \quad G_8 = \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} * A \tag{3}$$

4 Methodology

In this section we describe the dataset and process used to train the EfficientNetB0 CNNs. A flowchart summarizing the main training steps is shown in Figure 2.

4.1 Dataset

The dataset consists of 328 RGB images of palm fruit bunches at various resolutions and is an expanded version of that used by Tan et al. (containing additional images) [12]. The images are in JPEG format and taken with multiple devices under various lighting conditions. All images contain one harvested palm fruit bunch, laid upon backgrounds that vary in colour and texture across the dataset. The dataset is split into 3 classes: Unripe, Underripe, and Ripe, and all images were labeled by experts. One sample from each class is shown in

(a) Unripe   (b) Underripe   (c) Ripe

Figure 3: Sample from each class in the dataset

Table 2: Image augmentation parameters

Parameter	Value
Rotation Range (°)	30
Shear Range (°)	15
Vertical Mirroring	✓
Horizontal Mirroring	✓

Figure 3. We split the dataset into training, validation, and test sets, with 246, 32, and 50 images in each set respectively.

4.2 Image Augmentation & Preprocessing

Apart from applying the feature conversions described in Section 3, additional preprocessing was also performed in the form of image augmentation to help improve performance and mitigate the impacts of the small dataset size. This involved applying random degrees of affine transformations such as rotation, mirroring, and shear to each image at each epoch of training. Parameters used for image augmentation are listed in Table 2. Images were then resized to a 500x500 resolution before being input to the model.

4.3 Modeling

The EfficientNetB0 [13] CNN architecture was used as a base model to evaluate the effectiveness of the various input features. This particular architecture was chosen as it was the smallest network among the EfficientNet series of networks (5.3M parameters), and was best suited to the (small) size and complexity of our dataset. A total of 5 of these models were trained, one for each input feature. Due to the size of our dataset, we opted to perform transfer learning by fine-tuning the weights of an EfficientNet pre-trained on the ImageNet dataset, allowing us to take advantage of the pre-trained feature extraction filters. Note that while we have taken multiple mitigative measures, overfitting is still a concern with our method given the small size of our dataset by deep learning standards.

To adapt the base model to our task, a few minor changes were made. First, the top layer of the pre-trained EfficientNetB0 was removed to account for the difference in the number of output classes. The top of the network was then rebuilt with a global average pooling layer, and an output layer with 3 neurons (one for each class). The softmax activation function was used for the output layer.

4.4 Transfer Learning

Transfer learning was performed in two stages: Frozen and Unfrozen.

Frozen: All layers except the batch normalization layers in the base model were frozen (set to not trainable) and training was performed over 25 epochs at a learning rate of 1×10^{-4}. The Adam optimizer [5] was used and loss was calculated using categorical crossentropy. This phase of training allows the rebuilt top layers to undergo a few rounds of training without disrupting the weights of the base model.

Unfrozen: All layers were unfrozen and training was performed over 120 epochs at a reduced learning rate of 1×10^{-5}. As before, the Adam optimizer was used and loss was calculated with categorical crossentropy. A lower learning rate was used to prevent disruption of the pre-trained weights, and validation loss was observed to converge after around 100 epochs. A batch size of 16 was used for both stages.

5 Experiments & Results

To determine the performance characteristics of coloured edge maps, we conducted 3 experiments. In Section 5.1, we compare performance of the four edge map extraction techniques to the fully-coloured YCbCr benchmark. The experiments are carried out under unaltered lighting and thus act as a measure of baseline performance.

In Section 5.2, we then compare the performance of coloured edge maps to their grayscale counterparts to determine if the residual colour information in the former plays a significant role in classification performance.

Finally, we conduct tests with synthetically altered lighting conditions in Section 5.3 to determine if the reduced colour information in coloured edge maps relative to fully coloured images helps improve generalization.

5.1 Baseline Performance

All models were evaluated by their classification accuracy, defined as the number of correct predictions over the total number of predictions made. To obtain a more representative measure of accuracy, 25 batches of the 50-image test set were created, with image augmentation applied to each batch. Image augmentation parameters were identical to those used for training (Table 2). Each model received input features matching those it was trained on.

Figure 4 shows the distribution of results (across 25 batches) for all models. The Sobel model performs similarly to the YCbCr model, with the former achieving a mean accuracy of 91.4% and the latter 91.5%. The other edge map models performed slightly worse, with mean accuracies ranging from 85.5% (LoG) to 90.8% (Kirsch). There was no significant difference between the distribution of results across the models as illustrated by the shape and length of the violin plots, showing that all models performed at similar levels of consistency. However, the LoG model did perform noticeably worse than its peers.

All the models outperformed the approach used in [15] despite the introduction of an additional Underripe class. While from an absolute accuracy perspective our best performing models still slightly underperform those of [11], we believe our approach allows for more streamlined and practical data collection due to the use of full FFBs vs individual kernels, and that this approach results in a more reliable measure of ripeness for reasons discussed in Section 2.1.
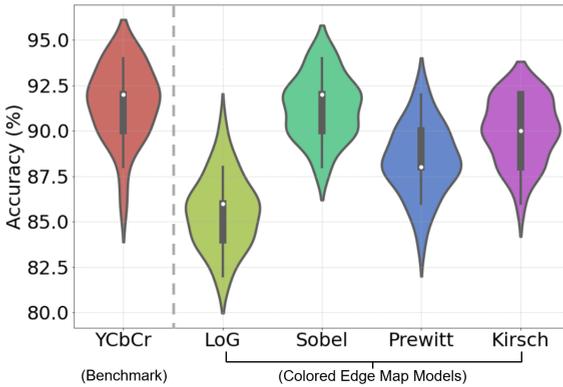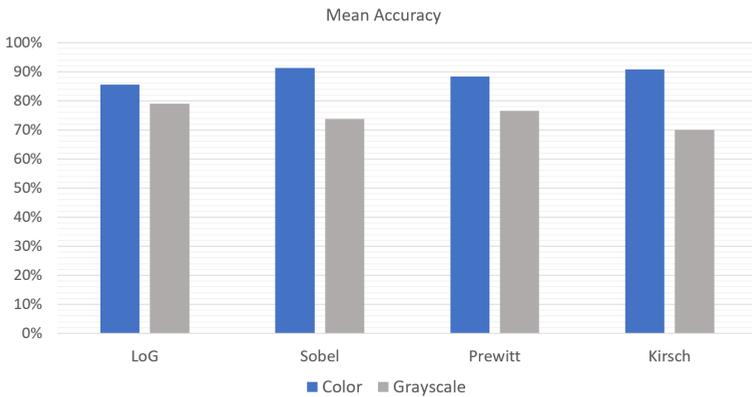
Figure 4: Violin plot of baseline model performance

Figure 5: Performance comparison between coloured and grayscale edge maps

5.2 Grayscale Edge Maps

To determine if the colour content present in coloured edge maps is significant to classification performance, we next train 4 models on the grayscale versions of the four edge map techniques used. This is done by first converting the RGB images to grayscale using OpenCV's *cvtcolour* function before applying the edge extraction algorithms as described in Section 3. The test setup is otherwise identical to that of Section 5.1.

Figure 5 shows the mean accuracies of the coloured and grayscale versions for each edge extraction method used. In all cases, the use of grayscale edge maps produced a significant negative impact on performance, showing that the residual colour content present in coloured edge maps does contribute to classification performance.

5.3 Tolerance to Variances in Lighting

As coloured edge maps remove most colour and lighting information present in the original image, we hypothesize that the resulting models would be less reliant on said features. To test this hypothesis, we synthetically apply brightness augmentation (in addition to the previously described affine transformations in Table 2) to determine if variances in lighting affect

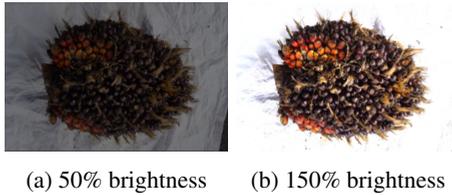(a) 50% brightness   (b) 150% brightness

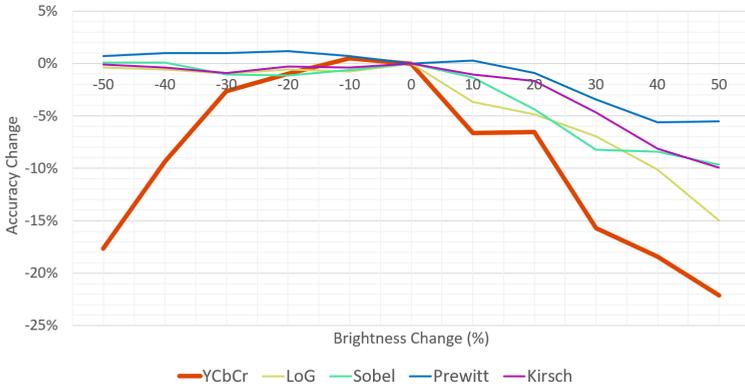Figure 6: Darkened and brightened versions of Figure 3c

Figure 7: Model performance across range of synthetically modified lighting conditions

the models' performance. For each test run, the brightness of each image was multiplied by a constant factor, ranging from 50% to 150% at 10% increments. Figure 6 shows the darkened and brightened versions of Figure 3c. The change in accuracy of each model in these conditions is shown in Figure 7.

In general, the results align with our hypothesis, with the fully-coloured YCbCr model performing the worst at both extremes of the graph. An interesting asymmetrical trend can be observed in all the edge map models, where darkening the images did not cause any observable drop in performance, but brightening the images caused reductions in accuracy of between 5% to 15% at a 50% increase.

6 Conclusion

Our results show that colored edge maps are able to match or surpass the performance of fully colored images when used as an input feature to a convolutional neural network for the task of oil palm ripeness classification. Specifically, the Sobel model matched the baseline accuracy of the benchmark YCbCr model, and all colored edge map models significantly outperformed the benchmark model in the synthetic lighting variance tests.

More generally, we showed that the residual color information left in colored edge maps plays a significant role in model performance, but that the removal of most of the color from the images allows the models to generalize more effectively over novel lighting conditions, outperforming models trained on fully colored images. This property could be of significance to various other applications of computer vision where uncontrolled lighting conditions are common. However, further work will be required to determine if these performance characteristics extend to larger and more generalized datasets.

References

[1] Dibya Jyoti Bora. An optimal color image edge detection approach. *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 2017. doi: 10.1109/icoei. 2017.8300945.

[2] Katharina Buchholz. The world's growing appetite for palm oil. *Statista*, Dec 2020. URL https://www.statista.com/chart/20114/global-consumption-of-palm-oil/.

[3] Norasyikin Fadilah, Junita Mohamad Saleh, Haidi Ibrahim, and Zaini Abdul Halim. Oil palm fresh fruit bunch ripeness classification using artificial neural network. *2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012)*, 2012. doi: 10.1109/icias.2012.6306151.

[4] Mohd Hafiz Mohd Hazir, Abdul Rashid Mohamed Shariff, Mohd Din Amiruddin, Abdul Rahman Ramli, and M. Iqbal Saripan. Oil palm bunch ripeness classification using fluorescence technique. *Journal of Food Engineering*, 113(4):534–540, 2012. doi: 10.1016/j.jfoodeng.2012.07.008.

[5] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[6] S N Mohanaraj and C R Donough. Harvesting practices for maximum yield in oil palm: Results from a reassessment at ijm plantations, sabah. *Oil Palm Bulletin*, 72, May 2016.

[7] B O.Sadiq, S.M. Sani, and S. Garba. Edge detection: A collection of pixel based approach for colored images. *International Journal of Computer Applications*, 113(5): 29–32, 2015. doi: 10.5120/19825-1667.

[8] Nurbaity Sabri, Zaidah Ibrahim, and Dino Isa. Evaluation of color models for palm oil fresh fruit bunch ripeness classification. *Indonesian Journal of Electrical Engineering and Computer Science*, 11(2):549, 2018. doi: 10.11591/ijeecs.v11.i2.pp549-557.

[9] Osama Mohammed Ben Saeed, Sindhuja Sankaran, Abdul Rashid Mohamed Shariff, Helmi Zulhaidi Mohd Shafri, Reza Ehsani, Meftah Salem Alfatni, and Mohd Hafiz Mohd Hazir. Classification of oil palm fresh fruit bunches based on their maturity using portable four-band sensor system. *Computers and Electronics in Agriculture*, 82:55 – 60, 2012. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2011.12. 010. URL http://www.sciencedirect.com/science/article/pii/S016816991100322X.

[10] A. Septiarini, H. Hamdani, H. R. Hatta, and A. A. Kasim. Image-based processing for ripeness classification of oil palm fruit. In *2019 5th International Conference on Science in Information Technology (ICSITech)*, pages 23–26, 2019. doi: 10.1109/ICSITech46713.2019.8987575.

[11] Muhammad Kashfi Shabdin, Abdul Rashid Mohamed Shariff, Mohd Nazrul Azlan Johari, Nor Kamilah Saat, and Zulkifly Abbas. A study on the oil palm fresh fruit bunch (ffb) ripeness detection by using hue, saturation and intensity (hsi) approach.

IOP Conference Series: Earth and Environmental Science, 37:012039, 2016. doi: 10.1088/1755-1315/37/1/012039.

[12] Ian K. Tan, Yue-Hng Lim, and Nyen-Ho Hon. Towards palm bunch ripeness classification using colour and canny edge detection. *Lecture Notes in Electrical Engineering*, page 41–50, 2021. doi: 10.1007/978-981-33-4069-5_4.

[13] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv e-prints*, art. arXiv:1905.11946, May 2019.

[14] Li Za Wong.   Colours   of   distinction: When   oil   palm   is   right
for   picking.   *TheStar*,   Sep   2014.   URL   https://www.
thestar.com.my/Lifestyle/Features/2014/09/01/
Colours-of-distinction-When-oil-palm-is-right-for-picking/.

[15] Zhin Yue Wong, Wei Jen Chew, and Swee King Phang. Computer vision algorithm development for classification of palm fruit ripeness. *13Th International Engineering Research Conference (13Th Eureca 2019)*, 2020. doi: 10.1063/5.0002188.