Utilization of Multi-Channel Hybrid Deep Neural Networks for Avocado Ripeness Classification

Sumitra Nuanmeesri
Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand
sumitra.nu@ssru.ac.th (corresponding author)

Received: 26 April 2024 | Accepted: 12 May 2024

 DOI: https://doi.org/10.48084/etasr.7651

ABSTRACT

Ripeness classification is crucial in ensuring the quality and marketability of avocados. This paper aims to develop the Multi-Channel Hybrid Deep Neural Networks (MCHDNN) model between Visual Geometry Group 16 (VGG16) and EfficientNetB0 architectures, tailored explicitly for avocado ripeness classification in five classes: firm, breaking, ripe, overripe, and rotten. Each feature extracted is concatenated in an early fusion-based to classify the ripeness. The image dataset used for each avocado fruit was captured from six sides: front, back, left, right, bottom, and pedicel to provide a multi-channel input image in of a Convolution Neural Network (CNN). The results showed that the developed fine-tuned MCHDNN had an accuracy of 94.10% in training, 90.13% in validation, and 90.18% in testing. In addition, when considering individual class classification in the confusion matrix of the training set, it was found that the 'ripe' class had the highest accuracy of 94.58%, followed by the 'firm' and 'rotten' classes with 94.50% and 93.75% accuracy, respectively. Moreover, compared with the single-channel model, the fine-tuned MCHDNN model performs 7.70% more accurately than the fine-tuned VGG16 model and 7.77% more accurately than the fine-tuned EfficientNetB0 model.

Keywords-avocado ripeness classification; convolutional neural networks; EfficientNetB0; hybrid deep neural networks; visual geometry group

I. INTRODUCTION

Avocado consumption has soared globally, fueled by its nutritional benefits and culinary versatility. The biggest supplier and distributor of avocados worldwide is Mexico, which exports the most avocados into the United States market [1]. Avocado is classified as a climacteric fruit that undergoes a distinct ripening phase characterized by a significant increase in respiration and ethylene production, leading to notable biochemical and physiological changes [2]. During this period, climacteric fruits undergo a crucial shift from maturation to ripening, undergoing various changes in taste, scent, consistency, and nutritional value. Some climacteric fruits are easily visible from the outside. For example, banana peels will become yellower as they ripen. Accurately determining the ripeness stage of avocados is crucial for ensuring optimal quality, taste, and shelf life. Although avocados are highly nutritious when eaten ripe, if the consumers eat the flesh of an avocado that is not yet ripe, health effects may occur. The green flesh of an unripe avocado has significant concentrations of a substance called persin, a fungicidal toxin mainly found in avocado leaves and fruit, which diminishes as the fruit matures [3]. Although it is not considered a danger to humans, it has been found that it can impact certain animals that consume it, including dogs, sheep, goats, rabbits, and horses [4-5]. Moreover, if resalers or merchants cannot plan the distribution of avocados according to the ripening period, can cause a loss of economic value. However, accurately assessing avocado ripeness remains challenging, impacting merchants and consumers. Traditional methods often rely on subjective human judgment or limited technology. Sometimes, an avocado is not ready to be eaten without the expertise of a person who can recognize its ripeness. To address these limitations, the integration of advanced technologies such as Deep Neural Networks (DNNs) has emerged, offering promising avenues for precise avocado ripeness classification. Convolutional Neural Networks (CNNs) are becoming a handy tool for jobs involving image analysis, including fruit ripeness classification. CNN is a deep learning algorithm that analyzes visual data such as images. It primarily focuses on generalization, which is their ability to process unseen data [6]. They excel at learning hierarchical features from images, making them well-suited for fruit ripeness classification, such as mangos [7], strawberries [8], and avocados [9-10]. By leveraging CNN, it becomes possible to automate and improve the accuracy of ripeness assessment, which is traditionally done manually and is subjected to human error. Once trained, CNN can classify new fruit images into their categories, such as unripe, ripe, or overripe. This automated classification saves time and labor and enables consistent and accurate assessment, improving quality control and customer satisfaction.

CNNs vary in architecture depending on their design and implementation on devices such as tablets and smartphones. The EfficientNetB0 is a CNN architecture [11] designed to run on low-demand operating systems with low speed and memory, especially smartphones. Due to its effectiveness and efficiency, EfficientNetB0 is often used as a base model for transfer learning and reused as a pre-trained model [12]. Pre-trained versions of EfficientNetB0 on large image datasets like ImageNet are available, allowing developers to fine-tune the model on specific tasks with relatively small datasets. This model is suitable for smartphones so that customers can use it to classify the ripeness of avocados by themselves. Although EfficientNetB0 offers a good balance between accuracy and model size, it often performs competitively with larger models while being more computationally efficient. Further, the Visual Geometry Group 16 (VGG16) architecture is a trendy deep CNN architecture that achieves good accuracy on various image classification tasks and is known for its simplicity and effectiveness. VGG16's depth allows it to learn hierarchical features from images. The network can collect high-level characteristics (object pieces and complicated patterns) and low-level characteristics (boundaries and texturing) through its convolutional layers, leading to rich and informative representations. At the same time, the EfficientNetB0 is based on compound scaling (wider, deeper, and higher resolution) and uses more complex building blocks [11]. Thus, combining the two architectures may help the extracted features to be more accurate for prediction.

Most studies use photographs of only one or two sides of the fruit to classify ripeness. Therefore, this article proposes the Multi-Channel Hybrid Deep Neural Network (MCHDNN) explicitly tailored for avocado ripeness classification using VGG16 and EfficientNetB0 architectures. By amalgamating the power of deep learning with multiple imaging data of six sides of avocado fruits, this approach aims to improved performance.

II. RESEARCH METHODOLOGY

The development of the MCHDNN for avocado ripeness classification consists of four main processes: data collection, image preprocessing, MCHDNN modeling, and model evaluation, as shown in Figure 1.

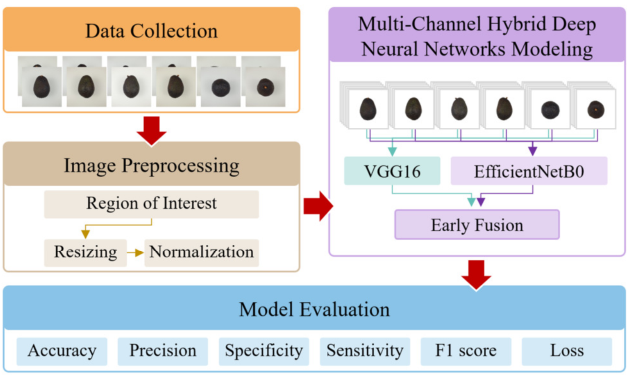


Fig. 1. The model framework for avocado ripeness classification.

A. Data Collection

The avocado fruit photograph data used in this work were captured using built-in cameras of Android smartphones. The standard parameters include an F-stop of f/2.4, exposure bias of 0 steps, no flash, and auto International Organization for Standardization (ISO) speed. In addition, the distance between the camera lens and an avocado placed on a flat surface was 20 cm with a natural light source. The captured image is a square image size of 3,024×3,024 pixels with a resolution of 72 dpi. The total number of avocados used for identification was 200 fruits harvested at a mature but unripe (firm) stage. Each fruit was photographed on six sides: front, back, left, right, bottom, and pedicel, as shown in Figure 2. The avocados were stored at room temperature, and were noted according to the ripening period by capturing the photos of five ripening periods (classes): firm (unripe), breaking (almost ripe), ripe, overripe, and rotten, as shown in Figure 3. Thus, one avocado was photographed from six sides and five ripening periods, to a total of 30 images. So, a total of 6000 photos of 200 avocados for five ripening periods was acquired. All images were labeled with accordingly.

Fig. 2. Photos of the six sides of an avocado fruit.

Fig. 3. The five ripening periods (classes) of an avocado.

B. Image Preprocessing

This work considered three techniques for image preprocessing: image Region of Interest (ROI), image resizing, and image normalization:

1) The Region of Interest

The ROI in image processing refers to a specific area or subset of an image selected for further analysis or processing. It is often used to focus computational resources and attention on the most relevant parts of an image, thereby improving efficiency and accuracy in tasks such as object detection, recognition, and classification. In the context of CNN and computer vision, the ROI typically refers to a bounding box or a defined region within an image that contains the object or feature of interest. This work applied the GrabCut [13] algorithm that separates the prominent object's boundaries from the underlying content based on graph cuts for image segmentation by iterative energy minimization and border matting. This algorithm makes it possible to select only avocado fruit images and then eliminate the background content from the image. The output of the image using GrabCut is shown in Figure 4.

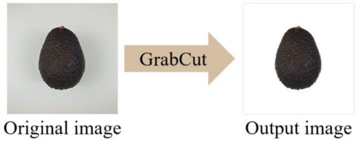


Fig. 4. GrabCut selects the foreground object.

2) Image Resizing

All collected images had a resolution of 3,024×3,024 pixels. However, the EfficientNetB0 and VGG16 architecture used in this work require an input image with a width and height of 224×224 pixels [14]. Thus, all images were resized to 224×224 pixels. This step ensures that the CNN can process the images efficiently.

3) Image Normalization

Image normalization offers several benefits when used as part of the preprocessing pipeline for CNN. Normalizing pixel values to a range of 0-1 reduces the scale of input data, which can lead to faster convergence during training. It helps prevent large gradients that may cause the optimization algorithm to oscillate or converge slowly. Further, it makes the network less sensitive to the initial values of weights and biases. This is especially important for deep networks where vanishing or exploding gradients can occur without proper normalization. Moreover, image normalization can improve CNN's generalization ability by reducing overfitting. By scaling the input data appropriately, normalization encourages the model to learn more meaningful and generalizable features rather than memorizing specific training examples. Therefore, in this work, all input images had their image pixels normalized to the range of 0 to 1, which involves dividing each pixel's value by the image's maximum pixel value, which is typically 255 for images with 8-bit color depth (0-255 range for each channel).

C. Multi-Channel Hybrid Deep Neural Network Modeling

MCHDNN is a deep learning model that integrates multiple ANN architectures to leverage their strengths and improve overall performance. In this work, the MCHDNN model combined the features learned by different architectures between VGG16 and EfficientNetB0 to enhance the accuracy of avocado ripeness classification. Each architecture is defined in detail below.

1) VGG16 Architecture

VGG16 is a CNN architecture that the Visual Graphics Group introduced at the University of Oxford. It accepts input images of 224×224 pixels with three color channels (RGB). Five convolutional (Conv) blocks are composed of 13 convolutional layers. Every block usually consists of a max-pooling layer (2×2 window, stride 2) after several convolutional layers for down-sampling [15]. To preserve spatial dimensions, the convolutional layers employ tiny kernel (k) 3×3 filters, which have a stride of 1. They also employ zero padding. There are 3 Fully Connected (FC) layers, two dense layers with 4,096 neurons and one Rectified Linear Unit (ReLU) activation function. The last layer of FC is the output layer that classifies probabilities for each class with 1,000 neurons and uses softmax activation, corresponding to 1,000 ImageNet classes. However, this work has five classes for this output layer.

2) EfficientNetB0 Architecture

EfficientNetB0 is part of a family of CNN architectures known as EfficientNets. These networks were designed to achieve high accuracy while being computationally efficient, meaning they can achieve state-of-the-art performance using fewer parameters and computations than other architectures. It accepts input images of variable size, typically resized to an input size of 224×224 pixels with three color channels. The initial layer of EfficientNetB0 is a stem Conv layer with 3×3 kernel that processes the input image. The core building blocks of EfficientNet are Mobile Inverted Bottleneck (MBConv) blocks, which consist of inverted residual connections [11]. Each MBConv block is structured as follows:

- Depthwise Convolution: Applies depthwise convolution to extract spatial features efficiently. It functions independently for every input channel.
- Pointwise Convolution: Uses 1×1 convolution to combine the depthwise features and project them into a new feature space with different channels.
- Expansion and Squeeze-Excitation (SE): Some versions of EfficientNet include an expansion phase that increases the number of channels before the depthwise convolution. The SE block helps the network to focus on important channels by adaptively recalibrating channel-wise feature responses.

EfficientNetB0 introduces scaling parameters (compound scaling) to balance model depth, width, and resolution. This allows for scaling the network architecture to different model sizes. Typically, the global average pooling layer, which creates a vector of constant size independent of the size of the input image, often completes the EfficientNetB0 design. It does this by averaging the spatial dimensions of the feature mappings. Finally, an FC layer with softmax activation for classification tasks follows the global average pooling layer. It has neurons equivalent to the number of classes to classify the ripeness of avocados.

3) Hybrid Deep Neural Networks

Six channels of input images are available for each avocado fruit. Thus, all features extracted from each channel were concatenated using the early fusion approach for both VGG16 and EfficientNetB0 networks. Then, the concatenated output of each network was concatenated together. This gives the output feature with a concatenated resolution of VGG16 and EfficientNetB0 as 7×7×10,752. Global average pooling was applied before the FC layers. However, the output still follows the VGG16 architecture with 3 FC layers and one output layer with the softmax activation function. The dataset was randomly divided, with class balancing for training, validation, and testing the models at ratios of 70%, 15%, and 15%, respectively. The model is configured to run the whole training dataset at a rate of 100 epochs. The batch size was set to 32. Further, the adaptive moment estimation (Adam) optimization was applied to dynamically adjust the learning rate to 0.001 for each parameter during training. However, in this research, we developed a model using VGG16 and EfficientNetB0, both pre-trained and fine-tuned models, to compare their performance and accuracy. In addition, the single-channel for each network model without hybridization was used in this work as a baseline for the proposed model. Therefore, this work has 10 models for avocado ripeness classification. The proposed model in this work is shown in Figure 5.

D. Model Evaluation

All 10 models were evaluated for their accuracy in classifying avocado ripeness based on accuracy, precision, specificity, sensitivity, F1 score, and categorical cross-entropy loss. Moreover, the confusion matrix was used to indicate which classes that have the highest classification accuracy.

III. RESULTS AND DISCUSSION

The results of the models' efficiency showed that the fine-tuned models had higher performance than the pre-trained models. For training, the fine-tuned MCHDNN model is more efficient than the other models. The obtained results of this model include 94.10% accuracy, 93.17% precision, 95.11% sensitivity, 93.10% specificity, and 94.64% F1 score. The next five most accurate models were pre-trained MCHDNN, fine-tuned MCEfficientNetB0, fine-tuned MCVGG16, pre-trained MCVGG16, and pre-trained MCEfficientNetB0, with accuracies of 90.63%, 90.58%, 90.53%, 87.18%, and 87.17%.

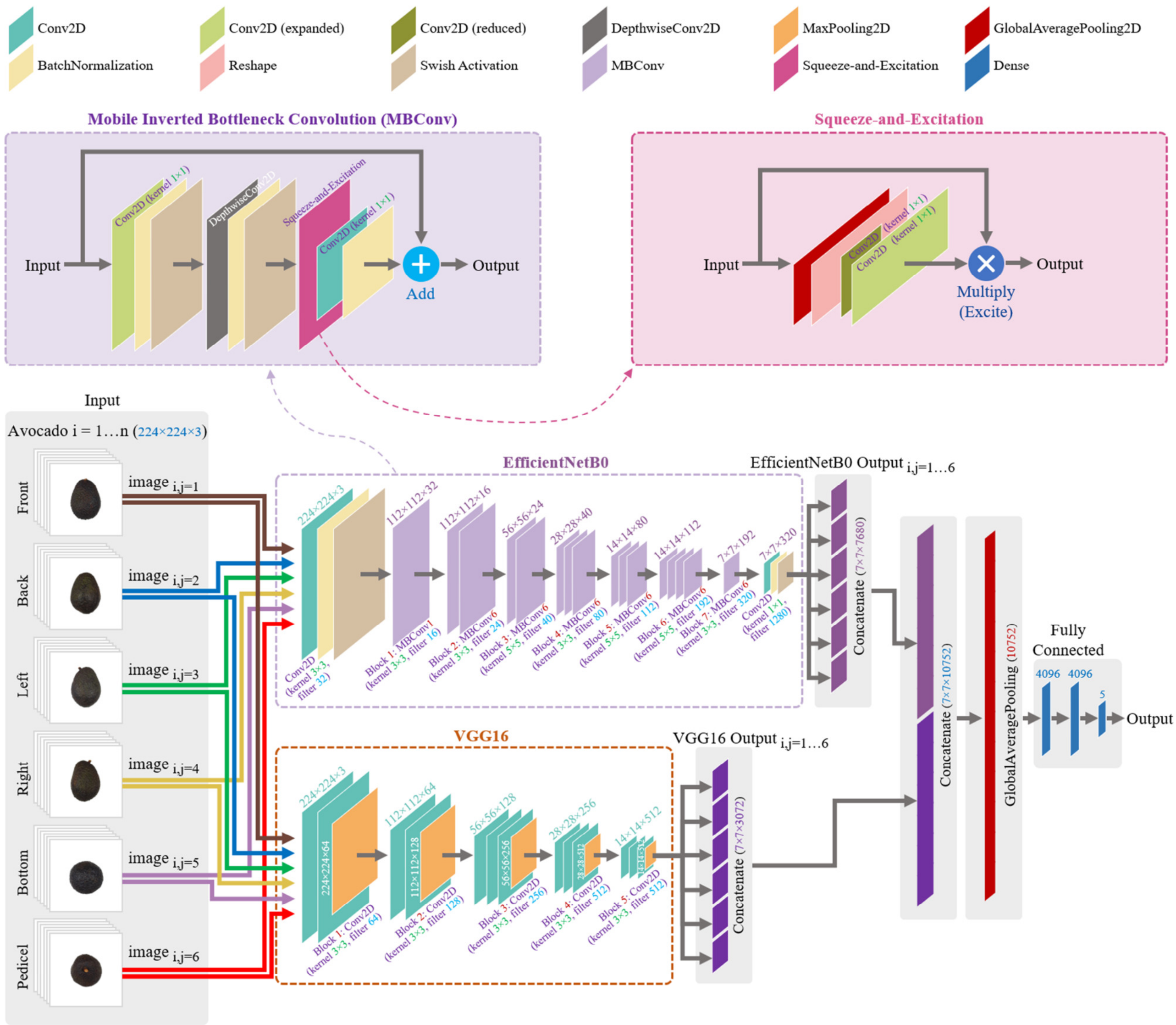


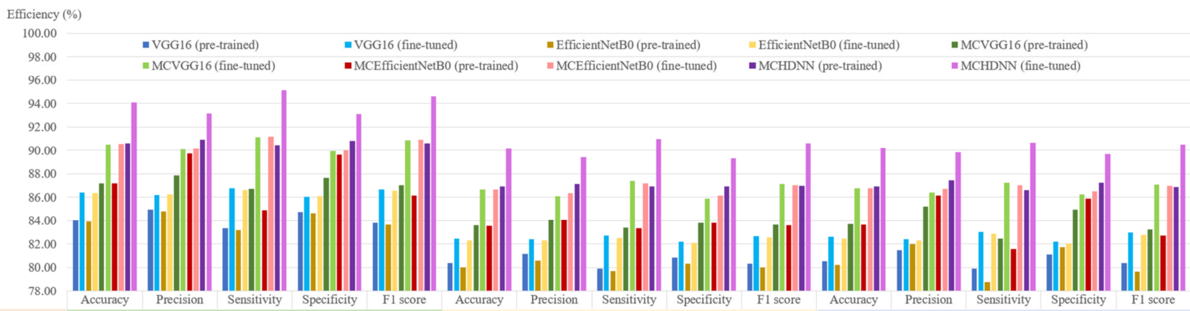
Fig. 5. The proposed model of MCHDNN architecture.

The top-6 most accurate models used Multi-Channel (MC) input data. For the remaining four single-channel models, the VGG16 model performs slightly better than the EfficientNetB0 model for both pre-trained and fine-tuned models. In addition, the experiment results found that the multi-channel model is around 3%-4% more accurate than the single-channel model.

Additionally, the fine-tuned models have more sensitivity than the pre-trained models, which means a lower proportion or number of false negatives and higher model performance. When considering model performance during training, validation, and test processing, it was found that the validation process was approximately 4% less efficient than the training process and less efficient than the test process, by no more than 0.2%. All the models' efficiency is shown in Figure 6. Further, the normalized confusion matrix result of fine-tuned MCHDNN showed that the 'ripe' class had the highest classification accuracy compared to other classes in training, validation, and test processing. The next classifications were 'firm', 'rotten', 'overripe', and 'breaking' classes, respectively, as shown in Figure 7. The consideration of values between training and validation with accuracy and categorical cross-entropy loss of 100 epochs is shown in Figure 8.

IV. CONCLUSION

The ripeness of avocados can be determined using deep learning algorithms and images of avocados. However, having multiple views of the same avocado increases the efficiency in classifying ripeness even more. This research develops a CNN model based on a data set of images of six sides of 200 avocado fruits and collects photos of five ripeness classes. The images of the six sides of the avocados were subjected to feature extraction on hybrid deep neural networks between VGG16 and EfficientNetB0 architectures. The results of each image were concatenated before being sent to the fully connected layers.



Model	Training					Validation					Testing				
	Accuracy	Precision	Sensitivity	Specificity	F1 score	Accuracy	Precision	Sensitivity	Specificity	F1 score	Accuracy	Precision	Sensitivity	Specificity	F1 score
VGG16 (pre-trained)	84.02	84.93	83.34	84.72	83.79	80.37	81.18	79.90	80.85	80.30	80.52	81.45	79.93	81.12	80.39
VGG16 (fine-tuned)	86.40	86.20	86.77	86.03	86.67	82.45	82.41	82.71	82.19	82.69	82.62	82.42	83.05	82.18	82.95
EfficientNetB0 (pre-trained)	83.90	84.82	83.20	84.62	83.65	80.03	80.57	79.72	80.35	79.99	80.23	81.99	78.79	81.75	79.64
EfficientNetB0 (fine-tuned)	86.33	86.24	86.59	86.08	86.54	82.30	82.31	82.53	82.07	82.54	82.47	82.28	82.88	82.05	82.78
MCVGG16 (pre-trained)	87.18	87.85	86.72	87.66	87.05	83.60	84.06	83.40	83.81	83.63	83.68	85.21	82.46	84.97	83.21
MCVGG16 (fine-tuned)	90.53	90.09	91.11	89.96	90.89	86.65	86.10	87.40	85.90	87.12	86.75	86.43	87.26	86.23	87.10
MCEfficientNetB0 (pre-trained)	87.17	89.74	84.89	89.61	86.12	83.57	84.06	83.34	83.80	83.58	83.65	86.13	81.55	85.90	82.74
MCEfficientNetB0 (fine-tuned)	90.58	90.13	91.18	89.99	90.95	86.68	86.37	87.20	86.16	87.04	86.77	86.72	87.01	86.52	86.98
MCHDNN (pre-trained)	90.63	90.94	90.46	90.81	90.61	86.92	87.11	86.90	86.93	86.99	86.93	87.43	86.62	87.25	86.87
MCHDNN (fine-tuned)	94.10	93.17	95.11	93.10	94.64	90.13	89.42	90.98	89.30	90.62	90.18	89.81	90.68	89.69	90.49

Fig. 6. The efficiency result of the models.

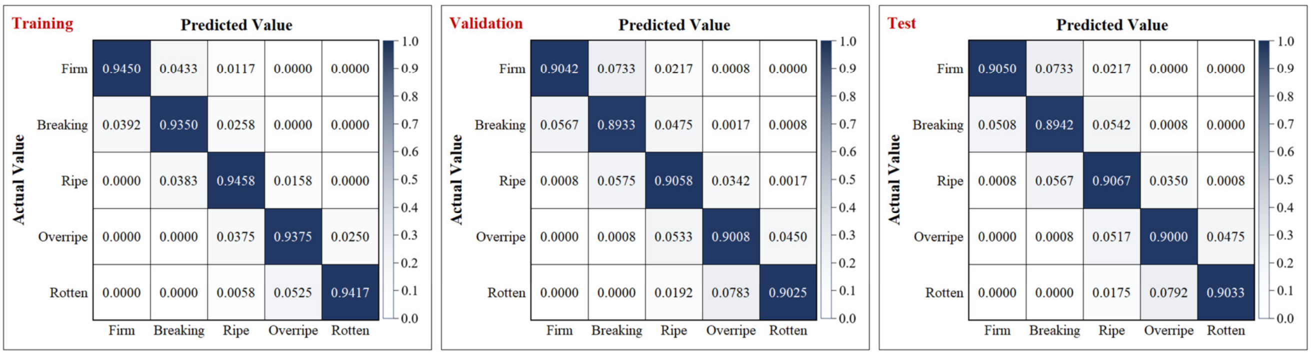


Fig. 7. The normalized confusion matrix of the fine-tuned MCHDNN model.

The research results revealed that the fine-tuned MCHDNN model that uses images on all six sides as multi-channel input and then uses it as a hybrid feature between VGG16 and EfficientNetB0 has the highest efficiency compared to the model that uses single-channel input images and non-hybrid network models. The proposed fine-tuned MCHDNN model has an accuracy of 94.10%, 90.13%, and 90.18% for training, validation, and testing, respectively. It has higher accuracy value than the single-channel models, such as the fine-tuned VGG16 and the fine-tuned EfficientNetB0, by 7.70% and 7.77%, respectively. Further, according to the confusion matrix, the 'ripe' class has the highest classification accuracy, followed by the 'firm' and 'rotten' classes.

These research findings extend the results of previous work [9, 10] by applying multi-channel calculations of six-sided images of the avocado fruit, which contributes to higher efficiency in classifying the ripeness of the avocado fruit, due to the apparent visible changes in the external physical characteristics. For example, if the fruit's stem is lost, it may be classified as having increased ripeness. Therefore, using images from multiple sides of an avocado as multi-channel input can make the ripeness classification more accurate.

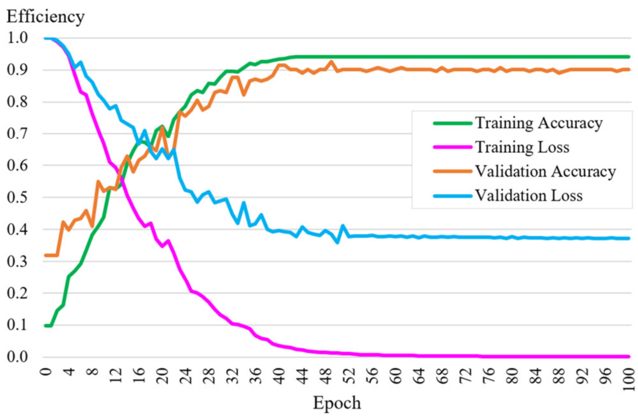


Fig. 8. The accuracy and loss during training and validation processes.

Moreover, hybridizing the feature maps obtained between the VGG16 and EfficientNetB0 models using the concatenate method will improve the model's efficiency and accuracy even more. These demonstrate that the proposed model significantly improves the efficiency of classifying avocado ripeness. Future research will use this model further developed into a mobile application. This may increase buyers' ability to choose the avocado fruit they want. It will also help sellers manage plan distribution or shelf placement efficiently, conveniently, and quickly. However, this image dataset consists of photographs of the appearance of avocado fruits. If internal data of the fruit, such as hyperspectral images, are obtained, it is expected to help further increase the accuracy of the avocado ripeness classification.

ACKNOWLEDGMENT

The author is grateful to the Institute for Research and Development, and the Faculty of Science and Technology, Suan Sunandha Rajabhat University, who supported this research.

REFERENCES

[1] J. C. López-Pimentel, M. Alcaraz-Rivera, R. Granillo-Macías, and E. Olivares-Benitez, "Traceability of Mexican avocado supply chain: A microservice and blockchain technological solution," *Sustainability*, vol. 14, no. 21, Nov. 2022, Art. no. 14633, https://doi.org/10.3390/su142114633.

[2] F. Liu *et al.*, "Time-series transcriptome of Cucumis melo reveals extensive transcriptomic differences with different maturity," *Genes*, vol. 15, no. 2, Jan. 2024, Art. no. 149, https://doi.org/10.3390/genes15020149.

[3] R. Permal *et al.*, "Converting avocado seeds into a ready to eat snack and analysing for persin and amygdalin," *Food Chemistry*, vol. 399, Jan. 2023, Art. no. 134011, https://doi.org/10.1016/j.foodchem.2022.134011.

[4] A.-L. Nagy *et al.*, "Emerging plant intoxications in domestic animals: A European perspective," *Toxins*, vol. 15, no. 7, Jul. 2023, Art. no. 442, https://doi.org/10.3390/toxins15070442.

[5] M. S. Freitas *et al.*, "Acetogenin-induced fibrotic heart disease from avocado (Persea americana, Lauraceae) poisoning in horses," *Toxicon*, vol. 219, Nov. 2022, Art. no. 106921, https://doi.org/10.1016/j.toxicon.2022.09.007.

[6] S. Nuanmeesri, S. Chopvitayakun, P. Kadmateekarun, and L. Poomhiran, "Marigold flower disease prediction through deep neural network with multimodal image," *International Journal of Engineering Trends and Technology*, vol. 69, no. 7, pp. 174–180, Jul. 2021, https://doi.org/10.14445/22315381/IJETT-V69I7P224.

[7] S. Nuanmeesri and L. Poomhiran, "Improved classification of intact ripe mango sweetness using fusion deep learning and enhanced near-infrared spectroscopy," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 60–67, Jul. 2022, https://doi.org/10.14445/22315381/IJETT-V70I7P207.

[8] T. K. Ameetha Junaina, R. Kumudham, B. E. Abishek, and M. Shakir, "Using Deep Learning-Based Features and Image Augmentation to Predict Brix Values of Strawberries for Quality Control," *International Journal of Engineering Trends and Technology*, vol. 71, no. 7, Jul 2023, https://doi.org/10.14445/22315381/IJETT-V71I7P231.

[9] Y. J. Davur, W. Kämper, K. Khoshelham, S. J. Trueman, and S. H. Bai, "Estimating the ripeness of Hass avocado fruit using deep learning with hyperspectral imaging," *Horticulturae*, vol. 9, no. 5, May 2023, Art. no. 599, https://doi.org/10.3390/horticulturae9050599.

[10] C. A. Jaramillo-Acevedo, W. E. Choque-Valderrama, G. E. Guerrero-Álvarez, and C. A. Meneses-Escobar, "Hass avocado ripeness classification by mobile devices using digital image processing and ANN methods," *International Journal of Food Engineering*, vol. 16, no. 12, Dec. 2020, Art. no. 20190161, https://doi.org/10.1515/ijfe-2019-0161.

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *The 36th International Conference on Machine Learning*, Long Beach, CA, USA, Jun. 2019, pp. 6105–6114.

[12] N. C. Kundur, B. C. Anil, P. M. Dhulavvagol, R. Ganiger, and B. Ramadoss, "Pneumonia detection in chest x-rays using transfer learning and TPUs," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11878–11883, Oct. 2023, https://doi.org/10.48084/etasr.6335.

[13] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, Aug. 2004, https://doi.org/10.1145/1015706.1015720.

[14] S. Nuanmeesri, "A hybrid deep learning and optimized machine learning approach for rose leaf disease classification," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7678–7683, Oct. 2021, https://doi.org/10.48084/etasr.4455.

[15] V. T. H. Tuyet, N. T. Binh, and D. T. Tin, "Improving the curvelet saliency and deep convolutional neural networks for diabetic retinopathy classification in fundus images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8204–8209, Feb. 2022, https://doi.org/10.48084/etasr.4679.