Convolutional Neural Network based Rotten Fruit Detection using ResNet50

Chai C. Foong
Department of Electrical and Electronics Engineering
Faculty of Engineering and Technology
Tunku Abdul Rahman University College,
Jalan Genting Klang,
53300 Setapak, Kuala Lumpur
charles_1119@live.com

Goh K. Meng
Department of Electrical and Electronics Engineering
Faculty of Engineering and Technology
Tunku Abdul Rahman University College,
Jalan Genting Klang,
53300 Setapak, Kuala Lumpur
gohkm@tarc.edu.my

Lim L. Tze
Department of Electrical and Electronics Engineering
Faculty of Engineering and Technology
Tunku Abdul Rahman University College,
Jalan Genting Klang,
53300 Setapak, Kuala Lumpur
limlt@tarc.edu.my

2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC) | 978-1-6654-4011-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICSGRC53186.2021.9515280

***Abstract*— Food contains essential nutrients for human beings to grow and develop. Out of so many type of food, vegetables and fruits are important for humans' daily healthy diet as they provide all the nutrients that helps human to prevent diseases. However, fruit will get rotten easily if not store properly due to the spread of bacteria. Therefore, it is important for food industry to perform inspection on fruits before selling to the consumers. The problem encountered in the human inspection is lower in consistency and accuracy as the manual inspection by humans' eye will consume time and energy. To solve this problem, the proposed method is to apply the deep learning technique which is Convolutional Neural Networks (CNNs) for feature extraction and classification of rotten fruits. The types of fruits that will be detected and classified in this paper are banana, apple and orange. The validation accuracy obtained in this paper is 98.89%. The total duration of training stage is 212.13 minutes. Hence, the required time to classify single fruit image is approximately 0.2 second.**

Keywords*— *Deep learning, CNN, rotten fruit detection, ResNet50.

I. Introduction

Not only human, other living being on earth is very dependent on food to stay alive. Food provides nutrients while it is the main sources to provide energy to human. Without energy, human hardly can perform their daily task or activity. It is important to acquire nutrients in order to provide energy for all functions of human body such as digesting food, breathing and making sure the wellness of immune system and so on [1]. Speaking of food, fruits is one of the food that could provide many essential nutrients. For example, fruits are rich in potassium, folate, dietary fibre, vitamin C and many other nutrients which provide health benefits. Therefore, it is important to make sure that the fruits consume by human is fresh. Regardless of whether the fruit can provide the nutrients, a rotten fruit will likely cause illness to human if it is infected with bacteria.

Other than that, the production of fruits is considered as part of the agriculture sector in Malaysia. For many years, agriculture sector is playing important role in the economy of Malaysia. Based on the statistic from Department of Statistic Malaysia, the agriculture sector contributed 7.6% in average to the Gross Domestic Product (GDP) over the years from 2015 to 2019. For example, out of RM1361.5 billion of GDP in 2018, there are RM99.5 billion is contributed by the agriculture sector [2]. Therefore, it is important to develop technologies to support food industry in Malaysia in order to generate more revenues to the economy of country.

In recent years, there are a few research work done on rotten fruit detection to detect rotten fruits with good accuracy. In the work presented by D. Karakaya, O. Ulucan and M. Turkan [5], they have investigated the performance of multiple feature extraction techniques on fruit freshness classification. After the feature extraction of fruits images, the classifier technique which is support vector machine (SVM) were used in the experiment. A total of 1200 fruits images were used where the images are categorized into three different fruits classes such as orange, banana and apple. Among all the feature extraction techniques, Convolutional Neural Networks Features (CNNsF) achieved the best performance with an overall accuracy of 97.61%. The performance of experiment is obtained by computing the success rates of different features SVM while each class is trained with a one-vs-all SVM classifier per feature [5].

In this work [6], the authors had investigated the performance of multiple type of classifiers which are Decision Tree, Artificial Neural Network (ANN) and Naïve Bayes on distinguishing the condition of orange. The condition is categorized as ripe, unripe, scaled or rotten. The features of orange images such as RGB color space and grey values are extracted using BIC in this paper. After the comparison of multiple type of classifiers, Decision Tree has the highest accuracy which is 93.13%. The precision and sensitivity when using Decision Tree classifier is 93.45% and 93.24% respectively. The results are obtained by testing on total of 335 orange images which including 125 unripe oranges, 85 ripe oranges and 125 scaled or rotten oranges [6]. In the work presented by Monika J. and Ashwani K. [7], they had introduced the concept of artificial neural network (ANN) to detect disease and grading of fruit. The disease grading of fruits is divided into 5 by identifying the percentage of infection on fruits. Moreover, it computed the percentage of diseases for some part out of the total pixel counts of the fruits. After that, all the data computed is provided to ANN for training purpose. Back propagation algorithm is used to compute the output value in grading the query images which are not included in learning database. The accuracy of the proposed method is 90% [7].

In this paper, the intelligent rotten fruit detection is designed and developed. The types of fruits that will be focused in the research is banana, apple and orange. The experimental work is to classify whether the banana, orange and apple are fresh or rotten. In this proposed work, the algorithm that will be used is convolutional neural networks (CNNs) of ResNet50. This is because after reviewing several research papers regarding CNN models, ResNet50 is the best model with the highest average accuracy acquired. All the images set are acquired from an open source website, namely

Kaggle [9]. For experimenting with the dataset to perform rotten fruits classification, the training option parameter of mini batch size and the number of epochs in deep learning are studied to investigate and validate their relationship to the validation accuracy. The batch size is a hyperparameter that consider the number of examples from the training dataset used in the estimate of the error gradient before updating the model parameters. When the batch size is set to more than one and less than the examples in the training set, the learning algorithm is called mini-batch gradient descent. A considerably large number of fruits images is acquired to use as the dataset for training with CNNs. The outline of the paper is as such the design methodology of the system is discussed in Section II, the result and discussion obtained from the studies will be demonstrated in Section III and lastly, the conclusion will be presented in Section IV with recommendation of future work.

II. DESIGN METHODOLOGY

A. Fundamental Processes in Image Processing

There are some fundamental steps in image processing such as image acquisition, enhancement, and segmentation. Generally, the image acquisition stage involves some pre-processing steps, such as image rescaling, denoise, smoothing edges and etc. [3]. Image enhancement is among the most appealing areas in digital image processing. This process is to focus on the details of features that is blur and to improve them through the variation of brightness and contrast. While, the segmentation procedures partition an image into its constituent parts or objects [3]. In general, image segmentation is to change the representation of an image to become more detail and easier to perform further processing. For example, it can remove the background of an image and take the foreground as the region of interest (ROI) for next process.

In this paper, the methodology in image processing is divided into 4 different stages as shown in Fig. 1. In the first stage, the input images which are the fruits images are loaded to the program to be processed. Then, the segmentation is performed on the images to extract the foreground of images. The foreground images will then perform with the feature extraction for classification in the system. Before processing the features extraction, all the images need to perform segmentation using colour thresholder function for foreground extraction on the images. The purpose is to reduce the processing time of the next process in feature extraction. The 'HSV' colour space technique is used to detect the background of the fruits images. The representation of colours is hue (H), saturation (S) and value (V). Hue expresses the colour from red to blue from the angle of 0 to 360 degrees. There are six different range of degrees which representing 6 different colours. Saturation controls the purity of colour used. It is represented by the value ranges from 0 to 1. For the value of the model, it means the intensity or brightness of colour [10]. The types of fruits to be classified are banana, apple and orange. Each of the fruits is categorized into fresh fruits and rotten fruits. The total amount of fruits images is 3000. 70% of the images are used for training dataset and 30% is for use as validation dataset. Table I shows the number of images as training and validation datasets for respective fruits. The training dataset is used to train the model, while the validation dataset is used to evaluate the model's performance.

TABLE I. Total number of fresh and rotten fruit images

Type of fruit	Training Set		Validation Set	
	Rotten	*Fresh*	*Rotten*	*Fresh*
Banana	350	350	150	150
Apple	350	350	150	150
Orange	350	350	150	150

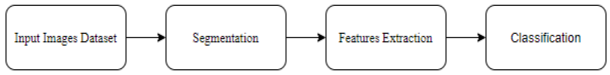

Fig. 1. General steps in image processing.

B. Convolutional Neural Networks (CNNs)

In the field of deep learning, CNNs is considered as a class of deep neural networks. This deep learning technique is mainly used to analyse images for classification, object recognition, object detection and so on. The function of convolutional neural networks (CNNs) is to train and test an image. The images will go through multiple convolutional layers which are the filters, pooling, fully connected layers and lastly applying the softmax function to classify the input image. Fig. 2 shows the flow of CNNs from input image to classify the image based on the feature extracted. Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data. In other words, two inputs such as image and a filter are used to perform a mathematical operation. Sometimes, the filter used in convolution layer does not fit perfectly with the input images. In this case, two options of padding are introduced. The first option is to pad the images with zeroes to make the pictures fit with the filter. Another option is valid padding which only take the valid part of images [4].

ReLU is the most common type of nonlinear activation function. The purpose of this function is to introduce non-linearity to the CNNs. This is because the real-world data may expect CNNs to learn non-negative linear values. Basically, this operation is to convert all the negative pixels of images to zero [4]. When the size of images is too big, pooling layer is used as a down sampling operation to reduce the spatial size while retaining the important parameters. The most commonly used type of pooling layer is max pooling which return the largest value from the rectified feature regions [4].

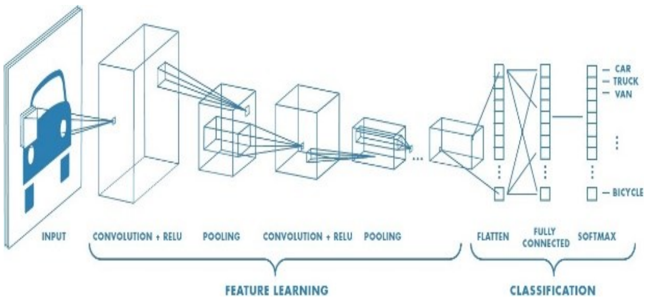

Fig. 2. Complete flow of CNNs [4].

In fully connected layer, all the features learned by the previous layer will be combined together to form a fully connected layer which is similar to neural network. The purpose of this is for the system to identify the larger patterns of images. Finally, the activation function such as softmax function is used to classify the output [4]. As the number of layer increases, the capability for the CNNs to classify complex images also increase.

In this paper, the features of images are extracted and classified using ResNet50. To reduce the computational time, ResNet50 is used as it is a pretrained convolutional neural network that consists of 50 layers deep. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and animals [11]. Fig. 3 shows the model network structure of Resnet50. There is large amount of feature representations extracted from the images has been learned by this pretrained networks. However, to increase the accuracy on classifying the rotten fruits, the pretrained network is taken as a starting point to learn a new task. Instead of training a network from scratch, the pretrained network is modified with transfer learning. Another benefits of doing this is to reduce the required time of training the network.

First, the pretrained network is loaded in MATLAB. All 50 layers of ResNet50 are loaded into a variable. The next step after loading the network is the replacement of final connected layer. To modify this pretrained network to classify the rotten fruits, the fully connected (FC) layer in the pretrained network need to be replaced by a new fully connected layer with the number of outputs equal to the number of classes to classify. In this case, the number of outputs is layer. Normally, the last layer of network which is the fully connected layer will be set as 6. In the network training stage, the size of images needs to be resized to meet the requirement of the pretrained network. For ResNet50, the size of input images is 244-by-244. Other than that, some training options need to be specified before the execution of training such as the learning rate factors, the number of epochs to train and the mini-batch size. The details of training option are explained in the following section. Lastly, the validation accuracy is computed during the validation of network.

C. Segmentation for Training

In this experiment, the segmentation process is performed on all the input images before the training stage. As shown in Fig. 4, it shows one of the rotten orange with white fungus as the foreground pixels in the image. After segmentation, the foreground pixels from the image does not extracted out perfectly, leaving some part of the white fungus unsegmented correctly. The segmented image has the background of images' pixels reduce to zero value (black color).

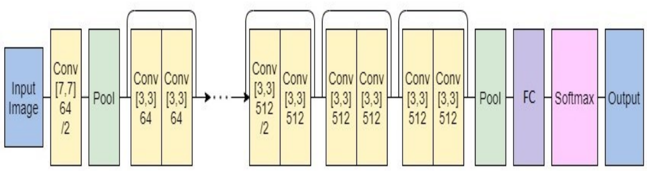

Fig. 3. Resnet50 model network structure [12].

Fig. 4. Rotten orange with white fungus before segmentation (left) and after segmentation (right).

Based on the table in Table II, the results show that the validation accuracy is the same for both cases which is 98.89%. It is observed that training the network using foreground detected images without segmentation has the similar accuracy with the fruit images after segmentation. The results also show the foreground detected images with segmentation requires more training duration to train. Since all the original images are monotonous background in white and grey colour, it is decided to use the original images without segmentation as the dataset for training purpose.

Next, the training option parameters for detecting rotten apple, orange and banana using 350 training datasets for each category are determined. They are listed as shown in Table III. Further experiments have also conducted to investigate the effect of varying mini-batch size and the number of epochs to the validation accuracy and training duration when training with Resnet50. The mini-batch size of 10, 50 and 100 are for investigation. Additionally, the number of epochs to be deployed in the experiment are 3 and 6 epochs. The reason for not using more number of epochs in this case is because if the number of epochs used to train a neural network is more than necessary, the training model will learn the pattern that are specific to a great extent with longer training time than 212.13 minutes using 6 number of epochs in this experiment. Increasing number of epochs will cause overfitting of the training data. Therefore, 6 number of epochs is considered as an optimal value to obtain little overfitting to a training data with good accuracy and acceptable training time. On the other hand, a smaller mini-batch size tends to give better validation accuracy as the frequency of updating the weights of model is higher. The model is able to learn faster and better when the weights are updated more frequently. The result will be discussed in next section to demonstrate the validation of the theoretical studies.

III. RESULT AND DISCUSSION

A graph in Fig. 5 shows the accuracy is plotted for 6 epochs after training and validation processes. The blue line in the graph indicates the training accuracy while black dots line is the validation accuracy. A training accuracy is the accuracy obtains after performing classification on the training dataset using the trained model. The training accuracy is generally higher than validation accuracy because the classification is applied on the same dataset where the model learned from. In this case, after performing classification on training dataset, the training accuracy obtained is 99.14%. However, the validation accuracy will be emphasized more in this case as it measures the capability and quality of the trained model. The required training duration is 212.13 minutes. It is observed that the accuracy started to get stable after the first epoch of training. With only single time of training on entire dataset of training images, the model is able to learn majority of the features and classify the images in validation set. As shown in Fig. 5, the validation accuracy obtained is 98.89%. The validation accuracy also slowly improves as the number of epochs increase. Finally, it reaches to approximately 99% after 6 repeated training on entire set. As plotted in Fig. 6, the losses after the prediction from training and validation are also investigated. A loss is a value indicating how bad the classification is done on the input image. If the prediction of the trained model is perfect, then the value of loss is 0. Otherwise, the value of loss is greater. By observing the graph, the value of loss is getting smaller as the number of training iteration increases. This means that the model is getting better on the classification of images after more training is applied to it. In this case, the value of loss is approximately 0.1 which is a small value of loss.

Fig. 7 shows the prediction result on 4 random images from validation set. The classification result for these four images are predicted correctly with a high probability. Taking the first image as example, the model has detected the image as rotten apple with 94.2% accuracy. There is only 5.8% that this image is recognized as other categories of fruits. Based on Table IV, it shows that by setting the mini-batch size as 10, the validation accuracy obtained is the highest among other cases which is 98.89%. The results obtained from the experiment has agreed well with the theory. With the mini-batch size as 10, the training duration is much more longer compared to the training duration for 50 and 100 mini-batch size. This is because the process of updating the weights of model consumes time while the weights is updated every iteration. For case 1, the number of iterations is 252, which indicates the weights of model is updated for 252 times. Table V investigates the relationship between the number of epochs and the validation accuracy with different number of iterations. It shows that the validation accuracy for 6 epochs is 98.89%, which is higher than using 3 epochs for validation. Based on theoretical study, large number of epochs results in higher accuracy as the larger number of epochs indicates more number times that the learning algorithm work through the entire training dataset. On the other hand, with large number of epochs, the training duration is longer due to that the iteration process will take more time to update the internal model parameters through forward pass and backward pass of all training datasets. In the results, the training duration recorded is 24.5 minutes for 3 epochs, while 49.6 minutes for 6 epochs.

TABLE II. Comparison of accuracy before and after segmentation

Input Images	Mini-batch Size	Number of Epochs	Validation Accuracy	Training Duration (minutes)
With Segmentation	10	6	98.89%	54.03
Without Segmentation	10	6	98.89%	49.60

TABLE III. Parameters of training option

Mini-batch Size	Number of Epochs	Initial Learning Rate	Validation Frequency
50	6	0.0001	3

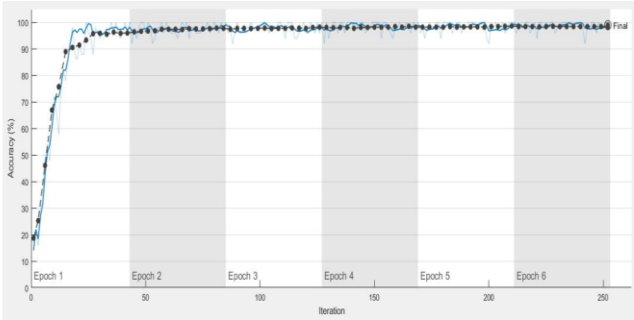

Fig. 5. Plot of training and validation accuracies with number of iterations.

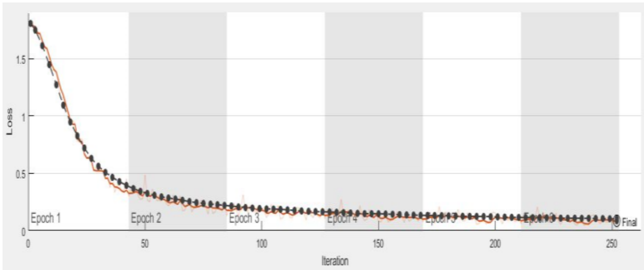

Fig. 6. Plot of training and validation losses with number of iterations.

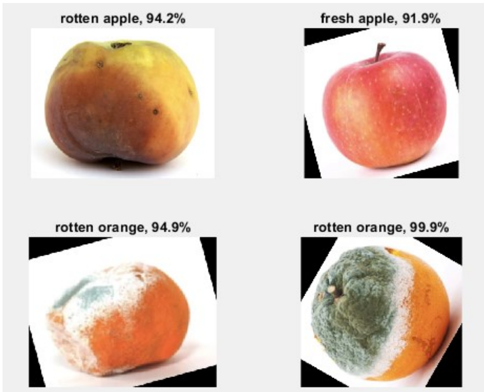

Fig. 7. Four random images from validation set with their prediction result.

Table IV. Comparison of results in validation accuracy for 10, 50 and 100 mini-batch sizes.

Case	Mini-batch Size	Number of Epochs	Number of Iterations	Validation Accuracy	Training Duration
1	10	6	252	98.89%	49.60 minutes
2	50	6	48	98.33%	13.97 minutes
3	100	6	24	96.67%	10.55 minutes

Table V. Comparison of results in validation accuracy for two different number of epochs

Case	Mini-batch Size	Number of Epochs	Number of Iterations	Validation Accuracy	Training Duration
1	10	3	126	97.78%	24.47 minutes
2	10	6	252	98.89%	49.60 minutes

Therefore, it is reasonable for the trained model to classify it wrongly. There are two alternate methods to solve this. The first method is to increase the amount of training dataset for rotten apple and rotten orange. This will allow the model to learn better and differentiate the features between the rotten apple and orange. However, this method depends on the availability of image dataset and also requires more time to train. The second method is to separate the model where one model is mainly trained to classify orange while another model is to classify apple. The method can address the problem easily with minor modification in the model coding. Furthermore, there are 2 images that shows the classification on the type of fruits are correct but with a wrong prediction on the condition of fruits. This error could bring an issue compared to another 7 images which only wrongly predicted the type of fruits. The reason is because it will cause the rotten fruits to pass over and sell to the customer, leading to possible customers' health issue if the system is deployed for use in the market.

Fig. 9 shows the probability of each class after classification for the selected image that have been classified wrongly as a rotten apple. In fact, the image is a fresh apple. The probability of the image predicted as fresh apple is 36.55% while rotten apple is 47.14%. The difference in the probability is 10.59%. Based on the observation on training dataset, majority of the images are apple at the best ripening stage which is fully red apple or green apple. In contrast, only minority of the images are apple with partially yellow and partially red colour on the apples' surface. With less amount of unripe apple images, the model could not learn well and recognize it correctly. Therefore, more images of this kind of apple should include in the training set. From our experiment studies, the probability of rotten green apple being recognized wrongly as fresh apple is also high. This is also due to the small amount of rotten green apple images available in the training dataset. Similar solution is to increase the number of rotten green apple images in the training dataset that can be trained to reduce the probability of wrong prediction. By doing this, the model could learn the features of rotten green apple and to recognize it correctly.

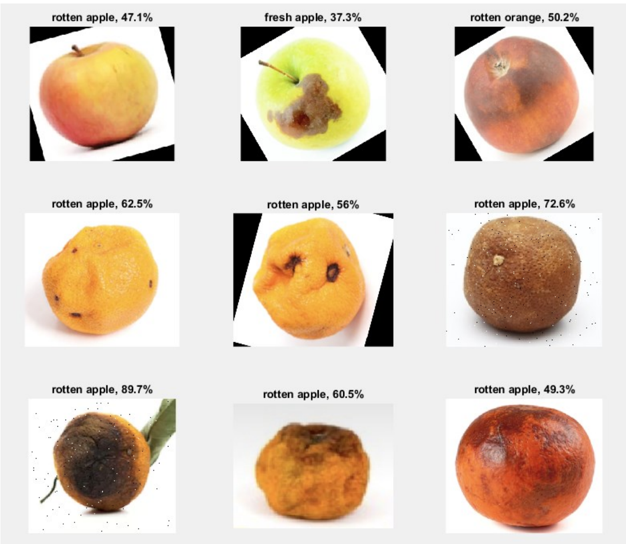

Fig. 8. Images that are wrongly predicted.

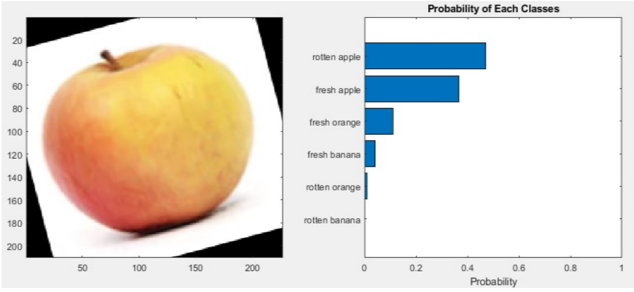

Fig. 9. Probability of each class after classification.

As a conclusion, the proposed method applied is to use transfer learning to detect the rotten fruits. A pretrained network, ResNet50 is used to retrain by changing the final layer to the desired output. The total training duration required to train the model with 2100 input images set is 212.13 minutes. The training duration is long, but the model will take even longer time for training if it is developed from scratch instead of using a pretrained network model. The duration of training can be shorter if higher performance of computer is used. The trained model has its own good flexibility because the trained model can be saved as a network to retrain it. When a new set of input images is desirable to be added into the training dataset to improve the accuracy, the training process with the use of previous original dataset does not need to repeat. Instead, just load the trained model and train it uses the new set of images data. In additional work, a new category of fruit such as strawberry can also be included for classification using this model. Firstly, load the pretrained network which has already contained the features of six categories of the original fruit images. Then, change the final layer to indicate there are seven output classes and then follow by training the network using the dataset with strawberry images. This shows the flexibility of the transfer learning technique.

Although a good accuracy of 98.89% is obtained in detecting the rotten apple, orange and banana. There is a limitation that affects the performance of classification using this proposed method, which is the requirement specifies on the image dataset. For training with the proposed model in this paper, it is important to make sure that the background of images is a monotonous colour. This is because if the background of a fruit image is non-monotonous and is contaminated with noises, it is required to be processed, filtered and segmented in order to extract the features of a fruit in the image properly. Else, it will act as an unwanted feature which will be learned by the model during the training stage. This could affect the result of classification. To require the model to classify a large varying pattern of a particular fruit, the fruit images with varying pattern is required for the model to learn the features of fruit image Based on the result of classification, majority of the wrongly predicted images are under the category of rotten orange and apple. This shown the difficulty of the trained model to classify between these two categories as the appearance of rotten orange and apple on the outer surface is having the similar pattern and colour. These similar features being obtained by the model for rotten apple and rotten orange led to wrong classification.

IV. CONCLUSION AND FUTURE WORKS

An intelligent rotten fruit detection system has successfully developed. The CNN algorithms was studied appropriately in order to detect the rotten fruits. The transfer learning technique was used where the model was retrained from a pretrained CNN network ResNet50. The trained model was then used to perform classification on the rotten and fresh fruits. A total of 2100 fruit images with monotonous background were trained. After comparing the accuracy from the experiment, the input dataset applied for both training and validation stage does not require segmentation. The training option parameter is determined as 6 number of epochs and 30 mini-batch size. With this settings, the required duration for training stage is 212.13 minutes. Six categories of classification were included in this paper including rotten and fresh fruits for apple, banana and orange respectively. The validation accuracy obtained is 98.89%. This indicates that there are only 9 images were classified wrongly from out of 900 images in the validation dataset. In terms of accuracy, classifying a fresh or rotten banana can attain a better result than orange and apple. This is because all the wrongly predicted images are in the category of apple and orange. Evaluating in terms of the speed of classification process, it requires approximately 2 seconds to complete the classification on 10 fruit images. After performing classification on 60 fruit images with non-monotonous background, it has been observed that the accuracy has dropped from 98.89% to 85%.

In future work, it is expected that the probability of correct prediction on green apple can be improved with more apple images having different level of ripeness to be included into the training dataset. This increased amount of fresh and rotten apple images can reduce the possibility of obtaining wrong classification. As the difficulty on classifying rotten apple and rotten orange was discussed. It is recommended to separate the trained model into two different models, which one model is to focus on classifying apple while another is to classify solely on orange. With this method, the accuracy can expect to be higher in classifying the apple and orange.

REFERENCES

[1] L. B. Aydilek, "Approximate estimation of the nutritions of consumed food by deep learning," in 2017 International Conference on Computer Science and Engineering (UBMK), 2017, 160-164.

[2] R. Hirschmann, GDP from the total agriculture industry in Malaysia 015 -2019 [Online]. Available: ttps://www.statista.com/statistics/952693/Malaysia-GDP-from- total-griculture-industry/.

[3] K. Dharavath, G. Amarnath, F. A. Talukdar and R. H. Laskar, "Impact of image preprocessing on face recognition: A comparative analysis," in 2014 International Conference on Communication and Signal Processing, 2014, pp. 631-635.

[4] Prabhu, "Medium - Understanding of Convolutional Neural Network (CNN) — Deep Learning," 2018. [Online].

[5] D. Karakaya, O. Ulucan and M. Turkan, "A Comparative Analysis on Fruit Freshness Classification," in 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), 2019, pp. 1-4.

[6] A. Wajid, N. K. Singh, P. Junjun and M. A. Mughal, "Recognition of ripe, unripe and scaled condition of orange citrus based on decision tree classification," in 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018, pp. 1-4.

[7] M. Jhuria, A. Kumar and R. Borse, "Image processing for smart farming: Detection of disease and fruit grading," *in* 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), 2013, pp. 521-526.

[8] I. Z. Mukti and D. Biswas, "Transfer Learning Based Plant Diseases Detection Using ResNet50," in 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 2019, pp. 1-6.

[9] K. R. Sriram, "Kaggle - Fruits fresh and rotten for classification," 2018. [Online]. Available: https://www.kaggle.com/sriramr/fruits-fresh-and-rotten-for-classification.

[10] S. Trambadia and H. Mayatra, "Food detection on plate based on the HSV color model," *in* 2016 Online International Conference on Green Engineering and Technologies (IC-GET), *2016, pp. 1-6*

[11] I. Z. Mukti and D. Biswas, "Transfer Learning Based Plant Diseases Detection Using ResNet50," in 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 2019.

[12] P. Yang, C. Dong, X. Zhao and X. Chen, "The Surface Damage Identifications of Wind Turbine Blades Based on ResNet50 Algorithm," Shenyang, China, 2020.