Application of Transfer Learning for Fruits and Vegetable Quality Assessment

*Sherzod Turaev
Department of Computer Science & Software Engineering
United Arab Emirates University
Al Ain 15551, UAE
*sherzod@uaeu.ac.ae

Ali Abd Almisreb
Department of Computer Science and Engineering & Software Engineering
International University of Sarajevo
Sarajevo 71210, Bosnia and Herzegovina
alimes96@yahoo.com

Mohammed A. Saleh
School of Computing and Creative Media
University College of Technology Sarawak
Sarawak, Malaysia
mohamedswm@yahoo.com

Abstract— **In this paper, we utilize the concept of transfer learning in fruits and vegetable quality assessment. The transfer learning concept applies the idea of reuse the pre-trained Convolutional Neural Network to solve a new problem without the need for large-scale datasets for training. Eight pre-trained deep learning models namely AlexNet, GoogleNet, ResNet18, ResNet50, ResNet101, Vgg16, Vgg19, and NasNetMobile are fine-tuned accordingly to evaluate the quality of fruits and vegetable. To evaluate the training and validation performance of each fine-tuned model, we collect a dataset consists of images from 12 fruits and vegetable samples. The dataset builds over five weeks. For every week 70 images collected therefore the total number of images over five weeks is 350 and the total number of images in the dataset is (12*350) 4200 images. The overall number of classes in the dataset is (12*5) 60 classes. The evaluation of the models was conducted based on this dataset and also based on an augmented version. The model's outcome shows that the Vgg19 model achieved the highest validation accuracy over the original dataset with 91.50% accuracy and the ResNet18 model scored the highest validation accuracy based on the augmented dataset with 91.37% accuracy.**

Keywords—Transfer learning, deep learning, fruits, vegetable, assessment

I. Introduction

Fruits and vegetables are consumed widely by human beings. However, humans do a kind of visual inspection before buying any items from the market. It is possible to evaluate fruits and vegetables visually, however, this is a hard and subjective process as it depends on an inconsistent evaluation and the fact that these items are being affected by several other factors.

The item's quality can also determine its price in the market besides the ability to consume it. Computer vision-based methods were used to investigate the quality assessment and measurement of fruits and vegetables. Several methods were applied for fruits grading and sorting such as in [1], [2], and [3].

In fact, there are several attributes that usually used to evaluate the fruits and vegetable quality in which, the appearance of the items, the color, the texture, nutritional value, and also the flavor [4]. Traditionally, the first three factors are easy can be captured by the human it is suitable for designing a machine learning application that can measure these factors and decide the quality of the fruits and vegetables. For our application, these three factors are the most contributed to the quality decision as we take images of the fruits and vegetables and classify it into five categories. These categories and, basically, reflect the weekly age of the items. This is due to that the color, appearance, and texture of the fruits and vegetables are changing over time.

The rest of this paper is organized as follows: Section 2 explains the literature review, section 3 includes the materials and methods, then section 4 presents the results and discussions.

II. Literature Review

There are methods investigated in the area of fruits and vegetables quality measurement such as hyperspectral imaging. These concepts are used by several researchers for example [5]and [6]. Another method in which is working is based on detecting the chemical forms and crystal locations using a signal, this method is called Raman imaging. This approach is used also by [7] and [8]. In fact, there are several other techniques reported in [4], such as Laser backscattering imaging used by [9] this method depends on the mechanism of transmitting, absorbing and reflecting the light when going through the fruit and vegetable samples, Magnetic resonance imaging (MRI) applied by [10] and it works to create 2D and 3D images through mapping the proton molecules density of an item, Thermal imaging (TI) applied by [11], this technique depends on collecting the temperature of the item surface. In contrast, there are also statistics based approaches used for fruits and vegetables quality evaluation such as principal component analysis (PCA), partial least squares (PLS), linear discriminant analysis (LDA), and artificial neural networks (ANN).

On the other hand, a study by [12] investigated the defect of the cucumber by using a stacked sparse autoencoder with Convolutional Neural Network. Another Deep learning-based approach applied by [13], this paper implements two steps method, firstly they capture an image of the plum and process it to detect the plum location then secondly applying CNN for classification. A method used to detect the defects of mangosteen applied by [14].

978-1-7281-8184-4/20/$31.00 ©2020 IEEE

III. Methodology and Materials

A. Dataset

In order to evaluate the proposed methods, we have collected a dataset of 11 items, including fruits and vegetables. the list of item's names presented in Table 1. This table shows that the number of images of each item is 350 images. Therefore the total number of images in this dataset is 4200 images. For the Grape, we collected images of a single grape and also grape in the shape of a cluster. A sample image of each item is presented in Fig. 1. These images are collected thought out for 5 weeks. A group of images collected every week to show the changes in every item. Table 2 lists a sample of Apple fruit, it shows also the number of images collected every week is 70 images. A weekly sample of these images is presented in Fig. 2. This figure displays the change of the peel of an apple fruit over 5 weeks. These changes in the peel can give us the impression that the apple quality has been changing over time. However, the challenge now is to measure the quality based on this change. In this paper, we have two assumptions. Firstly, that the first week starts when the customer buys the items from the market. Secondly, in the fifth week, most of the fruits and vegetables are in a condition that can be eaten. In total, as we have 12 items, therefore, the total number of classes is 60. (i.e. (12 items) * (5 weeks)).The bread is considered in this paper also due to its high consumption by human beings.

Table 1: Data collection of all items in the dataset

Item	Number of images
Apple	350
Banana	350
Bread	350
Cucumber	350
Single Grape	350
Grape/cluster	350
Kiwi	350
Lemon	350
Mandarin	350
Paprika	350
Pear	350
Tomato	350

Table 2: Data collection for Apple within 5 weeks

Item	Number of images
Apple/First Week	70
Apple/Second Week	70
Apple/Third Week	70
Apple/Fourth Week	70
Apple/Fifth Week	70

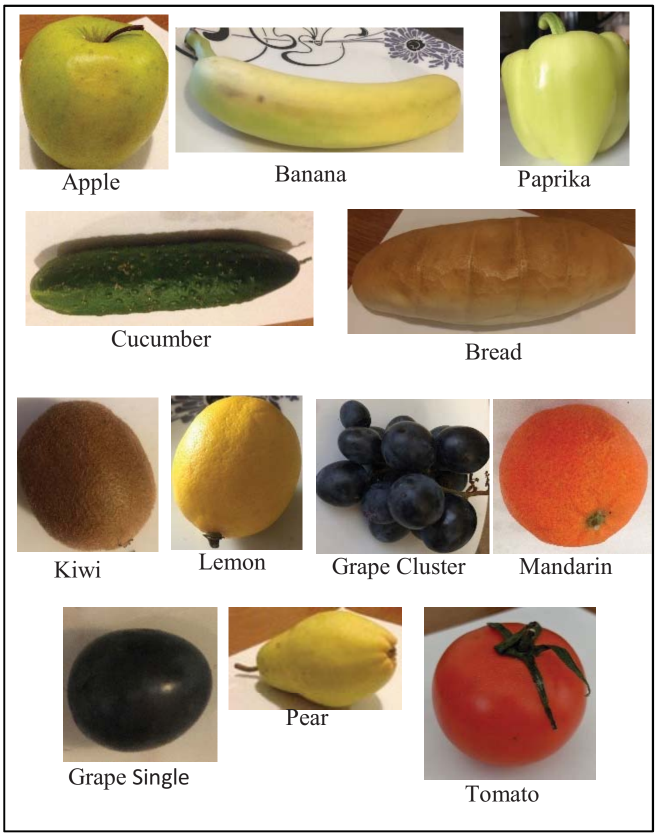

Fig. 1: Items covered in the dataset

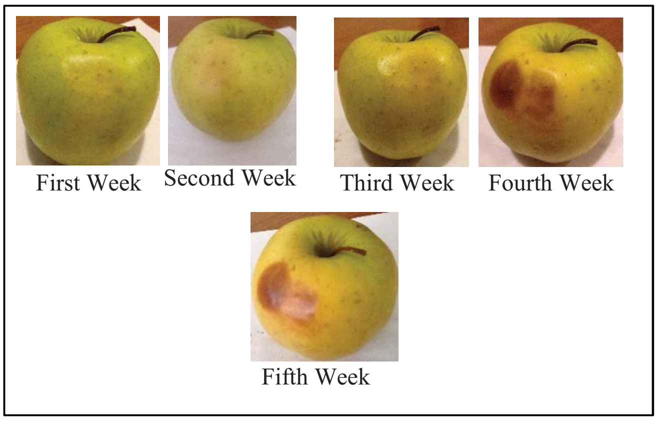

Fig. 2:Apple timeline over 5 weeks

B. Transfer Learning

It is well known that deep learning models can achieve their utmost performance when using large datasets. Therefore, to overtake this obstacle, we use the transfer learning concept. Transfer learning belongs to the machine learning area, in which the Convolutional Neural Networks (CNNs) models finetuned and reused for a new task rather than designing the CNN from scratch [15]. In this paper, we consider (8) eight pre-trained deep learning models. Namely, Alexnet [16], GoogleNet [17], ResNet18, ResNt50, ResNet101 [18], Vgg16, Vgg19 [19], NasNetMobile [20]. ImageNet massive dataset [21] used to train each of these CNNs to classify 1000 classes. Therefore, each of these pre-training models has to be fine-tuned to be training on our proposed dataset and recognize 60 classes.

C. Fine-tuning

Fine-tuning the CNNs aims to keep the learning weights from the previous training on the ImageNet dataset. However, it requires to remove the last three layers of every model (AlexNet, GoogleNet, ResNet18, ResNet50, ResNet101, Vgg16, Vgg19, NasNetMobile). The pre-trained models are divided into two groups. The first group is named as series CNNs in which the layers are stacked one after the other. This category includes; AlexNet, Vgg16, and Vgg19. The second group was named Directed acyclic graph (DAG). The layers of this family are arranged to have multiple inputs and outputs layers in the form of a graph. This category includes GoogleNet, ResNet18, ResNet50, ResNet101 and NasNetMobile. As for the series CNNs models fine-tuning, we added five new layers. Whereas for the DAG CNNs models, three new layers were added. Table 3 depicts the fine-tuning process. Up to this point, we have a new version of each of the pre-trained models.

TABLE 3: FINE-TUNING FOR SERIES AND DAG PRE-TRAINED MODELS

Type	Series			DAG				
Models	AlexNet	Vgg16	Vgg19	GoogleNet	NesNetMobile	ResNet18	ResNet50	ResNet101
New Layers	Fully Connected			Fully Connected				
	ReLU			Softmax				
	Fully Connected			Classification				
	Softmax							
	Classification							

D. Data Augmentation

Our experiments include training the fine-tuned models on the proposed dataset with and without data augmentation. Data augmentation is one of the techniques used in deep learning approaches when the size of training datasets is small scale. Its purpose is to increase the size of the dataset by transforming the original images via rotation, scaling, cropping, or changing the color characteristics. This technique emulates various changes of the original images relating to geometric and color transformation, and it results in expanding the model of the target [22]. The term itself refers to methods for constructing iterative optimization or sampling algorithms via the introduction of unobserved data or latent variables [23]. In this paper, data augmentation has been done in two steps. Firstly, the images are reflected randomly around the X-axis. Then, each image rotated around X-axis and Y-axis with an angle of up to 30 degrees. The last phase in the proposed methodology is partitioning the dataset into 60% of the images for training and 40% for validation and test.

IV. RESULTS AND DISCUSSION

The implementations and experiments of all fine-tuned CNNs models were conducted with Matlab. Models training, validation, and test performed on NVIDIA GeForce GTX 1060 Ti [24]. To evaluate the performance of the fine-tuned CNNs models, we divided the experiments into two parts. Firstly, evaluation is based on the original dataset without applying data augmentation. Secondly, evaluation with data augmentation. The outcome is presented based on the comparative experiments within the same model, i.e with and without data augmentation, and also performance comparison between all proposed CNN models. The evaluation of each model depends on the same criteria. This criterion involves four concepts namely, training accuracy, loss accuracy, validation accuracy, and validation loss for with and without data augmentation.

A. AlexNet

The experimental performance of the AlexNet fine-tuned model over the proposed dataset depicts in Figure 3. This figure shows that the AlexNet model has achieved better training accuracy in the case of the data without augmentation. Figure 3 (a) shows that the model achieved 100% training accuracy at the last training epochs. In terms of the training validation, the model performance has not reached 100% in both cases, with and without data augmentation. However, the model still performing better with the original dataset. In general, the model showed better performance if the dataset is not augmented.

B. GoogleNet

Figure 4 shows the performance of the GoogleNet finetuned model. The figure illustrates that the model still performed well when trained on the data without augmentation. However, the gap between the training accuracy of with and without data augmentation is smaller compare to the AlexNet model. It is also clear that the model started to achieve 100% training accuracy in the earlier epochs. However, the model could not sustain the same training accuracy as in the case of without augmentation. For the validation accuracy in Figure 4 (c), the model performed closely in both cases and that supported by Figure 4 (d) which represents the validation loss.

C. ResNet18

The behavior of the ResNet18 fine-tuned model is displayed in Figure 5. The training accuracy of this model achieved 100% accuracy earlier than AlexNet and GoogleNet models. Figure 5 (a) shows that the model in the case of with and without data augmentation has achieved 100% accuracy during the very beginning epochs and sustain this performance until the end of the training process. As depicts in Figure 5(c), in the case of the data augmentation, the model achieved higher accuracy when trained on the dataset with augmentation.

D. ResNet50

ResNet50 performance is presented in Figure 6. It has a similar behavior as ResNet18 in terms of training accuracy. In contrast, the validation accuracy performance with data augmentation registered less accuracy compared with the original dataset as shown in Figure 6 (d).

E. ResNet101

The outcome of this fine-tuned model is presented in Figure 7. As in Figure 7(a) and (c), the ResNet101 performance quite similar to ResNet50 but less efficient than ResNet18. However, the model with data augmentation could not reach 100% accuracy in any of the training epochs

F. Vgg16

Figure 8 shows the results of this model. There is a sufficient difference between the model performance on the data with and without augmentation. Figure 8(a) shows that the training accuracy is higher with data without augmentation. In fact, the validation accuracy of this model with data augmentation is lower than the validation accuracy of most of the previous models.

G. *Vgg19*

As in Figure 9, Vgg19 fine-tuned model is performing efficiently when trained with original data. In addition, Its training accuracy is more is higher than all other fine-tuned models when trained on unaugmented data as depicts in Figure 9 (a). Moreover, the model performance is considered more efficient compared to other models trained on augmented data as in Figure 9 (d).

H. *NasNetMobile*

Figure 10 presents the performance of this model. From Figure 10 (a) it is clear that the training accuracy with data augmentation is less than the accuracy when training the model on the original dataset. However, from Figure 10 (c) and (d), the validation accuracy and validation loss are performing similarly with slight advantages to training with data augmentation.

The overall comparison is presented in Table 4. This table shows that there are 3 fine-tuned models achieved more than 90% validation accuracy based on the original dataset in which, AlexNet, ResNet50, and Vgg19 which achieved 91.17%, 90.31 and 91.50 respectively. For the validation accuracy based on the augmented dataset, This table shows that ResNet18 has achieved the highest value in which 91.37%.

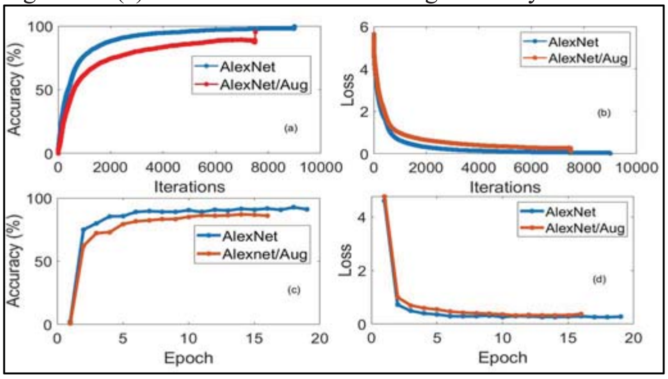

Fig. 3:AlexNet, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

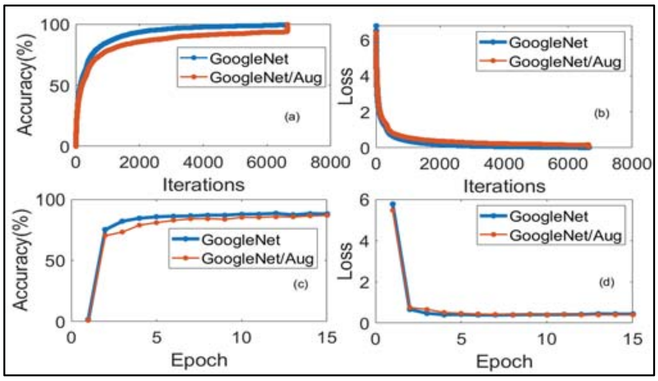

Fig. 4: GoogleNet, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

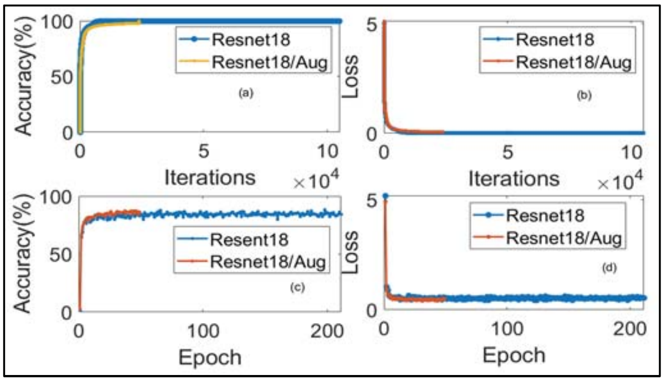

Fig. 5: ResNet18, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

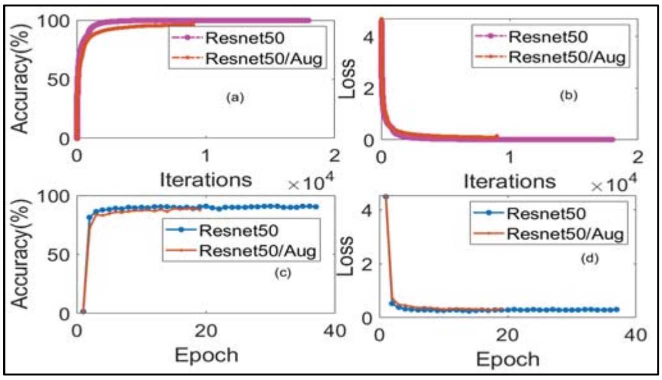

Fig. 6: ResNet50, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

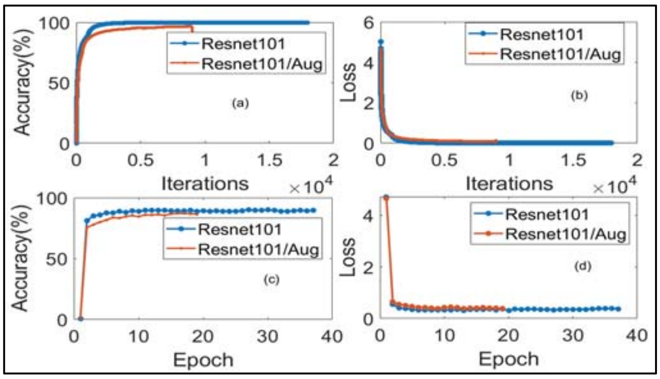

Fig. 7: ResNet101, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

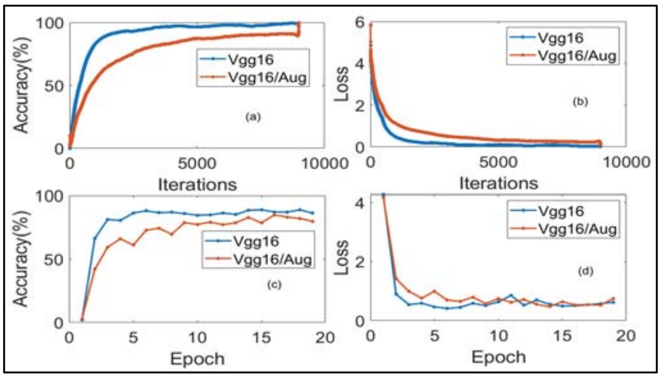

Fig. 8: Vgg16, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

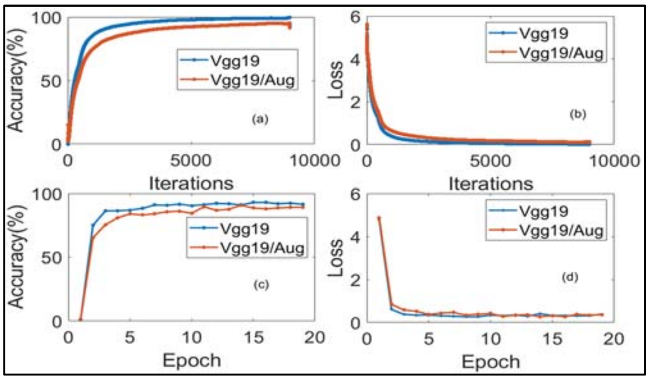

Fig. 9: Vgg19, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

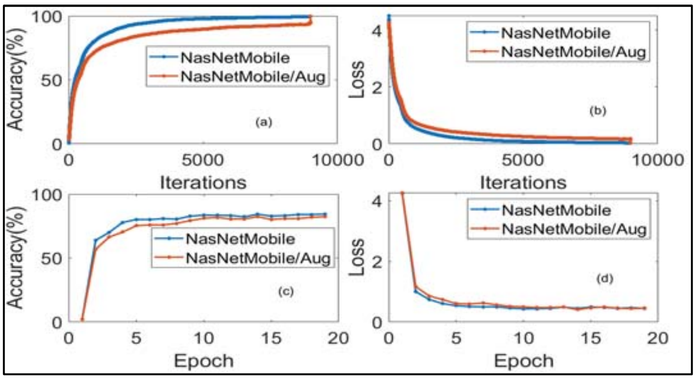

Fig. 10: NasNetMobile, (a) training accuracy. (b) training loss. (c) validation accuracy. (d) validation loss.

TABLE 4: COMPARISON OF FINE-TUNED MODELS BASED ON VALIDATION ACCURACY WHEN TRAINING CONDUCTED WITH AND WITHOUT DATA AUGMENTATIONS

Pre-trained model	without Data Augmentation Validation Accuracy %	With Data Augmentation Validation Accuracy %
AlexNet	91.17	86
GoogleNet	88.05	86.88
ResNet18	87.80	**91.37**
ResNet50	90.31	87.83
ResNet101	89.48	87.94
Vgg16	86.17	79.83
Vgg19	**91.50**	89.17
NasNetMobile	85.11	85.70

V. CONCLUSION

In this paper, we utilized the concept of transfer learning to tackle the problem of fruits and vegetable quality assessment. A dataset of 12 different items (fruits, vegetables, and bread) has been collected to evaluate eight pre-trained models. The fine-tuning process is conducted for each pre-trained model to make it ready to be trained on the collected dataset. The evaluation of these models is conducted based on the training and validation accuracy and also training and validation loss. The training of each model is performed on the original dataset and also on augmented data. The highest validation accuracy based on the original data is achieved by the Vgg19 model which is 91.50%. For the validation accuracy based on the augmented dataset, a validation accuracy of 91.37% is sustained by the ResNet18 model.

ACKNOWLEDGMENT

This work has been supported by the United Arab Emirates University Start-Up Grant 31T137.

REFERENCES

[1] S. Naik and B. Patel, “Thermal imaging with fuzzy classifier for maturity and size based non-destructive mango (Mangifera Indica L.) grading,” in 2017 International Conference on Emerging Trends and Innovation in ICT, ICEI 2017, 2017.

[2] S. Naik and B. Patel, “Machine Vision based Fruit Classification and Grading - A Review,” Int. J. Comput. Appl., 2017.

[3] S. R. Dubey and A. S. Jalal, “Application of Image Processing in Fruit and Vegetable Analysis: A Review,” Journal of Intelligent Systems. 2015.

[4] P. Pathmanaban, B. K. Gnanavel, and S. S. Anandan, “Recent application of imaging techniques for fruit quality assessment,” Trends in Food Science and Technology, vol. 94. pp. 32–42, 2019.

[5] L. M. Dale et al., “Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review,” Appl. Spectrosc. Rev., 2013.

[6] A. A. Gowen, C. P. O’Donnell, P. J. Cullen, G. Downey, and J. M. Frias, “Hyperspectral imaging - an emerging process analytical tool for food quality and safety control,” Trends Food Sci. Technol., 2007.

[7] M. Yamashita, H. Sasaki, and K. Moriyama, “Vapor Phase Alkyne Coating of Pharmaceutical Excipients: Discrimination Enhancement of Raman Chemical Imaging for Tablets,” J. Pharm. Sci., 2015.

[8] W. Huang, N. Zhu, C. Zhu, D. Wu, and K. Chen, “Morphology and cell wall composition changes in lignified cells from loquat fruit during postharvest storage,” Postharvest Biol. Technol., 2019.

[9] N. Zulkifli, N. Hashim, K. Abdan, and M. Hanafi, “Application of laser-induced backscattering imaging for predicting and classifying ripening stages of ‘Berangan’ bananas,” Comput. Electron. Agric., vol. 160, pp. 100–107, 2019.

[10] R. K. Srivastava, S. Talluri, S. K. Beebi, and B. Rajesh Kumar, “Magnetic Resonance Imaging for Quality Evaluation of Fruits: a Review,” Food Anal. Methods, 2018.

[11] S. Zolfagharnassab, A. R. Mohamed Shariff, and R. Ehsani, “Emissivity determination of oil palm fresh fruit ripeness using a thermal imaging technique,” in Acta Horticulturae, 2017.

[12] Z. Liu, Y. He, H. Cen, and R. Lu, “Deep feature representation with stacked sparse auto-encoder and convolutional neural network for hyperspectral imaging-based detection of cucumber defects,” Trans. ASABE, vol. 61, no. 2, pp. 425–436, 2018.

[13] F. J. Rodríguez, A. García, P. J. Pardo, F. Chávez, and R. M. Luque-Baena, “Study and classification of plum varieties using image analysis and deep learning techniques,” Prog. Artif. Intell., vol. 7, no. 2, pp. 119–127, 2018.

[14] L. M. rifatul Azizah, S. F. Umayah, S. Riyadi, C. Damarjati, and N. A. Utama, “Deep learning implementation using convolutional neural network in mangosteen surface defect detection,” in Proceedings - 7th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2017, 2018, vol. 2017-November, pp. 242–246.

[15] A. Lumini and L. Nanni, “Deep learning and transfer learning features for plankton classification,” Ecol. Inform., 2019.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Adv. Neural Inf. Process. Syst., pp. 1–9, 2012.

[17] C. Szegedy et al., "Going Deeper with Convolutions," in IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[20] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018.

[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in CVPR09, 2009.

[22] H. Jo and W. Y. Kim, "NIR reflection augmentation for deeplearning-based NIR face recognition," Symmetry (Basel)., vol. 11, no. 10, 2019.

[23] K. Tang, Y. Wang, H. Liu, Y. Sheng, X. Wang, and Z. Wei, "Design and Implementation of Push Notification System Based on the MQTT Protocol," pp. 116–119, 2013.

[24] "GeForce GTX 1060." [Online]. Available: https://www.nvidia.com/en-us/geforce/products/%0A10series/geforce-gtx -1060/specifications. [Accessed: 15-Aug-2020].