Fruit recognition based on near-infrared spectroscopy using deep neural networks

Duy Khanh Ninh*
The University of Danang – University of Science and Technology, Danang, Vietnam

Thi-Ngoc-Canh Doan
The University of Danang – University of Economics, Danang, Vietnam

Chi Khanh Ninh
The University of Danang – Vietnam-Korea University of Information and Communication Technology, Danang, Vietnam

Thanh Xuan Nguyen-Thi
The University of Danang – University of Science and Technology, Danang, Vietnam

Nhan Le Thanh
The University of Danang – Danang International Institute of Technology, Danang, Vietnam

ABSTRACT

Near-infrared (NIR) spectroscopy has been widely used to determine the varieties and chemical properties of agricultural and food products. The major advantage of NIR spectroscopy is that the analysis is carried out in a simple, fast, and non-destructive manner, making it suitable for food applications. As the first step in applying NIR spectroscopy for fruit recognition and analysis in Vietnam, this paper presents deep neural networks (DNNs) based solutions for automatic recognition of several kinds of fruits. We compared two proposed DNN architectures based on Convolutional Neural Network (CNN) and Residual Network (ResNet). Additionally, we proposed feature extraction methods using the first and second derivatives of the Savitzky-Golay (SG) filtered normalized NIR data. Experimental results show that the deep learning approach combined with reasonable feature extraction process can achive the accuracy of approximately 99% for the task of classifying five types of fruits including Apple, Avocado, Dragon Fruit, Guava, and Mango. The ResNet-based model is more compact and has slightly better recognition performance than the CNN-based one. The inclusion of the first and second derivatives of SG-smoothed normalized spectra improves the recognition accuracy of the proposed DNN models by more than 8%. Moreover, the recognition performance of the proposed DNN models surpasses that of traditional classifiers, including k-nearest neighbors, Naive Bayes, and support vector machine. Our proposed methods were proved to be robust against the freshness of fruits, the NIR device's calibration parameters, and the measurement position on the body of fruits.

*Corresponding author.

ICMLSC'21, January 29–31, 2021, Virtual Event, Vietnam

ACM ISBN 978-1-4503-8761-3/21/01…$15.00
https://doi.org/10.1145/3453800.3453817

CCS CONCEPTS

• **Computing methodologies**; • **Machine learning**; • **Machine learning approaches**;

KEYWORDS

Near-infrared Spectroscopy, Feature Extraction, Deep Neural Networks, Fruit Recognition

ACM Reference Format:
Duy Khanh Ninh, Thi-Ngoc-Canh Doan, Chi Khanh Ninh, Thanh Xuan Nguyen-Thi, and Nhan Le Thanh. 2021. Fruit recognition based on near-infrared spectroscopy using deep neural networks. In *2021 The 5th International Conference on Machine Learning and Soft Computing (ICMLSC'21), January 29–31, 2021, Virtual Event, Vietnam.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3453800.3453817

1 INTRODUCTION

Nowadays, the demand for fruit products tends to increase rapidly with health protection trend. An improvement in dietary diversity, which consumes more fruits and vegetables to replace food of animal origin, leads to improve overall health. Indeed, consumer choices and price highly depend on fruit appearances (such as size, color, and texture) and internal chemical attributes (such as freshness, total soluble solid, and acidity). As a result, there is a need for fast, non-contact, non-destructive analysis methods to extract internal quality features of fruits. Among them, the recent emergence is near-infrared (NIR) spectroscopy and near-infrared spectroscopic imaging (NIR spectral imaging) [1]. Many studies in recent years have provided evidences that NIR spectroscopy (NIRS) have significant potentials for detecting chemical, microbiological and physical hazards in food products as well as helping to distinguish varieties of agricultural products [2]. For fruits, it can be said that these techniques are viable options as the substitutes for traditional measurements which are time-consuming, expensive, destructive, and even dangerous for the environmental or human health.

As the first step in applying NIRS technology in combination with machine learning techniques for fruit freshness detection in Vietnam, this paper presents deep learning based solutions for identity recognition of five kinds of popular fruits including Apple, Avocado, Dragon Fruit, Guava, and Mango. This paper is organized as follows. Section 2 reviews related work on NIRS and machine learning applications in food and agriculture and describes the data collected in this study. Section 3 describes our proposed solutions, including feature extraction methods and deep learning model architectures. Experimental results are given in Section 4. Section 5 presents the conclusion and future work of the current research.

2 RELATED WORK AND DATA COLLECTION

2.1 Related Work and Motivation

Spectroscopy is a powerful technique for the recognition and characterization of physical materials. In general, it will measure the change in absorption or emission of a material for different wavelengths of light. This method determine how the light interacts with the material and produce a spectrum as the function of the light intensity reflecting towards the sensor. This spectrum is considered as the chemical fingerprint of the material.

Over the last two decades, the investigation of agricultural product quality by use of NIRS technique has been increasingly due to rapid, non-invasive and nondestructive ability on these products [1-3]. This technique has been employed with encouraging results for determining the internal quality of fruits and vegetables [4-8]. In addition, different research studies focus on variety discrimination of agricultural products such as Momotaro tomatoes using partial least squares regression with 96.85% classification success rate [9], or cultivar identification of barley, chickpea and sorghum in Ethiopia using support vector machine and partial least squares-discriminant analysis algorithms with an accuracy of 89%, 96% and 87%, respectively [10].

These studies show that NIRS has been used as a powerful tool for food variety recognition and quality assessment. When used in combination with traditional multi-dimentional statistics or machine learning techniques, they have shown great potentials with good classification rates. On the other hand, deep learning has developed rapidly in recent years due to its effectiveness in learning the essential features of a data set comprising a large number of redundant information such as images [11-12] or NIR spectra [13]. However, there has been no research on the combined use of NIRS and deep neural networks (DNNs) such as Convolutional Neural Network [11] or Residual Network [12] for the identity recognition and freshness detection of fruits. Therefore, the objective of this study is to explore the potential that these two technologies can be used to determine fruit type as the first step of fruit freshness detection problem. Note that available deep learning based fruit recognition [14] or fruit ripeness grading recognition [15] systems employ images acquired by cameras rather than NIR spectra acquired by spectrometers.

2.2 Data Collection

Although there are several NIR spectrometers available in the market, this study used Texas Instrument's DLP NIRscan Nano EVM for its high-performance, affordability and portability. This DLP (Digital Light Processing) based spectrometer supports a wavelength range of 900-1700 nm. A sample was placed directly against the sapphire window of the device so that it can perform an accurate scan. During a scan, the sample absorbs a specific amount of NIR light emitted from two broadband tungsten filament lamps and diffusely reflects the non-absorbed light into the collection lens and focuses it into the optical engine through the input slit. Details on how the device operates can be found in [16].

Figure 1: Settings of NIR data collection sessions.

We used this device and its accompanied software to collect NIRS of fruits. The NIR scanner was connected to a computer via a USB cable. These settings, illustrated in Figure 1, were kept unchanged throughout all data collection sessions. In each session, a .csv file containg the intensity spectrum of the fruit sample was stored in the computer for further data processing and modeling. Each spectrum consists of 224 wavelengths and their corresponding NIR intensity values.

In this study, we gathered the NIR data of five types of fruits, which are Apple, Avocado, Dragon Fruit, Guava, and Mango. These fruits are popular in Vietnam and have different physical properies (e.g., shape and color) and flavours. All samples in each fruit category were bought from the same distributor and have the same place of origin to filter out any effect of these factors on the collected data. However, we tried varying the freshness of fruit samples by carrrying out data collection sessions across several days, from the day of buying them in total fresh condition to six days later. Depending on the fruit type, the quality of samples degrades differently after several days being kept in room conditions. In addition, to investigate the robustness of our proposed methods against the NIR device's calibration parameters and the measurement position, we utilized three different NIR devices of the same type and repeated the scanning operation five times at the center points around the body and three times at each of the two ends of a fruit sample. Totally, we have collected 1,979 NIR spectrum measurements, each of which has the dimension of 224. These data were then divided into three sets for training, validating, and testing the proposed DNN models as shown in Table 1

3 PROPOSED METHODS

Our fruit recognition system was carried out by using supervised machine learning algorithms. The system includes two phases. In the training phase, the recorded NIR intensity spectra of multiple samples of five fruits are used as the training data, and the fruit types are used as the labels. Since the dynamic range of the

Table 1: Number of NIR spectrum measurements for each fruit type.

Fruit type	Training set	Validation set	Test set	All sets
Apple	246	102	66	414
Avocado	235	78	66	379
DragonFruit	240	78	66	384
Guava	251	73	66	390
Mango	266	80	66	412
Total	**1238**	**411**	**330**	**1979**

recorded spectra is quite fluctuated among fruit varieties, these data are then pre-processed with the normal distribution normalization technique so that all of them have the mean of zero and the standard deviation of 1.0. After that, relevant features are extracted for training predictive models. In the testing phase, a test data, (i.e., the NIR spectrum of a fruit sample) is pre-processed and applied feature extraction in the same way as the training phase. Then, the extracted features is inputted into the trained models for predicting the associated label of the test data. The feature extraction and model building steps are described in details below.

3.1 Feature Extraction Techniques

The original feature vector of a fruit sample only includes its normalized NIR intensity spectrum. However, initial training experiments with the normalized spectra show that the recognition accuracy is not quite high because the models fail to learn minor differences among the spectra of different types of fruits. Thus we propose to include the first and second derivatives of the normalized spectra to the original feature vector to enhance the differences between them. In fact, the inclusion of derivative (or dynamic) features of the spectrum has shown its effectiveness in generative problems such as statistical parametric speech synthesis [17]. Moreover, simply taking the derivatives of the normalized spectra will result in noisy data with lots of small variations, making the models to learn unuseful information. Consequently, we first utilize the Savitzky-Golay (SG) filter [18] to smooth the normalized spectra without distorting its tendency, then compute the first and second derivatives of the smoothed spectra to obtain the extended feature vector. The full feature vector is the combination of the original feature vector (which has the dimension of 224) and the extended one (which has the dimension of 448), resulting in the total dimension of 672. Figure 2 shows an example of the original spectra, normalized spectra, first and second derivatives of SG-smoothed normalized spectra of five sample fruits. For the SG filter used in this study, the length of the filter window was set to 25 and the order of the polynomial used to fit the samples was set to 5.

3.2 Deep Learning Model Architectures

We proposed two different DNN architectures as shown in Figure 3. The details of the two models are described as follows.

3.2.1 Convolutional Neural Network (CNN) based model. The model (shown in Figure 3a) includes three input layers where the first input layer is the normalized spectra, the second and third input layers are the first and second derivatives of the smoothed normalized spectra, respectively. Each input layer contains 224 neurons as input data, representing the feature vector of size 224×1. Each CNN-based block consists of four convolutional layers, each of them followed by pooling layers. The convolutional layers have consecutively 16, 32, 32, and 64 filters with kernels of size 2×1 and Rectified Linear Units (ReLUs) as the activation functions. In pooling layers, max pooling was used with a pool size of 2×1 along with a stride length of 2. A flatten layer was used to converts the 2D filter matrix into an 1D output feature vector of size 832. The output feature vectors of three CNN-based blocks were concatenated into a single output vector of size 2496, which was then entered into the two fully connected (FC) layers. The first FC layer consists of 128 neurons and a ReLU activation function. Finally, the second FC layer or the output layer contains 5 neurons where softmax classifier activation was used to predict the output (i.e., the fruit label) of the model.

3.2.2 Residual Network (ResNet) based model. Differing from the CNN-based one, the ResNet-based model (shown in Figure 3b) includes only one input layer containing 672 neurons as input data, representing the full feature vector of size 224×3. In a ResNet-based block (or a residual layer) [12], the input layer, after going through convolutional layers, will be concatenated with itself to form the output. This make the input layer be preserved on the way to the last layer, independent with how deep the model is. Our proposed model consist of four ResNet-based blocks, each of which consists of two convolutional layers with kernels of size 3×1 and Leaky ReLUs as the activation functions, followed by pooling layers where max pooling was used with a pool size of 2×1 along with a stride length of 2. The number of output filters in eight consecutive convolutional layers are 16, 32, 32, 64, 16, 32, 32, and 64. A flatten layer was used to converts the 2D filter matrix into an 1D output feature vector of size 896, which was then entered into the three FC layers. The first and second FC layers consists of 128 and 256 neurons, respectively, with a Leaky ReLU activation function and a dropout of 0.5 for reducing overfitting. Finally, the third FC layer or the output layer contains 5 neurons where softmax classifier activation was used to predict the output of the model.

4 EXPERIMENTAL RESULTS

4.1 Comparisons between Proposed DNN Models

For training the proposed DNNs, we used Adam optimizer with a learning rate of 0.0001. The experiments were carried out by using Keras, a deep learning API running on top of the machine

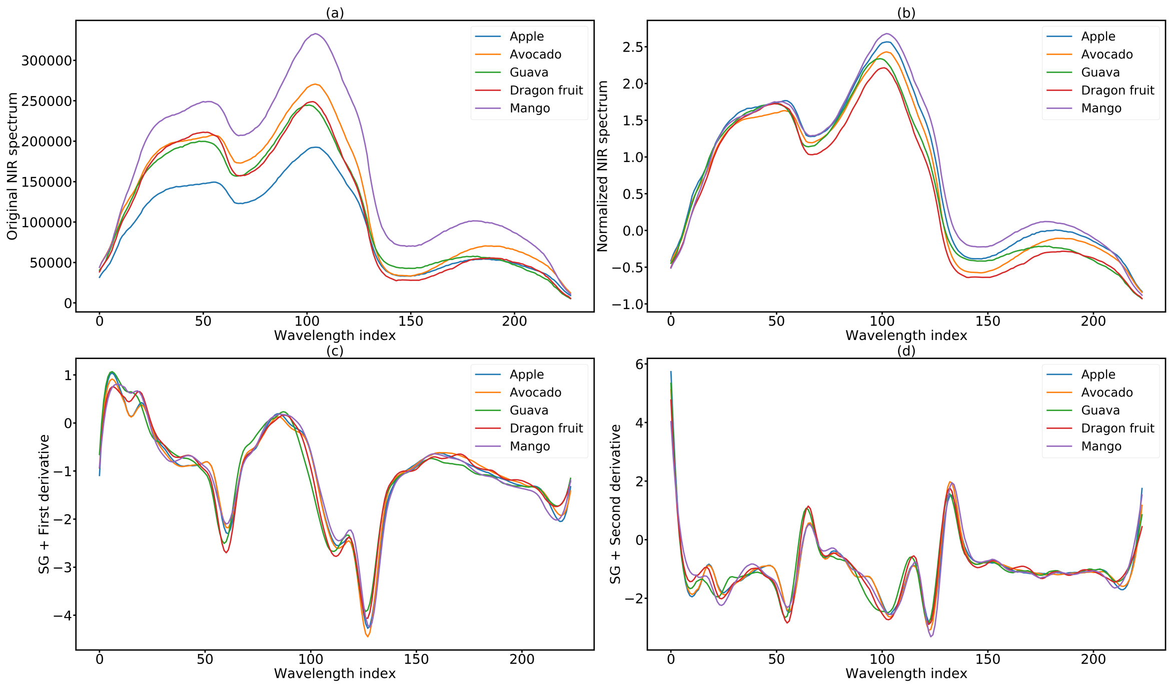
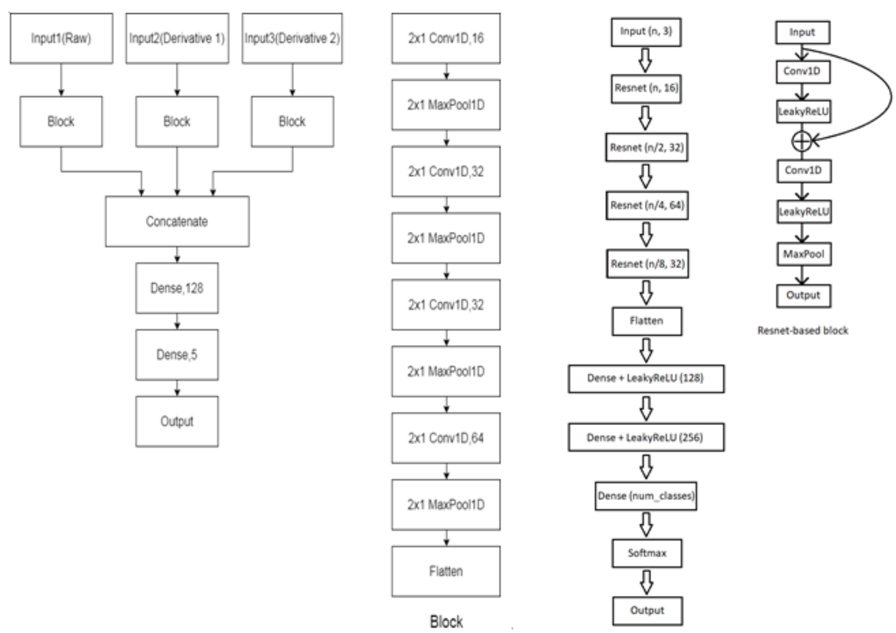
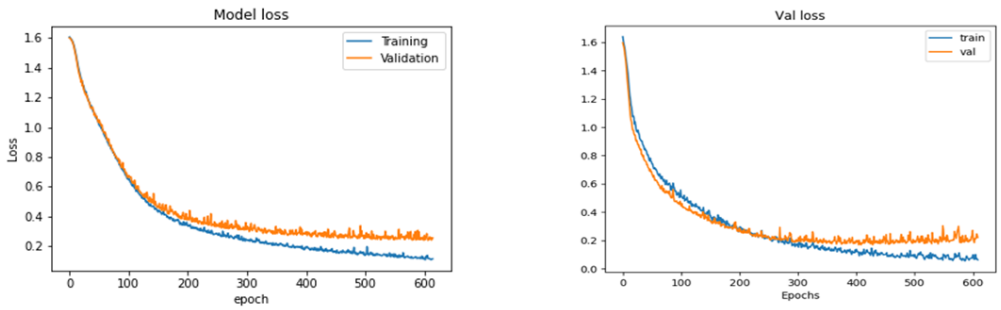


Figure 2: An example of original spectra (a), normalized spectra (b), first (c) and second (d) derivatives of normalized spectra after SG-smoothing of five sample fruits.



(a) CNN model (left) with CNN-based block (right) (b) ResNet model (left) with Resnet-based block (right)

Figure 3: Proposed deep learning model architectures.



(a) CNN-based model (b) ResNet-based model

Figure 4: Model loss variations on the training and validation sets after 600 training epochs.

Table 2: Accuracy of DNN models on three data subsets when using proposed feature extraction techniques.

Model	Training set	Validation set	Test set
CNN-based	99.4%	97.1%	98.2%
ResNet-based	97.0%	95.0%	99.0%

Table 3: Accuracy of DNN models on three data subsets when using only normalized NIR spectra.

Model	Training set	Validation set	Test set
CNN-based	91.6%	90.1%	89.8%
ResNet-based	90.9%	89.0%	90.2%

learning platform TensorFlow. The proposed feature extraction techniques (i.e., normalized NIR spectra combined with the first and second derivatives of SG-smoothed normalized spectra) were used. Figure 4 shows how the cross-entropy based loss function of the two DNN models varied on the training and validation sets over training epochs. We stopped the training process after 600 epochs to prevent over-fitting since both of the two models had their losses converged on the validation set at this point.

Table 2 reports the accuracy on three data subsets of the two DNN models. It can be observed that the ResNet-based model achives slightly higher correct classification rate on the test set compared to the CNN-based one (99.0% vs. 98.2%), though the former performs less efficiently on the training and validation sets compared to the latter. Regarding the model complexity, the total number of trainable parameters of the ResNet-based model is 192,853, while that of the CNN-based model is 342,293. In terms of the recognition time, it tooks both of the two models from 2 to 3 ms for giving the prediction on a fruit sample when tested on the same computer equipped with Intel Core i5-4590 CPU, 8GB RAM, and Nvidia GeForce GTX 1070 Ti Graphic Card. These results show that ResNet-based model is more compact and has slightly better recognition performance than the CNN-based one.

4.2 Effect of Proposed Feature Extraction Techniques

We investigated the effect of our proposed feature extraction techniques by replicating the above experiment with the data using only normalized NIR spectra. The training experiments were conducted in the same way as described in the previous section, except that the CNN-based model includes only one input layer containing 224 neurons, and the ResNet-based model's input layer only consists of 224 neurons for representing the feature vector of size 224×1. Table 3 shows the recognition accuracy on three data subsets of the two DNN models when using only normalized NIR spectra. Comparing with the results in Table 2, it can be seen that the inclusion of the first and second derivatives of SG-smoothed normalized spectra helps improve the recognition accuracy on the test set of the proposed DNN models by more than 8%.

4.3 Comparisons with Traditional Machine Learning Models

We further compared our proposed DNN models with other traditional classifiers including k-nearest neighbors (KNN), Naive Bayes (NB), and support vector machine (SVM) when these models were trained on the data extracted by the proposed feature extraction techniques. The training and tesing experiments were carried out by using the scikit-learn machine learning toolkit with the same data subsets as in the previous experiments. Table 4 gives the highest recognition accuracy on the test set of each model together with the corresponding main model setting. It can be observed that the KNN model achives the highest accuracy of 91.5% when the number of nearest neighbors K is set to 7. Meanwhile, the NB classifier gains the best recognition rate of 86.5% with Gaussian NB, and the SVM model achives the highest accuracy of 90.2% when the linear kernel is used. In summary, the recognition performance of these traditional classifiers is inferior compared to the two proposed DNN models.

5 CONCLUSION

This paper presents the first attemp in applying NIR spectroscopy for fruit recognition in Vietnam. We have proposed deep learning based solutions for automatic recognition of five kinds of fruits, which are Apple, Avocado, Dragon Fruit, Guava, and Mango. We have compared two proposed DNN architectures based on CNN

Table 4: Accuracy of traditional classifiers on test set when using proposed feature extraction techniques.

Model	Main model setting	Accuracy (%)
KNN	K=6	90.0
	K=7	**91.5**
	K=8	88.5
NB	Gaussian NB	**86.5**
	Multinomial NB	85.2
	Bernoulli NB	81.1
SVM	Linear kernel	**90.2**
	Polynomial kernel	88.9
	Radial Basis Function kernel	84.0

and ResNet. Additionally, we proposed feature extraction methods using the first and second derivatives of the collected NIR data in combination with the SG filter. Experimental results showed that the ResNet-based model achives slightly higher accuracy and is more compact than the CNN-based one. The inclusion of the first and second derivatives of SG-smoothed normalized spectra improves the recognition accuracy of the proposed DNN models by more than 8%. Besides, the recognition performance of the proposed DNN models surpasses that of traditional classifiers. Our proposed methods are robust against the freshness of fruits, the NIR device's calibration parameters, and the measurement position on the fruit's body. These results confirm the potential of applying NIR spectroscopy and machine learning technologies for fruit analysis. In the next step, we will examine machine learning methods for identifying the freshness of fruits based on their NIR spectroscopy.

ACKNOWLEDGMENTS

This work was supported by The University of Danang, University of Science and Technology, code number of Project: T2020-02-32, and the Agence Universitaire de la Francophonie (AUF), code number of Project: DRAP-4398. The authors would like to thank the students who help with several experiments in this research.

REFERENCES

[1] Jan U.Porep, Dietmar R. Kammerer, and Reinhold Carlea. 2015. On-line application of near infrared (NIR) spectroscopy in food production. Trends in Food Science & Technology 46, 2 (December 2015), 211–230. https://doi.org/10.1016/j.tifs.2015.10.002

[2] Jia-Huan Qu, Dan Liu, Jun-Hu Cheng, Da-Wen Sun, Ji Ma, Hongbin Pu, and Xin-An Zeng. 2015. Applications of near-infrared spectroscopy in food safety evaluation and control: a review of recent research advances. Critical Reviews in Food Science and Nutrition, 55, 13 (May 2015), 1939–1954. https://doi.org/10.1080/10408398.2013.871693

[3] Yongni Shao, Yidan Bao, and Yong He. 2011. Visible/near-infrared spectra for linear and nonlinear calibrations: A case to predict soluble solids contents and pH value in peach. Food and Bioprocess Technology 4 (November 2011), 1376–1383. https://doi.org/10.1007/s11947-009-0227-6

[4] Wenchuan Guo, Lijie Fang, Dayang Liu, and Zhuanwei Wang. 2015. Determination of soluble solids content and firmness of pears during ripening by using dielectric spectroscopy. Computers and Electronics in Agriculture 117 (September 2015), 226–233. https://doi.org/10.1016/j.compag.2015.08.012

[5] Jiangbo Li, Wenqian Huang, Chunjiang Zhao, Baohua Zhang. 2013. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. Journal of Food Engineering 116, 2 (May 2013), 324–332. https://doi.org/10.1016/j.jfoodeng.2012.11.007

[6] Paloma A. M. Nascimento, Lívia C. de Carvalho, Luis C. C. Júnior, Fabíola M. V. Pereira, Gustavo H. de Almeida Teixeira. 2016. Robust PLS models for soluble solids content and firmness determination in low chilling peach using near-infrared spectroscopy (NIR). Postharvest Biology and Technology 111 (January 2016), 345–351. https://doi.org/10.1016/j.postharvbio.2015.08.006

[7] Peano, C., Reita, G. and Chiabrando, V., 2006. Firmness and soluble solids assessment of nectarines by NIRS spectroscopy. ISHS Acta Horticulturae 713 (July 2006), 465–469. https://doi.org/10.17660/ActaHortic.2006.713.70

[8] Lingling Li; Yuan Wu, Lian Li, and Bingqing Huang. 2017. Rapid detecting SSC and TAC of peaches based on NIR spectroscopy. In 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA). IEEE, Beijing, China, 312–317. DOI: 10.1109/CIAPP.2017.8167229

[9] Panmanas Sirisomboon, MunehiroTanaka, Takayuki Kojima, and Phil Williams. 2012. Nondestructive estimation of maturity and textural properties on tomato 'Momotaro'by near infrared spectroscopy. Journal of Food Engineering 112, 3 (October 2012), 218–226. https://doi.org/10.1016/j.jfoodeng.2012.04.007

[10] Frédéric Kosmowski and Tigist Worku. 2018. Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. PLoS ONE 13, 3 (March 2018). https://doi.org/10.1371/journal.pone.0193620

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, 1097–1105

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, 770–778. DOI: 10.1109/CVPR.2016.90

[13] Zhe Xu, Xiaomin Zhao, Xi Guo, and Jiaxin Guo. 2019. Deep learning application for predicting soil organic matter content by VIS-NIR spectroscopy. Computational Intelligence and Neuroscience 2019 (November 2019), Article ID 3563761, 11 pages. https://doi.org/10.1155/2019/3563761

[14] Lei Hou, QingXiang Wu, Qiyan Sun, Heng Yang, and Pengfei Li. 2016. Fruit recognition based on convolution neural network. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Changsha, 18–22, DOI: 10.1109/FSKD.2016.7603144

[15] Zaidah Ibrahim, Nurbaity Sabri, and Dino Isa. 2018. Palm oil fresh fruit bunch ripeness grading recognition using convolutional neural network. Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 10, 3-2 (September 2018), 109–113

[16] DLP NIRscan Nano EVM User's Guide. Retrieved June 21, 2020 from https://www.ti.com/lit/ug/dlpu030g/dlpu030g.pdf?ts=1592754432447&ref_url=https%253A%252F%252Fwww.google.com%252F

[17] Duy K. Ninh, Masanori Morise, and Yoichi Yamashita. 2013. A generation error function considering dynamic properties of speech parameters for minimum generation error training for hidden Markov model-based speech synthesis. Acoustical Science and Technology 34, 2 (February 2013), 123–132. https://doi.org/10.1250/ast.34.123

[18] Abraham Savitzky and Marcel J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36, 8 (July 1964), 1627–1639. https://doi.org/10.1021/ac60214a047