Food Grading System Using Support Vector Machine and YOLOv3 Methods

Lili Zhu and Petros Spachos
School of Engineering, University of Guelph, Guelph, ON, Canada

***Abstract*—The quality and safety of food is a great concern to the whole society because it is the most basic guarantee for human health and social development and stability. Ensuring food quality and safety is a complex process, and all stages of food processing must be considered, from cultivating, harvesting and storage to preparation and consumption. Grading is one of the essential processes to control food quality. This paper proposed a two-layer image processing system based on machine learning for banana grading. Support Vector Machine is the first layer to classify bananas based on an extracted feature vector that is composed of colour and texture features and YOLOv3 follows up for further locating the defected area on the peel and determining if the inputs belong to mid-ripened or well-ripened class. The performance of the first layer achieved an accuracy of 98.5% and the accuracy of the second layer is 85.7%. The overall accuracy is 96.4%.**

***Index Terms*—Machine Vision, Image Processing, Food Processing, Machine Learning.**

I. Introduction

Food processing takes raw materials and converts them into more suitable forms for the dietary habits of modern people. Food processing includes a series of physical and chemical changes. During the entire process, not only the nutrition of the raw materials needs to be maintained to the greatest extent, but also the poisonous and harmful substances should be prevented from entering the food. Therefore, food processing is highly valued to food scientists, the food industry, and society [1].

To meet the increasing expectations and standards of food processing, the quality inspection of food and agricultural produce is arduous and labour-intensive. After years of rapid development, Machine Vision System (MVS) has now penetrated several aspects of people's lives. Its high efficiency and accuracy assist various industries to save a large amount of labour [2], [3]. In the field of agriculture, agri-technology and precision farming is an interdisciplinary science that integrates with MVS and utilizes data-intensive methods to achieve high agricultural yields while reducing environmental impact. MVS can acquire image data in a variety of land-based and aerial-based methods and can complete multiple types of tasks as well, such as quality and safety inspection, agriculture produce grading, foreign objects detection, and crop monitoring [4]. In food processing, MVS can collect a series of parameters such as size, weight, shape, texture, and colour of food, and even many details that human eyes cannot observe to monitor and operate the food processing. In this way, fatigue and mistakes of workers caused by a large number of repeated labours can be avoided [5].

Banana is one of the most important tropical fruits and a basic staple food for many developing countries. However, banana pests and diseases pose a threat to sustainable production, and banana yellow leaf disease caused by Panama disease is a destructive disease for bananas [6]. Additionally, the ripening process of bananas is so rapid that a large number of over-ripened bananas cannot enter the market. As a result, researchers are interested in developing automatic monitoring systems to assist banana management. In this paper, a novel two-layer system to realize banana grading and defect detection is proposed. The system is composed of the first-layer classifier Support Vector Machine (SVM) and the second-layer classifier YOLOv3. A feature vector that contains extracted colour and texture information is the input for the first-layer classifier and the output of the first-layer classifier is connected to the second-layer classifier. This network can provide both of the banana ripeness level classification and peel defected area detection so that it can be a solid foundation for further implementing into an Internet of Things (IoT) system. The proposed system is compared with three other popular systems and it can provide higher accuracy.

II. Related Works

MVS applications have been applied to multiple research areas of food processing, such as food safety and quality evaluation, food process monitoring, and foreign object detection. Regarding food safety and quality evaluation, in [7] and [8], hyperspectral reflectance imaging methods were applied to determine the bruise or damage of blueberry. Pattern recognition algorithm was adopted to separate stem and calyx and detected blueberries with diseases and the orientations of blueberries [9]. In [10], not only the proposed model realized grading the fruits with Multi-Attribute Decision Making (MADM) but also it successfully predicted the number of actual days that the harvested mangoes can be sent away with Support Vector Regression (SVR). In [11], the authors built a data set of rice – FIST-Rice with 30,000 rice kernel samples and developed a system called Deep-Rice for grading rice by extracting the discriminative features from several angles of the rice. In [12], the authors adopted Artificial Neural Networks (ANN) to classify the shapes of boiled shrimps by accepting the Relative Internal Distance (RID) values. For banana grading, in [13], the author developed a method to classify bananas into healthy and unhealthy groups based on

978-1-7281-8086-1/20/$31.00 ©2020 IEEE

image processing techniques and a neural network and they obtained an accuracy of 97%. In [14], the authors designed a method to detect at which ripening stages red bananas are by measuring the dielectric properties of red banana and sending the features to a Fuzzy C means (FCM) classifier. In [15], the authors also adopted a fuzzy model that was optimized with particle swarm optimization (PSO) technique to grade unripe, ripe and overripe bananas with the features of the bananas' peak hue and normalized brown area. The accuracy of this model is 93.11%. A random forest classifier was utilized to grade the bananas according to the colour features in [16] and accuracy arrived at 94.2%. In [17], the authors also adopted machine learning algorithms to classify different types of bananas and their ripeness levels. SVM achieved an accuracy of 99.1% to classify the banana types and has a 96.6% accuracy to distinguish the level of ripeness. In [18], an ANN outperforms other machine learning algorithms to detect the ripeness of bananas with a feature vector that consists of colour and texture features and the classification accuracy of this system is 97.75%. In comparison to the reviewed related work, this study proposes a two-layer mechanism to realize both of the banana grading tasks and defected area locating mission. Traditional data augmentation methods and a deep learning-based architecture called CycleGAN were adopted to enlarge the data set which was built by the authors to avoid overfitting. Then, a feature vector that contains colour and texture features was used to train the first layer classifier. Next, the YOLOv3 model can detect the defected areas on the fruit peel in the images in the ripened class, which is one of the first layer's output classes.

III. Data set and Methodologies

A. Data set

In this study, we created the data set as the existing online open-access banana data sets only contain bananas in good conditions. We took 150 images of bananas at different ripeness levels and labelled them into three main groups, which are unripened, ripened, and over-ripened (50 images for each group). The ripened class has two sub-classes that are mid-ripened and well-ripened. The bananas in the unripened group are those still in the green peel, while bananas in ripened and over-ripened groups have a yellowish peel and different levels of brown spots. However, 150 samples are not satisfactory for machine learning methodologies as it is easy to cause overfitting. As a result, we adopted data augmentation techniques, including traditional methods and a deep learning method – CycleGAN, to enlarge the data set.

Traditional data augmentation methods such as rotation, flipping, and shifting are widely used for machine learning training. We also adopted CycleGAN to generate images of defected bananas. Generative adversarial net (GAN) [19] is a generative model to learn the data distribution via an adversarial mode between a generating network and a discriminating network. The generating network generates samples that are similar to the true samples as much as possible, while the discriminating network tries to determine whether the samples

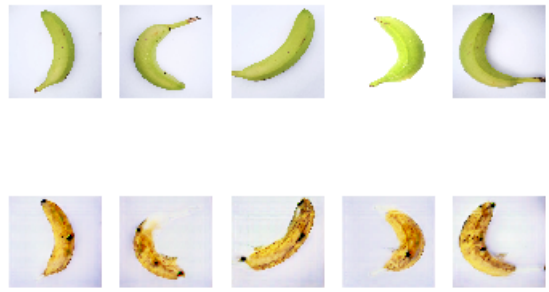


Fig. 1. Five ripened banana images generated by CycleGAN (the second row) based on the unripened banana images (the first row).

are real samples or generated false samples. CycleGAN [20] makes the principle of GAN apply to the image generation with ease. Based on GAN, CycleGAN adds another pair of the generator - discriminator and cycle consistency loss as well to determine whether the style of the generated images are consistent with the images in the original data set. If two GAN are being trained at the same time, one of the generator - discriminator pairs is G_{A2B} and D_B and the other pair is G_{B2A} and D_A. Then, an image x of style A should be able to transform back to itself after two transformations and image y of style B is the same as:

$$\begin{aligned} G_{B2A}(G_{A2B}(X)) &\simeq x \\ G_{A2B}(G_{B2A}(X)) &\simeq y \end{aligned} \tag{1}$$

The first-order distance between the two graphs can be expressed as:

$$\begin{aligned} L_{cyc}(G_{A2B}, G_{B2A}, A, B) = \mathbb{E}_{x\sim A}[\| G_{B2A}(G_{A2B}(x)) - x \|_1] \\ + \mathbb{E}_{y\sim B}[\| G_{A2B}(G_{B2A}(y)) - y \|_1] \end{aligned} \tag{2}$$

Eq. 2 is the cycle consistency loss and is one of the terms of the total loss function:

$$\begin{aligned} L(G_{A2B}, G_{B2A}, D_A, D_B) = L_G(G_{A2B}, D_B, A, B) \\ + L_G(G_{B2A}, D_A, B, A) \\ + \lambda L_{cyc}(G_{A2B}, G_{B2A}, A, B), \end{aligned} \tag{3}$$

where $L_G(G_{A2B}, D_B, A, B)$ is the loss of G_{A2B} and D_B and $L_G(G_{B2A}, D_A, B, A)$ is the loss of G_{B2A} and D_A. The expectation of CycleGan model is :

$$\begin{aligned} G^*_{A2B}, G^*_{B2A} = arg \min_{G_{A2B}, G_{B2A}} \max_{D_A, D_B} \\ L(G_{A2B}, G_{B2A}, D_A, D_B) \end{aligned} \tag{4}$$

Figure 1 shows the comparison between the original unripened banana images and the generated ripened images. 150 new ripened banana images were created by CycleGAN model. The total data set after data augmentation is as Table I.

TABLE I
THE TOTAL DATA SET AFTER DATA AUGMENTATION.

Original	Rotation	Flipping	Shifting	CycleGAN	Total
150	150	150	150	50	650

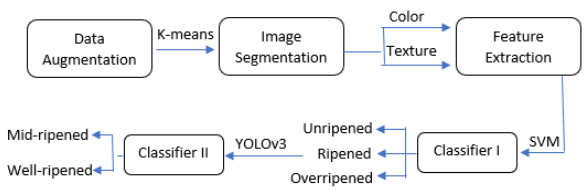


Fig. 2. The flowchart of the proposed banana grading system.

Fig. 3. The sample inputs (the upper row), the masks (the middle row) and the outputs (the lower row) of image segmentation step.

B. Methodologies

The proposed banana grading system includes data augmentation, image segmentation, feature extraction, and classification. Figure 2 shows the flowchart of the system.

1) Image segmentation: Image segmentation is to separate the target from the background in an image. This is the first step of image analysis, the basis of computer vision, an important part of image understanding, and one of the most difficult problems in image processing. For a grayscale image, the pixels inside the region generally have intensity similarity but have discontinuous intensities at the boundaries of the region. Methods for processing image segmentation include thresholding, region-based segmentation, edge detection-based algorithms, and machine learning-based methods. In this work, to get closer to practical applications, image acquisition was performed in natural light, so there is inconsistent brightness in the background and shadows as well. As a result, it is difficult to find a suitable threshold and complete and accurate edges to segment the target. Therefore, K-means is used here to address this task. K-means utilizes distance as the evaluation index of similarity. The basic idea is to cluster the samples into different clusters according to the distance. The closer the two points are, the greater the similarity. Finally, all the data are allocated to the closest cluster center so that the sum of the squares of the distances between each point and its corresponding cluster center is minimized. Before applying K-means, rank filter and log transformation were adopted to reduce noise and improve image contrast. Figure 3 shows the sample segmentation results.

2) Feature Extraction: For images, each image has its characteristics that can be distinguished from other types of images. Some are natural features that can be intuitively felt, such as brightness, edges, texture, and colour. Some require transformation or processing to obtain, such as moments, histograms, and principal components. These features will be extracted in the form of numerical values or vectors so that the computer can identify images. Common image features are colour features, texture features, shape features, and spatial features.

a) Color Features: The colour feature is a global feature, which describes the surface properties of the targets corresponding to the image or image area. The general colour feature is based on the characteristics of the pixels and all pixels that belong to the image or image area have their contributions. Colour features can be extracted by colour histograms, colour sets, colour moments, colour coherence vector and so on. In this work, because the unripened, ripened, and overripened bananas have distinctive colour features (green, yellow, and brown) and it is unnecessary to consider the colour space distribution, the colour feature is one of the components that are extracted to train the classifier. Traditionally, RGB (Red, Greem, Blue) colour space is prevalent in digital image processing. However, HSV (Hue, Saturation, Value) colour space is closer to how humans perceive colour and more suitable for statistical analysis than RGB colour space. Therefore, the colour features of the proposed data set were extracted in the HSV colour space. Eq. 5, Eq. 6, and Eq. 7 can explain that how RGB color space converts to HSV color space.

$$V = max(\frac{R}{255}, \frac{G}{255}, \frac{B}{255}) \tag{5}$$

$$S = 1 - \frac{3}{(R+G+B)}[min(R,G,B)] \tag{6}$$

$$H = \begin{cases} \theta, & G \geq B \\ 2\pi - \theta, & G < B, \end{cases} \tag{7}$$

$$where\ \theta = cos^{-1}\left[\frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right].$$

Due to the colour characteristics of the three groups of bananas, the corresponding H, S, and V value ranges are acquired from the analogy between HTML colour codes and the natural colours of different banana peels [13]. Table II illustrates that H and V value ranges are distinct to be two of the input features to classify the bananas.

TABLE II
THE RANGE FOR H, S, V VALUES IN HSV COLOR SPACE FOR TWO BANANA GROUPS.

	Unripened	Ripened
H	$72° \leq H \leq 78°$	$39° \leq H \leq 72°$
S	$85\% \leq S \leq 100\%$	$70\% \leq S \leq 100\%$
V	$27\% \leq V \leq 50\%$	$69\% \leq V \leq 100\%$

b) Texture Features: Texture is another natural characteristic of the surface of an object. It describes the distribution of the gray space between the image pixels and the image area, and it will not change with different illumination. The texture feature is a global feature as well. However, because texture is only a characteristic of the surface of an object and cannot fully reflect the essential attributes of the object, it is impossible to obtain high-level image representation using only texture features. Unlike colour features, texture features are not pixel-based features, they require statistical calculations in an area containing multiple pixels.

As a statistical feature, texture features often have rotation invariance and are robust to noise. Local Binary Pattern (LBP) [21] is an operator used to describe the local texture features of an image and it has significant advantages such as rotation invariance and gray invariance. The basic LBP operator is defined as a 3×3 size texture unit and the value of the center pixel is the threshold. The grayscale value of the adjacent 8 pixels is compared with the pixel value of the center of the unit. If the adjacent pixel values are greater than the center pixel value, the position of the pixel is marked as 1, otherwise, it is 0. In this way, 8 pixels in the 3×3 unit can generate 8-bit binary numbers after compared to the center pixel. These 8-bit binary numbers are arranged in sequence to form a binary number. This binary number is the LBP value of the center pixel. Therefore, there are 256 LBP values and the LBP value of the center pixel reflects the texture information of the area around the pixel. Mathematically, the process can be expressed as Eq. 8 and Eq. 9, where g_c is the center pixel value and g_0 is the adjacent pixel value.

$$s(g_0 - g_c) = \begin{cases} 1, g_0 - g_c \geq 0 \\ 0, g_0 - g_c < 0 \end{cases} \tag{8}$$

$$LBP = \sum_{p=0}^{7} s(g_0 - g_c) 2^p \tag{9}$$

Figure 4 illustrates that bananas at different ripeness levels can show distinctive texture features that were extracted by the LBP operator.

3) Classification: In this work, the classification task is divided into two steps. The first step is to feed extracted features into SVM to classify unripened, ripened, and over-ripened banana groups. The brown spots on banana peel will not be detected here as there is no consistency of what level of brown colour that should be included and for over-ripened bananas, the detection of multiple irregular areas that are caused by the connected brown areas will result in inaccurate results. Additionally, it is unnecessary to detect the brown spots for over-ripened bananas due to the peel is mainly brown. As a result, the bananas will be classified into the basic group. The next step is to feed the output ripened fruit images from SVM into the YOLOv3 [22] transfer learning model to detect the brown spots and separate the bananas into mid-ripened and well-ripened groups according to how many the brown areas they have.

Fig. 4. A sample image: (left) segmented image; (middle) grayscale image; (right) the texture feature that was extracted by the LBP operator.

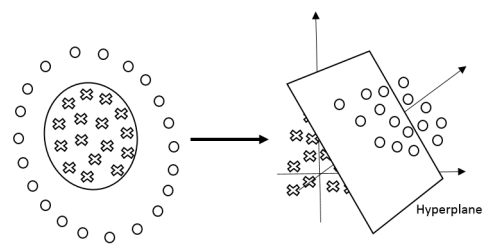


Fig. 5. Nonlinear data classification in high-dimensional feature space with the SVM.

a) Support Vector Machine: Cortes and Vapnik proposed SVM in 1995 [23] that is a supervised learning method and can be widely applied to statistical classification and regression analysis. Its basic model is a linear classifier defined on the feature space to find the hyperplane that has the maximum interval between two types of data. The learning strategy of support vector machines is to maximize the interval, which can be formalized as a problem to solve the convex quadratic programming, which is also equivalent to the minimization problem of a regularized hinge loss function. However, the data is not linearly separable for most of the time. Under this circumstance, the hyperplane that meets the condition does not exist at all. For nonlinear situations, the SVM approach is to choose a kernel function. The SVM first completes the calculation in the low-dimensional space and then maps the input space to the high-dimensional feature space through the kernel function. Finally, the optimal separating hyperplane is constructed in the high-dimensional feature space, so that the nonlinear data are separated, as shown in Figure 5.

For multi-classes tasks such as the one in this work, a nonlinear SVM with Radial Basis Function (RBF) kernel can

be applied. In this case, the feature vector that was input into the SVM classifier is:

$$A = [H\ V\ LBP] \tag{10}$$

b) YOLOv3: YOLO [24] is an object recognition and localization algorithm based on deep neural networks. The most distinct feature of YOLO is that it runs fast and can be used in real-time systems. However, its mPA towards small objects is not satisfactory. On the premise of maintaining the speed advantage of YOLO, YOLOv3 [22] adopted the residual network structure to form a deeper network level and uses multi-scale features for object detection. Also, object classification uses Logistic instead of softmax, which improves prediction accuracy, especially for small Object recognition capabilities. In YOLOv3, there are only convolution layers, and the size of the output feature map is controlled by adjusting the convolution step. Therefore, there is no particular limitation on the size of the input picture. YOLOv3 draws on the idea of Feature Pyramid Networks (FPN) – small size feature maps are used to detect large-sized objects while large size feature maps are used to detect small-sized objects.

YOLOv3 model [25] was adopted because many of the brown spots on the banana peel are small targets that need to be detected and located. When three or less defected areas are detected by the YOLOv3 model, this sample will be considered as a mid-ripened sample. Also, a well-ripened sample is determined by whether there are more than three defected areas that are found by the model.

4) Method Evaluation: The accuracy, sensitivity/recall, and precision, all of which are common evaluation methods in statistics, are used to evaluate the first-layer classification results, as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$Sensitivity/Recall = \frac{TP}{TP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$TP = True\ positive, TN = True\ negative,$$
$$FP = False\ positive, FN = False\ negative.$$

To further assess the performance of the YOLOv3 model, mAP (mean Average Precision), Intersection over Union (IoU), and recall are applied to evaluate the predicted locations of the target. The definitions of the evaluation methods are expressed in Figure 6.

IV. Experiments Results and Discussion

The experiments were conducted on the Ubuntu 18.04.4 LTS system, with an Intel® Core™ i7-8700K CPU @ 3.70GHz × 12 processor, 32G memory, and GeForce GTX 1080 Ti/PCIe/SSE2 graphic.

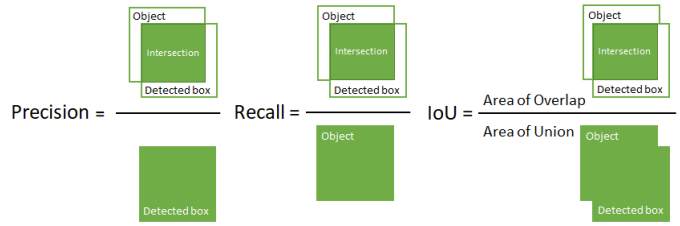
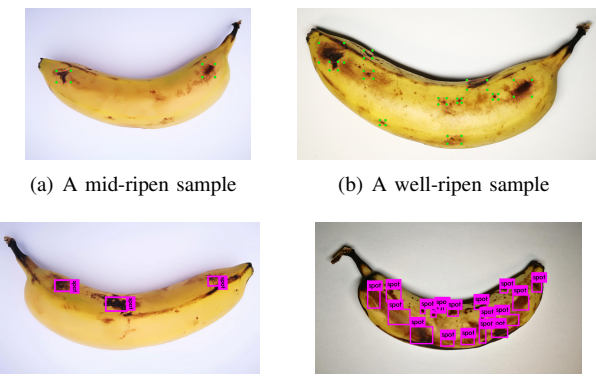


Fig. 6. The definitions of the evaluation methods.

TABLE III
THE CONFUSION MATRIX FOR THE FIRST-LAYER CLASSIFIER.

True Class	Predicted Class			Sensitivity
	Unripened	Ripened	Overripened	
Unripened	39	0	0	100
Ripened	0	48	1	98.0%
Overripened	0	1	41	97.6%
Precision	100	98.0%	97.6%	Acc=98.5%



(a) A mid-ripen sample (b) A well-ripen sample

(c) A predicted mid-ripen sample (d) A predicted well-ripen sample

Fig. 7. The sample ground truth data and the predicted samples in the mid-ripen and well-ripen classes

For the first classification layer, 520 images were used to train the SVM classifier and 130 images were for testing. After forming the input feature vector $A = [H\ V\ LBP]$ with the H value, V value, and LBP features, A of all the training images were fed into the SVM classifier for training. Table III shows the confusion matrix for the first layer classifier's testing result. This confusion matrix demonstrates that the overall predicting accuracy of the SVM classifier achieved 98.5% (with g = 0.005 and C = 1000).

For the second layer, the defected areas of the images in the ripened group were labelled with an open-source tool called "labelImg" [26] manually and the 48 images that were predicted as ripened were fed to the second predictor. Figure 7 shows one sample ground truth data from each class.

After 100,000 iterations of training, the mAP of the testing results is 0.8137, the average IoU is 76.34%, the average recall and precision of the testing results is 91.32% and 74.93%, respectively. The average processing time to predict a testing image is 0.051 seconds. The reason for the high recall and low precision could be that the model detected some spot areas that

TABLE IV
THE CONFUSION MATRIX FOR THE SECOND-LAYER CLASSIFIER.

True Class	Predicted Class		Sensitivity
	Mid-ripened	Well-ripened	
Mid-ripened	18	2	90.0%
Well-ripened	3	25	89.3%
Precision	85.7%	92.6%	Acc=89.5%

were not labelled on the ground truth data. According to the results of the detected areas, to which sub-class this sample belongs will be determined by the number of the detected areas. When the detected areas are more than five, this banana will be classified to the well-ripened group. The IoU result indicates that the predicted areas shifted from the ground truth labels to some extent, but this will not affect the number of predicted areas. As a result, the confusion matrix, shown in Table IV, that is based on the number of predicted areas is still valid.

V. CONCLUSION

This paper proposed a novel two-layer classifier to grade bananas in terms of their ripeness levels. Because there is redundant information in the original images and this information will decrease the classification accuracy, a feature vector that is composed of the essential colour and texture information was formed. The experiment results illustrate that the extracted feature vector assisted the SVM classifier to achieve an accuracy of 98.5%. Then the YOLOv3 system continued to detect the defected area of the ripened bananas and separated them into the mid-ripened and the well-ripened groups. This system accomplished the banana grading mission in terms of ripeness and also conquered the difficulty of detecting and outputting small defected areas on bananas. The future work will continue to improve the performance of the second-layer classifier by labelling the defected areas more precisely and fine-tuning the model and develop an Internet of Things system to integrate with the grading system and monitor the quality of fruits in real-time.

REFERENCES

[1] "The impact of food processing on antioxidants in vegetable oils, fruits and vegetables," *Trends in Food Science Technology*, vol. 9, no. 8, pp. 336 – 340, 1998.

[2] L. Zhu and P. Spachos, "Towards image classification with machine learning methodologies for smartphones," *Machine Learning and Knowledge Extraction*, vol. 1, no. 4, p. 1039–1057, Oct 2019. [Online]. Available: http://dx.doi.org/10.3390/make1040059

[3] L. Zhu and P. Spachos, "Butterfly classification with machine learning methodologies for an android application," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019, pp. 1–5.

[4] "Machine vision technology for agricultural applications," *Computers and Electronics in Agriculture*, vol. 36, no. 2, pp. 173 – 191, 2002.

[5] S. Cubero, N. Aleixos, E. Molto, J. Gmez-Sanchis, and J. Blasco, "Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables," *Food Bioprocess Technol.*, pp. 1–18, 01 2010.

[6] R. Ploetz, "Panama disease:a classic and destructive disease of banana," *Plant Health Progress*, vol. 10, 01 2000.

[7] S. Fan, C. Li, W. Huang, and L. Chen, "Detection of blueberry internal bruising over time using nir hyperspectral reflectance imaging with optimum wavelengths," *Postharvest Biology and Technology*, vol. 134, pp. 55–66, 12 2017.

[8] M.-H. Hu, Q. Dong, and B.-L. Liu, "Classification and characterization of blueberry mechanical damage with time evolution using reflectance, transmittance and interactance imaging spectroscopy," *Computers and Electronics in Agriculture*, vol. 122, pp. 19–28, 03 2016.

[9] G. Leiva-Valenzuela and J. Aguilera, "Automatic detection of orientation and diseases in blueberries using image analysis to improve their postharvest storage quality," *Food Control*, vol. 33, p. 166–173, 09 2013.

[10] C. Nandi, B. Tudu, and C. Koley, "A machine vision technique for grading of harvested mangoes based on maturity and quality," *IEEE Sensors Journal*, vol. 16, pp. 1–1, 08 2016.

[11] Y. Wu, Z. Yang, W. Wu, X. Li, and D. Tao, "Deep-rice: Deep multi-sensor image recognition for grading rice*," in *2018 IEEE International Conference on Information and Automation (ICIA)*, Aug 2018, pp. 116–120.

[12] P. Poonnoy, P. Yodkeaw, A. Sriwai, P. Umongkol, and S. Intamoon, "Classification of boiled shrimp's shape using image analysis and artificial neural network model," *Journal of Food Process Engineering*, vol. 37, no. 3, pp. 257–263.

[13] F. Mazen and A. Nashat, "Ripeness classification of bananas using an artificial neural network," *Arabian Journal for Science and Engineering*, 01 2019.

[14] A. Mohapatra, S. Shanmugasundaram, and R. Malmathanraj, "Grading of ripening stages of red banana using dielectric properties changes and image processing approach," *Computers and Electronics in Agriculture*, vol. 143, pp. 100–110, 2017.

[15] S. Marimuthu and S. M. M. Roomi, "Particle swarm optimized fuzzy model for the classification of banana ripeness," *IEEE Sensors Journal*, vol. 17, no. 15, pp. 4903–4915, Aug 2017.

[16] E. Piedad, J. Larada, G. Pojas, and L. Ferrer, "Postharvest classification of banana (musa acuminata) using tier-based machine learning," *Postharvest Biology and Technology*, vol. 145, pp. 93–100, 11 2018.

[17] I. A. Sabilla, C. S. Wahyuni, C. Fatichah, and D. Herumurti, "Determining banana types and ripeness from image using machine learning methods," in *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, March 2019, pp. 407–412.

[18] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Intelligent grading system for banana fruit using neural network arbitration," *Journal of Food Process Engineering*, vol. 40, no. 1, p. e12335, 2017.

[19] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017.

[21] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1, Oct 1994, pp. 582–585 vol.1.

[22] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[23] C. Cortes and V. Vapnik, "Support-vector networks," in *Machine Learning*, 1995, pp. 273–297.

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015.

[25] J. Redmon, "Darknet: Open source neural networks in c," http://pjreddie.com/darknet/, 2013–2016.

[26] Tzutalin, "Labelimg," *Git code*, 2015. [Online]. Available: https://github.com/tzutalin/labelImg