Analysis and Detection of Tomatoes Quality using Machine Learning Algorithm and Image Processing

Haichun Zuo
School of Electrical and Computer Engineering,
Nanfang College Guangzhou, Guangzhou
Guangdong, 510970, China

***Abstract*—Grading of agricultural products methods based on artificial intelligence is more important. Because these methods have the ability to learn and thus increase the flexibility of the system. In this paper, image processing systems, detection analysis methods, and artificial intelligence are used to grade tomatoes, and the success rate of grading these methods is compared with each other. However, the purpose of this study is to obtain a solution to detect appearance defects and grade and sort the tomato crop and provide an efficient system in this field. A visual dataset is created, to investigate the approach of image processing and machine learning based on a tomato image. Tomato models are placed individually under the camera and samples are classified in a lighting box away from the effects of ambient light. Data sets have been used in three types of first, second, and third quality categories. It should be noted that quality category one has the best quality and quality category two has the medium quality and category three has the worst quality, Also, each data class contains 80 samples. Using tomato appearance such as size, texture, color, shape, etc. Image processing is performed for extract features. Tomato images are pre-processed for optimization. Then, to prepare for classification, the dimensions of the images are reduced by principal component analysis (PCA). Three categories of an artificial neural network, a support vector machine, and a decision tree are compared to show the most efficient support machine. The analysis is examined in two classes and three classes. The support vector machine has the best accuracy compared to other methods so this rate is 99.9% for two classes and 99.79% for three classes.**

***Keywords*—Machine learning; image processing; product category; tomato quality rating**

I. Introduction

Farmers usually separate healthy and damaged tomatoes according to their size and quality [1,2]. The acceptable quality of tomatoes makes separating the healthy from the bad more accessible, and it prevents the spread of rotten tomatoes among healthy ones [3,4]. Damaged tomatoes are either sold at a lower price, or discarded. Sorting tomatoes in the traditional way are still done by old people, which takes a lot of time. Relying on a product defect alone can lead to a fundamental error by relying on human inspection. The desire to carry quality tomatoes with undivided defects has a great impact on customer feedback and satisfaction. In this study, Image processing and machine learning technology are used. In this way, an image of a tomato is taken, and the presence of defects is automatically detected in a computer vision system to distinguish healthy from damaged tomatoes, using the concept of a developed system [5,6].

Every day, intelligent systems with industrial applications are increasing. The most important biological processes in the production of crops are the classification of vegetables and fruits. Still, such processes are done manually in a country like Egypt based on a database [7]. After wheat, tomato is the eighth most used crop. The global production of tomatoes is reported to be about 159 million tons for 144 countries. Fresh fruit revenue in 2011 is about $582 trillion [8]. The performance, quality, and weight of the products are excellent in five countries including the United States, China, India, Turkey, and Egypt. Although very important in the production of previously mentioned tomatoes, there is very little literature on the works. In recent years, tomato grading has become very important, especially with the detection of diseases, due to new market restrictions. Therefore, the need for new technologies in the process of separation and product quality monitoring has increased.

Daily, millions of people use tomatoes and vegetables, that's why it is one of the most popular food products in human life. However, labor costs have increased as the workforce ages, making many farms less profitable. As mentioned, tomato is a popular food product among people, so with the increase in population, tomatoes need to be produced more. An effective solution to control quality and reduce costs is to use a robot instead of human labor to harvest tomatoes. Therefore, most researchers have spent the past few decades building robots to harvest fruits and vegetables [9,10]. Tomato color is the main indicator for the detection of ripening. Different steps are taken to produce tomato fruit. These stages change the tomato's color from green to light pink, then to pink, then to bright red, and finally to red, which categorizes them into different categories [11,12]. The closer the tomato is red, the better its quality. and leads to an increase in prices for companies. The storage time for a quality harvest is about 70 to 75 days in total. The time stages of tomato ripening include 21 to 28 days for green, 15 to 20 days for light pink, 7 to 14 days for pink, 5 to 6 days for bright red, and 2 to 4 days for red [13,14]. Therefore, improving the tomato classification system is an important task for designing a crop robot. In recent years, machine vision and pattern recognition methods have been well received by researchers, especially in the processing or sorting of many intelligent agricultural products. According to EU regulations on processed vegetables and fruits, the main condition is that the tomatoes arrive intact, completely red, fresh, without rotting and cracks in satisfactory condition, and without damage. The sorting and grading method is done manually. As mentioned, the manual method has low accuracy, high cost, and high effort [15,16]. Efficiency sorting governs how the product is delivered through packaging lines and quality standards. Therefore, as stated above, it can be concluded that the importance of this research is the requirement of high accuracy, high speed, and low cost is very important for sorting using machinery tomato detection. The effective solution is the use of mechanical sorting machines for tomatoes. But mechanical sorting machines can only classify tomatoes according to size and weight.

Now, with the advancement of technology and the introduction of vision devices, it covers the limitations of mechanical devices. New generation machines classify tomatoes with high accuracy, high speed, and low cost [17,18]. In 1985, Sarkar et al. reported the first tomato classification based on machine vision. This study is about the classification of tomatoes according to their size, shape, color, and defects [19-21]. Classification of products is an important step in the packaging and storage processes, which image processing is one of the practical tools in the field of post-harvest technologies. Izadi et al., obtain an algorithm for detecting physical defects [22] and degree tomato product classification and providing an efficient system in this field [23-25]. The specific objectives of this research are to develop a computer vision system and develop an efficient image processing algorithm, to create an accurate and accurate detection algorithm, and to create different classifications for classifying tomatoes into different grading categories [26-30]. One of the purposes of this study is to develop an image processing algorithm and develop a computer vision system, which is used immediately after harvesting tomatoes. Therefore, the time of the process of picking and classifying tomatoes is greatly reduced, in addition, due to the high accuracy of the optical device, quality sorting is done.

Given that image processing and machine vision are one of the current technologies in the world, the main task of a researcher in the field of using machine vision in agriculture is to extract the properties of objects in digital images using these technologies [31-33].

The most important sensory feature of fruits and vegetables is the appearance that affects the market value, priority, and consumer choice. Although sorting and grading are done by humans, it is inconsistent, time-consuming, variable, subjective, heavy, expensive, and easily affected by the environment. Hence, a clever fruit grading system is needed. Therefore, the motivation of this study is to provide a method of quality detection and automatic grading for many tomato disorders.

Most agronomic disorders are observed according to roots, stems, leaves, and fruits based on the type and causes of damage. Fruit images are used for diagnostic purposes. The image acquisition method used to collect the data set ensures that the entire fruit is examined. In the first part, general fruits, especially tomatoes, are introduced. The second part presents the general method of grading tomatoes based on fruit, image processing algorithm, classification methods, machine learning systems, neural network methods, and application of appearance features, color recognition, and texture recognition in detail. In the third section, the proposed methods are introduced and the steps related to the method are described from an algorithmic point of view. In the fourth part, the proposed method is implemented on the set of images, and the evaluation criteria are evaluated. Finally, the fifth part will end with a conclusion.

II. Related Works

The most advanced image-based tomato detection method is shown in this section.

Malik et al [49] proposed a Watershed segmentation method was utilized to separate the clustered fruits according to ripe tomato identification algorithms, which were based on the HSV color space. The Otsu segmentation algorithm was used by Arum Sari et al. [50] to multiply the Cb and V channels from the YCbCr and YUV color spaces, respectively, when categorizing tomatoes into 6 separate groups.

Using the color space L*a*b*, Yin et al. [51] segmented ripe tomatoes using K-means clustering, and Indriani et al. [52] divided tomatoes into 5 groups using the HSV color space, Gray Level Co-occurrence Matrix, and K-Nearest Neighbor. To segment and localize ripe tomatoes in a greenhouse, Huang et al. [53] employed the L*a*b* color space. Goel and Sehgal [54] used fuzzy rule-based classification through the RGB color space to distinguish the fruits from the background.

Some studies use edge detection operations like the Canny [55] or the Sobel [56] operators as suggested in the Zhang [57] and Benavides et al. [58] investigations, respectively, to enhance detection and produce better segmentation.

III. Criteria and Data Sets

In this section, in the first stage, the criteria used in this study, including sensitivity, specificity, and accuracy are examined. The data set to be evaluated is then introduced.

A. Criteria

Contrast is one of the most important criteria for determining image quality. The contrast shows the difference between the darkest and lightest parts of an image. When the contrast of the image is low, the difference between the brightness of those points is small. As a result of low image contrast, the image will fade. One way to increase the contrast of a low-quality image is the Histogram Equalization smoothing technique. It changes the values of the gray surfaces of the image to cover the entire range from 0 to 255 per pixel.

1) Sensitivity: Sensitivity is one of the performance evaluation criteria of binary classification tests, and it is also used in artificial intelligence. Sensitivity measures the ratio of positively identified positives, meaning that the classifier can correctly identify the data in classes. The terms "positive" and "negative" do not mean profit, but the existence or non-existence of a condition. For example, if caries is caries, "positive" means "caries" and "negative" means "healthy". In many sensitivity tests, the degree to which positive cases are ignored is called positive, so false negatives are rare.

Sensitivity is considered to be the correct positive ratio (the percentage of tuples that are positively labeled and their class is actually positive). The sensitivity parameter is obtained based on Equation (1) [34].

$$Sensitivity = \frac{TP}{TP+FN} \qquad (1)$$

Where *TP* is the correct number of first-class predictions, and *FN* is the incorrect number of second-class predictions.

2) Specificity: Specificity measures the proportion of negatively identified negatives (for example, the percentage of healthy tomato samples that are correctly identified as not rotten). Characteristic is the category in which real negatives are classified in this way, so false positives are rare. A sensitive test seldom ignores a real positive (for example, no caries is shown despite caries). The specificity parameter is obtained based on Equation (2) [35].

$$Specificity = \frac{TN}{FP+TN} \qquad (2)$$

Where *TN* is a true negative, or healthy samples of tomatoes are properly identified as healthy. The *FN* is right, it's wrong.

3) Accuracy: Accuracy is the most intuitive measure of performance and is simply the ratio of the predicted correct observations to the total observations. It may be thought that our model is the best if we are very careful. In general, accuracy is a valuable measure when there is a symmetric data set where the negative, positive, and false positives are approximately the same. Therefore, to evaluate the performance of your model, you should look at other parameters [36].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad (3)$$

Accuracy is also used as a statistical criterion that a binary classification test correctly identifies or removes a condition. In a classification task, the accuracy for a class is the number of actual positives divided by the total number of elements labeled positively related to the class (i.e. the sum of the actual positives and False positives are items that are incorrectly labeled as belonging to the class). The recall is defined as a positive number divided by the total number of elements that actually belong to the positive class, (i.e. the sum of false positives and false negatives are those that do not have a positive classification label) [37,38].

B. Data Sets

One of the important factors in quality recognition and model evaluation is the use of appropriate data sets. Given that in this study, with the help of image processing and machine learning, the proposed solution to determine the quality of tomatoes is presented. Therefore, it uses data sets in three types of first, second, and third-degree quality categories. It should be noted that quality category one has the best quality and category three has the worst quality, and also each data class contains 80 samples. An example of images related to different types of tomatoes can be seen in Fig. 1. The data set is graded into three categories in terms of quality, which include first grade (best quality), second grade (medium quality, or small healthy-slightly spoiled-immature), and third grade (worst quality). In this study, in order to improve the quality of images and eliminate their potential noise, the data preprocessing method has been used. Methods including applying the median filter [39], calculating the area in an image [40], the optimal local pattern and the matrix of gray surface occurrence [41], the Local Binary Pattern method (LBP) [42], and LAB color channel [43] are used in the step of extracting the shape, color, and texture characteristics. Then, the dimensions are reduced by Principal Component Analysis (PCA) to prepare the vector for classification [44]. Three classification models are used, including the decision tree [45], the support vector machine [46], and the artificial neural network [47]. These classifiers are compared to measure the performance of different classifiers on datasets.

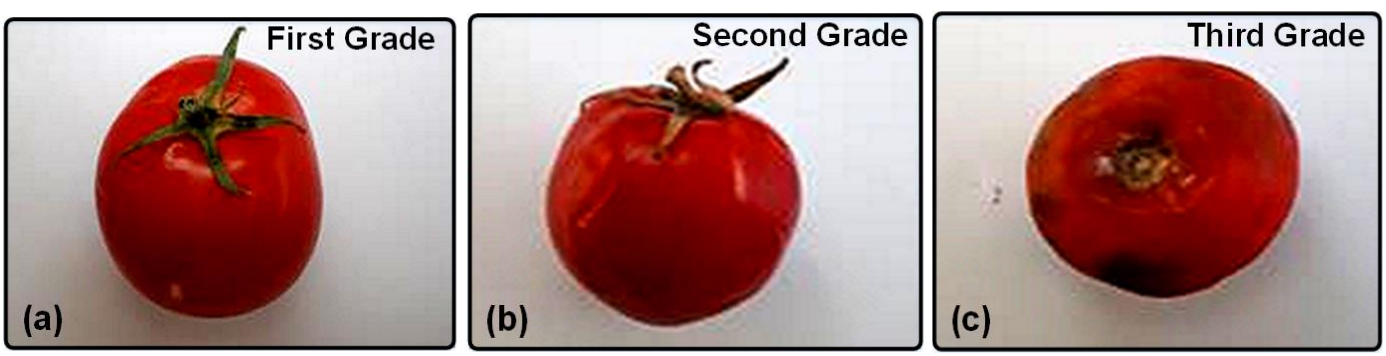

Fig. 1. Tomato sample image of class (a) first grade, (b) second grade, (c) third grade.

IV. Proposed Method

This article, with the help of features related to the shape, texture, and color of tomatoes, offers a suitable solution for quality recognition and analyzes the effect of using image processing methods in quality recognition using machine learning. The purpose of using preprocessing operations on the data set is to improve its quality and then extract the appropriate feature from them to compare with other methods and also by selecting the appropriate features finally with the help of the proposed algorithms the ability of the machine learning model in support vector machine algorithms, the artificial neural network and the decision tree are examined.

With the decreasing trend of the labor force and also the advancement of technology in order to mechanize agricultural operations, there is a possibility to distinguish the quality of agricultural products. Product quality ultimately leads to post-evaluation economic justification for continuing to grow crops. There are several factors in identifying tomatoes in different quality categories. Classification of agricultural products causes, in addition to harvesting at the right time, also causes the supply of products to be done in a timely manner, to reduce their quality and eliminate first-class products and turn them into second-class products. And three and prevent losses due to waste of time. One of the most important processes in the classification and maintenance of products is their grading and sorting, for which the use of image processing techniques is an important step in this area. The use of machine vision systems makes the post-harvest stages fast, easy, cheap, and with high accuracy, and on the other hand, reduces inspection time, quality assurance, and product grading.

For this purpose, the proposed design has been developed according to Fig. 2 in order to determine the quality of tomatoes. Accordingly, in order to increase the quality of images by applying preprocessing operations by changing the dimensions of images to the size of 400 × 300, converting color images to gray, Contrast Enhancement, and applying a median Filter. In the next step, the properties related to texture, shape, and color are extracted, so that in the next step, with the help of Principal Component Analysis (PCA), the most useful features are evaluated according to certain parameters using machine learning. Using the right data set plays an important role in achieving a reliable model. The use of appropriate hardware to some extent causes the image obtained from the cameras to be of good quality. In addition, the application of a series of algorithms and software methods causes images in the data set to have better quality. To improve the quality of images and eliminate their possible noise, preprocessing methods are used. First, the dimensions of the images are changed so that all the images by different dimensions are converted to 400 × 300, then the color images are converted to gray, the contrast is improved and the median filter is applied.

A. Feature Extraction

The median filter is the non-linear filters used to remove noise. To calculate the median in the 3 × 3 neighborhood, we have 9 cells, which after sorting them, we put the value in the middle of the array in output. One of the important features of this filter is maintaining the edges and their position without displacement. Fig. 3 shows a good example of the function of this filter in simple language.

To extract the shape feature, the appearance of the tomato and the space occupied by it has been used. In other words, according to Fig. 4, the path area within the boundary of an area is declared as a scale.

The LBP and Gray Level Co-Occurrence Matrix (GLCM) are some of the items used to extract image texture information. The LBP is one of the methods that can easily produce suitable features for high-precision tissue classification. In the usual LBP method, the histogram is used to extract features. It is a powerful tool for tissue analysis because the local binary model uses both statistical and structural characteristics of the tissue. In the LBP operator, local texture patterns are extracted by comparing the value of the adjacent pixels and the value of the central pixel and are finally represented by binary codes.

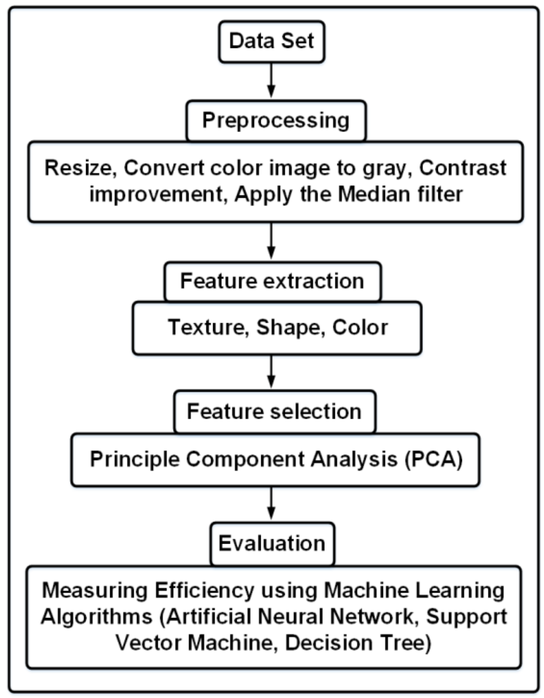

Fig. 2. Proposed method for detection the quality of tomatoes.

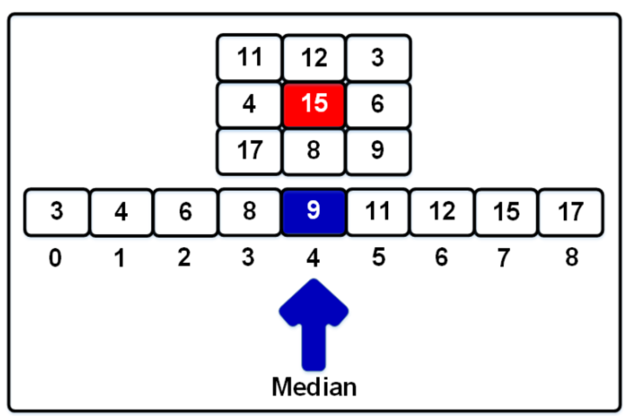

Fig. 3. The example of how to apply the median filter.

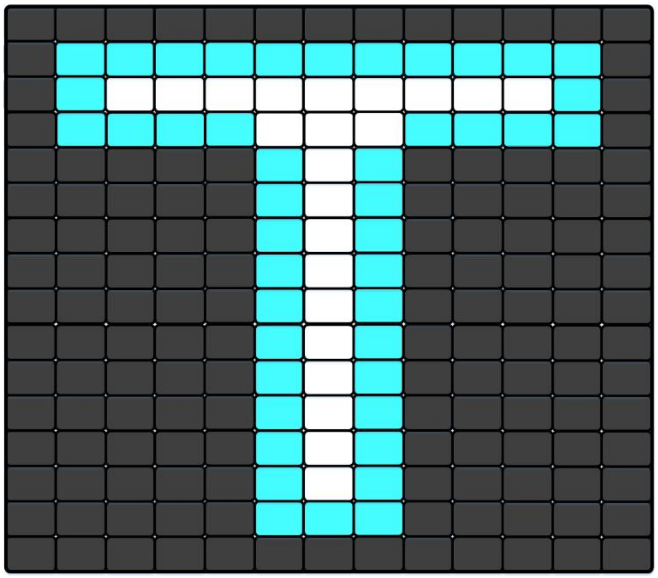

Fig. 4. How to calculate the area in an image.

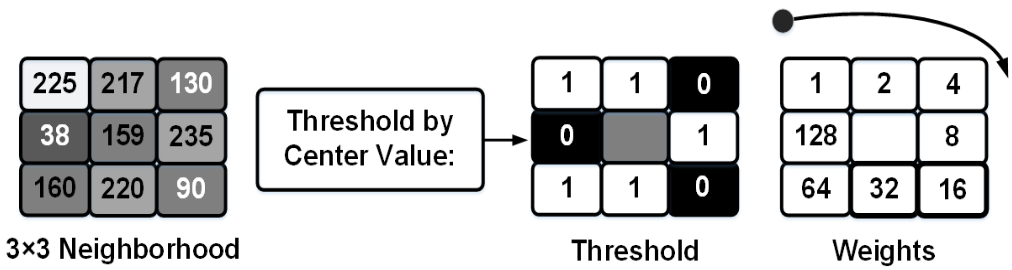

Fig. 5. The LBP basis method.

Local Binary Pattern (LBP) is another development of the local binary patterns. In order to identify the maturity stage of object feature representation, textural information provided by the local binary pattern (LBP) method was utilized [59]. The LBP transforms any local 3 3 picture region into a particular binary pattern. LBP basically contrasts the intensities of a center pixel (ic) and its eight surrounding neighbors (ik; where k = 1, 2,..., 8). If ikic is true, the k-th nearby pixel would receive a value of 1; otherwise, it would receive a value of 0. When an image rotates on the screen, all the neighbors will rotate around the center pixel in one direction. This rotation effect leads to different values for the local binary pattern. To this end, researchers have introduced a new development of the binary local pattern that is independent of rotation (the LBP with constant rotation). The expression independent of rotation here does not refer to spatial changes related to changes in light or different objects. Independent of rotation to eliminate the adverse effect of rotation, a circular bit rotation operator is used to the right to obtain all the binary codes that can be generated with these bits with several repetitions and then the minimum decimal value of the binary pattern (Fig. 5) [48].

In tissue statistical analysis, tissue properties are calculated through statistical distributions of compounds observed for grayscale in a particular position relative to each other. A matrix is also a matrix event, whose number of rows and columns is equal to the number of grayscale levels used in the image. That is, if the number of gray surfaces is up to G, then the dimensions of the matrix of the event in question are equal to one matrix G×G. The matrix P, P(i,j|Δx,Δy) represents the number of iterations of the relationship between two pixels separated by a pixel distance (Δx,Δy), and is defined by the neighborhood relation that one of the pixels has a gray *i* degree and the other has a gray *j* degree, as shown in Fig. 6.

In tissue statistical analysis, tissue properties are calculated through statistical distributions of compounds observed for grayscale in a particular position relative to each other. A matrix is also a matrix event, whose number of rows and columns is equal to the number of grayscale levels used in the image. That is, if the number of gray surfaces is up to G, then the dimensions of the matrix of the event in question are equal to one matrix G×G. The matrix P, P(i,j|Δx,Δy) represents the number of iterations of the relationship between two pixels separated by a pixel distance (Δx,Δy), and is defined by the neighborhood relation that one of the pixels has a gray *i* degree and the other has a gray *j* degree, as shown in Fig. 6.

L, A, B color mode is used to extract the color feature. This color fashion is the most complete color space designated by the International Committee on Lighting and describes all colors visible to the human eye. The three coordinates L, A, B are shown in Fig. 7.

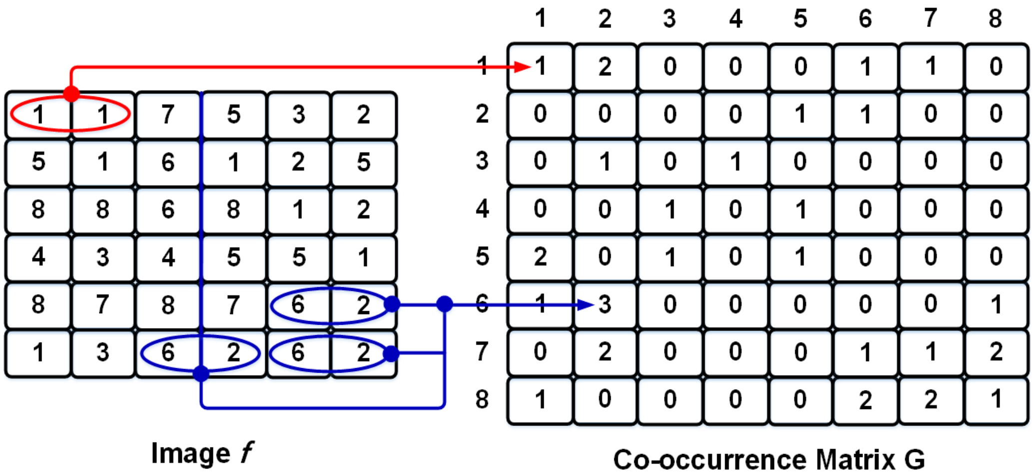

Fig. 6. Gray Level Co-occurrence Matrix (GLCM).

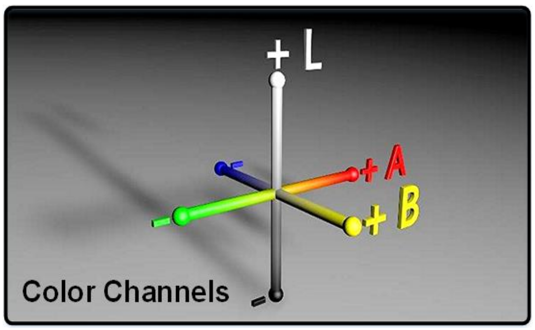

Fig. 7. L,A,B color channels.

The L indicates the intensity of light. The L=0 black and L=100 indicate the scattering of full light. A's position varies between green and red, negative values represent green and positive values represent red. The B Its position varies between blue and yellow, negative values B indicate blue colors and positive values represent yellow colors.

In order to reduce the dependence and noise of bands in images with a large number of bands, the methods of reducing the dimensions of the feature space are used. One of the most well-known methods of reducing the size of the feature space is the principal component analysis method. As the name implies, it can identify key components and help us analyze a series of features that are more valuable instead of examining all the features. In fact, PCA extracts those features that give us more value. After collecting samples and performing preprocessing steps and after extracting and selecting features, machine learning techniques such as support vector machine algorithms, artificial neural networks, decision tree, and k-means clustering are used to evaluate the model.

V. Results and Discussion

What has been presented in the previous sections is a description of the problem of examining the quality of tomatoes using machine learning and image processing. In fact, the background, challenges, and necessities of the study have been analyzed. In this section, by presenting a method based on machine learning along with image processing techniques, the problem of quality recognition of tomatoes is improved, and the proposed method is examined in terms of implementation and evaluation. The proposed steps are then evaluated step by step.

A. Preprocessing and Feature extraction

As mentioned earlier, after applying any type of processing algorithm and pre-processing evaluation, the data set is introduced. For this purpose, preprocessing goals are performed by changing the dimensions of the images to 400 × 300, changing the color channels from three channels to one channel, or in other words, turning the color image to gray, improving the contrast, and applying a median filter. In fact, after these steps, the images in the data set become images like Fig. 8.

According to the data set and based on their differences in different qualitative categories, different types of features are obtained according to Table I.

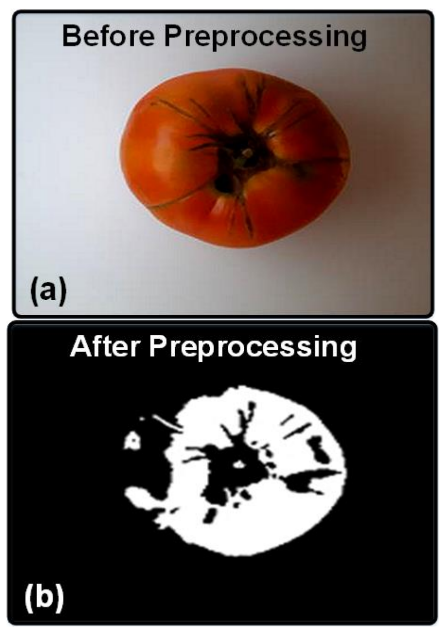

Fig. 8. An example of a tomato image, (a) before preprocessing, (b) after preprocessing.

TABLE I. TYPES OF FEATURE EXTRACTION USED

Feature	Category	Output
Area	Shape	A numerical value
Local Binary Pattern	Texture	A 300 × 400 matrix
Gray Level Co-Occurrence Matrix	Texture	A 255 × 255 matrix
L,A,B color model	Color	A matrix of 300 × 400 × 3

Agricultural products such as tomatoes in the ripe state, as well as in the post-ripening period may be damaged. Based on this, the characteristics of color, shape, and texture cause the quality of tomatoes to be recognizable in different periods, and tomatoes of suitable quality have the best shape, color, and texture. On the other hand, the size of the tomato causes the difference in the first and second quality degrees to be distinguishable. Applying these features together causes the learning model to be able to distinguish each of these quality degrees in the later stages.

B. Feature Selection

As the number of features increases, the dimensions of the feature space also increase. On the other hand, not all of these features are useful, but their presence may even cause the model to be too compatible with the data set, or so-called overfitting. In the PCA method, with the help of feature removal and extraction methods, an attempt is made to separate important features from less important features. And further, reduce the size of the feature by removing minor features. Fig. 9 shows well how the data set is distributed under different PCA components.

PCA has three different components, each of which is well represented in each data class in the data class. A basic component is a normalized linear combination of the main predictions in the data set. Component # 1, component # 2, and component # 3 are the basic components. Component # 1 is a linear combination of the main predictions that contains the largest variance in the data set. This component determines the direction of maximum changes in the data. The higher the range of changes in component 1, the more information there is in this component. Component # 2 and then component # 3 will be important. Fig. 9(a) shows that with the prediction components, the accuracy has reached 89.8% and shows the high accuracy of the diagnosis with only components # 3. Fig. 9(b) is created by the Principal Component Analysis Tool. The principal component coefficients for each variable, the principal component score, are also displayed.

C. Validation

After extracting and selecting the appropriate features, the conditions for learning the model are provided. Accordingly, the evaluation criteria of each of these techniques are expressed as a prediction matrix and the parameters of accuracy, sensitivity, and specificity. It is important to note that in this article we classify two classes for the quality class by merging the first and second class into one class (these two classes differ slightly in size and appearance of products) and the third class in another class, as well as three classes for First, second and third-degree qualitative grading, has been used in three different data classes. The prediction matrix is used to evaluate the performance of the proposed classifier. The predicted label for each sample can have four possible states, including true negative (TN), true positive (TP), false negative (FN), and false positive (FP), as shown in Table II.

The detect components represent the number of samples that are correctly classified, while the main diameter components in the table represent errors. The prediction matrix is the result of evaluating the results of two neural network classes according to Table III. For example, 159 instances of class one has been correctly identified and two instances of data class one has been incorrectly predicted to belong to class two.

Fig. 10 shows a comparative trend of the accuracy of different classifications. Accordingly, the support vector machine has the best accuracy compared to the others, so that even its accuracy in two-class data in the case of merging first- and second-class products is better than three-class classification.

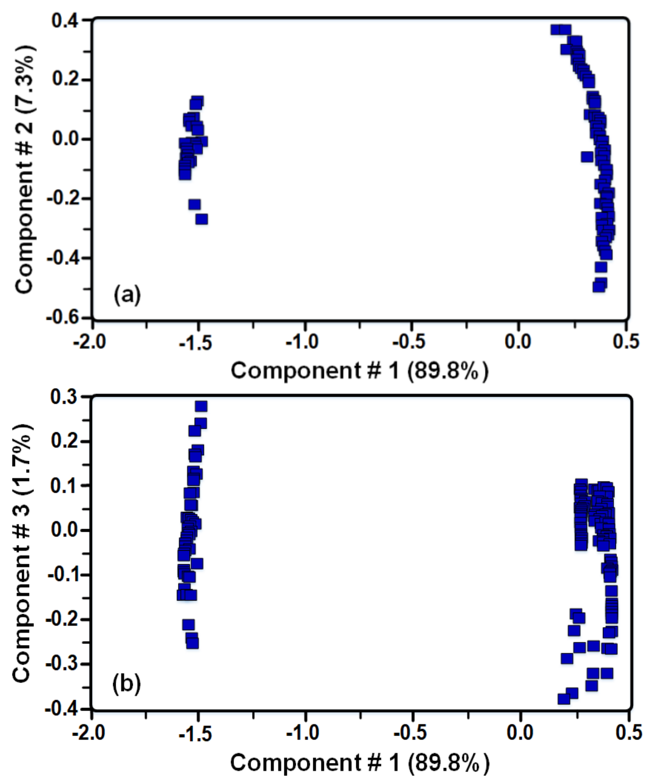

Fig. 9. Accuracy of classification performance of data set (a) component # 1 and component # 2, (b) component # 1 and component # 3.

TABLE II. PREDICTION MATRIX

		Prediction class	
		Class 1	Class 2
Real class	Class 1	TN	FP
	Class 2	FN	TP

TABLE III. NEURAL NETWORK PREDICTION MATRIX

		Prediction class	
		Class 1	Class 2
Real class	Class 1	159	2
	Class 2	1	58

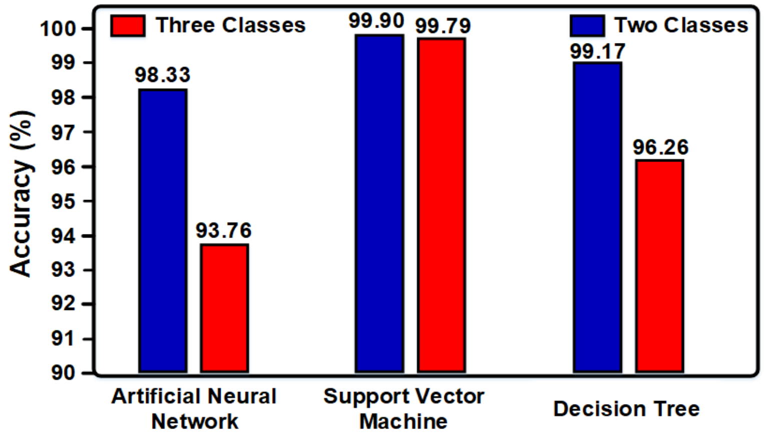

Fig. 10. Performance accuracy of data classification.

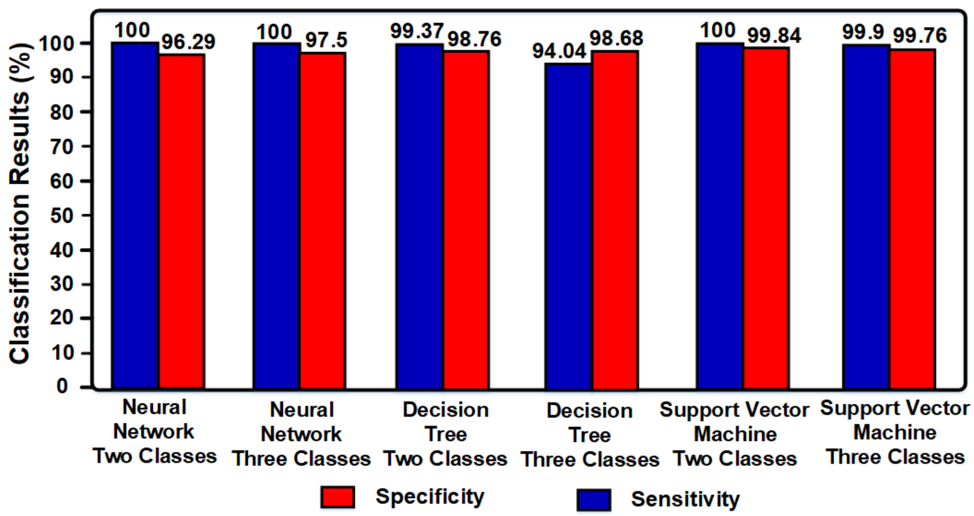

Fig. 11. Comparison of sensitivity and specificity of data classification performance.

According to Fig. 11, after the vector machine, the artificial neural network support performs better in the sensitivity criterion and the decision tree in the specificity criterion. But as a general result, it can be seen that compared to the neural network, the decision tree has the same performance as the support vector machine.

In this section, in the first step, in order to prepare the data set for the next stages of preprocessing operations. In the next step, by extracting the feature with the help of features related to texture, shape, and color and by selecting the features using principal component analysis and support vector machine classifications, artificial neural network, and tree, the decision is applied. The results show that the use of this method has better results compared to other methods and also the integration of the first and second quality grade also improves the accuracy of identification.

VI. Conclusion

The most important sensory feature of fruits and vegetables is the appearance that affects the market value, priority, and consumer choice. Although sorting and grading are done by humans, it is inconsistent, time-consuming, variable, subjective, heavy, expensive, and easily affected by the environment. Hence, a clever fruit grading system is needed. This paper provides an in-depth overview of various methods such as preprocessing, segmentation, feature extraction, and classification that focus on fruit and vegetable quality based on color, texture, size, shape, and defects. Using the appearance characteristics of tomatoes such as color, shape, size, and texture, image processing operations have been performed with MATLAB software to extract the mentioned features. Furthermore, the proposed algorithm can be used for some other types of image data because the proposed LPB feature representation enables to represent the object which can be presented better performance. In the future, a system can be set up that reads images on the rails simultaneously and processes them in video images. Algorithms such as deep learning are introduced that can be more efficient in real-time processing. Thus, the study method can be developed for future work with deep learning algorithms.

Acknowledgment

2021 Guangzhou Science and technology plan project (202102080497).

2021 school level teaching quality and teaching reform project (XJJG2110).

REFERENCES

[1] Pogonyi, A., Pék, Z., Helyes, L., & Lugasi, A. (2005). Effect of grafting on the tomato's yield, quality and main fruit components in spring forcing. Acta Alimentaria, 34(4), 453-462.

[2] Afifah, E. N., Murti, R. H., & Wahyudhi, A. (2021). Evaluation of a promising tomato line (Solanum lycopersicum) derived from mutation breeding. Biodiversitas Journal of Biological Diversity, 22(4).

[3] De Luna, R. G., Dadios, E. P., Bandala, A. A., & Vicerra, R. R. P. (2019). Tomato fruit image dataset for deep transfer learning-based defect detection. In 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM) (pp. 356-361). IEEE.

[4] Sanaeifar, A., ZakiDizaji, H., Jafari, A., & de la Guardia, M. (2017). Early detection of contamination and defect in foodstuffs by electronic nose: A review. TrAC Trends in Analytical Chemistry, 97, 257-271.

[5] da Costa, A. Z., Figueroa, H. E., & Fracarolli, J. A. (2020). Computer vision based detection of external defects on tomatoes using deep learning. Biosystems Engineering, 190, 131-144.

[6] Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. Food Research International, 62, 326-343.

[7] Pfister, S., Vionnet, S., Levova, T., & Humbert, S. (2016). Ecoinvent 3: assessing water use in LCA and facilitating water footprinting. The International Journal of Life Cycle Assessment, 21(9), 1349-1360.

[8] Semary, N. A., Tharwat, A., Elhariri, E., & Hassanien, A. E. (2015). Fruit-based tomato grading system using features fusion and support vector machine. In Intelligent Systems' 2014 (pp. 401-410). Springer, Cham.

[9] Sembiring, A., Budiman, A., & Lestari, Y. D. (2017). Design and control of agricultural robot for tomato plants treatment and harvesting. In Journal of Physics: Conference Series (Vol. 930, No. 1, p. 012019). IOP Publishing.

[10] Oktarina, Y., Dewi, T., Risma, P., & Nawawi, M. (2020). Tomato harvesting arm robot manipulator; a pilot project. In Journal of Physics: Conference Series (Vol. 1500, No. 1, p. 012003). IOP Publishing.

[11] Lili, W., Bo, Z., Jinwei, F., Xiaoan, H., Shu, W., Yashuo, L., ... & Chongfeng, W. (2017). Development of a tomato harvesting robot used in greenhouse. International Journal of Agricultural and Biological Engineering, 10(4), 140-149.

[12] Feng, Q., Zou, W., Fan, P., Zhang, C., & Wang, X. (2018). Design and test of robotic harvesting system for cherry tomato. International Journal of Agricultural and Biological Engineering, 11(1), 96-100.

[13] Syahrir, W. M., Suryanti, A., & Connsynn, C. (2009). Color grading in tomato maturity estimator using image processing technique. In 2009 2nd IEEE International Conference on Computer Science and Information Technology (pp. 276-280). IEEE.

[14] Rupanagudi, S. R., Ranjani, B. S., Nagaraj, P., & Bhat, V. G. (2014). A cost effective tomato maturity grading system using image processing for farmers. In 2014 International Conference on Contemporary Computing and Informatics (IC3I) (pp. 7-12). IEEE.

[15] Arjenaki, O. O., Moghaddam, P. A., & Motlagh, A. M. (2013). Online tomato sorting based on shape, maturity, size, and surface defects using machine vision. Turkish Journal of Agriculture and Forestry, 37(1), 62-68.

[16] Rokunuzzaman, M., & Jayasuriya, H. P. W. (2013). Development of a low cost machine vision system for sorting of tomatoes. Agricultural Engineering International: CIGR Journal, 15(1).

[17] Gould, W. A. (1975). A Preliminary Report on Mass Sorting of Mechanically Harvested Tomatoes.

[18] Duarte, A. C. (1965). The economics of mechanical harvesting of tomatoes in the Salinas Valley, California.

[19] Chen, Y. R., Chao, K., & Kim, M. S. (2002). Machine vision technology for agricultural applications. Computers and electronics in Agriculture, 36(2-3), 173-191.

[20] Mahendran, R., Jayashree, G. C., & Alagusundaram, K. (2012). Application of computer vision technique on sorting and grading of fruits and vegetables. J. Food Process. Technol, 10, 2157-7110.

[21] Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. Food and bioprocess technology, 4(4), 487-504.

[22] Izadi, H., Kamgar, S., & Raoufat, M. H. (2016). Tomato grading system using machine vision technology and neuro-fuzzy networks (ANFIS). Journal of Agricultural Machinery, 6(1).

[23] Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y. D. (2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. Scientia Horticulturae, 263, 109133.

[24] Ismail, N., & Malik, O. A. (2022). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. Information Processing in Agriculture, 9(1), 24-37.

[25] Kukreja, V., & Dhiman, P. (2020). A Deep Neural Network based disease detection scheme for Citrus fruits. In 2020 International conference on smart electronics and communication (ICOSEC) (pp. 97-101). IEEE.

[26] Minten, B., Tamru, S., & Reardon, T. (2021). Post-harvest losses in rural-urban value chains: Evidence from Ethiopia. Food Policy, 98, 101860.

[27] Granatstein, D., & Kupferman, E. (2006). Sustainable horticulture in fruit production. In XXVII International Horticultural Congress-IHC2006: International Symposium on Sustainability through Integrated and Organic 767 (pp. 295-308).

[28] Ireri, D., Belal, E., Okinda, C., Makange, N., & Ji, C. (2019). A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artificial Intelligence in Agriculture, 2, 28-37.

[29] Clement, J., Novas, N., Gazquez, J. A., & Manzano-Agugliaro, F. (2012). High speed intelligent classifier of tomatoes by colour, size and weight. Spanish Journal of Agricultural Research, 10(2), 314-325.

[30] Patel, K. K., Kar, A., Jha, S. N., & Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. Journal of food science and technology, 49(2), 123-141.

[31] Chen, Y. R., Chao, K., & Kim, M. S. (2002). Machine vision technology for agricultural applications. Computers and electronics in Agriculture, 36(2-3), 173-191.

[32] Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. Food and bioprocess technology, 4(4), 487-504.

[33] Patel, K. K., Kar, A., Jha, S. N., & Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. Journal of food science and technology, 49(2), 123-141.

[34] Hintze, H., Lussi, A., Cuisinier, F., & Nyvad, B. (2015). Additional caries detection methods. In Dental caries: the disease and its clinical management. Oxford: Wiley Blackwell.

[35] Nasiri, A., Taheri-Garavand, A., & Zhang, Y. D. (2019). Image-based deep learning automated sorting of date fruit. Postharvest biology and technology, 153, 133-141.

[36] Stein, R. M. (2002). Benchmarking default prediction models: Pitfalls and remedies in model validation. Moody's KMV, New York, 20305.

[37] Ebenuwa, S. H., Sharif, M. S., Alazab, M., & Al-Nemrat, A. (2019). Variance ranking attributes selection techniques for binary classification problem in imbalance data. IEEE Access, 7, 24649-24666.

[38] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the operational research society, 54(6), 627-635.

[39] Topno, P., & Murmu, G. (2019). An improved edge detection method based on median filter. In 2019 Devices for Integrated Circuit (DevIC) (pp. 378-381). IEEE.

[40] Freeman, T. G. (1991). Calculating catchment area with divergent flow based on a regular grid. Computers & geosciences, 17(3), 413-422.

[41] Park, Y., & Guldmann, J. M. (2020). Measuring continuous landscape patterns with Gray-Level Co-Occurrence Matrix (GLCM) indices: An alternative to patch metrics?. Ecological Indicators, 109, 105802.

[42] Basavaiah, J., & Arlene Anthony, A. (2020). Tomato leaf disease classification using multiple feature extraction techniques. Wireless Personal Communications, 115(1), 633-651.

[43] Khan, S., & Narvekar, M. (2020). Novel fusion of color balancing and superpixel based approach for detection of tomato plant diseases in natural complex environment. Journal of King Saud University-Computer and Information Sciences.

[44] Sivakumar, J., Prashanth, J. E. P., Rajesh, N., Reddy, S. M., & Pinjari, O. B. (2020). Principal component analysis approach for comprehensive screening of salt stress-tolerant tomato germplasm at the seedling stage. Journal of biosciences, 45(1), 1-11.

[45] Basavaiah, J., & Arlene Anthony, A. (2020). Tomato leaf disease classification using multiple feature extraction techniques. Wireless Personal Communications, 115(1), 633-651.

[46] Das, D., Singh, M., Mohanty, S. S., & Chakravarty, S. (2020). Leaf disease detection using support vector machine. In 2020 International Conference on Communication and Signal Processing (ICCSP) (pp. 1036-1040). IEEE.

[47] Küçükönder, H., Boyaci, S., & Akyüz, A. (2016). A modeling study with an artificial neural network: developing estimationmodels for the tomato plant leaf area. Turkish Journal of Agriculture and Forestry, 40(2), 203-212.

[48] Tan, L., Lu, J., & Jiang, H. (2021). Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. AgriEngineering, 3(3), 542-558.

[49] Malik, M. H., Zhang, T., Li, H., Zhang, M., Shabbir, S., & Saeed, A. (2018). Mature tomato fruit detection algorithm based on improved HSV and watershed algorithm. IFAC-PapersOnLine, 51(17), 431-436.

[50] Sari, Y. A., Adinugroho, S., Adikara, P. P., & Izzah, A. (2017, November). Multiplication of V and Cb color channel using Otsu thresholding for tomato maturity clustering. In 2017 International Conference on Sustainable Information Engineering and Technology (SIET) (pp. 209-214). IEEE.

[51] Yin, H., Chai, Y., Yang, S. X., & Mittal, G. S. (2009, December). Ripe tomato recognition and localization for a tomato harvesting robotic system. In 2009 International Conference of Soft Computing and Pattern Recognition (pp. 557-562). IEEE.

[52] Indriani, O. R., Kusuma, E. J., Sari, C. A., & Rachmawanto, E. H. (2017, November). Tomatoes classification using K-NN based on GLCM and HSV color space. In 2017 international conference on innovative and creative information technology (ICITech) (pp. 1-6). IEEE.

[53] Huang, L., Yang, S. X., & He, D. (2012). Abscission point extraction for ripe tomato harvesting robots. Intelligent Automation & Soft Computing, 18(6), 751-763.

[54] Goel, N., & Sehgal, P. (2015). Fuzzy classification of pre-harvest tomatoes for ripeness estimation–An approach based on automatic rule learning using decision tree. Applied Soft Computing, 36, 45-56.

[55] Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, (6), 679-698.

[56] Nausheen, N., Seal, A., Khanna, P., & Halder, S. (2018). A FPGA based implementation of Sobel edge detection. Microprocessors and Microsystems, 56, 84-91.

[57] Zhang, F. (2015, March). Ripe Tomato Recognition with Computer Vision. In 2015 International Industrial Informatics and Computer Engineering Conference (pp. 466-469). Atlantis Press.

[58] Benavides, M., Cantón-Garbín, M., Sánchez-Molina, J. A., & Rodríguez, F. (2020). Automatic tomato and peduncle location system based on computer vision for use in robotized harvesting. Applied Sciences, 10(17), 5887.

[59] Zhuang, J., Hou, C., Tang, Y., He, Y., Guo, Q., Miao, A., ... & Luo, S. (2019). Assessment of external properties for identifying banana fruit maturity stages using optical imaging techniques. Sensors, 19(13), 2910.