A Study of Combining KNN and ANN for Classifying Dragon Fruits Automatically

Nguyen Minh Trieu and Nguyen Truong Thinh
Department of Mechatronics, Ho Chi Minh City University of Technology and Education, Thu Duc Citty, Ho Chi Minh City, Vietnam
Email: minhtrieucdt@gmail.com, thinhnt@hcmute.edu.vn

***Abstract*—The export of agricultural products is expanding in developed agricultural countries like Vietnam. The dragon fruit is a fruit that accounts for a large percentage of exports in order to meet the import standards of nations throughout the world, dragon fruit is sorting according to each importing country's standard. The paper describes a dragon fruit classification system automatically that solves the pricing and accuracy difficulties compared with the manual one. Using a combination of models such as KNN to identify dragon fruit, CNN to extract dragon fruit features, and ANN to calculate dragon fruit classification based on the data provided, proposed a model self-training model to improve the accuracy of the system due to insufficient or not various data for training. The dragon fruit database was collected from 1287 dragon fruit supplied from dragon fruit farms, or from export dragon fruit packaging and sorting facilities. Dragon fruit is divided into 3 groups of G1, G2, G3 with different standard in length, width, weight, defects outside the dragon fruit skin. The automatic dragon fruit classification system proposed after the test achieved 98.5% accuracy 6 times more yield than the manual classification.**

***Index Terms*—classifying, sorting, grading, CNN, ANN, KNN, machine learning, AI**

I. INTRODUCTION

In recent years, along with the development of technology, many applications of science and technology have been applied to live in many different fields such as medicine, traffic, education [1]-[3]. Vietnam is an agricultural country so the need for the classification of agricultural products is truly necessary. Dragon fruit or pitaya is a tropical plant, that grown in many different countries around the world due to its high drought tolerance, adaptability to high light and temperature conditions. In recent years, in Vietnam, the area and production of dragon fruit have increased rapidly year by year from 5 512 ha (in 2000) to 55 419 ha (in 2018), dragon fruit is grown throughout the country but concentrated largely in 3 provinces Binh Thuan, Long An, Tien Giang [4], according to the Vietnam Import and Export Report 2020 of the Ministry of Industry and Trade of the Socialist Republic of VietNam for dragon fruit exports 1,363.8 thousand tons. The quality of dragon fruit is graded largely by the external feature of dragon fruit, each country has different standards for classifying dragon fruit. At the present time, dragon fruit export companies use human-powers to classify, which takes a lot of time and costs high, but accuracy does not high. The automatic dragon fruit classification system is really necessary, and it is also a need of the current dragon fruit packing classification facilities. There is a lot of research in the field of agriculture, namely the classification of fruits that achieve high performance published as chili classification using image processing and fuzzy with up to 85% accuracy [5], fruit classification via eight layers [6] proposed model of 8 layers convolution to classify fruits with the accuracy of the model up. To 95.67%. Along with that, some studies use a combination of image processing and artificial intelligence to recognize dragon fruit such as detecting the ripeness of dragon fruit [7] using HSV color space to recognize 3 states of dragon fruit unripe, half, rip with the ripeness of 86.6%, identify the mellowness of dragon fruit [8] using the CNN network, proposed automatic mango classification system using a combination of four models RF, LDA, SVM, KNN to evaluate mango classification [9] achieved accuracy up to 98.1%.

Due to dragon fruits which have different from other fruit in outside shape with green scaly spikes, tails, and red skin shown Fig. 1, dragon fruit recognition is an important classification problem so a self-learning model for the KNN model is proposed to detect the color threshold of dragon fruit in this study. We propose the convolutional neural network to identify the characteristics of dragon fruit and propose ANN for classification system because the standard of some countries doesn't have a clear tie between the standard for features of the dragon fruit such as length, width, defect.

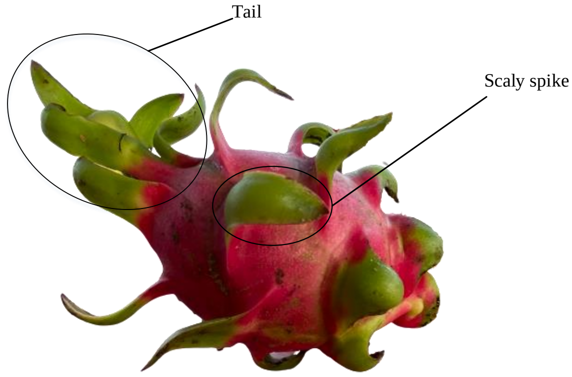

Figure 1. The structure of the dragon fruit.

Manuscript received November 17, 2021; revised February 15, 2022.

doi: 10.18178/joig.10.1.28-35

Therefore, the result of actual surveys from workshops, sorting dragon fruit by the artificial neural network is used and the accuracy of the whole classification system is up to 98.5%, it witnessed by VietNam agricultural experts. In brief, this study has two main contributions. In the first, propose model self-learning for the KNN model to detect the dragon fruit. Secondly, an automatic classification of dragon fruit is built with an accuracy of 98.5% for whole sorting processing after applicate self-learning model.

II. Detect Dragon Fruit Using KNN Self-Training

In this study, the K-Nearest Neighbor (KNN) model is used to detect the color threshold of dragon fruit. The color threshold is one of the important factors affecting the process of identifying dragon fruit. It is necessary to find an appropriate color threshold to segment dragon fruit. It is difficult to determine the common color threshold for different types in different growing areas and conditions. This study proposed a self-learning algorithm for the KNN model to detect dragon fruit in images (include the colors of dragon fruit and background). The self-training model uses the error on the external shape of the dragon fruit to generate new training data and it makes increase the accuracy of the classification system. Although in some cases classification using CNN, it is not necessary to remove the background, the defect criteria (cracks, white spots, …) outside of the dragon fruit skin greatly affects the group of fruit. The colors space of dragon fruit and background are surveyed, there are color values of defects close to the color of the background so we propose the KNN model for removing the background. Then proceed to identify the characteristics of dragon fruit by CNN extract the scaly spikes and the tail, the image processing is used to detect features of dragon fruit such as the length and width, and the proportional area of the defects on the dragon fruit to conduct classification according to the import standard of the countries by the neural network.

In Vietnam, dragon fruit has 2 main types for export are red flesh dragon fruit and white flesh dragon fruit, which have two different types of dragon fruit in terms of composition and flesh color of the fruit, but they all have red outer skin with green dragon fruit scaly spike. There are many different standards to assess the quality of dragon, but they focus on the classification of shape (dragon fruit must be balanced, not too short or too long), weight, volume, area, or rate of the defect on the fruit, the number of broken dragon fruit scaly spikes because it directly affects the quality of the dragon fruit during cleaning and storage. For the field of artificial intelligence, data is one of the important factors determining the success of the model. There are many open-source data sources available, which makes it easier and easier to train the models. However, due to different climate and environmental characteristics in different countries, open data sources on the agricultural sector are still limited. The data used in this study is collected from farms and packaging companies for dragon fruit in Vietnam.

The database is acquired from 1287 dragon fruits from the above sources to generate the data including 6500 dragon fruit images for training the model. The small contribution of this study is the generation of the dragon fruit data. Because several cases will lead to overtraining of model training, there are some data augmentation methods such as image rotation, gamma correction [10], scale, ... which increase the number of databases for the model, helping to increase the accuracy of the training model to avoid over-training cases. The purpose of using the KNN model [11] is to decide whether the color is color of dragon fruit or background color. The color values of dragon fruit are similar to each other, so KNN is a suitable model to make predictions based on votes of neighboring points. The system uses original data, where $X^0 = \{X^0_1, X^0_2, X^0_3, \ldots, X^0_m\}$ is the color set of dragon fruit and background according to the RGB color, and $Y^0 = \{Y^0_1, Y^0_2, Y^0_3 \ldots, Y^0_m\}$ is labeled set of X^0, where m is the number of color values extracted from all shapes. KNN is used in this section to classify X^0 into 2 groups, with "1" being the dragon fruit (S_1) and "0" being the background (S_2). The main concept of KNN is easy to understand. The i^{th} object X^1_j in the unlabelled set X^1 is classified according to its neighborhood point X^0, which assigns the object to the class most frequently among its nearest neighbors. The distance between X^1_j and X^0_j is calculated in equation (1).

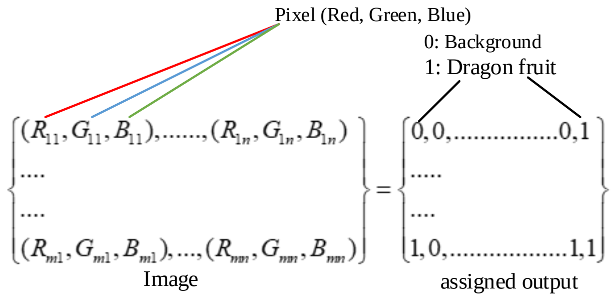

Figure 2. A description of how to obtain data set X^0.

$$d = \sqrt{(R_{X^0_i} - R_{X^1_j})^2 + (G_{X^0_i} - G_{X^1_j})^2 + (B_{X^0_i} - B_{X^1_j})^2} \quad (1)$$

where: X^0_i is i^{th} observation in X^0, $i \in \{1,2,m\}$; X^1_j is j^{th} object in X^1, $j \in \{1,2,m\}$; and R, G, B is the value of red, green, and blue respectively.

The probability that the object X^1_j is class S_i is calculated in equation (2). The class with the highest probability becomes the predicted class.

$$Y^1_j = \frac{1}{K}\sum_{k=1}^{K} I(X^1_j = \underset{i \in \{0,1\}}{S_i}) \quad (2)$$

where: K is the number of neighboring points; I is indicator function (1 is true, 0 is false); and Y^1_j is predicted label set of X^1_j.

The object recognition is a technology used to recognize and monitor objects found in images and videos using computer vision and machine learning. The pixels of the image will be considered according to each color value to identify the background color and the color of the dragon fruit based on the output of the KNN model used to create a binary image to identify the dragon fruit. On each pixel extract 3 RGB color channels (Red, Green, Blue). All color values from acquired images of 1287 dragon fruits were extracted as X^0 dataset these values were manually labeled with "0" as background "1" as dragon fruit (Fig. 2). Then, the KNN model is trained with the labeled data set to continue to combine the self-learning algorithm to label the unlabeled color values to achieve high accuracy for the model.

III. Extract Feature Using CNN

The dragon fruit contrasts with other fruits in that it has an outside shape with scaly spikes and tails so it is difficult to extract the dragon fruit's feature. To help extract the length and width of dragon fruit more accurately, the CNN network was used to recognize the external features of the fruit. To avoid errors in the extraction, this study proposes removing the scaly spikes and tails of dragon fruit. The system processes the images of the dragon fruit through steps like Fig. 3 to extract the characteristics such as length, width, defects. After removing the background and identifying the dragon fruit, the system is also divided into 3 stages. In stage 1, the images are cropped to the corrected size to identify the characteristics using CNN for recognizing the scaly spikes and tail of dragon fruit and interpolating the boundary of the fruit after the dragon fruit background is removed. Stage 2, the image processing methods are used such as thresholding, converting to binary images, finding contours, … to extract the external characteristics of dragon fruit: length, width, area ratio defect rate. Stage 3 classifies dragon fruit using neural network into 3 groups G1, G2, G3 with G1 is the best group.

A. *Stage 1: Using CNN to Extract Features of the Dragon Fruit*

Due to the unusual properties of dragon fruit, we propose to use CNN to identify the scaly spikes and the tail of the dragon fruit. In machine learning, when the CNN is used to identify the feature, data is one element for deciding the accuracy of the network. The structure of CNN contains convolution layer, nonlinear activation function, pooling layer, Fully-connected layer, and dropout. The input is an RBG image (214×214×3) with (L×W×C) with L: Length, W: Width, C: Chanel then convolution with kernels (L_k×W_k×C). In the deep convolutional neural network, to add a bias then pass through the activation function.

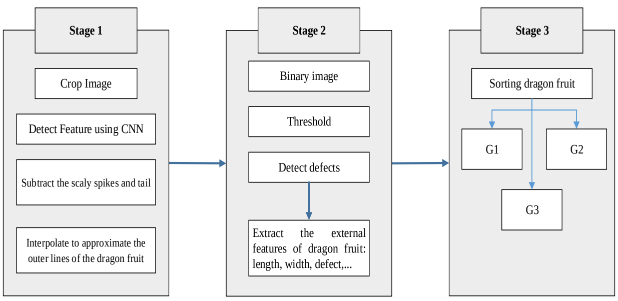

Figure 3. Stage of the processing system.

There are many activation functions, it is common to use is rectified linear unit (ReLU), it is defined via formula (3) or leaky rectified linear unit (LeakyReLU) [12] defined via Equation (4) with a is leak coefficient ($a>0$).

$$f(u) = \mathrm{ReLU}(u) = \begin{cases} u & if\ \ u \geq 0 \\ 0 & if\ \ u < 0 \end{cases} \tag{3}$$

$$f(u) = leaky\mathrm{ReLU}(u) = \begin{cases} u \text{ if } u \geq 0 \\ a{\cdot}u \text{ if } u < 0 \end{cases} \tag{4}$$

The pooling layer has the main task to reduce the complexity for the next layer [13], reduce the number of neurons, reduce overfitting. Max-pooling is one of the popular pooling methods, which returns the maximum value in each scan region defined $P_{max} = Max(x)$ with x as the pooling region.

There are many different studies using CNN to recognize objects in images such as fruit classification [6], the authors propose to use eight-layer convolutional neural network with adjustment of parameters with high resolution. Model accuracy is up to 95.67%. In this study, we propose a CNN network with 8 convolutional layers, 2 fully connected, 4 layers max -pooling, the last layer is softmax layer, which capitalize on softmax function, which is also called multinomial logistic regression to extract feature of the dragon fruit. The input image layer allows receiving RBG images, in the model using only the 3x3 kernel to reduce the number of parameters to increase the computational speed for the system. Adam optimizer [14] is used in this study, and learning rates is set to 0.001, and is decreased by factor of 10 every 100 epochs. After identifying the characteristics of the dragon fruit, such as the scaly spikes and tail, then the system subtracts the scaly spikes to avoid cases that the dragon fruit scaly spikes increase the size of the fruit, causing errors in the classification.

B. *Stage 2 Image Processing – Extract Features*

Image denoising is an important image processing operation because image noise from the camera increases the inaccuracy of extract features. There are many methods to remove the noise of the image introduced in the paper [15] which are proposed to remove the noise in this study. In contour finding, the contour is found using the contour finding algorithm [16], [17]. A contour can be simply defined as a curve that combines all consecutive points (along the contour of the object). With the same color or intensity, the contour of the object is defined $v(s) = (x(s), y(s))$, where $x(s)$ and $y(s)$ are the set of (x) and (y) coordinates of points on the contour line. Find the contour applied to calculate the area of the whole dragon fruit with the area of defects on the dragon fruit.

The image of the dragon fruit obtained from the image processing chamber is cropped to reduce the size of the image to increase the processing speed of the system. According to the actual survey, the cracks of the dragon fruit are well recognized by the system when converting to the binary image and applying the threshold method to determine the cracked part of the dragon fruit and this is the most dangerous defect. Therefore, a deep crack is detected, the system will be eliminated immediately. Color examination of other defects such as burn marks, insect bites, white spots is acceptable in the importing country within the permitted range. To classify by defects, we need to calculate the total pixels of the entire dragon fruit and the pixels of the defective part to customize according to the standard of each importing country.

Let K_C be the adjustment coefficient depending on the camera, P_L, P_R, P_U, P_B are the extreme points of the dragon fruit image with coordinates *(x, y)* in the order of left point, right point, upper point, bottom point. Length (*L*) and width (*W*) are calculated according to the equation (5-6), respectively, at ε_K is the error of K:

$$L = K_C * \sqrt{(x_L - x_R)^2 + (y_L - y_R)^2} + \varepsilon_K \quad (5)$$

$$W = K_C * \sqrt{(x_U - x_B)^2 + (y_U - y_B)^2} + \varepsilon_K \quad (6)$$

In this stage, the area of the defects A_{de} and the total area of the dragon fruit A_{total} are calculated. In dragon fruit identification, all defects are detected and accumulated on the entire surface of the dragon fruit to obtain the final defect level of each dragon fruit given in equation (7), where A_i is number of pixels of the itth defect region. The defects of dragon fruit are superficial damage caused by insects or impacts during its growth, which can be scars, bruises, spots, etc. Dragon fruit is classified according to the ratio of the defective area and the total area of the whole dragon fruit. Depending on the standard, the system classifies according to the area of the defect, or the percentage of the defect. The length and width are extracted to classify the shape of the dragon fruit. The dragon fruit has a balanced shape to meet export requirements, the dragon fruit must not be too long or too short. Besides, thanks to the length, width and height to approximate the volume of the dragon fruit, we can calculate and predict the ripeness of the dragon fruit, which will be analyzed more clearly in the following article. To calculate the weight of the dragon fruit, we use a loadcell sensor arranged on the classification conveyor to calculate the actual weight of the dragon fruit (see Table I).

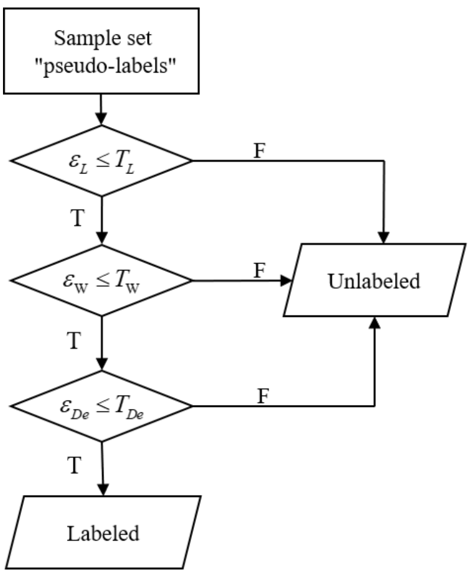

Figure 4. The principles of the self-learning process.

TABLE I. STANDART IMPORT THE DRAGON FRUIT

	Group 1	Group 2	Group 3
Weight (g)	>460	350 - 450	300-350
Defect area (cm^2)	< 2	2 - 4	4 - 5

$$A_{de} = (K_C{}^2 + 2\varepsilon_K K_C)\sum_{j=0}^{t} A_i \quad (7)$$

C. Stage 3 Update Data

During this stage, a sample set is formed containing samples X^1 with dummy labels Y^1 selected through error thresholds for fruit characteristics. It is difficult to learn a model and to optimize approximate labels on data without annotations together. Thus, using self-learning to generate estimated labels also known as "pseudo-labels" from the predictions with the highest confidence, trusting that they are mostly accurate and approximate to truth labels (meet the conditions on the permissible error thresholds). The labels with low accuracy are kept for future prediction. The conditions to check whether used to update the data are expressed through Fig. 4 with T_L is the threshold error of the length feature, T_W is the threshold error of the width feature, T_{De} is the threshold error of the defect feature. Updating data helps the system to improve accuracy with different types of dragon fruit in different seasons of the year. Decreasing T_L, T_H, and T_{De} result in greater confidence of the predicted label. However, because of the production of an analysis that corresponds too exactly to a particular set, the model fails to fit additional mango data or predict future observations reliably. Therefore, the choices of T_L, T_H, and T_{De} are the balance between the stability and precision of the KNN model.

IV. GRADING THE DRAGON FRUITS ACCORDING TO STANDARD

Depending on the import standard of each country, dragon fruit has different classification standard, but in general, it is considered with the standard of weight, shape, and defects. In the largest import market of Vietnam's dragon fruit, China accounted for 73% of Vietnam's total dragon fruit exports (2014), and 80% (The Asia Foundation 2019), and according to a survey at packaging and exporting dragon fruits in Tien Giang province and Long An province, most of the establishments export dragon fruits to the Chinese market, so this study will analyze in-depth the criteria in the import China market, the system can adjust the parameters of the standard to suit different countries.

There are many studies that propose artificial neural networks to classify fruits such as sorting apples [18], sorting oranges [19], sorting the date [20], …. The Multilayer Perceptron (MLP) algorithm or Artificial Neural Network (ANN) [21] is a computational system inspired by biological neural networks that use a network of functions to understand and transform a wide range of input data into the desired output. ANN minimizes errors for nonlinear inputs and can obtain relationships between inputs and outputs without complex mathematical equations. The model of this study has three main components: input layer, hidden layer and output layer shown in Fig. 5. In the input and output layer, the number of nodes corresponds to the number of input and output variables. In this case the first layer has 5 nodes including weight, defect, length, width, number of broken scaly spike.

Let *W, b, p*, and *y* be weight matrix, bias vector, input vector, and output vector of each layer, respectively. The output of the j^{th} node is calculated in equation (8).

$$\bar{y}_j^L = f(W_{ij}^L x_i^{L-1} + b_j^L) \tag{8}$$

where: $\bar{y}_j^L$ is the output of the j^{th} node of L^{th} layer; W_{ij}^L is the weight matrix between the i^{th} output node of the $(L-1)^{th}$ layer and the j^{th} node of the L^{th} layer; x_i^{L-1} is the i^{th} input node of the $(L-1)^{th}$ layer; b_j^L is the bias vector between the i^{th} output node of the $(L-1)^{th}$ layer and the j^{th} node of the L^{th} layer; f is the activation function.

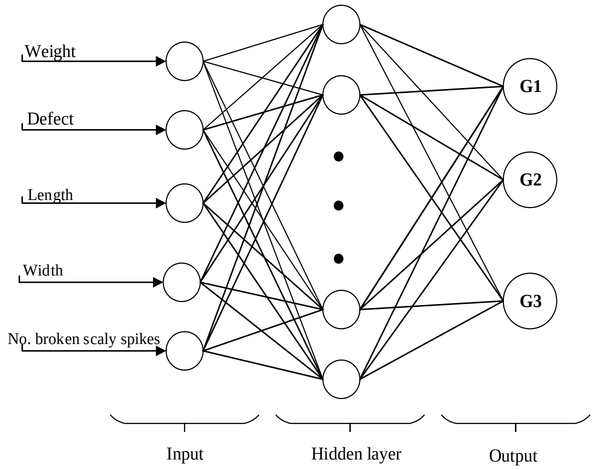

Figure 5. The architecture of the ANN model.

The performance of the training process is identified by finding the difference between the ANN output and the exact output. The loss function (*Ls*) is calculated (equation (9)) based on the mean square error (MSE).

$$Ls = \frac{1}{2s}\sum_{i=1}^{s}(\bar{y}_j^L - y_j^L)^2 \tag{9}$$

where: *s* is the number of outputs, $\bar{y}_j^L$ is the predicted output of the j^{th} neuron of the L^{th} layer; y_j^L is the fitness output of the j^{th} neuron of the L^{th} layer.

Subsequently, gradient descent is applied to update the weights and biases, which minimizes the loss function. The updated weights and biases are computed in equations (10-11), respectively, where α is the learning rate. In each iteration, the learning rate controls the speed of a move to the minima point. Through the iterations, gradient descent converges at the minima, which provides the best weights and biases.

$$W_{ij}^L = W_{ij}^L - \alpha \frac{\partial}{\partial W_{ij}^L} Ls(W_{ij}^L) \tag{10}$$

$$b_j^L = b_j^L - \alpha \frac{\partial}{\partial b_j^L} Ls(b_j^L) \tag{11}$$

The predicted group of the output layer is determined in equation (12):

$$G = \sum_{i=1}^{m}(\mathrm{W}_k^{N_L} x_k^{N_L}) + b_k^{N_L} \tag{12}$$

where: *G* is the predicted group of the dragon fruit; *m* is the number of nodes in the final layer (N^{th} layer) of the hidden layer; N_L the number of layers; $x_k^{N_L}$ is input; $\mathrm{W}_{\mathrm{ij}}^{N_L}$ is the weight matrix; $b_j^{N_L}$ is the biases vector.

V. Results and Experiments

The study was carried out in Viet Nam with the support of agricultural experts and dragon fruit sorting and packing workshops. Data collected from January to October of the year included 1287 dragon fruits. The results of the automatic dragon fruit classification process are satisfying. Combining the signal from the load cell to determine the weight of the dragon fruit, extracting features of the dragon fruit such as length, width, and defects to classify dragon fruit automatically achieves the system's accuracy up to 98.5% for 3 groups G1, G2, G3 with G1 being the best group. The combination of KNN and CNN models are used to identify the features of dragon fruit, because there is not any common color threshold for all types of dragon fruit through the seasons of the year, this study proposes the self-learning model for KNN model to segment dragon fruit, by creating a color threshold to identify dragon fruit and background color. Initially, the pixel values are hand-selected and labeled by agricultural experts in Vietnam so that the samples are labeled correctly. Self-learning solution designed to sample pixel values in images of dragon fruit through error threshold to make increased accuracy of the model KNN. After 2 months of experimentation, the result presented that the accuracy of the automatic dragon fruit classification system is increased. The self-training process will be stopped when the accuracy of the model reaches the accuracy threshold.

The study was conducted based on survey results from workshops in Vietnam. The hardware of the system is designed in accordance with the dragon fruit described in Fig. 6. The actual weight of the dragon fruit is obtained from the loadcell sensor signal. The signal from the load cell is filtered through the Kalman filter. The structure is divided into two parts: the first part is the image processing chamber, the dragon fruit images will be collected and the processing center extracts the features of the dragon fruit such as length, width, fruit defects. Part 2 is a conveyor with loadcell sensor arrangement to get the actual weight signal of the dragon fruit. After the image of the dragon fruit is collected from the image processing chamber, it is combined with the load cell signal from the conveyor, the central processor calculates the group of dragon fruit according to its own standards pre-set according to the standards import standards of countries.

The classification is mainly based on the external feature of the fruit such as length, width, shape, and external defects of dragon fruit. The dragon fruit is put in by the worker from the input of the image processing chamber, the images are obtained through a camera arranged in the image processing chamber with reasonable light to minimize the shadow of the dragon fruit to avoid noise for image processing. The central processor calculated to sorting group of dragon fruit. The weight of the fruit was obtained from the load cell through Kalman noise filtering. The signals about shape and weight are combined to sorting the group of dragon fruit according to each standard of importing countries, implementation diagram as Fig. 7. Using CNN network to identify the features of dragon fruit such as the scaly spikes and the tail with the results of recognizing the features. Using the input image with size 214×214 convolution with 3×3 kernels, not use larger kernels because the system's processing speed is one of the requirements for the automatic dragon fruit classification system. The accuracy of the proposed CNN model achieves an accuracy of 97.38%. All model code is written in Python programming language with the support of the available libraries such as OpenCV, NumPy, pandas, TensorFlow, etc. Use the standard of importing Chinese dragon fruit presented in the sorting section to test the accuracy of the system.

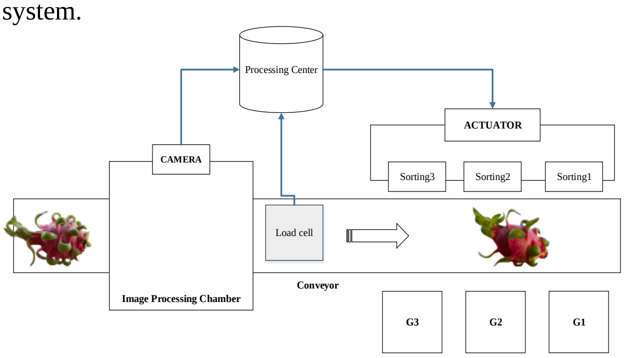

Figure 6. Structure of automatic sorting dragon fruit system.

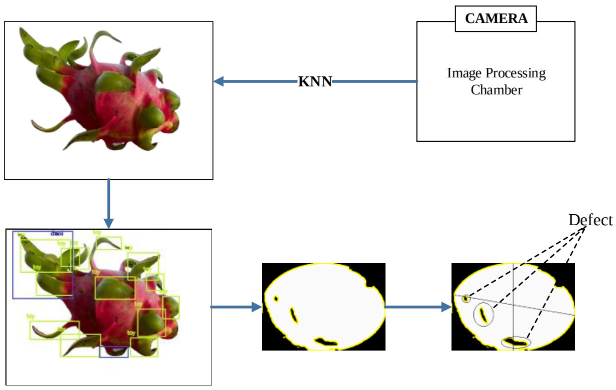

Figure 7. The process of extracting dimensions and defects.

A neural network is proposed to classify dragon fruit according to the features extracted from processing such as length, width, defects, number of broken scaly spikes, which are combined with the weight of the fruit to predict group of dragon fruit. The results of the system are satisfied because the system is built to solve the sorting problems of dragon fruit that can be applied to reality, the test process is performed on 387 dragon fruits. The data provided from the workshops were divided into 2 datasets, data 1 with 200 dragon fruits, and data 2 with 187 dragon fruits to compare before and after the self-training process for the KNN model to identify dragon fruit.

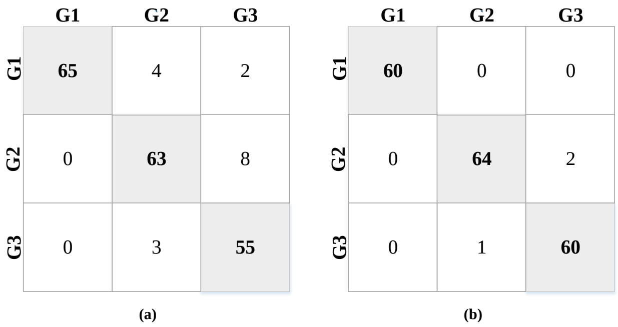

Figure 8. Confusion matrix represents the accuracy of the model, (a) before using the self-training for KNN model, (b) after using the self-training for KNN model.

Experimental results show that the accuracy of the automatic dragon fruit classification system meets the requirements with high accuracy, Fig. 8(a) shows the results of the classification process when not yet used with the self-model for KNN model with system accuracy of 91.7%, the accuracy is lower because there is a confusion of background color and fruit defects causing the group of the dragon to be reduced, the training data for the dragon fruit recognition model is not diverse. After being collected color data from the test dragon fruit and during test operation. Re-testing the accuracy of the system, the results are shown in Fig. 8(b), for group 1 with 100% accuracy, for group 2 with 97% accuracy, for group 3 with 98.4% accuracy. and the average accuracy of the whole system is 98.5%. The system results achieved higher accuracy than expected of dragon fruit packing and grading facilities in Vietnam. The system uses a combination of 3 models KNN, CNN, ANN to identify and extract the characteristics of dragon fruit for classification. The background is removed to avoid confusion with the defects on the fruit, the actual length and width of the dragon fruit are extracted, ANN is used to calculate and predict the group of dragon fruit according to the actual surveyed training data. The self-learning model is proposed, which helps the system improve the accuracy when changing different test data.

TABLE II. COMPARE WITH OTHERS APPROACH

Approach	Overall accuracy
Image processing + ANN	83.1%
CNN	87.7%
Our approach	98.5%

TABLE III. COMPARE WITH DEEP LEARNING METHODS

Models	Overall accuracy
8-layers conv [6]	95.67%
Our – CNN	97.38%

The accuracy of other deep models is compared in Table II, it presents the proposed CNN accuracy as 97.38%. Our approach is to use KNN to segment the dragon fruit, CNN and image processing are proposed to extract features of the dragon fruit. The accuracy of this approach is 98.5%, we compared with others approach and the result is presented in Table III, this approaches the highest.

VI. Conclusion

The article proposes an automatic dragon fruit classification system based on extracting the characteristics of dragon fruit such as length, width, defects on dragon fruit combined with the weight from the load cell. This study is proposed the self-training model for the KNN model to segment dragon fruit from the background, the CNN is used to recognize the characteristics of dragon fruit such as dragon fruit scaly spikes and tail, increase the accuracy of the dragon fruit classification system automatically. Sorting dragon fruit with a neural network, optimize parameters for high accuracy. The results of the study include the following conclusions. First, an automatic dragon fruit classification system with an accuracy of up to 98.5% is proposed to solve difficult problems for exporting dragon fruit classification. The accuracy of CNN is proposed is 97.38% thanks to data collected from dragon fruit packing facilities and farms. Based on a neural network to classify dragon fruit with 3 types G1, G2, G3 from the real database. The system can be applied in reality in the classification facilities after being adjusted to suit each different establishment. Solving problems in dragon fruit classification systems, which reduces the cost increases the accuracy of sorting. Developing the smart dragon fruit classification system with the commercial scale.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Nguyen Truong Thinh and Nguyen Minh Trieu conceptualization and methodology; Nguyen Minh Trieu wrote the manuscript; Nguyen Truong Thinh writing—review and editing; Nguyen Minh Trieu software; Nguyen Truong Thinh supervised; Nguyen Truong Thinh project administration; All authors had approved the final version.

Acknowledgment

The authors would also like to thank the Ho Chi Minh City University of Technology and Education in Vietnam for providing financial support and study facilities.

References

[1] A. Haleem, M. Javaid, and I. H. Khan, "Current status and applications of Artificial Intelligence (AI) in medical field: An overview," *Current Medicine Research and Practice*, vol. 9, no. 6, pp. 231-237, 2019.

[2] H. S. Alenezi and M. H. Faisal, "Utilizing crowdsourcing and machine learning in education: Literature review," *Education and Information Technologies*, vol. 25, no. 4, 2020.

[3] H. Alizadeh, A. Khoshrou, and A. Zuquete, "Traffic classification and verification using unsupervised learning of Gaussian mixture models," in *Proc. IEEE International Workshop on Measurements & Networking*, October 2015, pp. 1-6.

[4] T. L. Le, N. Huynh, and P. Quintela-Alonso, "Dragon fruit: A review of health benefits and nutrients and its sustainable development under climate changes in Vietnam," *Czech Journal of Food Sciences*, vol. 39, no. 2, pp. 71-94, 2021.

[5] E. J. Olaes, E. R. Arboleda, J. L. Dioses, and R. M. Dellosa, "Bell pepper and chili pepper classification: An application of image processing and fuzzy logic," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 4832-4839, 2020.

[6] S. H. Wang and Y. Chen, "Fruit category classification via an eight-layer convolutional neural network with parametric rectified linear unit and dropout technique," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15117-15133, 2020.

[7] I. S. Khisanudin, "Dragon fruit maturity detection based-HSV space color using naive bayes classifier method," in IOP Conference Series: Materials Science and Engineering, IOP Publishing, March 2020, vol. 771, no. 1, p. 012022.

[8] T. Vijayakumar and M. R. Vinothkanna, "Mellowness detection of dragon fruit using deep learning strategy," *Journal of Innovative Image Processing*, vol. 2, pp. 35-43, 2020.

[9] N. T. M. Long and N. T. Thinh, "Using machine learning to grade the mango's quality based on external features captured by vision system," *Applied Sciences*, vol. 10, no. 17, 2020.

[10] Y. Miki, *et al.*, "Classification of teeth in cone-beam CT using deep convolutional neural network," *Computers in Biology and Medicine*, vol. 80, pp. 24-29, 2017.

[11] D. Coomans and D. L. Massart, "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules," *Analytica Chimica Acta*, vol. 136, pp. 15-27, 1982.

[12] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," arXiv preprint arXiv:1505.00853, 2015.

[13] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. International Conference on Engineering and Technology*, August 2017, pp. 1-6.

[14] S. Naskar and T. Bhattacharya, "A fruit recognition technique using multiple features and artificial neural network," *International Journal of Computer Applications*, vol. 116, no. 20, 2015.

[15] R. Verma, and J. Ali, "A comparative study of various types of image noise and efficient noise removal techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 10, 2013.

[16] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP: Image Understanding*, vol. 55, no. 1, pp. 14-26, 1992.

[17] J. Subur, T. A. Sardjono, and R. Mardiyanto, "Braille character recognition using find contour method," in *Proc. International Conference on Electrical Engineering and Informatics*, August 2015, pp. 699-703.

[18] F. Dara and A. Devolli, "Applying Artificial Neural Networks (ANN) techniques to automated visual apple sorting," *Journal of Hygienic Engineering and Design*, vol. 17, pp. 55-63, 2016.

[19] A. S. Simões, A. R. Costa, A. R. Hirakawa, and A. M. Saraiva, "Applying neural networks to automated visual fruit sorting," in *Proc. World Congress of Computers in Agriculture and Natural Resources*, 2002.

[20] K. M. Alrajeh and T. A. Alzohairy, "Date fruits classification using MLP and RBF neural networks," *International Journal of Computer Applications*, vol. 41, no. 10, 2012.

[21] J. J. Hopfield, "Artificial neural networks," *IEEE Circuits and Devices Magazine*, vol. 4, no. 5, pp. 3-10. 1988.

Nguyen Truong Thinh is Dean of Faculty of Mechanical Engineering at Ho Chi Minh City University of Technology and Education, Vietnam. He is also Associate Professor of Mechatronics. He obtained his PhD in 2010 - Mechanical Engineering from Chonnam National University. His work focuses on Robotics and Mechatronic system. Projects include: Service robots, Industrial Robots, Mechatronic system, AI applying to robot and machines, Agriculture smart machines.

Nguyen Minh Trieu is a graduate student at Ho Chi Minh City University of Technology and Education, Viet Nam. His fields are robotic, artificial intelligence, automatic systems. Current, he is a member of a Mechatronics Laboratory in Ho Chi Minh City University of Technology and Education – VietNam.