

A Raspberry Pi-Guided Device Using an Ensemble Convolutional Neural Network for Quantitative Evaluation of Walnut Quality



Turab Selçuk^{*}, Mustafa Nuri Tütüncü

Department of Electrical and Electronics Engineering, Kahramanmaraş Sütçü İmam University, Kahramanmaraş 46000, Turkey

Corresponding Author Email: tselcuk@ksu.edu.tr

<https://doi.org/10.18280/ts.400546>

ABSTRACT

Received: 20 March 2023
Revised: 15 July 2023
Accepted: 5 September 2023
Available online: 30 October 2023

Keywords:

walnut quality, deep neural networks, YOLO V5

In this study, a device, augmented by artificial intelligence and controlled by Raspberry Pi, has been engineered for estimating the yield of walnut trees and assessing walnut quality. The device, equipped with a camera, identifies walnuts in real-time using the YOLO V5 detection system. For each detected image of a walnut, feature extraction, selection, and classification were conducted employing a Support Vector Machine (SVM). This methodology facilitated the development of a system capable of determining and recording the quality of all walnuts within a tree or orchard. By leveraging deep neural networks for the analysis of 1800 walnut samples, the device demonstrated an accuracy of 98% in ascertaining walnut quality. This innovative device holds the capacity to swiftly analyze a considerable quantity of walnuts, thereby providing a numerical representation of the quality classes of walnuts cultivated by growers. This quantitative evaluation of walnut quality could subsequently streamline agricultural activities such as irrigation and fertilization, enabling a more efficient and informed approach to these processes. The findings presented in this study underscore the potential of integrating artificial intelligence with practical devices for enhancing the productivity and quality control in agriculture.

1. INTRODUCTION

The quality of food products, and hence their market value, is significantly influenced by the processes of cultivation, harvesting, storage, and packaging. Therefore, it becomes imperative to monitor all stages of food processing, from production to consumption. Agricultural engineers play a pivotal role in overseeing these processes [1]. Product quality assessment can generally be divided into two approaches, namely manual control, which involves direct observation by relevant personnel, and automated control systems facilitated by computer assistance. Image and video processing are widely utilized in determining fruit quantity, size, and visual structures, thereby enabling quality evaluation [2]. However, traditional image and video processing systems often struggle with fruit tracking, size determination, and counting, due to relatively low accuracy criteria, thereby leaving a significant gap in the field.

The commercial value of the walnut, a key food commodity, is heavily contingent on its quality. Walnuts of high quality are characterized by a uniform, bright brown shell, ideal dimensions, and the absence of any stains or cracks. In accordance with the Chilean Walnut Commission's technical standards, walnuts are categorized into three classes: Extra, Category I, and Category II, based on size and shell color. These properties not only influence fruit quality and thus market prices but also serve as a reflection of the yield of the walnut tree and its orchard. Factors such as timely and adequate garden maintenance, proper drying conditions for the fruits, and suitable storage conditions directly affect yield and quality. Consequently, a higher yield and fruit quality,

resulting from planned and regular maintenance, positively impact the market and marketing power of a walnut orchard.

This paper presents the design of a device capable of real-time detection of the quality classes of walnuts in a tree. The device is controlled by Raspberry Pi, with real-time object detection performed using YOLO V5. Convolutional Neural Network (CNN)-based feature extraction is conducted from the images of the detected walnuts. The most decisive features are selected and classified using a Support Vector Machine (SVM).

1.1 Related works

In the past decade, various object detection methods have seen successful implementation in numerous agricultural applications [3-6]. They have classified *Fortunella margarita* into three growth stages: mature, immature, and overgrown. The dataset they generated can serve as a valuable resource for researchers utilizing different machine learning or deep learning algorithms for object detection, image segmentation, and multi-class classification [7].

Another study employed Mask-RCNN and YOLOv3, real-time object detection models, to detect peppers using thermal and RGB color images. A mean Average Precision (mAP) value of 1.0 was achieved with the YOLOv3 architecture. The study demonstrated that the YOLOv3 algorithm has superior capabilities compared to Mask-RCNN, with a ten-fold increase in computational speed in the chili dataset [8].

A novel fruit counting method was introduced that integrates deep segmentation, frame-to-frame tracking, and 3B localization for precise fruit counting. This model is capable

of working with image streams from a monocular camera under controlled lighting, in both natural light and at night. The Full Convolutional Network (FCN) based method traces the fruits along the squares using the Hungarian Algorithm. To correct the estimated count from the monitoring process, the Structure-from-Motion (SFM) algorithm is utilized to disregard outliers and double-counted fruit [9].

There are also studies that assess their performance in object detection under varying conditions. For instance, an efficient method based on YOLO V5 has been proposed to detect apple fruits, achieving a recall, precision, F1 score, and false detection rate of 87.6%, 95.8%, 91.5%, and 4.2% respectively, with an average detection time of 8 ms per image [10].

A Faster R-CNN based method has been developed to accurately detect and position apples. This method achieved an Average Precision (AP) of 88.12% with an average detection rate of 0.32 s per image. The mean standard deviation and mean localization precision of apples are 0.51 cm and 99.64%, respectively. They proposed an enhanced binocular localization method for apple detection using deep learning. An improved YOLOv3 model was proposed for real-time detection of apples in orchards, assessing the growth stages of apples and predicting yield. With DenseNet-based YOLOv3, the average detection time per 3000×3000 resolution image was calculated as 0.304 [11].

In another study, a deep learning-based vision system was developed that combines two neural networks and color thresholds for real-time detection, tracking, and localization of strawberries [12]. They proposed an advanced active obstacle separation method that employs a push and drag-push process to separate obstacles from the target in three stages.

The YOLO detection models were applied to automatically detect reliable and accurate fruit yield and fruit count in an orange orchard with hundreds of trees [13]. Model performance and accuracy of yield estimation were investigated for a garden with 1115 trees, whose RGB images were trained in the Google Colaboratory environment to detect and count ripe fruit. Precision, recall, F1 score, and mAP values were 91.23%, 92.8%, 92%, and 90.8%, respectively.

Zhang et al. [14] proposed a YOLO-based method for simultaneous automatic identification, counting, and measurement of maize leaf stomata. They used the entropy rate superpixel algorithm for accurate measurement of stoma parameters. Based on the characteristics of the stoma image dataset, the architecture of YOLO V5 was modified, resulting in a corn leaf stoma recognition accuracy of 95.3% with an average accuracy of 90%.

Finally, another study proposed a two-layer image processing system based on machine learning to grade bananas. The Support Vector Machine (SVM) is the first layer, classifying bananas based on a feature vector of color and texture attributes. Next, the YOLO V3 model was used to further identify the defect area of the shell and determine whether the inputs belonged to the medium matured or well matured class. According to the experimental results, the accuracy for the first layer is 98.5%, the accuracy for the second layer is 85.7%, and the overall accuracy is 96.4% [15].

1.2 Contributions of our study

The contributions of the proposed system are summarized as follows:

- A novel, more effective artificial intelligence-based approach to determining the quality of walnuts

compared to existing methods.

- Enabling comprehensive quality assessment of all walnut trees in a given orchard.
- Establishing the quality of the tree and providing concrete insights regarding the annual maintenance of the orchard (such as irrigation, fertilization, etc.).
- Giving preliminary information about the harvest to the producers and presenting a report on whether the product is grown correctly and on the determination of its commercial value.

2. MATERIALS AND METHODS

2.1 Dataset

In the study, two datasets, dataset 1 and dataset 2, were created. The dataset will be used to detect 1 object. There are two folders, 50 original images and tag images of those images. Images are 1250×1250 in size. In data set 2, there are a total of 450 walnut images belonging to 3 different quality groups (150 Extra I, 150 Category I, 150 Category II). These walnuts were divided into three quality classes by the expert food engineer according to the standards specified in the Chilean walnut commission report. Images are 100×100 in size. The data set is divided into two groups as training and testing. The training images are 375 (125 Extra I, 125 Category I, 125 Category II), and the test images are 75 (25 Extra I, 25 Category I, 25 Category II). By applying symmetry to the x-axis, symmetry to the y-axis and 90° rotation, one of the data augmentation techniques, the number of images was increased to 4 times. In total, 1500 of 1800 images are training images and 300 are test images. Figure 1 shows sample images in dataset 1 and dataset 2.

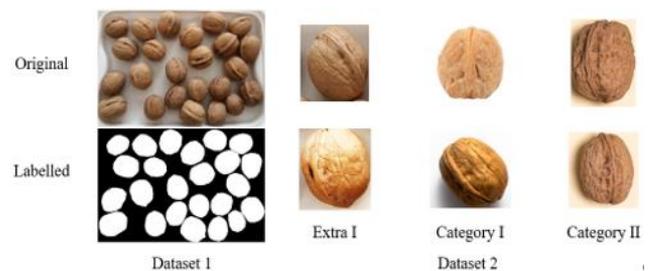


Figure 1. Sample images from the dataset

2.2 Methods

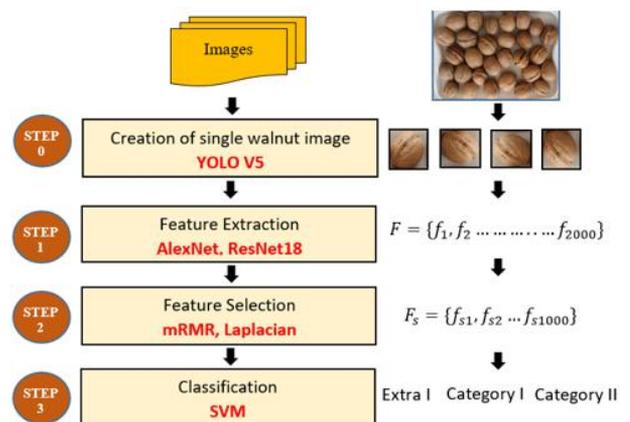


Figure 2. Flow chart of the methods used

In this study, a device was designed to determine the quality of each walnut in an image containing a large number of walnuts. The operation of the device will be explained in the experimental studies section. In this section, the methods used to determine walnut quality will be explained. These methods consist of 4 steps. These steps are detection, feature extraction, feature selection and classification of walnuts. These stages are given in the flow chart in Figure 2.

2.3 Step 0 creation of single walnut image

In this step, the walnuts on the platform of the designed device were detected. For this process, the Yolo V5 [16] network trained with the original and labelled images in Dataset1 was used. In this way, walnuts were detected from images containing a large number of walnuts. This detection process is for obtaining single walnut images.

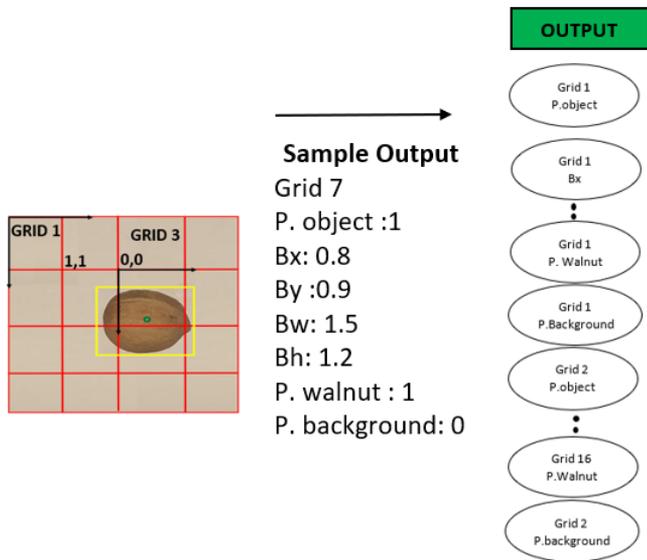


Figure 3. YOLO splitting the image into grids

Since the system will work in real time, speed is an important factor besides high performance. For this, the YOLO V5 network model was used. The feature that makes this model stand out from the models used in object detection such as U-Net [17], Mask R-CNN [18] is that it is very fast. In

the feature extraction of the model, there is DarkNet-53 in the model itself.

The aim here is to keep the performance high, especially in the detection of walnut edge pixels. Figure 3 shows the detection of each walnut. As seen in Figure 3, the image is passed through the neural network in one go in the YOLO model. It predicts the class and coordinates of all objects in the image. At the core of this estimation process is that it treats object detection as a single regression problem. To do this, we first divide the input image into $S \times S$ grids. These grids are 3×3 , 5×5 , 19×19 etc. it could be. Certain properties are obtained for each grid. Information such as whether the object is in the relevant grid, whether it is in the middle point, if it is in the middle point, its length, height, and which class it is can be obtained. For example, since the midpoint of the walnut coincides with the 7th grid in Figure 3, that grid was held responsible for detecting the walnut/drawing boxes around it. Accordingly, YOLO creates a separate prediction vector for each grid. In each of this vector:

Confidence score: This score shows how confident the model is whether there are objects in the current grid. (If 0 is definitely not, if 1 is definitely present) If it thinks it is an object, it shows how sure it is whether this object is really that object and the coordinates of the box around it.

B_x : x coordinate of the object's midpoint

B_y : y-coordinate of the object's midpoint

B_w : width of the object

B_h : Height of the object

Associated Class Probability: As many predictive values as there are different classes in the model.

2.4 Step 1 feature extraction

In this step, feature extraction will be done for the images obtained in step 0. For this, feature vectors obtained from the fully connected layer of AlexNet and ResNet 18 architectures are used. The fc8 layer of the AlexNet architecture is used. This layer is the last layer before the classifier. Therefore, it is more decisive than other layers (fc6, fc7). In this step, when the images are given as input to two different architectures, the feature vectors obtained from each mesh are combined to obtain the final feature vector. In total, 2000 features were obtained from each image. These processes are summarized in Figure 4.

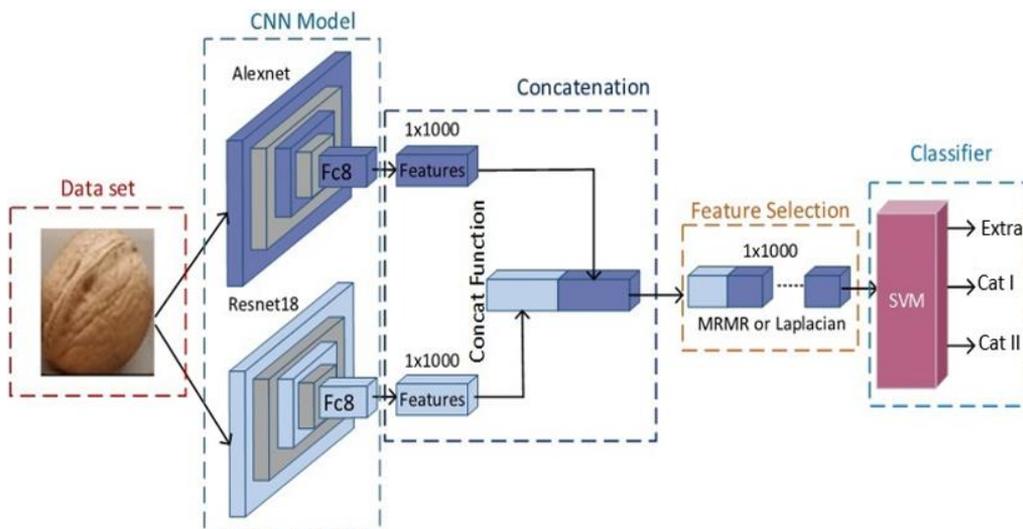


Figure 4. Flow chart of the proposed system

2.5 Step 2 feature selection

In this step, feature selection was made according to the rank values of the feature vector obtained in step 1. The feature vector, which is 1×2000 in total, has been reduced to 1×1000 dimensions. In this study, Minimum Redundancy Maximum Relevance Mrmr [19] and Laplacian [20] selection method was used.

2.5.1 Minimum redundancy maximum relevance (mRMR) algorithm

The mRMR algorithm is an effective method in finding answers with a high degree of variation. The algorithm reduces the feature set to the minimum length, leaving the features with the highest correlation with the response variable. With S being the feature set to be selected, $|S|$ Let be the number of elements in this set. In order for the set to be selected from the S vector to have the highest features, two conditions must be met. The first of these (minimum redundancy and the other is maximum relevance. Minimum redundancy is shown in Eq. (1) and maximum relevance in Eq. (2)).

$$\min W = \frac{1}{|S|^2} \sum_{F_i, F_j \in S} I(F_i, F_j) \quad (1)$$

$$\max V = \frac{1}{|S|} \sum_{F_i \in S} I(F_i, H) \quad (2)$$

$$MID = (V - W) = (I(F_i, H) - \frac{1}{|S|} \sum_{F_i \in S} I(F_i, F_j)) \quad (3)$$

$$MIQ = (V/W) = (I(F_i, H) / \frac{1}{|S|} \sum_{F_i \in S} I(F_i, F_j)) \quad (4)$$

Fulfilment of the conditions stated in Eq. (1) and Eq. (2) is done as shown below by common knowledge difference MID (Eq. (3)) and common knowledge ratio (MIQ) (Eq. (4)).

2.5.2 Laplacian score

The Laplacian Score algorithm is based on detecting the closest ones among the feature values belonging to the same class. Let (D_{ij}) be the distance between any x_i and x_j points calculated with the nearest neighbor. First of all, a similarity matrix is created by using Eq. (5) from these distance values.

$$S_{ij} = \exp\left(-\frac{D_{ij}^2}{t}\right) \quad (5)$$

where, t represents the scale factor of the kernel function. The Laplacian score is determined by the mean \tilde{f}_r (Eq. (6)) and S_{ij} of each feature.

$$\tilde{f}_r = f_r - \frac{f_r^T D 1}{1^T D 1} \quad (6)$$

where, $f_r = [f_{r1}, f_{r2}, \dots, f_{rm}]^T$ $D = \text{diag}(S1)$ denotes matrix degree 1^T , and $[1, 1, \dots, 1]^T$. In this case, the Laplacian score is calculated as in Eq. (7).

$$L_r = 1 - \frac{\tilde{f}_r^T S \tilde{f}_r}{f_r^T D_g \tilde{f}_r} \quad (7)$$

2.6 Step 3 classification

The SVM classification algorithm was trained using the mRMR method specified in step 2 and the first 1000 features among the most defining features. In this algorithm, each data item is plotted as a point in the n -dimensional space (where n is the number of features you have) with the value of each feature being the value of a particular coordinate. Next, the classification is performed by finding the hyperplane that distinguishes the two classes quite well. The SVM algorithm has high performance in classification problems.

2.7 Step 4 calculating performance parameters

The confusion matrix is used and performance parameters are calculated to determine the success of the proposed model in detecting quality class of walnuts. The pseudo-code of the proposed method is as follows (Algorithm 1):

Algorithm 1. Pseudo code of proposed method

```

Input: Walnut Patch Image (I)
Output: Quality of Walnuts
// Feature extraction of images from AlexNet and ResNet
18
F1 = AlexNet(I), // I: Input Image
F2 = ResNet18(I),
// Concatenate features
F=Concat (F1, F2)
// Determine the score or rank value of the
obtained 1_2000 dimensional feature vectors
Fs = mRMR(F)
// Determine the feature that maximizes the classification
accuracy (FSVM)
for i:1 to 1000
for j:1 to i
FSVM = Fs (1, j)
SVM (FSVM, class) // Classify the new feature
vector using the SVM classification algorithm.
end
end

```

3. EXPERIMENTAL RESULTS



Figure 5. The image of the designed device

In this study, a device that performs the operations described in the methods section is designed. This device

determines the quality classes of walnuts in real time thanks to the camera on its own. The walnuts in the device hopper fall onto a platform. A camera that sees the platform from the top first detects each walnut separately. Then, the quality class of these walnuts is determined and saved in an excel file. The image of the device is shown in Figure 5.

The device consists of a platform with a white floor, a camera that sees the platform from above, and a hopper with walnuts. All necessary code is created in python environment. Raspberry Pi single board computer is used. The flow diagram of the operation of the device is shown in Figure 6.

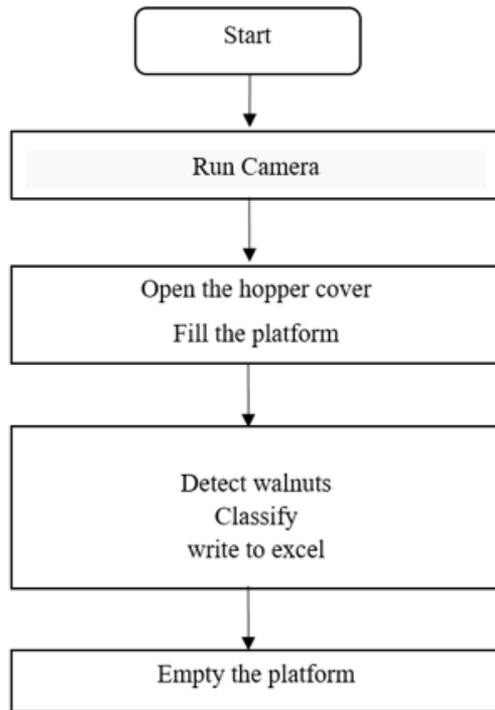


Figure 6. Flow chart of the operation of the device

Since a multi-class classification was made in this study, the above scales were evaluated separately for three classes (Extra, Cat 1, Cat2). Confusion matrix was used in the performance evaluation of the proposed method. 10 scales were used. These are accuracy (A), sensitivity (SN), specificity (SP), precision (P), F1-score (F1), misclassification rate (M), negative predictive value (NPV), false positive rate (FPR), false discovery rate (FDR), false negative rate (FNR), and diagnostic odds ratio (DOR).

$$A = \frac{TP + TN}{TP + TN + FN + FP} \quad (8)$$

$$SN = \frac{TP}{TP + FN} \quad (9)$$

$$SP = \frac{TN}{TN + FP} \quad (10)$$

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$F1 = \frac{2XSNXP}{SN + P} \quad (12)$$

$$M = \frac{FP + FN}{TP + TN + FN + FP} \quad (13)$$

$$NPV = \frac{TN}{TN + FN} \quad (14)$$

$$FPR = \frac{FP}{FP + FN} \quad (15)$$

$$FDR = \frac{FP}{FP + TN} \quad (16)$$

$$L_r = 1 - \frac{\tilde{f}_r^T S \tilde{f}_r}{f_r^T D_g \tilde{f}_r} \quad (17)$$

Since a multi-class classification was made, performance scales for each quality class were obtained. These values are shown in Table 1. E stands for extra quality class, CI stands for category I and CII stands for category two. 1000 features obtained with AlexNet and ResNet were classified separately by SVM and the results are shown in Table 1.

When Table 1 is examined, it is seen that the performance increases when the most determinant 1000 of the features obtained with both CNN models are selected and classified in SVM. It is seen that the highest performance is obtained in the detection of the Extra class (around 98%). It is seen that mRMR is more successful than Laplacian in the selection of determinants. 98.6% accuracy, 98.0 sensitivity, 98.9 specificity values were obtained when the mRMR values belonging to the Extra class were examined. Since Category II is the middle class, the performance of this class is as low as 1% compared to the other two classes. Figure 7 shows the confusion matrix for the three quality classes.

Table 1. Performance values

| | AlexNet | | | ResNet18 | | | Proposed Method | | | | | |
|---------|---------|------|------|----------|------|------|-----------------|------|------|-----------|------|------|
| | | | | | | | mRMR | | | Laplacian | | |
| | E | CI | CII | E | CI | CII | E | CI | CII | E | CI | CII |
| A (%) | 96.2 | 94.7 | 96.1 | 95.3 | 93.4 | 95.2 | 98.6 | 97.8 | 98.4 | 98.0 | 96.5 | 98.0 |
| SN (%) | 96.7 | 94.1 | 96.4 | 95.7 | 93.7 | 95.5 | 98.0 | 97.2 | 97.8 | 98.6 | 95.9 | 98.3 |
| SP (%) | 96.8 | 94.7 | 96.7 | 95.8 | 93.8 | 95.7 | 98.9 | 98.1 | 98.7 | 98.7 | 96.5 | 98.6 |
| P (%) | 96.7 | 94.6 | 96.4 | 95.8 | 93.7 | 95.5 | 98.0 | 96.9 | 97.8 | 98.7 | 96.4 | 98.4 |
| F1 (%) | 94.7 | 92.6 | 94.5 | 93.2 | 91.4 | 93.1 | 96.9 | 96.2 | 96.8 | 96.5 | 94.4 | 96.3 |
| M (%) | 2.2 | 2.4 | 2.3 | 2.1 | 2.0 | 2.1 | 1.3 | 1.3 | 1.3 | 2.2 | 2.0 | 2.2 |
| NPV (%) | 96.8 | 94.6 | 96.7 | 95.9 | 93.7 | 95.8 | 98.9 | 98.1 | 98.7 | 98.7 | 96.4 | 98.2 |
| FPR (%) | 1.9 | 2.2 | 2.1 | 2 | 2.4 | 2.2 | 1.0 | 1.0 | 1.0 | 1.9 | 2.0 | 2.0 |
| FDR (%) | 2.1 | 2.4 | 2.2 | 2.2 | 2.5 | 2.4 | 1.9 | 2 | 1.9 | 1.9 | 2.1 | 2.0 |
| FNR (%) | 2.1 | 2.3 | 2.2 | 2.2 | 2.4 | 2.3 | 1.8 | 1.9 | 1.8 | 1.9 | 2.1 | 1.9 |

| | | Predicted Values | | |
|-------------|-------------|------------------|------------|-------------|
| | | Extra I | Category I | Category II |
| True Values | Extra I | 99 | 1 | 1 |
| | Category I | 1 | 97 | 2 |
| | Category II | 1 | 1 | 98 |

Figure 7. Confusion Matrix of three quality classes

The highest performance was obtained with mRMR. Accuracy values of Eksta, category I and category II were obtained as 98.6%, 97.8%, and 98.4%, respectively. The next high performance was achieved with Laplacian. The accuracy values obtained in this method were obtained as 98%, 96.5% and 98%, respectively. It has been seen that the proposed method, that is, combining the features obtained from two different CNN models and making feature selection, increases the performance.

4. DISCUSSION

In this study, quality analysis of walnuts in shell was made with the help of camera. This analysis was performed with a Raspberry Pi based device. Some of walnuts in the category I are very similar in size and appearance to category II or extra I class walnuts. For this reason, the performance in the detection of Category I is lower than the other two classes. The device does not separate high quality walnuts and low walnuts into different sections. There are manual systems that do this. However, these systems make classification according to the size of the walnut. A conveyor belt can be formed by combining two different devices. Thus, while the yield analysis of a garden is being done, the workload in the separation of quality walnuts will be reduced, and the walnut quality will be reported numerically.

5. CONCLUSION

In this study, a device was designed to determine the yield of a walnut tree or walnut orchard. The device determines and reports the quality class of each walnut in real time on a platform with CNN models. A two-stage method was applied. In the first stage, a YOLO-based model was used to detect walnuts on the platform, and in the second stage, a combined CNN model was used.

Features obtained from two different CNN (AlexNet, ResNet18) architectures were combined and the most decisive one was selected. The results showed that the proposed model has higher performance than traditional CNN models. Thanks to this developed device, the quality classes of walnuts are expressed numerically. In this way, the productivity of walnut trees in a garden will be determined individually. In this way, agricultural activities such as fertilization and irrigation of walnut orchards will be carried out more accurately.

ACKNOWLEDGMENTS

This study is related to “A Device Design that Automatically Determines Walnut Quality” project. And it was supported by Kahramanmaraş Sutcu Imam University

Scientific Research Unit (Grant No.: 2022/4-9 YLS). We would like to thank Prof. Dr. Mehmet Sütüyemez for her technical support regarding walnut quality.

REFERENCES

- [1] Fomunyan, K.G. (2019) The role of agricultural engineering in ensuring food security in Nigeria. *International Journal of Mechanical Engineering and Technology*, 10(11): 22-27.
- [2] Meenu, M., Kurade, C., Neelapu, B.C., Kalra, S., Ramaswamy, H.S., Yu, Y. (2021). A concise review on food quality assessment using digital image processing. *Trends in Food Science & Technology*, 118: 106-124. <https://doi.org/10.1016/j.tifs.2021.09.014>
- [3] Capizzi, G., Sciuto, G.L., Napoli, C., Tramontana, E., Woźniak, M. (2015). Automatic classification of fruit defects based on co-occurrence matrix and neural networks. In 2015 Federated Conference on Computer Science and Information Systems (FedCSIS), Lodz, Poland, pp. 861-867. <https://doi.org/10.15439/2015F258>
- [4] Chetan, R., Ashoka, D.V., Ajay Prakash, B.V. (2022). IMLAPC: Interfused machine learning approach for prediction of crops. *Revue d'Intelligence Artificielle*, 36(1): 169-174. <https://doi.org/10.18280/ria.360120>
- [5] Nada, H., Omer, O., Esmail, H., Ashour, M., Arafa, A. (2023). Deep learning networks for non-destructive detection of food irradiation. *Revue d'Intelligence Artificielle*, 37(3): 551-556. <https://doi.org/10.18280/ria.370303>
- [6] Wang, S., Zhang, Y., Ji, G., Yang, J., Wu, J., Wei, L. (2015). Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization. *Entropy*, 17(8): 5711-5728. <https://doi.org/10.3390/e17085711>
- [7] Huang, M.L., Wu, Y.S. (2021). A dataset of fortunella margarita images for object detection of deep learning based methods. *Data in Brief*, 38: 107293. <https://doi.org/10.1016/j.dib.2021.107293>
- [8] Hespeler, S. C., Nemati, H., Dehghan-Niri, E. (2021). Non-destructive thermal imaging for object detection via advanced deep learning for robotic inspection and harvesting of chili peppers. *Artificial Intelligence in Agriculture*, 5: 102-117. <https://doi.org/10.1016/j.aiaa.2021.05.003>
- [9] Liu, X., Chen, S. W., Aditya, S., Sivakumar, N., Dcunha, S., Qu, C., Taylor, C.J., Das, J., Kumar, V. (2018). Robust fruit counting: Combining deep learning, tracking, and structure from motion. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, pp. 1045-1052. <https://doi.org/10.1109/IROS.2018.8594239>
- [10] Wang, D., He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, 210: 271-281. <https://doi.org/10.1016/j.biosystemseng.2021.08.015>
- [11] Tian, Y., Yang, G., Wang, Z., Li, E., Liang, Z. (2019). Detection of apple lesions in orchards based on deep learning methods of cyclegan and yolov3-dense. *Journal of Sensors*, 2019: 7630926. <https://doi.org/10.1155/2019/7630926>

- [12] Xiong, Y., Ge, Y., From, P. J. (2021). An improved obstacle separation method using deep learning for object detection and tracking in a hybrid visual control loop for fruit picking in clusters. *Computers and Electronics in Agriculture*, 191: 106508. <https://doi.org/10.1016/j.compag.2021.106508>
- [13] Mirhaji, H., Soleymani, M., Asakereh, A., Mehdizadeh, S.A. (2021). Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. *Computers and Electronics in Agriculture*, 191: 106533. <https://doi.org/10.1016/j.compag.2021.106533>
- [14] Zhang, F., Ren, F., Li, J., Zhang, X. (2022). Automatic stomata recognition and measurement based on improved YOLO deep learning model and entropy rate superpixel algorithm. *Ecological Informatics*, 68: 101521. <https://doi.org/10.1016/j.ecoinf.2021.101521>
- [15] Zhu, L., Spachos, P. (2021). Support vector machine and YOLO for a mobile food grading system. *Internet of Things*, 13: 100359. <https://doi.org/10.1016/j.iot.2021.100359>
- [16] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [17] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*, Munich, Germany, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [18] He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. In 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 2980-2988. <https://doi.org/10.1109/ICCV.2017.322>
- [19] Radovic, M., Ghalwash, M., Filipovic, N., Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1): 9. <https://doi.org/10.1186/s12859-016-1423-9>
- [20] He, X., Cai, D., Niyogi, P. (2005). Laplacian score for feature selection. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, Columbia, Canada, pp. 507-514.