IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 22, 2020, accepted November 10, 2020, date of publication November 16, 2020,
date of current version November 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037948

Deep Learning and Computer Vision for Estimating Date Fruits Type, Maturity Level, and Weight

MOHAMMED FAISAL[1], (Member, IEEE), FAHAD ALBOGAMY[2], HEBAH ELGIBREEN[3], MOHAMMED ALGABRI[4], AND FATTOH ABDU ALQERSHI[4]

[1]College of Applied Computer Sciences, King Saud University, Riyadh 145111, Saudi Arabia
[2]Turabah University College, Taif University, Taif 26571, Saudi Arabia
[3]Information Technology Department, College of Computer and Information Science, King Saud University, Riyadh 145111, Saudi Arabia
[4]College of Computer and information Science, King Saud University, Riyadh 145111, Saudi Arabia

Corresponding author: Mohammed Faisal (mfaisal@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research (DSR) with King Saud University, Riyadh, Saudi Arabia, through a research group program under Grant RG-1441-503.

ABSTRACT According to the Food and Agriculture Organization, the world production of date fruits is 8,526,218 tons and around 1,302,859 tons in the Kingdom of Saudi Arabia (KSA) in 2018. There are several types of date fruits, and the most common in KSA are Barhi, Khalas, Meneifi, Naboot Saif, and Sullaj. Moreover, there are around five main maturity levels: Immature, Khalal, Khalal with Rutab, Pre-Tamar, and Tamar. Harvesting date fruits is performed according to its maturity level and type, which is a critical decision that significantly affects profit. In this paper, we propose a smart harvesting decision system to estimate date fruits type, maturity level, and weight using computer vision (CV) and deep learning (DL) techniques. The proposed system consists of three sub-systems: Dates maturity estimation system (DMES), type estimation system (DTES), and dates weight estimation system (DWES). We utilized four DL architectures, including ResNet, VGG-19, Inception-V3, and NASNet for both DMES and DTES and support vector machine (SVM) (regression and linear) for DWES. We evaluated the performance of the proposed system using the dataset collected by the Center of Smart Robotics Research. Using multiple performance metrics, DTES achieved maximum performance of 99.175% accuracy, an F1 score of 99.225%, 99.8% average precision, and 99.05% average recall. The maximum performance of DMES was 99.058% accuracy, F1 score of 99.34%, 99.64% average precision, and 99.08% average Recall. DWES achieved a maximum performance of 84.27% using SVM-Linear.

INDEX TERMS Date fruit type classification, date fruit maturity classification, deep learning, date fruit classification, neural networks, computer vision.

I. INTRODUCTION AND BACKGROUND

Dates are fruits of the date palm trees, which are famous for its high nutritional value and is a widespread summer fruit in the Arab world. Historically, Arabs relied on dates in their daily lives. Dates are oval in shape with size ranging between 20 to 60 mm in length and 8 to 30 mm in diameter. Furthermore, dates are the major fruit of the Kingdom of Saudi Arabia (KSA). According to the Saudi Arabian Ministry of Agriculture, date palm trees cover about 72% of the total cultivated area in KSA. Based on the reports of the Food and Agriculture Organization (FAO), the global production of dates increased from 3,430,883 tons in 1990 to 8,526,218 tons in 2018 [1]. Particularly, in KSA, dates production grew from around 527,881 tons in 1990 to 1,302,859 tons in 2018 [1]. Furthermore, the average production share of dates between 1994 and 2018 is 4,154,484 tons in Asia, 2,579,209 tons in Africa, 28,109 tons in the Americas, and 11,034.52 tons in Europe [1] (FIGURE 1).

According to FAO, the top 10 producers of dates in 2018 are: Egypt with 1,562,171 tons, Saudi Arabia with 1,302,859 tons, Iran with 1,204,158 tons, Algeria

The associate editor coordinating the review of this manuscript and approving it for publication was Hiu Yung Wong.

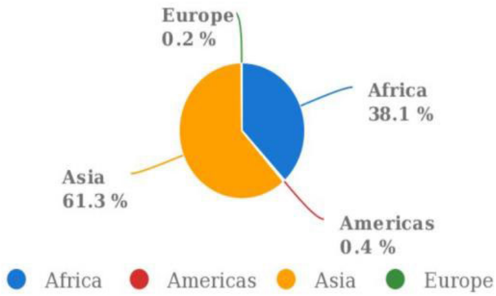


FIGURE 1. The average production share of dates between 1994 and 2018 [1].

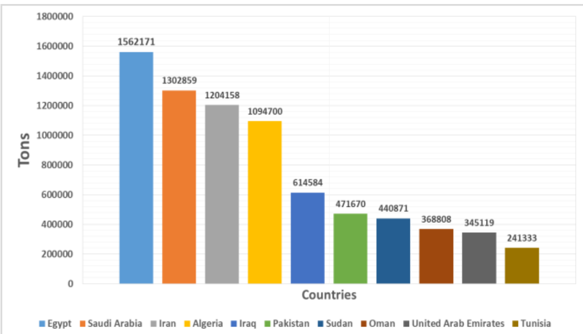


FIGURE 2. Top 10 producers [1].

with 1,094,700 tons, Iraq with 614,584 tons, Pakistan with 471,670 tons, Sudan with 440,871 tons, Oman is 368,808 tons, the United Arab Emirates with 345,119 tons, and Tunisia with 241,333 tons, as shown in FIGURE 2. Harvesting date fruits at an appropriate time based on its type and specific maturity level significantly improve profit. There are three stages involved in harvesting date fruits from the palm trees, and each stage consists of several tasks, as shown in FIGURE 3.

A. PRE-HARVESTING TASKS

1) DETHRONING

Dethroning is the process of removing long thorns, as shown in FIGURE 4, in order to facilitate the climbing of the palm tree during future operations.

2) THINNING THE PALM DATE TREE

This process consists of three distinct operations: First, some bunches are usually removed to reduce the number of bunches and increase quality. Second, from each bunch, middle strands are removed to increase airflow and exposure to sunlight. Third, each strand is thinned (60% of the dates are removed) to enable the dates to grow bigger and become tastier, as shown in FIGURE 5.

3) BUNCH ALIGNMENT

Date palm tree workers prepare the bunches after pollination by applying some force to stretch them and seclude them from the trunk, in order to prevent the growth of the fruits in a nest and to improve exposure to sunlight and airflow, as shown in FIGURE 6.

4) BUNCH ATTACHING

In this task FIGURE 7, some fruit bunches are attached to non-fruitful bunches to prevent the concentration of the bunches in one place or their dispatching from the trunk. It allows the reinforcement of the next date bunches and positions them in a similar radius around the trunk. Therefore, making the date harvest easier and faster.

5) DUST REMOVAL

This operation is necessary to remove dust from the date bunches in order to improve the penetration of required sun-light and air to the fruits. A typical sand storm and cleaning process are shown in FIGURE 8.

6) DATE SPIDER REMOVAL

Similar to dust removal, a water compressor is used to remove small insects (date spider) from the strands, as shown in FIGURE 9. This process uses medium pressure water and soap (in small quantities) to remove spiders and their webs.

7) BAGGING

Bagging is the process of protecting the date palm bunches from insects and birds, as shown in FIGURE 10. This process uses a bagging system instead of the cutting device in the robotic arm.

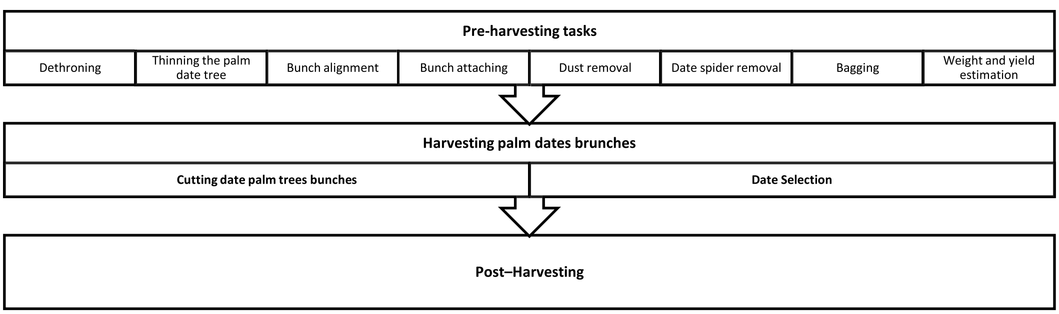


FIGURE 3. Stages of harvesting.

FIGURE 4. Thorns of the Palm date tree.

(a) Before (b) After

FIGURE 5. Strand thinning.

FIGURE 6. Dispatching of the bunches for future easy access and good exposure to sunlight and airflow.

(a): Bunches free

(b): bunches reinforced

FIGURE 7. Bunch attaching.

8) WEIGHT AND YIELD ESTIMATION

Usually, the weight of the date fruit bunches is estimated visually, as farmers rely on experienced experts to estimate the weight of fruits per palm from simple visual screening. We aim to apply CV and DL techniques to estimate the weight of bunches for improved in-depth study and effective cost analysis.

FIGURE 8. Dust covering the date palm trees and cleaning process.

FIGURE 9. Spider web covering the date palm trees.

FIGURE 10. Bagging bunches against insects and birds.

B. HARVESTING PALM DATES BUNCHES

There are different types of harvesting methods, including selecting date fruits one by one, shaking the bunch and most of the dates fall off, or cutting the bunch at a specific time. In this work, we focus on bunch cutting.

1) CUTTING DATE PALM TREES BUNCHES

In the current manual harvesting, a manual worker climbs the palm tree to cut bunches using a saw. We aim to automate these steps, including bunch gripping, cutting, and collection into a dedicated basket in an effective manner. At the end of the harvesting process, bunches are weighed, and pictures are taken for future reference and processing.

2) DATE SELECTION

This operation deals with some types of dates, as the harvest does not collect all the dates at the same time, but collect only matured or half matured dates, while others are left for future harvest.

C. POST-HARVESTING

Post-harvesting includes several operations that come after the dates have been removed. During this step, the palm tree has no more dates. The brown dead leaves are cut using a circular saw at an exact angle, avoiding sharp cuttings to prevent wounding manual workers while climbing. In KSA, the tradition is to avoid cutting many leaves (six on average per tree) to avoid the elongation of the palm trees and keep them as short as possible in order to ease the next harvest for the manual workers.

Some of the cleaning operations can be automated, such as brown leaves cutting and trunk cleaning. These operations require less effort and precision, unlike the harvesting process; FIGURE 11 shows a trunk that requires cleaning.

FIGURE 11. **Date Palm trunk requiring cleaning.**

In order to overcome the inability of estimating the date fruit type, maturity level, and weight per palm tree in the pre-harvesting stage, we proposed a computer vision (CV) and deep learning (DL) based smart system to predict date types (Barhi, Khalas, Meneifi, Naboot Saif, and Sullaj), date maturity levels (Immature, Khalal, Khalal with Rutab, Pre-Tamar, and Tamar), and weight of date per palm tree in the pre-harvesting stage.

The rest of this paper is organized as follows: the literature review is presented in Section II. The dataset is explained in Section III. In Section IV, the proposed system is presented. In Section V, we explain the experimental results and evaluation. Section VI compares the proposed system with other systems, and the conclusion is presented in Section VII.

II. LITERATURE REVIEW

Several research works have studied the estimation of the type, maturity, and weight of various fruits using traditional image-processing technologies. In 2013, Prabha *et al.* developed a system to estimate the maturity of banana fruit using traditional image-processing techniques by extracting the size and color from their images. [2]. They categorized the maturity of banana into three types over-mature, mature, and under-mature. Another study [3] used a color-grading method to estimate the maturity and the quality of date fruits using 2-D histograms of colors in the grading category to identify the co-occurrence frequency. In 2015, another study [4] introduced a method to determine the maturity of sweet lime using image processing and the RGB color. However, most of these methods used thresholds, such as shape, color, and size. In 2014, YAMAMOTO, Kyosuke *et al.* [5] utilized machine learning (ML) to identify tomato organic product development stages without a threshold value. They proposed a system with three stages: an "X" means clustering, pixel-based, and blob-based segmentation.

Multiple studies utilized advanced mechanics innovation such as robotics and machine vision in agricultural applications, referred to as harvesting robots. As a result of the shortage of skilled labor, the application of agriculture harvesting robots has the potential to increase productivity, reduce waste, and improve agriculture sustainability [6]. Recently in 2020, Lin, Guichao, *et al.* introduced a harvesting robots for automatic fruit detection in natural environments, yield estimation, and quality monitoring systems [7]. Collecting robots can be utilized for organic product picking [6] and identifying fruit-bearing branches [8]. Recently, several studies introduced methods to classify date fruits using CV and ML techniques. In 2020, Faisal *et al.* [9] introduced the IHDS system, which consists of six DL systems to estimate seven maturity stages. The IHDS system used date fruit bunches in orchard datasets [10] and archived accuracy of 99.4%, F1 score of 99.4%, 99.7% recall, and 99.7% precision.

Altaheri *et al.* [11] used the ML and vision system to categorize date fruits. The system proposed in [11] was used to classify date fruits' type and maturity using the VGG-16 and Alexnet architectures. It achieved an accuracy of 97.25% for five-stage maturity and 99.01% for date type classification. Studies in [9] and [11] used the same dataset [10]. In 2019, Nasiri *et al.* [12] utilized CV and ML to detect three development stages (Tamar, Rutab, and Khalal). The dataset was gathered using single dates with a uniform background using a cell phone. Their proposed method used VGG-16 architecture and achieved an overall accuracy of 96.98%. Various studies introduced methods to characterize natural products other than dates. In 2020, Behera *et al.* [13] presented two ML-based techniques to classify papaya development stages. They utilized a small dataset with 300 papaya organic product pictures, comprising 100 pictures of each development stage. They utilized seven pre-prepared models: GoogleNet, VGG-16, ResNet18, ResNet50, AlexNet, VGG-19, and ResNet101. Multiple studies used ML-based techniques to classify the maturity of Milano and Chonto tomatoes [14] and Philippine coconut [15]. The study in [15] used the support vector machine (SVM) and random forest techniques to classify the Philippine coconut into three maturity stages (over-mature, mature, and pre-mature).

Regarding fruit weight estimation, several approaches for fruit weight estimation have been introduced using image processing techniques. Some of the studies direct estimate the fruit weight using traditional image processing techniques. Eoh, C., and AR Mohd Syaifudin [6] used the linear regression to determine the linear relation between the measured area and the actual weight of the mangoes. Other studies [7] and [8] used 3-D images gathered by multiple cameras to estimate the weight of mango fruit. A more mathematical approach was introduced by Dang, Nhan T., *et al.* In 2016, where the volume is estimated using the 3D bounding box of the mangoes and the Monte Carlo. Recently CNN have been used for food volume estimation. In 2017, Liang *et al.* [16] introduced Faster-R CNN based regression to estimate the food volume with an average error of 20%. in 2018, Li *et al.* [17] developed a CNN system for volume estimation using two phases. First, the fruit is recognized by a pre-trained detection net, then the fruit image passes into a ResNet based regression relation between food image and volume estimation, with an average error of less than 15% for the various viewpoints. 2020, Kalantar *et al.* [18] used RetinaNet deep convolutional neural network to estimate the number of melons and the weight of each melon using colored images acquired by a digital camera mounted on an unmanned aerial vehicle. This system included three phases: melon detection, geometric feature extraction, and individual melon yield estimation.

III. DATASET

In this study, we use a dataset that was built by the Center of Smart Robotics Research (**www.CS2R.ksu.edu.sa**) called ''DATE FRUIT DATASET FOR AUTOMATED HARVESTING AND VISUAL YIELD ESTIMATION'' [10]. This dataset was developed to be used in pre-harvesting stages. It consists of two different datasets. Dataset-1 has about 8079 pictures taken using two Canon cameras EOS-1100D and EOS-600D, with resolutions of 4272 × 2848 and 5184 × 3456, respectively. Dataset-1 covered all maturity levels. The palms and bunches of dataset-1 belong to five date types: Meneifi, Khalas, Sullaj, Barhi Naboot, and Saif. FIGURE 12 shows the five maturity levels that the proposed system will estimate, and FIGURE 13 shows the date types.

FIGURE 12. Simples of maturity levels.

FIGURE 13. Simples of date types (Meneifi, Khalas, Sullaj, Barhi, and Naboot Saif).

Dataset-2 was developed for weight estimation and consisted of 152 date bunches of 13 palm trees. Bunches were weighed after harvesting, and their images were captured with a white background. Multiple data were extracted from the palm trees and date bunches during measurements, including palm type, code, height, trunk circumference, number of bunches, harvest date, and the state of the recording, as shown in Table 1. This dataset is available online at https://ieee-dataport.org/open-access/date-fruit-dataset-automated-harvesting-and-visual-yield-estimation.

IV. PROPOSED SYSTEM

The proposed system consists of three sub-systems: dates type estimation system (DTES), dates maturity estimation system (DMES), and dates weight estimation system (DWES). DTES is used to estimate five different date types (Barhi, Khalas, Meneifi, Naboot Saif, and Sullaj). Moreover, DMES is used to estimate five different maturity levels of date fruit (Immature, Khalal, Khalal with Rutab, Pre-Tamar, and Tamar). DWES is used to estimate the weight of date fruit. As shown in FIGURE 14, the proposed system takes the video stream from video sources (unmanned aerial vehicles or any other source) as input, then images are extracted from the video. Subsequently, it performs an image manipulation on the images. Then, the manipulated images are fed into the sub-systems. In the proposed system, we used four different DL architectures: ResNet, VGG-19, Inception-V3, and NASNet for both DMES and DTES, and we used SVM (regression and linear) for DWES. DTES and DMES (FIGURE 14) used an end-to-end DL to estimate date types and maturity level from images, without feature extraction. As shown in FIGURE 14, the proposed DTES and DMES started by gathering dataset images (thousands of date fruits images bunches in the orchard environment). We augmented the images and manipulated them by resizing them according to the standard size of their respective CNN models. Then, we divided the dataset into training and testing datasets. To estimate date type and maturity levels, we applied the trained CNN models (ResNet, VGG-19, Inception-V3, and NASNet). For DWES, as shown in FIGURE 14, we used SVM-regression (SVR) to estimate the weight of date fruit. DWES used the images generated by the system, which was extracted from the video stream. It selects the area of interest, calculates the values of the image bytes, and normalizes the calculated value. Then, DWES applies the SVR to estimate the weight of the date fruits.

Harvesting date fruits is performed based on its maturity level, type, and weight, which is a critical decision that significantly affects profit. In this paper, we proposed a smart harvesting decision system to estimate date fruits type, maturity level, and weight.

A. SELECTED CNN ARCHITECTURE

We used four CNN architectures ResNet [16], VGG-19 [17], Inception-v3 [18], and NASNet [19] in the proposed system

TABLE 1. Extracted information of one palm in database-2.

Palm type:	Barhi	
Palm code	B3.K.BW	
Palm height	2.85 m	
Palm trunk circumference	195 cm	
No. of bunches in the palm	15	
Harvesting date	18/8/2016	

No.	Bunch code	Bunch weight (kg)	Recording state of the bunch			
			Image on palm	Image on graph paper	Video	**Marked on palm**
1	B3.K.BW.19	33.64	yes	yes	yes	yes
2	B3.K.BW.20	12.06	yes	yes	yes	yes
3	B3.K.BW.21	21.90	yes	yes	yes	yes
4	B3.K.BW.22	36.94	yes	yes	yes	yes
5	B3.K.BW.23	20.70	yes	yes	yes	yes
6	B3.K.BW.24	18.30	yes	yes	yes	yes
7	B3.K.BW.25	37.32	yes	yes	yes	yes
8	B3.K.BW.26	19.08	yes	yes	yes	yes
9	B3.K.BW.27	12.00	yes	yes	yes	yes
10	B3.K.BW.28	33.00	yes	yes	yes	yes
11	B3.K.BW.29	36.76	yes	yes	yes	yes
12	B3.K.BW.30	23.48	yes	yes	yes	yes
13	B3.K.BW.31	42.16	yes	yes	yes	yes
14	B3.K.BW.32	29.60	yes	yes	yes	yes
15	B3.K.BW.33	19.12	yes	yes	yes	yes
Palm Yield (kg)		396.06				
Average (kg)		26.40				
Max (kg)		42.16				
Min (kg)		12.00				

for DTES and DMES, and used the SVM (regression and linear) for DWES. ResNet is a CNN model, which is based on the idea of skip blocks of convolutional layers by using shortcut connections (FIGURE 15). Its basic blocks called ''bottleneck'' blocks follow two design rules: the number of filters is doubled if the output feature map size is halved or uses the same number for layers and filters if they have the same output feature map size.

In ResNets, the down-sampling is performed directly by convolutional layers with a stride of 2, and batch normalization is performed after each convolution before ReLU activation. The identity shortcut is used in ResNets if the input and output have the same dimensions; otherwise, the projection shortcut is used to match dimensions. VGG-19 [17] model is developed with minimum pre-processing to identify graphic patterns from pixel images. The ImageNet project has been configured for use in applications for visual object detection research work. VGG network is characterized by its simplicity, using only 3×3 convolutional layers stacked in increasing depth. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes, are then followed by a softmax classifier. FIGURE 15 illustrates an overview of VGG-19 and ResNet.

Neural Architecture Search Network (NASNet) is a Google DL model presented in 2017, as shown in FIGURE 16. It achieves good results with lower complexity (FLOPs) and smaller network architecture size. NASNet was introduced by Google primarily for image classification applications.

In 2015, GoogleNet announced a CNN architecture called Inception-v1. Inception-v2 was subsequently released with the addition of batch normalization [20]. In late 2015, Szegedy *et al.* improved the Inception-v2 by adding factorization, which was released as Inception-v3. [18]. The Inception model's core functionality is to discover the optimal local construction of the convolutional network and recurrence [18]. Generally, the Inception works based on numerous redundant connections between layers with excessive information due to their correlation. Therefore, the Inception model used only 22 layers in a parallel organization, as shown in FIGURE 17. It utilizes several auxiliary classifiers inside the intermediate layers to improve the discrimination capacity in the lower layers [21].

For ResNet and VGG-19 architectures, we froze all layers from 1 to 15. Then, we added five more layers (Global average pooling, Dropout (0.3), Dense (128), Dense (64), and Softmax (5 lasses)) before the last layer. TheResNet architecture has total 26,748,805 parameters, 26,691,589 trainable parameters, and 57,216 non-trainable parameters, and VGG-19 architecture has total 20,098,629 parameters, 2,434,053 trainable parameters, and 17,664,576 non-trainable parameters. For Inception-v3 and NASNet, we added five more layers (Global average pooling, Dense (1,024), Batch normalization, Dense (1,024), and Softmax (5 classes)) before

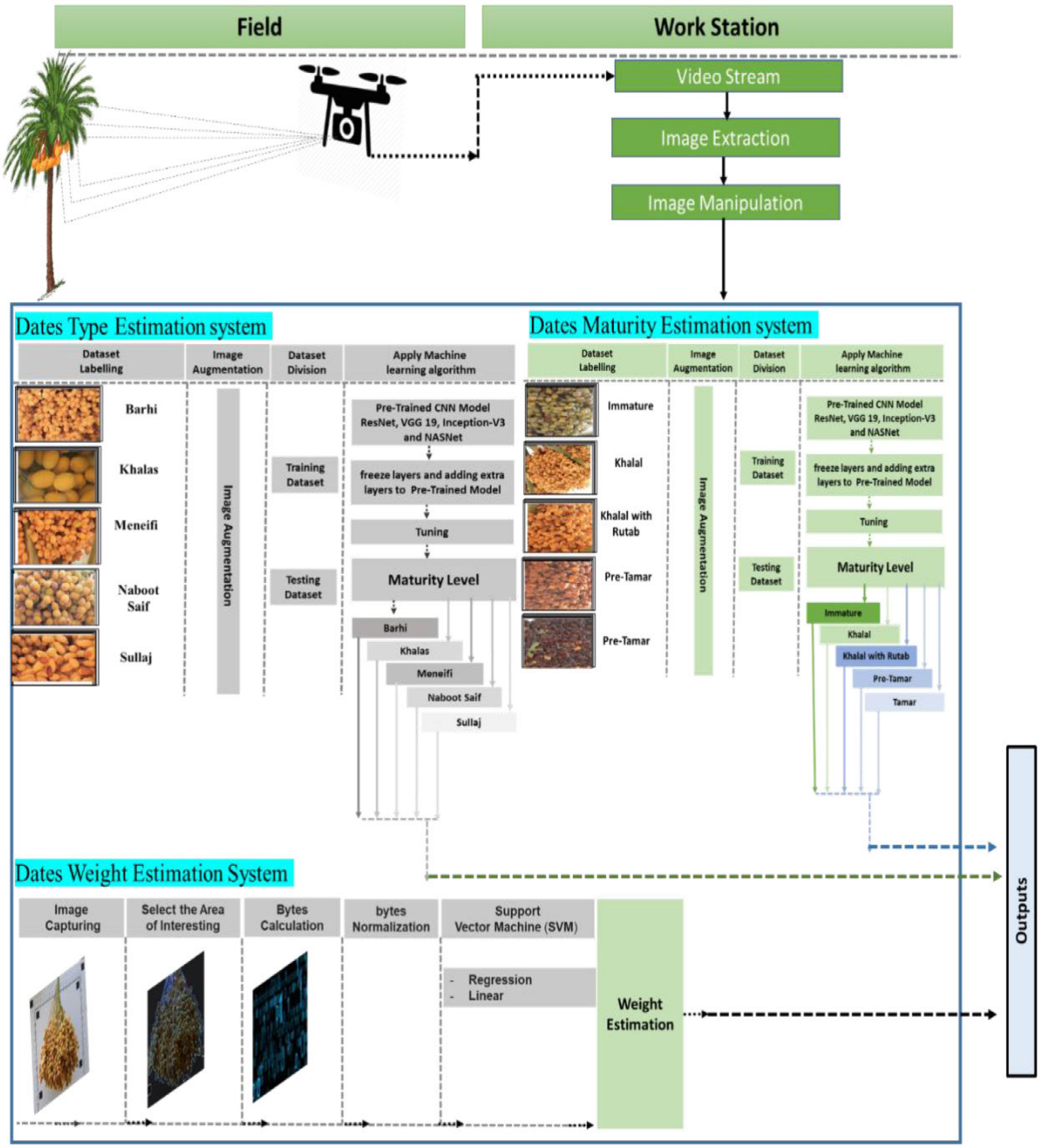


FIGURE 14. Proposed harvesting decision system.

the last layer. The Inception-v3 architecture had a total 23,916,327 parameters, 23,875,749 trainable parameters, and 38,528 non-trainable parameters for the seven-stage MLDS, and the Inception-v3 architecture had a total 6,415,001 parameters, 6,374,167 trainable parameters, and 40,834 non-trainable parameters for the seven-stage MLDS.

For DWES, we used SVM [22] for DWES. SVM is a supervised ML model introduced in **1992** by Vapnik at AT& T laboratories, used for classification and regression.

The main idea of SVM is to measure a maximum marginal hyperplane that best distributes the dataset into several classes. As shown in FIGURE 18, SVM classifies the faces into happy face and sad face classes.

Furthermore, SVM is used as a regression technique and called support vector regression (SVR). SVR has the same principles as the SVM for classification, with some variances due to the output, which is a real number that becomes very difficult to predict the information at hand, which has infinite possibilities. Therefore, a margin of tolerance is used to reduce the error. Generally, the objective of linear regression is to minimize the sum of squared errors. FIGURE 19 shows the linear regression of random data points.

B. CNN ARCHITECTURE PARAMETERS

Four pre-trained models ResNet, VGG-19, Inception-v3, and NASNet, were used in the proposed system. Training, evaluating, and testing was done using a machine with

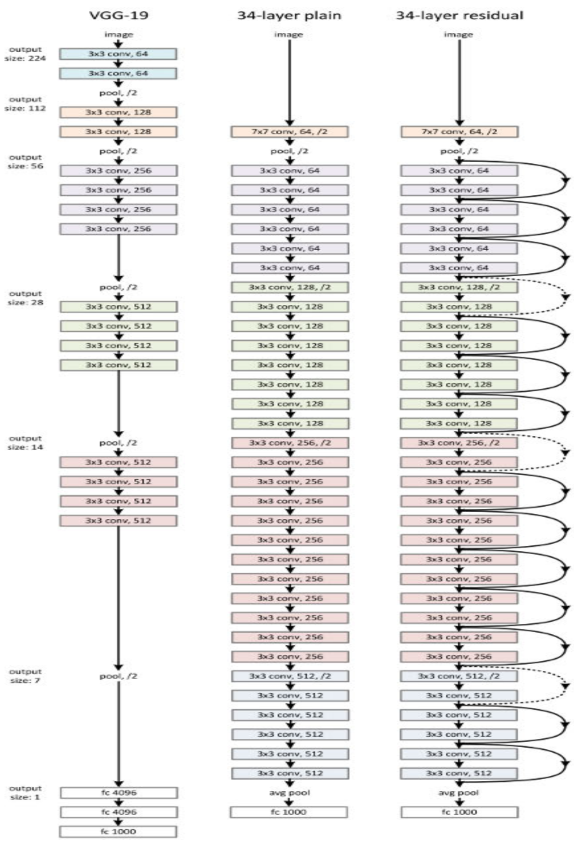


FIGURE 15. General overview of ResNet and VGG-19 architectures) [16].

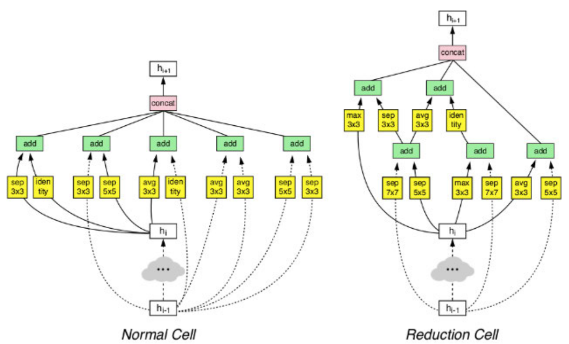


FIGURE 16. NASNet architecture.

Intel©i9-9880H core @ 2.3 GHz, NVIDIA GeForce RTX 2080 (8 GB) GPU, 32 GB RAM, and 64-bit Windows 10. In the training of all models, we used ImageDataGenerator for augmentation with the following parameters: rotation range = 40, height shift range = 0.2, width shift range = 0.2, shear range = 0.2, and zoom range = 0.2. Furthermore, we resized all images to 224 × 224. We used Keras 2.2.4 framework with Tensorflow 2.1.0 backend in Anaconda 4.8.3 environment using Spyder 3.7 development. Moreover, we used the training parameters: number of epochs = 50, batch size = 16, and ADAM optimizer with

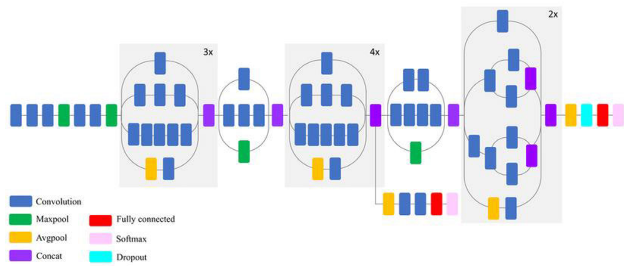


FIGURE 17. Inception-v3 architecture [21].

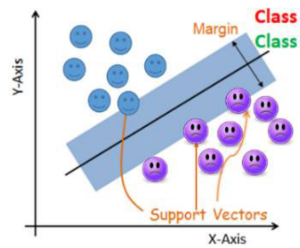


FIGURE 18. SVM classification example of two class (happy and sad face).

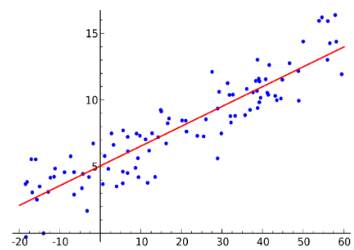


FIGURE 19. Linear regression of random data points (f(x) = 3* x/20 + 5, where x is real uniform' distribution in the period [– 20, 60].

learning rate = 0.0001. For training and testing, we used the four-fold cross-validation method.

V. EXPERIMENTAL RESULTS AND EVALUATION

The proposed system was evaluated based on the three subsystems: DTES, DMES, and DWES, using dataset-1 and dataset-2 [10]. We used ResNet, VGG-19, Inception-v3, and NASNet models to train, evaluate and test DTES to estimate five different dates types (Barhi, Khalas, Meneifi, Naboot Saif, and Sullaj), DMES to estimate the five maturity stage (Immature, Khalal, Khalal with Rutab, Pre-Tamar, and Tamar), and SVR to estimate the weight of date fruit. We used standard metrics, such as F1 score, accuracy, recall, precision, and confusion matrix, to evaluate the ResNet, VGG-19, Inception-v3, NASNet, and SVR models of DTES, DMES, and DWES, and compared DTES and DMES with other referenced studies.

For DTES and DMES, we achieved four-fold with 50 epochs for each cross-validation for all ResNet, VGG-19, Inception-v3, and NASNet models, and took the overall average of all results. In DTES, we used 1500 images,

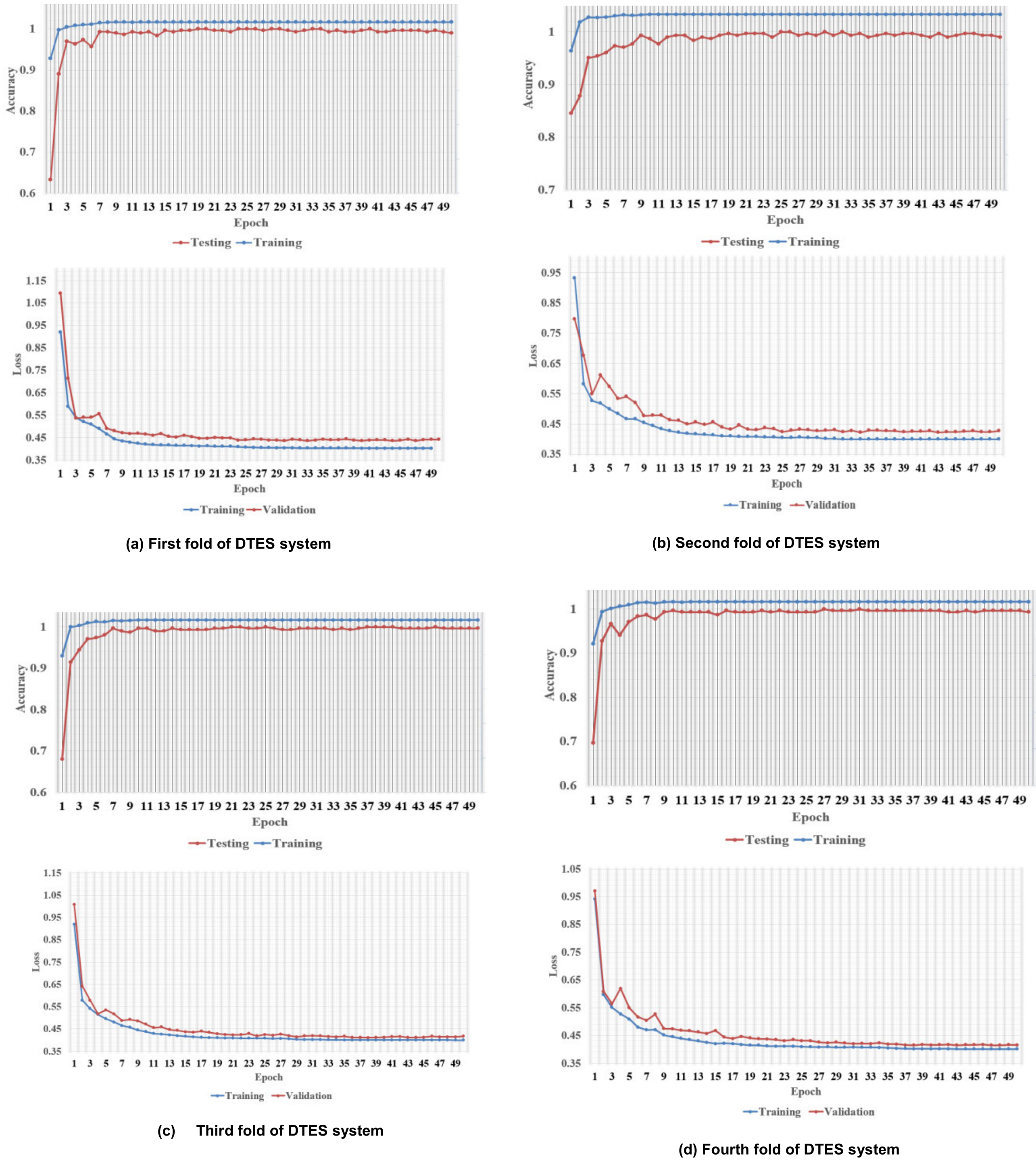
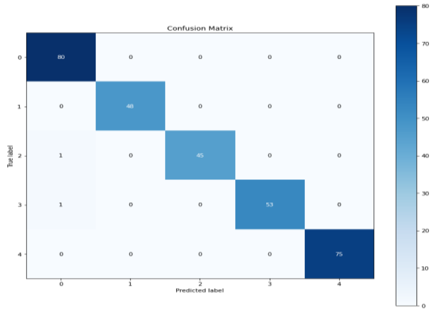


FIGURE 20. Learning performance accuracy, train, and validation learning curves of ResNet of DTES in four folds with 50 epoch.

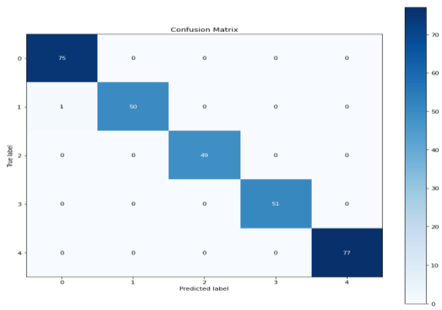
300 images per each type from different maturity states. For DWES, we used SVR to estimate the weight of the date fruit. **Error! Not a valid bookmark self-reference.** summarizes the execution performance of DTES and DMES in all four models tested using datasets-1 [10] and DWES tested using datasets-2 [10]. As shown in, DTES and DMES ResNet model outperformed the VGG-19, Inception-v3, and NASNet in all the performance metrics (F1 score, accuracy, sensitivity (recall), precision, and confusion matrix). For DTES, the ResNet achieved 99.175% accuracy, 99.225% F1 score, 99.8% average precision, and 99.05 % average recall, will the VGG-19 given 86.25% accuracy, 91% F1 score, 98.25% average precision, and 86.5% average recall; the Inception-V3 achieved 83.47% accuracy, 86.825% F1 score, 92.9%

	Barhi	Khalas	Meneifi	Naboot	Sullaj
Barhi	**# of TPs**	# of FPs	# of FPs	# of FPs	# of FPs
Khalas	# of FPs	**# of TPs**	# of FPs	# of FPs	# of FPs
Meneifi	# of FPs	# of FPs	**# of TPs**	# of FPs	# of FPs
Naboot	# of FPs	# of FPs	# of FPs	**# of TPs**	# of FPs
Sullaj	# of FPs	# of FPs	# of FPs	# of FPs	**# of TPs**

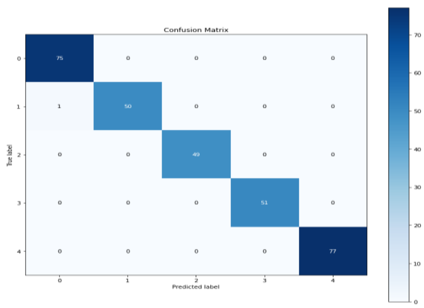
(a) Confusion matrix overview representation for DTES system, TP: True Positive, FN: False Negative



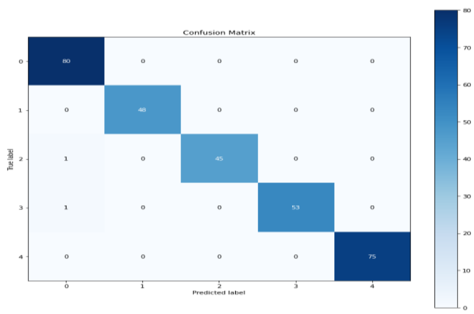
(b) Confusion matrix of first fold of DTES system



(c) Confusion matrix of second fold of DTES system



(d) Confusion matrix of third fold of DTES system



(e)

Confusion matrix of fourth fold of DTES system

FIGURE 21. The confusion matrices of ResNet of DLES with 50 epoch.

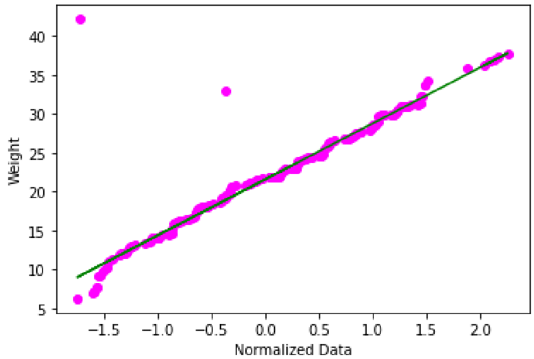


FIGURE 22. Visualizing SVM-Linear Results.

average precision, and 83.3% average recall; and NASNet achieved 77.5% accuracy, 83.75% F1 score, 92.8% average Precision, and 77.8% average recall.

For DTES, the ResNet achieved 99.058 % accuracy, 99.34 % F1 score, 99.64 % average precision, and 99.08% average recall. VGG-19 achieved 98.6% accuracy, 98.96% F1 score, 98.24% average precision, and 99.4% average recall; the Inception-V3 achieved 94.9% accuracy, 98.8% F1 score, 97.9 % average precision, and 99.4 % average recall. Furthermore, NASNet achieved 94.9% accuracy, 95.48% F1 score, 96.24% average precision, and 95.08 average recall. For DWES, we archived 84.27% and 83.78% score with SVM-Linear and SVM-Regression, respectively.

As mentioned earlier, we achieved four-fold cross-validation with 50 epochs for DTES and DMES and computed the overall average of the results. FIGURE 20 shows the learning performance accuracy of ResNet in four-fold cross-validation with 50 epochs of DTES.

As shown in FIGURE 20, ResNet has an excellent fit and stable performance. The training and validation loss decreased to the point of stability with a minimal gap between the two final loss values in all folds. FIGURE 21 shows the confusion matrix for ResNet for all four folds of DTES. FIGURE 22 shows the excellent fit of DWES.

VI. COMPARISON AND DISCUSSION

To compare the efficiency of the proposed system, we compared it with reference studies using the same dataset (dataset-1) and other datasets. The performance metrics used are accuracy, F1 score, recall, and precision. The studies by Faisal *et al.* [9] and Altaheri *et al.* [11] used the same datasets in a farm environment and the date fruit bunches in the orchard. In contrast, the study in [12] used a different dataset collected based on a single date type with uniform background. Table 3 shows the comparison results of the proposed system and the reference studies [9], [12], and [11]. In the proposed system, the ResNet model outperformed the other models and achieved outstanding results in all the performance metrics for both DMES and DTES.

TABLE 2. Performance metrics for DTES and DMES.

DTES (Barhi, Khalas, Meneifi, Naboot Saif, and Sullaj)		ResNet	VGG-19	NASNet	Inception-V3
Accuracy		0.99175	0.8625	0.775	0.83475
F1-score		0.99225	0.91	0.8375	0.86825
Avg. Precision		0.998	0.9825	0.928	0.929
Avg. Recall		0.9905	0.865	0.778	0.833
Precision	Barhi	0.9975	0.9775	0.92	0.9275
	Khalas	1	0.995	0.9475	0.935
	Meneifi	1	1	0.9825	0.925
	Naboot Saif	1	0.9775	0.9125	0.9275
	Sullaj	0.9925	0.9625	0.8775	0.93
Recall	Barhi	0.995	0.925	0.85	0.93
	Khalas	0.985	0.905	0.815	0.8625
	Meneifi	0.9925	0.87	0.65	0.8225
	Naboot Saif	0.9875	0.7675	0.76	0.75
	Sullaj	0.9925	0.8575	0.815	0.8

DMES (Immature, Khalal, Khalal with Rutab, Pre-Tamar, and Tamar)		ResNet	VGG-19	NASNet	Inception-V3
Accuracy		0.99058	0.986	0.981	0.949
F1-score		0.9934	0.9896	0.988	0.9548
Avg. Precision		0.9964	0.9824	0.979	0.9624
Avg. Recall		0.9908	0.994	0.994	0.9508
Precision	Immature	1	0.98	0.976	0.949
	Khalal	0.998	0.986	0.986	0.954
	Khalal with Rutab	0.994	1	0.99	0.96
	Pre-Tamar	0.998	0.984	0.996	0.97
	Tamar	0.992	0.996	0.998	0.954
Recall	Immature	1	0.976	0.974	0.976
	Khalal	0.978	0.982	0.978	0.936
	Khalal with Rutab	0.99	0.984	0.986	0.932
	Pre-Tamar	0.998	0.974	0.962	0.97
	Tamar	0.988	0.986	0.981	0.94

TABLE 3. Comparison with reference study.

System	Dataset	Maturity level	best Model in study	Accuracy	F1-score	Precision	Recall
Maturity level	Same Dataset	Our proposed (DMES)	ResNet	99.05	99.34	99.64	99.08
		Reference study ([9])	VGG-19	98.30	98.60	98.90	98.24
		Reference study (Five maturity) [4]	VGG-16	97.25	89.56	96.10	97.20
	Diff. Dataset	Reference study (Four Maturity)[12]	VGG-16	98.49		97.33	97.33
Data type	Same Dataset	Our proposed DTES	ResNet	99.10	99.20	99.80	99.05
		Reference study [4]	VGG-16	99.01	98.92	98.82	99.01

As shown in Table 3, the proposed DMES and DTES based on ResNet outperformed other studies. For DMES, the reference study [4] used VGG-16 and achieved an accuracy of 97.25%, F1 score of 89.56%, sensitivity (recall) of 96.1%, and precision of 97.2%. The reference study [12] achieved 98.49%, 97.33%, and 97.33, accuracy, sensitivity (recall), and precision, respectively, using VGG-16 for four maturity levels. In contrast, our proposed system achieved 99.05% accuracy, an F1 score of 99.34%, 99.64% sensitivity (recall), and 99.08% precision. For DTES, the reference study [4] used VGG-16 and achieved 99.01% accuracy, 98.92% F1 score, 98.82% sensitivity (recall), and 99.01% precision, where our proposed system achieved 99.1% accuracy, 99.2% F1 score, 98.8% sensitivity (recall), and 99.05% precision. For DWES, we did not find any system that estimates the weight of date fruit in the literature.

VII. CONCLUSION

In this paper, we proposed a real-time CV and DL-based system for estimating date fruits type, maturity level, and weight in an orchard. The proposed system is made up of three sub-systems: DTES, DMES, and DWES. We used four DL architectures: ResNet, VGG-19, Inception-V3, and NASNet for DMES and DTES, and SVM (regression and linear) for DWES. Based on multiple performance metrics, DTES achieved maximum performance of 99.175% accuracy, 99.225% F1 score, 99.8% average precision, and 99.05% average recall. DMES achieved a performance of 99.058% accuracy, 99.34% F1 score, 99.64% average precision, and 99.08% average recall, and DWES was 84.27% SVM-linear. Furthermore, we compared the proposed system with existing systems, and it comparably outperformed the others. We plan to extend the system and the dataset to cover more date types and to use another DL method for weight estimation in future work. In additional, in the future, we are planning to apply the proposed system with other datasets of different fruit types. Besides, we are planning to use other machine learning algorithms with weight, maturity, and type estimation.

REFERENCES

[1] (2015). *Food and Agriculture Organization of the United Nations*. [Online]. Available: http://faostat3.fao.org/browse/Q/QC/E

[2] D. Surya Prabha and J. Satheesh Kumar, "Assessment of banana fruit maturity by image processing technique," *J. Food Sci. Technol.*, vol. 52, no. 3, pp. 1316–1327, Mar. 2015.

[3] D. Zhang, D.-J. Lee, B. J. Tippetts, and K. D. Lillywhite, "Date maturity and quality evaluation using color distribution analysis and back projection," *J. Food Eng.*, vol. 131, pp. 161–169, Jun. 2014.

[4] P. R. Gokul, S. Raj, and P. Suriyamoorthi, "Estimation of volume and maturity of sweet lime fruit using image processing algorithm," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2015, pp. 1227–1229.

[5] K. Yamamoto, W. Guo, Y. Yoshioka, and S. Ninomiya, "On plant detection of intact tomato fruits using image analysis and machine learning methods," *Sensors*, vol. 14, no. 7, pp. 12191–12206, Jul. 2014.

[6] Y. Tang, M. Chen, C. Wang, L. Luo, J. Li, G. Lian, and X. Zou, "Recognition and localization methods for vision-based fruit picking robots: A review," *Frontiers Plant Sci.*, vol. 11, p. 510, May 2020.

[7] G. Lin, Y. Tang, X. Zou, J. Cheng, and J. Xiong, "Fruit detection in natural environment using partial shape matching and probabilistic Hough transform," *Precis. Agricult.*, vol. 21, no. 1, pp. 160–177, Feb. 2020.

[8] J. Li, Y. Tang, X. Zou, G. Lin, and H. Wang, "Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots," *IEEE Access*, vol. 8, pp. 117746–117758, 2020.

[9] M. Faisal, M. Alsulaiman, M. Arafah, and M. A. Mekhtiche, "IHDS: Intelligent harvesting decision system for date fruit based on maturity stage using deep learning and computer vision," *IEEE Access*, vol. 8, pp. 167985–167997, 2020.

[10] H. Altaheri, M. Alsulaiman, M. Faisal, and G. Muhammed, "Date fruit dataset for automated harvesting and visual yield estimation," in *Proc. IEEE DataPort*, 2019.

[11] H. Altaheri, M. Alsulaiman, and G. Muhammad, "Date fruit classification for robotic harvesting in a natural environment using deep learning," *IEEE Access*, vol. 7, pp. 117115–117133, 2019.

[12] A. Nasiri, A. Taheri-Garavand, and Y.-D. Zhang, "Image-based deep learning automated sorting of date fruit," *Postharvest Biol. Technol.*, vol. 153, pp. 133–141, Jul. 2019.

[13] S. K. Behera, A. K. Rath, and P. K. Sethy, "Maturity status classification of papaya fruits based on machine learning and transfer learning approach," *Inf. Process. Agricult.*, May 2020.

[14] W. D. N. Pacheco and F. R. J. Lopez, "Tomato classification according to organoleptic maturity (coloration) using machine learning algorithms K-NN, MLP, and K-means clustering," in *Proc. 22nd Symp. Image, Signal Process. Artif. Vis. (STSIVA)*, Apr. 2019, pp. 1–5.

[15] J. A. Caladcad, S. Cabahug, M. R. Catamco, P. E. Villaceran, L. Cosgafa, K. N. Cabizares, M. Hermosilla, and E. J. Piedad, "Determining philippine coconut maturity level using machine learning algorithms based on acoustic signal," *Comput. Electron. Agricult.*, vol. 172, May 2020, Art. no. 105327.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[19] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[21] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, no. 7, p. 1119, Jul. 2018.

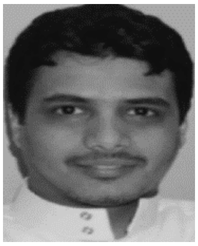
[22] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. workshop Comput. Learn. Theory (COLT)*, 1992, pp. 144–152.

MOHAMMED FAISAL (Member, IEEE) received the master's and Ph.D. degrees from King Saud University, in 2002 and 2016, respectively. He works as an Assistant Professor and supervises the unit of Innovation and Entrepreneurship in the College of Applied Computer Science, King Saud University, where he is also a Robotics Consultant with the Center of Smart Robotics Research. His research is currently focusing on the use of robots and artificial intelligence to improve the quality of life and find innovative solutions. In recent years, he has published scores of scientific research in refereed and classified scientific journals and conferences. He won the President of the Republic of Yemen Youth Award for the Applied Sciences Branch for the year 2013, as well as many scientific/research awards and medals.

FAHAD ALBOGAMY is currently an Assistant Professor with Taif University, Al-Hawiya, Saudi Arabia. He is interested in artificial intelligence, NLP, and digital image processing.

HEBAH ELGIBREEN is currently an Assistant Professor with the Department of Information Technology, King Saud University, and also a Research Affiliate at the MRL Laboratory, Mechanical Engineering Department, MIT. She is currently the Deputy Director of the Alumni Center. She has developed a strong interest in artificial intelligence and machine learning, and is currently working on robotics. She is also a member of the Center of Smart Robotics Research, KSU.

MOHAMMED ALGABRI received the master's degree from King Saud University, where he is currently pursuing the Ph.D. degree with the Computer Science Department, College of Computer and Information Science. His research interests include deep learning, speech recognition, soft computing techniques, and robotics.



FATTOH ABDU ALQERSHI received the master's degree in information systems from the College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia, in 2012, where he is currently pursuing the Ph.D. degree. From 2007 to 2009, he was an Assistant Lecturer with the Department of Software Engineering, Faculty of Engineering and Information Technology, Taiz University, Yemen. His research interests include online social networks, social media analysis, and crowdsourcing.