IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 December 2023, accepted 26 January 2024, date of publication 5 February 2024, date of current version 12 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3362294

RESEARCH ARTICLE

Highly Efficient Machine Learning Approach for Automatic Disease and Color Classification of Olive Fruits

NASHAAT M. HUSSAIN HASSAN[1,2], A. A. DONKOL[3,4], M. MOURAD MABROOK[5], AND A. M. MABROUK[6]

[1]Faculty of Engineering and Technology, Badr University in Cairo (BUC), Badr City 62511, Egypt
[2]Electronics and Communication Engineering Department, Fayoum University, Fayoum 63514, Egypt
[3]Electrical Engineering, Communication and Electronics Department, Faculty of Engineering, South Valley University, Qena 83523, Egypt
[4]Electrical Engineering Department, Faculty of Engineering, Nahda University in Beni Suef, Beni Suef 46511, Egypt
[5]Space Communication Department, Faculty of Navigation Science and Space Technology, Beni-Suef University, Beni Suef 62511, Egypt
[6]Faculty of Engineering, King Salman International University (KSIU), South Sinai, El Tor, Egypt

Corresponding author: M. Mourad Mabrook (mohamed.mourad@nsst.bsu.edu.eg)

ABSTRACT The following ends have been established via an in-depth examination and assessment of numerous prior studies on olive fruit classifications: First, several of these researches rely on the use of an unrelated image library. Since every image features a single fruit with a background that contrasts sharply with the fruit's hue, they are all ready for testing. As was previously stated, this issue is unrelated to reality. In practical application, one must deal with a frame that holds hundreds of fruits. To keep the fruits steady, they are put on a conveyor with multiple channels. It's also notable that the majority of this study offered suggestions for useful technology that could yet be developed. Finally, it is important to emphasize that processing speed data is essential in this type of application and has not been collected in many of these experiments. The presented work deals with a new strategy based on two principles: first, a successful extraction of the fruits from the background; and second, the classification of olive fruits into eight categories based on colors and defects. The fruits were extracted from the backdrop using a modified version of the K-Means technique. The outcomes of the suggested fruit extraction were examined utilizing several assessment techniques. By contrasting the outcomes of pertinent procedures with the suggested proposal for fruit extraction, the efficacy and precision of the proposed method were verified. Depending on why the fruit needed to be separated, there were two stages to the process. Three colors were separated using the SVM algorithm, and five distinct defects were separated using the ANN algorithm Approximately 15,000 photos of olive fruits that were shot straight from the fruit conveyor were included in a robust database that was used in the proposed study to validate the effectiveness of the suggested technology. Efficiency was further validated by contrasting our outcomes with those of related technology. When the fruits were set on a white backdrop, the test accuracy results of the suggested approach showed that it was highly efficient in classifying the fruits in the shortest period; the suggested method had an effectiveness of 99.26% for fruit classification. The most important discovery was that it could classify fruits with an efficiency of 97.25% while they were being put on a fruit conveyor, which was in contrast to other approaches. The unique findings of the study that was presented hold promise for practical implementation.

INDEX TERMS Olive fruit pre-processing, olives detecting and extracting, features extractions, SVM classifier, ANN classifier, hyper parameters tuning.

I. INTRODUCTION

In recent times, investment companies involved in the trading of agricultural products have placed double the importance on the automatic separation of olive fruits and vegetables.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongquan Sun.

For a variety of reasons, including the wide distribution of olive fruits worldwide, the rising demand for eating them raw or cooked due to their high nutritional value, their excellent application in numerous industries, the decreased overall cost of manual separation, the ability to increase the amount of separation per hour, and the improvement of separation quality, these companies are becoming more and more dependent on the automatic separation of olive fruits [1], [2], [3]. Apart from that, the prices of these fruits vary according to their shapes and colors. The fruit's defect-free status also affects how much these fruits cost. In addition, the cost can vary depending on the kind and size of the defect. Even with the significant advancements in recent years in the technology of automatic fruit and vegetable separation, the processes of automatic olive fruit separation continue to confront numerous obstacles for a variety of reasons. These include the fact that these fruits are not readily available as a database, their small size, their wide variety of colors within the same type, their striking similarity amongst different types, the impact of their colors on the surrounding environment (rain, wind, sun, dust), and the significant interference that frequently arises between the fruits and the conveyor surface. Additionally, it's not a simple task to manage a whole frame with more than 300 olives at once, as Figure 1 illustrates. Ultimately, it's not always simple to combine the separation method with the highest accuracy, quickest speed, and least amount of expense. It has been demonstrated that there has been a lot of recent interest in this topic. A large number of these research were dependent on the application of deep learning algorithms [16], [17], [18], [19], [20], [21], [22], [23] and machine learning [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. The majority of these methods rely on a sequence of sequential steps, beginning with pre-processing, which comprises numerous preparatory operations including resizing, color conversion, noise reduction, and image enhancement that support the primary task. Image segmentation is the next step (the accuracy of the findings is highly dependent on the approach selected). The quantity and caliber of the unique characteristics in relation to what makes the application the best fit for implementation—this element could have an impact on the primary work's accuracy. To sum up, choosing the best classifier from a huge selection is one of the most crucial factors in determining performance quality After reviewing and evaluating numerous earlier studies, we came to the following conclusions: first, the majority of these studies relied on applying their technology to a database, which frequently has few images and is insufficient to confirm the effectiveness of the suggested technology, in order to verify the accuracy of their technology. Furthermore, most of these investigations rely on the use of an unrelated picture database. Since each image features a single fruit with a background that is entirely distinct from the fruit's hue, they are all ready for testing. As previously stated, this issue is unrelated to actuality. In practical uses, one must deal with a frame that has hundreds of fruits in it. To keep the fruits stable, they are arranged on a conveyor with multiple channels. Notable is also the fact that the majority of these research offered suggestions for effective technologies that are still open to improvement. Lastly, it is noteworthy that processing speed data—which has not been obtained in many of these studies—is crucial in this kind of application.

FIGURE 1. 1 example of a frame of olive fruits.

All of the prior research' inadequacies were considered in the current investigation, both with regard to the quantity and quality of photos used from the database because they were taken straight from the fruit conveyor. Additionally, the effectiveness of the suggested technique was confirmed by using a method that involved splitting the categorization procedure for olive fruits into many stages. Ensuring that the right feature extractions are chosen for each stage and that each classifier undergoes testing, validation, and training to ensure that the classification process is accurate at every level. A comparative analysis is conducted between the suggested technology and related technologies. Additionally, the computation times for the suggested technique were examined in the proposed study.

The presented work is constructed as the following: in section two, a literature review has been studied and analyzed. An explanation of the proposed strategy is discussed in detail in the third section. In the fourth section, the comparison between the outcomes of the suggested technique and those of comparable procedures is explained. Finally, the conclusion and future work is demonstrated

II. LETERETURE REVIEW

This section reviews and analyzes several recent papers that deal with offering suggestions for methods that further the development of automatic fruit and vegetable separation in general, and olive fruit separation in particular. Starting with a study report written by Pablo Cano Marchal and published in the MPDI journal in 2021 [24], let's begin discussions. The suggestion made by this author was to just separate the contaminated olive fruits from the healthy ones. In order to do this work, the authors used an infrared camera. What follows in relation to this paper: First off, using infrared camera images is limited to applications that can distinguish between just two fruit varieties. Second, since handling fruit picks while they are on the stand is more difficult than handling them off the stand, it was preferable to remove the fruit picks from the fruit conveyor rather than from somewhere else. Third, the processing is too slow; the author stated that an image takes an average of 100 seconds to process, which is too lengthy for making decisions in real time. Fourth, the quality of the suggested technique's performance still has to be improved; the report states that it ranged from 79% to 91%. A 2022 article written by Simone Figorilli and published in the Journal of MPDI [25] The author proposed separating infected green olive fruits from infected black fruits and comparing the resulting separation quality at varying carrier speeds (low, high). CNN techniques were utilized by the authors of this work. What this study incorrectly performs is the following: First off, as we previously mentioned, handling the fruit picks while they are on the stand is more difficult than handling them directly, so it would have been preferable to take them from the fruit conveyor rather than from another location when evaluating the efficiency of the suggested technique. She is not pregnant with her. Second: The suggested technique's performance quality still need work; according to the report, it varied from 63% to 69% for black fruits to 88–89% for green fruits. Third: Because the suggested technology depends so heavily on Deep Learning (CNN) techniques, it is anticipated to be slow.

Additionally, Hussain, an author, published an article in the Springer Journal in 2018 [26]. This author proposed dividing olive fruits into healthy and defective categories, then dividing the defects into small and large categories based on their size. This paper's author relies on a variety of image-processing methods. This paper makes mention of the following: I reduced the amount of samples used for testing (there are 350 original photos total in the dataset, each representing four things). offers a high processing speed and efficiency even while using photographs from the stand.

Another technology that was released in 2020 was that of author Mazen and Nashat's [19], [27]. This method was reliant on ANN techniques. This book presented the idea of differentiating between the four types of banana fruits: green, yellowish green, mid-ripe, and overripe. This study report has flaws because the writers used a basic database with no more than 300 samples per item and they failed to provide the results of their suggested technique's processing speed.

Lastly, a study presented by Ponce [28] and published in the 2017 IEEE ACCESS journal reviewed the accuracy of the performance of the six most well-known deep learning approaches for classifying olive fruits into seven different kinds. Table 1 provides an overview of the findings from the evaluation of these methods' performance accuracy in fruit grading. The average accuracy of performance for these techniques ranged between 89.9% in favor of AlexNet and 95.9% in favor of Inception-V3, indicating a variation in performance accuracy between these different techniques. The study's first disagreeable aspect is that, according to the research, only 2,800 photographs were used. Secondly, these aren't screenshots from the fruit conveyor. As was previously indicated, handling fruit samples that are removed from the conveyor is far more challenging than handling fruit that is placed on ordinary surfaces. Thirdly, no private data computation times were addressed in the investigation. Fourth, there's still room for improvement in the performance quality.

TABLE 1. Test accuracy results obtained by six different CNN architectures to classify the olive fruits into seven categories.

Olive fruits categories	CNN Architectures					
	AlexNet	Resnet-50	Resnet-101	InceptionV1	Inception-ResnetV2	InceptionV3
Arbequina	0.9333	0.9333	0.9467 0.9533	0.9533	0.9400	0.9267
Arbosana	0.9133	0.9067	0.9467	0.9400	0.9133	1.0000
Changlot	0.8667	0.9400	0.9200	0.9600	0.9600	0.9333
Lechin	0.8333	0.8867	0.9267	0.9067	0.9200	0.9067
Picual	0.9467	0.8933	0.9600	0.9600	0.9400	0.9867
Ocal	0.9133	0.9333	0.9667	0.9933	0.9667	0.9733
Verdial	0.8867	0.9333	0.9733	0.9600	0.9600	0.9867
Overall	0.8990	0.9181	0.9486	0.9533	0.9400	0.9591

III. PROPOSED METHODOLOGY

The purpose of this work is to introduce a novel technique for the quick and effective separation of eight distinct olive fruit classes. The method for classifying and dividing olive fruits into eight categories healthy, with three different colors, and defective, with five different defects that has been suggested relies on two stages of implementation of the classification process: the first stage involves dividing the fruits into two categories (damaged and healthy fruits), and the second stage involves multiple separations. Fruits can be categorized as either healthy or defective. While healthy fruits can have three distinct colors (green, black, or burgundy), damaged fruits can have five different types of defects (serpeta, rehu, molestado1, wrinkled, and molestado2). The novel technique includes seven stages specifically dataset collection, pre-preprocessing (feature scaling, dataset augmentation, dataset splitting), image segmentation for detecting and extracting the olive fruit from the background (pregnant), features extractions (to increase the accuracy of the classifier models by removing the redundant data, and extracting distinctive features, which will increase training speed, and also improve the accuracy of the decision), model classifier construction, training, validation, hyper-parameters tuning, and testing. The workflow of our proposed approach is illustrated in figure 2.

A. DATASET COLLECTION AND DESCRIPTION

The dataset utilized in our proposed work comprises over 3750 different types of olive fruits, both healthy and defective. This dataset was gathered from ''invistigacion assitida de vision artificial,'' a Seville, Spain-based business. The division of these photos was as follows: Twenty percent of these photos were used for validation, sixty percent were used for training the suggested method, and twenty percent were used to assess the suggested technique's correctness. During the augmentation procedure, this quantity of photographs was duplicated four times, totaling fifteen thousand images used in this study. The Basler aviator camera avA1900-60km/kc (digital monochromatic-color, 1920 × 1080 pixels) is used for capturing all of the material being tested. These images distributed over the different eight classes as shown in table 2

B. IMAGE PRE-PROCESSING

Our idea is first treated using data dividing, augmenting the data, and scaling of features. In our situation, the feature scaling technique involved resizing the datasets into 128 × 128 dimensions using RGB formatted pictures. Other than to image resizing, our team additionally normalized the gathered dataset as part of our feature scaling process. This is to reduce the impact of illumination differences, additionally, the CNN ends faster when data is provided between [0, 1] than it does when data is provided between [0, 255]. The second phase of our proposal to pre-process the collected data is a data augmentation. The phrase ''augmentation'' describes the process of making the dataset larger. To prevent over-fitting, it is therefore utilized to increase the number of data samples and possibly the rate of variance in our dataset. In our case, all the collected data was rotated at ten different angles: 0°, 90°, 180°, and 270°. Table 3 displays some of our augmented phase results.

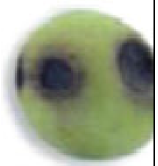
C. SEGMENTATION PHASE

In real applications, the fruits are handled while they are on the conveyor, which increases the difficulty of handling the fruits due to the overlap and similarity in many cases between the fruits and the interface of the conveyor. Therefore, one of the important processes for grading the fruits is good extraction of the fruits from the conveyor. Therefore, the purpose of our segmentation phase is the good extracting and separation of the olive fruits from the conveyor. Three approaches have been tested in order to guarantee the accuracy of our segmentation: Fuzzy C-means, modified K-means

TABLE 2. Description of the used subset of healthy and defective olives in the sample.

Classes Name		Total	Training Number	Validation Number	Testing Number	Samples
Healthy	Healthy_Green	6000	3600	1200	1200	
	Healthy_Purple	2000	1200	400	400	
	Healthy_Black	2000	1200	400	400	
Defected	Serpeta	1000	600	200	200	
	Granizo	1000	600	200	200	
	Rehu	1000	600	200	200	
	Molestado -1	1000	600	200	200	
	Molestado -2	1000	600	200	200	

TABLE 3. Samples of our augmented phase results.

Original Images	Rotated by 90°	Rotated by 180°	Rotated by 270°
 IMG1			
 IMG2			

clustering, and Otsu's thresholding technique. The following was the basis for the establishment of the modified-K-means algorithm and its implementation: 1. Selecting k = 3 to divide the three primary colors—black, burgundy, and green—so that each appears in a distinct channel. 2- Each channel undergoes a separate thresholding process due to potential interference between the fruits and the conveyor. 3. The three channels are then combined to create the image depicted in the accompanying figure. Based on the data collected, it was determined that the Modified-K-Means method is the most effective of the three in terms of extracting the fruits fully and without losing any of them, as well as totally eliminating the background and leaving no trace of the fruits. Even if the fruits have a nice appearance, it is observed in the Fuzzy C-Means technique findings that some background appears and does not entirely disappear, which could have a detrimental impact on the decision's accuracy. Otsu's approach did not resolve the issue; quite the reverse. Not only did portions of the backdrop with the fruits appear, but it was also discovered that portions of the objectives (olive fruits) vanished. This phenomenon increases the likelihood of making incorrect selections about how to divide the various fruit varieties. The results of the three tried approaches—the original image, the Modified K-means clustering algorithm, the fuzzy C-means algorithm, and Otsu—are displayed in Figure 3. A review of all the results of the process of separating the fruits from the stand and an explanation of these results will be discussed in the results section.

Following the extraction of the olive fruits from the background, the olive fruit batch is split into trails in accordance with the training work's needs. The olives are then separated for each trail, and as seen in figure 4, each individual fruit is treated as a single image. 3. This procedure will only be used when training the olive fruit; however, the entire frame—rather than just each item for a unit—will be tested. What is actually done is deal with the entire frame.

D. FEATURE EXTRACTION PHASE

The chosen set of features for the proposed study was made after a thorough review of numerous prior publications in the field to determine which features produced the greatest results for texture analysis as well as the best outcomes for color analysis. Consequently, it was determined to select the features GLCM [29], LBP [30], and MLBP [31] for fruit texture analysis and CCT [32] and CCV [33] for fruit color analysis because at this point the healthy and diseased fruits have been separated, necessitating an examination of both the fruits' color and texture. However, the features In a subsequent step, CCT and CCV were selected to divide colors. Finally, classify the various flaws using GLCM, LBP, and MLBP characteristics.

E. PHASE OF OLIVE CLASSIFICATION

The goal of this stage is to distinguish between the many types of olive fruits. According to the suggested technique, the separation process is carried out in two steps: To distinguish between healthy and defective olives is the first step. The healthy olives are divided into three categories of color (Green, Burgundy, and Black), and the defective olives are divided into five categories (Serpeta, Granizo, Rehu, Molestado, and Molestado2). The first stage of separation employed the SVM classifier, and the second stage employed the ANN. The next two paragraphs offer a succinct overview of the separation methods that were employed.

1) SUPPORT VECTOR MACHINE

An algorithm for supervised learning is the SVM classifier [34]. SVM is a preferable option for binary classification, a technique employed for linearly separable issues.

$$y_i = W \times X_i + b \tag{1}$$

where is xi is referred to as a feature vector with n dimensions that can be plotted in n dimensions. These feature vectors each have a class yi label on them. Where W and b are line parameters, the class yi can either be 1 or −1. To ensure accurate classification, SVM maximizes the boundaries between different classes. It places a strong emphasis on lowering structural risk while learning. The ideal hyperplane is given by the equation w.x+b = 0. Equation w.x+b=−1 for the left support vector and w.x+b=1 for the right support vector [35]. In our case, the database file consisted of 3500 samples of olive fruits distributed randomly, each sample containing five features (Color classification technique, CCV, LTP, HOG, and LBP) for two classes of olive fruits (Healthy and defected olives). 70% of this data was used to make the training, and the rest of the images were used to test the accuracy of the algorithm.

2) ARTIFICIAL NEURAL NETWORKS (ANN)

In the second stage of the olives classifications, the ANN classifier [36] is used. Two different architectures of the proposed ANN classifier were used the first ANN architecture

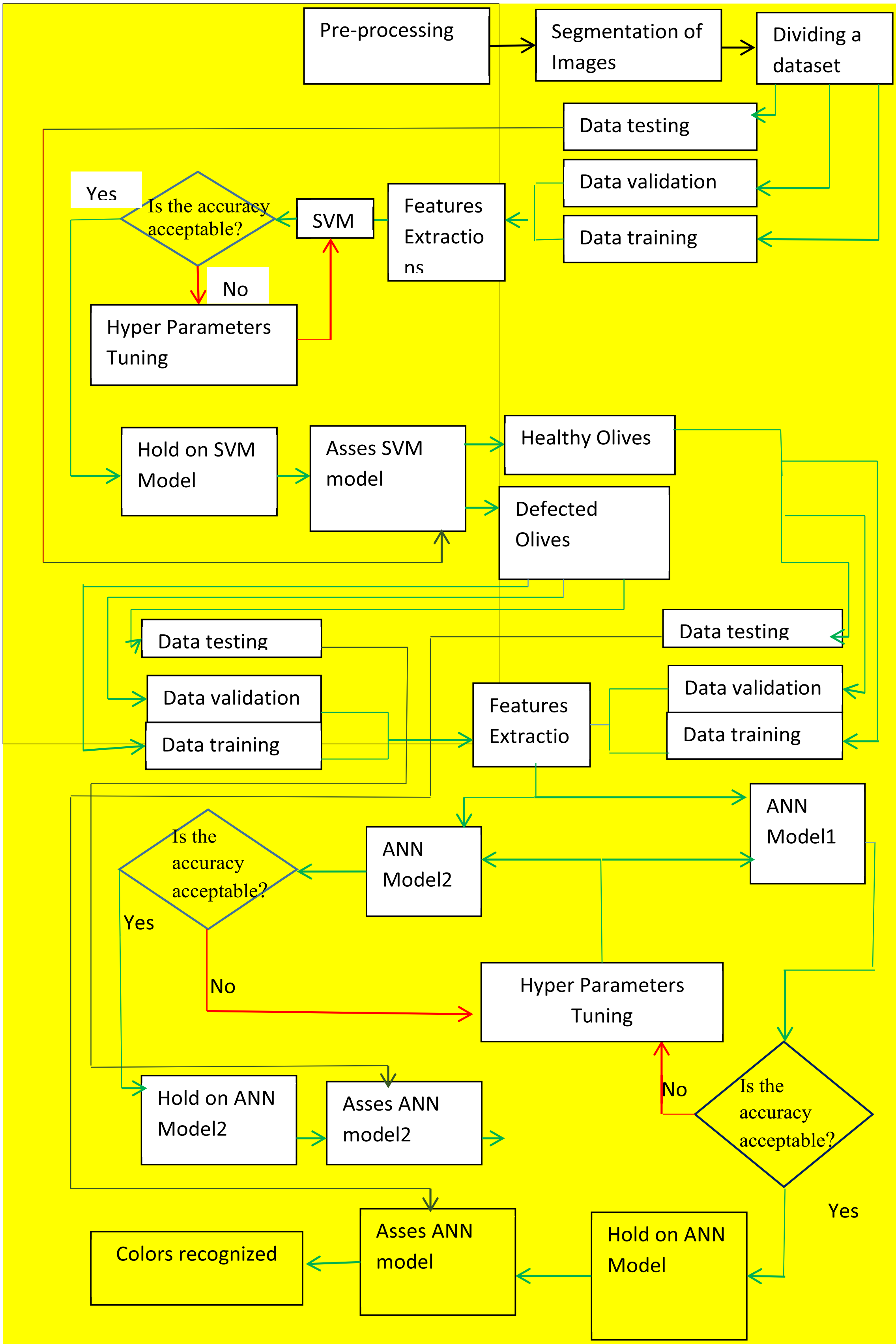


FIGURE 2. A complete workflow of the planned method.

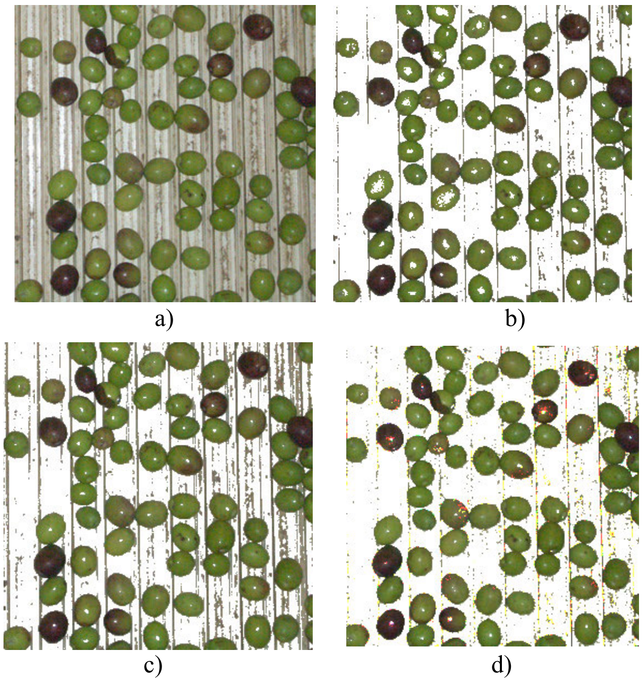


FIGURE 3. **Samples of the segmentation results: a) original image, b) Otsu thresholding results, c) Fuzzy-C-Means results, d) Modified-K-Means results.**

FIGURE 4. **Sub-images of defected and healthy olives.**

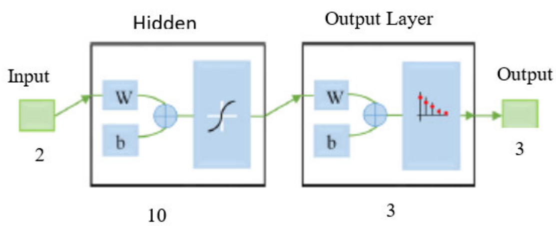


FIGURE 5. **The first design of the suggested ANN1 model.**

is used to classify among the three different colors of the healthy olive fruits the second ANN architecture is used to classify among the five different defects of the defective olive fruits. The first and second ANN architectures are shown in Figures 5 and 6.

Two neurons make up the input layer of the first ANN architecture. These are the two features color classification

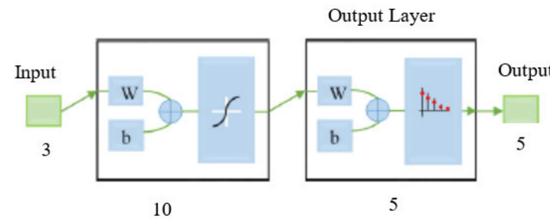


FIGURE 6. **The second design of the suggested ANN2 model.**

technique, and CCV. The output layer of this ANN model has three neurons, which stands for the three levels of olive fruit colors. While three neurons make up the input layer of the second ANN architecture. These are the three features (LTP, LBP, and HOG). The output layer of this ANN model has five neurons, which stands for the five levels of olive fruit defects. Each one of the two architectures has one hidden layer with ten neurons which make up the model. Due to the sigmoid function's straightforward derivative and soft switching capabilities, it is used as the activation function. In each case, the error is produced by deducting the goal output from the neural network's actual output, which is created by the output layer. The mean squared error is sent back into the hidden layers of the network as a performance function, updating the weighted sum of each neuron's input and bias.

F. HYPER PARAMETERS TUNING

Hyper parameters play a crucial role in creating reliable and precise models. They assist us in striking a balance between variance and bias, preventing the model from being over- or under fitted. Finding the optimum potential sets of hyper parameters to construct the model from a given dataset requires adjusting hyper parameter. In the presented study, as mentioned previously, the classification process was carried out in two stages, the first stage used the SVM classifier, and the ANN classifier was used in the second stage. Therefore, as shown in the diagram of the proposed method, hyper parameters tuning was done for both classifiers separately, as follows: First, respecting to the SVM classifier, In SVM techniques, drawing or determining the decision border is an especially important step. A decision border distinguishes between multiple categories. The C and gamma are the two most important parameters that control the drawing or determining the decision border. Since the C parameter and the margin size have an inverse relationship, a bigger value of C indicates a smaller margin, and a smaller value of C indicates a larger margin. Conversely, the gamma parameter has an inverse relationship with distance. The decision boundary is determined by taking into account points that are closer together, as indicated by a higher gamma value, and farther apart, as indicated by a lower gamma value. In the presented study, It was decided to use an experimental range for the C parameter, which goes from 0.0001 to 10, and an experimental range for the gama parameter, which goes from 1 to 100. Should ultimately choose the values 0.0001 for the c parameter and 10 for the gama parameter while keeping in mind that over fitting won't occur if the best quality is achieved. Secondly, with regard to the ANN classifier, there are several things that are taken into consideration, including the type of the neural network, the activation function, learning rate, and error function. in addition to the number of hidden layers, and the number of neurons on each layer. In our recommended ANN model, the following hyper-parameters were addressed using the validation set for ANN model 1 and ANN model 2: Feed forward neural network, including one input layer with 2-neurons, one hidden layer with 10 neurons and one output layer with 3-neuron for ANN model1. Feed forward neural network, including one input layer with 3-neurons, one hidden layer with 10 neurons and one output layer with 5-neuron for ANN mode2. ReLU activation for both the two models (ReLU function is calculated in mathematical function 7) [36].

$$f(x) = \max(0, x) \tag{2}$$

where (x) is the input data

Learning rate for both models is more than 0.001 and less than 0.01, respecting the error function, the average absolute percentage error (AAPE) is used for both models, where it is calculated as follows [36]:

$$AAPE = \frac{\sum_1^n \left|\frac{y_i - y_i'}{y_i}\right|}{n} 100\% \tag{3}$$

where yi is the actual value, yi' is the estimated value.

IV. RESULTS AND DISCUSSION

The outcomes of using the suggested study to extract and categorize olive fruits are examined in this section. These outcomes cover the following two areas: First, the outcomes of the olive fruit extraction process from the conveyor belt, whereby two different kinds of mechanisms were examined in order to gauge the effectiveness of the Modified-K-means-Clustering technology. This is accomplished by both quantitative and visual assessments, as well as a comparison between the outcomes of the Otsu clustring and Fuzzy C-Means approaches and the Modified-K-Means Clustering results. Second, the results of training and testing the suggested method for classifying fruits are examined, along with a comparison of the suggested method's outcomes with those of other pertinent procedures. All of these outcomes are examined below.

A. RESULTS OF OLIVES FRUIT SEGMENTATION

Two forms of assessment mechanisms were utilized to ensure the effectiveness of the technology used in the separation process (Modified-K-Means-Clustering): evaluation by observation (vision) and also by quantitative evaluation, as illustrated below.

1) VISION ASSESSMENTS

Table 4 shows the outcomes of the Modified-K-Means-Clustering algorithm used to extract the olive fruits in comparison to the outcomes of Otsu-Thresholding and Fuzzy C-Means. In real applications, fame is dealt with completely. This frame may contain more than 300 olives in one shot. Dealing with this amount of fruit in one shot, which moves at high speeds, to separate the largest possible amount of fruit in one hour is not an easy matter. Therefore, in this study, care was taken to deal with the frames and not to deal with the fruits separately. Table 3 displays a number of sample frames that were used in this study.

In this instance, the technique's correctness is determined by its capacity to remove and exhibit all of the fruits from the conveyor without any of them corroding, irrespective of how much the fruits resemble the background (i.e., not treating any portion of the fruits as a background). However, the method can conceal the carrier entirely and display no visible portions of it. The most successful of the three methods in terms of completely removing the background and leaving no trace of the fruits, as well as extracting all of the fruits without losing any of them, was found to be the Modified-K-Means approach based on the data gathered. Despite the attractive appearance of the fruits, the Fuzzy C-Means approach findings show that some background occurs and does not completely disappear, which could negatively affect the accuracy of the judgment. The problem was not solved by Otsu's method—quite the contrary. It was found that some of the objectives (olive fruits) disappeared in addition to the backdrop with the fruits appearing in it.

2) QUANTITATIVE EVALUATIONS

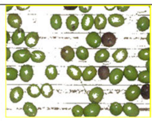
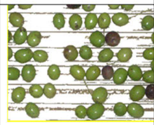
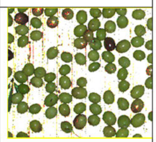
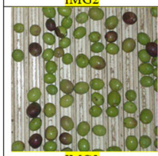
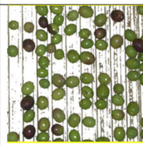
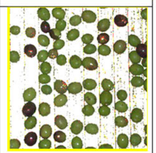
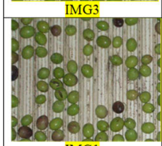
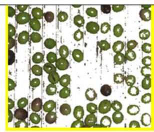
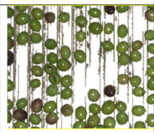
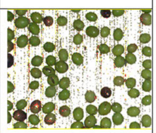
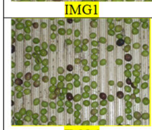
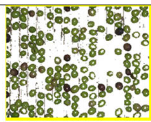
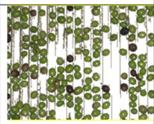
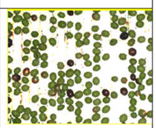
A meaningful study of the comparison between the suggested approach and other traditional methods is anticipated to be provided by quantitative analysis in terms of accuracy. The study that was presented used two different methods of quantitative analysis: confusion matrix and effectiveness measures, and dice similarity coefficient index. Keeping in mind that care was taken in selecting the pertinent techniques so that they were not slow techniques in data analysis, we review the findings of the quantitative evaluation using the aforementioned mechanisms in the following to determine the superiority of the technique used in separating olive fruits compared to other relevant techniques.

a: DICE SIMILARITY COEFFICIENT INDEX

The level of similarity among a planned segmentation mask and the ground truth segmentation mask can be assessed in the context of image segmentation using the Dice score. The Dice score goes from 0, which denotes no overlap, to 1, which denotes complete overlap. The following mathematical function can be used to calculate the dice similarity coefficient index:

$$d = 2 \times \frac{|R_{seg} \cap R_{gt}|}{|R_{seg}| + |R_{gt}|} \tag{4}$$

TABLE 4. Obtained results by Modified-K-Means approach fir olive fruits segmentation compared to the Otsu method, and Fuzzy C-Means.

Image Name	Otsu Clustering	Fuzzy C-means	Modified K-Means Clustring
 IMG5			
 IMG2			
 IMG3			
 IMG1			
 IMG6			

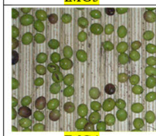
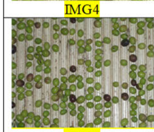
Where R_{seg} is the segmented result of the proposed algorithm and R_{gt} is manually segmented (ground truth segmentation mask).

The following table 5 displays the findings of an analysis of three fruit separation techniques using the Dice mechanism: the Otsu-Clustering algorithm, Fuzzy-C-means, and Modified-K-Means-Clustering. The results of evaluating three methods of fruit separation utilizing the Dice mechanism—Modified-K-Means-Clustering, Fuzzy-C-means, and Otsu-Clustering—are shown in the following table. The results demonstrate a significant advantage of the Modified-K-Means-Algorithm's efficiency over the others. The Modified-K-Means technique achieved an average efficiency of 0.958125, while the Fuzzy C-Means technique achieved an average efficiency of 0.6991, and the Otsu technique achieved an average efficiency of 0.520225.

b: CONFUSION MATRIX AND EFFECTIVENESS MEASURES

For quantitative categorization, the confusion matrix—also called the error matrix—is mostly utilized. Visualizing the efficiency of a method is made possible by a particular table arrangement. The actual value, or vice versa, is represented by the column in the matrix, whereas each row represents an instance of a forecast value. There are four cells in the final result matrices: false negative (FN), false positive (FP), true positive (TP), and true negative (TN). The following codes indicate different scenarios: TP denotes a positive value for

TABLE 5. Obtained results of dice coefficients of the proposed algorithm and the four conventional algorithms.

Original Image	Otsu Thresholding	Fuzzy C-Means	Modified-K-Means
 IMG1	0.5269	0.6420	0.9896
 IMG2	0.5944	0.6267	0.9493
 IMG3	0. 6450	0.7625	0. 9789
 IMG4	0.4727	0.7223	0.9240
 IMG5	0.4869	0.7420	0.9696
Average	0.520225	0.6991	0.958125

both the actual and predicted values; TN denotes a positive value for the truly but an unfavourable result for the model's prediction expected; FP denotes a negative value for the truly but a favourable model anticipated value; and FN denotes an unfavourable value for both the actual as well as the predicted values. In this instance, the frame is split into two regions: the stand surface region (background) was considered the negative zone, and the fruit portion of each frame was considered the positive region. Results of the Confusion Matrix of the Modified-K-means-Clustering, and Fuzzy C-Means algorithms are shown in Tables 6, and 7. The results confirmed the superiority of the Modified-K-Means-Clustering algorithm over the Fuzzy C-Means algorithm. The results show the efficiency of the proposed technique to distinguish between olives fruits and background (conveyer surface). It achieved an average efficiency of fruit detection of 96%, and an average efficiency of removing the conveyer surface reached 95. On the other side, although the Fuzzy C-Means can extract fruits with an efficiency of 87.5%, it finds it difficult to remove the background well, as it achieved an accuracy of only 70%. A metric known as an effectiveness measurement,

TABLE 6. Results of the confusion matrix of the modified-K-means-Clustering algorithm.

<table>
<tr><td rowspan="5">Predicted Value</td><td></td><td colspan="2">Actual Value</td><td>Total Accuracy</td></tr>
<tr><td rowspan="2">Positive</td><td>Accuracy of True Positive (TP)</td><td>Accuracy of false positive (FP)</td><td rowspan="2">100%</td></tr>
<tr><td>92%-99% Average = 96%</td><td>2%-8% Average = 4%</td></tr>
<tr><td rowspan="2">Negative</td><td>Accuracy of False Negative (FN)</td><td>Accuracy of True Negative (TN)</td><td rowspan="2">100%</td></tr>
<tr><td>2%-10% Average = 6 %</td><td>90%-98% Average = 94%</td></tr>
</table>

TABLE 7. Results of the confusion matrix of the fuzzy C-Means algorithm.

<table>
<tr><td rowspan="5">Predicted value</td><td></td><td colspan="2">Actual Value</td><td>Total Accuracy</td></tr>
<tr><td rowspan="2">Positive</td><td>Accuracy of True Positive (TP)</td><td>Accuracy of false positive (FP)</td><td rowspan="2">100%</td></tr>
<tr><td>80%-95% Average = 87.5%</td><td>5%-20% Average =12,5%</td></tr>
<tr><td rowspan="2">Negative</td><td>Accuracy of False Negative (FN)</td><td>Accuracy of True Negative (TN)</td><td rowspan="2">100%</td></tr>
<tr><td>24%-35% Average = 29 %</td><td>65%-76% Average = 71%</td></tr>
</table>

TABLE 8. Effectiveness of the proposed algorithm.

Statistical Test	*Formula*	*Result*
Accuracy	(TP+TN) / (TP+FP+FN+TN)	95%
Precision	TP / (TP+FP)	96%
Sensitivity	TP / (TP + FN)	94%
Specificity	TN / (FP + TN)	96%
Positive predictive value	TP / (TP + FP)	96%
Negative predictive value	TN / (FN + TN)	94%
False positive rate (α)	FP / (FP + TN)	4%
False negative rate (β)	FN / (TP + FN)	6%

or measures of effectiveness (MOEs), evaluates how well a system is able to satisfy the requirements of a particular scenario. Using the evaluation measure Effectiveness measure, Tables 8 and 9 show the test results of the Modified-K-Means-Clustering algorithm over the Fuzzy C-Means algorithm. As indicated in tables 7 and 8, the effectiveness measure incorporates a number of statistical tests, including sensitivity, accuracy, precision specificity, and others. The tables contain the results of the tests, the formula for each measure, and eight from the statistical tests. A great advantage is shown by the results in Tables 7 and 8 in favor of the proposed Modified-K-Means-Clustering technique compared to the Fuzzy-C-Means technique. When comparing the suggested Modified-K-Means-Clustering technique to the Fuzzy-C-Means technique, the results in Tables 7 and 8 demonstrate a clear benefit. In light of this, all of the evaluated data attest to the effectiveness of the suggested Modified-K-Means-Clustering methodology and its suitability for application in the proposed study as a method of removing olive fruits from the stand's surface.

B. RESULTS OF OLIVES FRUIT CLASSIFICATIONS

In this section, the results of training, validation and testing of the proposed technique for classifying olive fruits are reviewed. Given that the proposed study to conduct the classification work includes two stages: the first stage in which the SVM classifier is used to separate between healthy and defective fruits, the second stage in which the ANN classifier is used with two models, the first model is ANN1 to separate between the colors of healthy fruits, and the second is the model ANN2 which is used to separate between different types of fruits. Therefore, in the following, the results of training and testing for each model separately will be reviewed, then a review of the results of comparing the accuracy of the proposed technique compared to related techniques.

1) RESULTS OF THE FIRST PHASE OF THE PROPOSED CLASSIFIER

The performance analysis of training, validating, and testing the suggested SVM approach for dividing olive fruits into healthy and defective categories is described, along with the findings and discussion, in this subsection.

TABLE 9. Effectiveness of fuzzy c-means algorithm.

Statistical Test	*Formula*	*Result*
Accuracy	(TP+TN) / (TP+FP+FN+TN)	86%
Precision	TP / (TP+FP)	87%
Sensitivity	TP / (TP + FN)	84%
Specificity	TN / (FP + TN)	84%
Positive predictive value	TP / (TP + FP)	87%
Negative predictive value	TN / (FN + TN)	71%
False positive rate (α)	FP / (FP + TN)	15%
False negative rate (β)	FN / (TP + FN)	25%

TABLE 10. Results of training the first phase of the proposed classifier.

Actual classes	Predicted classes		
	Training Losses	Training Accuracy	Training Time (s)
Healthy	1.44%	98.56%	14.79
Defective	2.46%	97.54%	

a: RESULTS OF TRAINING THE FIRST PHASE OF THE PROPOSED CLASSIFIER (SUGGESTED SVM APPROACH)

During the initial stage of our suggested classification method for olive fruits, the suggested Support Vector Machine (SVM) model was trained on a total of 12,000 photos of olive fruits split into two folders: one with 9,000 healthy olives and the other with 3,000 damaged ones. Table 10 displays the training results of the suggested SVM. The results of training losses, training accuracy, and training duration for dividing fruits into two categories—healthy and defective—are shown in the table. The outcomes demonstrate the training's great accuracy in classifying fruits. The findings also demonstrate how quickly the classification was finished. On the other hand, it took less than fifteen seconds to finish the categorization task with this level of accuracy.

b: RESULTS OF VALIDATING THE FIRST PHASE OF THE PROPOSED CLASSIFIER

Validation was carried out for the initial classification step to ensure that there was no over-feting of the suggested categorization process. 3,000 olive fruits were used for validation; 2000 of the fruits were deemed healthy, while 1,000 were found to be flawed. According to the findings in table 11, there was little difference between the training and a validation set of results, which suggests a strong likelihood that over-feting does not happen.

TABLE 11. Results of validating the first phase of the proposed classifier.

Actual classes	Predicted classes		
	validating Losses	validating Accuracy	validating Time (s)
Healthy	2.52%	97.48%	4.58
Defective	3.62%	96.38%	

TABLE 12. Results of the overall accuracy of testing the first phase of the proposed algorithm.

Actual class	Predicted class				Class sensitivity %
	Class 1			Class 2	
Class 1	2000			0.	100
Class 2	20			980	98
Class precision %	99			100	Overall correctness= 99.5

2) CONFUSION MATRIX RESULTS FOR TESTING THE FIRST PHASE OF THE PROPOSED CLASSIFIER

Sensitivity and accuracy are two quantitative indicators of effectiveness used in categorization testing. Sensitivity is the ability of the prediction model to select an instance of a particular class from the available data. What counts is the proportion of true positive categories that are correctly recognized. On the other hand, the definition of correctness is the proportion of correctly identified expected positive classes. The following equations are used to calculate them:

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{5}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{6}$$

For the classes under discussion, the forecasts for true positive, false positive and false negative are represented by the symbols TP, FP, and FN, respectively. Table 12 displays the total testing accuracy of our suggested method for dividing about 3000 olive fruits into two groups. According to the data, the technique we proposed in the first phase of our plan can identify healthy olives with 100% accuracy; on the other hand, it can identify faulty olives with 98% accuracy, meaning that an overall correctness score is 99.5. This demonstrates the suggested technique's excellent accuracy in the initial categorization step.

TABLE 13. Training results of the ANN1 from the second phase of our proposed classifier.

Actual classes	Predicted classes		
	Training Losses	Training Accuracy	Training Time (s)
Healthy-Green	0.56%	99.44%	6.68
Healthy-Black	2.68%	97.32%	
Healthy-Purple	3.32%	96.68%	

TABLE 14. Validating results of the ANN1 from the second phase of our proposed classifier.

Actual classes	Predicted classes		
	Validating Losses	Validating Accuracy	Validating Time (s)
Healthy-Green	1.7%	98.32%	4.05
Healthy-Black	3.55%	96.45%	
Healthy-Purple	4.88%	95.12%	

3) RESULTS OF THE SECOND PHASE OF THE PROPOSED CLASSIFIER

This subsection describes the findings and discussion of the performance analysis of training, validating, and testing the proposed ANN1 and ANN2 approaches for classifying healthy olive fruits into three colors (Green, Black, and Purple) by ANN1, and the defective ones into five categories (Serpeta, Granizo, Rehu, Molestado, and Molestado2) by ANN2 as shown below.

a: RESULTS OF TRAINING THE ANN1 MODEL OF THE PROPOSED CLASSIFIER

In the second phase of our proposed olive fruit classification method, 6,000 images of healthy olive fruits were used to train the Artificial Neural Network 1 (ANN1) model. These images were divided into three folders, each containing 1,200 images of healthy olives in black, 1,200 images of healthy olives in purple and 3,600 images of healthy olives in green. The training results of the proposed ANN1 are shown in Table 13. The table displays the outcomes of training duration, training accuracy, and training losses for classifying fruits as either healthy or faulty. The results show how accurate the training is at classifying fruits. The findings reveal that the approach can be trained to achieve the accuracy displayed in table 13 in a relatively short amount of time. Achieving a 97.8% accuracy rate just required seven seconds of effort.

b: RESULTS OF VALIDATING THE ANN2 MODEL OF THE PROPOSED CLASSIFIER

In order to make sure that there was no over-feting of the recommended categorization procedure, validation was done for the ANN1 model in the second classification phase. Using 2,000 photos of olive fruits, the Artificial Neural Network1 (ANN1) model was verified. These photos were sorted into three files, each including 400 pictures of black olives, 400 pictures of purple olives, and 1200 pictures of green olives. The results in Table 14 showed minimal variation between the training and validation sets of data, indicating a high probability that over-feting does not occur.

TABLE 15. Results of the overall accuracy of testing the the ANN1 model in the second phase of the proposed classifier.

Actual class	Predicted class					Class sensitivity %
	Health y-green	Healthy -black			Healthy -purple	
Healthy -green	1200		0.		0	100
Healthy -black	0		39 0		10	97.5
Healthy -purple	0		12		388	97
Class precisio n %	100		97		97.4	Overall correctness = 98.16

c: CONFUSION MATRIX RESULTS FOR TESTING THE ANN1 MODEL IN THE SECOND PHASE OF THE PROPOSED CLASSIFIER

Table 13 presents the overall testing accuracy of our ANN1 in the second phase of our proposed approach for classifying around 2000 olive fruits into three groups (healthy-green, healthy-black, and healthy-purple), in the same manner as the results are shown in Table 15. The results show that our suggested ANN1 model can correctly identify healthy green olives with 100% accuracy, healthy black olives with 97.5% accuracy, and healthy purple olives with 97% accuracy. This results in an overall correctness score of 98.16. This illustrates the great accuracy of the proposed ANN1 approach in the second step of categorization.

d: RESULTS OF TRAINING THE ANN2 MODEL OF THE PROPOSED CLASSIFIER

Three thousand photos of faulty olive fruits were utilized to train the Artificial Neural Network 2 (ANN2) model in the second stage of our suggested olive fruit categorization technique. The 600 photos in each of the five folders—Serpeta, Granizo, Rehu, Molestado1, and Molestado2—represent the five different kinds of olive fruit faults. These photos were

TABLE 16. Training results of the ANN1 from the second phase of our proposed classifier.

Actual classes	Predicted classes		
	Training Losses	Training Accuracy	Training Time (s)
Serpeta	4%	96	5.48
Granizo	4%	96	
Rehu	5%	95	
Molestado1	4%	96	
Molestado2	3%	97	

TABLE 17. Validating results of the ANN2model in the second phase of our proposed classifier.

Actual classes	Predicted classes		
	Training Losses	Training Accuracy	Training Time (s)
Serpeta	4%	95%	3.38
Granizo	4%	96%	
Rehu	5%	96%	
Molestado1	4%	96%	
Molestado2	3%	96%	

TABLE 18. Results of the overall accuracy of testing the the ANN1 model in the second phase of the proposed classifier.

Actual class	Predicted class					Class sensitivity %
	Serpeta	Granizo	Rehu	Molestado1	Molestado2	
Serpeta	193	7	3.	0	0	96.5
Granizo	7	190	3	0	0	95
Rehu	3	2	195	0	0	97.5
Molestado1	0	0	0	192	8	96
Molestado2	0	0	0	5	195	97.5
Class precision %	95	95.5	97	97.5	96	Overall correctness= 96.5

organized into five categories. Table 16 displays the training outcomes of the suggested ANN2. Training losses, training accuracy, and training time data are displayed, as previously, to show the outcomes of training the ANN2 model. The results show how accurate the training is at classifying five different types of fruits. The findings reveal that the approach can be trained to achieve the accuracy displayed in table 15 in a relatively short amount of time. Achieving a 96% accuracy rate just required six seconds of effort.

e: RESULTS OF VALIDATING THE ANN2 MODEL OF THE PROPOSED CLASSIFIER

Similar to the prior instance, validation was carried out for the ANN2 model in the second classification phase to ensure that there was no over-feting of the advised categorization approach. An initial 1,000 images of olive fruits were used to validate the Artificial Neural Network1 (ANN2) model. These images were organized into five files representing the five categories of defective olive fruits, with 200 images each category. The findings, in Table 17, also revealed no difference across the data sets used for training and validation, suggesting a high likelihood that over-feting does not happen.

f: CONFUSION MATRIX RESULTS FOR TESTING THE ANN2 MODEL IN THE SECOND PHASE OF THE PROPOSED CLASSIFIER

The testing accuracy of our ANN2 in the second step of our suggested approach, which involves classifying about 1000 olive fruits into five faulty categories (Serpeta, Granizo, Rehu, Molestado1, and Molestado2), is shown in Table 17. Similarly to how the outcomes in Table 18 are displayed in Tables 12 and 15. The outcomes demonstrate that, with 96.5% accuracy, our proposed ANN2 model can successfully detect Serpeta olives, Granizo olives, Rehu olives, Molestado1 olives, with 96% accuracy, and 97.5% accuracy of Molestado2 olives. An overall accuracy score of 96.5 is obtained as a consequence. This proves that the suggested ANN2 method is incredibly accurate.

C. COMPRESSION BETWEEN ACCURACY TEST RESULTS OF THE PROPOSED ALGORITHM AND THE RELATED ALGORITHMS

This section presents a comparison of the findings from an efficiency test conducted on the suggested method with the related machine learning and deep learning techniques. These machine learning algorithms include DT, NB, KNN, LR, RF, and SVM methods. The related deep learning algorithms include VGG19, ResNet-50, MobileNetV2, InceptionV3, DenseNet201, EfficientNetB0). By applying the suggested technique over two sorts of images—one of which is an image of an olive fruit with a white backdrop, and the other is an image of the fruits while they are over the conveyor—a comparison is made to determine how effective it is in identifying eight different types of olive fruits when compared to previous techniques.

Comparison of the suggested algorithm's test accuracy results with those of related machine learning methods utilizing olive images with a Wightbackground

The obtained results of a comparison between the test accuracy of the suggested method and six from the related machine learning strategies to classify the olive fruits into eight different categories are shown in Table 19. This com-

TABLE 19. Comparison of the suggested algorithm's accuracy test results with those of related machine learning methods utilizing olive images on a Wight background.

Olive Classes	SVM	KNN	NB	DT	LR	RF	Proposed Model
Class 1	96	95	95	96	94	93	100
Class 2	94	94	94	94	93	92	100
Class 3	98	97	97	97	95	96	100
Class 4	89	87	87	86	83	86	99.33
Class 5	90	88	88	90	89	91	99.02
Class 6	91	89	86	91	90	92	98.2
Class 7	93	95	95	93	94	93	98.56
Class 8	96	96	96	96	94	95	99
Overall correctness=	92.37	91.89	91.18	92.13	91.5	92.25	99.26375

parison was made using 3000 images of the fruits on a white background. The results, which are displayed in Table 19, demonstrate how much the test accuracy of the suggested strategy differs from the outcomes of related machine learning approaches. It was observed that the planned method's accuracy was higher than 99% (99.26) than that of the other procedures, which varied from 91% to 92%.

1) COMPARISON OF THE SUGGESTED ALGORITHM'S ACCURACY TEST RESULTS WITH THAT OF THE RELATED MACHINE LEARNING METHODS USING OLIVE PHOTOS ON A CONVEYOR BELT (WITH A REAL BACKGROUND)

The capacity of the suggested technology to handle the fruits and categorize them into eight different types while on the conveyor is a huge advantage over related machine learning technologies. This is what Table 20's results indicate. 6,000 olive fruits were evaluated while they were on the carrier to determine how accurate each technology was—both the suggested way and the others. In comparison to other strategies that ranged in accuracy from77.25% of the LR technique to 83.75% of the SVM technique, the suggested technique attained an accuracy of.97.25%. These findings, as shown in Table 20, highlight the usefulness of the suggested method due to its exceptional capacity to handle fruits while they are on the stand—a feat that other methods find challenging.

2) COMPARISON OF THE SUGGESTED ALGORITHM'S TEST ACCURACY RESULTS WITH THOSE OF RELATED DEEP LEARNING ALGORITHMS UTILIZING OLIVE IMAGES WITH A WIGHT BACKGROUND

Table 21 displays the findings of a comparison between the test accuracy of the recommended approach and six of the relevant deep-learning strategies for categorizing olive fruits into eight distinct groups. Three thousand pictures of the fruits against a white background were used to create this comparison. The findings, which are presented in Table 21, show how much the recommended strategy's test accuracy deviates from related machine learning methodologies' results. Concerning accuracy, it was found that the planned approach outperformed the other methods by more than 99% (99.26), ranging from 93.625% of the MobileNetV2 algorithm to 97.375% of the InceptionV3 algorithm.

TABLE 20. Compression between accuracy test results of the proposed algorithm and the related algorithms.

Olive Classes	SVM	KNN	NB	DT	LR	RF	Proposed Model
Class 1	87	86	82	81	80	82	99.2
Class 2	85	85	81	80	81	79	98.8
Class 3	86	86	84	79	78	82	98
Class 4	84	83	77	76	75	78	96
Class 5	79	78	78	78	74	76	96
Class 6	81	78	76	74	78	81	95
Class 7	85	84	85	76	77	79	97
Class 8	83	85	76	79	75	82	98
Overall correctness=	83.75	83.125	79.875	77.875	77.25	79.87	**97.25**

TABLE 21. Comparison of the suggested algorithm's accuracy test results with those of related deep learning methods utilizing olive images on a wight background.

Olive Fruits Classes	VGG19	Resnet-50	MobileNetV2	DenseNet201	EfficientNetB0	InceptionV3	**Proposed Algorithm**
Class 1	100	97	95	96	94	100	100
Class 2	100	94	93	93	91	100	100
Class 3	98	95	91	96	97	97	100
Class 4	96	93	93	91	93	94	99.33
Class 5	95	94	95	95	94	98	99.02
Class 6	96	95	94	98	97	97	98.2
Class 7	96	94	95	97	95	98	98.56
Class 8	95	95	93	94	94	95	99
Overall correctness=	97	94.625	93.625	95.250	94.375	97.375	99.263

TABLE 22. Compression between accuracy test results of the proposed algorithm and the related algorithms.

Olive Fruits Classes	VGG19	Res net-50	MobileNetV2	DenseNet201	EfficientNetB0	InceptionV3	**Proposed Algorithm**
Class 1	98	92	91	94	94	97	99.2
Class 2	98	91	90	92	91	98	98.8
Class 3	95	93	92	93	94	97	98
Class 4	92	90	92	91	91	93	96
Class 5	95	91	93	90	90	92	96
Class 6	92	93	90	95	94	93	95
Class 7	94	94	92	93	92	94	97
Class 8	94	92	89	90	91	93	98
Overall correctness=	94.75	92	91.125	92.25	92.125	94.625	**97.25**

3) COMPARISON OF THE SUGGESTED ALGORITHM'S ACCURACY TEST RESULTS WITH THAT OF THE RELATED DEEP LEARNING ALGORITHMS USING OLIVE PHOTOS ON A CONVEYOR BELT (WITH A REAL BACKGROUND)

The proposed system has a significant advantage over comparable technologies in that it can handle the fruits and classify them into eight different types while they are on the conveyor. The findings presented in Table 22 suggest this. 6,000 olive fruits were assessed while they were in the carrier to find the accuracy of each technology, including the one that was recommended. The suggested strategy achieved an accuracy of 97.25 percent, whereas alternative techniques ranged in accuracy from 91.125% of the MobileNetV2 technique to 94.75% of the VGG19 technique. These results, which are displayed in Table 20, demonstrate the value of the recommended approach since it can handle fruits while they are on the stand, which is a difficult task for other methods.

V. CONCLUSION AND FUTURE LINES

The proposed work in this study dealt with a new strategy for classifying the colors and diseases of olive fruits. The proposed strategy based on firstly, good extraction of fruits from the background, secondly, separation of olive fruits into eight types according to color and defect. The modified K-Means algorithm was used to extract fruits from the background. The results of proposed fruit extraction were reviewed using different evaluation methods. The efficiency and accuracy of the proposed proposal for extracting fruits was confirmed by comparing it with the results of relevant techniques. The fruit separation process was divided into two stages, according to the purpose of the separation process. The SVM algorithm was used to separate between three colors, and the ANN algorithm was relied upon to separate between five different defects. The results of the test accuracy of fruits classification were reviewed using different evaluation methods (machine learning and deep learning algorithms). The most important thing that distinguishes the proposed method from other methods presented in the previous publication for classifying olive fruits is that it dealt realistically, through the use of a powerful database consisting of a total of 15,000 images. The proposed technique achieved separation accuracy of 99.26% for images with a white background, and 97.25% for images with the background of the surface of the fruit carrier. The high-performance correctness of the suggested strategy, after training it on about 9000 samples of healthy and damaged olive fruit images (real frames) and then testing it on about 6000 samples, indicates the high possibility of using it in computer vision applications [3], [37]. All experiments in this paper were conducted using MATLAB (2016 b) on a computer running Windows 10 64-bit with an Intel 2.7 GHz processor and 8 GB of RAM. The future work of this study is to train the proposed technology to classify the largest number of olive fruits classes. The hardware implementation of this technology is also one of our interests.

VI. AVAILABILITY OF DATA AND MATERIALS

"not applicable"

VII. AUTHOR CONTRIBUTIONS

We declare that the manuscript entitled "Highly Efficient Machine Learning Approach for Automatic Disease and Color Classification of Olive Fruits" is original, has not been full or partly published before, and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order updating of authors listed in the manuscript has been approved.

We understand that the Corresponding Author is the sole contact for the editorial process. The corresponding author "Mohamed Mourad Mabrook" is responsible for communicating with the other authors about process, submissions of revisions, and final approval of proofs.

Nashaat Mohamed Hussain Hassan
Ahmed Abdelbaset Donkol
Mohamed Mourad Mabrook
Ahmed Mosaad Mabrouk

REFERENCES

[1] A. A. A. El-aziz, A. Darwish, D. Oliva, and A. E. Hassanien, "Machine learning for olive fruit diseases classification system," in *Proc. Int. Conf. Artif. Intell. Comput. Vis.* Marrakesh, Morocco: Springer, 2020, pp. 16–25.

[2] H. Ayaz, E. Rodríguez-Esparza, M. Ahmad, D. Oliva, M. Pérez-Cisneros, and R. Sarkar, "Classification of olive disease based on non-linear deep features," *Appl. Sci. J., MDPI, Appl. Sci.*, vol. 11, no. 14, pp. 1–13, 2021.

[3] P. Moallem, A. Serajoddin, and H. Pourghassem, "Computer vision-based apple grading for golden delicious apples based on surface features," *Inf. Process. Agricult.*, vol. 4, no. 1, pp. 33–40, Mar. 2017.

[4] F. Femling, A. Olsson, and F. Alonso-Fernandez, "Fruit and vegetable identification using machine learning for retail applications," in *Proc. 14th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Las Palmas de Gran Canaria, Spain, Nov. 2018, pp. 9–15, doi: 10.1109/SITIS.2018.00013.

[5] A. Kader, S. Sharif, P. Bhowmick, F. HaqueMim, and A. Y. Srizon, "Effective workflow for high-performance recognition of fruits using machine learning approaches," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 7, no. 2, pp. 1516–1521, Feb. 2020.

[6] A. Nosseir and S. E. A. Ahmed, "Automatic classification for fruits' types and identification of rotten ones using k-NN and SVM," *Int. J. online Biomed. Eng.*, vol. 15, no. 3, pp. 47–61, 2019.

[7] H. M. W. M. Hippola, D. P. WaduMesthri, R. M. T. P. Rajakaruna, L. Yasakethu, and M. Rajapaksha, "Machine learning based classification of ripening and decay stages of mango (Mangifera indica L.) cv. Tom EJC," in *Proc. 2nd Int. Conf. Image Process. Robot. (ICIPRob)*, Colombo, Sri Lanka, Mar. 2022, pp. 1–6, doi: 10.1109/ICIPRob54042.2022.9798722.

[8] D. Gerdan Koç and M. Vatandas, "Classification of some fruits using image processing and machine learning," *Turkish J. Agricult.-Food Sci. Technol.*, vol. 9, no. 12, pp. 2189–2196, Dec. 2021.

[9] A. A. A. El-aziz, A. Darwish, D. Oliva, and A. E. Hassanien, "Machine learning for apple fruit diseases classification system," in *Proc. Int. Conf. Artif. Intell. Comput. Vis.*, vol. 1153. Marrakesh, Morocco: Springer, 2020, pp. 16–25.

[10] K. Sabanci, "Different apple varieties classification using kNN and MLP algorithms," *Int. J. Intell. Syst. Appl. Eng.*, vol. 4, pp. 166–169, Dec. 2016.

[11] N. Chaithra, B. Madhu, K. Umamaheswari, and S. Balasubramanian, "Classification and regression tree (CART) algorithm for the prediction of ischemic heart disease," *Int. J. Future Gener. Commun. Netw.*, vol. 13, no. 4, pp. 196–206, 2020.

[12] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Commun.*, vol. 30, no. 2, pp. 169–190, May 2017.

[13] S. Wang and Q. Chen, "The study of multiple classes boosting classification method based on local similarity," *Algorithms*, vol. 14, no. 2, p. 37, Jan. 2021.

[14] D. A. McCarty, H. W. Kim, and H. K. Lee, "Evaluation of light gradient boosted machine learning technique in large scale land use and land cover classification," *Environments*, vol. 7, no. 10, p. 84, Oct. 2020.

[15] F. R. Lumbanraja, E. Fitri, A. Junaidi, and R. Prabowo, "Abstract classification using support vector machine algorithm," *J. Phys.*, vol. 1751, no. 1, pp. 1–7, 2021.

[16] A. Pande, M. Munot, R. Sreeemathy, and R. V. Bakare, "An efficient approach to fruit classification and grading using deep convolutional neural network," in *Proc. IEEE 5th Int. Conf. for Converg. Technol. (ICT)*, Bombay, India, Mar. 2019, pp. 1–7, doi: 10.1109/I2CT45611.2019.9033957.

[17] M. S. Hossain, M. Al-Hammadi, and G. Muhammad, "Automatic fruit classification using deep learning for industrial applications," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 1027–1034, Feb. 2019, doi: 10.1109/TII.2018.2875149.

[18] Z. M. Khaing, Y. Naung, and P. H. Htut, "Development of control system for fruit classification based on convolutional neural network," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, Jan. 2018, pp. 1805–1807.

[19] H. S. Gill, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and F. Alassery, "Fruit image classification using deep learning," *Comput., Mater. Continua*, vol. 71, no. 3, pp. 5135–5150, 2022, doi: 10.32604/cmc.2022.022809.

[20] J. L. Rojas-Aranda, J. I. Nunez-Varela, J. C. Cuevas-Tello, and G. Rangel-Ramirez, "Fruit classification for retail stores using deep learning," in *Proc. Mex. Conf. Pattern Recognit.*, Morelia, Mexico, 2020, pp. 3–13, doi: 10.1007/978-3-030-49076-8_1.

[21] H. Muresan and M. Oltean, "Fruit recognition from images using deep learning," *Acta Universitatis Sapientiae, Inf.*, vol. 10, no. 1, pp. 26–42, Aug. 2018.

[22] N. Ismail and O. A. Malik, "Real-time visual inspection system for grading fruits using computer vision and deep learning techniques," *Inf. Process. Agricult.*, vol. 9, no. 1, pp. 24–37, Mar. 2022.

[23] S. Sakib, Z. Ashrafi, and M. Abu Bakr Siddique, "Implementation of fruits recognition classifier using convolutional neural network algorithm for observation of accuracies for various hidden layers," 2019, *arXiv:1904.00783*.

[24] P. Cano Marchal, S. Satorres Martínez, J. Gómez Ortega, and J. Gámez García, "Automatic system for the detection of defects on olive fruits in an oil mill," *Appl. Sci.*, vol. 11, no. 17, p. 8167, Sep. 2021.

[25] S. Figorilli, S. Violino, L. Moscovini, L. Ortenzi, G. Salvucci, S. Vasta, F. Tocci, C. Costa, P. Toscano, and F. Pallottino, "Olive fruit selection through AI algorithms and RGB imaging," *Foods*, vol. 11, no. 21, p. 3391, Oct. 2022.

[26] N. M. Hussain Hassan and A. A. Nashat, "New effective techniques for automatic detection and classification of external olive fruits defects based on image processing techniques," *Multidimensional Syst. Signal Process.*, vol. 30, no. 2, pp. 571–589, Apr. 2019.

[27] F. M. A. Mazen and A. A. Nashat, "Ripeness classification of bananas using an artificial neural network," *Arabian J. Sci. Eng.*, vol. 44, no. 8, pp. 6901–6910, Aug. 2019.

[28] J. M. Ponce, A. Aquino, and J. M. Andujar, "Olive-fruit variety classification by means of image processing and convolutional neural networks," *IEEE Access*, vol. 7, pp. 147629–147641, 2019, doi: 10.1109/ACCESS.2019.2947160.

[29] W. M. Thu and D. W. T. Tun, "GLCM and LTP based classification of food types," *Int. J. Sci. Eng. Appl.*, vol. 7, no. 8, p. 5, 2018.

[30] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.

[31] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proc. 4th ACM Int. Conf. Multimedia*, 1997, pp. 65–73.

[32] J. Gola, J. Webel, D. Britz, A. Guitar, T. Staudt, M. Winter, and F. Mücklich, "Objective microstructure classification by support vector machine (SVM) using a combination of morphological parameters and textural features for low carbon steels," *Comput. Mater. Sci.*, vol. 160, pp. 186–196, Apr. 2019.

[33] E. M. Lara-Espinoza, M. Trejo-Duran, R. A. Lizarraga-Morales, E. Cabal-Yepez, and N. Saldana-Robles, "Determination of the ripeness state of guavas using an artificial neural network," *Res. Comput. Sci.*, vol. 121, no. 1, pp. 105–111, Dec. 2016.

[34] L. R. Sekar, N. Ambika, V. Divya, and T. Kowsaly, "Fruit classification system using computer vision: A review," *Int. J. Trend Res. Develop. (IJTRD)*, vol. 5, pp. 22–26, Feb. 2018.

[35] M. Horea and M. Oltean, "Fruit recognition from images using deep learning," *Comput. Vis. Pattern Recognit.*, vol. 17, no. 12, pp. 576–580, 2017.

[36] A. A. Nashat and N. M. H. Hassan, "Automatic segmentation and classification of olive fruits batches based on discrete wavelet transform and visual perceptual texture features," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 1, Jan. 2018, Art. no. 1850003.

[37] S. Naik and B. Patel, "Machine vision based fruit classification and grading—A review," *Int. J. Comput. Appl.*, vol. 170, no. 9, pp. 22–34, Jul. 2017.

NASHAAT M. HUSSAIN HASSAN was born in Quena, Egypt, in 1977. He received the B.Sc. degree in communication and electronics engineering from Al-Azhar University, Egypt, in 2002, and the M.Sc. degree in communication and electronics engineering and the Ph.D. degree in digital integrated circuit design for the applications of image processing from the National Center of Microelectronics (CNM), Seville University, Spain, in 2005 and 2009, respectively. In October 2019, he was promoted to an Associate Professor. Currently, he is an Associate Professor with the Department of Electronics and Electrical Communication, Faculty of Engineering, Fayoum University, Egypt. His research interests include algorithms development (analysis, design and improvement) and full-cycle software and hardware product development (MATLAB, C, C++, VHDL, FPGA, and Xilinx). For more information visit the link: nmh01@fayoum.edu.eg.

A. A. DONKOL received the B.S.E.E. degree in electrical and electronics engineering from Assiut University, Egypt, in 2010, the Diploma degree (nine months) in cloud administrator track from the Information Technology Institute (ITI), Ministry of Communications and Information Technology (MCIT), in 2011, and the pre-master's degree. He is currently a Lecturer with the Electrical Engineering, Communication and Electronics Department, Faculty of Engineering, South Valley University (SVU), Egypt. He is a part-time with Nahda University in Beni Suef (NUB), Egypt. His research interests include machine learning, biomedical and genomic signal processing, speech processing, data compression, wavelet-transforms, and cloud computing. For more information visit the link: ahmed.donkol@nub.edu.eg and ahmed.donkol@eng.svu.edu.eg.

M. MOURAD MABROOK received the B.Sc. and M.Sc. degrees in electrical and communication engineering from Assiut University, Egypt, in 2008 and 2013, respectively, and the Ph.D. degree from the Communication Department, Menia University, Egypt, in 2017.

He is currently an Associate Professor with the Space communication Engineering Department, Faculty of Navigation Science and Space Technology (NSST), Beni-Suef University, Beni Suef, Egypt. He was the Head of Applied Sciences of Space and Navigation Program Science, in 2021. He is a part-time Lecturer with the Faculty of Engineering, Nahda University in Beni Suef (NUB), Beni-Suef. His research interests include wireless communications, 5G, 6G networks, cognitive radio, artificial intelligence, and sensors. He is a reviewer in many Elsevier and Springer publishers' journals. For more information visit the link: mohamedmourad2008@gmail.com and mohamed.mourad@nsst.bsu.edu.eg.

A. M. MABROUK was born in Elmofya, Egypt, in 1988. He received the B.Sc. and M.Sc. degrees in electronics and electrical communications from Menoufia University, Egypt, in 2011 and 2019, respectively, and the Ph.D. degree in electronics and electrical communications from Minia University, Egypt, in 2021. He was an Assistant Professor with the Department of Electronics and Electrical Communication, Faculty of Engineering and Technology, Badr University in Cairo (BUC), Egypt. Currently, he is an Assistant Professor with the Department of Electronics and Electrical Communication, Faculty of Engineering, King Salman International University (KSIU). His research interests include electronics, communications, antenna, and microwaves. For more information visit the link: eng_amosaad44@yahoo.com and ahmed.mosaad@ksiu.edu.eg.