VOL. 18, NO. 16, AUGUST 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

DATE FRUIT CLASSIFICATION USING DEEP LEARNING CONVOLUTIONAL NEURAL NETWORK

Abdulkarem H. M. Almawgani, Alhassan A. Alharazi, Naif A. Alalhareth, Saleh Mo. Alyami and Mohsen Q. Mohammed
Department of Electrical Engineering, College of Engineering, Najran University, Najran, Kingdom of Saudi Arabia
E-Mail: ahalmwgani@nu.edu.sa

ABSTRACT

A visual system for classifying different types of dates is extremely important for the buyer and the customer. Practical success in this field is still limited, and better outcomes are essentially required. This paper aims to classify all kinds of date fruits to help the customer and the buyer identify these types. The framework of this paper consists of a single model for the classification of date fruits, which is based on using the CNN method. This method is utilized with the transfer of learning and fine-tuning depending on the previous models and new models. To build an accurate and robust visual system, a rich image data set of the date fruit varieties is generated. The dataset contains significant differences, including differences in angles, dimensions, and lighting. In this paper, the CNN method is used to identify and classify various types of date fruits. Using this method produced high-accuracy results and a few errors. In this work, the database includes nine types of dates, whereby only six types were taken for training. High accuracy results were obtained from training these types of dates. The achieved accuracy of the results is 99%.

Keywords: date fruit; convolutional neural network CNN Database; dataset; classification of date fruits.

Manuscript Received 16 March 2023; Revised 17 October 2023; Published 27 October 2023

1. INTRODUCTION

The date fruit has been a primary source of food in the Middle East, North Africa, and particularly in the Arabian Peninsula. Dates are cheap to produce and preserve and can also be considered very nutritious. There is a total number of 6,924,975 date palm trees, as estimated in 2005. Worldwide date fruit production will continue to grow, especially in the Middle East [1]. This huge number shows the demand for dates worldwide. Moreover, there are more than 200 varieties of dates with different shapes, colors, and textures. Statistical data highlight the market value of date palms worldwide from 2014 to 2023. In 2018, for example, the global dates market was valued at about 13 billion U.S. dollars and is forecasted to reach 18.2 billion U.S. dollars by 2023 [2]. This enormous amount of date fruits with all varieties should be easily classified for manufacturers and consumers. This can be achieved through image classification methods.

Computer vision has significantly developed over the past few years, whereby there are many applications in image processing and image classification. Image processing proved to be an effective tool for analyzing various fields and applications. In the agricultural sector, these parameters like canopy, yield, and quality of product are important measures from the farmers' point of view. On the other hand, experts' advice in many cases may not be affordable and the availability of experts with their services may consume time. Image processing along with the availability of communication networks can change the situation of getting advice from experts more effectively, timely, and at affordable costs since image processing represents an effective tool for analyzing parameters [3].

The authors [4] investigated the design and implementation of a computer vision-based date grading and sorting system on real-life data. The system consists of two modules: 1) the image processing module and 2) the pattern recognition module. A binarization threshold is estimated from the image intensity histogram. The computer vision system captures the image, and then presents it in a pattern to assess the quality of the dates and classify them according to features, such as flabbiness, size, shape, intensity, and defects. Then, the Sobel edge operator is applied to extract the edges that surround the binarized regions.

Two Back Propagation Neural Network (BPNN) models are used for classification. Classifying different types of dates using computer vision without the need for complex physical measurements can be done [5] by using a total of 140 images taken for seven date types evenly divided. The size of the images is (480x640). Segmentation is carried out for the date fruit regions after implementing a thresholding method to separate the date fruit from the rest of the image. To represent each image, 15 features were used, including color-related, shape and size-related, and texture-related features. The histogram was filtered with a one-dimensional finite impulse response filter. Then, three classification methods were used, including K-nearest neighbor, linear discriminant analysis, and neural networks. The neural network was trained with 98 images, tested, and validated with 21 images for each. Artificial Neural Network (ANN) obtained the highest accuracy, while K-nearest neighbor and linear discriminant analysis produced approximately the same results. The authors [6] have proposed a new color space for grading the date fruit in the maturity stages. The objective of the work involved grading dates

and separating immature dates from the rest. The sorting was performed by machine vision, 100 images were taken for each maturity stage of one date type and the size of the images was reduced to (300x300). The features of color and texture were extracted by using the image processing toolbox of MATLAB software. After that, a statistical method in the Statistical Package for Social Sciences (SPSS) software was used to analyze each parameter and a taxonomy method was used to compute the coefficients of each component. High accuracy results were achieved. However, the accuracy of the Rotab stage was insufficient due to the feature extraction within the image.

The authors [7] classified dates (i.e., Khalas Variety) into three color categories and four size categories using a computer vision system based on image processing for date fruit classification according to color and size. The system can classify dates into 10 grades according to color and size. All the samples of the experiment are of Khalas variety, and they were obtained from the Date Palm Research Center of Excellence. A total of 200 images size of (640x480), including 2 images for each sample with static and various positions were used in the experiment. Half of the images were used for the database and the rest were used for testing. Since the color of the date ranges from yellow to dark red depending on the date's maturity level. Segmentation for the fruit color was developed using MATLAB code, making the program's main function to calculate the histogram. The date fruits based on size were classified into four categories, jumbo, big, medium, and small. The MATLAB algorithm "graythresh" is implemented to calculate the area covered by the fruit image to estimate the size. Also, different image processing techniques were evaluated to extract different features from the images of date. The prototype for sorting dates' time processing is 0.3 seconds for one fruit.

Furthermore, an automatic system, which classifies different types of dates from their images is used [8]. Features, including color, texture, and shape are used to recognize and distinguish the dates. Four types of dates are selected using 800 images: 200 images for each type of the four types. There was no restriction on the size of the date. The images were taken with a uniform (non-texture) color background, and the image color is passed through RGP (Red Green Blue) color filters. Color spaces RGP, YCbCr, and HSI (i.e., hue saturation and intensity) were investigated. YCbCr color space stores the color in terms of luminance and chrominance. HSI attributes are the closest approximation to human interpretation of color. By using the image histogram, a copy of the image color is converted into a black-and-white image, and the minimum and maximum vertical and horizontal coordinates of the black image are identified. Then, the best least-square fitting ellipse method is used to fit the image into an ellipse to define the shape and the size of the examined dates by using texture descriptors, Local Binary Patterns (LBP) and Weber, and Local Descriptor (WLD). After that, to select the important features from LBP or WLD, Fisher Discriminant Ratio (FDR) is applied. Finally, a Support Vector Machine SVM is created during the training phase so that the vectors can be used to classify the samples during the testing phase. Radial Basis Function (RBF) kernel is also added to help SVM. The highest accuracy was obtained. Both LBP and WLD Texture descriptors were better compared with shape and size features. However, WLD outperformed LBP. YCbCr and HSI color spaces provided more discriminative information than RGB color space.

Also, [9] proposed an automatic method based on machine learning techniques for date fruit classification. A total of 5,000 date fruit images were acquired with a resolution of (4,128 x3, 096) for ten varieties each. 4,000 images (400 per variety) and the remaining 1,000 images (100 per variety) were randomly selected and used for testing noting that some varieties contain some deformed dates. To describe the images of dates, RGB-based histogram, standard deviation, GLCM, and four geometrical features were used. Principal Component Analysis (PCA) is used to remove the correlation between different features and reduce their dimensionality. Overall high accuracy was obtained, and the negative effect of the outlier deformed samples on the classification results was reduced.

Researchers [10] proposed a method for automatically recognizing different date varieties. The presence of outlier samples could significantly degrade the recognition outcomes. Therefore, prune samples of each variety were separated from outliers using the Pruning Local Distance-based Outlier Factor (PLDOF) method. A total of 660 samples were collected from 11 varieties. The images, size of (4,128 x3,096), were divided into training and testing sets. The images were represented with RGB color space. For the texture descriptor, the Gray-Level Co-occurrence Matrix (GLCM) was used. To identify the shape, the principal axes method is used to calculate the minor and major axis length and eccentricity. The PLDOF method has pruned this set from those outliers. In the PLDOF, the samples were clustered, then outliers were removed from each cluster. Davies-Bouldin index (DBI) is used to automatically determine the optimal number of clusters. Mardia's multivariate skewness and kurtosis tests check if it is normally distributed. After that, to represent the visual model of the date's variety, the probability density functions that correspond to each cluster are combined in a Gaussian Mixture Model (GMM). The Expectation-Maximization (EM) algorithm is used for parameter estimation. For recognizing new samples, the probability that the sample belongs to each variety is calculated. With the inclusion of the outlier removal module, the average accuracy is better than without the outlier removal module's average accuracy. The proposed method achieved a high recognition rate, whereby it can distinguish between similar varieties.

The proposed method in [11] is based on machine learning technologies. Four types of dates are selected for the dataset for testing; however, the dataset is not balanced in each class (Ajwa: 105, Mabroom:92, Sagai:47, and Sukkary:81). The images were collected from Google pics. 70% of the dataset was for training, 20% for testing, and 10% of the data was used as validation data. Four

classification methods were used: 1) SVM, 2) Decision Tree Classifier (DT), 3) Random Forest Classifier (RF), and 4) Neural Networks (NN), and Python was used for implementation. Since the number of images is very low, the Neural Networks performance showed weakness. SVM achieved the best result and accuracy. Accordingly, the results, which were produced by the experiment, could be improved by obtaining more data.

The main objective in [12] is to develop an automated system, which can improve the dates industry. The system counts the number of dates, classifies their category, and recognizes defects if any. The main advantage is the ability to execute all the functions through one single system. First, the system counts the number of dates. Then, by using color and size features, the dates were classified to determine the maturity stage of the date. Finally, the system recognizes the defective dates by thermal images. A thermal infrared camera FLIER E5 is used for thermal images. The size of the images was reduced by MATLAB. RGB color space is converted to HSV (Hue, Saturation, Value) color space. The date color range is used to determine the maturity of the dates. In image segmentation, the pixels are divided into different sets to analyze and extract dates from the images. Thresholding is applied for segmentation. A labeling image function is used to check the connected component in a binary image. Labeling helps count the number of dates within one image. Those labeled dates are used to extract the desired features, such as the mean intensity, area, perimeter, and centroid as a structured array. The proposed system has successfully performed its functions.
A framework for the inspection of the external quality of the date fruits through the image processing method is proposed in [13] using the Bag of Feature (BOF) method, whereby the date fruits are classified into good quality and sugar-defected. The dates were of Khalas type. The size of the images is (459x332); 566 images in total were taken: 275 images for good quality dates and 291 images for sugar-defected dates, and 30% of the images for the training dataset and 70% for testing were randomly selected. The Grid method for griding the image and the Custom Speeded Up Robust Features (SURF) detector method were tested. SURF uses wavelet responses in horizontal and vertical directions. The histogram was built by counting every feature in the dataset using the approximate nearest neighbor. For classification, SVM was used. Significant features were obtained from the Bag of the feature classification algorithm. The method achieved 99% classification accuracy with 133 features.

The main contribution of [14] is to utilize the convolutional neural network CNN to build an automated sorting system for dates. A total of 1300 images of healthy and defective Shahani dates with 3 maturity stages were collected for the dataset. The examined stages are Khalal, Rutab, Tamar, and defective with 327, 288, 284, and 458 images, respectively. The size of the images is (150×150 RGB Image). For the network training process, 30,688 images for training and 6,368 for validation were utilized. Also, 199 images were used for the testing process. The utilized model structure has been modified by adding a classifier block instead of the fully connected classifier, version of the VGG-16 CNN. The modified classifier block included max-pooling, dropout, batch normalization, and dense layers. For the fully connected layer, the SoftMax activation function was applied. The CNN model was trained five times on the training dataset to increase the classification accuracy. For every training time, 25 epochs; one Epoch is when an entire dataset is passed forward and backward through the neural network only once, were used. After epoch 15, training the process was terminated to achieve the highest classification accuracy without any over-fitting. The classification accuracy of the training and validation dataset at the 15th epoch recorded 0.9794 and 0.9846. The CNN model obtained an overall classification accuracy rate of 96.98% and a deep CNN has overcome the complexity of the traditional systems.

2. METHODOLOGY

A Convolutional Neural Network (CNN) is utilized in this paper to classify the types of date fruits. The CNN type of neural network that has been created proved its effectiveness and robustness in various fields, such as image recognition and classification. This work aims to classify the types of dates and identify the name of each type according to several extracted features by CNN. In deep learning, CNN is a type of deep neural network, which is the most widely used in classifying and analyzing images. CNN is inspired by specific biological processes, as the pattern of communication between neurons is like the organization of the animal's visual cortex [15]. As for the deeper layers, CNN will be learned to distinguish or find simple features, such as edges, gradations, and variations of color, while the higher layers will merge simple features to form more complex features. Finally, the density layers at the top of the grid will incorporate the features into the plane very high and produce a forecast for the classification to be found [16]. In this study, the database is constructed from dates of various kinds. The database is then processed and programmed using MATLAB by the CNN method to classify the types of dates.

In this paper, dates of various kinds were photographed from different angles with a camera. Then, these pictures were saved to a computer. After that, the computer-saved images were processed and programmed using MATLAB by CNN to classify the different types of dates. In this study, the proposed framework, as shown in Figure-1 and Figure-2, of CNN consist of one classification model for the different types of dates. Every convolutional and fully connected layer was attached by ReLU as non-linear activation. Adding non-linearity using ReLUs helps CNNs train much faster. The first, second, and third convolutional layers are followed by both Local Response Normalization (LRN) and max-pooling layers, but only a max-pooling layer is used after the convolutional layer [16]. After that, the flattening process takes place and moves from the max pooling layers to the flat layers [17]. In the fully connected layer, the neural network will be fed, and this layer will be the last process

for classifying the types of date fruit. The classification stages are shown in Figure-1.

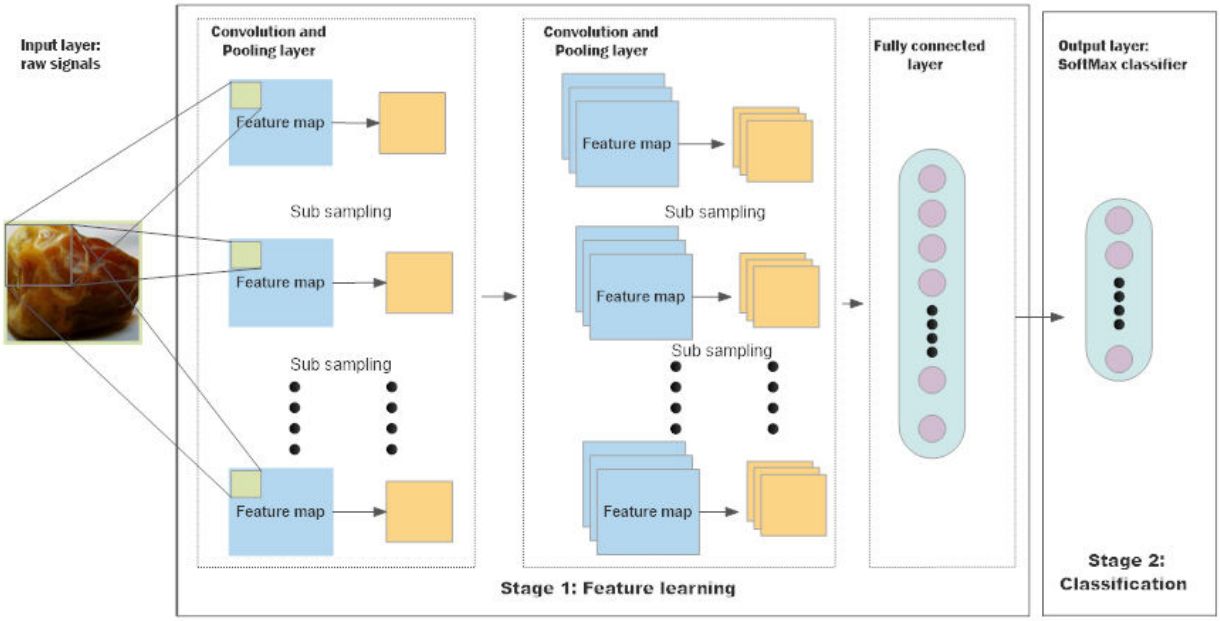

Figure-1. The CNN system.

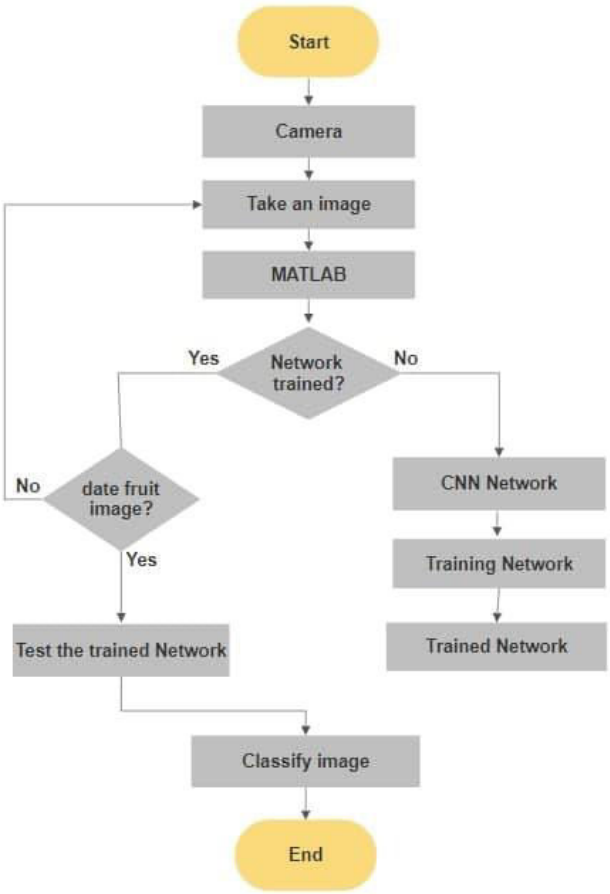

Figure-2. Flowchart of the proposed system.

The database contains 9 types of date fruits, as shown in Figure-3. Each type has 1000 images of 100 samples. Therefore, each sample has 10 images. The images were taken in five different backgrounds, as shown in Figure 6 and Figure-3 with different lighting conditions, as shown in Table-1.

Figure-3. Types of the date fruit.

Table-1. Database of the images' backgrounds and lighting conditions.

Type	Samples	Background of images					Lighting of sample			images
		White	Green	Woody	Sand	hand Palm	Sun lighting	Outdoor Lighting	Indoor lighting	
AJWA	100	200	200	200	200	200	25	25	50	1000
BARNI	100	200	200	200	200	200	25	25	50	1000
BAYADH	100	200	200	200	200	200	25	25	50	1000
KHLAS	100	200	200	200	200	200	25	25	50	1000
MAWAKIEL	100	200	200	200	200	200	25	25	50	1000
NABTAT_ALI	100	200	200	200	200	200	25	25	50	1000
RUTHANH	100	200	200	200	200	200	25	25	50	1000
SHISHI	100	200	200	200	200	200	25	25	50	1000
SOQUAI	100	200	200	200	200	200	25	25	50	1000
Total samples	900						Total images			9000

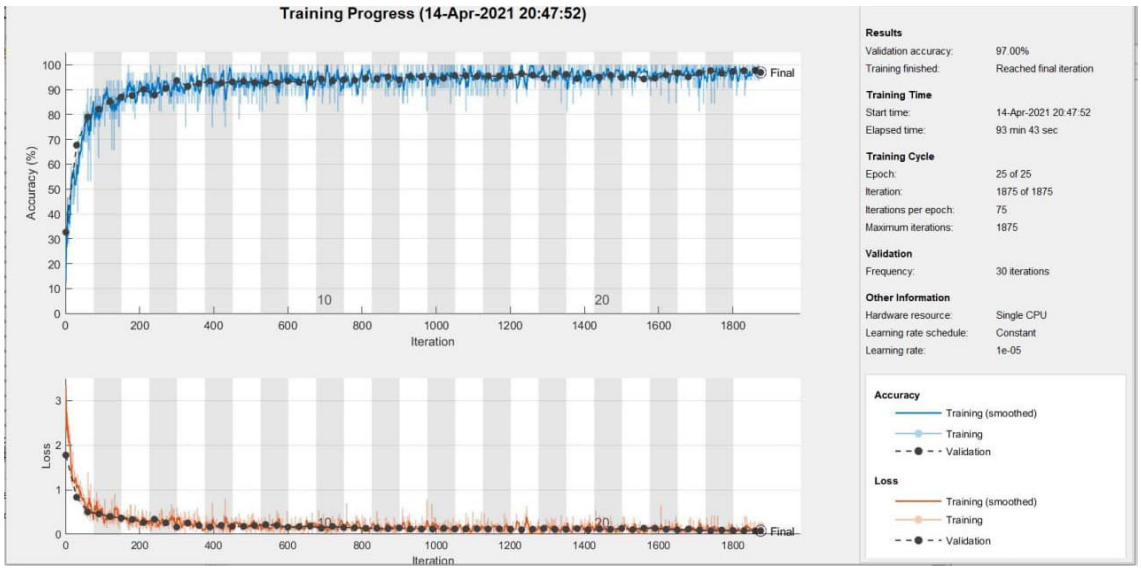

Figure-4. Results of training from the first computer.

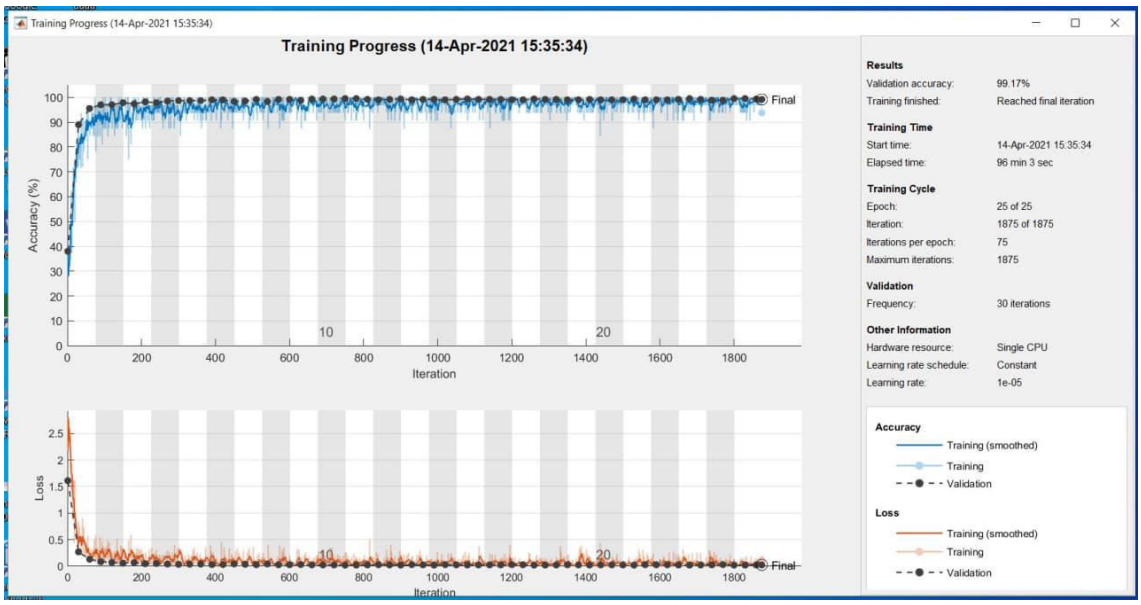

Figure-5. Results of training from the second computer.

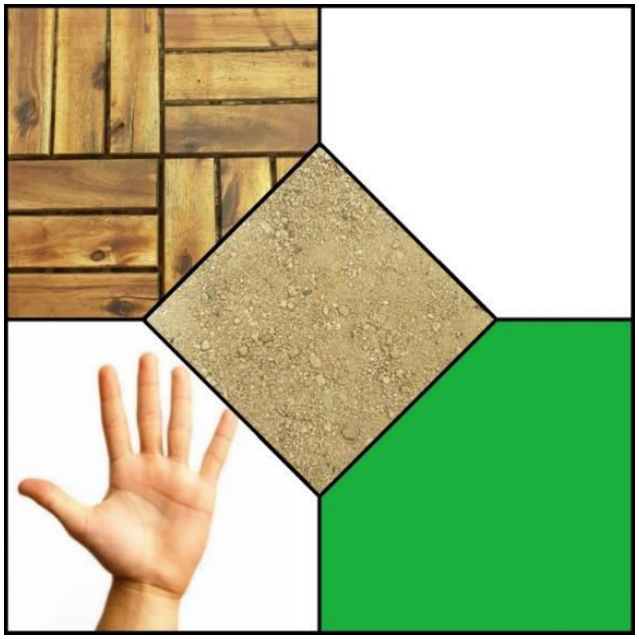

Figure-6. Different backgrounds of date fruit samples.

3. RESULTS AND DISCUSSIONS

3.1 Building the Network of this Paper

The previously trained network "AlexNet" has been used while replacing the last layers with new layers to change the old data with new ones to classify the date fruit types.

3.2 Training

Training the database on 25 epochs, a learning ratio of 1 * 10 ^ -6, with a frequency between 1000 and 2000. The database was trained on two separate devices, each containing three different types. The first training includes three types of dates, which are BARNI, BAYADH, and NABTAT-ALI. The second training includes three date fruit types, which are MAWAKIEL, AJWAH, and RUTHANA. Training in the first device achieved an accuracy of 97% and the accuracy of the second training reached 99%, as shown in Figure-4 and Figure-5.

3.3 Testing

A GUI was designed to facilitate the use of the program for users and to clarify the features of the program, as shown in Figure-7.

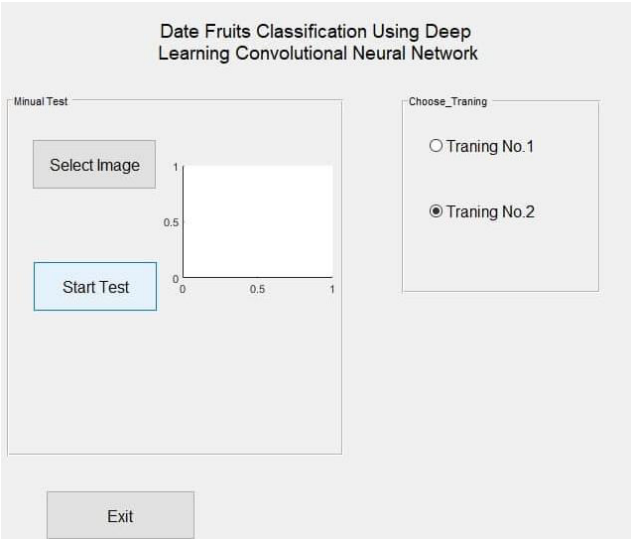

Figure-7. GUI design.

When the program starts, a ready-made image is selected from the computer, and when the sample image is selected, it is entered into the trained network to be classified as a date fruit or not, as shown in Figure-8.

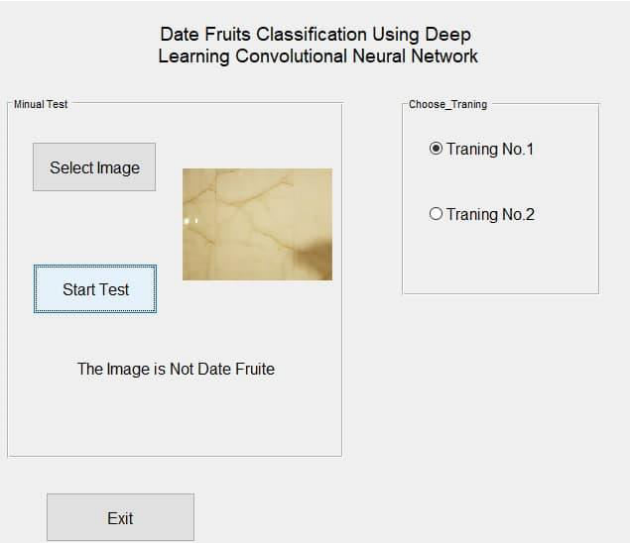

Figure-8. Result of the entered sample when it is not a date fruit type.

If the sample is a date fruit, it is entered into the network; another coach is to be categorized and the result of this rating appears in the GUI, as shown in Figure-9 and Figure-10.

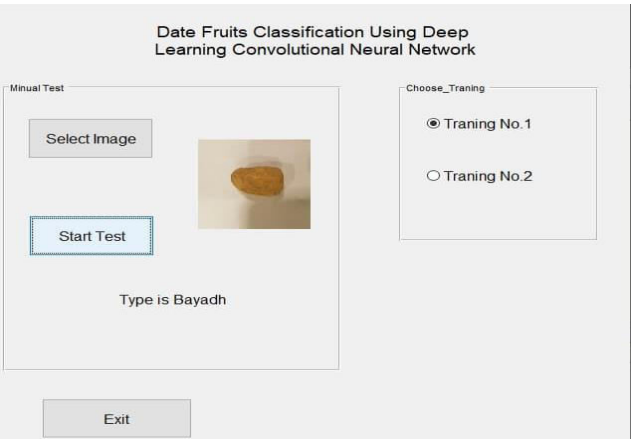

Figure-9. Result of the entered sample when it is not a date fruit type from the first computer.

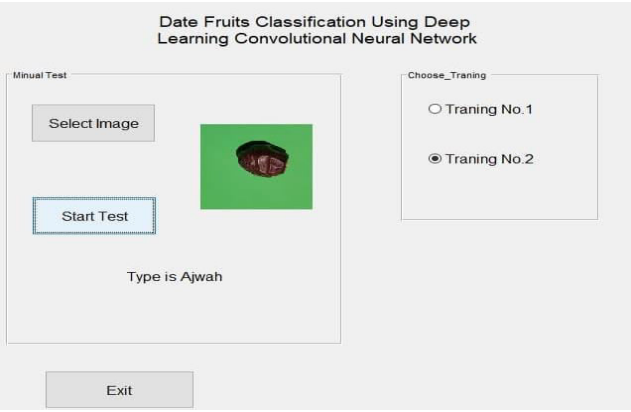

Figure-10. Result of the entered sample when it is not a date fruit type from the first computer.

The advantage of using the CNN method in reproducing the images' features of different types of date fruits is that the results are characterized with high accuracy. The results in this paper were positive in terms of training and testing samples. The accuracy of the results in the first training from the first computer was comparable to the second training in the second device. The experimental results for the first and second computers were repeated 50 times for each date type. The results showed that the proposed system is highly effective and efficient with an average detection percentage of 99%.

4. CONCLUSIONS

Dates are one of the most popular fruits in the Arab World. Dates have many types, and they are either different or similar shapes. Because of the large number of dates, it is difficult for the buyer to know their types. In this paper, a solution was orchestrated to classify the types of dates to make it easier for the buyer to distinguish between a wide array of different date varieties. About nine types of dates were classified and for each type, a hundred samples were taken and classified using the CNN method. The results from this method were good and with an accuracy of 99%.

4.1 Further Studies

There are several promising research works related to this paper:

- Developing a database by introducing multiple types and increasing the number of samples for each type.
- Improving the date fruit varieties with a high-end computer.
- Designing a program for mobile devices to facilitate using the system for the user.

ACKNOWLEDGMENT

The authors extend gratitude to the Deanship of Scientific Research at Najran University for funding this work under the Future Funding program grant code (NU/SRP/SERC/12/1).

REFERENCES

[1] C. C. T. Chao and R. R. Krueger. 2007. The date palm (Phoenix dactylifera L.): Overview of biology, uses, and cultivation. HortScience, 42(5): 1077-1082, doi: 10.21273/hortsci.42.5.1077.

[2] M. Shahbandeh. Market value of date palm industry worldwide from 2014 to 2023. Www.Statista.Com. https://www.statista.com/statistics/960213/date-palm-market-value-worldwide/.

[3] A. Vibhute and S. K. Bodhe. 2012. Applications of Image Processing in Agriculture: A Survey. Int. J. Comput. Appl., 52(2): 34-40, doi: 10.5120/8176-1495.

[4] Y. Al Ohali. 2011. Computer vision based date fruit grading system: Design and implementation. J. King Saud Univ. - Comput. Inf. Sci., 23(1): 29-36, doi: 10.1016/j.jksuci.2010.03.003.

[5] A. Haidar, H. Dong and N. Mavridis. 2012. Image-based date fruit classification. Int. Congr. Ultra Mod. Telecommun. Control Syst. Work., no. October, pp. 357-363, doi: 10.1109/ICUMT.2012.6459693.

[6] R. Pourdarbani, H. R. Ghassemzadeh, H. Seyedarabi, F. Z. Nahandi and M. M. Vahed. 2015. Study on an automatic sorting system for Date fruits. J. Saudi Soc. Agric. Sci., 14(1): 83-90, doi: 10.1016/j.jssas.2013.08.006.

[7] A. A. A. Ibrahim, A. H. Eissa and A. R. O. Alghannam. 2014. Image processing system for automated classification date fruit. International J. Adv. Res., no. June, pp. 1-55, [Online]. Available: https://www.researchgate.net/publication/270955882_IMAGE_PROCESSING_SYSTEM_FOR_AUTOMATED_CLASSIFICATION_DATE_FRUIT.

[8] G. Muhammad. 2015. Date fruits classification using texture descriptors and shape-size features. Eng. Appl. Artif. Intell., 37(April): 361-367, doi: 10.1016/j.engappai.2014.10.001.

[9] O. Aiadi and M. L. Kherfi. 2017. A new method for automatic date fruit classification. Int. J. Comput. Vis. Robot., 7(6): 692-711, doi: 10.1504/IJCVR.2017.087751.

[10] O. Aiadi, M. L. Kherfi and B. Khaldi. 2019. Automatic date fruit recognition using outlier detection techniques and Gaussian mixture models. Electron. Lett. Comput. Vis. Image Anal., 18(1): 52-75, doi: 10.5565/rev/elcvia.1041.

[11] A. A. Abi Sen, N. M. Bahbouh, A. B. Alkhodre, A. M. Aldhawi, F. A. Aldham and M. I. Aljabri. 2020. A classification algorithm for date fruits. Proc. 7th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2020, no. May, pp. 235-239, doi: 10.23919/INDIACom49435.2020.9083706.

[12] T. Najeeb and M. Safar. 2018. Dates maturity status and classification using image processing. 2018 Int. Conf. Comput. Sci. Eng. ICCSE 2018 - Proc., pp. 1-6, doi: 10.1109/ICCSE1.2018.8374209.

[13] A. Hakami and M. Arif. 2019. Automatic Inspection of the External Quality of the Date Fruit. Procedia Comput. Sci., 163: 70-77, doi: 10.1016/j.procs.2019.12.088.

[14] A. Nasiri, A. Taheri-Garavand and Y. D. Zhang. 2019. Image-based deep learning automated sorting of date fruit. Postharvest Biol. Technol., 153(January): 133-141, doi: 10.1016/j.postharvbio.2019.04.003.

[15] 2018. Convolutional Neural Network. Encyclopedia of Social Network Analysis and Mining.

[16] R. E. Neapolitan and X. Jiang. 2018. Neural Networks and Deep Learning.

[17] J. Koushik. 2022. Understanding Convolutional Neural Networks. WILDML, 2016. http://arxiv.org/abs/1605.09081 (accessed Jan. 12, 2022).