

Date Grading using Machine Learning Techniques on a Novel Dataset

Hafsa Raissouli¹, Abrar Ali Aljabri² Sarah Mohammed Aljudaibi³
Fazilah Haron⁴ and Ghada Alharbi⁵

College of Computer Science and Engineering, Taibah University, Medina, KSA^{1,2,3,5}
College of Computer and Cyber Sciences, Prince Muqrin University, Medina, KSA⁴

Abstract—Dates grading is a crucial stage in the dates' factories. However, it is done manually in most of the Middle Eastern industries. This study, using a novel dataset, identifies the suitable machine learning techniques to grade dates based on the image of the date. The dataset consists of three different types of dates, namely, Ajwah, Mabroom, and Sukkary with each having three different grades. The dates were obtained from Manafez company and graded by their experts. The color, size and texture of the dates are the features that have been considered in this work. To determine the color, we have used color properties in RGB (red, green, and blue) color space. For measuring the size, we applied the best least-square fitting ellipse. To analyze the texture, we used Weber local descriptor to distinguish between texture patterns. In order to identify the suitable grading classifier, we have experimented three approaches, namely, k-nearest neighbor (KNN), support vector machine (SVM) and convolutional neural network (CNN). Our experiments have shown that CNN is the best classifier with an accuracy of 98% for Ajwah, 99% for Mabroom, and 99% for Sukkary. Hence, the CNN classifier has been incorporated in our date grading system.

Keywords—Date grading; machine learning; k-nearest neighbor; support vector machine; convolutional neural network

I. INTRODUCTION

The world produces more than four million metric tons of dates annually [1]. Fig. 1 shows the top ten countries in dates production [2]. Each kind of dates (Ajwah, Safawi, Sukkary, etc.) has three different grades. For example, Ajwah has: Ajwah grade 1, Ajwah grade 2, and Ajwah grade 3. The first grade is known with the best quality.

In companies that treat and package dates, sorting and grading is a crucial stage. However, only a limited number of factories in the Middle East adopted automated grading machines to grade dates. The process of grading dates relies on many criteria. The most relevant one is the size of the date. Typically, the first grade comes with the largest size. The existing grading machines grade based on the size criterion. However, grading dates does not base on the size only. The Ministry of Environment Water and Agriculture published in KSA the standards of grading dates. The standards covered eleven kinds of dates (Burji, Khudri, Khalas, Raziz, Sukkary, Shishi, Safawi, Safari, Sagui, Ajwah, and Anbar). The criteria that are considered for grading each type to three grades are: weight, size (height and width), color, shape, humidity content, and sugar level.

According to Manafez international company, the grading machines that base on the size of the date to differentiate between grades have a low grading accuracy. In dates factories,

the grading is manually done by trained experts, which is time consuming. "We are still using manual grading. The main reason is that grading machines are expensive but with a low accuracy" a remark made by the CEO of Manafez International (AIWasel, W. 2018, personal communication, 1 October).

Besides the size of the date, the texture is an important feature. Generally, the first grade of dates, that is the highest quality, is less textured. The texture is an index of the water content of the date. The thirist grade, that is the lowest quality, usually has less water content and hence is drier and more textured. The color is also considered an important criterion in grading dates. Even though the color from grade to grade varies within a small range, considering the color as a feature is helpful.

There are tens of kinds of dates around the world. However, they are similar in shape and within a small range of colors. Therefore, intraclass dates grading presents a challenge. Dates that are of the same type, but different grades are not easy to recognize. Separating different grades manually is time consuming and requires experience and skills. Since this process is needed so frequently in industries, a date grading system will be of great use to reduce time and human intervention.

The purpose of this study is to identify suitable machine learning algorithm to grade dates using a novel dates dataset based on three key features (color, size, and texture). The machine learning algorithms of interest are; the k-nearest neighbor (KNN), support vector machine (SVM), and convolutional neural network (CNN). The rest of the paper is organized as follows: Section II presents a review on the existing work on both classification and grading of dates. Section III demonstrates the development framework. Section IV discusses the experimental settings used. Section V shows the results. A discussion of those results is presented in Section VI, and the conclusion is in Section VII.

II. LITERATURE REVIEW

Researches on dates grading and classification started almost two decades ago. Many studies in the literature have addressed the problem of classification of different types of dates [3] [4] [5] [6] [7] [8]. However, only a handful of researchers focused on dates grading [9] [10]). In the relevant work, intra-class classification is more focussed on types of fruits other than dates. In [11], the authors conducted an extensive review on the quality evaluation of different fruits and vegetables. In [12], the authors have used multilayer perceptron (MLP) and random forest (RF) to grade grapes

while in [13] they have adopted neural networks (NN) to grade apples. Other works include the use of SVM for mango grading [14], NN for tomatoes [15] and SVM for strawberries [16].

The earliest work on dates' research is in 2003 [3]. The authors proposed a system that classifies seven types of dates using multilayer perceptron (MLP) and a statistical method with 100 images for each date type. The most accurate model was the MLP based model. The features used are a combination of physical features (size, shape, texture) and the color feature of the dates. In [4], probabilistic neural networks (PNN) are applied to classify five types of dates. The research used a dataset that consists of 40 images per type. The obtained accuracy was 60% for Bomaan date type, 80% for Khalas, Lolo, and Berhi, and 100% for Fard. In [5], KNN, linear discriminant analysis (LDA) and back propagation neural networks (BPNN) are experimented to classify different types of dates using 140 images per type. The achieved accuracy was 99% using BPNN. The study in [6] used a dataset of 220 images per class. MLP was used with backpropagation (BP) and radial basis function (RBF) to classify dates according to the shape and colour of the date. They obtained the highest accuracy using the RBF, which was 91.1%. The author in [7] proposed a system to classify dates using SVM based on the colour, size, and texture of the dates. They compared two algorithms for texture extraction, namely, local binary patter (LBP) and Weber local descriptor (WLD). The WLD showed slightly better results. This research achieved an accuracy of 98%. In [8], a real time data analytic framework is proposed. It uses 5G technology to instantly analyse dates pictures and classify them using CNN. The input can be in a variety of types: single, rectangular box, round box, piled up, and wrapped. The dataset was extracted from google search engine and consists of 2000 images per class. The achieved accuracy was 99%.

There are only two studies that focused on date grading [9] [10]. The author in [9] used BPNN to grade dates into three grades according to the following features: flabbiness, size, shape, intensity, defects using 400 images per grade. This research achieved an accuracy of 80%. The study in [10] graded the dates into six intra classes (soft small, soft large, semi hard small, semi hard large, hard small, hard large) based on the shape and texture of the dates. A dataset of 960 images was used with SVM, KNN and LDA. The The KNN method provides the best performance with an accuracy of 96.45%. Table I summarizes the related work on both date grading and date classification.

In the light of the literature, two algorithms were used in texture analysis, namely, LBP and WLD. LBP was used in dates grading [10]. In [7], both LBP and WLD were experimented with dates classification and the results showed that WLD performed slightly better than LBP. As for the classifiers, NN was used in [9] [5]. NN performed better than both KNN and LDA in the latter. SVM was used in [10] [7]. SVM showed good results for dates classification [7], but not dates grading [10].

III. THE DEVELOPMENT FRAMEWORK

The approach is divided to two phases (see Fig. 2). The first phase starts by uploading images. The grayscale conversion of the image permits it to be thresholded in the next step.

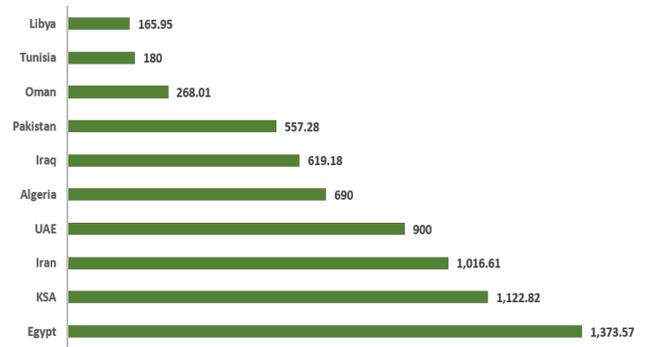


Fig. 1. Top Ten Countries in Dates Production 2019 (from EECO)

TABLE I. SUMMARY OF RELATED PAPERS ON DATES CLASSIFICATION AND GRADING

Author and year	Method	Accuracy	Features	Images per type
Hobani, et al. (2003) [3]	<ul style="list-style-type: none">• Statistical Method• MLP	<ul style="list-style-type: none">• 88.4% to 99.4%• 88.7% to 99.6%	<ul style="list-style-type: none">• Moisture• Volume• Weight• Size• Shape• Color	100
Fadel (2007) [4]	<ul style="list-style-type: none">• PNN	<ul style="list-style-type: none">• 60% to 100%	<ul style="list-style-type: none">• Color	40
Haidar, et al. (2012) [5]	<ul style="list-style-type: none">• KNN• LDA• NN	<ul style="list-style-type: none">• 90%• 96%• 99%	<ul style="list-style-type: none">• Color• Size• Shape• Texture	140
Alrajeh and Alzohairy (2012) [6]	<ul style="list-style-type: none">• MLP (BP)• MLP (RBF)	<ul style="list-style-type: none">• 87.5%• 91.1%	<ul style="list-style-type: none">• Shape• Color• Size	220
Muhammad (2015) [7]	<ul style="list-style-type: none">• SVM	<ul style="list-style-type: none">• 98.1%	<ul style="list-style-type: none">• Size• Texture	NA
Hossain, et al. (2018) [8]	<ul style="list-style-type: none">• CNN	<ul style="list-style-type: none">• 99%	<ul style="list-style-type: none">• Size• Texture	2000
Al Ohali (2011) [9]	<ul style="list-style-type: none">• BPNN	<ul style="list-style-type: none">• 80%	<ul style="list-style-type: none">• Flabbiness• Size• Shape• Intensity• Defect	1200
Mohana and Prabhakar (2014) [10]	<ul style="list-style-type: none">• KNN• SVM• LDA	<ul style="list-style-type: none">• 96.46%• 91.66%• 89.58%	<ul style="list-style-type: none">• Size• Shape• Texture	960

Thresholding the image is needed to apply the best least fitting ellipse for size measurements. The color identification and the texture analysis happen on the original images. Feature selection is then performed to reduce the high dimensionality of features by identifying the most relevant ones. Two feature selection methods are experimented: Fisher discriminant ratio (FDR) and sequential feature selection. The obtained feature set serves for KNN and SVM classification. While for CNN, only normalization of the images is done before passing the images to the classifier. In order to implement the actual grading system, the most accurate model out of the three is chosen for phase 2. In phase 2, which represents the grading system, the uploaded image is classified by the trained model that is integrated to the system. Results are saved to the database.

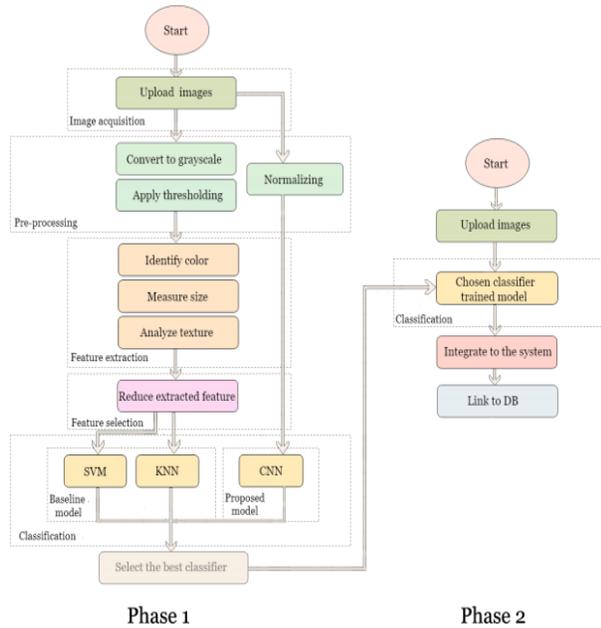


Fig. 2. The Date Grading Development Framework

IV. EXPERIMENTAL SETTINGS

A. Dataset

The dates that were used to construct the dataset are obtained from Manafez International company¹. Manafez is one of the main exporters of dates in Saudi Arabia. The company provided us with the grades of three types of dates, namely, Ajwah (grade 1, grade 2, and grade 3), Mabroom (grade 1, grade 2, and grade 3), Sukkary (grade 1 and grade 2 only as this type of dates usually comes in two grades). To build the dataset, a video of rotating dates placed on a dish in a lighted compartment was captured. The lighted compartment is illustrated in Fig. 3. A uniform distance of 20 centimeters was fixed with a relatively constant lighting. The distance and lighting were fixed according to the two following experimental observations; changing the distance affects the size of the date in such a way that images taken far from the date appear smaller (see Fig. 6). The size of the date is an important feature in grading. So, changing this parameter may lead to inaccurate results. For the lighting, when enough light is supplied, the details of the date, especially color and texture, are clearly visible. The experiments have shown that low lighting makes the dates appear similar. The wrinkles of the date are not clearly visible. In addition, the color feature is sensitive to the lighting, especially in grading, because the color variance is slight between grades. Fig. 5 illustrates these differences. After capturing the videos, the frames corresponding to the different sides of a date are extracted. The dataset produced, namely Taibah University-Dates Grading dataset (TU-DG), consists of the grades corresponding to three types of dates: Ajwa, Mabroom, and Sukkary. Samples of TU-DG dataset are shown in Fig. 4. Table II shows the number of images in each date type for all the grades in TU-DG dataset. TU-DG dataset is

¹<http://www.manafezinternational.com/>



Fig. 3. The Lighted White-Compartment and the Rotating Dish used to Capture the Dataset

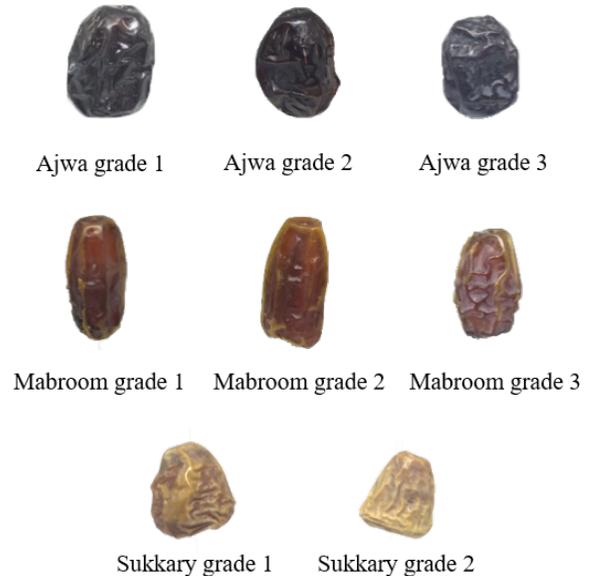


Fig. 4. Samples of the TU-DG Dataset

available online on the IEEE dataport.²

B. Preprocessing

The preprocessing stage converts the date image to a greyscale image and then threshold it to obtain the region of interest. This serves for applying the best least fitting ellipse to measure the size of the dates. When the thresholded image has some imperfections due to shadows, applying the best least fitting ellipse fails. The algorithm ends up taking ellipses for the date along with other parts of the images that were wrongly thresholded as part of the region of interest. Fig. 7 demonstrates that problem. In order to find a thresholding

²<https://iee-dataport.org/open-access/tu-dg-dataset-date-grading>
DOI: 10.21227/qhrr-m850

TABLE II. NUMBER OF IMAGES IN TU-DG DATASET FOR EACH DATE TYPE AND GRADE

Date type	#Grade 1	#Grade 2	#Grade 3	Total
Ajwah	300	345	492	1137
Mabroom	300	500	496	1296
Sukkary	492	458	-	950
Total	1092	1303	988	3383

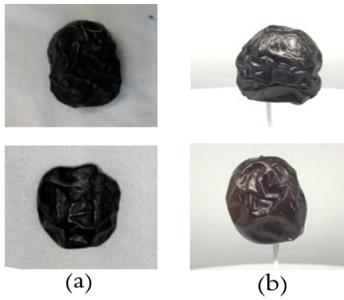


Fig. 5. Dates Captured in (a) Room Light, and (b) TU-DG Dataset

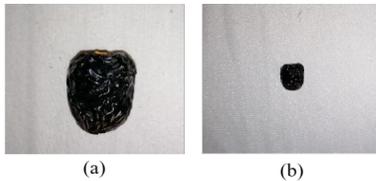


Fig. 6. A Date Captured with Different Distance, (a)10cm and (b) 30cm

algorithm that works for all the date types, six algorithms were experimented: setting a threshold value manually, weight of intensity difference, active contour, local thresholding, adaptive thresholding, and Otsu thresholding. The number of images that failed in thresholding varies from one method to another. Fig. 9 summarizes the performance of the six algorithms. From the figure, the percentage does not exceed 81% of images that were thresholded correctly for the first five algorithms (setting a threshold value manually, weight of intensity difference, active contour, local thresholding, and adaptive thresholding). Otsu thresholding has succeeded to threshold all the images of all date types. Yet, this type of thresholding resulted in some holes and imperfections in the thresholded image that are recoverable using postprocessing (morphological) techniques. Fig. 8 shows a thresholded image before and after postprocessing. The obtained image after Otsu thresholding shows some small dots (holes) on the white thresholded date. The Matlab built-in functions *imfill* and *bwareaopen* were used to perform the postprocessing. *imfill* fills the small dots while *bwareaopen* ignores any other dots that were not filled by *imfill* function.

C. Features

Most of the related work used the size of the date as in [9] [10]. This is because the size is important to distinguish different types of dates and also different grades. Grade 3 usually appears smaller than grade 1 and 2, while grade 1 and 2 have a slight difference in size. The color was used by many studies in dates classification in [3] [6] and in [10] in dates grading.

The texture was used in [10] for grading and many other studies for classification such as [5], [7], and [8]. The texture is an index of the water contents in the date. Usually grade 1 and 2 have more water contents and hence have less textured skin than grade 3. Some studies used other physical features such as weight, moisture and volume for classification [3]. These features involve manual measurements and are more time consuming. For that reason, the most relevant features

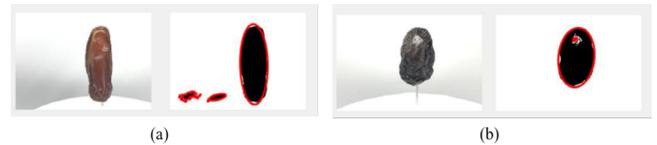


Fig. 7. Best Least Fitting Ellipse Taken for Wrong Areas: (a) Outside and (b) Inside the Date



Fig. 8. Thresholded Image using Otsu Thresholding (a) Before Postprocessing, and (b) After Postprocessing

are extracted in this research that are size, color, and texture. The size was extracted using the best least fitting ellipse. The color was extracted from the RGB color channels of the image. For texture, WLD [17] was used. These extracted features were stored in a matrix with 900 rows, which represents the number of input images, and 873 columns of extracted features. As demonstrated in Table III, 6 features were extracted for size, 771 for color, and 96 for texture. The number of features for each color channel is equal to the values of the channels' histogram bins which is 256. The intensity value for each color channel was also extracted. For the size, the parameters extracted are the length, width, area, perimeter, eccentricity, and equidiameter. WLD descriptor outputs a vector of 32 elements. This descriptor was used with three color channels, RGB, HSV, and YcbCr as in [7].

D. Feature Selection

Feature selection can improve the performance of the model by reducing the dimensionality of the feature set, two feature selection methods were experimented, FDR and sequential.

1) *FDR*: This method calculates the mean and variance of the variables to increase their separability according to the following equation:

$$F = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 - \sigma_2^2} \quad (1)$$

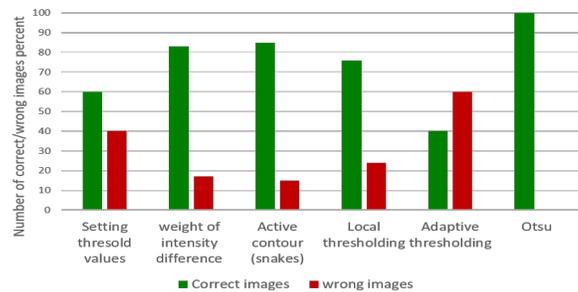


Fig. 9. Percentage of Correct and Wrong Thresholded Images using Different Thresholding Methods

TABLE III. NUMBER OF VALUES FOR EACH EXTRACTED FEATURE

Features	Method used	Description	#Features	Total
Color	Color histogram of RGB channels	Red channel	256	771
		Intensity of red	1	
		Green channel	256	
		Intensity of green	1	
		Blue channel	265	
		Intensity of blue	1	
Size	Best least fitting ellipse	Length	1	6
		Width	1	
		Area	11	
		Perimeter	1	
		Eccentricity	1	
		Equidiameter	1	
Texture	WLD	RGB	32	96
		HSV	32	
		YcbCr	32	
Total number of features			873	

Where μ is the mean and σ^2 is the variance. After performing Fisher discriminant ratio (FDR), the feature selection dimensionality is reduced to two dimensions.

2) *Sequential*: This selection method selects the best features in the feature set. It takes a number of features to select as an argument and returns a set of best features that is passed to the training process.

E. Classification

1) *Baseline Models*: A baseline model represents the simplest ways that can be used to obtain results for a given problem. It usually bases on the previous work and serves for comparison with the proposed model. The baseline models in this research are the KNN and SVM. In [7] SVM was used with dates classification and achieved an accuracy of 98%. In [10] KNN and SVM are compared for dates grading. The research shows that KNN gives better results than SVM. KNN and SVM are usually not time consuming in the training phase. They have also relatively simple structure which makes them good to be a baseline model. The baseline models were trained using 300 images per grade for each date type. The loss is estimated using a 10-fold validation for the model. The *FoldLoss* Matlab built-in function calculates the loss for each fold in the validation set and returns the model's loss (or error). This value is between zero and one (for simplicity and ease of comparison with the accuracy value, we state the loss in percentage). The model was tested using 100 images.

- *K-Nearest Neighbor*

KNN is a classifier that uses similarity to classify new instances. It considers the k closest points in the feature space to assign a class to a given instance. KNN involves several parameter categories such as the value of k (number of neighbors), the distance metric, and the search value. Two search methods are experimented namely, k-d tree and exhaustive search. For the distance metric seven distance metrics were experimented with the exhaustive search method (Minkowski, cityblock, Chebychev, correlation, cosine, Euclidean, and Hamming) and five distance metrics with k-d tree search method that are: Minkowski, Chebychev, cityblock, Euclidean.

- *Support Vector Machine*

Multi-class SVM was used with RBF kernel. The kernel is required to deal with non-linear problems. This method maps the data points to a higher dimensional space where the classes are linearly separable then converts the dimensions back to the original space. The RBF kernel has two parameters that are C and Gamma. Since those two parameters are not learnable, the Bayesian optimization was used for parameter tuning.

2) *Proposed Model*: The proposed model is CNN. This classifier requires relatively large dataset. In [18], a dataset of 14828 images was used for a nine class tomato disease classification. This classifier requires more training time compared with the baseline models. CNN has been proven to perform well with many image classification problems. Many studies such as [19] [20] [21] [22] [23] [24] have used CNN to classify fruits. In [25] and [26] CNN was used for vegetables classification. The authors in [18] used CNN for tomato disease classification. The research achieved an accuracy of 99.18%. The study in [27] used CNN in mango classification and achieved an accuracy of 99%. The authors in [28] classified flowers using CNN with an accuracy of 97%. Even though many of the relevant work that applied CNN used big datasets, there are some researches that achieved good results using small datasets. The study in [26] used a dataset of 50 images for cucumber classification task and achieved an accuracy of 96.08%. We have experimented CNN using TU-DG dataset that has 3383 images, and in order to maximize the use of this dataset, a 10-fold cross validation is applied.

3) *Evaluation Metrics*: Evaluating the result of a classification model using a suitable evaluation metric is important. It shows the performance of a given ML algorithm and eases comparison. Most related works on dates grading used accuracy as an evaluation metric. Since accuracy can be insufficient [29], the loss function is used to support the evaluation of the classification models.

A. Accuracy (ACC)

Accuracy equation is defined by equation 2

$$ACC = \frac{\text{Number of correct classifications}}{\text{Total number of images}} \quad (2)$$

Accuracy is usually extracted from confusion matrix according to the above equation. A confusion matrix is a table that shows the error of a classification based on the test set. Accuracy works well when the dataset is balanced.

B. Loss Function

The loss function is a method of calculating how well an algorithm models the data. Unlike accuracy that calculates the performance of the model in terms of the test set, the loss function is calculated for the trained model using the validation set. Hence, it is a form of prediction of how well the model will do. In this research, the validation set is divided into ten folds. The loss function calculates the misclassifications of each fold and returns a loss value between 0 and 1, where 0 represents no loss.

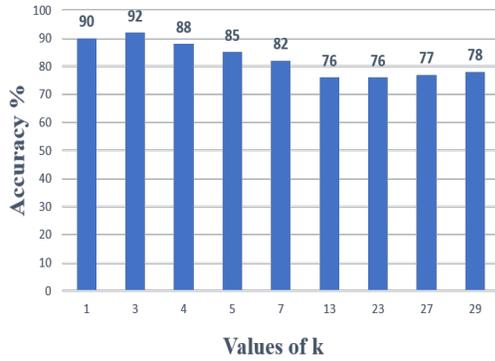


Fig. 10. Change of Accuracy with Different Values of k

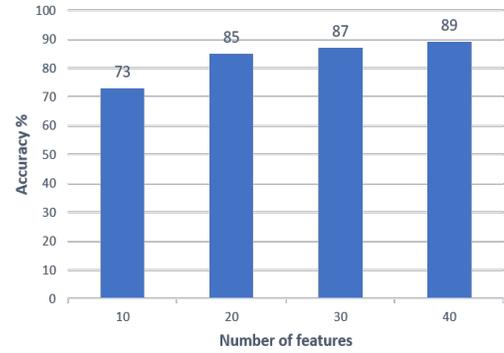


Fig. 12. The Effect of Different Number of Features on Accuracy

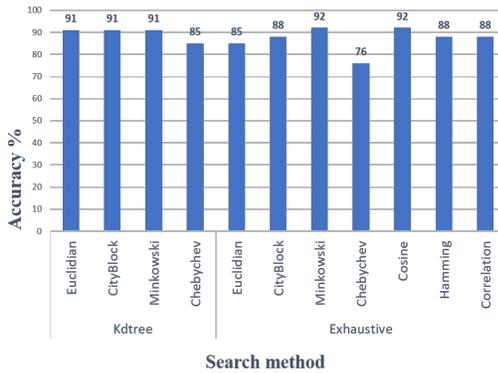


Fig. 11. K-d Tree and Exhaustive Search Methods with Different Distance Metrics

V. EXPERIMENTS AND RESULTS

A. Baseline Models Results

1) *K-Nearest Neighbor*: The experiments were conducted first with the raw features, that is, without feature selection. Fig. 10 and 11 show the accuracy variation with different values of k and different distance metrics using Ajwa. The results in the figures suggest that the best search method is exhaustive with either cosine or Minkowski distance metric and with k equal to 3. The same parameters worked the best with the other date types (Mabroom and Sukkary). The three types of dates were run in individual experimentations. The accuracy and loss of the KNN classification for the three types of dates is shown in Table VI.

- *FDR Feature Selection*

When applying the FDR feature selection, the numbers were mapped to complex numbers that cannot be used by the classifier. This problem was solved by cutting off the imaginary part of the numbers and keeping the real part only. Table IV shows the accuracy with and without feature selection. The accuracy has dropped almost half.

- *Sequential Feature Selection*

When applying the sequential feature selection, the accuracy increases using 20 features and then starts to plateau, indicating minimum benefit in increasing the number of features. Fig. 12 shows varying number of features and the resulting accuracy. Based on the

TABLE IV. ACCURACY WITHOUT AND WITH FDR FEATURE SELECTION

Date type	Accuracy before feature selection	Accuracy after feature selection
Ajwa	92.12%	72.11%
Mabroom	90.29%	67.38%
Sukkary	88.29%	52.14%

results, this feature selection method also has a negative effect on accuracy. According to the literature review, not every case needs a feature selection. This step can be omitted, and the extracted features can be passed directly to the classifier. This indicates that this problem performs better with raw features.

2) *Support Vector Machine*: The multiclass SVM was trained using RFB kernel and with default values of C and Gamma. The Bayesian optimization was then used to tune the parameters. For feature selection, the behavior of the model showed again that it performs better with the raw features. Table VI illustrated the obtained results.

From the results, we observe the following:

TABLE V. THE PERFORMANCE OF THE PROPOSED MODEL (CNN)

Date Type	Validation Accuracy	Test Accuracy
Ajwah	99%	98%
Mabroom	100%	99%
Sukkary	99%	99%

TABLE VI. THE PERFORMANCE OF THE BASELINE (KNN AND SVM) AND PROPOSED MODEL (CNN)

Date type	# training and validation	# testing images	correct classifications	Acc	Loss	
KNN	Ajwa	800	100	92	92%	9.88%
	Mabroom	800	100	90	90%	18%
	Sukkary	800	100	88	88%	17%
SVM	Ajwa	800	100	94	94%	10%
	Mabroom	800	100	97	87%	13%
	Sukkary	800	100	90	90%	11%
CNN	Ajwa	810	90	89	98%	2%
	Mabroom	918	102	101	99%	1%
	Sukkary	864	96	95	99%	1%

- The two baseline models, KNN and SVM, have the same accuracy with Ajwah and Sukkary.
- KNN performs better than SVM for Mabroom.
- Ajwah has the highest performance with both KNN and SVM.
- KNN and SVM results are roughly comparable for all the three date types. The performance of a classification method is highly affected by the quality of dataset supplied and the features extracted. The dataset and the methods used in feature extraction were the same for both KNN and SVM. This might be a reason for the comparability of the results of these two classifiers.

B. Results of Proposed Model

For the three date types, the model was trained with a 10-fold cross validation using six convolutional layers with batch normalization and maxpooling. The initial learning rate is set to 0.01 and stochastic gradient descent (SGD) is used for optimization. Table V shows the validation and test accuracy of the model. The results show that both Mabroom and Sukkary were classified with an accuracy of 99% and Ajwa was classified with an accuracy of 98% which is a significantly higher result than the baseline models. Table VI presents the total images used in training and testing, and also the number of correct classifications and testing accuracy for the baseline.

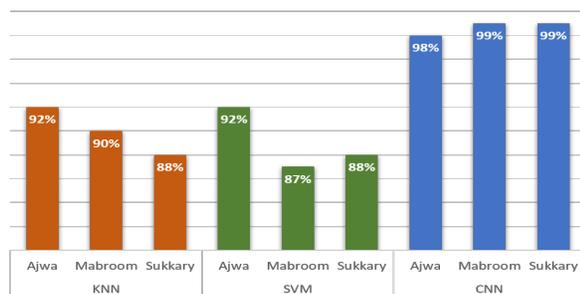


Fig. 13. The Accuracy of KNN, SVM and CNN with the Three Date Types

and proposed models. Fig. 13 shows the accuracy of the baseline models (KNN and SVM) and the proposed model (CNN).

VI. DISCUSSION

In this research, a novel dataset of three grades of three date types was produced (Ajwah (grade 1, 2, and 3), Mabroom (grade 1, 2, and 3), Sukkary (grade 1, 2, and 3)). The dataset was captured under relatively constant lighting and from a distance of 20 cm. Since the details of the date (size, color, and texture) are crucial in grading and vary between the grades within a small range. Changing the lighting or the distance of capturing would affect the appearance of the date details in the image and hence result in images that are hard to grade even by human experts. For the three features that were considered (size, color, and texture), in order to measure the size using the best least fitting ellipse as in [10] and [7], thresholding the image was necessary to define the region of interest. Among the experimented thresholding algorithms such as weight of

intensity difference, active contour, local thresholding, adaptive thresholding, and Otsu thresholding, Otsu thresholding outperformed the other algorithms and successfully thresholded all the images of the three types of date. However, as in [10], postprocessing was still needed to fill the holes inside the date and to ignore the small dots in the background. Following preprocessing, the best least fitting ellipse was applied on the binary thresholded image returning six parameters (length, width, area, perimeter, eccentricity, equidiameter). The color was extracted from the RGB channels along with the intensity of each channel. The texture was extracted using WLD which was applied on three color models (RGB, HSV, YcbCr) [7]. The extracted features were passed to the baseline models which are KNN and SVM. For KNN, different search methods and distance metrics along with different values of k were experimented. Using Minkowski distance and with k=3, the highest accuracy was achieved and ranged between 88% and 92% depending on the date type. For SVM, the RBF kernel was used, and the Bayesian optimization was applied to tune the parameters. The highest accuracy achieved ranged between 86% and 92% depending on the date type. In [10] both KNN and SVM were experimented on dates grading and the study showed that KNN gave much better results than SVM. These results hold somehow in this research, with SVM slightly less than KNN. For the proposed model, the CNN, the images were normalized to have same size across them before passing to the classifier. This model resulted in an accuracy of 98% for Ajwa and 99% for Mabroom and Sukkary. The accuracy of Mabroom and Sukkary is slightly higher than Ajwa because these two kinds show more variation in colour and size from one grade to another. Unlike Ajwa, that has a dark color and less textured skin (refer to Fig. 4). In general, CNN has performed better than KNN and SVM because this model learns high-level features without the need of feature extraction.

VII. CONCLUSION

This paper presents an image based date grading approach using a novel dataset that has three kinds of dates, namely Ajwah, Mabroom and Sukkary. The dataset consists of 3383 images of the three types of dates with their grades; Ajwa (grade 1, grade 2, and grade 3), Mabroom (grade 1, grade 2, and grade3), and Sukkary (grade 1 and grade 2). The size of the date was measured using the best least fitting ellipse after thresholding the image. Among six experimented thresholding algorithms, Otsu thresholding has successfully thresholded all the images of the dataset. The color was extracted from the RGB color space. The texture of the dates was extracted with WLD using three color models that are RGB, HSV, and YcbCr. Three classification techniques were experimented, KNN and SVM as baseline models and CNN as the proposed model. The baseline models achieved an average accuracy of 90% for KNN and 88% for SVM. The proposed model, that is CNN, outperformed the baseline models and achieved an accuracy of 98% for Ajwah, and 99% for Mabroom and Sukkary.

As a future work, this work can be extended in two directions; to build a more comprehensive date types in the dataset to represent the different varieties of dates, and to develop a real-time system that automates the grading process using the CNN algorithm.

ACKNOWLEDGMENT

The authors would like to thank Mr. Wasel Abdullah AlWasel, CEO of Manafez international company and Dr. Nadia Al-Kaff for supporting this research.

REFERENCES

- [1] Johnson, D., Al-Khayri, J. and Jain, S., 2015. Introduction: Date Production Status and Prospects in Africa and the Americas. *Date Palm Genetic Resources and Utilization*, pp.3-18.
- [2] Egyptian Economic and Commercial Office in Brazil. 2019. TOP 10 DATE PRODUCING COUNTRIES IN THE WORLD. [online] Available at: <https://ecob.com.br/top-10-date-producing-countries-in-the-world/>.
- [3] A. Hobani, A. Thottam, K. Ahmed, Development of a neural network classifier for date fruit varieties using some physical attributes, *King Saud University Agricultural Research Center* 126 (2003).
- [4] M. Fadel, Date fruits classification using probabilistic neural networks, *Agricultural Engineering International: the CIGR Ejournal* 7 (3) (2007).
- [5] A. Haidar, H. Dong, N. Marvridis, Image-based date fruit classification, *IV International Congress on Ultra Modern Telecommunications and Control Systems* (2012).
- [6] M. Alrajeh, A. Alzohairy, Date fruits classification using mlp and rbf neural networks, *International Journal of Computer Applications* 41 (10) (2012).
- [7] G. Muhammad, Date fruits classification using texture descriptors and shape-size features, *Engineering Applications of Artificial Intelligence* 37 (2015).
- [8] M. Hossain, G. Muhammad, S. Amin, Improving consumer satisfaction in smart cities using edge computing and caching: A case study of date fruits classification, *Future Generation Computer Systems* 88 (2018).
- [9] A. O. Y., Computer vision based date fruit grading system: Design and implementation., *Journal of King Saud University - Computer and Information Sciences* 23 (1) (2011).
- [10] S. Mohana, C. Prabhakar, A novel technique for grading of dates using shape and texture features, *Machine Learning and Applications: An International Journal (MLAIJ)* 1 (2) (2014).
- [11] B. Anuja, B. Atul, Fruits and vegetables quality evaluation using computer vision: A review, *Journal of King Saud University - Computer and Information Sciences* (2018).
- [12] F. e. a. Vazques, Mango classification using convolutional neural networks, *Lecture Notes in Computer Science* (2009).
- [13] X. Li, W. Zhu, Apple grading method based on features fusion of size, shape and color, *Procedia Engineering* 15 (2011).
- [14] C. Nandi, Computer vision based mango fruit grading system, *International conference on Innovative Engineering Technologies (ICIET'2014)* 28 (29) (2014).
- [15] M. Arakeri, Lakshmana, Computer vision based fruit grading system for quality evaluation of tomato in agriculture industry, *Procedia Computer Science* 79 (2016) .
- [16] O. Mahendra, H. Pardede, R. Sustika, K. Suryo, Comparison of features for strawberry grading classification with novel dataset, *International Conference on Computer, Control, Informatics and its Applications (IC3INA)*. (2018).
- [17] C. e. a. Jie, Wld: A robust local image descriptor, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (9) (2009).
- [18] M. Brahim, K. Boukhalfa, A. Moussaoui, Deep learning for tomato diseases: Classification and symp- toms visualization, *Applied Artificial Intelligence* 31 (4) (2017).
- [19] S. Jeong, H. Yoe, Fruit classification system using deep learning, *Journal of Knowledge Information Technology and Systems* 13 (5) (2018).
- [20] I. Hussain, Q. He, Z. Chen, Automatic fruit recognition based on dcnn for commercial source trace system, *International Journal on Computational Science and Applications* 8 (2/3) (2018).
- [21] H. Muresan, M. Oltean, Fruit recognition from images using deep learning, *Acta Universitatis Sapientiae Informatica* 10 (1) (2018).
- [22] J. Soo-ho, L. Meonghun, Fruit classification system using deep learning, *Journal of Knowledge Information Technology and Systems* 13 (5) (2018).
- [23] I. e. a. Sa, Deepfruits: A fruit detection system using deep neural networks, *Sensors* 16 (8) (2019).
- [24] K. Sanhdhiya, M. Vidhya, M. Saranya, Smart fruit classification using neural networks, *International Journal of Trend in Scientific Research and Development* 2 (1) (2017).
- [25] O. Patil, Classification of vegetables using tensor ow, *International Journal for Research in Applied Science and Engineering Technology* 6 (4) (2018).
- [26] K. e. a. Lin, Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network, *Frontiers in Plant Science* 10 (2019).
- [27] R. Sriram, A. Tejas, J. Girija, Mango classification using convolutional neural networks, *International Research Journal of Engineering and Technology (IRJET)* 5 (11) (2018).
- [28] H. Hiary, H. Saadeh, M. Saadeh, M. Yaqub, Flower classification using deep convolutional neural networks, *IET Computer Vision* 12 (6) (2018).
- [29] J. Brownlee, Classification accuracy is not enough: More performance measures you can use, *Machine Learning Mastery* (2014).