ATATÜRK ÜNİVERSİTESİ YAYINLARI
ATATURK UNIVERSITY PUBLICATIONS

Research Article *Araştırma Makalesi*

DOI: 10.5152/AUAF.2024.23171

Date Fruit Classification by Using Image Features Based on Machine Learning Algorithms

Makine Öğrenme Algoritmalarına Dayalı Görüntü Özellikleri Kullanılarak Hurma Meyvelerinin Sınıflandırılması

Öznur ÖZALTIN

Department of Mathematics, Atatürk University, Faculty of Science, Erzurum, Turkey

Received/Geliş Tarihi: 14.08.2023

Accepted/Kabul Tarihi: 13.11.2023

Publication Date/Yayın Tarihi: 25.01.2024

Corresponding Author/Sorumlu Yazar: Öznur ÖZALTIN
E-mail: oznur.ozaltin@atauni.edu.tr

Cite this article as: Özaltın, Ö. (2024). Date fruit classification by using image features based on machine learning algorithms. *Research in Agricultural Sciences*, 55(1), 26-35.

ABSTRACT

The date fruit, scientifically known as *Phoenix dactylifera*, is a significant dietary component due to its high nutritional value and abundance of essential vitamins and minerals. The process of discerning the classification of this fruit, which exhibits a multitude of variations within its natural domain, needs a specialized skill set. The automated recognition of species based on images of agricultural goods has gained significant prevalence in recent times. In this objective, the present study employed machine learning algorithms to automatically identify seven types of date fruit. In the investigation, decision tree, K-nearest neighbor, artificial neural networks, and support vector machine through their different hyperparameters are employed for the purpose of classifying date fruit. The dataset was divided into ratios of 80% and 20% for training and testing, respectively, and the training process employed the five-fold cross-validation technique to avoid overfitting. In summary, the results indicate that the best algorithm is neural network with a layer size of 25. In this study, this proposed algorithm achieved a test accuracy rate of 93.85%. Given the absence of computational complexity in the investigation, it can be effortlessly incorporated into diverse tools, thereby facilitating the identification of the types of date fruit.

Keywords: Automatic detection, classification, date fruit, machine learning algorithms, neural networks

ÖZ

Bilimsel olarak *Phoenix dactylifera* olarak bilinen hurma meyvesi, yüksek besin değeri ve temel vitamin ve minerallerin bolluğu nedeniyle önemli bir diyet bileşenidir. Doğal ortamında çok sayıda varyasyon sergileyen bu meyvenin sınıflandırılmasını ayırt etme süreci, özel bir yetenek gerektirir. Tarımsal ürünlerin görüntülerine dayalı türlerin otomatik olarak tanınması son zamanlarda önemli bir yaygınlık kazanmıştır. Bu amaçla, mevcut çalışma, yedi tür hurma meyvesini otomatik olarak tanımlamak için makine öğrenme algoritmalarını kullandı. Araştırmada hurma meyvelerinin sınıflandırılması amacıyla farklı hiperparametreler ile karar ağaçları, K-En Yakın Komşu, yapay Sinir Ağları ve Destek Vektör Makinesi kullanılmıştır. Veri seti, eğitim ve test için sırasıyla %80 ve %20 oranında bölünmüştür ve eğitim sürecinde, fazla uydurmayı önlemek için 5 katlı çapraz doğrulama tekniği kullanılmıştır. Özetle, sonuçlar en iyi algoritmanın katman boyutu 25 olan Sinir Ağları olduğunu göstermektedir. Bu çalışmada önerilen bu algoritma %93,85'lik bir test doğruluk oranı elde etmiştir. Araştırmada hesaplama karmaşıklığının olmaması göz önüne alındığında, çeşitli araçlara zahmetsizce dahil edilebilir, böylece hurma türlerinin tespiti kolaylaşır.

Anahtar Kelimeler: Otomatik tespit, sınıflandırma, hurma meyveleri, makine öğrenimi algoritmaları, sinir ağları

Introduction

The yearly global production of date fruit is estimated at 8.46 million tons (Albarrak et al., 2022). The date fruit is prized for its use in sweets and as a fruit crop. Dates are mostly grown in the hot, dry regions of Southwest Asia, North Africa, and the Middle East (Albarrak et al., 2022). From 1961 to 1985, date output went from 1.8 million tons to 2.8 million tons. Production of dates has increased from 5.4 million tons in 2001 to 8.46 million tons in recent years (Albarrak et al., 2022). There exist more than 40 distinct types of dates, along with over 400 variations, encompassing a vast spectrum of tastes, shapes, and hues, as well as varying in terms of cost and significance (Haidar et al., 2012). The classification of date fruits is a crucial process, especially considering that a significant portion of consumers lack the ability to distinguish between various types (Haidar et al., 2012). The lack of a fully automated system for classifying date fruit is a persistent issue in the market, forcing workers to rely instead on their own knowledge and judgment, which can be time-consuming, costly, and subject to bias (Albarrak et al., 2022). Therefore, it is of utmost significance to possess the capability to visually categorize date fruits for the purpose of automated factory classification.

Machine learning is a rapidly expanding field within the realm of computer science, characterized by its extensive range of practical applications (Osisanwo et al., 2017). In recent times, the field of image processing has gained significant popularity (Garcia et al., 2021; Koklu et al., 2021a; Ozaltin et al., 2022; Ozaltin & Yeniay, 2023a, 2023b; Wróbel et al., 2022). The acquisition of diverse information from images is a prevalent technological practice. The study utilized images of date fruits belonging to seven distinct classes, namely, Berhi, Deglet, Dokol, Iraqi, Rotana, Safavi, and Sogay, for the purpose of detecting the different varieties. The features of these images were extracted using the Otsu approach as described by Koklu et al. (2021b), resulting in the creation of the dataset. Thirty-four different features and 898 samples of date fruits were analyzed using machine learning algorithms. Machine learning algorithms make data processing easy, especially for features. If just image data is available, deep learning-derived convolutional neural networks (CNN) may be chosen. However, the construction of CNN would be laborious and complex computationally. Convolutional neural networks can extract several features that may not be obvious to the user. One notable advantage of the study described herein lies in its ability to properly explain the many sorts of features and afterward achieve the desired outcome without extra time spent.

Many researchers aimed to detect types of date fruit based on artificial intelligent as follows: Albarrak et al. (2022) developed a dataset of eight date fruit images for the purpose of classification. They utilized MobileNetv2 for this task. Their experimental findings demonstrated that their proposed approach achieved an accuracy of 99%. Muhammad (2015) introduced an algorithmic framework for the automated categorization of date fruit images. Initially, the author employed the feature extraction technique to derive features from the images of dates. Next, a feature selection method is employed to decrease the dimensionality of these features. In conclusion, SVM algorithm was assigned the task of classifying reduced features with an accuracy rate exceeding 98%. Koklu et al. (2021b) utilized a dataset consisting of 898 images of date fruit. They extracted features from these images using the Otsu method. They developed a stacking model comprising logistic regression and artificial neural networks. The method proposed by them achieved a precision of 92.8%. Altaheri et al. (2019) applied CNN using transfer learning approach for classification of five types of date fruit images. In final, their proposed approach had an accuracy of 99.01%. Abi Sen et al. (2020) proposed an automatic system for classifying four types of date fruit using SVM with an accuracy of 73.8%. Alsirhani et al. (2023) implemented DenseNet based on transfer learning approach to detect types of date fruit and obtained a test accuracy rate of 95.21%. Alhadhrami et al. (2023) utilized pretrained CNN to class date fruit images and they obtained a testing accuracy of 98.33%. Nasiri et al. (2019) proposed a deep CNN for detection types of date fruit from images. According to their experimental findings, their proposed approach was able to class effectively with an accuracy of 96.98%. Faisal et al. (2020) suggested a decision system including computer vision and deep learning approach and obtained an accuracy of 99.4%. Adige et al. (2023) aimed to detect apple types from images by using SVM and ResNet-50 via different optimal hyperparameters. Their experimental results demonstrated that SVM with an accuracy of 96% was superior to ResNet-50 with an accuracy of 90%. Arshaghi et al. (2023) used deep learning algorithms (AlexNet, GoogleNet, VGG, R-CNN, and transfer learning) to diagnose potato diseases from 5000 images. They obtained successful outcomes in their study. Gencturk et al. (2023) classified three types hazelnut based on InceptionV3 + ResNet50 data fusion model. In their study, 1024 features were obtained via their suggested model and then, they achieved 100% accuracy rate.

The literature review shows that the studies included computational complexity. Therefore, this study aims to introduce a novel framework for the classification of date fruit, employing a machine learning methodology without computational complexity. The dataset experienced training and validation utilizing decision tree, K-nearest neighbor (KNN), neural networks (NN), and support vector machine (SVM) with varying hyperparameters. The optimal hyperparameter and algorithm selection are achieved through the utilization of a wide range of performance metrics. As a result, the NN (with a layer size of 25) approach is regarded as a successful algorithm in the realm of date fruit identification and classification. Moreover, contributions and advantages of this study as follows:

1. In order to automatically classify date fruit based on features extracted from images, machine learning algorithms are utilized.
2. The hyperparameter that yields the best results for each classifier is chosen.
3. The dataset is divided into a training set of 80% and a testing set of 20%. Validation of the training set is accomplished by the five-fold cross-validation.
4. The testing results are offered to demonstrate the reliability of this study.
5. According to the findings of the experiments, it is possible to use machine learning algorithms to determine the species of an agricultural product. Therefore, these algorithms can be applied to devices and there will be an improvement in the quality of agricultural products. Figure 1 shows a framework of this study.

The following section of this study is outlined as follows: Section 2 clarifies the material and methodology that includes the overview of the dataset, machine learning algorithms along with their hyperparameters, cross-validation, and performance evaluation. In the next, experimental findings are presented, and discussed in Section 3. Finally, this study is concluded and the following investigations are explained briefly in Section 4.

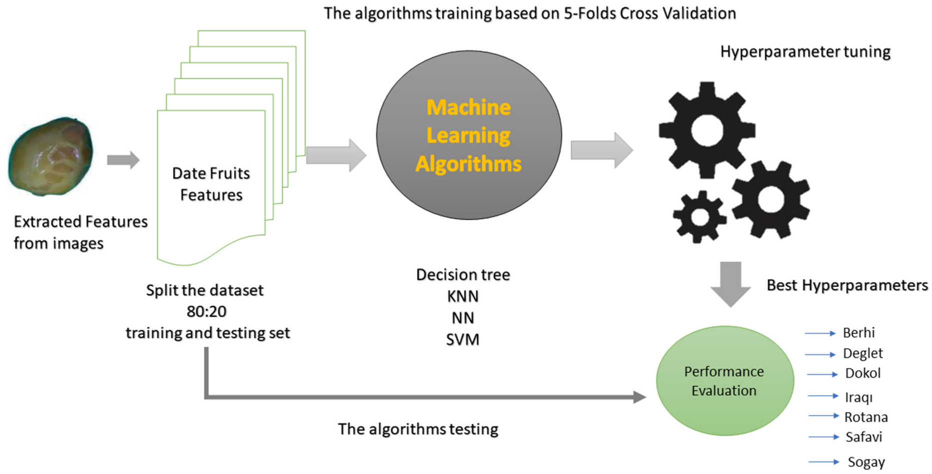


Figure 1.
Flowchart of the presented study.

Methods

In the presented study, features of date fruit images were classified using decision tree, KNN, NN, and SVM based on various hyperparameters to find the type of date fruit. Moreover, the dataset was split as 80:20 training–testing set and then the training set was validated by using five-fold cross-validation. All experiments were evaluated with diverse metrics. More details are given in this section.

Date Fruit Dataset

The date fruit dataset is downloadable in .xlsx format from the website https://www.muratkoklu.com/datasets/. This dataset comprises a collection of seven distinct varieties of date fruit which are obtained through the computer vision system (Koklu et al., 2021b). Based on the image processing approach, 34 features were achieved from date fruit images by Koklu et al. (2021b). They extracted morphological features from images and applied the image processing method. First, they converted color images to grayscale and binary for feature extraction. Then, they used threshold and pixel information methods. After image processing, date fruit were analyzed individually and features were retrieved by them. They employed the Otsu method, a common image thresholding approach and explained the Otsu method as follows: it uses a variable to distinguish between nature's groupings. The method operates on gray-level images, checking how many times each color is present on the image. The image color distributions are calculated first, and then other procedures are done on this distribution sequence.

These 34 features include genetic varieties such as morphological features, shape, and color. The main feature details are as follows: morphological features: area, perimeter, major axis, minor axis, eccentricity, roundness, equivalent diameter, solidity, convex area, extent, aspect ratio, and compactness. The other main feature is shape features: shapefactor_1, shapefactor_2, shapefactor_3, and shapefactor_4. The last main feature is shape features: shapefactor_1, shapefactor_2, shapefactor_3, and shapefactor_4, and the last main feature is color features: mean RR, std. dev RR, skew RR, kurtosis RR, entropy RR, all daub4 RR, mean RG, std. dev RG, skew RG, kurtosis RG, entropy RG, all daub4 RG, mean RB, std. dev RB, skew RB, kurtosis RB, entropy RB, all daub4 RB. (R: red, G: green, B: blue). More details about the dataset are given in Table 1. Further information can be found in reference (Koklu et al., 2021b).

In total, in the dataset, there are 898 date fruit classes and 34 different features obtained from images. Hence, the dataset possesses an 898×35 size.

Machine Learning Algorithms

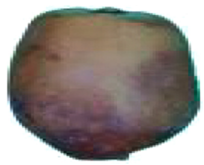
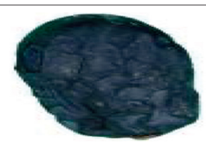
Decision Tree

Decision tree are widely utilized in the field of machine learning as a prominent category of methods (Zhou, 2021). A decision tree often comprises a single root node, several internal nodes, and multiple leaf nodes. The terminal nodes represent the decision outcomes, while all other nodes represent feature tests. The samples contained within each node are partitioned into child nodes based on the outcomes of feature splitting. Every trajectory from the primary node to the terminal node can be considered as a succession of decisions. The objective is to generate a tree structure that possesses the ability to make accurate predictions on samples that have not been previously encountered (Zhou, 2021). The essence of the decision tree learning algorithm is in the process of identifying and selecting the most advantageous splitting criteria (Zhou, 2021). In this study, some splitting criteria, Gini's diversity index, maximum deviance reduction, and towing role, were examined to determine the best one. In addition, the maximum number of splits was determined as 100.

K-Nearest Neighbor

The nearest neighbor classifiers do not necessitate any preprocessing of the labeled sample set before their utilization. The crisp nearest-neighbor classification rule allocates an input sample vector y, whose classification is unknown, to the class of its nearest neighbor (Cover & Hart, 1967; Keller et al., 1985). The concept described can be generalized to the K-nearest neighbor algorithm, where the vector y is assigned to the class that is most frequently represented among its K-nearest neighbor (Keller et al., 1985). In the context of K-nearest neighbor, it is important to acknowledge the potential occurrence of ties among classes when multiple neighbors are taken into account. One straightforward approach to addressing this issue is to impose limitations on the feasible values of K. Given a binary classification problem, if we impose a constraint on the value of K such that it can only be odd, it ensures that there will be

Table 1.
The characteristics contained within the dataset (Koklu et al., 2021b)

Classes	Images	Number of Instances	Features
Barhee origin is Basra, Iraq.		65	Morphological features: 12 Shape features: 4 Color features: 18 Details: When it is ready to be picked, it is a golden-brown color. It has a hard shell and is small to medium in size.
Deglet Nour origin is not specified.		98	Morphological features: 12 Shape features: 4 Color features: 18 Details: It is a type of date fruit that ranges in size from medium to large and turns from yellow to dark brown after being picked.
Sukkary origin is Al Qassim region, Saudi Arabia.		204	Morphological features: 12 Shape features: 4 Color features: 18 Details: It is a medium-sized, golden-colored date fruit variety.
Rotap Mozafati origin is Kerman, Iran.		72	Morphological features: 12 Shape features: 4 Color features: 18 Details: It possesses a dense, dark brown look. It is a variety of medium-sized, succulent dates.
Ruthana origin is Madinah, Saudi Arabia.		166	Morphological features: 12 Shape features: 4 Color features: 18 Details: It possesses brown and gold hues. It is a species of medium-sized date fruit.
Safawi origin is Madinah, Saudi Arabia.		199	Morphological features: 12 Shape features: 4 Color features: 18 Details: It is a dark black cherry color with brown ends. It possesses medium-sized.
Sagai origin is Arabian Peninsula, particularly Saudi Arabia.		94	Morphological features: 12 Shape features: 4 Color features: 18 Details: The ends are dry and golden in hue, while the undersides are soft and brown in color. It is a date variety of medium-sized.

no possibility of a tie. When the number of classes exceeds two, this technique becomes ineffective. One possible approach for managing the situation of a tie is as follows. The sample vector is assigned to the class that has the minimum sum of distances from the sample to each neighbor in the class, among the classes that are tied. In the event that a tie occurs, the assignment will be given to the last class encountered among those that are tied. This assignment is considered arbitrary. It is evident that there will exist instances in which the categorization of a vector is subject to an arbitrary assignment, irrespective of the inclusion of supplementary procedures within the algorithm (Keller et al., 1985).

There exists a predetermined value for K, which is utilized in the process of determining the K-nearest neighbor through distance computation (Chomboon et al., 2015). In this study, K is identified as 10, and distance metrics: cosine, Euclidean, and Minkowski (cubic) were respectively analyzed to measure the performance of date fruit classification. Moreover, distance weights were equal for each process of classification.

Neural Networks

Neural networks (McCulloch & Pitts, 1943) are computational algorithms that aim to emulate certain aspects of the biological brain, such as the ability to learn, generalize, and abstract from past experiences. These computational systems possess the capability to perceive and analyze patterns in order to establish connections when presented with factual information. They are essentially comprised of fundamental computing units that are linked together in various manners to construct a network (Bourquin et al., 1997; Hecht-Nielsen, 1988; Kohonen, 1988). The capability to extract unconscious information from data renders NN intriguing tools for the purpose of modeling. NN can also be perceived from a mathematical perspective as an extensive category of versatile, nonlinear regression and discriminant algorithms (Bourquin et al., 1997). In this study, NN are employed for the purpose of image classification by leveraging their extracted features.

When designing a functional model of the biological neuron, there are three fundamental components that hold significance.

Initially, the synapses of the neuron are represented as weights. The weight value denotes the magnitude of the connection strength between an input and a neuron. In the realm of neural networks, it is widely acknowledged that weight values that are negative in nature represent inhibitory connections, whereas weight values that are positive in nature are indicative of excitatory connections (Dongare et al., 2012). The subsequent pair of components simulate the intrinsic dynamics occurring within the neuronal cell. A computational unit known as an adder operates by summing up all the inputs, which are subject to modification by their corresponding weights. This particular operation is commonly denoted as a linear combination. In conclusion, an activation function governs the magnitude of the output of the neuron. Typically, the acceptable range of output is bounded by the values of 0 and 1, inclusively, or by the values of −1 and 1, inclusively (Dongare et al., 2012). In the presented study, rectified linear activation function (ReLU), which is frequently used in this study, was chosen. Moreover, different size of hidden layers was determined and renamed NN. When the hidden layer size was 10, 25, and 100, the algorithm was named Narrow NN, Medium NN, and Wide NN, respectively. Meanwhile, Medium NN are proposed algorithm in terms of obtaining the highest success for classifying date fruit types. Additionally, maximum iteration was limited to 1000 in the presented study.

Support Vector Machine

Support vector machine is a type of machine learning algorithms that were initially developed to address classification problems and have now been extended to handle a range of additional scenarios. These algorithms are grounded in the ideas of statistical learning theory and convex optimization. They are presently employed in diverse fields such as bioinformatics, text categorization, and computer vision (Mammone et al., 2009). Support vector machine was first introduced by Vapnik and colleagues (Boser et al., 1992) in the 1990s (Mammone et al., 2009). They belong to a group of algorithms designed to learn two-class discriminant functions based on a given collection of training instances (Mammone et al., 2009). The initial introduction of the simplest model of SVM was referred to as the maximal margin classifier or hard margin SVM. While its practical application is limited due to its reliance on linearly separable data, this method serves as a foundational component for complex SVM. Furthermore, it is comprehensible and easy (Mammone et al., 2009). Due to the presence of noise in many real-world datasets, the maximal margin approach was not applicable as it generates a hypothesis that exactly aligns with the training data, making it unable to identify a linear separation between classes. This issue provided the basis for the advancement of a more robust version of the algorithm initially proposed by Cortes and Vapnik (Cortes & Vapnik, 1995). This enhanced version was designed to withstand the presence of noise and outliers in the dataset while minimizing significant changes to the answer. A higher-dimensional space, known as feature space, can be used to remap the data points for a more accurate representation of the data (Mammone et al., 2009). The explicit functional form of the mapping is not required to be known, as it is implicitly defined by the selection of a kernel function. In the presented study, three different kernel functions: linear, cubic, and quadratic were investigated to find the best classifying function. Further, other hyperparameters were as follows: Box constraint level = 1, multi-class method = one vs. one.

Cross-Validation

Cross-validation is a technique that has been devised to enhance the robustness of classification by minimizing potential security vulnerabilities (Koklu & Ozkan, 2020). Cross-validation is a technique that involves randomly partitioning the dataset into a pre-determined number of sets, each of which has an equal size. The system is trained using the remaining sets, wherein one of the subsets is selected as the test set. The above process is iterated until all sets of numbers have been tested within the system. The outcomes derived from these procedures are generalized (Koklu et al., 2021b; Koklu & Ozkan, 2020). In the presented study, the dataset was first divided into 80:20 training and testing sets, randomly. Then the training set was validated based on cross validation approach where the number of folds was determined as 5. Therefore, overfitting is overcome.

Performance Evaluation

In this study, to identify which machine learning algorithm is the best, the performance metrics which are accuracy (acc), sensitivity (sens), specificity (spe), precision (pre), F1-score (F1), geometric mean (G-Mean), Matthews correlation coefficient (MCC) (Matthews, 1975), and the kappa value (κ) (Cohen, 1960) are used. The metrics are displayed in Equations (1)–(9) (Chicco et al., 2021; Ozaltin et al., 2023a, 2023b; Rajinikanth et al., 2020; Sharifrazi et al., 2021; Singh et al., 2022; Wang et al., 2019):

$$Acc = p_A = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

$$Sens = \frac{TP}{(TP + FN)} \tag{2}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{3}$$

$$\text{Pr}ecision = \frac{TP}{(TP + FP)} \tag{4}$$

$$F1 - Score = \frac{(2 \times TP)}{(2 \times TP + FP + FN)} \tag{5}$$

$$G - Mean = \sqrt{Sensitivity \times Specificity} \tag{6}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{7}$$

$$Expected\ acc = p_E = \frac{(TP + FP) \times (TP + FN) + (FP + TN) \times (TN + FN)}{(TP + TN + FP + FN)^2} \tag{8}$$

$$\kappa = \frac{p_A - p_E}{1 - p_E} \tag{9}$$

where TP , FP , TN , and FN are true positive, false positive, true negative, and false negative, respectively.

Results

In this section, the findings of the generated models are shown. The models were constructed in order to recognize different types of date fruit based on features that were extracted from images. The suggested model has been built in the MATLAB (2022b) environment, which was run on a personal computer. The dataset has dimensions of 898 × 35, and it was first split into training and testing sets with a ratio of 80:20. As a result, the dimensions of the training set were 719 × 35, whereas the dimensions of the testing set were 179 × 35. Four different machine learning methods were then used to analyze the training set based on five-fold cross-validation. All of the features in the study were used, and none were ever dropped from the dataset. In fact, a few feature selection techniques were tested, however, the performance in this study did not improve with identified hyperparameters. The findings of the performance evaluations obtained by cross-validation and through tests are presented in Table 2 and 3, respectively.

While Table 2 is reviewed, the performance of four distinct algorithms, each of which is based on a different combination of three hyperparameters, is measured using various metrics. When a decision tree using Gini's diversity index, maximum deviance reduction, and towing role were chosen as the classifier, successful performance was attained using maximum deviance reduction splitting criteria with a validation accuracy of 84.42%. Based on the KNN classifier, the maximum validation accuracy with 86.65% and other metrics were obtained by using the Euclidean distance. The performance outcomes of neural networks employing various layer sizes, specifically 10, 25, and 100, exhibited striking similarities, with validation accuracies above 89%. Therefore, the evaluation of this method will be based on the outcomes obtained from the tests. In the case of SVM, a comparable scenario to that of NN was seen, wherein the performance outcomes exhibited a high degree of similarity, with validation accuracies above 89.8%.

In this research, a subset of the dataset consisting of 179 instances was tested. Testing results are shown in Table 3. According to this table, based on decision tree' results, Gini's diversity index and maximum deviance reduction achieved almost the same testing

Table 2.
The performance values of machine learning algorithms using five-fold cross-validation.

Model	Hyperparameter	Acc	Sens	Spe	Pre	F1	G-Mean	MCC	Kappa
Decision Tree	Gini's Diversity Index	0.8317	0.7740	0.9723	0.7763	0.7744	0.8675	0.7473	0.3128
	Max.Deviance Reduction	0.8442	0.7833	0.9745	0.7893	0.7849	0.8736	0.7605	0.3639
	Towing Role	0.8081	0.7364	0.9685	0.7425	0.7385	0.8445	0.7077	0.2163
KNN	Cosine	0.8401	0.7819	0.9724	0.8178	0.7864	0.8720	0.7685	0.3469
	Euclidean	0.8665	0.8195	0.9777	0.8345	0.8247	0.8951	0.8042	0.4548
	Minkowski	0.8595	0.8098	0.9765	0.8297	0.8150	0.8892	0.7949	0.4264
NN	Narrow	0.8901	0.8611	0.9819	0.8656	0.8632	0.9195	0.8452	0.5513
	Medium	0.8901	0.8634	0.9818	0.8691	0.8654	0.9207	0.8478	0.5513
	Wide	0.8971	0.8734	0.9831	0.8763	0.8746	0.9266	0.8577	0.5797
SVM	Linear*	**0.9096**	**0.8859**	**0.9851**	**0.8894**	**0.8864**	**0.9342**	**0.8723**	**0.6309**
	Cubic	0.8983	0.8677	0.9833	0.8710	0.8693	0.9237	0.8527	0.5848
	Quadratic	0.9082	0.8849	0.9849	0.8865	0.8853	0.9336	0.8704	0.6252

Note: ***Bold metrics indicate the highest performance in this part of the study.

Table 3.
The performance values of machine learning algorithms testing results.

Model	Hyperparameter	Acc	Sens	Spe	Pre	F1	G-Mean	MCC	Kappa
Decision Tree	Gini's Diversity Index	0.8659	0.8303	0.9783	0.8329	0.8243	0.9012	0.8070	0.4525
	Max.Deviance Reduction	0.8603	0.8216	0.9773	0.8263	0.8212	0.8961	0.8001	0.4297
	Towing Role	0.8380	0.7950	0.9737	0.7987	0.7930	0.8798	0.7689	0.3385
KNN	Cosine	0.8715	0.8280	0.9779	0.8685	0.8351	0.8999	0.8216	0.4753
	Euclidean	0.9218	0.9087	0.9869	0.9124	0.9095	0.9470	0.8970	0.6806
	Minkowski	0.9106	0.8971	0.9850	0.8971	0.8967	0.9400	0.8819	0.6350
NN	Narrow	0.9050	0.8875	0.9847	0.8775	0.8811	0.9348	0.8664	0.6122
	Medium*	**0.9385**	**0.9290**	**0.9897**	**0.9292**	**0.9278**	**0.9589**	**0.9183**	**0.7491**
	Wide	0.9330	0.9241	0.9890	0.9189	0.9195	0.9560	0.9095	0.7263
SVM	Linear	0.9106	0.8963	0.9853	0.8982	0.8948	0.9398	0.8815	0.6350
	Cubic	0.9218	0.9041	0.9872	0.9073	0.9034	0.9447	0.8919	0.6806
	Quadratic	0.9274	0.9188	0.9882	0.9132	0.9138	0.9528	0.9031	0.7034

Note: ***Bold metrics indicate the highest performance in this part of the study.

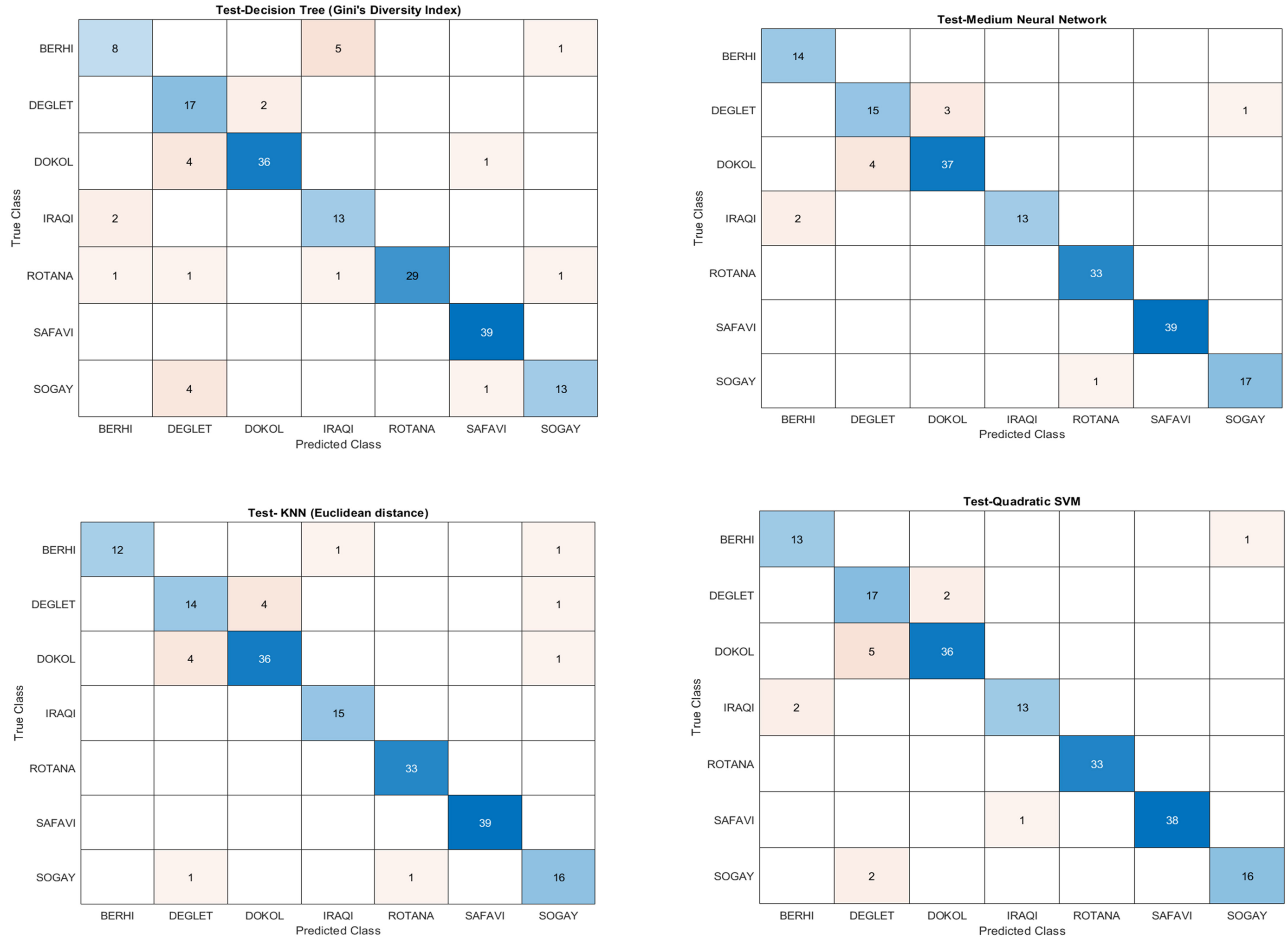


Figure 2.
Confusion matrices of machine learning algorithms.

results. The algorithm achieved the highest possible rate of testing accuracy, which was 86.59%; this is considered to be satisfactory. However, when the kappa values were analyzed, the results showed that values below 0.50 indicated that the method in question should not be used (Wang et al., 2019).

Based on the results of KNN, the maximum testing accuracy rate that was 92.18% was obtained using Euclidean distance. Moreover, the kappa value was 0.68, and also it can be said that the result was acceptable to detect the type of date fruit.

In the presented study, it was observed that the Medium NN with a layer size of 25 achieved the highest testing accuracy of 93.85%. Additionally, the technique also yielded a maximum kappa value of 0.75.

Based on the findings of SVM, it was observed that the Quadratic kernel function achieved a testing accuracy of 92.74%. Additionally, the corresponding kappa value was determined to be 0.70, which falls within an acceptable range. Nevertheless, the findings of this study indicate that Medium NN yielded the highest performance metrics. Hence, this study introduces the concept of Medium NN as a means to achieve optimal performance in classifying different species of date fruit. Figure 2 and Figure 3 display different scenarios of confusion matrix and ROC (receiver operating characteristic) curves. Additionally, these curves show AUC (area under curve) for each class.

Figure 2 displays four confusion matrices based on the most effective hyperparameters for detecting date fruit types. The first matrix is produced by a decision tree algorithm that uses Gini's diversity index. This scenario incurs 24 costs. The second matrix is generated by using the KNN algorithm, which utilizes the Euclidean distance. This scenario has a total cost of 14. The Medium NN method yields these findings, and there are 11 costs in total. The ultimate confusion matrix is obtained using the SVM algorithm via a quadratic kernel function and there are 13 costs in this scenario. As a result, confusion matrix of Medium NN has minimum cost, and it is the best algorithm to detect the types of date fruit.

The ROC curve demonstrates the relationship between sensitivity (true positive rate) and specificity (one minus the false positive rate). Classifiers with curves positioned closer to the top-left corner of the graph typically indicate better performance. In the presented study, Figure 3 indicates ROC curves and AUC values for each class using four classifiers (decision tree, KNN, Medium NN, and SVM) with the most optimal hyperparameters. Each AUC value falls within the range of 0.75 to 1.00, and the decision tree, KNN, Medium NN, and SVM have average AUC values of 0.9388,

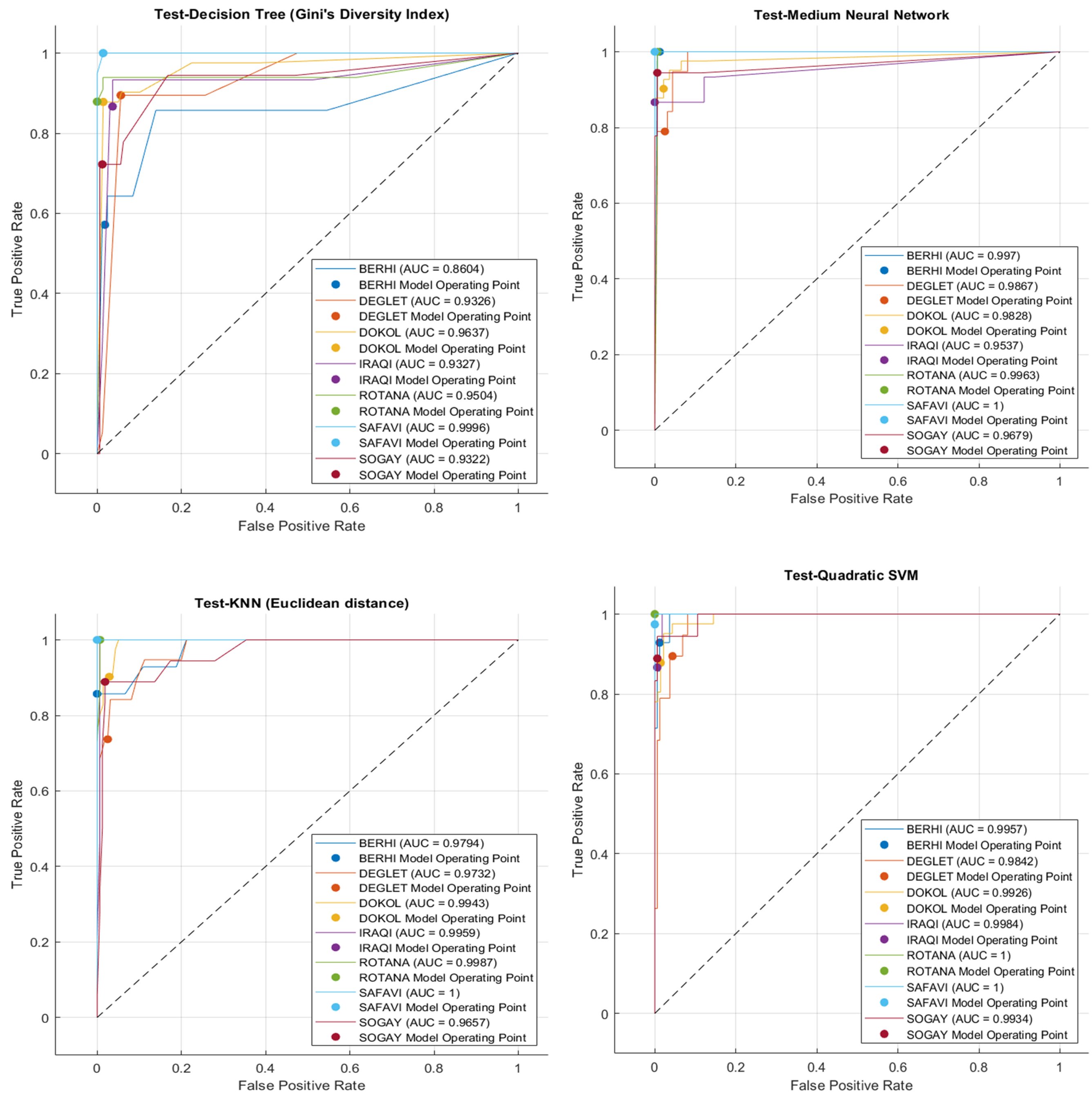


Figure 3.
Roc Curves of machine learning algorithms.

0.9867, 0.9835, and 0.9949 respectively. Classifiers have a remarkable ability to accurately identify different species of date fruit.

As a consequently, this investigation should have an emphasis on both of these different scenarios. The first recommendation is that research be conducted on the hyperparameters of machine learning algorithms. The second recommendation is that while presenting the data set, not only the result of the validation technique but also the results of the test should be included. In final, the presented study has the maximum testing accuracy of 93.9 % via Medium NN.

Machine learning algorithms offer a user-friendly approach to handling data, particularly in relation to incorporating various features. Convolutional neural networks derived from deep learning methods may be selected if solely the image data were accessible. Nevertheless, the process of constructing CNN would entail significant labor and computational complexity. Furthermore, CNN are capable of extracting a multitude of features, the significance of which may not be discernible to the observer. One notable benefit of this study is its significant departure from computational complexity. Other benefits of this study are: the dataset was evaluated for training–validation–testing and fine-tuning hyperparameters were effectively determined to class date fruit. One limitation of this study is the unavailability of the dataset images, which restricts the ability to perform comparisons with CNN.

Discussion, Conclusion, and Recommendations

Agricultural goods are easily categorized by machines, which benefits customers as well as vendors. This study proposes a method for automatically detecting and classifying various date fruit varieties. When features are extracted from images, machine learning algorithms can effectively identify seven types of date fruit using various hyperparameters.

In the presented study, various machine learning algorithms such as decision tree, KNN, artificial NN, and SVM are used to classify date fruit based on their different hyperparameters. The dataset was split into proportions of 80% for training and 20% for testing. To prevent overfitting, the training process utilized the 5-fold cross-validation technique. The dataset includes 34 features from images of date fruit. These features involve genetic variations like morphological features, shape, and color. Here are the main features in detail: These are some of the morphological features that can be examined: area, perimeter, major axis, minor axis, eccentricity, roundness, equivalent diameter, solidity, convex area, extent, aspect ratio, and compactness. Another main aspect is the shape features. Experimental results show some important findings as follows: (i) While the decision tree is selected as a classifier; it shows Gini's diversity index is the best hyperparameter and the algorithm obtains 86.59% testing accuracy. (ii) KNN achieves the highest performance via Euclidean distance with 92.18% testing accuracy. (iii) Artificial NN consisting of 25 layers (called Medium NN) achieved the best test accuracy rate, reaching an impressive value of 93.9%. Therefore, the presented study suggests Medium NN to detect types of date fruit (iv) SVM gets the successful performance based on quadratic kernel function. Additionally, this study also involved the calculation of the Kappa value and other relevant criteria. A kappa value exceeding 0.70 signifies that the proposed approach demonstrates strong classification performance. The proposed algorithm used the ReLU activation function and 1000 maximum iteration limits.

This study has some limitations. Owing to the unavailability of images of the dataset, a comparison with deep learning algorithms could not be presented. Furthermore, due to the inability to evaluate various image features or extraction methods, the effectiveness of machine learning algorithms in these circumstances remains uncertain.

The objective of future research is to conduct comparative analyses on agricultural products and to devise tools that utilize deep learning and machine learning algorithms to class images and features of the same images.

Peer-review: Externally peer-reviewed.

Declaration of Interests: The author has no conflicts of interest to declare.

Funding: The author declared that this study has received no financial support.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazar çıkar çatışması bildirmemiştir.

Finansal Destek: Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

References

Abi Sen, A. A., Bahbouh, N. M., Alkhodre, A. B., Aldhawi, A. M., Aldham, F. A., & Aljabri, M. I. (2020). A classification algorithm for date fruits. 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom). [CrossRef]

Adige, S., Kurban, R., Durmuş, A., & Karaköse, E. (2023). Classification of apple images using support vector machines and deep residual networks. *Neural Computing and Applications*, *35*(16), 12073–12087. [CrossRef]

Albarrak, K., Gulzar, Y., Hamid, Y., Mehmood, A., & Soomro, A. B. (2022). A deep learning-based model for date fruit classification. *Sustainability*, *14*(10), 6339. [CrossRef]

Alhadhrami, N., Abobakr, A., Alhammadi, A., & Shatnawi, M. (2023). Multiple classification of date fruit using transfer learning. 2023 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 1-5, [CrossRef]

Alsirhani, A., Siddiqi, M. H., Mostafa, A. M., Ezz, M., & Mahmoud, A. A. (2023). A novel classification model of date fruit dataset using deep transfer learning. *Electronics*, *12*(3), 665. [CrossRef]

Altaheri, H., Alsulaiman, M., & Muhammad, G. (2019). Date fruit classification for robotic harvesting in a natural environment using deep learning. *IEEE Access*, *7*, 117115–117133. [CrossRef]

Arshaghi, A., Ashourian, M., & Ghabeli, L. (2023). Potato diseases detection and classification using deep learning methods. *Multimedia Tools and Applications*, *82*(4), 5725–5742. [CrossRef]

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory, 144-152, [CrossRef]

Bourquin, J., Schmidli, H., van Hoogevest, P., & Leuenberger, H. (1997). Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development. *Pharmaceutical Development and Technology*, *2*(2), 95–109. [CrossRef]

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. *IEEE Access*, *9*, 78368–78381. [CrossRef]

Chomboon, K., Chujai, P., Teerarassammee, P., Kerdprasop, K., & Kerdprasop, N. (2015). An empirical study of distance metrics for k-nearest neighbor algorithm. Proceedings of the 3rd international conference on industrial application engineering, Japan, 280-285, [CrossRef]

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. [CrossRef]

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. [CrossRef]

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. [CrossRef]

Dongare, A., Kharde, R., & Kachare, A. D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, *2*(1), 189–194.

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

Garcia, L. J., Timmermans, M., Pozuelos, F. J., Ducrot, E., Gillon, M., Delrez, L., Wells, R. D., & Jehin, E. (2021). prose: A Python framework for modular astronomical images processing. *Monthly Notices of the Royal Astronomical Society*, *509*(4), 4817–4828. [CrossRef]

Gencturk, B., Arsoy, S., Taspinar, Y. S., Cinar, I., Kursun, R., Yasin, E. T., & Koklu, M. (2023). Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models. *European Food Research and Technology*, 1–14. [CrossRef]

Haidar, A., Dong, H., & Mavridis, N. (2012). Image-based date fruit classification. 2012 IV International Congress on Ultra Modern Telecommunications and Control Systems, IEEE, St. Petersburg, Russia, 357-363, [CrossRef]

Hecht-Nielsen, R. (1988). Neurocomputing: Picking the human brain. *IEEE Spectrum*, *25*(3), 36–41. [CrossRef]

Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC–15*(4), 580–585. [CrossRef]

Kohonen, T. (1988). An introduction to neural computing. *Neural Networks*, *1*(1), 3–16. [CrossRef]

Koklu, M., Cinar, I., & Taspinar, Y. S. (2021a). Classification of rice varieties with deep learning methods. *Computers and Electronics in Agriculture*, *187*, 106285. [CrossRef]

Koklu, M., Kursun, R., Taspinar, Y. S., & Cinar, I. (2021b). Classification of date fruits into genetic varieties using image analysis. *Mathematical Problems in Engineering*, *2021*, 1–13. [CrossRef]

Koklu, M., & Ozkan, I. A. (2020). Multiclass classification of dry beans using computer vision and machine learning techniques. *Computers and Electronics in Agriculture*, *174*, 105507. [CrossRef]

Mammone, A., Turchi, M., & Cristianini, N. (2009). Support vector machines. *WIREs Computational Statistics*, *1*(3), 283–289. [CrossRef]

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, *405*(2), 442–451. [CrossRef]

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*(4), 115–133. [CrossRef]

Muhammad, G. (2015). Date fruits classification using texture descriptors and shape-size features. *Engineering Applications of Artificial Intelligence*, *37*, 361–367. [CrossRef]

Nasiri, A., Taheri-Garavand, A., & Zhang, Y.-D. (2019). Image-based deep learning automated sorting of date fruit. *Postharvest Biology and Technology*, *153*, 133–141. [CrossRef]

Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128–138. [CrossRef]

Ozaltin, O., Coskun, O., Yeniay, O., & Subasi, A. (2022). A deep learning approach for detecting stroke from brain CT images using OzNet. *Bioengineering*, *9*(12), 783. [CrossRef]

Ozaltin, O., & Yeniay, Ö. (2023a). Detection of monkeypox disease from skin lesion images using Mobilenetv2 architecture. *Communications Faculty of Science University of Ankara Series A1Mathematics and Statistics*, *72*(2), 482–499. [CrossRef]

Ozaltin, O., & Yeniay, O. (2023b). A novel proposed CNN–SVM architecture for ECG scalograms classification. *Soft Computing*, *27*(8), 4639–4658. [CrossRef]

Ozaltin, O., Yeniay, O., & Subasi, A. (2023a). Artificial intelligence-based brain hemorrhage detection. In *Accelerating strategic changes for digital transformation in the healthcare industry* (pp. 179–199). Elsevier.

Ozaltin, O., Yeniay, O., & Subasi, A. (2023b). OzNet: A new deep learning approach for automated classification of COVID-19 computed tomography scans. *Big Data*. [CrossRef]

Rajinikanth, V., Joseph Raj, A. N., Thanaraj, K. P., & Naik, G. R. (2020). A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection. *Applied Sciences*, *10*(10), 3429. [CrossRef]

Sharifrazi, D., Alizadehsani, R., Roshanzamir, M., Joloudari, J. H., Shoeibi, A., Jafari, M., Hussain, S., Sani, Z. A., Hasanzadeh, F., Khozeimeh, F., Khosravi, A., Nahavandi, S., Panahiazar, M., Zare, A., Islam, S. M. S., & Acharya, U. R. (2021). Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomedical Signal Processing and Control*, *68*, 102622. [CrossRef]

Singh, D., Taspinar, Y. S., Kursun, R., Cinar, I., Koklu, M., Ozkan, I. A., & Lee, H.-N. (2022). Classification and analysis of pistachio species with pre-trained deep learning models. *Electronics*, *11*(7), 981. [CrossRef]

Wang, J., Yang, Y., & Xia, B. (2019). A simplified Cohen's Kappa for use in binary classification data annotation tasks. *IEEE Access*, *7*, 164386–164397. [CrossRef]

Wróbel, M., Stan-Kłeczek, I., Marciniak, A., Majdański, M., Kowalczyk, S., Nawrot, A., & Cader, J. (2022). Integrated geophysical imaging and remote sensing for enhancing geological interpretation of landslides with uncertainty estimation—A case study from Cisiec, Poland. *Remote Sensing*, *15*(1), 238. [CrossRef]

Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.