Selcuk Journal of Agriculture and Food Sciences
https://dergipark.org.tr/tr/pub/selcukjafsci
Research Article

(2024) 38(3), 403-413 DOI:10.15316/SJAFS.2024.036 e-ISSN: 2458-8377 Selcuk J Agr Food Sci

Classification of Orange Features for Quality Assessment Using Machine Learning Methods

Talha Alperen Cengel[1], **Bunyamin Gencturk**[2], **Elham Tahsin Yasin**[2], **Muslume Beyza Yildiz**[1], **Ilkay Cinar**[1], **Osman Ozbek**[3], **Murat Koklu**[1,*]

[1]Technology Faculty, Department of Computer Engineering, Selcuk University, Konya, Türkiye

[2]Graduate School of Natural and Applied Sciences, Selcuk University, Konya, Türkiye

[3]Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Selcuk University, Konya, Türkiye

HIGHLIGHTS

- Oranges are valued for their high vitamin C, sweet taste, fiber, and antioxidant qualities.
- Quality assurance is vital for orange market competitiveness and customer satisfaction.
- Traditional quality assessment methods are costly and prone to human error.
- The study found k-NN to be the most accurate algorithm for orange quality assessment at 69.38%.
- Machine learning improves orange quality control, benefiting consumers and producers.

Abstract

Oranges are a member of the citrus family and are eaten in large quantities due to their high vitamin C content, sweet and tart taste, and useful fiber and antioxidant qualities. Orange quality assurance is essential to market competitiveness and customer satisfaction. Conventional approaches to evaluating quality are costly and susceptible to mistakes made by people. This research investigates how well different machine learning algorithms automate and improve the orange quality assessment procedure. A dataset containing 241 samples and 11 features (size, weight, sweetness (Brix), acidity (pH), and color) was used to evaluate the effectiveness of the Random Forest (RF), XGBoost, and k-Nearest Neighbors (k-NN) algorithms. According to the findings, k-NN acquired the maximum accuracy of 69.38%, with RF coming in second at 67.34% and XGBoost third at 63.26%. These results demonstrate how machine learning models may be used to improve quality control in the orange industry by offering a more dependable and effective approach. According to this study, machine learning can significantly improve the quality control procedures for oranges, resulting in higher-quality goods for customers and more productivity for providers. The orange sector can enhance product quality and expedite operations by utilizing these technologies, eventually benefiting producers and consumers.

Keywords: Artificial Intelligence; Quality Classification; Orange Quality; Machine Learning Algorithms.

Citation: Cengel TA, Gencturk B, Yasin ET, Yildiz MB, Cinar I, Ozbek O, Koklu M (2024). Classification of Orange Features for Quality Assessment Using Machine Learning Methods. *Selcuk Journal of Agriculture and Food Sciences*, 38(3), 403-413. https://doi.org/10.15316/SJAFS.2024.036
Correspondence: mkoklu@selcuk.edu.tr
Received date: 30/06/2024
Accepted date: 22/08/2024

1. Introduction

Orange is a fruit belonging to the citrus family and is usually orange in color and oval in shape. It contributes to strengthening the immune system with its abundant amount of vitamin C. With a sweet and sour flavor, oranges are mostly consumed fresh or used for juices and desserts. They are also rich in fiber and antioxidants, which have positive effects on digestive health and overall body function. In order to increase sales of orange fruits and improve market competitiveness, the selection of high quality oranges is of great importance (Cayuela and Weiland 2010). During this selection process, many criteria such as size, shape, acidity, sweetness and color are used to determine quality (Asriny et al., 2020). However, this process can still be time-consuming to perform manually and is vulnerable to human error. Therefore, the development of artificial intelligence technologies that automate the orange selection process and objectively determine quality can offer a more reliable and efficient system for both producers and consumers (Denata et al. 2021; Ganesh et al. 2019; Kumar et al. 2024). Literature reviews on the quality of oranges and external factors affecting the quality of oranges are included in this section:

Santiago et al. (2019) presented two different methods for assessing the freshness and caliber of oranges. In the first study, they classified fresh and rotten orange images with 78.57% accuracy using CNN algorithm. The second study classified oranges as small, medium and large using LDA and k-NN algorithms. The LDA algorithm outperformed k-NN with an overall accuracy of 82% and a kappa coefficient of 0.66 (Santiago et al., 2019). Ganesh et al. (2019) introduced a deep learning method called Deep Orange for orange detection and segmentation in orange groves. This method is based on the image segmentation framework Mask R-CNN. The system uses both RGB and HSV color information from images taken under natural light conditions from an orange grove in Florida. The addition of HSV data significantly improves the model's accuracy in orange identification. The accuracy increases from 89.47% when only RGB data is used to 97.53% when HSV data is added. The overall effectiveness of the system is measured by the F1 score of close to 89% using both RGB and HSV data (Ganesh et al. 2019). In another study, Asriny et al. (2020) used a Convolutional Neural Network (CNN) to determine the quality of oranges. They categorized the dataset into 5 classes according to the quality status of the orange. The dataset contains a total of 1000 orange images using a smartphone camera. For each class, it consists of 200 images divided into 60% training data, 20% validation data and 20% test data. In this study, they achieved the highest classification accuracy of 96% with CNN. Leelavathy et al. (2021) developed a method for assessing the freshness of orange fruits using a convolutional neural network (CNN). The dataset consists of fresh and rotten orange images. The CNN model obtained 78.57% accuracy for fresh and rotten oranges. Denata et al. (2021) developed an image-based classification system for the classification of orange varieties in Sambas Regency. The dataset used consists of 2250 images of oranges, including madu oranges, madu susu and siam oranges. They used AlexNet architecture based on Convolutional Neural Network (CNN) as a deep learning method. They allocated 1575 images for training and 675 images for testing. In total, 10 periods of training were performed, and one model was produced in each period. The best performance was achieved with the model obtained in the 9th period with 94.81% accuracy. Momeny et al. (2022) presented a robust and generalizable method for disease and maturity detection by fine-tuning pre-trained deep learning models. The dataset consists of 1896 images collected from the farm, categorized into 4 classes (immature, semi-mature, mature and black spot diseased). The results show that the fine-tuned ResNet50 model performs best with a learning replication strategy with 99.5% accuracy and 100% F-1 score measurement success when considering black spot diseased images as a positive class. In this study, Kumar et al. (2024), compared machine learning models for classification of orange diseases. They tested K-Nearest Neighbor (k-NN), Convolutional Neural Network (CNN) with MobilNetV2 architecture, Random Forest (RF) and Support Vector Machine (SVM) models on four disease classes, namely Black spot, Mingy disease, Healthy and Green crack. As a result, the CNN model using MobilNetV2 architecture showed the best performance in all disease categories with an accuracy of 98.22.

Machine learning algorithms can make predictions about quality by analyzing the physical and chemical properties of oranges. In this study, orange size, weight, sweetness (Brix), acidity (pH), softness, harvest time, ripeness, color, variety, presence of blemishes and overall quality are input and output data for machine learning models. Using this data, the quality of the orange can be predicted. The dataset used in the study consists of 11 features and 241 data. Random Forest, k-NN and XGBoost machine learning algorithms were used to predict the quality of oranges. Together with the results obtained, this study highlights the potential of machine learning algorithms to optimize quality control processes in the orange industry and increase the production of quality products. With this technology, consumers can get better quality products while producers can increase their productivity. In addition, the study analyzes the important characteristics that affect quality in the available dataset.

2. Materials and Methods

This study evaluates the performance of various machine learning algorithms for determining orange quality. The dataset used contains a total of 241 samples of 11 features. The continuous variables in the dataset were converted into categorical variables using the np.digitize() function. Then the dataset was divided into two parts, 80% training and 20% testing. Random Forest, k-NN and XGBoost algorithms were used to predict the data set. The prediction processes were compared by looking at the performance metrics results. Then, the strengths and weaknesses of the machine learning models were observed by looking at the confusion matrix results. The methodology of the study is shown in Figure 1.

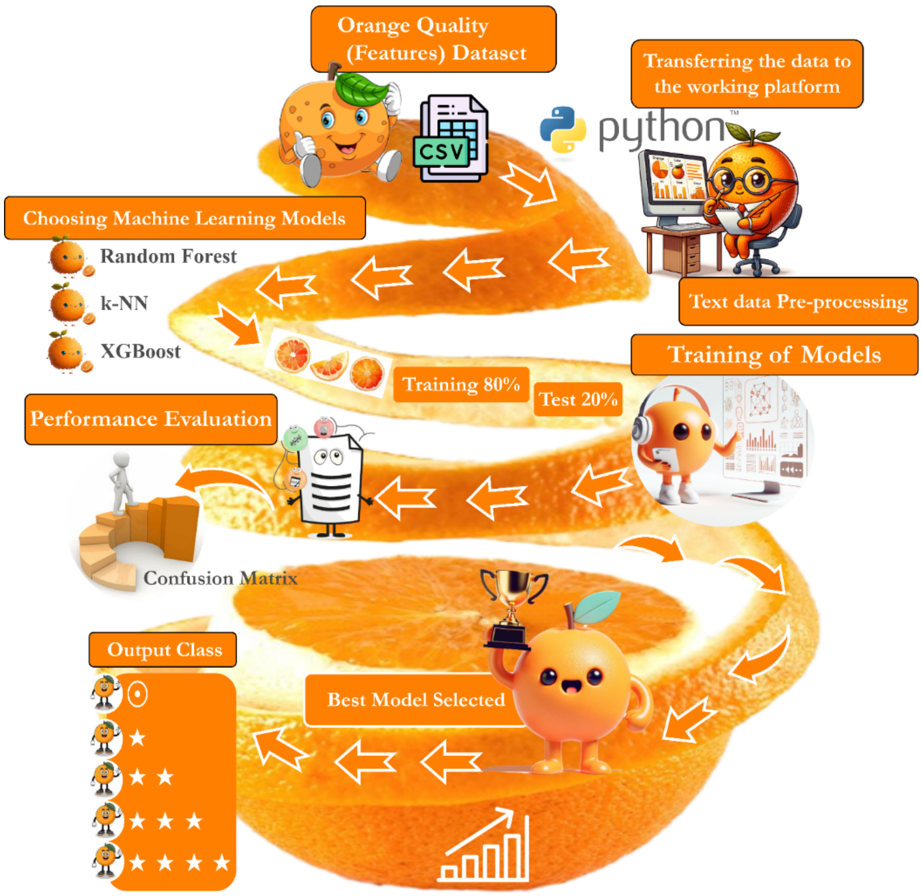

Figure 1. Flow Diagram for The Orange Classification Research

2.1. Dataset Description

The dataset "Orange Quality Analysis Dataset" contains 241 data with 11 different features. In Table 1 the dataset attributes and their description are presented for the Orange dataset. These attributes include numerical characteristics such as size, weight, sweetness (Brix), acidity (pH), softness, time of harvest and degree of ripeness. There are also categorical characteristics such as color, variety, presence of blemishes and overall quality (Shruthi 2024). The first five examples in the dataset are shown in Table 2. The dataset is divided into two separate sets, 80% train and 20% test, for quality estimation. The table describes various features of oranges along with their descriptions. The features include the size of the orange in centimeters, the weight of the orange in grams, the sweetness level measured in Brix, and the acidity level indicated by pH. Additionally, it covers the softness rating on a scale of 1 to 5, the number of days since harvest, and the ripeness rating also on a scale of 1 to 5 (1 represents the lowest and 5 for the highest rating value). Other characteristics listed are the fruit color, the variety of the orange, and whether there are any blemishes (Yes/No). The table also includes the overall quality rating of the orange, which serves as the target variable, rated on a scale of 1 to 5 (1 represents the lowest rating value and 5 represents the highest).

Table 1. Dataset Properties and Descriptions

No.	Features	Descriptions
1	Size	Size of orange in cm
2	Weight	Weight of orange in g
3	Brix	Sweetness level in Brix
4	pH	Acidity level (pH)
5	Softness	Softness rating (1-5)
6	HarvestTime	Days since harvest
7	Ripeness	Ripeness rating (1-5)
8	Color	Fruit color
9	Variety	Orange variety
10	Blemishes	Presence of blemishes (Yes/No)
11	Quality (Target)	Overall quality rating (1-5)

Table 2. First Five Samples of The Dataset

NO	Size (cm)	Weight (g)	Brix (Sweetness)	pH (Acidity)	Softness (1-5)	HarvestTime (days)	Ripeness (1-5)	Color	Variety	Blemishes (Y/N)	Quality (1-5)
0	7.5	180	12.0	3.2	2.0	10	4.0	Orange	Valencia	N	4.0
1	8.2	220	10.5	3.4	3.0	14	4.5	Deep Orange	Navel	N	4.0
2	6.8	150	14.0	3.0	1.0	7	5.0	Light Orange	Cara Cara	N	5.0
3	9.0	250	8.5	3.8	4.0	21	3.5	Orange-Red	Blood Orange	N	3.5
4	8.5	210	11.5	3.3	2.5	12	5.0	Orange	Hamlin	Y (Minor)	4.5

2.2. Random Forest

The Random Forest algorithm is a widely used ensemble method for classification and regression problems in machine learning (Koklu et al. 2012; Yong 2019). Random Forest is a tree-based learning method (Sarica et al., 2017) that combines multiple decision trees to create a more powerful model (Cinar et al. 2023; Yildiz et al., 2024a). The benefits of Random Forest algorithms as redusing the overfitting, aiding in feature selection, ensembility provided to be a popular algorithm and used for the classification. The Random Forest architecture is shown in Figure 2 (Yasin and Koklu 2023). In this study, the parameter settings for the Random Forest algorithm are n_estimators=100, random_state=42 and class_weight='balanced'.

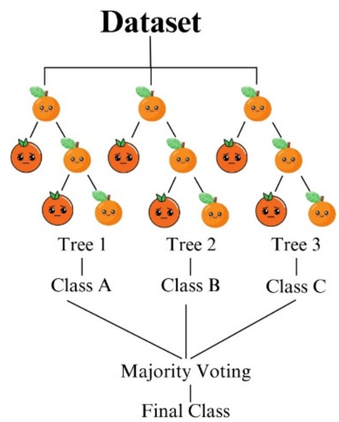

Figure 2. Random Forest Architecture

2.3. k-NN (k-Nearest Neighbors)

k-Nearest Neighbors (k-NN) is a simple and fundamental machine learning algorithm used for classification and regression problems (Cinar et al. 2023; Koklu et al. 2022). k-NN is classified as an instance-based learning method and travels around the surrounding neighbors of data to make predictions (Guo et al. 2003; Yildiz et al. 2024b). The k-NN architecture is shown in Figure 3. In this study, the neighbours parameter is set to 5 for the k-NN algorithm.

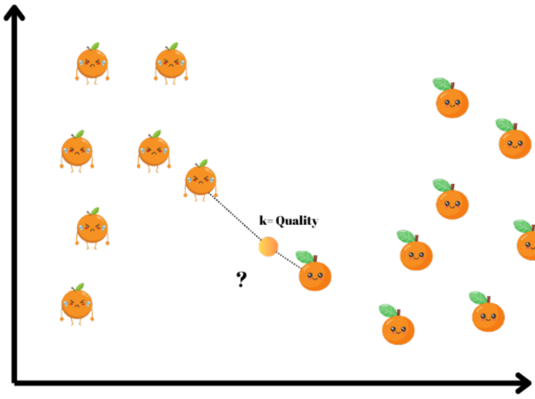

Figure 3. k-NN Architecture

2.4. XGBoost

XGBoost (Extreme Gradient Boosting) can be used for both classification and regression problems. XGBoost is based on an ensemble learning technique called Gradient Boosting. Ensemble learning (Koklu et al. 2014; Tutuncu et al. 2022) is the combination of multiple weak predictors (such as decision trees) into a stronger model. Gradient Boosting works by training weak predictors sequentially, with each predictor trying to correct the errors of the previous predictor (Chen and Guestrin 2016). In this study, default parameter settings were made for the XGBoost algorithm. The XGBoost architecture is shown in Figure 3.

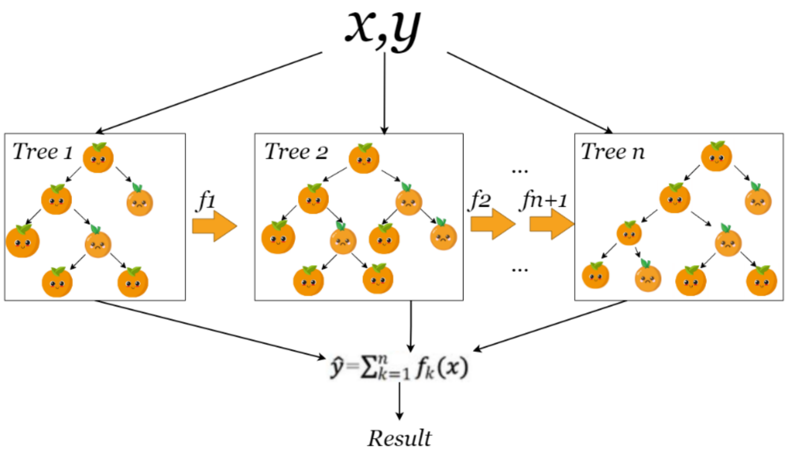

Figure 4. XGBoost Architecture.

2.5. Confusion Matrix and Performance Metrics

Confusion matrix is a table used to evaluate the performance of a classification model. This table provides a comparison between the actual and predicted classes of the model (Tutuncu et al. 2022). Confusion matrix is very useful for visually analyzing classification results and contributes to the calculation of many performance metrics (Gencturk et al. 2024; Heydarian et al. 2022). The two-class confusion matrix and multiclass confusion matrix are shown in Figure 5 and Figure 6. TP is the number of instances that the model correctly classifies as positive, TN is the number of instances that the model correctly classifies as negative, FP is the number of instances that the model incorrectly classifies as positive, FN is the number of instances that the model incorrectly classifies as negative.

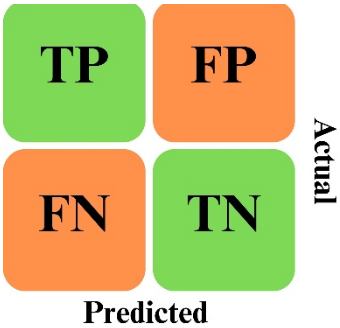

Figure 5. Two Class Confusion Matrix

	Predicted					
	C_1	C_2	C_3	...	C_n	
C_1	T_1	F_{12}	F_{13}	...	F_{1n}	Actual
C_2	F_{21}	T_2	F_{23}	...	F_{2n}	
C_3	F_{31}	F_{32}	T_3	...	F_{3n}	
...	...	...	...	...	...	
C_n	F_{n1}	F_{n2}	F_{n3}	...	T_n	

Figure 6. Muti Class Confusion Matrix

Performance metrics are measures used to evaluate the success of a machine learning model. Various metrics can be used to evaluate the performance of a model (Kursun et al. 2023; Tutuncu et al. 2022). The formulas of performance metrics are shown in Table 3.

Table 3. Formulas of Performance Metrics

Metrics	Formula	
Accuracy	$\dfrac{\sum_{i=1}^{1} \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l}$	(1)
Precision	$\dfrac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fp_i}}{l}$	(2)
Recall	$\dfrac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fn_i}}{l}$	(3)
F1-Score	$\dfrac{2 * \frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fp_i}}{l} * \frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fn_i}}{l}}{\frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fp_i}}{l} + \frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fn_i}}{l}}$	(4)
Mean Square Error	$\frac{1}{N}\sum_{i=1}^{N}(y_i - \acute{y}_i)^2$	(5)

2.6. Correlation Matrix

A correlation matrix is a matrix that measures the relationship between each pair of features in a dataset. This relationship is usually determined using statistical measures such as the Pearson correlation coefficient. The correlation matrix is an important tool for understanding the relationships between variables in a data set, assessing how variables are related to each other, and transforming or modeling variables when necessary (Steiger 1980). A value close to -1 in the matrix indicates a strong negative correlation between the two attributes. If the value is closer to 0, there is a weak correlation between the two attributes. If the value in the matrix is close to 1, there is a strong positive correlation (Lorenzo-Seva and Ferrando 2021).

3. Results

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

In this section, 3 different models created with different machine learning algorithms to determine orange quality are analyzed. The dataset used to determine orange quality consists of 241 samples and 11 features.

Random Forest, k-NN and XGBoost algorithms were used to predict this dataset. The obtained accuracy and Mean Square Error (MSE) values are shown in Table 4 and the correlation matrix result is shown in Figure 6.

Table 4. Accuracy Presentation for Used Machine Learning Algorithms

	Random Forest	k-NN	XGBoost
Accuracy (%)	67.34%	69.38%	63.26%
Precision	73.91%	67.99%	64.52%
Recall	67.34%	69.38%	63.26%
F-1 Score	67.89%	68.43%	63.61%
MSE	0.55	0.55	0.77

In this section, we compare the performance of different machine learning algorithms for predicting orange quality. The results show that the k-NN algorithm stands out in this field by providing the highest accuracy rate 69.38%. This shows that the k-NN algorithm is able to predict orange quality in the best way and is more reliable in this area compared to other algorithms. In addition, the Random Forest algorithm achieved a precision value of 73.91%, indicating that the model has a high probability of accurately predicting orange quality. The Random Forest algorithm ranked second after k-NN with a classification accuracy of 67.34%. This result shows that the Random Forest algorithm is also successful in predicting orange quality. However, the XGBoost algorithm showed a lower performance compared to the other two algorithms with a classification accuracy of 63.26%. This shows that the XGBoost algorithm is not as successful as the other two algorithms in predicting orange quality.

k-NN and Random Forest algorithm obtained an MSE value of 0.55, indicating that the mean squared error between the predictions of the model and the actual values is low. This shows that k-NN and Random Forest algorithms are the algorithms that predict orange quality well. However, the XGBoost algorithm has a higher error rate compared to the other two algorithms with an MSE of 0.77. This again shows that the XGBoost algorithm is not as successful as the other two algorithms in predicting orange quality. Correlation matrix values are given in Figure 7. In Figure 8 the confusion matrix values illustrated.

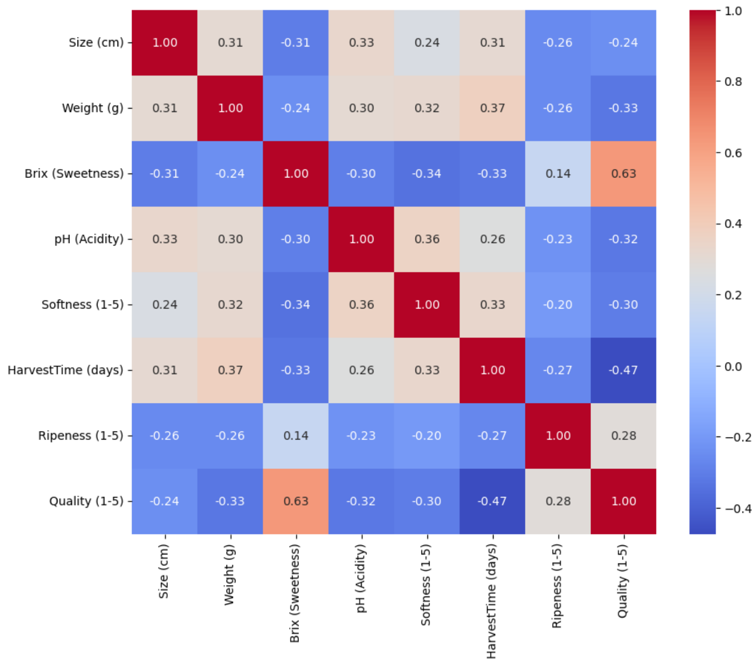

Figure 7. Correlation Matrix Result

When the correlation matrix given in Figure 7 is examined, it is seen that the most important feature affecting quality is Brix (Sweetness) with a correlation value of 0.63. In other words, the higher the taste value of the orange, the higher the quality grade. Then there is a strong and negative relationship between HarvestTime and Quality with a correlation value of -0.47. The correlation values of Weight (-0.33), pH (-0.32) and softness (-0.30) are very close to each other. According to the observations, "Weight," "pH," and "softness" have an inverse effect on the quality of oranges. As the values of these three attributes increase, the quality of the orange decreases. "Size" and "ripeness" have weak correlation values. It cannot be said that they affect orange quality as much as other variables.

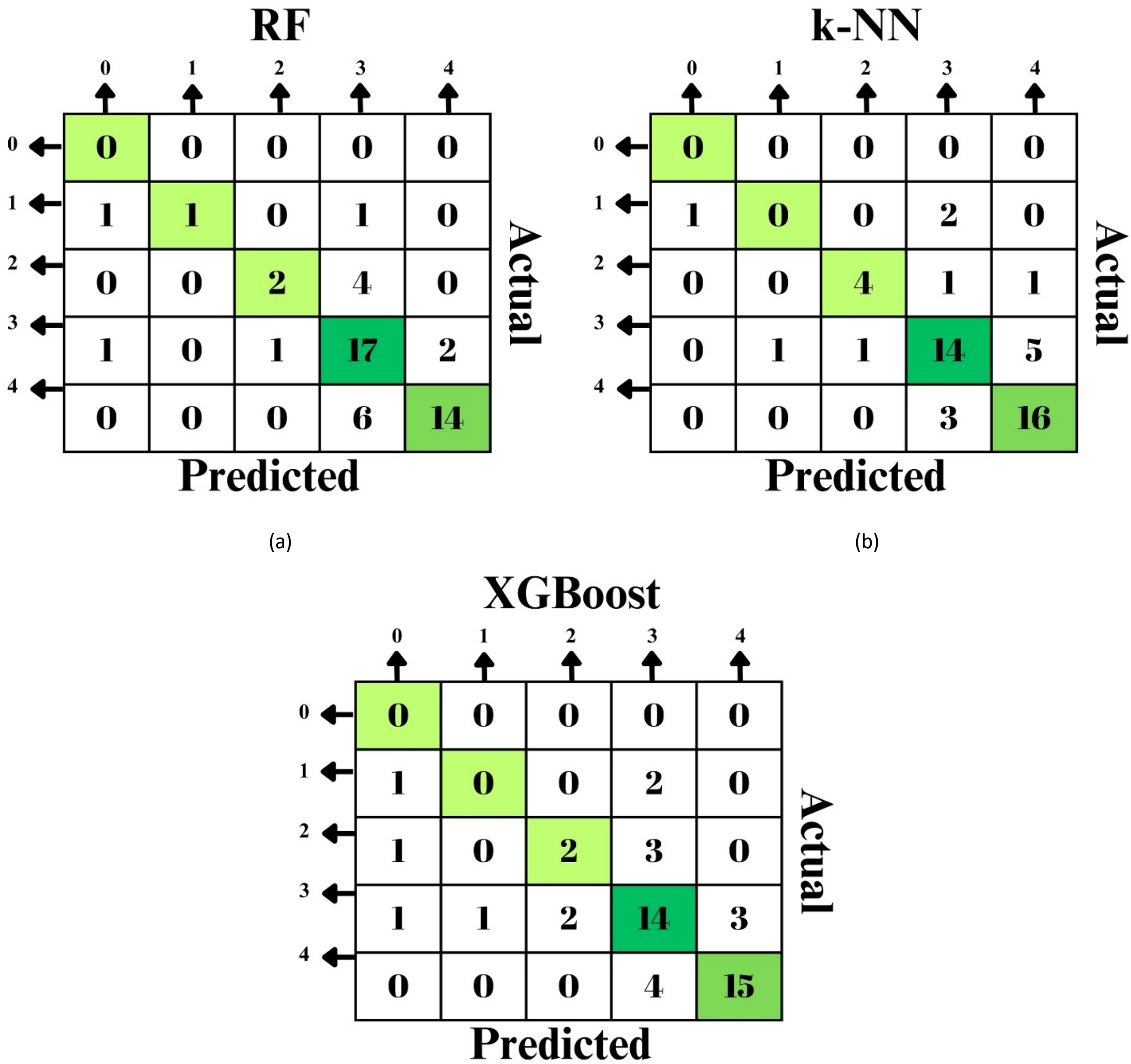

Figure 8. Confusion Matrix Results

When the confusion matrix results are analyzed, it is seen that k-NN is the algorithm that most accurately predicts the quality oranges in the dataset. However, it is observed that Random Forest and XGBoost algorithms confuse oranges with quality values of 2 and 3. This indicates that these two algorithms perform more poorly in distinguishing oranges of certain quality levels.

4. Discussion and Conclusions

Random Forest, k-NN and XGBoost machine learning algorithms were used to make predictions on the orange quality dataset. Each of these algorithms attempted to predict orange quality by capturing different features and patterns in the dataset in different ways. Random Forest used a method of decision trees and combined the prediction of each tree to create a robust prediction model, while k-NN used a similarity-based approach to predict the label of a sample based on its surrounding neighbors. XGBoost used gradient boosting to combine weak prediction models into a strong prediction model.

We used the np.digitize function to convert the continuous variables in the dataset to categorical. We performed this step because we wanted to categorize continuous variables into specific categories when preparing input data for regression or classification models. We categorized each sample categorically according to thresholds.

In the future, orange quality detection systems developed using this technology can scan agricultural fields more effectively and detect poor quality or diseased oranges, thus reducing agricultural labor and time loss and increasing productivity. At the same time, accurate and timely diagnostics can reduce the amount of pesticides needed, which can contribute to the spread of environmentally friendly agricultural practices. Machine learning and artificial intelligence techniques can also offer solutions to more complex agricultural problems. Trait engineering studies can be conducted to identify traits that affect orange quality. Perhaps better results can be obtained by creating new traits or transforming existing traits. Such studies can provide valuable guidance to orange growers and industry experts to make more accurate and reliable decisions.

Based on the results of the study, it seems that machine learning algorithms can be used effectively in predicting orange quality. It is thought that these algorithms can provide important information to orange producers in quality control processes and help them to improve quality

---

Author Contributions: Conceptualization, Talha Alperen CENGEL, Bunyamin GENCTURK, Elham Tahsin YASIN, Osman OZBEK, and Murat KOKLU; methodology, Talha Alperen CENGEL, Bunyamin GENCTURK, Elham Tahsin YASIN, Muslume Beyza YILDIZ, Ilkay CINAR, Osman OZBEK, and Murat KOKLU; software, Talha Alperen CENGEL, Bunyamin GENCTURK, and Elham Tahsin YASIN; validation, Talha Alperen CENGEL, Bunyamin GENCTURK, Elham Tahsin YASIN, Muslume Beyza YILDIZ, Ilkay CINAR, Osman OZBEK, and Murat KOKLU; formal analysis, Talha Alperen CENGEL, and Elham Tahsin YASIN; investigation, Talha Alperen CENGEL, and Elham Tahsin YASIN; writing—original draft preparation, Talha Alperen CENGEL, Bunyamin GENCTURK, and Elham Tahsin YASIN; writing—review and editing, Talha Alperen CENGEL, Elham Tahsin YASIN, Ilkay CINAR, Osman OZBEK, and Murat KOKLU; visualization, Talha Alperen CENGEL, Bunyamin GENCTURK, and Elham Tahsin YASIN; supervision, Ilkay CINAR, Osman OZBEK, and Murat KOKLU. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data can be obtained from the public data website of Kaggle through the given link: https://www.kaggle.com/datasets/shruthiiiee/orange-quality.

Conflicts of Interest: The authors declare no conflict of interest.

References

Asriny DM, Rani S, Hidayatullah AF (2020). Orange fruit images classification using convolutional neural networks. *IOP Conference Series: Materials Science and Engineering.*

Cayuela JA, Weiland C (2010). Intact orange quality prediction with two portable NIR spectrometers. *Postharvest Biology and Technology 58*(2): 113-120. https://doi.org/10.1016/j.postharvbio.2010.06.001

Chen T, Guestrin C (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*.

Cinar I, Taspinar YS, Koklu M. (2023). Development of early stage diabetes prediction model based on stacking approach. *Tehnički Glasnik 17*(2): 153-159. https://doi.org/10.31803/tg-20211119133806

Denata I, Rismawan T, Ruslianto I (2021). Implementation of deep learning for classification type of orange using the method convolutional neural network. *Telematika: Jurnal Informatika dan Teknologi Informasi 18*(3): 297-307. https://doi.org/10.31315/telematika.v18i3.5541

Ganesh P, Volle K, Burks T, Mehta S (2019). Deep orange: Mask R-CNN based orange detection and segmentation. *Ifac-papersonline 52*(30): 70-75. https://doi.org/10.1016/j.ifacol.2019.12.499

Gencturk B, Arsoy S, Taspinar YS, Cinar I, Kursun R, Yasin ET, Koklu M (2024). Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models. *European Food Research and Technology 250*(1): 97-110. https://doi.org/10.1007/s00217-023-04369-9

Guo G, Wang H, Bell D, Bi Y, Greer K (2003). KNN model-based approach in classification. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE, November 3-7, 2003*. Catania, Sicily, Italy.

Heydarian M, Doyle TE, Samavi R (2022). MLCM: Multi-label confusion matrix. *IEEE Access* 10: 19083-19095. https://doi.org/10.1109/ACCESS.2022.3151048

Koklu M, Cinar I, Taspinar YS, Kursun R (2022). Identification of sheep breeds by CNN-based pre-trained InceptionV3 model. *11th Mediterranean Conference on Embedded Computing (MECO)*,

Koklu M, Kahramanli H, Allahverdi N (2012). A new approach to classification rule extraction problem by the real value coding. *International Journal of Innovative Computing, Information and Control 8*(9): 6303-6315.

Koklu M, Kahramanli H, Allahverdi N (2014). A new accurate and efficient approach to extract classification rules. *Journal of the Faculty of Engineering and Architecture of Gazi University 29*(3). https://hdl.handle.net/20.500.12395/30550

Kumar S, Pandey AK, Raghav D, Gupta G, Srivastava V (2024). A deep learning approach for multiclass orange disease classification. *2nd International Conference on Disruptive Technologies (ICDT)*.

Kursun R, Bastas KK, Koklu M (2023). Segmentation of dry bean (*Phaseolus vulgaris* L.) leaf disease images with U-Net and classification using deep learning algorithms. *European Food Research and Technology 249*(10): 2543-2558. https://doi.org/10.1007/s00217-023-04319-5

Lorenzo-Seva U, Ferrando PJ (2021). Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution. *Structural Equation Modeling: A Multidisciplinary Journal 28*(1): 138-147. https://doi.org/10.1080/10705511.2020.1735393

Santiago WE, Lopes GD, Faceto AD, Lobo Júnior A, Caldeira R (2019). Classification of oranges by digital image processing.

Sarica A, Cerasa A, Quattrone A (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. *Frontiers in Aging Neuroscience 9*: 329. https://doi.org/10.3389/fnagi.2017.00329

Shruthi (2024). *Orange Quality Analysis Dataset*. https://www.kaggle.com/datasets/shruthiiiee/orange-quality

Steiger JH (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin 87*(2): 245. https://doi.org/10.1037/0033-2909.87.2.245

Tutuncu K, Cinar I, Kursun R, Koklu M (2022). Edible and poisonous mushrooms classification by machine learning algorithms. *11th Mediterranean Conference on Embedded Computing (MECO)*,

Yasin E, Koklu M (2023). Utilizing Random forests for the classification of pudina leaves through feature extraction with inceptionV3 and VGG19. *Proceedings of the International Conference on New Trends in Applied Sciences*.

Yildiz MB, Hafif MF, Koksoy EK, Kursun R (2024a). Classification of diseases in tomato leaves using deep learning methods. *Intelligent Methods In Engineering Sciences 3*(1): 22-36. https://doi.org/10.58190/imiens.2024.84

Yildiz MB, Yasin ET, Koklu M (2024b). Fisheye freshness detection using common deep learning algorithms and machine learning methods with a developed mobile application. *European Food Research and Technology* 1-14. https://doi.org/10.1007/s00217-024-04493-0

Yong CC (2019). Random forest for image classification. https://hdl.handle.net/10356/78792