sciendo International Journal of Mathematics and Computer in Engineering 2(1) (2024) 1–12



International Journal of Mathematics and Computer in Engineering

https://sciendo.com/journal/IJMCE

Original Study

Prediction modeling using deep learning for the classification of grape-type dried fruits

Md Nurul Raihen[1†], Sultana Akter[2]

[1]Fontbonne University, Department of Mathematics and Computer Science, Saint Louis, MO, 63105, USA

[2]Western Michigan University, Department of Statistics, Kalamazoo, MI, 49008-5278, USA

Communicated by Haci Mehmet Baskonus; Received: 20.08.2023; Accepted: 19.09.2023; Online: 31.10.2023

Abstract

Dried grapes (or Raisins) are among the most frequently grown and consumed cereal crops worldwide. They are also an important source of nutrition and nourishment in a variety of countries including Türkiye, the United States, Greece, etc. In addition to that, raisins consist of 15% water, 79% carbs (including 4% fiber), 3% protein, and very little fat. In our study, there were a total of 900 raisin grains used, with 450 pieces from each type: Kecimen and Besni raisin. Seven morphological features were taken from these images after going through several steps of pre-processing. Since machine learning algorithms can analyze large datasets quickly, automatic classification is made possible. With enough training and testing, machine learning models can attain a high degree of precision in classifying raisin grains. They are able to detect variations in size, shape, color, and texture that would be difficult for humans to detect consistently. Eleven machine learning and five different types of artificial intelligence have been used to classify these features. As part of this study, we look into different machine learning and deep learning methods: GaussianNB, Decision Tree, K-Nearest Neighbor, Random Forest, Support vector machine (SVM), XGBoost, LightGBM, and AdaBoost, Logistic Regression, Artificial Neural Network and Deep Learning Network. Study efficacy is evaluated using standard metrics as F1 score and ROC area under the curve (AUC). Using the caret, H2O, neuralnet, and keras packages, AdaBoost and LightGBM, two of the fourteen models, achieve an accuracy of 90.30% and 98.40%, respectively, and a ROC curve score of around 90%.

Keywords: Raisin grains, AdaBoost, random forest, classification algorithms, LightGBM, machine learning technique.
AMS 2020 codes: 68T09; 68T07; 62H10; 62R07.

1 Introduction

The antioxidants, potassium, fiber, and iron included in raisins make them a healthy snack and a concentrated source of carbs for energy [1]. Globally, raisin output is comparable to that of rice, wheat, and corn in terms of importance to the agricultural industry [2]. Carbohydrates and starch abound in raisins. Because of its low cost and high satiety value, raisins play a crucial role in human diets around the world. Additionally, they have found extensive application in manufacturing. As one of the world's leading producers of grapes, Türkiye is among the world's top producers [3]. About 30% of Türkiye's grape harvest is used for table consumption, 37% for drying, 3% for wine production, and the remaining 30% for other purposes [4].

Machine learning is a cutting-edge technology that can be used to identify and categorize a wide range of conditions, such as fire forest prediction, rice grain classifications, seed classifications and raisin varieties [5, 6].

†Corresponding author.
Email address: nraihen@fontbonne.edu

sciendo
ISSN 2956-7068
doi:10.2478/ijmce-2024-0001
sciendo

Open Access.© 2024 Raihen et al., published by Sciendo.
This work is licensed under the Creative Commons Attribution alone 4.0 License.

This technique is used in a wide variety of fields, from food products to renewable energy [7–9]. The application of machine learning (ML), which makes use of computer algorithms to discover patterns in vast datasets with many complex features, has been substantially accelerated in the field of food industry by recent breakthroughs in big data [10]. ML makes use of computer algorithms to identify patterns in enormous datasets with numerous complex characteristics.

The intelligent application of ML prediction algorithms can aid in an early, more affordable approach to a variety of raisin grains. Our work here contributes to the field by allowing us to forecast raising grain classification using ML methods, which first requires the identification of important characteristics in the raw dataset through preprocessing. The results of this research could potentially make it easier to maintain quick and accurate classification of raisin grains within the framework of any realistic and reliable identification of this type of raisins. The results provided in this research are compared to those that have already been published [11].

The present paper is organized in the following manner: The literature review concerning the utilization of machine learning techniques in the context of raisin grains is exposed in part 2. The dataset utilized for this research endeavor is offered in section 3. The methodology employed in this study is delineated in section 4. Lastly, the findings derived from this investigation are deliberated about in section 5. In sections 6, 7, and 8, we provide our findings experimental results and discuss prospective avenues for future research and conclusion.

2 Literature review

According to a study conducted in 1995 by Valiente et al. in [12], raisins are widely considered to be one of the most important sources of dietary fiber. Martin-Carron et al. in [13] have found that when there was a higher degree of tannin polymerization, there was also a higher inhibitory effect on digestive enzymes. This was the case regardless of the type of tannin. This contributes to the diet being more easily digested than it would be otherwise. Yeung et al. in [14] have demonstrated that Golden Thompson Seedless raisins have significantly higher antioxidant activity in addition to a higher total phenolics content than dipped and sun-dried Thompson Seedless raisins. Additionally, they found that the total phenolics content of these raisins was higher. Because of this discovery, the hypothesis that enzymatic activity plays a role in the reduction of phenolic content and antioxidant activity receives additional support.

When working in the food industry, it is not always simple to differentiate between the various kinds of food and levels of quality that are on the market. This manual process requires a lot of labor, which adds to the expense and extends the amount of time it takes. On the other hand, due to an inefficient method of classifying and sorting the products, there has been a reduction in the total number of products that have been shipped. It is possible that the utilization of an automatic raisin sorting system that is based on machine vision will have the potential to improve product quality, eliminate inconsistencies that are caused by manual evaluation, and lessen the reliance on readily available manpower.

Researchers have already started to utilize ML algorithms in order to carry out a wide variety of computations and conduct data analysis thus leading to relevant findings that will help improve people's lives. The findings of studies that employ machine learning algorithms are beneficial to a wide variety of fields, including business, medicine, science, and mathematics, amongst others.

When it comes to evaluating and figuring out the quality of food, there are a lot of different ways in which traditional methods can be used. Nevertheless, these activities can be both labor-intensive and financially burdensome. Because of these drawbacks, researchers have been working hard on new techniques to swiftly and accurately assess most fundamental qualities of products, like the sweetness of their raisins. One such option is a machine learning system. Machine learning allows us to extract features from images that can be used to quantitatively assess the state of a product's quality [15, 16].

As a result of this, many different aspects of food goods, including raisins, have been studied in recent years utilizing machine learning and image processing techniques. These aspects include color, texture, quality, and size. Several sensory aspects determine the quality of trading mass products. Consumers' decisions and preferences are heavily influenced by aesthetic qualities, making them more crucial [17]. The raisins can be blended with other raisin varieties (such as seeded raisins) and low-quality raisins due to the high profit margins and the inconvenience of standard drying processes.

Okamura et al. in [18] were able to extract characteristics of raisins by using images of the fruit. The naive Bayes method was utilized by them in order to classify these characteristics. These were more effective than the

traditional method of manually classifying raisins, according to Okamura et al. in [18].

Using this approach, Cinar et al. in [4] have photographed the raisins inside the box. The gathered statistical data was then classified using three separate machine approaches. Classification outcomes were obtained with the employment of the LR, MLP, and SVM algorithms. One of these methods, the support vector machine (SVM), has the highest success rate of 86.44 % [4].

Dirik et al. in [19] have proposed the hybrid PSO-ANN approach, which is compared to k-Nearest Neighbor (KNN) and Random Tree (RT) on a dataset of 900 raisin grains from two different categories. The hybrid PSO-NN strategy outperformed the other approaches in the evaluation, achieving a fantastic classification performance with an accuracy rate of 100%. KNN and RT had respective accuracies of 87.39% and %94.91.

Karimi et al. in [15] have developed an expert system to evaluate the consistency and freshness of raisins. The images of 1400 individual raisins were analyzed for their textural characteristics. As a result, the combinations were categorized using an Artificial Neural Network (ANN) and a Support Vector Machine (SVM). Using the top 50 features, the SVM classifier outperformed the ANN in terms of efficiency and accuracy of classification (an average of 92.71%).

For four distinct types of raisins, Mollazade et al. in [16] collected color pictures and extracted 36 colors and 8 shape attributes. They started with 44 different features and narrowed it down to 7 utilizing a correlation-based feature selection process. They have attempted to categorize raisins using artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), and Bayes networks (BN). With the ANN method, they were able to achieve a classification accuracy of 96.33% [20].

Omid et al. in [21] have created a system for imaging and categorizing raisins to extract their size and color characteristics using image processing. They obtained approximately 96% classification accuracy.

Tarakci et al. in [22] have conducted a classification algorithm comparison of KNN and WKNN. The K closest neighbor (kNN) algorithm, the most widely used classification algorithm in machine learning, and the weighted kNN (WKNN) algorithm, which accounts for the importance of each feature's index, are both employed. Using the inverse squared distance $(w = \frac{1}{d^2})$ as the weighting factor produces a weighting that is proportional to the square of the distance. The results showed that the WKNN algorithm was more effective at classification than the kNN.

On the basis of the raisin grain dataset, a system for predicting the classification of raisin grains was developed by Yavuj et al. [23]. Using Machine Learning Methods, the 7-attribute, 2-class data set was acquired as-it was classified using Decision Trees and Random Forest. Analysis results showed that 85.44% and 85.22% accuracy were achieved with 020 Random Forest and Decision Trees, respectively.

Angadi et al. [24] have used the application of MATLAB for the classification of raisins through the use of image processing techniques. Through the use of color and size features, they were able to achieve an average accuracy of 95% in their study.

Khojastehnazhand and other researchers [25] have conducted research on the overall quality of raisins in bulk, using both high-quality and lower-quality raisins in addition to wood. The data model that was utilized contains a division into two modes, the first of which contains six classes and the second of which contains fifteen classes. The support vector machine (SVM) model and the linear discriminant analysis (LDA) algorithm are both examples of the proposed classification methods for the supervised learning category. Based on the results, it can be deduced that the SVM model that incorporates the grey level run length matrix (GLRM) performs better and more accurately.

According to Hu et al. [26], the implementation of classification methods that are founded on data mining (DM) techniques can serve as an efficient instrument for the production of accurate classifier models. There are powerful algorithms for selecting and extracting useful features, as well as training various DM models to adapt to difficult input output mappings [16,21,27]. These algorithms were developed by Kirkos et al. [27,28]. Image processing techniques have been successfully used in recent years to differentiate between the various types and sizes of food [29].

In a different study, Koklu et al. [30] have used the morphological traits that were found by looking at images of pumpkin seeds to carry out classification procedures. They employed five different machine learning techniques LR, MLP, SVM, Random Forest (RF), and k-Nearest Neighbor (k-NN) to carry out classification tasks. According to the findings of the study, the SVM approach, which produced an accuracy percentage of 88.64%, was the most accurate. Table 1 below illustrates the performance of previously stated results.

Table 1 Outcomes from prior studies performances.

Authors	Collection of Datasets (Samples)	Methods that are applied	Model Performance
Cinar et al., 2020	Used the system's camera to snap images of the raisins.	LR, MLP and **SVM**	86.44%
Dirik et al., 2023	900 raising grains and 2 classes	KNN, RT and **PSO-ANN**	100%
Karimi et al., 2017	1400 images of raisins and top 50 features	ANN and **SVM**	92.71%
Mollazade et al., 2012	Four different types of raisins' color pictures were analyzed, yielding 36 colors and 8 form attributes	**ANN**, SVM, DT and NB	96.33%
Omid et al., 2010	Obtaining raisins' size and color characteristics using classification techniques	Image Processing Technique	96%
Tarakci et al., 2021	Database for machine learning at UCI different	KNN and **WKNN**	91.70%
Yavuj et al., 2023	Raisin dataset 2022 from UCI machine Learning repository	**RF** and DT	85.44%
Yu et al., 2011	Separated the data into four groups based on appearance, texture, and wrinkling	SVM	95%
Koklu et al. 2021	Images of pumpkin seeds	LR, MLP, **SVM**, RF, KNN	86.64%

3 Data set

The "Raisin" data set was obtained from the website https://www.muratkoklu.com/datasets/. The two types of grapes used to construct the 900-piece data set were cultivated in Türkiye in equal quantities. On the basis of seven distinguishing features [4], raisins were divided into two broad categories: Kecimen and Besni. In Table 2, we see that 7 morphological characteristics were determined for each raisin.

Table 2 Names and descriptions of dataset attributes.

Attribute Name	Attribute Description
Area	Determine how many pixels are contained within the raisin and return that value.
MajorAxisLength	The maximum length of a line that can be drawn on a raisin.
MinorAxisLength	The minimum length of a line that can be drawn on a raisin.
Eccentricity	An ellipse, which shares the same moments as a raisin, can be described by its degree of roundness.
ConvexArea	Provide the size in pixels of the smallest convex shell that contains the raisin region.
Extent	Provide the fraction of the bounding box's pixels that are within the raisin region.
Perimeter	Its circumference can be determined by measuring the distance between the pixels that make up the raisin's circumference.
Class	Kecimen and Besni raisin.

Figure 1 shows the data for both types of raisin samples, including the minimum, average, and maximum values, as well as the standard deviation, skewness, and Kurtosis.

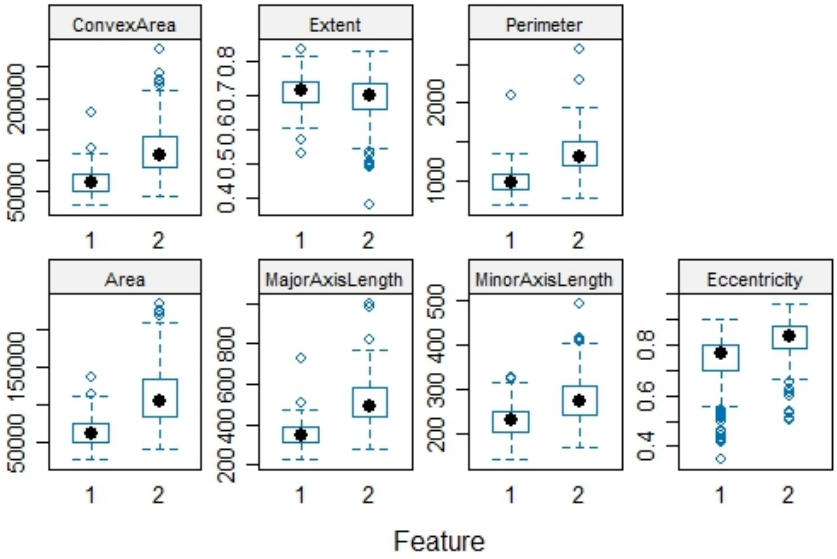


Fig. 1 Box plots of raisin varieties on the features.

4 Research methodology

The primary goal of our investigation is to develop a robust and precise algorithm for the identification of raisin grains. The work entails extensive data gathering, cleaning, processing, training, and testing. The algorithm 1 depicts the procedures involved in machine learning classifiers.

Algorithm 1 Decision Making-ML routine Algorithm

Start
Require: $n \geq 0$ data set collection
Ensure: Training of machine learning classifier
while Evaluation of classifier with dataset? **do**
if Yes **then**
Prediction for raisin grains based on machine learning
else if No **then**
Parameter tuning
end if
end while

Predictions can be made after finishing the three phases of gathering datasets, training machine learning classifiers, and evaluating those classifiers. In this step, the data is divided into training and testing data in a 70-30 ratio. As a result, 70% of the data is used for training, with the remaining 30% used to test the performance of the previously developed model. The more training data there is, the easier it is to learn from. As shown in the flow chart, the methodology of decision making system is based on supervised machine learning classifiers. As shown in the flow chat, Figure 2 shows that the methodology is observed.

sciendo

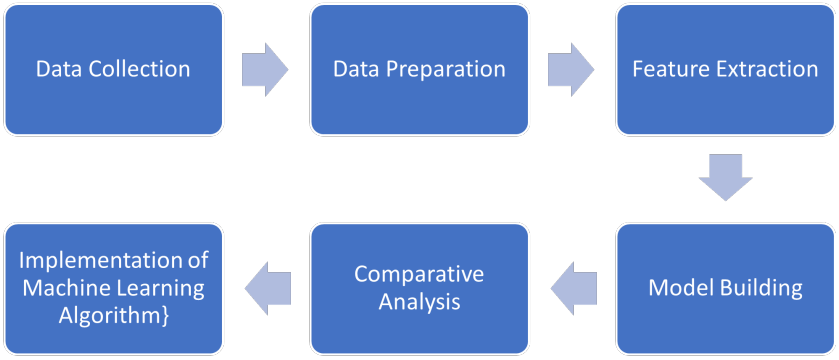


Fig. 2 Decision Making Methodology.

4.1 Dataset collection

We analysed data from the raisin grains dataset available in the open UCI machine learning repository [31]. There were 450 occurrences of raisin grains, and 8 attributes were extracted from them based on their physical attributes, nutritional content, quality criteria, origin, and production. There are no missing variables in the dataset, and all feature values are recorded to four significant digits and include multivariate properties that are relevant to classification tasks. For the two-sample distribution, we have 450 Kecimen and 450 Besni. In order to accurately anticipate raisin grains, it is necessary to use attributes that accurately determine the symptoms.

4.2 Data pre-processing

The variables in the dataset each have each their own unique distribution of values, which may make it more challenging for the models to produce accurate results. Data reduction to a single order improves precision when there is a substantial disparity between the data. To do this, min-max normalization was employed to uniformly distribute the attributes present in the data sets utilized in the analysis between 0 and 1. Equation (1) provides the formula for min-max normalization [32].

$$x' = \frac{x - \min(x)}{max(x) - min(x)}. \tag{1}$$

Figure 3 illustrates the correlation between each of the morphological features.

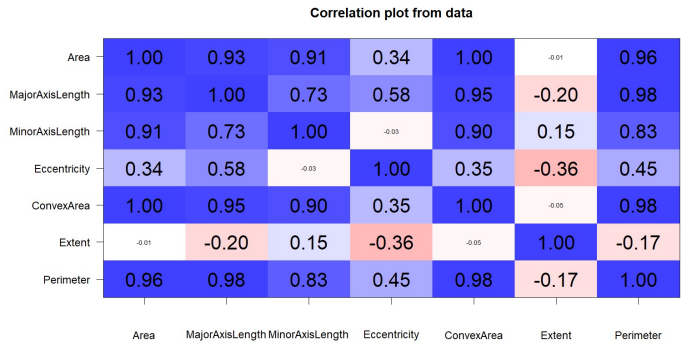


Fig. 3 Correlation matrix for the Raisin dataset.

4.3 Validation process

When working with a certain dataset, it is paramount importance to choose a validating method that is suitable for the situation. Because it generates reliable findings, hold-out validation is frequently the method of choice when dealing with huge datasets [33]. In this investigation, we utilized the hold-out validation strategy to test 30% of the dataset while simultaneously training 70% of it. With the help of this validation process, we were able to compute performance metrics for each machine learning approach, such as precision, recall, and F1-score. In the result analysis section devoted to the examination of the results, there is a comprehensive presentation of the performance indicators and output graphs [34]. The complete research process is illustrated in a step-by-step flowchart.

4.4 Flowchart of proposed methodology

Reading a variety of papers and becoming familiar with the nature of the work involved is the first step in gaining domain expertise [35]. The proposed approach is applied to the 8-column dataset. Eleven different machine learning methods, including five deep learning-based neural networks, are used. Using the R programming language, the code is written in a Jupyter Notebook. In this study, the LightGBM Classifier achieves a 98.40% accuracy and a 92.57% ROC AUC with adequate parameter tuning utilizing cross-validation in the training set. The flow chart depicts the steps in our methodology workflow, Figure 4.

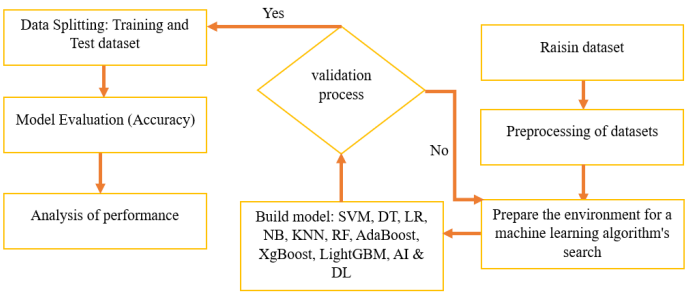


Fig. 4 The proposed steps for creating the machine learning classification model.

4.5 Performance evaluation

In order to fortify machine learning classifications, cross validation has become a regular method of evaluation. In cross-validation, a dataset is split into k groups for separate training and testing purposes. The system is trained using the remaining sets while one set is used as a test set. This is done until k is reached, and then the system is put through its paces [36]. The following Figure 5 depicts the results of a cross-validation [4].

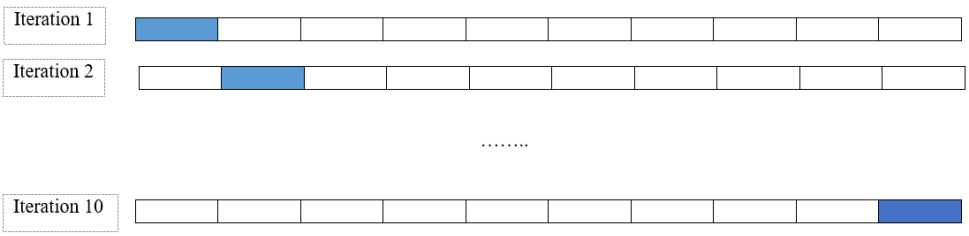


Fig. 5 10-fold cross Validation of the Training Dataset.

4.6 Training of machine learning classifiers

Classification is the process of training supervised machine learning classifiers. The task of constructing a classifier from a labelled raisin dataset serves as the basis for training. An effective classifier for predicting raisin dataset is generated by an algorithm based on supervised machine learning (C-Section Classification Database Report, UCI Machine Learning Repository, University of California, Irvine, USA) [31].

5 Evaluation of classifiers

Classification algorithms can be evaluated using standard measures including accuracy, specificity, sensitivity, recall, and F1 score. The confusion matrix is used to determine these values. Confusion matrices are tables used to describe the performance of a classification model on a set of test data that is already well-known. There are 4 variables that make up the confusion matrix. True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are the four possible outcomes. The format of the confusion matrix is displayed in Table 3 [37].

Table 3 Confusion matrix.

	Predictive Positive	**Predictive Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

In order to perform an accurate evaluation of the machine learning classifiers, important measures were collected from the confusion matrix. In addition to the correct classification rate or accuracy, other metrics such as True Positive Rate (TPR), False Positive Rate (FPR), Precision, Recall, F1 score, and ROC area were used to evaluate the machine learning classifiers. These metrics were used to evaluate the classifiers' performance (See Table 4).

Table 4 The metric used in evaluating the performance of machine and deep learning classifiers.

Performance Measure Name	Formula
Correct Classification Rate	$CCR = \frac{TP+TN}{TP+FP+FN+TN}$
Precision	$PPV = \frac{TP}{TP+FP}$
Recall	$= \frac{TP}{TP+FN}$
F1-score	$F_1 = \frac{2TP}{2TP+FP+FN}$
True Positive Rate	$TPR = \frac{TP}{TP+FN}$
False Positive Rate	$FPR = \frac{FP}{TN+FP}$
Specificity	$TPR = \frac{TN}{TN+FP}$
Negative Predictive Value	$NPV = \frac{TN}{TN+FN}$

6 Experimental results

The data were computed and reviewed to determine the optimum technique for predicting varieties of raisin grains concerns. The performance criteria such as accuracy, precision, recall, F1-score, and area under the curve (AUC) were used to evaluate eleven machine learning and five deep learning methods.

It is insufficient to rely just on accuracy as a performance measure for a model. The AUC value is an important indicator for assessing a model's performance and ability to discriminate across classes. Machine learning algorithms such as SVM, DT, LR, NB, RF, AdaBoost, XgBoost, LightGBM, and Deep learning were

used to achieve performance values. In all models used, the cross-validation k value was set to 10 in Table 5 and 5 in Table 6. Table 5 and 6 show the confusion matrix for all of the algorithms employed in the study.

At various points along the probability curve, the True Positive Rate and False Positive Rate are calculated. The performance of GaussianNB, DecisionTree, K-Nearest Neighbor, RandomForest, SVM, XGBoost, LightGBM, AdaBoost, Logistic Regression, Conviluation Neural Network, Radiaal Basis Function Network, Recurrence Neural Network, Artificial Neural Network, and Deep Neural Network ML algorithms on the raisin grains dataset is represented in Table 5.

Table 5 Model performance with a 10-fold cv.

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC-ROC score (%)
Support Vector Machine	87.78%	90.37%	85.92%	88.00%	87.88%
Decision Tree	86.3%	91.85%	82.67%	87.02%	86.75%
Logistic Regression	87.41%	88.15%	86.86%	87.5%	87.42%
Naive Bayes	84.81%	90.37%	81.33%	85.62%	85.25%
K-nearest Neighbours	87.04%	90.37%	84.72%	87.46%	87.20%
Random Forest	86.30%	88.15%	85.00%	86.55%	86.35%
AdaBoost	**90.30%**	87.41%	85.51%	86.45%	86.31%
XgBoost	83.70%	85.19%	82.73%	83.95%	83.73%
LightGBM	**98.40%**	97.41%	89.19%	93.10%	92.57%
Convolution Neural Net.	81.35%	83.15%	79.08%	81.23%	-
Radial Basis Function Net.	83.41%	84.57%	81.73%	83.78%	-
Recurrence Neural Net.	73.94%	74.52%	72.50%	73.11%	-
Artificial Neural Net.	65.00%	67.01%	64.86%	65.35%	-
Deep Neural Net.	69.00%	70.56%	67.19%	68.89%	-

Table 6 Model performance with a 5-fold cv.

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC-ROC score (%)
Support Vector Machine	88.52%	91.85%	86.11%	88.88%	88.69%
Decision Tree	85.93%	88.89%	83.92%	86.33%	86.05%
Logistic Regression	86.67%	88.89%	88.37%	86.95%	86.74%
Naive Bayes	85.93%	91.11%	82.55%	86.61%	86.32%
K-nearest Neighbours	86.30%	91.11%	83.11%	86.92%	86.64%
Random Forest	85.56%	89.63%	82.88%	86.12%	85.79%
AdaBoost	**89.15%**	92.59%	85.03%	82.41%	88.45%
XgBoost	83.33%	87.41%	80.82%	83.98%	83.56%
LightGBM	**96.31%**	97.11%	88.21%	93.35%	91.83%
Convilution Neural Net.	78.51%	82.22%	75.54%	78.91%	-
Radial Basis Function Net.	81.11%	83.31%	79.02%	82.37%	-
Recurrence Neural Net.	71.29%	74.10%	68.53%	67.40%	-
Artificial Neural Net.	64.81%	65.34%	62.13%	63.92%	-
Deep Neural Net.	68.27%	70.59%	66.19%	67.78%	-

7 Discussion

According to the model performance with a 10-fold cv compared to 5-fold cv, LightGBM has the maximum accuracy score of 99.00%, while Artificial Neural Networks has the lowest accuracy score of 64.81%, as shown in Figure 6. Furthermore, AdaBoost performed well, scoring 90.30%. The SVM, Logistic regression, and Knn values are 87.78%, 87.41%, and 87.04%, respectively.

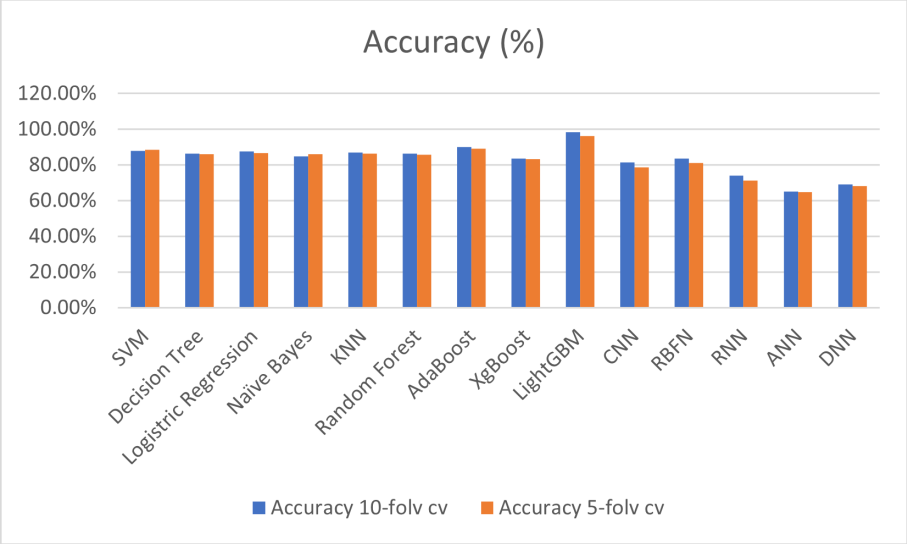


Fig. 6 Accuracy of the models in predicting the Raisin dataset using a 10-fold and 5-fold cv.

How well the models can tell the difference between the positive and negative classes is quantified by the AUC. The AUC can take on values between 0 and 1, with higher values indicating more favorable outcomes. Incorrect results are indicated by an AUC of 0.5, while a perfect score of 1 shows that the test was absolutely accurate. AUCs between 0.7 and 0.8 are generally accepted as being below the raisin classification or variety detection threshold. A performance rating of 0.8 to 0.9 is considered outstanding, and a rating of 0.9 or above is exceptional [38].

8 Conclusions

It can be difficult to obtain a large and diverse dataset of identified raisin images for training. Manual work is required to gather, classify, and annotate images appropriately, which can be time-consuming and expensive. Identifying meaningful indicators that effectively differentiate between different quality aspects of raisin grains might be difficult. It is critical for accurate classification to develop adequate feature extraction methods that capture the key properties of raisins.

The dataset for this study was gathered from Turkish raisin grains. Using a real-world dataset of raisin types, this study investigates how to forecast raisin kinds grown in Türkiye or any other country. Machine learning and deep learning are used in this study to create a model for predicting the seed of a raisin dataset. Several studies were examined to assess objectives, analyze the methods used, and identify any problems in the research.

Moreover, during the process of our research, we made use of fourteen different machine learning algorithms and evaluated the level of accuracy achieved by each. When we looked at all of the different methods for predicting raisin grain based on the information that we utilized, LightGBM had the greatest accuracy of 98.40%.

The Supervised machine learning technique has been utilized successfully in a variety of applications, including women career and work happiness in constructing appropriate algorithms for predicting employees' job satisfaction with a variety of individual characteristics [39].

The work can be expanded in the future by gathering more data from different regions of the world to analyze. The accuracy of the machine learning model predictions would improve with the addition of more data. Some additional deep learning approaches may also be used.

9 Declarations

9.1 Conflict of interest:

The authors hereby declare that there is no conflict of interests regarding the publication of this paper.

9.2 Funding:

The authors hereby declare that there is no funding regarding the publication of this paper.

9.3 Author's contribution:

M.N.R.-Writing original draft, Writing review editing, Methodology; S.A.-Writing review editing, Analysis. All authors read and approved the final submitted version of this manuscript.

9.4 Acknowledgement:

The authors deeply appreciate the anonymous reviewers for their helpful and constructive suggestions, which can help further improve this paper.

9.5 Data availability statement:

All data that support the findings of this study are included within the article.

9.6 Using of AI tools:

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

References

[1] Cinar I., Koklu M., Classification of rice varieties using artificial intelligence methods, International Journal of Intelligent Systems and Applications in Engineering, 7(3), 188–194, 2019.
[2] Şapaloğlu A., The Structure of the Marketing Channels and Rice Marketing Margins in the Chain of Rice Production-Consumption: An Example of Edirne Province, MSc. Thesis, Namik Kemal University, Tekirdağ, Türkiye, 2015.
[3] Semerci A., Kiziltuğ T., Çelik A.D., Kiraci M.A., General overview of viticulture in Turkey, Journal of Mustafa Kemal University Faculty of Agriculture, 20(2), 42–51, 2015.
[4] Cinar I., Koklu M., Tasdemir S., Classification of raisin grains using machine vision and artificial intelligence methods, Gazi Journal of Engineering Sciences, 6(3), 200–209, 2020.
[5] Sammut S.J., Crispin-Ortuzar M., Chin S.F., Provenzano E., Bardwell H.A., Ma W., Cope W., Dariush A., Dawson S.J., Abraham J.E., Dunn J., Hiller L., Thomas J., Cameron D.A., Bartlett J.M.S., Hayward L., Pharoah P.D., Markowetz F., Rueda O.M., Earl H.M., Caldas C., Multi-omic machine learning predictor of breast cancer therapy response, Nature, 601(7894), 623–629, 2022.
[6] Chaulagain R.S., Liza F.T., Chunduri S., Yuan X., Lang M., Achieving the performance of global adaptive routing using local information on dragonfly through deep learning, ACM/IEEE SC Tech Poster, 1–3, 2020.
[7] Yao Z., Lum Y., Johnston A., Mejia-Mendoza L.M., Zhou X., Wen Y., Aspuru-Guzik A., Sargent E.H., Seh Z.W., Machine learning for a sustainable energy future, Nature Reviews Materials, 8(3), 202–215, 2023.
[8] Khan M., Das R.C., Casey J., Reese B.L., Akintunde B., Pathak A.K., Near room temperature magnetocaloric properties in Ni deficient $(Mn_{0.525}Fe_{0.5})$ $Ni_{0.975}$ $Si_{0.95}AI_{0.05}$, AIP Advances, 12(035227), 1–5, 2022.
[9] Stein I., Raihen M.N., Convergence rates for Hestenes' Gram–Schmidt conjugate direction method without derivatives in numerical optimization, AppliedMath, 3, 268–285, 2023.
[10] Baykan Ö.K, Babalik A., Botsali F.M., Recognition of wheat species using artificial neural network, 4 International Symposium on Advanced Technologies, Konya, Türkiye, 28–30 September 2005, 188-190.
[11] Amin M.Z., Ali A., Performance evaluation of supervised machine learning classifiers for predicting healthcare operational decisions, Technical Report, DOI: 10.13140/RG.2.2.26371.25127, 1-7, 2017.
[12] Valiente C., Arrigoni E., Esteban R.M., Amado R., Grape pomace as a potential food fiber, Journal of Food Science, 60(4), 818–820, 1995.
[13] Martin-Carron N., Garcia-Alonso A., Goñi I., Saura-Calixto F., Nutritional and physiological properties of grape pomace as a potential food ingredient, American Journal of Enology and Viticulture, 48(3), 328–332, 1997.
[14] Yeung C.K., Glahn R.P., Wu X., Liu R.H., Miller D.D., In vitro iron bioavailability and antioxidant activity of raisins,

Journal of Food Science, 68(2), 701–705, 2003.
[15] Karimi N., Kondrood R.R., Alizadeh T., An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms, Measurement, 107, 68–76, 2017.
[16] Mollazade K., Omid M., Arefi A., Comparing data mining classifiers for grading raisins based on visual features, Computers and Electronics in Agriculture, 84, 124–131, 2012.
[17] Zareiforoush H., Minaei S., Alizadeh M.R., Banakar A., A hybrid intelligent approach based on computer vision and fuzzy logic for quality measurement of milled rice, Measurement, 66, 26–34, 2015.
[18] Okamura N.K., Delwiche M.J., Thompson J.F., Raisin grading by machine vision, Transactions of the ASAE, 36(2), 485–492, 1993.
[19] Dirik M., Improving raisin grains classification with a hybrid PSO-NN approach, 1st International Conference on Contemporary Academic Research (ICCAR 2023), 17–19 May 2023, Konya, Türkiye, 1, 34–41, 2023.
[20] Raihen N., Akter S., Forecasting breast cancer: A study of classifying patients' post-surgical survival rates with breast cancer, Biomedical Journal of Scientific and Technical Research, 50(1), 41310–41319, 2023.
[21] Omid M., Abbasgolipour M., Keyhani A., Mohtasebi S.S., Implementation of an efficient image processing algorithm for grading raisins, International Journal of Signal and Image Processing, 1(1), 31–34, 2010.
[22] Tarakci F., Ozkan I.A., Comparison of classification performance of kNN and WKNN algorithms, Selcuk University Journal of Engineering Sciences, 20(02), 32–37, 2021.
[23] Unal Y., Kaplan H., Bektas Y., Caglar M.B., Classification of raisin grains variety using some machine learning methods, New Trends in Computer Sciences, 1(1), 62–69, 2023.
[24] Angadi S.A., Hiregoudar N., A cost effective algorithm for grading raisins using image processing, International Journal of Recent Trends in Engineering Research, 2, 2455–2457, 2016.
[25] Khojastehnazhand M., Ramezani H., Machine vision system for classification of bulk raisins using texture features, Journal of Food Engineering, 271, 109864, 2020.
[26] Hu M.H., Dong Q.L., Liu B.L., Malakar P.K., The potential of double K-means clustering for banana image segmentation, Journal of Food Process Engineering, 37(1), 10–18, 2014.
[27] Kirkos E., Spathis C., Manolopoulos Y., Data mining techniques for the detection of fraudulent financial statements, Expert Systems with Applications, 32(4), 995–1003, 2007.
[28] Raihen M.N.I., A Bifurcation Phenomenon of Regularized Free Boundary Problems of Two-Phase Elliptic-Parabolic Type, Ph.D. Thesis, Wayne State University, USA, 2022.
[29] Francis F.J., Clydesdale F.M., Food Colorimetry: Theory and Applications, AVI Publishing, USA, 1975.
[30] Koklu M., Sarigil S., Ozbek O., The use of machine learning methods in classification of pumpkin seeds (Cucurbita pepo L.), Genetic Resources and Crop Evolution, 68(7), 2713–2726, 2021.
[31] U.C.M.Learning, "Raisin Grains Dataset", March, 2020 [online], available: https://www.muratkoklu.com/datasets/ , Accessed: 05 July, 2023.
[32] Jain S., Shukla S., Wadhvani R., Dynamic selection of normalization techniques using data complexity measures, Expert Systems with Applications, 106, 252–262, 2018.
[33] Raihen M.N., Akter S., Tabassum F., Jahan F., Sardar M.N., A statistical analysis of positive excess mortality at Covid-19 in 2020-2021, Journal of Mathematics and Statistics Studies, 4(3), 07–17, 2023.
[34] Raihen M.N., Akter S., Sardar M.N., Food satisfaction among students: A study of present public university students in Bangladesh, Journal of Mathematics and Statistics Studies, 4(1), 1–18, 2023.
[35] Khoshroo A., Arefi A., Masoumiasl A., Jowkar G.H., Classification of wheat cultivars using image processing and artificial neural networks, Agricultural Communications, 2(1), 17–22, 2014.
[36] Bergmeir C., Benítez J.M., Forecaster performance evaluation with cross-validation and variants, 2011 11th International Conference on Intelligent Systems Design and Applications, IEEE, Cordoba, Spain, 849–854, 2011.
[37] Altay O., Ulas M., Alyamac K.E., Prediction of the fresh performance of steel fiber reinforced self-compacting concrete using quadratic SVM and weighted KNN models, IEEE Access, 8, 92647–92658, 2020.
[38] Omar N., Al-zebari A., Sengur A., Deep learning approach to predict forest fires using meteorological measurements, 2021 2nd International Informatics and Software Engineering Conference, IEEE, Ankara, Türkiye, 1–4, 2021.
[39] Raihen M.N, Akter S., Sardar M.N., Women's career challenges and opportunities (A study of career and job satisfaction among Bangladeshi Women), Academic Journal of Research and Scientific Publishing, 5(51), 05–22, 2023.