BABT

Brazilian Archives of **Biology** and **Technology**

Article - Agriculture, Agribusiness and Biotechnology

Rice Classification and Quality Detection Success with Artificial Intelligence Technologies

Gökalp Çınarer[1]
https://orcid.org/0000-0003-0818-6746

Nizamettin Erbaş[2*]
https://orcid.org/0000-0002-6379-3023

Abdurrahman Öcal[3]
https://orcid.org/0000-0002-6960-7156

[1]Yozgat Bozok University, Faculty of Engineering, Yozgat, Turkey; [2]Yozgat Bozok University, Yozgat Vocational School, Yozgat, Turkey; [3]Yozgat Bozok University, Yozgat Vocational School, Yozgat, Turkey.

HIGHLIGHTS

- Accelerating the process of product classification and to classifying correctly crops by qualities.

- Provides the opportunity to offer products in accordance with different consumer groups and market conditions.

- It is views that a very high classification result of 99.18% is accomplished by machine learning algorithms analysis values.

- This model shows that it can also be applied to other crops.

Abstract: Rice is the most consumed and the most traded food in the world, and so it is very important for it to be classified correctly by its qualities. In this study, the success situation in the classification of rice by qualities with information technologies systems was aimed. In the study, the feature selection process was applied by making statistical analyzes of the features obtained from the images of two different rice species. The classification process was carried out with five different Artificial Intelligence (AI) algorithms using 6 different morphological features. When the results and performance values are examined, it was viewed that the Support Vector Machine (SVM) algorithm gave the highest accuracy in classification with 93.53%. The obtained Area Under the Curve (AUC) values showed that a very high classification result of 99.18% was accomplished. It was detected that morphological features were very important parameters in classifying rice varieties with the AI algorithms. It is accepted that this study will be important in accelerating the process of product classification which is one of the main components of agricultural marketing and classifying correctly crops.

Keywords: agribusiness; agricultural marketing; artificial intelligence; crop classification; quality detection.

INTRODUCTION

Rice is a cereal product that is produced and consumed on a global scale, and is the most grown after wheat and corn. It is the world's principal energy and protein source. It is produced mostly in Asian countries at the global level. World rice production is 498 million tons in the 2019/20 period, 90% of which belongs to Asian countries [1].

Asia's share in world rice production is increasing every year due to traditional dietary habits, climate and land suitability, and population sizes. The fact that more than 60% of the world population lives in Asia and 1/3 of the total rice consumption is subject to foreign trade increases the economic importance of the rice. Apart from its economic contribution such as employment and foreign income, rice has many benefits in terms of nutrition and health. It has many benefits to the human body. It regulates intestinal systems, regulates blood sugar, prevents aging and is a source of vitamin B1. Rice is also very rich in carbohydrates and starch and is widely used in industry [2].

The continuous increase in the world population increases the importance of food even more [3]. Rice is among the most consumed foods in the world, it is necessary to produce more. It is the most traded crop, and it underlies the economy of producer countries. The manual classification of rice, which is the subject of trade, in machines today is costly and takes a lot of time. In this case, AI or deep learning methods used in all fields can be used to classify rice with ease. This method has many benefits in terms of the agricultural economy. These are;

Contributions at the individual level: Supplying better and quality rice for the consumer at a cheaper price, giving better prices to the producer and faster marketing, avoiding waste of resources through specialization increase the income of the producer.

Contributions at the enterprise level: Reducing marketing costs, improving the distribution channels, facilitating sales processes, and increasing sales.

Contributions at the national economy level: Sharing justly of national income, the best allocation of scarce resources, and increasing national income.

There are various quality criteria for crops. These quality features are also valid for rice in the same way. These are; matters such as physical appearance, size cooking properties, aroma, taste and smell of crop [4]. From the point of view of the end consumer, the first thing feature that comes to mind is the physical appearance feature of the rice varieties sold as packages on the market shelves. It is seen that the need for technological methods has increased due to the difficulty and time-consuming separation of rice, especially for species with a high consumption volume after production, the difficulties experienced in determining the varieties and the problems experienced in classifying them according to various quality elements.

The quality classification of crops has various contributions in terms of marketing [5]. These are;

- protects the producer and consumer from deceptions,
- reduces or prevents friction to arise between the buyer and the seller,
- facilitates communication and agreement in marketing,
- provides convenience in comparing the market price,
- provides convenience and effectiveness in the advertisement of the product,
- provides the development of the industry and trade-related to the product,
- provides the opportunity to offer products in accordance with different consumer groups and market conditions.

Even if the other production processes of the rice product are perfect, the classification not being in accordance with the market demands will cause a decrease in sales. Classification of rice with modern technological methods instead of manual methods will provide great advantages in terms of trade and cost. The classical and manual methods used in the classification of cereals force the people who will perform these tasks in terms of both time and cost. With AI technologies, these difficulties are completely eliminated. Therefore, advanced technologies are needed in the classification and quality determination of agricultural products with high accuracy [6, 7].

Classification of rice using AI technologies will offer a more advantageous classification. Today, modern technologies such as AI, deep learning, and machine learning used in very much fields are also applied with great success at every stage of agricultural activity [8]. From the cultivation of the soil and the care of the crops to the quality control and marketing processes, AI use is gradually increasing. In this context, AI is of much essence for the sustainability of agriculture. AI technologies, which adapt quickly to agricultural processes, especially in developed countries, apply by most producers. AI applications provide with a chance

for more profitable and productive agriculture by detecting the problems in agricultural production and marketing [9].

The study consists of six main headings, including the introduction. In the second part, literature studies are given. Materials and methods after the related works in which general information about the subject is given are presented in the third section. In materials and methods, the materials used in the study and the methods applied are explained. Results are given in the fourth section of the study. In this section, information on the reasons for producers' abandonment of farmland and sustainability is presented. The discussion, in which all the results of the research are compared and explained with the results of other studies, is given in the fifth section. The sixth chapter consists of conclusions.

This study aims to determine the success of the classification of agricultural products (rice) with AI technologies and its importance in terms of the agricultural economy.

LITERATURE REVIEW

[10] classified basmati rice images according to their morphological and textural features with a backpropagation-based neural network. In another study, [11] used basmati rice, classified with a method based on principal component analysis. When the earlier studies are examined, product features were revealed by various image processing techniques of products, using shape and color features as well as their morphological features [12]. In addition, rice types have been classified by different methods such as color, shape, visuality and morphological status. The results obtained in the study support the effectiveness of the importance of model and methodology in the classification of these species in obtaining successful results. In recent years, these methods have been replaced by new algorithms.

The methods and models applied for rice classification in the literature have increased with the developing technology.

[13] used the Sparse Representation-Based Classification (SRC) method to classify rice images they obtained with a microscope. In the study made with the obtained features, they distinguished the rice varieties with 89.1% accuracy. [14] in their study in Pakistan, classified rice quality and varieties according to some morphological characteristics and attributes by the Principal Component Analysis (PCA) method. In the study conducted with different rice varieties, they achieved 92.3% classification and 89.5% quality analysis success. [15], classified seven different features with machine learning algorithms for rice classification with the data set used in this study. The highest accuracy was obtained with the (Logistic Regression) LR algorithm with a rate of 93.02%.

Another study [16] used morphological, shape and color data obtained from 75,000 images in their study with images obtained from five different rice species. They have achieved very high results with these features with the Convolutional Neural Network (CNN) model.

[17] determined the physical properties of rice species with machine vision techniques in their study, which stated that grain quality is very important for human health. They graded each of the rice grains with the data they obtained. [18] made a proposal to detect the chalky area in rice by image processing to improve rice quality. They evaluated the quality of rice in terms of length, width and area in their study, where they stated that it is difficult to find all the properties in rice grains. They concluded that quality was important in all three cases.

The studies in the literature and the results obtained, the studies in which classification algorithms and statistical analyzes are evaluated in a common model are limited. [19] distinguished rice types with 90.5% accuracy in their study in four classes. In another study [20] classified rice grains using the Weka portal. Features extracted from rice images were tested with (Navie Bayes) NB tree and (Sequential minimal optimization) SMO classifiers. As a result of 10 cross folds, they reached 95.78% accuracy with the NBtree classifier. [21] classified rice types according to morphological, color, texture and wavelet characteristics and gave comparative accuracy values. In the results obtained, they concluded that morphological features play a more active role in classifying rice grains than other features.

As a result of the literature review, it shows that the correct determination of morphological features with statistical analyzes is very important in classifying rice species. For this reason, statistical analyzes were first included in the study and the morphological features obtained were used in classification.

MATERIAL AND METHODS

Dataset

In this study, two different types of rice, Osmancık and Cammeo type, were examined. In the dataset divided into two classes, there are 7 different quality data of 3810 rice grains as Area, Perimeter, Majoraxis,

Minoraxis, Eccentricity, Convex area, Extend. The dataset created by [15] was downloaded from the UCI machine learning Repository portal.

The 1000-grain weight of Osmancık rice is 23-25 grams, and they have a wide, long, glassy and dull appearance, whereas the 1000-grain weight of Cammeo rice is 29-32 grams and they are wide, long, glassy and dull [15]. The distributions and descriptive statistics by the characteristics of the rice belonging to the data set developed by Cınar and Köklü (2019) [15] and used in this study were given in Table 1.

Table 1. Dataset descriptive statistics

Rice Name	Piece	Features	Minimum	Maximum	Mean	Std. Deviation
Osmancık	2180	Area	7551.00	15420.00	11549.78	1041.91
		Perımeter	359.10	503.46	429.42	20.15
		Majoraxis	145.26	209.65	176.29	9.36
		Minoraxis	59.53	101.76	84.48	5.30
		Eccentricity	0.78	0.94	0.88	0.02
		Convex_area	7723.00	15800.00	11799.59	1062.80
		Extent	0.50	0.83	0.67	0.07
Cammeo	1630	Area	9908.00	18913.00	14162.89	1286.77
		Perımeter	410.51	548.45	487.44	22.18
		Majoraxis	170.78	239.01	205.48	10.33
		Minoraxis	67.70	107.54	88.77	5.35
		Eccentricity	0.84	0.95	0.90	0.01
		Convex_area	10205.00	19099.00	14494.43	1309.42
		Extent	0.50	0.86	0.65	0.08

Performance Parameters

There are some key indicators to measure their classification performance when using AI algorithms. The performance of the model created with machine learning algorithms was evaluated according to accuracy, precison, sensitivity and F1 score parameters. Accuracy indicates how close the results obtained as a result of the measurement are to the correct value. It is an important parameter used in evaluating model performance. But just guessing based on it is not enough. For this reason, it is necessary to look at the correctly estimated positive observation values and the total predicted positive observation values. This process is called precision. Sensitivity (recall) values are expressed as the ratio of correctly predicted positive results to all results in the real class. Another important classification parameter is the F1 score value. F1 score is the value obtained using the weight average result of the precision and sensitivity values, which take into account all the results obtained. This value is very important as it is used here as a criterion for both false positives and false negatives.

Evaluation criteria are very important in determining the performance of the created model. Confusion matrix table is used for this process. The formulas of the confusion matrix and calculation parameters were given in Table 2. True Positives (TP) in the table means that the result for both the actual and predicted classes is positive. True Negative (TN) is interpreted to mean that the value of both the true and predicted classes gives the negative. False Positives (FP) is when the value in the true class is no and the value in the predicted class is yes. Finally, the False Negative (FN) statement is the answer of yes for those in the real class and no for those in the predicted class.

Classification Algorithms

Depending on the rapid developments in technology, agricultural studies have also been positively affected by these developments and the concept of precision agriculture has emerged; [6, 22, 23]. Precision agriculture aims to maximize productivity in agriculture. Because one of the biggest goals of the farmer is that the product he obtains is of high quality and efficiency. In this context, AI technologies have started to be used in this field in recent years. While doing all these operations, machine learning algorithms are preferred. The most important advantage of these algorithms is that they perform classification and quality classification easily, quickly and automatically [24].

Table 2. Confusion Matrix and Parameter Formules

		Predicted		
		Cammeo (+)	**Osmancık (-)**	
Actual	Cammeo (+)	True Positive (TP)	False Negative (FN)	Sensitivy (recall) $\dfrac{TP}{TP + FN}$
	Osmancık (-)	False Positive (FP)	True Negative (TN)	Specificity $\dfrac{TN}{TN + FP}$
		Precision $\dfrac{TP}{TP + FP}$	F$_1$ score $2*\dfrac{precision * recall}{precision + recall}$	Accuracy $\dfrac{TP + TN}{TP + TN + FP + FN}$

The common features of the data are used to decompose the existing data to be realized in classification [25]. In this study, models were created from the data obtained by feature selection using different classification algorithms. In the study, the Random Forest (RF), the Support Vector Machine (SVM), the Linear Discriminant Analyzes (LDA), the K-Nearest Neighbor (K-NN), and the Logistic Regression (LR) algorithms were used.

Random Forest (RF)

The RF [26], which is also shown among the ensemble methods, is a model using the introduced regression trees. In this algorithm, the regression trees are trained independently and the tree outputs are averaged for prediction [27]. It takes the decision of each tree in the trees it has created a random forest and uses it to increase accuracy. This algorithm is preferred in studies because of its simplicity and ease of measuring estimation factors [28].

Support Vector Machine (SVM)

The SVM is a learning algorithm that makes classification with supervised learning, which was introduced to the literature by Cortes and Vapnik in 1995 [29], which is used in classification and regression problems, dividing the n-dimensional space into classes and creating the best line or decision boundary to classify data. In this algorithm, the basic logic allows minimizing a limit on the generalization-error of the model instead of minimize the mean-square-error in the train data [30]. The SVM was initially used only in classification problems, but after the loss function and insensitive loss function were defined, it started to be used in regression applications [31]. The fact that it can be used in both cases is one of the reasons why the SVM algorithms are preferred in many test datasets [32]. For this reason, the SVM algorithm was used in classification.

Linear Discriminant Analyses (LDA)

The LDA is a discrimination method in which the intra-class variance is studied by looking at the intra-class variance in cases where the frequency distributions within the class are not equal [33]. It is generally preferred in data classification methods. By choosing this algorithm for classifying two different rice types, the variance between classes will be determined and the classification process will be facilitated. In the LDA algorithm, it is necessary to perform the operations by calculating the distribution matrices within and between classes.

K-Nearest Neighbor (K-NN)

The K-NN algorithm, first developed by Evelyn Fix and Joseph Hodges in 1951 [34] and later with additions by Thomas Cover. The K-NN is known as non-parametric supervised learning method which is used for both classification and regression. It obtains the closest sample to the query sample and detects the single most common class label in the training samples and creates the test group [35]. This algorithm, which is based on distance calculation, has been used in this study as it is more accurate to use in numerical data sets.

Logistic Regression (LR)

The LR within machine learning is an algorithm that belongs to the family of supervised machine learning models. Logistic regression estimates the probability that a voted or unvoted event will occur in a defined independent variable dataset. Since all these operations are within the probabilities, the value of the dependent variable obtained is kept between 0-1 [36]. It can be easily used in more than one class and probabilistic view of different predictions. On the other hand, the assumption of linearity between the dependent variable and the independent variable is one of its biggest limitations.

RESULTS

Rice production, consumption and foreign trade in the world

World rice production increased by 5 million tons and reached 504 million tons in the period 2020/21 to the previous period. When the world rice production is examined, China comes first and India comes second [37]. These two countries meet more than half of the global production. More than half of the total rice production and consumption is in Asian countries due to traditional dietary habits, suitability of climate, and population sizes.

World rice consumption increased by 3 million tons compared to the previous period and reached 504 million tons due to population growth and the pandemic. In addition, stock figures decreased as consumption exceeded production in the related period.

World rice trade increased with the increasing demand due to the Covid-19 outbreak and reached 45.8 million tons in the period 2020/21. Especially the rice imports of Sub-Saharan Africa and South Asia increased.

World rice production, consumption, trade and stock statistics by years were given in Table 3. As shown in the table, rice production increases yearly, and it gains value commercially and economically.

Table 3. Rice production, consumption, trade and stock in the world (million tons)

Situation	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Production	449	470	478	481	484	478	492	494	498	499	504
Consumption	445	456	466	478	476	473	483	484	488	501	504
Trade	36	39	37,8	42,7	42	39,5	47	46,5	42,6	44	45,8
Stock	100	117	128	132	139	145	155	165	176	174	173

Source: Turkish Grain Board (TGB). (2021). Cereal Sector Report For 2020. Directorate General of Turkish Grain Board:1-42.

In the world rice market, production in orange, consumption in gray, trade in yellow and stock status in blue are shown in Figure 1. It is seen that rice production and consumption increase every year.
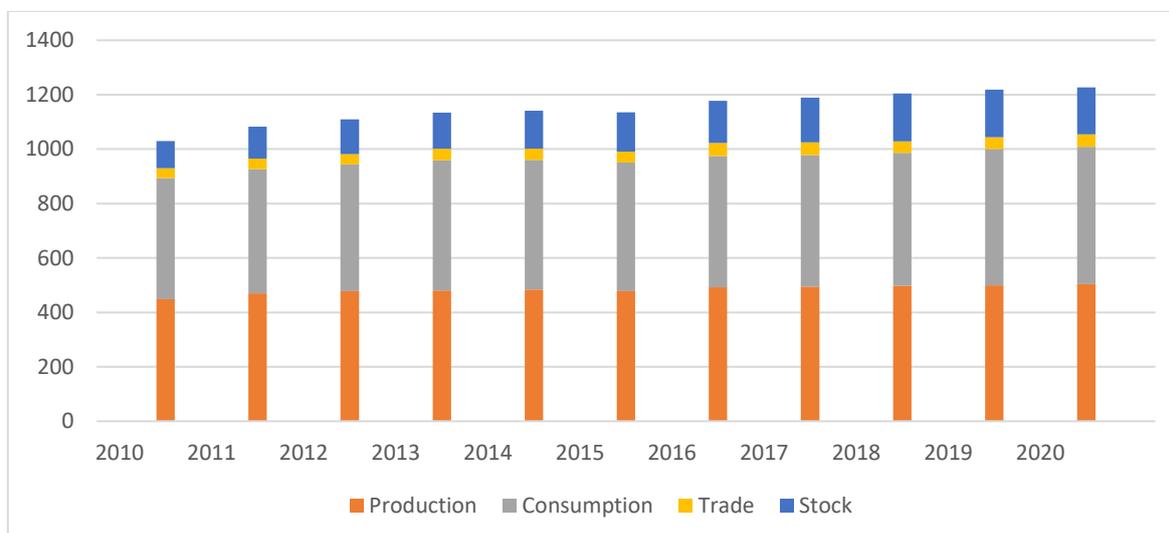

Figure 1. Rice production, consumption, foreign trade and stock in the world in 2010-2020

Statistical analysis

When the suitability of Cammeo and Osmancık rice types with normal distribution is examined graphically, it can be observed that the attributes other than Extent more compatible with the normal distribution (Figure 2). For datasets with more than 300 samples, the Normal distribution depends on the absolute values of skewness and kurtosis. An absolute skewness value greater than 2 or an absolute kurtosis value greater than 7 can be used as reference values to indicate that there is no normal distribution [38, 39].
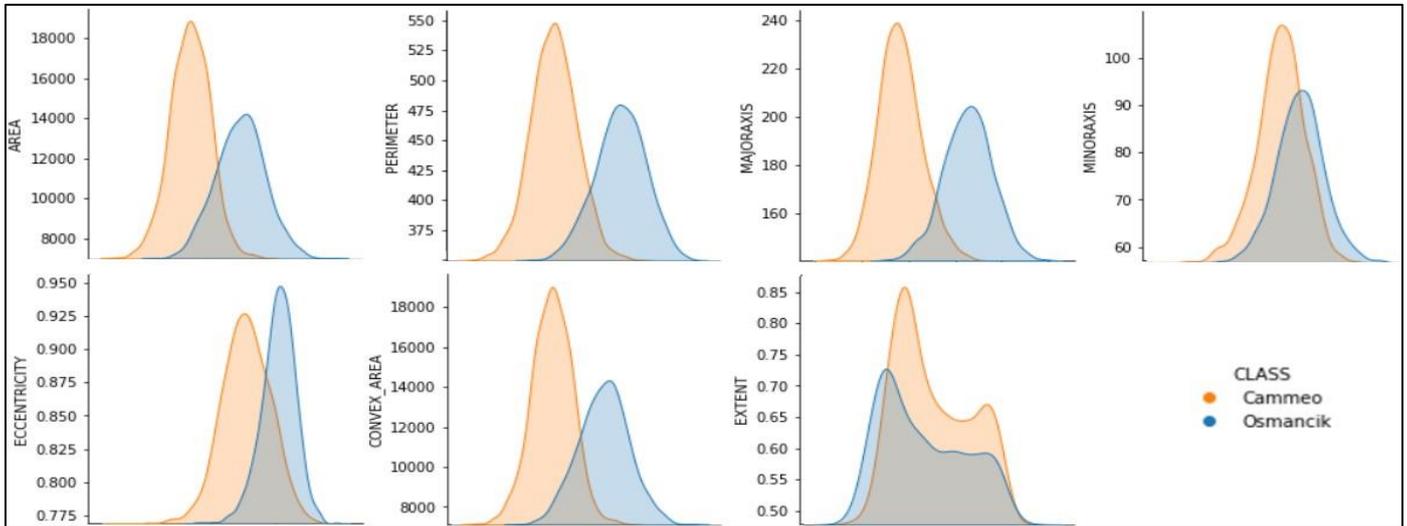

Figure 2. Distribution of rice by classes and qualities

Considering the skewness and kurtosis values of the attribute values of rice species, it can be stated that the data are compatible with the normal distribution. It is observed that the Extent and Eccentricity attributes in Cammeo type rice and the Extent attribute in Osmancık type rice have more skewness and kurtosis than the other attributes (Table 4).

Table 4. Skewness and kurtosis coefficients of data set attribute values

	Cammeo		Osmancik	
	Skewness	**Kurtosis**	**Skewness**	**Kurtosis**
Area	0.005	-0.033	-0.061	0.353
Perımeter	-0.169	-0.107	0.045	0.232
Majoraxis	-0.086	0.026	0.285	0.152
Minoraxis	-0.003	0.456	-0.347	0.594
Eccentricity	-0.518	1.149	-0.191	0.519
Convex_area	-0.021	-0.049	-0.051	0.382
Extent	0.413	-1.052	0.375	-1.044

Factor analysis was used to explain the relationship between variables with the correlation model. Factor analysis is the reduction of the variance seen in too many variables to a small number of factors by applying data reduction [40]. Factor analysis can also be used to screen for variables for subsequent analysis, such as determining collinearity before performing a linear regression analysis [41]. Factor analysis attempts to establish the common features underlying the relationships between variables in a dataset. Another essential feature of factor analysis is inference. Inference techniques enable the identification of factors underlying the relationship between a set of variables. There are many inference procedures, but the most common to use is the Principal Component Analysis (PCA) [42]. In the factor analysis applied to the data set, the PCA was used as the inference method.

Table 5 shows the values obtained as a result of the Kaiser-Meyer-Olkin (KMO) test. With this process, it was checked whether the sample size was sufficient or not. The KMO value appears as a value between

0-1, and this value approaching 1 indicates the suitability of the sample size for factor analysis. A value greater than 0.5 is sufficient for a reliable factor analysis [43]. In the analysis, the KMO value was found to be 0.668 and it was observed that the sample size was sufficient. In addition, Bartlett's test of sphericity was used to determine the suitability of the model. This test tests the case where the correlation matrix is an identity matrix. In the table, the significance value (p value) of Bartlett's test was found as 0.000 and it was seen that the correlation matrix was not a unit matrix. In this case, it is concluded that the factor model is appropriate [44].

Table 5. KMO and Bartlett's Test

Kaiser Meyer Olkin Measure of Sampling Adequacy		0.668
Bartlett's Test of Sphericity	Approx. Chi-Square	82352.1
	Df	21
	Sig.	0.000

The common points of all variables are given in Table 6. The fact that the variable associates with high-level features indicates that most of the variables are supported by all the factors described in the analysis. If the common point value of the variable is <0.4, this feature is considered unusable in the study and should naturally be removed from the model [44]. The extent attribute commonality is less than 4, therefore it should not be included in the model.

Table 6. Communalities

Features	Initial	Extraction
Area	1.000	0.997
Perimeter	1.000	0.992
Majoraxis	1.000	0.984
Minoraxis	1.000	0.956
Eccentricity	1.000	0.916
Convex_area	1.000	0.999
Extent	1.000	0.234

The model was rebuilt by excluding the extend attribute. The commonalities of the new model were shown in Table 7. It is viewed that the commonality of all variables was greater than 4, and therefore all variables are useful for the model.

Table 7. Communalities of Model

Features	Initial	Extraction
Area	1.000	0.997
Perimeter	1.000	0.993
Majoraxis	1.000	0.997
Minoraxis	1.000	0.997
Eccentricity	1.000	0.996
Convex_area	1.000	0.999

Figure 3 shows the scatter plot obtained with the factors and their eigenvalues. It can be seen from the figure that only two factors had eigenvalues above elbow bent. Thus, only two factors were preserved in this model.

In Table 8, the factors for rice and the corresponding variance of these factors were given. It is seen in the table that the first and second factors explain 76.118% and 23.531% of the total variance, respectively. These two factors correspond to 99.649% of the variance.

In other words, it was seen that the Extent attribute was useless for the model created by factor analysis, other attributes were useful in creating the model, and these attributes explained 99.649% of the total variance in the model.
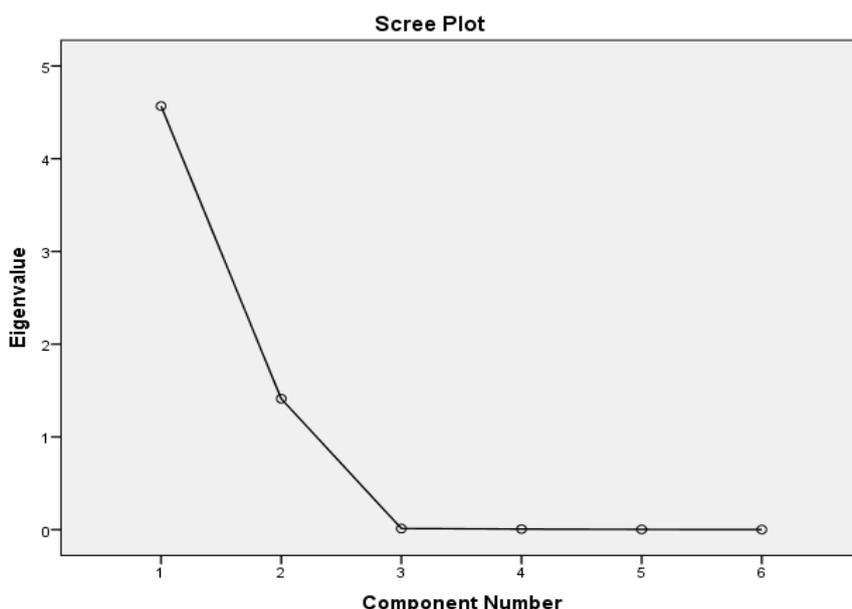
Figure 3. Scree plot for the factors

Table 8. Variance Distribution

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.567	76.118	76.118	4.567	76.118	76.118
2	1.412	23.531	99.649	1.412	23.531	99.649
3	0.012	0.203	99.852			
4	0.006	0.107	99.958			
5	0.002	0.034	99.993			
6	0.000	0.007	100.000			

In the study, classification process was carried out by using 6 different morphological features of rice varieties obtained as a result of statistical analysis. In addition, models created with five different machine learning algorithms were examined according to their performance using the Python programming language. All models were cross-validated. Cross validation is the preferred sampling method to avoid overfitting [39]. In k-fold cross validation, the data is trained by creating k-1 subsets for training and testing without changing the learning set. Thus, a more reliable average accuracy value is obtained [45]. The average accuracy is given according to the results obtained from each cross validation value by applying the 5-fold cross validation process. However, the data were divided into two groups as training and testing. The dataset was run as 70% training and 30% testing.

The accuracy, recall, precision, f1-score parameters of all algorithms are given in Table 9. When the table is reviewed in detail, it is seen that the SVM, LDA and LR algorithms give more successful results than other algorithms. Classification accuracy values are 93.53% for the SVM, 93.44% for the LR, 93.0% for the LDA, 92.83% for the RF and 87.58% for K-NN, respectively. According to the accuracy value, which is one of the basic parameters in classification, the SVM algorithm gave the highest accuracy in classifying rice varieties.

Table 9. Algorithms

Algorithm	Precision	Sensitivity	Specificity	F_1-Score	AUC	Accuracy
LR	95.25	93.04	93.95	94.14	99.03	93.44
RF	95.57	91.79	94.23	93.64	97.41	92.83
k-NN	91.46	86.79	88.68	89.06	97.86	87.58
SVM	95.25	93.19	93.96	94.21	99.18	93.53
LDA	96.36	91.44	95.18	93.84	96.07	93.00

Confusion matrix results and thermal distributions were obtained to estimate the classification performance in machine learning algorithms. The confusion matrices of the algorithms are given in Figure 4. With confusion matrices, errors during classification and the number of images classified in the wrong label

can be seen. According to the results of the confusion matrix of the logistic regression algorithm, it is seen that the algorithm predicts 1068 images out of 1143 images in the test group with an accuracy of 93.44%. On the other hand, it is seen that 1069 images are classified correctly in the SVM algorithm, which gives the best results. When the algorithms are examined in detail, it is showed that the highest precision value is 96.36% in the LDA algorithm. The sensitivity value results, which are calculated as proof that there is Cammeo rice at the true value of the test and the same rice at the estimated value, are also very important in classification. Considering the sensitivity values of the algorithms, it gave the highest result of the SVM algorithm with 93.19%. On the other hand, the specificity value is the highest in the LDA algorithm with 95.18%. Another critical parameter, f1-score value, is among the most important success criteria used together with accuracy in classification. Accordingly, the highest f1-score value was obtained with the SVM algorithm with 94.21%.
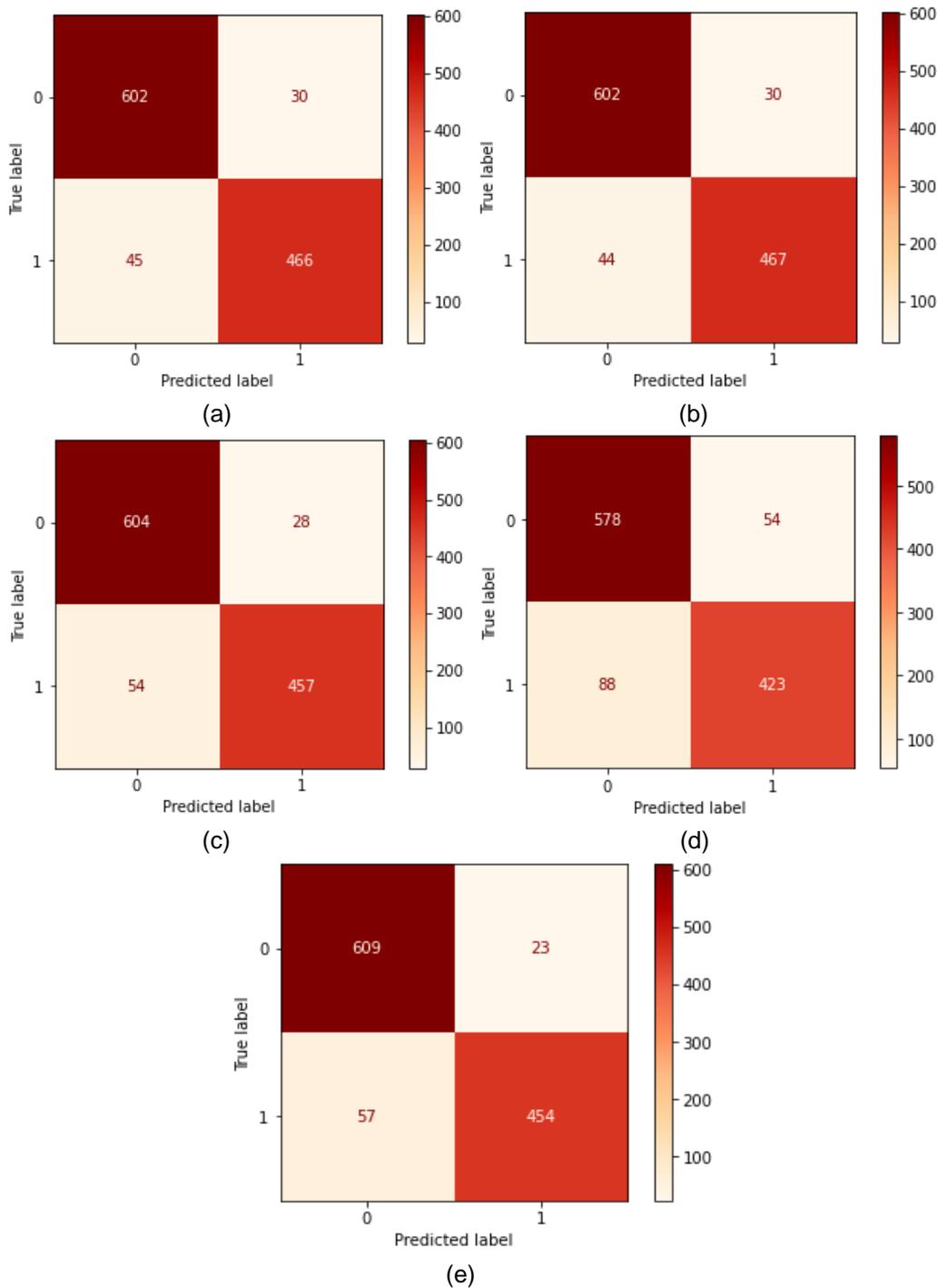
Figure 4. Confusion Matrix of (a) LR, (b) SVM, (c) RF, (d) k-NN, (e) LDA

The Receiver Operating Characteristic (ROC) curve of algorithms is a metric used in solving classification problems. The ROC is the curve with TP values (sensitivity) with the y and FP values (specificity) with the x axis to be calculated by assuming the cut-off point, respectively, of all values taken within the range of variation of the classification variable. While determining a threshold value, sensitivity and specificity value are determined by considering the point closes the up side to left corner of the ROC. Each ROC curve also has an AUC value. This AUC value is considered an indicator of overall accuracy [45]. When the AUC values are examined, the SVM algorithm has the highest AUC value of 99.18%, the sensitivity is 93.19% and the specificity is 93.96%. In other words, a very high result was accomplished in the classification. Thus, it appears that AI can be successfully applied in the product classification process, which is one of the main components of agricultural marketing.

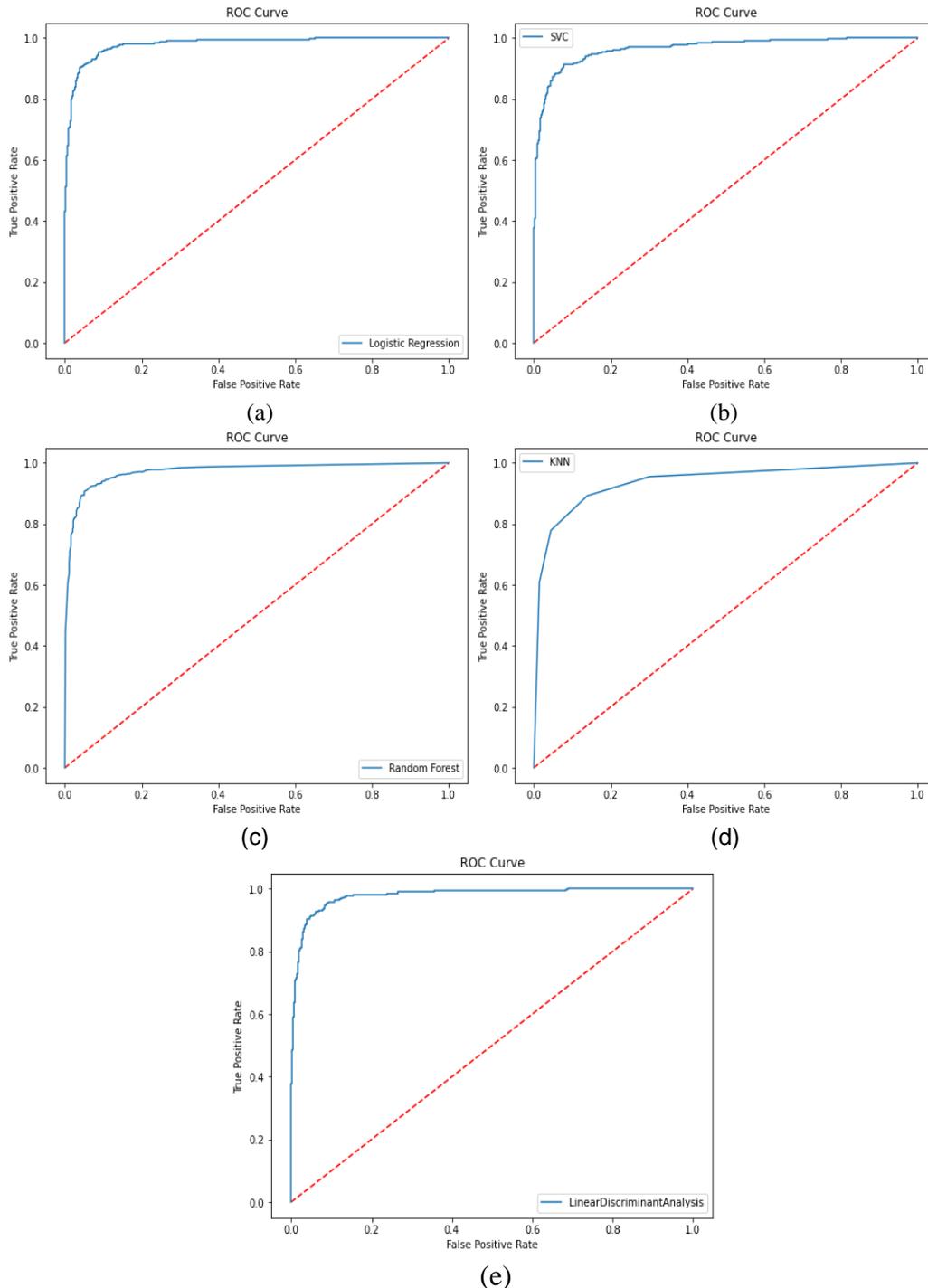The ROC curves of the algorithms are given in Figure 5.

Figue 5. Roc Curve of (a) Logistic Regression, (b) Support Vector Machine, (c) Random Forest, (d) k-NN, (e) LDA algorithms

DISCUSSION

Rice types were analyzed with models created using ML algorithms. Experimental data show that the selected features affect the accuracy performance of classification algorithms. It is seen that TP values are quite high with all algorithms. This also shows that the proposed model can successfully identify both types in classifying rice types. In addition, the classification accuracy value of any algorithm did not fall below 0.9283. When the results given in the literature review are examined, it is seen that the studies performed have lower accuracy. However, the fact that the AUC values of the three algorithms that give the best results are higher than the other algorithms supports the validity of the accuracy results obtained. Compared to similar studies, this study has higher performance in distinguishing 2 different classes. It has been observed that the accuracy values are higher than the study performed with the same dataset and using similar algorithms. The main reason for this is the application of a different feature selection method compared to other studies. This shows how valuable the selection of morphological features is for classification. The data obtained with statistical analyzes made the study more powerful. It has been shown by analysis that 6 features used in classification are important parameters that can be used in rice-type classification. When the morphological features obtained from the analyzes were used, a model with high classification success emerged. In addition, the k-fold cross validation process was applied in the adjustment of hyperparameters on this study. Training data was divided into five parts and each of them was tested separately to fit the model. Thus, it is prevented from memorizing the dataset and using the learned data as test data is prevented. On the other hand, this situation has been ignored in some literature studies. All these processes applied in the model show that the proposed feature selection method and the model used to contribute positively to the classification.

The high-performance values demonstrated by the model can be easily used in the classification of many different agricultural products. Thus, it appears that the AI can be successfully applied in the fields of agricultural economics. Likewise, [6], in their study on the use of the AI in agriculture and rural development, stated that the AI can be applied successfully. Using the feature selection process in classification has increased the accuracy. With this research study, it is concluded that feature selection and feature extraction methods are important factors in determining rice type with ML algorithms. The most important disadvantage of the model is that the rice types used in the data set are limited. The use of the model in classifying different rice types may pave the way for the use of this method in sectoral enterprises. Finally, this study shows that the use of technology is very important in crop classification and marketing processes.

CONCLUSION

In this study, a multidisciplinary study was put forward in the fields of agricultural economics and information technology, and a new AI-based model was tried to be developed in product classification. It was tried to show through a sample data set that AI algorithms can also be used in the field of agricultural economics and can be an alternative to known classical methods. In this context, in the study, it was revealed that the contribution of agriculture to the economy could be increased with the use of modern technologies.

As a result, how to perform the classification process on a sample agricultural data set, which criteria to use, and how to interpret it was discussed in detail, and the results were interpreted by comparing the actual observation values with the values produced by modern technology.

In future studies, comparative analyzes can be made with more different types of rice. In addition, models can be developed for the automatic classification of other agricultural products with different algorithms. With the use of AI-based classification methods in the classification of crops, both labor costs will decrease and the quality of production will be increased. In short, the added value of agriculture in the economy will increase with technological developments and smart systems.

REFERENCES

1. Chatnuntawech I, Tantisantisom K, Khanchaitit P, Boonkoom T, Bilgic B, Chuangsuwanich E. Rice classification using spatio-spectral deep convolutional neural network. arXiv preprint 2018;arXiv:1805.11491. Available from: https://arxiv.org/ftp/arxiv/papers/1805/1805.11491.pdf
2. Şapaloglu A, Güngör G. The structure of the marketing channels and rice marketing margins in the rice production-consumption chain: An example of the province of Edirne Province. Namık Kemal University Graduate School of Natural and Applied Sciences Department of Agricultural Economics MSc. Thesis. 2015; 164 p.

3. Erbas N. Comparative economic analysis of farms in Turkey and a critical assessment of the annual profitability: The case of Yozgat Province. Custos e agronegocio on line. 2021;17(1):332-U460.

4. Albayrak M. ZTE 315 Agricultural Marketing. Ankara University Faculty of Agriculture Department of Agricultural Economics. Turkey: 2022.

5. Ministry of National Education. (MNE). Classification and Packaging of Fruits. Republic of Türkiye Ministry of National Education. Agricultural Technologies. 2015; 46 p.

6. Gómez-Chova L, Tuia D, Moser G, Camps-Valls G. Multimodal classification of remote sensing images: A review and future directions. Proceedings of the IEEE. 2015;103(9):1560-84. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7182258

7. Uzun Y, Bilban M, Arıkan H. Use of artıfıcıal ıntellıgence in agrıculture and rural development. VI. KOP Regional Development Symposium; October 26-28, 2018. Konya, Turkey: 2018;1-6.

8. Terzi İ, Özgüven MM, Altaş Z, Uygun T. Use of artificial ıntelligence in agriculture. International Erciyes Agriculture, Animal & Food Sciences Conference; April, 2019. Erciyes University, Kayseri, Turkey: 2019;245-55.

9. Erbaş N, Çınarer G, Kılıç K. Classification of hazelnuts according to their quality using deep learning algorithms. Czech J of Food Sci. 2022 Jun 40(3):240-248. https://doi.org/10.17221/21/2022-CJFS

10. Gujjar HS, Siddappa DM. A method for identification of basmati rice grain of india and its quality using pattern classification. Int J of Eng Res. 2013 Jan;3(1):268-73.

11. Kambo R, Yerpude A. Classification of basmati rice grain variety using image processing and principal component analysis. 2014);arXiv preprint arXiv:1405.7626. Available from: https://arxiv.org/ftp/arxiv/papers/1405/1405.7626.pdf

12. Silva CS, Sonnadara U. Classification of rice grains using neural networks. Proceedings of Technical Sessions, Institute of Physics Sri Lanka: 2013;29:9-14. Available from: https://citeseerx.ist.psu.edu/viewdoc/download?repid=rep1&type=pdf&doi=10.1.1.401.8889

13. Kuo TY, Chung CL, Chen SY, Lin HA, Kuo YF. Identifying rice grains using image analysis and sparse-representation-based classification. Comput Electron Agric. 2016 Sep;127:716-25. Available from: https://doi.org/10.1016/j.compag.2016.07.020

14. Asif MJ, Shahbaz T, Rizvi STH, Iqbal S. Rice Grain Identification and quality analysis using image processing based on principal component analysis. In 2018 International symposium on recent advances in electrical engineering (RAEE) IEEE 2018; 6 p. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8706891

15. Cinar I, Koklu M. Classification of Rice Varieties Using Artificial Intelligence Methods. Int J of Intell. 2019 Sep;7(3):188-94. https://doi.org/10.18201/ijisae.2019355381

16. Koklu M, Cinar I, Taspinar YS. Classification of rice varieties with deep learning methods. Comput Electron Agric. 2021 Aug;187, 106285. https://doi.org/10.1016/j.compag.2021.106285

17. Kuchekar NA, Yerigeri V V. Rice grain quality grading using digital image techniques. IOSR J Electronics Communication Eng 2018;13(3):84-8.

18. Parveen Z, Alam MA, Shakir H. Assessment of quality of rice grain using optical and image processing technique. In 2017 International Conference on Communication, Computing and Digital Systems (C-Code) IEEE 2017;265-70. Available from: https://ieeeprojects.eminents.in/uploads/basepaper/ETSIP007-2017.pdf

19. Akı O, Güllü A, Uçar E. Classification of rice grains using image processing and machine learning techniques. In International scientific conference. 2015;20-1.

20. Philip TM, Anita HB. Rice Grain Classification using Fourier Transform and Morphological Features. Indian J Sci Technol. 2017 Apr;10(14):1-6. Available from: https://sciresol.s3.us-east-2.amazonaws.com/IJST/Articles/2017/Issue-14/Article9.pdf

21. Robert Singh K, Chaudhury S. A cascade network for the classification of rice grain based on single rice kernel. Complex & Intelligent Systems 2020;6(2):321-34. https://doi.org/10.1007/s40747-020-00132-9

22. Hütt C, Koppe W, Miao Y, Bareth G. Best accuracy land use/land cover (LULC) classification to derive crop types using multitemporal, multisensor, and multi-polarization SAR satellite images. Remote sensing 2016;8(8):684. https://doi.org/10.3390/rs8080684

23..Üstüner M, Abdikan S, Bilgin G, Balik Şanli F. Crop Classification Using Light Gradient Boosting Machines. Turk J Remote Sensing and GIS. 2020;1(2):97-105.

24. Mahesh B. Machine learning algorithms-a review. Int J Sci Res (IJSR). 2020;9(1):381-386. Available from: DOI: 10.21275/ART20203995

25. Ozkan Y. Data Mining Techniques. Papatya Publishing and Education Inc. Istanbul, Turkey: 2008.

26. Breiman, L. Random forests. Machine Learning 2001;45(1):5–32.

27. Yoon J. Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. Computational Economics 2021;57(1):247-65. https://doi.org/10.1007/s10614-020-10054-w

28. Sandhu AK, Batth RS. Software reuse analytics using integrated random forest and gradient boosting machine learning algorithm. Software: Practice and Experience 2021;51(4):735-47. https://doi.org/10.1002/spe.2921

29. Cortes C, Vapnik, V. Support-vector networks. Machine learning 1995;20(3):273-297. Available from: https://link.springer.com/content/pdf/10.1007/bf00994018.pdf

30. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 2020;408:189-215. https://doi.org/10.1016/j.neucom.2019.10.118

31. Zhou J, Qiu Y, Zhu S, Armaghani DJ, Li C, Nguyen H, Yagiz S. Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. Engineering Applications of Artificial Intelligence 2021;97, 104015. https://doi.org/10.1016/j.engappai.2020.104015

32. Ibrahim S, Zulkifli NA, Sabri N, Shari AA, Noordin MRM. Rice grain classification using multi-class support vector machine (SVM). IAES J Artificial Intell. 2019;8(3):215. Available from: https://doi.org/10.1016/j.engappai.2020.104015

33. Balakrishnama S, Ganapathiraju A. Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing 1998;18:1-8.

34. Silverman BW, Jones MC. E. fix and jl hodges. An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges. International Statistical Review/Revue Internationale de Statistique 1951;1989;233-8.

35. Ali M, Jung LT, Abdel-Aty AH, Abdel-Aty AH, Abubakar MY, Elhoseny M, Ali I. Semantic-k-NN algorithm: an enhanced version of traditional k-NN algorithm. Expert Systems with Applications 2020;151, 113374. https://doi.org/10.1016/j.eswa.2020.113374

36. Schober P, Vetter TR. Logistic regression in medical research. Anesthesia and analgesia 2021;132(2):365. https://doi.org/10.1213%2FANE.0000000000005247

37. Turkish Grain Board (TGB). Cereal Sector Report For 2020. Directorate General of Turkish Grain Board: 2021;1-42. Available from: https://www.tmo.gov.tr/Upload/Document/sektorraporlari/hububat2020.pdf

38. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restorative dentistry & endodontics 2013;38(1):52-54. DOI: https://doi.org/10.5395/rde.2013.38.1.52

39. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction Vol. 2, New York: 2009;1-758 springer.

40. Vapnik V. The nature of statistical learning theory. Springer science & business media. 1999.

41. IBM. IBM SPSS Statistics Base 28. IBM Corp. 2022; 270 p. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Base.pdf.

42. Miller RL, Acton C, Fullerton DA, Maltby J. SPSS for social scientists. Bloomsbury Publishing, United Kingdom: 2009; 334 p.

43. Berrar D. Cross-Validation. Data Science Laboratory, Tokyo Institute of Technology. 2019; 8 p.

44. Verma JP. Data analysis in management with SPSS software. Springer Science & Business Media, New Delhi. 2012; 480 p.

45. Hoo ZH, Candlish J, Teare D. What is an ROC curve?. Emergency Medicine Journal 2017;34(6):357-359. http://dx.doi.org/10.1136/emermed-2017-206735