International Journal of
Intelligent Systems and Applications in Engineering

ISSN:2147-6799 www.ijisae.org Original Research Paper

Classification of Rice Varieties Using Artificial Intelligence Methods

Ilkay CINAR*[1], Murat KOKLU[2]

Submitted: 24/08/2019 **Accepted : 30/09/2019**

Abstract: Rice being one of the most widely produced and consumed cereal crops in the world, is also the one of the main sustenance in our country because of its economical and nutritious nature. Rice, starting from farm to our table, goes through some manufacturing steps such as a cleaning process, color sorting and classification. If these stages are to be mentioned briefly, cleaning is the process of separating rice from foreign substances, classification is the process of separating broken ones with sturdy ones; color extraction is the process of separating the stained and striped ones except the whiteness on the rice surface. In this study, a computerized vision system was developed in order to distinguish between two proprietary rice species. A total of 3810 rice grain's images were taken for the two species, processed and feature inferences were made. 7 morphological features were obtained for each grain of rice. With these features, models were created using LR, MLP, SVM, DT, RF, NB and k-NN machine learning techniques and performance measurement values were obtained. Success rates in the classification were obtained 93.02% (LR), 92.86% (MLP), 92.83% (SVM), 92.49% (DT), 92.39% (RF), 91.71% (NB), 88.58% (k-NN). When we look at the results of the success rate of obtain, it is possible to say that the study achieved success.

Keywords: *Classification of rice, computer vision system, image processing, machine learning system*

1. Introduction

When we look at the production of grain products throughout the world, rice is the most important product following wheat and corn. Rice is quite rich in carbohydrates and starch. Rice has great importance in human nutrition in our country as well as in the world in terms of being nutritious and economical. It is also widely used in industry [1].

Various quality criteria for rice production in our country is made available. These are physical appearance, cooking characteristics, aroma, taste and smell are issues such as efficiency next to the properties. From the perspective of the end consumer's, it is the first feature physical appearance that comes to mind from the criteria that stand out in the rice varieties that are sold packaged on market shelves [2]. After production, it is seen that the need for technological methods increases because the calibration of rice, determination of its types, and separation of various quality elements are inefficient and time-consuming, especially in terms of those with high production volume.

Therefore, when we look at the recent studies on cereal products using machine vision systems and image processing techniques, it is seen that the products are examined in terms of many physical properties such as color, texture, quality and size.

In their work, Yadav and Jindal have calculated quantifications on ground rice by subtracting the perimeter, length, and shape features of the rice grain by means of digital image analysis [3].

Visen et al. in their study using artificial neural network with image processing techniques barley, rye, oats, wheat and durum wheat including images of the grain of five types variety have developed algorithms to analyze. Using the resulting images, over 150 color and textural features, they have achieved a success rate of over 90% for all grain types with the classifier they developed in the identification of grains [4].

Baykan et al. in their studies using wheat grains, they obtained a grey level average of the grain with 9 morphological features for the images they have obtained. They have achieved 72.62% success in the classification for the 5 different species used. However, when they removed a species that was difficult to identify in terms of classification, the success has been 82.65% [5].

Dubey et al. in their studies, they have used three different types of bread wheat. They have obtained about 88% accuracy as a result of their classification by subtracting 45 morphological features for the artificial neural network classification [6].

Zapotoczny et al. they have inferred 74 pieces morphological features using an image analysis technique for the classification of five different barley species. They have used basic component analysis, linear and nonlinear discriminant analysis as a method of classification. As a result, they stated that LDA is the best method in classification methods [7].

Chen et al. in their work, they have obtained 13 shapes, 17 geometric and 28 color features from the images obtained from 5 pieces corn varieties. As a result of their classification, they have detected that the accuracy rate among corn varieties was 90% on average [8].

Babalik et al. in their study, they have used multi-class support vector machines and binary particle swarm optimization algorithms to classify 9 geometric and 3 color properties from 5 pieces wheat species. They have achieved an average classification accuracy of 91.5% with M-SVM and 92.02% with BPSO algorithm [9].

Ouyang et al. using the image processing technique, they have designed an automated image capture band system to distinguish 5 different rice seeds. As the seeds passed through the tape system, their pictures have been taken with the help of a CCD camera. They

[1] *Graduate School of Natural and Applied Sciences, Selcuk University, TURKEY, ORCID ID : 0000-0003-0611-3316*
[2] *Faculty of Technology, Selcuk University, TURKEY. ORCID ID : 0000-0002-2737-2360*
* *Corresponding Author Email: ilkay_cinar@hotmail.com*

have used Visual C++ 6.0 for image analysis. They have achieved an average success rate of 86.65% for 5 different rice varieties using back propagation classification [10].

In his study, Farahani has used five clusters of features for linear discrimination analysis by subtracting morphological features of 5 pieces Durum wheat variety with image processing technique. In the analysis conducted, 11 morphological features used and achieved a classification accuracy of 67.66%, which was the best result [11].

Silva and Sonnadara classified rice varieties using artificial neural network in their work. In total, they developed algorithms to extract 13 morphological features, 6 Color features and 15 texture features by taking images from 9 different rice varieties. They have made different classifications for these features, both separately and together. When the classification was done separately, it was observed that tissue features had higher success compared to morphological and color features. A 92% success rate was achieved as a result of the classification with all features together [12].

In their study, Kaur and Singh studied rice classification using multi-class support vector machines. By separating the rice in which they extracted their geometric features, they calculated their percentages for each one and determined their features. They achieved a higher success rate than 86% as a result of the classification [13].

Abirami et al. in their study, they have made use of image processing and neural network pattern recognition techniques for the basmati rice species and carried out classification. They removed various morphological features of rice grains through pretreatment such as filtering, thresholding, edge detection, and classified them by neural network pattern recognition. As a result, they achieved a 98.7% success rate [14].

Pazoki et al. using ANN and Neuro-Fuzzy machine learning techniques, they classified five different rice varieties using 24 color, 11 morphological and 4 shape factor features and achieved 99.46% success as a result [15].

Szczypiński et al. in their studies, they have evaluated the effectiveness of Cultivar determination on barley varieties based on their shape, color and texture features. They have made use of linear distinctive analysis and artificial neural networks for classification. As a result, they have achieved varying success rates between 67% and 86% [16].

When the studies were examined, product features have been extracted using shape and color features besides morphological features by utilizing various image processing techniques on the resulting images. These features have been classified with different methods and it is observed that success rates are higher than 90% especially in studies where color, shape and morphological features are evaluated together.

In the second part of our study, the process of obtaining the image in the material and method section, the operations performed on the resulting images, feature extraction from processed images, data set, performance measurements and about cross-validation has been provided. In the third part, the creation of LR, MLP, SVM, DT, RF, NB and k-NN models and information about these models are given. In the fourth part, the results of our study are given. In the fifth chapter, which is the last chapter, the study has been evaluated, discussed and suggestions are given.

2. Materials And Methods

In our study, images of rice samples were obtained first and images were processed using various image processing techniques. The resulting images are first converted to grayscale image, then converted to binary image and removed from the noise on the image. In the next phase, various morphological feature inference processes were applied on the obtained images. The classification phase of rice is given in Figure 1. During the modeling phase, the rice classification process was carried out using seven pieces machine learning techniques. In the last step, the performance of the models used was evaluated.

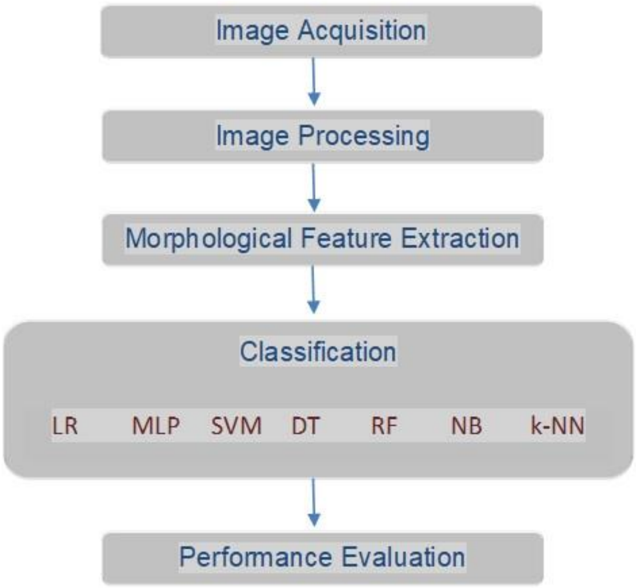

Fig. 1. Rice classification phases

2.1. Image Acquisition

In order to obtain images of the rice used in the study, a mechanism similar to the system given in Figure 2 was established. The camera used to take images in the system is placed on a box with a lighting mechanism. The box is designed so that it does not receive light from the outside and is intended to prevent shadow formation on the image to be obtained in this way. The box floor is selected in black for easy processing of the image. The captured rice samples were transferred to the computer system and recorded.

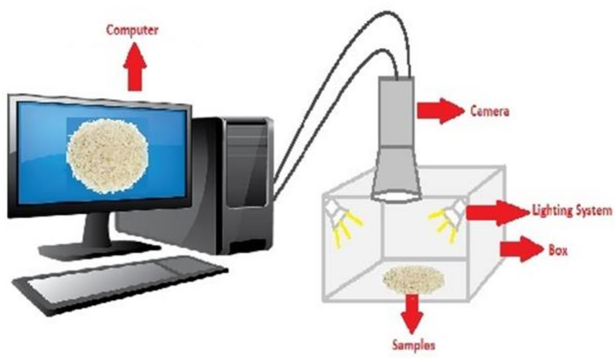

Fig. 2. Computer vision system

Both types of rice used in the study were set to 50 gr. A total of 3810 images of rice grains were obtained for the two species. Examples of rice images obtained in Figure 3 are given.

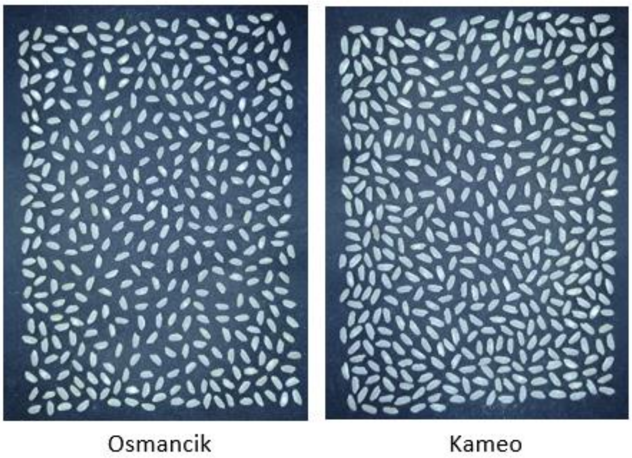

Fig. 3. Rice samples used in the study

2.2. Image Processing

In this section, preliminary operations on images are explained in order to obtain feature extraction and classification processes in the most accurate way. Image processing is critical as it directly affects the outcome of feature inference and classification. For this reason, these were taken into consideration when designing the image processing phase. The processing of the image was carried out with the help of MATLAB application. The images taken from the camera have been converted to grayscale and binary images in order to prepare for morphological feature inference and then stripped of their noise. In Figure 4, the preliminary process stages on the rice image are given.

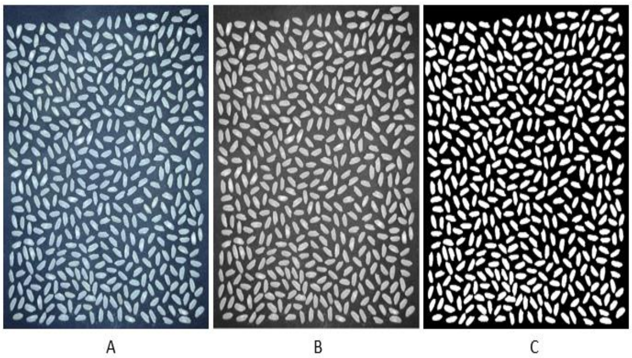

Fig. 4. Image processing stages

(a) Original image (b) Grayscale image (c) Binary image

2.3. Morphological Feature Extraction

After the processing phase of the image, the rice grains on each image were treated separately and a number of features were inferred. Feature extractions have been studied from morphological point of view. In total, 7 morphological features were inferred for each grain. Morphological feature extraction is a wide range of image processing processes that process based on the shapes found on the image. In this process, each pixel in the image is adjusted according to the value of the other pixels around it [17]. The most effective morphological features and explanations used in feature extraction for both types of rice are given in Table 1.

Table 1. The most effective morphological features and explanations used in feature extraction

No	Name	Explanation
1	Area	Returns the number of pixels within the boundaries of the rice grain.
2	Perimeter	Calculates the circumference by calculating the distance between pixels around the boundaries of the rice grain.
3	MajorAxisLength	The longest line that can be drawn on the rice grain, i.e. the main axis distance, gives.
4	MinorAxisLength	The shortest line that can be drawn on the rice grain, i.e. the small axis distance, gives.
5	Eccentricity	It measures how round the ellipse, which has the same moments as the rice grain, is.
6	ConvexArea	Returns the pixel count of the smallest convex shell of the region formed by the rice grain.
7	Extent	Returns the ratio of the region formed by the rice grain to the bounding box pixels.

2.4. Rice Features Dataset

Among the certified rice grown in our country, the Osmancik species, which has a large planting area since 1997, and the Cammeo species grown since 2014 have been selected for the study. When looking at the general characteristics of Osmancik species, they have a wide, long, glassy and dull appearance. One thousand pieces of grain weight is 23-25 gr [18]. When looking at the general characteristics of the Cammeo species, they have wide and long, glassy and dull in appearance. One thousand pieces of grain weight is 29-32 gr. In this study, the distribution of 3810 rice grain obtained as a result of processing images of both species is given in Table 2.

Table 2. Distribution of rice grains by species

No	Name	Piece	Grain Weight (mean)
1	Osmancik	2180	0,023 gr
2	Cammeo	1630	0,031 gr
Total		3810	0,027 gr

The minimum, average, maximum, standard deviation, skewness and kurtosis data for both types of rice samples are given in Table 3.

Table 3. Descriptive statistics of rice species data

No	Features	Min	Mean	Max	Std. Dev.	Skewness	Kurtosis
1	Area	7551	12667,73	18913	1732,37	0,3252	-0,4311
2	Perimeter	359,1	454,2392	548,446	35,5971	0,2214	-0,8402
3	MajorAxisLength	145,2645	188,7762	239,0105	17,4487	0,2602	-0,9518
4	MinorAxisLength	59,5324	86,3138	107,5424	5,7298	-0,1349	0,5621
5	Eccentricity	0,7772	0,8869	0,948	0,0208	-0,4492	0,0711
6	ConvexArea	7723	12952,50	19099	1776,97	0,3198	-0,4658
7	Extent	0,4974	0,6619	0,861	0,0772	0,3438	-1,0301

2.5. Performance Measures

The creation of a new model required for classification problems, or the use of existing models and achieving success on that model are calculated by means of the number of accurate estimates. This is effective on the accuracy of the classification rather than the estimation of whether the model is good or not. This is why the complexity matrix is used to explain the predictive evaluations of classification. The matrix that provides information about the real classes with the estimated classes performed via a classification model on the Test data is the complexity matrix [19]. In Table 4, the complexity matrix used in rice classification is given.

Table 4. Complexity matrix used in the classification of rice

		Predicted Rice	
		Cammeo	Osmancik
Actual Rice	Cammeo	tp	fp
	Osmancik	fn	tn

The complexity Matrix has four parameters, as seen in Table 4. These are;

- tp: true positive
- fp: false positive
- fn: false negative
- tn: true negative

For two-class classification performance measurements, success criteria such as Accuracy, Sensitivity, Specificity, Precision, F1-Score, Negative Predictive Value, False Positive Rate, False Discovery Rate, False Negative Rate are calculated using Table 4. Calculation formulas are given in Table 5 [20].

Table 5. Performance measurements and calculation formulas for two-class classification

No	Performance Measure	Formula
1	Accuracy	$ACC = \frac{tp+tn}{tp+fp+tn+fn} x100$
2	Sensivitiy	$TPR = \frac{tp}{tp+fn} x100$
3	Specificity	$SPC = \frac{tn}{tn+fp} x100$
4	Precision	$PPV = \frac{tp}{tp+fp} x100$
5	F1-Score	$F1 = \frac{2tp}{2tp+fp+fn} x100$
6	Negative Predictive Value	$NPV = \frac{tn}{tn+fn} x100$
7	False Positive Rate	$FPR = \frac{fp}{tn+fp} x100$
8	False Discovery Rate	$FDR = \frac{fp}{tp+fp} x100$
9	False Negative Rate	$FNR = \frac{fn}{tp+fn} x100$

2.6. Cross Validation

Cross-validation is a method of error prediction developed with the aim of improving the security of classification. Cross-validation divides the dataset such that it is random to a determined number of subsets for training and testing. It accepts one of the subsets as a test set, and the system is trained with the remaining sets. This process is repeated up to the number of dataset and the system is tested [21]. Figure 5 shows how cross-validation works.

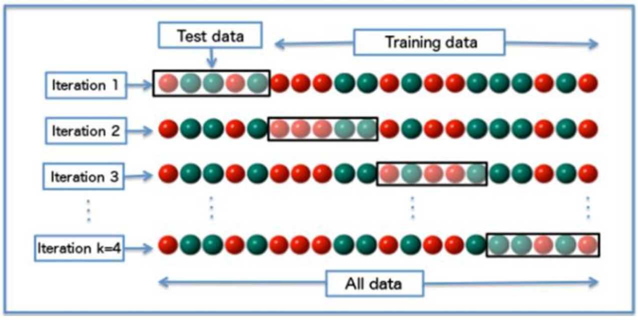

Fig. 5. An example for cross validation [22]

In the example given in Figure 5, the number of repeats (k) has been selected as 4. 25% of the dataset is divided for testing and 75% for training. The process is repeated for all data sets and the system test is completed.

3. Development of Modelling

Classification models are a method of high importance used in various fields. In class determination, classification models are used to determine which class the data belongs to. The classification model is a model that works by making predictions. The purpose of the classification is to make use of the common characteristics of the data to parse the data in question [23]. In our study, models were created using LR (Logistic Regression), MLP (Multilayer Perceptron), SVM (Support Vector Machine), DT (Decision Tree), RF (Random Forest), NB (Naive Bayes) and k-NN (K Nearest Neighbor) systems in order to classify rice grains according to their characteristics.

3.1. Logistic Regression (LR)

Logistic regression (LR) is one of the most widely used statistical models. Dependent variable in the LR is estimated from one or more variables. LR clarifies the relationship between dependent variables and independent variables. In LR there is no need for variables to require a normal distribution [24, 25]. Because the predicted values in LR are probabilities, LR is bounded by 0 and 1. The reason for this is that LR predicts probability rather than itself in the results [26].

3.2. Multiplayer Perceptron (MLP)

Today, many artificial neural network models have been developed for specific purposes. MLP is also one of the most widely used between these models. In MLP, neuron sequencing is in the form of layers. There are two main layers, as well as a hidden layer between them. MLP can contain multiple hidden layers. The input layer, the first of the main layers, contains information about the problem that needs to be solved. The second main layer, the output layer, gives out printouts of information processed within the network [27]. In Figure 6, MLP general structure is given.

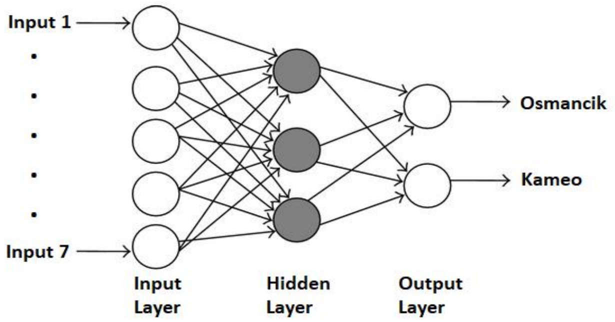

Fig. 6. MLP general structure

3.3. Support Vector Machine (SVM)

Support vector machine (SVM) is a core-based method that forms a hyper plane for classification and regression. By means of separation mechanisms, SVM can classify data as linear in two-dimensional space, planar in three-dimensional space and hyper plane in multidimensional space [28, 29]. SVM performs the classification process by finding the best hyper plane that separates the data from each other. The best hyper plane for an SVM is the one with the largest margin between the two classes, as shown in the Figure 7 [30].

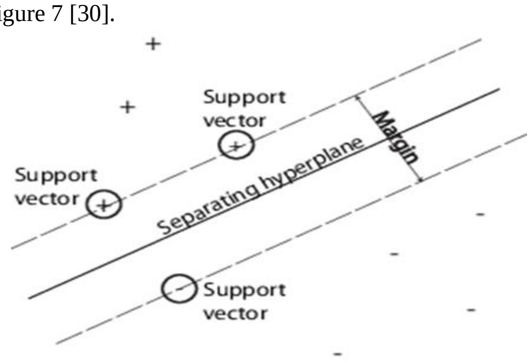

Fig. 7. SVM hyper plane

SVMs carry features similar to other machine learning algorithms. It is particularly similar to neural networks but bears more resemblance to the k-NN algorithm. Like the k-NN algorithm, SVM determines its neighbors based on the sample data presented to the algorithm and assumes that predictions are made for the new data [31].

3.4. Decision Tree (DT)

Decision trees (DT) is one of the first classification methods that come to mind along with neural networks in data mining. If DT is generally thought of as a tree diagram, it is branched out with a classification query in each of its branches and nodes [32, 33]. Figure 8 shows the general structure of DT.

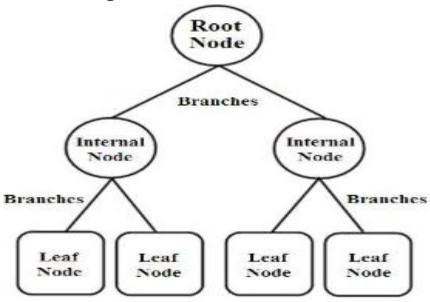

Fig. 8. DT general structure [34]

In DT, the root node represents the attribute. The inner node indicates the test or evaluation of a property. The branch represents the quality and shows the result of the evaluation in the internal node. The result node or leaf node are represents the classes and shows the final decision.

The properties of DT's in dealing with complex problems and their inferences in logical classification rules are seen as advantages [35]. It also makes DT stand out among other classification models due to its convenient installation, easy integration into databases and high reliability.

3.5. Random Forest (RF)

Random forest (RF) is a classifier consisting of many DT's. To make a new classification, each DT provides a classification for the inputs. Then RF evaluates the classifications and selects the estimate which has the most votes [36]. RF is capable of managing a large number of variables in a data set. It is also very good at estimating missing data. The biggest disadvantage of RF is the lack of reproducibility. In addition, it is also difficult to interpret the final model and its subsequent results. This is due to the fact that it contains many independent decision trees [37]. Figure 9 shows the general structure of RF.

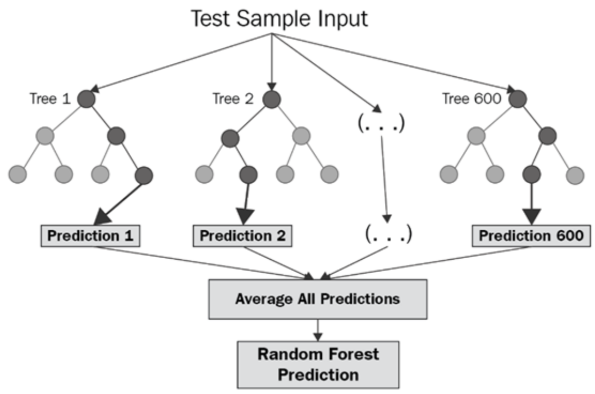

Fig. 9. RF general structure [38]

3.6. Naive Bayes (NB)

Naive Bayes (NB) is a classical probability classifier based on the Bayes theorem. The process of training NB can be carried out efficiently in a supervised learning environment, depending on the structure of the probability model. NB stands out as the most accurate classifier that can be used in situations where there is no adherence between a particular feature present in the system and other features [36]. In NB the learning module uses this model to predict the classification of a new sample by creating an estimated model of existing features [39].

3.7. K-Nearest Neighbor (k-NN)

K-Nearest Neighbors (k-NN) is an expanded standard machine learning method for large-scale training kits. In the k-NN algorithm, each point is conceptually plotted in a wide-dimensional space, where each axis in space corresponds to a different variable. Given that there is a certain amount of data to Test, the test data is processed one by one with all of the available data. The Test data will have many neighbors that are close to itself in terms of all the characteristics measured. For this reason, the closest k piece data to the test data is selected. As a result, from the selected data is looked at which class has more data and the data tested is said to belong to that class [40, 41]. In our study, the k value was chosen as 1.

The algorithm has some advantages and disadvantages. One of the most important advantages of the K-NN algorithm is efficiency and the other is flexibility. The algorithm is efficient in terms of simplicity, speed and scalability. The flexibility of the algorithm makes it easier to manage data sets that cannot be explained by linear or nonlinear relationships by the complexity of the relationships. The most important disadvantage of the k-NN algorithm is that, unlike other data mining algorithms, it does not give an idea of which variables are important in new predictions [31, 42].

4. Results

In order to classify the rice varieties used in our study, preliminary processing was applied to the pictures obtained with CVS and a total of 3810 rice grains were obtained.
Furthermore, 7 morphological features have been inferred for each grain. A data set has been created for the properties obtained. Models have been created and performances have been achieved with LR, MLP, SVM, DT, RF, NB and K-NN machine learning techniques for classification. Cross validation k value is selected 10 on all models used. In Table 6, the complexity matrix values for all the algorithms used are given.

Table 6. Complexity matrix of algorithms used

Algorithms	Confusion Matrix	
LR	1489	141
	125	2055
MLP	1484	146
	126	2054
SVM	1492	138
	135	2045
DT	1488	142
	144	2036
RF	1480	150
	140	2040
NB	1461	169
	147	2033
k-NN	1419	211
	224	1956

For classification performance metrics, success criteria such as Accuracy, Sensitivity, Specificity, Precision, F1-Score, Negative Predictive Value, False Positive Rate, False Discovery Rate and False Negative Rate are calculated using complexity matrix for each model. Classification performance measurement results are given in Table 7.

Table 7. Measurement results of classification performance

Measure	LR	MLP	SVM	DT	RF	NB	k-NN
Accuracy	93,02	92,86	92,83	92,49	92,39	91,71	88,58
Sensivitiy	92,26	92,17	91,70	91,18	91,36	90,86	86,37
Specificity	93,58	93,36	93,68	93,48	93,15	92,33	90,26
Precision	91,35	91,04	91,53	91,29	90,80	89,63	87,06
F1-Score	91,80	91,60	91,62	91,23	91,08	90,24	86,71
Negative Predictive Value	94,27	94,22	93,81	93,39	93,58	93,26	89,72
False Positive Rate	6,42	6,64	6,32	6,52	6,85	7,67	9,74
False Discovery Rate	8,65	8,96	8,47	8,71	9,20	10,37	12,94
False Negative Rate	7,74	7,83	8,30	8,82	8,64	9,14	13,63

As shown in Table 7, all models except k-NN have been achieved a classification success of over 90%. The 93.02% accuracy achieved in the LR model has the highest value among other models.

5. Discussion

Osmancik-97 and Cammeo rice species used in the study are patented rice grown in our country. By using the models and image processing techniques created for this study, dataset can be created about the species by making feature inferences for other patented species produced in our country. In this way, research can be carried out on the issues affecting the production result by obtaining information about the same type of rice cultivated in different regions.
In addition, automatic systems can be designed for many processes such as calibration of rice types and the separation of species from unwanted substances that may be present. In order to increase the success rate in classification, more images can be obtained from species and it is thought that success rates can be increased by using morphological features as well as color and shape features.

References

[1] Sapaloglu, A., *The structure of marketing channels and rice marketing margins in the rice production-consumption chain: the example of the province of Edirne.* 2015, Namik Kemal University.

[2] Mutafcilar, Z.C., *Molecular characterization of registered paddy varieties in Turkey*. 2019, Trakya University Graduate School Of Science.

[3] Yadav, B. and V. Jindal, *Monitoring milling quality of rice by image analysis.* Computers and Electronics in Agriculture, 2001. **33**(1): p. 19-33.

[4] Visen, N.S., et al. *Image analysis of bulk grain samples using neural networks*. in *2003 ASAE Annual Meeting*. 2003. American Society of Agricultural and Biological Engineers.

[5] Baykan, O., A. Babalik, and F. Botsalı, *Recognition of wheat species using artificial neural network.* 4. International Symposium On Advanced Technologies, Konya, September, 2005: p. 28-30.

[6] Dubey, B., et al., *Potential of artificial neural networks in varietal identification using morphometry of wheat grains.* Biosystems Engineering, 2006. **95**(1): p. 61-67.

[7] Zapotoczny, P., M. Zielinska, and Z. Nita, *Application of image analysis for the varietal classification of barley:: Morphological features.* Journal of Cereal Science, 2008. **48**(1): p. 104-110.

[8] Chen, X., et al., *Combining discriminant analysis and neural networks for corn variety identification.* Computers and electronics in agriculture, 2010. **71**: p. S48-S53.

[9] Babalik, A., et al. *Effects of Feature Selection Using Binary Particle Swarm Optimization on Wheat Variety Classification*. in *International Conference on Advances in Information Technology*. 2010. Springer.

[10] OuYang, A.-G., et al. *An automatic method for identifying different variety of rice seeds using machine vision technology*. in *Natural Computation (ICNC), 2010 Sixth International Conference on*. 2010. IEEE.

[11] Farahani, L., *Discrimination of some cultivars of durum wheat (Triticum durum Desf.) using image analysis.* International Research Journal of Applied and Basic Sciences, 2012. **3**(7): p. 1375-1380.

[12] Silva, C.S. and U. Sonnadara, *Classification of rice grains using neural networks.* 2013.

[13] Kaur, H. and B. Singh, *Classification and grading rice using multi-class SVM.* International Journal of Scientific and Research Publications, 2013. **3**(4): p. 1-5.

[14] Abirami, S., P. Neelamegam, and K.H. Thanjavur India, *Analysis of rice granules using image processing and neural network pattern recognition tool.* 2014.

[15] Pazoki, A., F. Farokhi, and Z. Pazoki, *Classification of rice grain varieties using two Artificial Neural Networks (MLP and Neuro-Fuzzy).* The Journal of Animal & Plant Sciences, 2014. **24**(1): p. 336-343.

[16] Szczypiński, P.M., A. Klepaczko, and P. Zapotoczny, *Identifying barley varieties by computer vision.* Computers and Electronics in Agriculture, 2015. **110**: p. 1-8.

[17] MathWorks. *Morphological Operations*. 2019 [cited 2019 09 September 2019]; Available from: https://ch.mathworks.com/help/images/morphological-filtering.html.

[18] PDD. *Osmacik and Baldo Rice*. 2019 08 September 2019]; Available from: http://www.pdd.org.tr/osmancik-ve-baldo-pirinc.

[19] Cataloluk, H., *Disease diagnosis using data mining methods on actual medical data.* 2012, Bilecik University, Graduate School Of Science.

[20] Ozkan, I. A., and Koklu, M. , *Skin Lesion Classification using Machine Learning Algorithms.* International Journal of Intelligent Systems and Applications in Engineering, 2017. 5(4): p. 285-289.

[21] Gupta, P. *Cross-Validation in Machine Learning*. 2017 [cited 2019 10 September 2019]; Available from: https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f.

[22] Wikipedia. *Cross-validation (statistics)*. 2019 [cited 2019 12 September 2019]; Available from: https://en.wikipedia.org/wiki/Cross-validation_(statistics).

[23] Ozkan, Y., *Data Mining Methods,* Papatya Publishing and Education Inc. Istanbul, Turkey, 2008.

[24] Cruyff, M.J., et al., *A review of regression procedures for randomized response data, including univariate and multivariate logistic regression, the proportional odds model and item response model, and self-protective responses*, in *Handbook of Statistics*. 2016, Elsevier. p. 287-315.

[25] Kalantar, B., et al., *Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN).* Geomatics, Natural Hazards and Risk, 2018. **9**(1): p. 49-69.

[26] Wikipedia. *Logistic regression*. 2019 [cited 2019 11 September 2019]; Available from: https://en.wikipedia.org/wiki/Logistic_regression.

[27] Sabanci, K. and M.F. Unlersen, *Classification of Apples of Different Sizes using KNN and MLP Algorithms.* Proceedings Book, 2016: p. 781.

[28] Abhang, P., B. Gawali, and S. Mehrotra, *Technical aspects of brain rhythms and speech parameters.* Introduction to EEG-and Speech-Based Emotion Recognition, 2016: p. 51-79.

[29] Kaynar, O., et al. *Emotion Analysis with machine learning methods.* in *International Artificial Intelligence and Data Processing Symposium (IDAP'16), September*. 2016.

[30] Mathworks. *Support Vector Machines for Binary Classification*. 2019 [cited 2019 12 September 2019]; Available from: https://ch.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html#bsr5b42.

[31] Miner, L., et al., *Practical predictive analytics and decisioning systems for medicine: Informatics accuracy and cost-effectiveness for healthcare administration and delivery including medical research*. 2014: Academic Press.

[32] Soylu, E. *Data Mining*. 2019 [cited 2019 08 September 2019]; Available from: http://web.karabuk.edu.tr/emelkocak/indir/MTM326/veri_madenciliği.pdf .

[33] Ozkan, I. A., Koklu, M., and Sert, I. U., *Diagnosis of urinary tract infection based on artificial intelligence methods.* Computer methods and programs in biomedicine, 2018. 166, p. 51-59.

[34] Sá, A., et al. *Lightning Forecast Using Data Mining Techniques on Hourly Evolution of the Convective Available Potential Energy*. in *Brazilian Congress on Computational Intelligence, Fortaleza, November*. 2011.

[35] Amor, N.B., S. Benferhat, and Z. Elouedi, *Qualitative classification with possibilistic decision trees*, in *Modern Information Processing*. 2006, Elsevier. p. 159-169.

[36] Mao, W. and F. Wang, *New advances in intelligence and security informatics*. 2012: Academic Press.

[37] Dunham, K., *Chapter 6-Phishing, SMishing, and Vishing.* Mobile Malware Attacks and Defense, 2009: p. 125-196.

[38] Chakure, A. *Random Forest Regression*. 2019 [cited 2019 12 September 2019]; Available from: https://towardsdatascience.com/random-forest-and-its-implementation-71824ced454f.

[39] Panigrahi, R. and S. Borah, *Classification and Analysis of Facebook Metrics Dataset Using Supervised Classifiers.* Social Network Analytics: Computational Research Methods and Techniques, 2018: p. 1.

[40] Nadkarni, P., *Clinical Research Computing: A Practitioner's Handbook*. 2016: Academic Press.

[41] Richman, J.S., *Multivariate neighborhood sample entropy: a method for data reduction and prediction of complex data*, in *Methods in enzymology*. 2011, Elsevier. p. 397-408.

[42] Sabanci, K., and Koklu, M., *The classification of eye state by using kNN and MLP classification models according to the EEG signals.* International Journal of Intelligent Systems and Applications in Engineering, 2015. 3(4), p. 127-130.