Journal of Agricultural Sciences
(Tarim Bilimleri Dergisi)

J Agr Sci-Tarim Bili
e-ISSN: 2148-9297
jas.ankara.edu.tr

JAS

Meta-Learning-Based Prediction of Different Corn Cultivars from Color Feature Extraction

Abdullah BEYAZ[a], **Dilara KOC**[a]

[a]*Ankara University, Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, 06110, Dışkapı, Ankara, TURKEY*

ARTICLE INFO
Research Article
Corresponding Author: Abdullah BEYAZ, E-mail: abeyaz@ankara.edu.tr
Received: 18 May 2019 / Revised: 20 August 2019 / Accepted: 24 September 2019 / Online: 18 January 2021

ABSTRACT

Image analysis techniques are developing as applicable to the approaches of quantitative analysis, which is aimed to determine cultivar grains. Additionally, corn (*Zea mays*) grain processing companies evaluate the quality of kernels to determine the price of these cultivars. Because of this reason, in the study, a computer image analysis technique was applied on three corn cultivars. These were *Zea mays L. indentata*, *Zea mays L. saccharata* and a hybrid corn (Yellow sweet corn). These cultivars are commercially important as dry grains in Turkey. In the study, the grain color values were tested in the cultivars from Turkey's collection. One hundred samples were used for each corn cultivar, and 300 corn grains in total were used for evaluations. Each of nine color parameters (R_{min}, R_{mean}, R_{max}, G_{min}, G_{mean}, G_{max}, B_{min}, B_{mean}, B_{max}) which were obtained from original RGB color channels with maximum and minimum values was evaluated from the digital images of three different corn cultivar grains. The values were analyzed with the help of the Multilayer Perceptron (MLP), Decision Tree (DT), Gradient Boost Decision Tree (GBDT) and Random Forest (RF) algorithms by using the Knime Analytics Platform. The majority voting method was applied to MLP and DT for prediction fusion. All algorithms were run with a 10-fold cross-validation method. The success of prediction accuracy was found as 99% for RF and GBDT, 97.66% for MLP, 96.66% DT and 97.40% for Majority Voting (MAVL). The MAVL method increased the accuracy of DT while decreasing the accuracy of MLP partly for the fusion of MLP and DT.

Keywords: *Corn identification; Color measurement; Multilayer Perceptron; Decision Tree; Gradient Boost Decision Tree; Random Forest*

1. Introduction

In recent years, agricultural production has become a more critical sector, because human needs have increased in terms of food, energy and raw industrial materials. This increase has resulted in another increase in agriculture as technological practices. In modern agricultural production, in general, experts take an important role in getting information and making decisions, but this is not always possible. Because of this reason, expert systems are used for solving problems in agricultural production. One area of agricultural production which uses these expert systems is corn production. Corn is an essential product in the world's agriculture. This is because it may be used as human food and for animal breeding. The reason for this issue may be understood from FAOSTAT production quantities of corn for countries as the average of the period of 1994-2017 (Figure 1). In Figure 1, it can be seen that Northern America and Europe have the mass production level of corn of more than 5 863 440 tons. Additionally, the production share of corn by region as the average of the period of 1994-2017 may be seen in Figure 2. It can be observed that America produces 52.5% of corn in the world.

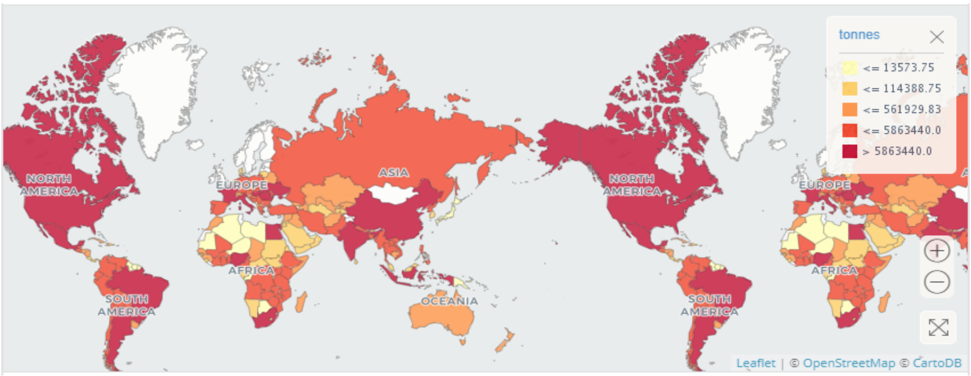

Figure 1- Production quantities of corn by country as the average of 1994 – 2017 (FAO 2019a)

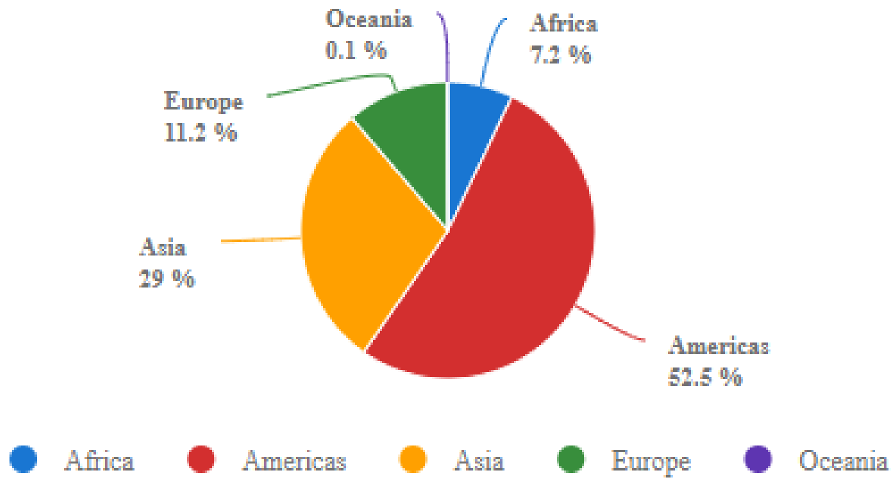

Figure 2- Production share of corn by region as the average of 1994 – 2017 (FAO 2019b)

The production of corn in agriculture reaches big amounts. This value varies from country to country in the world. The production area and production amount of corn in Turkey for 2001-2017 may be seen in Table 1 (TSI 2018). As seen in Table 1, Turkey produces 5 900 000 tons of corn from 639 084 hectares of production area. This statistic also shows that corn production is critical and has a large share amounts in Turkey's agricultural production like the case in the world. With this amount of production, the issue of corn, also corn cultivars, is becoming important for seed production of the world, as well as Turkey. Seed properties effect all agricultural production processes, as well as the quality and quantity of corn. The properties of corn cultivars are essential, and knowing about corn cultivars directly affects the agricultural production based on their production needs like water, soil and climate. It is the basic issue of seed studies in the scientific literature and the agriculture industry.

Table 1- Area and production of corn in Turkey (TSI 2018)

	Corn	
Year	*Sown area (Hectare)*	*Production (Tons)*
2001	550 000	2 200 000
2002	500 000	2 100 000
2003	560 000	2 800 000
2004	545 000	3 000 000
2005	600 000	4 200 000
2006	536 000	3 811 000
2007	517 500	3 535 000
2008	595 000	4 274 000
2009	592 000	4 250 000
2010	594 000	4 310 000
2011	589 000	4 200 000
2012	622 609	4 600 000
2013	659 998	5 900 000
2014	658 645	5 950 000
2015	688 169	6 400 000
2016	680 019	6 400 000
2017	639 084	5 900 000

In the scientific literature, Dechao (1996) worked on 11 different corn cultivars and their broken corn kernels positioned randomly in one frame, and then, they selected the morphological parameters detected with the BP moment learning algorithm. For this aim a three-layer feed forward neural network was formed, which could identify multiple or whole and broken corn kernels. Their classification accuracy was 93%. Ni et al. (1997) developed a machine-vision system to identify different types of corn kernels as crown and shapes. They used image processing techniques to enhance the object detection properties of images, and they reduced noise in the acquired images. Their system provided an average accuracy, approximately 87%, when they compared their results to human inspection. Tan´ska et al. (2005) and Stanisavljević et al. (2019) worked on measurement of the geometrical characteristics, size and surface colors of rapeseeds with the help of digital image and color analysis, and they expressed that an analysis of color of rape and sticky willy seeds in an RGB (red/green/blue) model showed differences in value ranges for distinguishing the seeds. Zhang et al. (2007) studied digital image processing for identification and detection of corn kernel surface cracks. A detection experiment was carried out by 50 kernels with cracks and 50 kernels without cracks. Their results indicated 90% to 94% detecting accuracy. Draganova et al. (2010) investigated an approach for identifying fusarium-infected maize grains by spectral analysis in the visible and near-infrared region, SIMCA models, parametric classifiers and neural classifiers, they also reported their system accuracy was good. Zhao et al. (2011) used a genetic algorithm and support vector machine (SVM) to determine species. Their methods were optimized for cultivar recognition, and their algorithm was based on machine vision, which also improved determination accuracy with a performance percentage of 94.4%. Jiang et al. (2012) worked on corn seed purity, where their system was based on machine vision, and it was divided into three steps: the first step was image segmentation, the second step was feature extraction, and the third step was classification of corn seeds. Based on a classifier designed with SVM, their results showed 97.3% identification accuracy in the Nongda108 and 98% accuracy in the Ludan981 cultivars, better than 95% in previous studies. Yang et al. (2015) worked on classification of the purity of waxy corn seed varieties. Their study was based on the combined spectral, morphological and texture features extracted from visible and near-infrared (VIS/NIR) hyperspectral images. 150 kernels of each variety were captured and analysed with the images of both sides of corn kernels. Support vector machines (SVM) and a model of partial least squares–discriminant analysis (PLS-DA) were used to build the classification models for classification of seed varieties. The recognition accuracy was 98.2% and 96.3% for the germ side and the endosperm side, respectively, in the SVM model. It was more satisfactory than in the PLS-DA model according to their research. In their research, they also stated that their procedure has the potential for use as a new method for seed purity testing.

In the light of scientific literature and corn production needs of the world, also Turkey's agriculture, this study aimed to classify some corn cultivars based on their color values. Moreover, it was aimed to demonstrate the success of a potential industrial application with a pre-application by using image processing and analysis techniques in corn cultivar grains. It is also of great importance that the findings will be a source of inspiration for future studies and its contribution to the literature.

2. Material and Methods

In the study, an image analysis technique was applied on three corn cultivars. These were *Zea mays L. indentata*, *Zea mays L. saccharata* and hybrid corn (Yellow sweet corn). These cultivars are commercially important as dry grains in Turkey. One hundred samples were used for each corn cultivar. Corn grains were randomly selected from each variety, and a total of 300 corn grains were used for the evaluations. The corn cultivar grain samples can be seen in Figure 3.

Figure 3- Corn grain samples

2.1. Image processing and analysis

Corn grain images were captured by using a DSLR camera (Nikon D800) with an illuminated background. Images were acquired at a resolution setting of 300 dpi (1600 x 2400 pixels). In the study, color features were used for classifying grain varieties because it is known that color channels are widely used for this aim in the literature, while it is also known that the colors of a corn grain are not quite uniform. Because of this reason, each of the three-color channels with nine color parameters (R, G, B values which are obtained from the original RGB color channel with mean, maximum and minimum values) were evaluated from digital images of the grains of three different corn cultivars. These corn cultivar color channel values were evaluated by using the NI Vision Assistant software for classification of corn grains (Figure 4.). The reason for usage of this software was that the software has a magic wand tool for selection of corn grains, in which way the method does not need to isolate corn grains from their background by using a given threshold in contrast to the literature (Chen et al. 2010). The magic wand tool was set as 20 for edge detection of corn grains. Additionally, with the use of this software, the normalization needs of each color channel were eliminated in contrast to the literature (Chen et al. 2010). In this study, as a pre-application, before developing an industrial automation system, it was aimed to only determine the efficiency of the measurements with the help of the LabVIEW Vision Assistant module and other modules of the software for developing an optimum image measurement system.

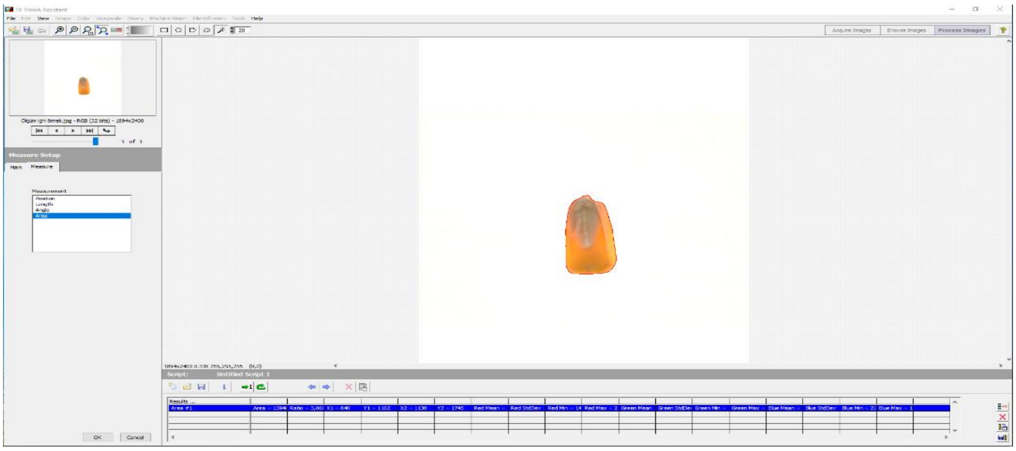

Figure 4- The evaluation process of color channel values of corn grain by using NI Vision Assistant software

All classification processes can be summarized as the steps of getting original corn images from corn cultivars, transferring these images to the NI Vision Assistant software, measurement setup process of the NI Vision Assistant software, area measurements, obtaining RGB data from digital images with mean, maximum and minimum values, saving the database as a txt file, txt to xls datasheet conversion for meta-learning operations, meta-learning of corn database, and getting corn cultivar classification results. These steps may also be seen in Figure 5.

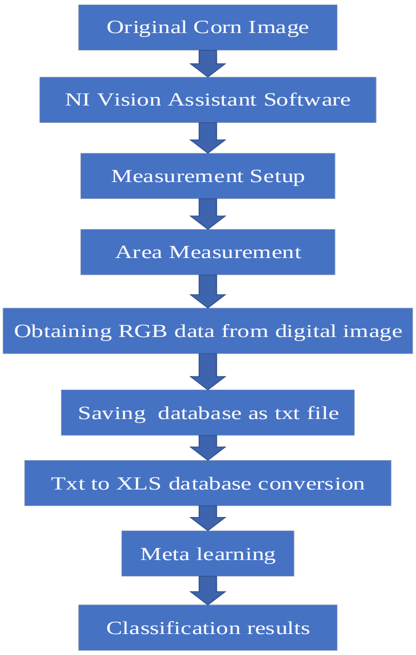

Figure 5- Classification steps of corn cultivars from digital images

2.2. Multilayer perceptron neural network

Since 1980, ANN has made progress thanks to developments in computer science. In addition to classification operations, it is also used in clustering and pattern definition operations. ANN refers to computer systems that perform the learning function of the human brain, just like the biological nervous system. In artificial intelligence, neural networks are frequently used in a wide range of implementations. A multilayer perceptron (MLP) is a feed-forward neural network structure that maps input training objects to target labels (Öztemel 2012; Silahtaroğlu 2016). In ANN, information is distributed over the connection weights between neurons. The hidden layer input is expressed as follows (Fakı 2015)

$$Hidden\ layer_{input} = w_1x_1 + w_2x_2 + w_3x_3 \tag{1}$$

In order to increase the ability of the ANN classifier to generalize learning and prevent overfitting, a small number of neurons and hidden layers were used to obtain the most appropriate network structure. Different ANNs were trained according to the algorithms used in this study. The ANN structure used in the study consisted of three hidden layers with 100 iterations and 10 neurons in each hidden layer (Table 2).

Table 2- Different ANN training results

Number of hidden layers	*Accuracy (%)*
1	96
2	97.33
3	97.66
4	96.66
5	97
6	92
7	77.33
8	84
9	79.66
10	70
15	61
25	33.66

2.3. Decision tree

Decision trees (DT) constitute an estimation model that represents the relationships between the properties and object values. They are frequently used in data mining methods. Decision trees, by sorting important attributes down the tree from the root to a leaf node, provided the classification of the instance. While each node is a decisive decision, each branch is finalized and results in a leaf. This process is then repeated for the subtree rooted at the new node. No branch continues with another branch (Mitchell 1997; Pandya & Pandya 2015; Köse 2018). In this study, for the Decision Tree, the gain ratio was selected as a measured quality, and Minimum Description Length was selected as a pruning technique.

The C4.5 tree is an improved version of the ID3 tree. It is an algorithm based on entropy and information gain same as the ID3 algorithm. Unlike the ID3 algorithm, pruning is performed in this algorithm (Köse 2018). Entropy is calculated by the following equation:

$$H(S) = \sum_{t=1}^{n} \rho i * \log_2(\rho i) \quad (2)$$

H: Entropy,
S: Source,
p: Probability (Silahtaroğlu 2016).
Gain information is obtained by calculating the differences between the weighted sums of the entropies of each sub-section (Silahtaroğlu 2016). The gain information is calculated by the following equation:

$$D = H(D) - \sum_{t=1}^{n} P(D_i)\, H(D_i) \quad (3)$$

D: Gain information,
H: Entropy,
p: Probability (Silahtaroğlu 2016).

2.4. Gradient boosted decision trees

Boosting is a prediction algorithm in machine learning based on the idea of combining a set of weak learners to create a single strong learner. Boosted trees is a classifier that is a combination of Boosting and Decision Trees in Meta-learning algorithms (Gupte et al. 2014; Kim et al. 2015). Likewise, GBDT is an ensemble method and a powerful supervised machine-learning technique that has been widely used in recent times mainly due to its high accuracy (Si et al. 2017). Formula 4 was used for multiple classification (Li et al. 2008):

$$\sum_{i=1}^{N} \sum_{k=0}^{K-1} -\log\left(P_{i,k}\right) 1_{yi=k} \quad (4)$$

2.5. Random forest

The random forest algorithm was developed by Leo Bieman, and it produces more than one tree to solve a question and creates different decision trees. RF is an ensemble learning method in meta-learning. These trees are an advanced version of the CART algorithm where many trees are created based on subsets of data. It is one of the popular meta-learning methods that provides simple and fast results in terms of understanding and application based on the collection of estimates from many decision trees.

It selects the observations and properties randomly to build various decision trees and takes the average of the results (Mitchell 2011). The Gini index is calculated through the following equation for producing a tree with the criteria of the RF (Classification and Regression tree) algorithm (Küçükönder et al. 2015).

$$\sum\sum j \neq i(f(C_i,T)/T)(f(C_i,T)/T) \tag{5}$$

2.6. Majority voting method in meta-learning

Meta-learning is an advanced technique in data mining that deals with the problem of computing a global classifier from the distributed database. In other words, this term is generally defined as learning from learned knowledge. Meta-Learning aims to learn from the predictions of classifiers on a common validation data set. At the end of classification, the meta-classifier works to train from the common validation data set (Figure 6) (Prodromidis et al. 2000). The selected Majority Voting (MAVL) method was deemed appropriate for showing correct high predictions. For this purpose, the 'Prediction Fusion' node in KNIME was connected between MLP and DT for majority voting application (Şeker & Erdoğan 2018). All nodes were connected as demonstrated in Figure 7. All algorithms were run with a 10-fold cross-validation method. At the end of the analysis, the accuracy, True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), precision, recall, and F-means values were given in the relevant part of the study. Accuracy shows how accurately the system can predict by using the following formula (Takran et al. 2017):

$$Accuracy = \frac{TP}{(TP+TN+FP+FN)} \tag{6}$$

Recall or sensitivity or true positive rate (TPR) is the value that the system is predicting what rate is the passing assessment of the overall passing assessments. The recall is evaluated by using the following formula (Takran et al. 2017):

$$Recall = \frac{TP}{(TP+FP)} \tag{7}$$

Precision or Predicted Position Value (PPV) is the value which indicates how truly or correctly the system can predict (Takran et al. 2017):

$$Precision = \frac{TP}{(TP+FP)} \tag{8}$$

F-Measure is the overall efficiency assessment derived from the means of Precision and Recall (Takran et al. 2017):

$$F-Measure = \frac{2xPrecisionxRecall}{Precision+Recall} \tag{9}$$

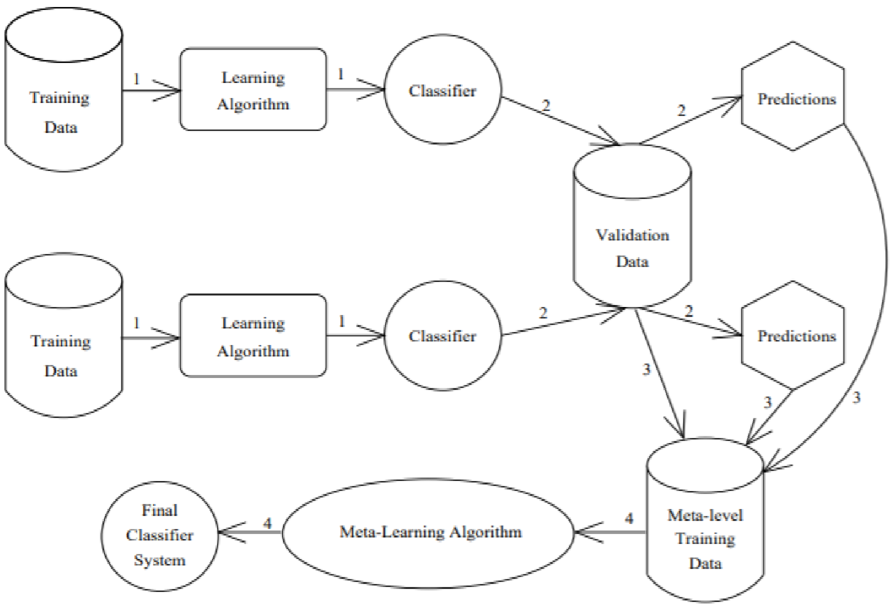

Figure 6- Meta-learning (Prodromidis et al. 2000)

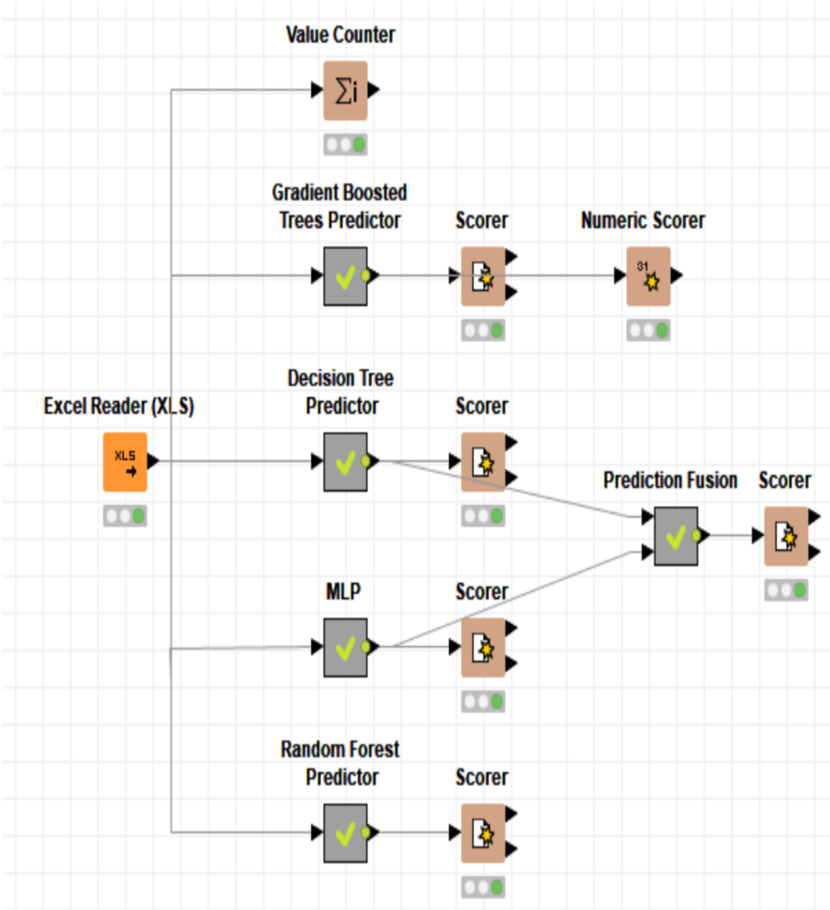

Figure 7- Workflow in KNIME

2.7. Training, validation and testing sets

Most machine learning techniques tend to overfit (Dietterich 1995; Shin et al. 2019). A way to prevent overfitting is to determine the optimum hyperparameter based on the data structure and use a large amount of training data. Moreover, adding validation steps or using well generalized architectures may also achieve success (Piotrowski & Napiorkowski 2013; Shin et al. 2019). In the study, an outer k-fold cross-validation loop was used for separating the data into training and testing folds in nested cross-validation and an inner loop for k-fold cross-validation with validation folds in the training folds. In order to evaluate the model, an independent testing dataset for the final model was kept, and these data were not provided before. The final accuracy was defined as the mean value of ten test accuracies obtained from ten training sessions, and this was 10-fold cross-validation. Whenever the training was running, the training datasets were randomly selected from the whole dataset (Shin et al. 2019).

The 10-Fold Cross Validation method refers to trying verification k times. In this method, for each 1/k of the data set, the previously unused part of the data set is used for testing, while the rest is used for training. Before applying the method, the k parameter must be determined. The k parameter specifies how many parts the data set will be divided into. k classification procedures are performed, and one of the parts divided at each step is reserved for the test process, while the remaining k-1 are used for the training of the classifier. The general classification result is obtained from the average of the classification results after k steps.

3. Results and Discussion

In the dataset, the R, G and B color channels were averaged to show the differences of all corn cultivars. The average color differences of corn cultivars can be seen in Figure 8. The confusion matrix and accuracy statistics of the supervised machine learning algorithms that were used are given in Tables 3 respectively. The success of prediction accuracy was found as 99% for RF and GBDT, 97.66% for MLP, 96.66% for DT and 97.40% for MAVL. The GBDT, RF, MLP, DT and MAVL algorithms were used, and in the comparison of the classification models, the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Root Relative Squared Error (RRSE), Relative Absolute Error (RAE) values were low, while the classification accuracy rate was high. As a result of the comparison, it was found that the classification model [MAE: 0.004, RMSE: 0.045% RAE:

0.915 and % RRSE: 9.458], which was formed based on the Random Forest algorithm, was a better classifier than the other algorithms with a prediction rate of 99%.

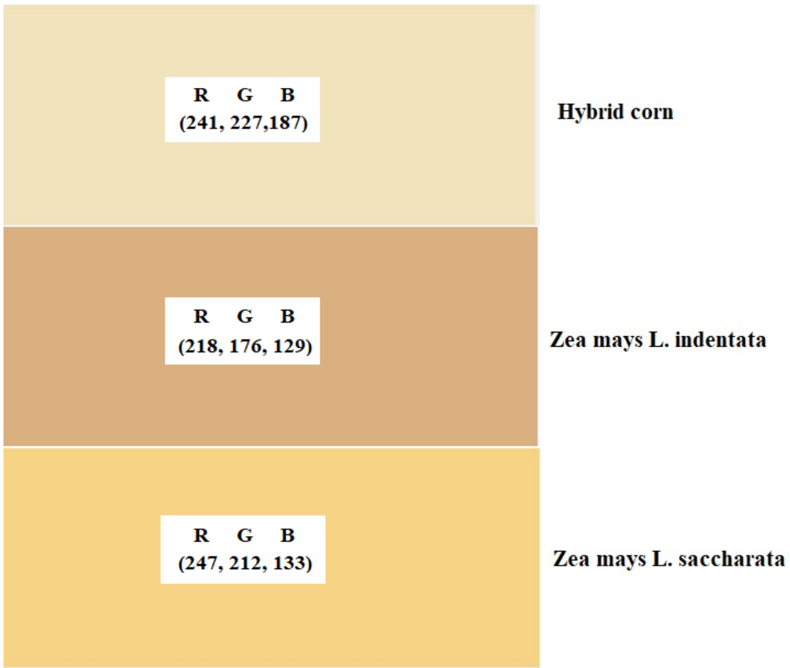

Figure 8- The average color values of corn cultivars for showing color differences (Anonymous 2019)

Table 3- Confusion matrix of the algorithms that were used

Algorithms	*Corn Various*	*Hybrid Corn*	*Indentata*	*Saccharata*	*Accuracy*
	Hybrid Corn	100	-	-	
GBDT	Indentata	1	99	-	99%
	Saccharata	2	-	98	
	Hybrid Corn	100	-	-	
Random Forest	Indentata	-	100	-	99%
	Saccharata	3	-	97	
	Hybrid Corn	97	2	1	
MLP	Indentata	3	97	-	97.66%
	Saccharata	1	-	99	
	Hybrid Corn	99	1	-	
Decision Tree	Indentata	4	95	1	96.66%
	Saccharata	3	1	96	
	Hybrid Corn	10284	385	106	
MAVL	Indentata	192	9342	-	97.40%
	Saccharata	97	-	9603	

The purpose of the meta-learning method is to contribute to high accuracy results when the low accuracy algorithm is evaluated together with another high accuracy algorithm. For this reason, Decision Tree, which has a low accuracy value, and Artificial Neural Network, which provides high accuracy values, were combined. As a result of MAVL, the decision tree prediction which had low accuracy was close to the artificial neural network prediction. The ANN and DT algorithm result was partly increased as from 96.66% to 97.40%. The disadvantage of the fusion process was that the ANN result was partly decreased as from 97.66% to 97.40%.

In the literature, Chen et al. (2010) worked on variety identification based on machine vision and pattern recognition for identifying five Chinese corn varieties according to their external features. They expressed that their classification accuracies of corn varieties (BAINUO 6, NONGDA 86, NONGDA 108, GAOYOU 115, and NONGDA 4967) were 100, 94, 92, 88 and 100%, respectively, according to discriminant and neural networks analysis. Kurtulmus & Ünal (2015) investigated distinguishing rapeseed varieties using computer vision and machine learning based on SVM. They reported that the developed computer vision system provided an overall accuracy rate of 99.24% for the best predictive model in discriminating rapeseed variety. Additionally, Draganova et al. (2010) examined an approach for identifying fusarium-infected maize grains by spectral analysis in the visible and near-infrared region, SIMCA models, parametric classifiers and neural classifiers, and they stated that their recognition accuracy which was achieved for both classes of grains was 99.89% for the healthy and 93.7% for the infected specimens.

According to DT, *Zea mays L. saccharata* was a root node, and the distinctive attribute was G_{mean}. The decision tree visual results of Hybrid (Yellow sweet corn), *Zea mays L. Indentata* and *Zea mays L. saccharata* grains are given in Figure 9. As seen in Figure 9, the $Green_{mean}$ parameter provided a distinguishing feature in identification of the corn cultivars. It was seen that the cultivars could be identified at the values under and above the average value of 187.27, and then, the $Blue_{min}$ parameter was taken as the second order. It is possible to say that all cultivars under $Green_{mean}$ 90 and $Blue_{min}$ 5.50 could be classified as Hybrid corn.

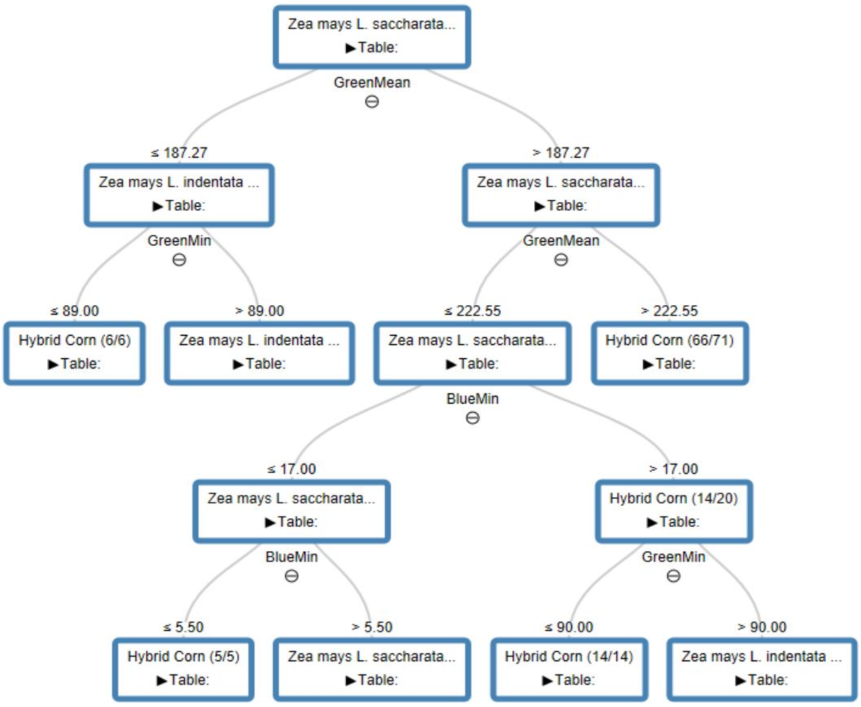

Figure 9- Decision tree visual results

4. Conclusions

As a conclusion, according to our research results and the literature, corn cultivars can be identified with the help of RGB color data, and these data may also be used for detecting the cracked and diseased ones among the healthy products. So, here, it can be clearly emphasized that, based on the differences and similarities, this algorithm can be used to distinguish different cultivars from other regions in future studies. However, it is not easy to argue that it could be possible to describe bulk products for increasing agricultural product quality. This is because, in bulk products, there are a lot of different parameters for color extraction like depth and 3D measurements. Furthermore, it may be stressed that the particular advantage of using meta learning here was making an optimization. Additionally, we think our results will be a source for future studies, and the contribution of this study to the literature is essential.

Abbreviations and Symbols	
TP	True positives
FP	False positives
TN	True negatives
FN	False negatives
GBDT	Gradient boosted decision trees
MLP	Multilayer perceptron neural network
MAVL	Majority voting
DT	Decision tree
RF	Random forest

References

Anonymous (2019). Colors RGB. Retrieved in January, 14, 2019 from https://www.w3schools.com/colors/colors_rgb.asp

Chen X, Xunb Y, Li W, Zhanga J (2010). Combining discriminant analysis and neural networks for corn variety identification. *Computers and Electronics in Agriculture* 71S (2010): 48-53

Dechao S T Z (1996). An investigation on morphological discrimination of corn kernels based on neural network. *Transactions of The Chinese Society of Agricultural Engineering 1*

Dietterich T (1995). Overfitting and undercomputing in machine learning. *ACM Computing Surveys* 27(3): 326-327

Draganova T, Daskalov P, Tsonev R (2010). An approach for identifying of fusarium infected maize grains by spectral analysis in the visible and near infrared region, simca models, parametric and neural classifiers. *International Journal Bioautomation* 14(2): 119-128

Fakı B M (2015). Classification of anemia using data mining methods: an application. MSc. Thesis, Istanbul Technical University, Graduate School of Natural and Applied Sciences, Department of Industrial Engineering, Engineering Management Division, TURKEY

FAO (2019a). Production quantities of maize by country as the average of 1994-2017 years. Retrieved in January, 11, 2019 from http://www.fao.org/faostat/en/?#data/QC/visualize

FAO (2019b). Production share of maize by region as the average of 1994-2017 years. Retrieved in January, 11, 2019 from http://www.fao.org/faostat/en/?#data/QC/visualize

Gupte A, Joshi S, Gadgul P, Kadam A (2014). Comparative study of classification algorithms used in sentiment analysis. *(IJCSIT) International Journal of Computer Science and Information Technologies* 5 (5): 6261-6264

Jiang J, Wang Y, Yang R & Mei S (2012). Variety identification of corn seed based on bregman split method. *Transactions of the Chinese Society of Agricultural Engineering* 28(1): 248-252

Kim T, Lee D, Choil J, Spurlock A, Sim A, Todd A, Wu K (2015). Extracting baseline electricity usage with gradient tree boosting. *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, Chengdu, China

Köse İ (2018). Data mining theory practice and philosophy. Papatya Education Publishing, ISBN: 6059594349, Istanbul, Turkey

Kurtulmus F & Ünal H (2015). Discriminating rapeseed varieties using computer vision and machine learning. *Expert Systems with Applications* 42(2015):1880-1891

Küçükönder H, Vursavuş K K & Üçkardeş F (2015). Determining the effect of some mechanical properties on color maturity of tomato with k-star, random forest and decision tree (C4. 5) classification algorithms. *Turkish Journal of Agriculture-Food Science and Technology* 3(5): 300-306

Li P, Wu Q & Burges C J (2008). Mcrank: Learning to rank using multiple classification and gradient boosting. In Advances in Neural Information Processing Systems, pp. 897-904

Mitchell M W (2011). Bias of the random forest out-of-bag (OOB) error for certain input parameters. *Open Journal of Statistics* 2011(1):205-211

Mitchell T M (1997). Machine learning. McGraw-Hill Inc., New York, USA

Ni B, Paulsen M R & Reid J F (1997). Corn kernel crown shape identification using image processing. *Transactions of the ASAE* 40(3): 833-838

Öztemel E (2012). Artificial Neural Networks. 3rd Edition, Papatya Publishing, ISBN: 975-6797-39-8, Istanbul, Turkey

Pandya R & Pandya J (2015). C5.0 Algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications* 117(16): 18-21

Piotrowski A P & Napiorkowski J J (2013). A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. *Journal of Hydrology* 476: 97-111

Prodromidis A, Chan P, Stolfo S (2000). Meta-learning in distributed data mining systems: Issues and approaches, In Advances in Distributed and Parallel Knowledge Discovery, AAAI/MIT Press.

Shin J, Chang Y K, Nguyen-Quang T, Heung B & Ravichandran P (2019). Optimizing parameters for image processing techniques using machine learning to detect powdery mildew in strawberry leaves. *American Society of Agricultural and Biological Engineers, In 2019 ASABE Annual International Meeting,* pp. 1

Si S, Zhang H, Keerthi S S, Mahajan D, Dhillon I S, Hsieh C-J (2017). Gradient boosted decision trees for high dimensional sparse output. *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, PMLR 70

Silahtaroğlu G (2016). Data Mining - Concepts and Algorithms. 3. Printing, Papatya Publishing, ISBN: 978-975-6797-81-5, Istanbul, Turkey

Şeker Ş E & Erdoğan D (2018). End-to-end data science with KNIME. 1. Edition, Demet Erdoğan Publishing House

Stanisavljević R, Knežević J, Tomić D, Jovanović D, Tmušić N, Štrbanović R, Poštić D (2019). Seed Quality of Oilseed Rape Varieties with Different Size and Colors After Three and Fifteen Months Storage. *Journal of Agricultural Sciences* 25(4): 449-458 DOI: 10.15832/ankutbd.442650

Takran T, Chartrungruang B, Tantranont N, Somhom S (2017). Constructing a thai homestay standard assessment model by implementing a decision tree technique. *International Journal of the Computer, the Internet and Management* 25(2): 106-112

Tan´ska M, Rotkiewicz D, Kozirok W, Konopka I (2005). Measurement of the geometrical features and surface color of rapeseeds using digital image analysis. *Food Research International* 38 (2005): 741-750

Turkish Statistics Institute (TSI) (2018). Area and production of cereals and other crop products (For selected products). Retrieved in January, 10, 2019 from http://www.tuik.gov.tr/UstMenu.do?metod=temelist

Yang X, Hong H, You Z, Cheng F (2015). Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors* 2015(15): 15578-15594

Zhang J X, Xun Y & Li W (2007). Detection of surface cracks of corn kernel based on morphology. *Optics and Precision Engineering* 15(6): 951-956

Zhao M, Wu W, Zhang Y Q & Li X (2011). Combining genetic algorithm and SVM for corn variety identification. *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, 19-22August, Jilin, China, pp. 990-993