Research Article

Classification of Wheat Rootstock and Their Hybrids According to Color Features by Machine Learning Algorithms

Mesut Ersin Sonmez [a],* iD, Kadir Sabanci , [b] iD, Nevzat Aydin[a] iD

aDepartment of Bioengineering, University of Karamanoglu Mehmetbey, Karaman 70200, Turkey
bDepartment of Electrical Electronics Engineering, University of Karamanoglu Mehmetbey, Karaman 70200, Turkey

ABSTRACT

Accurate classification of wheat varieties has a large economic market in the world is enabled both high income in the market and the development of new fertile hybrids for changing weather conditions due to global warming. In this study, instead of using the conventional classification method, we extracted color features of the 1400 wheat grain samples, consisting of Ahmetbugdayi, Cesare and their hybrids BC1F6 and BC2F5, by using image processing techniques. For the color features, each twelve channels of four different color spaces were used and square shaped samples were taken from the center of all the grains in these channels of images. the averages of the channel pixels values were used as color features. Then six different machine learning algorithms were employed for the classification task. Artificial Neural Networks (ANN), Support Vector Machine (SVM) and Decision Tree (DT) models achieved more than 99% accuracy. On the other hand, k-Nearest Neighborhood (k-NN) and Random Forest (RF) model reached approximately 99% accuracy. According to our results, in addition to different wheat varieties, also sibling hybrid seeds can be classified with high accuracy according to their color characteristics by the methods we proposed.

1. Introduction

The temperature of our world is increasing day by day. From 2000 to 2020, the world temperature increased by about 1 °C. [1]. This increase prolongs the residence time of water in the atmosphere. This atmospheric phenomenon leads to two natural disasters. The first is the regional droughts, which are caused by the decrease in the number of rainy days in the year. The other is flood disasters that occur when the water accumulated in the atmosphere suddenly descends. [2], [3]. These two natural disasters reduce agricultural production [4]. Another factor affecting agricultural production is the Covid-19 pandemic, which still continues today. The rules applied by the countries for precautionary purposes lead to the lack of labor and input supply required for food production. This causes inadequacy of agricultural production and deterioration of the distribution network and increases the need for food. [5]–[7]. On the other hand, food demands are increasing due to the increasing world population.The number of people living on Earth is approximately 7.79 billion. 8.9% of this population,

means 690 million, are people who are starving. In order to end this situation, it is aimed to end hunger by 2030 and worldwide studies are carried out for this.[8]. Increasing food production and ensuring food security are of vital importance for humanity. Therefore, new strategies are being developed.

At the beginning of these strategies is to increase the production efficiency of cereals that meet a large part of the needed nutrients. It is aimed to increase the production and crop yield of cereals, which are cheaper and more satisfying than other food products. Wheat is one of the most cultivated cereals and resistant to environmental stresses and has a rich biodiversity. It contains carbohydrates, essential amino acids, vitamins, beneficial phytochemicals and fibrous components in its structure. [9], [10]. It meets 20% of our daily energy needs [11]. Considering the price and nutritional value of prune, it is an important grain that can be used to meet the increasing food demand. Therefore, it has a large market around the world. Global wheat production was 760.9 million tons on an area of approximately 219 million

*Corresponding Author: Mesut Ersin Sonmez mesutersinsonmez@kmu.edu.tr

hectares in 2020. In the world, 43.9% of wheat production is produced in Asia, 33.1% in Europe, 16.3% in America, 3.3% in Africa and 3.3% in Oceania. The countries that produce the most wheat in the world have the largest wheat market with China (114.58 million tons), India (84.29 million tons) and Russia (56.18 million tons). [12]. Wheat market has economic value for many countries in the world. The increase in the yield of the wheat crop means that the income is obtained from market increases, too. The suitability of the varieties planned to be produced for the growing environment increases the plant yield. In this way, more production and profit are achieved. On the other hand, it is important to choose the right rootstocks for new varieties to be developed in breeding studies. By crossing the right cultivars, new high yielding cultivars can be developed. In general, molecular scans are performed to determine the varieties. It is detected by methods such as Gel electrophoresis or Near Infrared Reflectance (NIR) spectroscopy. [13], [14]. These methods require expensive instrumentation and long analysis times. On the other hand, the analyzed sample is not reused.

With the developments in image processing techniques and artificial intelligence in recent years, it offers the opportunity to analyze quickly and free from human-induced errors without damaging the product. [15]–[17]. These techniques, which have been widely used in wheat classification studies, aim to increase crop yield on a country basis. Ronge and Sardeshmukh [18] made a study to classify the wheats according to texture of grain. Wheat images of four different Indian local varieties were obtained and extracted 131 texture features with five different methods. For the classification process, k-NN and ANN algorithms are used. Their results showed that classification accuracy reached 100% success via ANN architecture. In another study, the width, height, area, perimeter, kernel asymmetry and diameter information of 210 wheat grains were extracted using image processing techniques. These data were processed in WEKA program using Neural Network (NN), Naive Bayes (NB), J48 and multilayer perceptron (MLP) machine learning algorithms. The highest classification success achieved was obtained with an MLP with 97.17% success. [19]. In a similar study, the physical properties, which were extracted via image processing, were optimized with the artificial bee colony (ABC) algorithm and used in the classification process to detect the bread or durum wheat with an artificial neural network. Wheats were classified with 100% success with optimized network architecture. [20]. In another study in which wheat was classified according to its physical properties, The classification process was carried out according to the similarity of the wheat grains to the known geometric shapes. In this study, in which Emmer (tetraploid) and Einkorn (diploid) varieties were used as well as modern varieties. It was observed that wheat with an aspect ratio equals to 1.6 was bread wheat, and wheat with a 2.4 aspect ratio was durum wheat. [21]. The successes from deep learning studies that have been popularly used in the past few years are also high. Laabassi et al [22] classified simeto, Vitron, ARZ, and HD varieties from Algerian local wheat by using deep learning networks. DensNet201, MobilNet and inception V3, are called the state of art, were used in the classification made with the deep learning architectures of wheat grains. According to the results, the best classification was performed with the DensNet201 network with 95.68% success.

In this study, unlike other studies, Ahmetbugdayi and Cesare and also BC1F6, BC2F5 hybrids, obtained from Ahmetbugdayi x Cesare, were classified only according to their color characteristics. For this, images of the samples were taken in the imaging cabinet we developed. With image processing techniques, data consisting of 12 color channels belonging to four different color spaces were obtained. Squares were drawn in the center of each wheat grain in the images in these channels, and the average of the pixels in these squares was accepted as the color of that grain for machine learning. By using the created data set, high classification success was obtained with ANN, SVM, k-NN, DT, RF and Naive Bayes algorithms. It has been shown that using this method, wheat classification can be done without the need for fast and expensive analyzes and equipment. It can also be used as an alternative method for color analysis in wheat.

2. Material Method

2.1. Plant material

In this study, Ahmetbugdayi and Cesare durum wheat varieties used as rootstocks were grown in Karamanoglu Mehmetbey University engineering faculty speed breeding greenhouse. (location: 37°10'35.8"N, 33°15'09.7"E). Two different hybrids were obtained from these rootstocks by using the back-cross method. F1 hybrid was obtained by crossing Ahmetbuğdayi and Cesare. Then, BC1F6 was obtained by backcrossing F1 and Ahmetbugdayi one time and self- inbreeding of the hybrid six times. In the same way, BC2F5 was obtained by backcrossing F1 and Ahmetbugdayi and self- inbreeding of the hybrid five times. The hybrids were bred in the research land in Karaman. (location: 37°14'40"N, 33°20'08"E). Breeding studies BC1F6 and BC2F5 hybrids were sown in September 2019 and harvested in July 2020.

2.2. Collecting images

A custom imaging system is designed to get images (Fig.1). The box is produced with a 3d printer to prevent environmental light noise and to take photos at constant light intensity. The box can move up and down. Light intensity and movement of the box are controlled by the Atmega328p chip. Samsung 4000 K Strip LED is used for lighting. Basler MP acA2500-14uc / um CMOS color camera and Basler C125-0618-5M F1.8 f6 mm are located on the box. The camera is fixed at a distance of approximately 15 cm from the viewing area. LattePanda minicomputer card (Intel Cherry Trail Z8300 Quad Core 1.8GHz CPU, 4GB DDR3L RAM, Intel HD Graphics, 12 EUs @200-500 MHz GPU) is used to get and collect the images. Specially designed black stands were used by using double-sided tape that prevents the wheat from slipping on it.

From each wheat sample, 70 grains were randomly selected for each image. For preparing the image data five different images of each Ahmetbugdayi, Cesare, BC1F6 and BC2F5 were used. A total of 1400 wheat grains are included in the images. All the images were gotten at the same light intensity. Pylon Viewer 6.2.0.8205 program was used to obtain images and saved in Tagged Image File Format (TIFF) format.

2.3. Image processing

Images were processed by using the python (version 3.8) programming language and the OpenCV-4.1.3 library. First, RGB images were preprocessed in order to extract the color features of each wheat grain. For this, the images were converted to gray scale images. Then the Otsu threshold was applied. Salt-pepper noises caused by Otsu thresholding were eliminated by dilution with a 3x3 kernel size. The position of each wheat grain was determined using the contour finder library in OpenCV. Then, frames with a size of 25 x 25 pixels were drawn in the center of each grain (Fig.2). After determining the position of the square, images were converted to RGB, HSV, Lab, YCrCb color spaces. Afterwards, the determined positions of the squares were drawn at the same position of all four color spaces applied images.

a.   b.
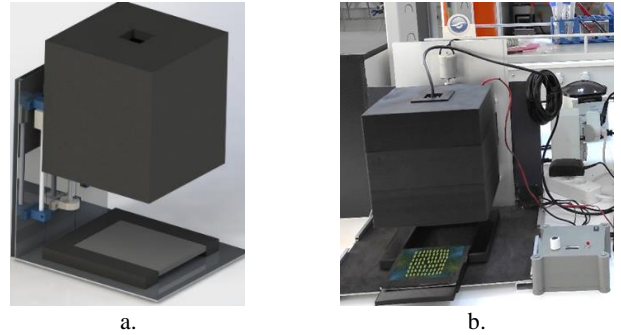
Fig.1 Imaging system, a. 3D designed model, b. Manufactured imaging system
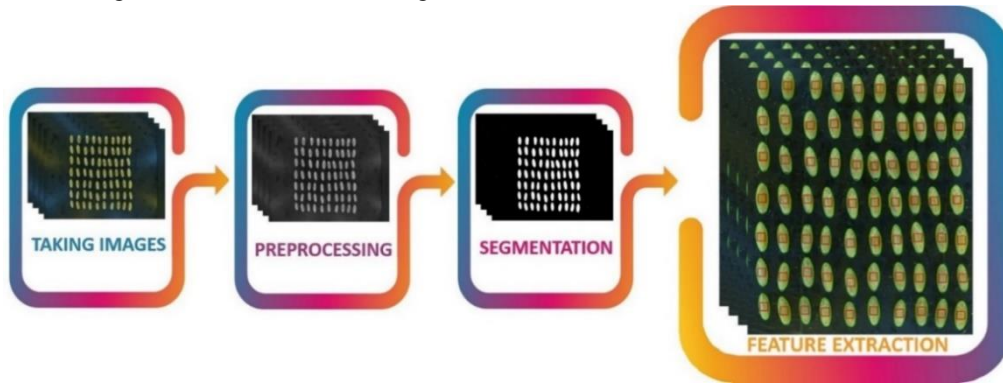
Fig.2 The process of finding the color frames of each wheat grains
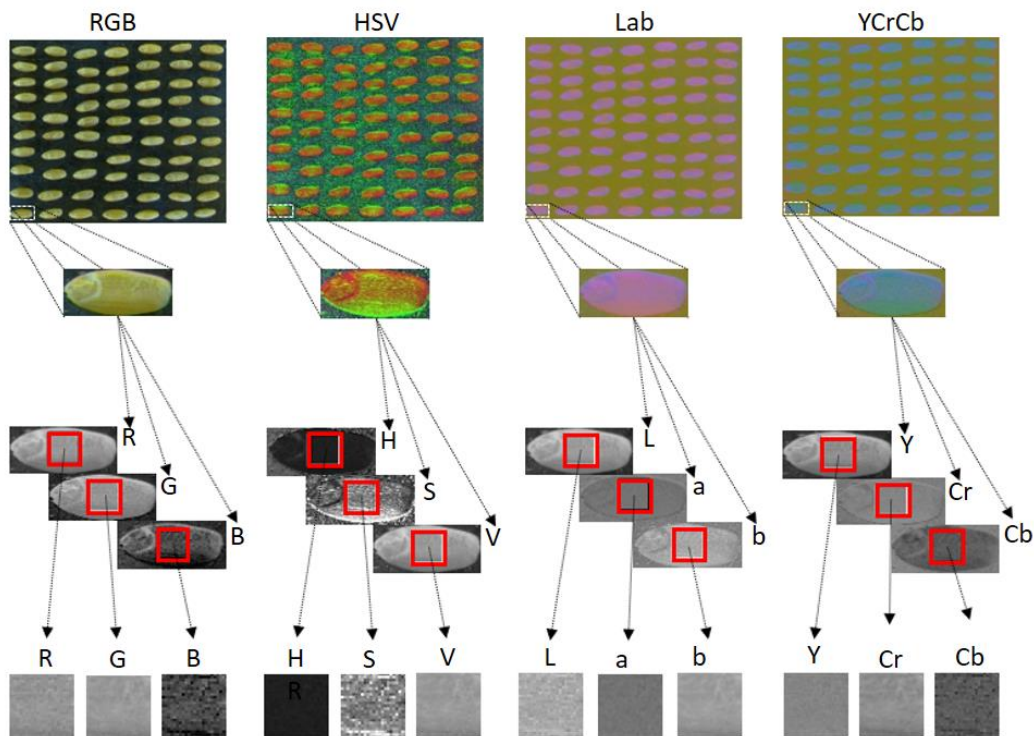
Fig.3 Extracting color features

Every pixel of inside the taken pieces were listed then the average of them was calculated. So, twelve color channels values were calculated one by one for 1400 grains in the images.

2.4. Dataset

The dataset consists of Ahmetbuğdayi, Cesare, BC1F6 and BC2F5 wheat grains. Each variety and hybrids comprise 350 grains. According to four different color spaces, totally 1400 grain samples have twelve color channel features. out of the total data, 840 data (60%) was randomly selected for training and the remaining 560 were data used as test data. We standardized all the values in the datasets by the z-score method before developing machine learning algorithms.

2.5. Classification Method

Confusion matrices were used to determine the performance of the machine learning algorithms [23]. To obtain the confusion matrices values the formulas are shown in Eqs 1-5 [24], [25]were employed with the python SciKit-learn library.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} x100 \qquad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (2)$$

$$\text{F1 Score} = \frac{2TP}{2TP + FP + FN} \qquad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (5)$$

Additionally, receiver operating characteristic (ROC) score, is one of the probabilistic forecasting performance measurement method was determined [26]. The roc curve has a false positive rate on the x-axis and a true positive rate on the Y-axis. The intersection of these two ratios gives the ROC value. It provides a lot of convenience in evaluating the performance of the machine learning algorithm, especially in an analysis with unbalanced data sets. The ROC score is defined as the area under the curve that ranges from 0 to 1. In general, prediction models with higher scores are considered more skillful.

2.6. Machine Learning Algorithms

Six machine learning algorithms were employed in this study. These Algorithms are ANN, SVM, k-NN, DT, RF and NB (Fig.4). The algorithms used were compiled using the SciKit-learn library. Confusion matrices and ROC graph were drawn for all the algorithms. All the results were saved to excel file during the running program. Also, Confusion matrices and ROC graph were saved as jpeg file.

2.7. Artificial Neural Network (ANN)

Artificial neural networks mimic the biological neural structure of the human brain. ANN can derive or classify new information using previously learned or classified information. ANN consists of input layer, hidden layers that include computational neurons, and an output layer [27]–[29]. In this study we used multi-layer perceptron (MLP). Our ANN architecture included two hidden layers. In our code, we used nested for loops. In every for loop the accuracy of the network was determined and listed. For each index of list was checked with if statements to find neurons number that had the higher accuracy result.
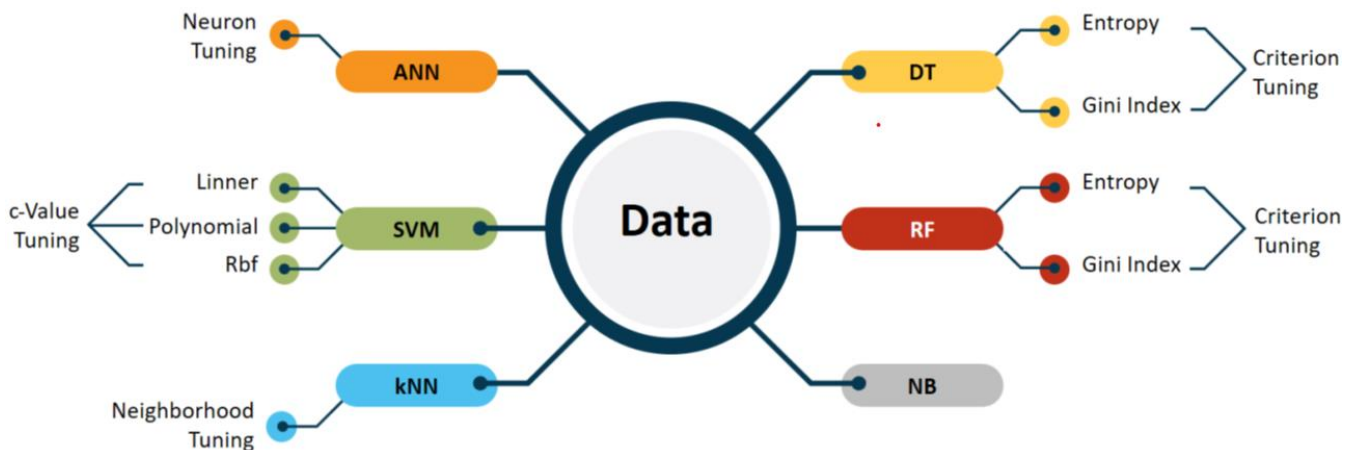
Fig.4 Machine learning algorithms used in the study

The neurons with the highest accuracy was determined and printed. The network initiated with 10 neurons for the first and second hidden layer and the neuron numbers decreased until 1 for each hidden layers. So, the best performed neurons number in the hidden layer were determined by the program automatically. ReLU activation function and 200 epochs were used. The architecture of the ANN included two hidden layers, twelve inputs (color channels) and four outputs (wheat varieties and hybrids).

2.8. Support Vector Machine (SVM)

Support vector machine is one of the non-parametric and supervised learning algorithms. SVM is characterized by using the kernel and acting on the margins [30]. The most important advantage of SVM needs low processing time. Kernel functions and c values are the hyper-parameters that need tuning. In this study, we used Linear, Polynomial and Rbf kernel function. For all these kernels, c values were tuned automatically according to accuracy. We used for loop to test the c values. All the accuracy results were listed by the append function. The index had the highest accuracy gave the best c value.

2.9. k-Nearest Neighborhood (k-NN)

K nearest neighborhood is an instance based and supervised learning algorithm. The related classes is placed on the coordinate plane. Test data was placed later on the coordinate plane. Test data is classified by looking the nearest neighbors [31]. In order to run k-NN algorithm k value is needed by the user. In this study, for the best k-NN algorithm performance, k value was determined automatically. Again, we used for loops and list method, the maximum k values were taken from user. In for loop all the k values are run and the accuracies were listed. According to the accuracy list index had the highest accuracy was determined the fine-tuned k value. Minkowski distance was used for metric parameter.

2.10. Decision Tree (DT)

Decision tree algorithm consists of roots, nodes and several leaves. Nodes makes binary decision to determine the categories. This algorithm is a non-parametric can affectively use for classification and regression studies [32]. Here, we test Gini index and Entropy criterion, Eqs (6,7) were used to calculate, to find the optimum spilt of roots to reach the highest performance.

$$Gini\ Index = \quad 1 - \sum_j p_j^2 \qquad (6)$$

$$Entropy = \quad 1 - \sum_j p_j^2 \cdot log_2 \cdot p_j \qquad (7)$$

2.11. Random Forest (RF)

Random forest is an algorithm based on classification and regression tree, consisting of many decision trees. Each trees are built using a different set of randomly selected input data. This trees are voted to label the most likely class. The final label is determined by the majority voting [33], [34]. In this study we used Gini and Entropy criterion to find the successful model.

2.12. Naïve Bayes (NB)

Naïve Bayes algorithm is a powerful probabilistic classification algorithm usually employed while working with huge data set. Naïve Bayes algorithm is based on Bayes theorem. This algorithm is work to find the probability of occurrence of each category under the condition of this occurrence. Classification is done according to the highest probability [35], [36]. We used to see the wheat sample classification success of NB algorithm and compare the other machine learning algorithm.

3. Results and Discussions

In this study Ahmetbugdayi, Cesare, BC1F6 and BC2F5 varieties and their hybrids wheat grains were used for classification according to color features. For the first step, image processing techniques were used to convert the four different color space and separate all the color channels to make a data set for machine learning algorithms. After separating color channels, 25x25 square portions of the image, as named ROI, were extracted from all the grain centers. All the pixel values in the ROIs were listed and the average of them were determined. As a result, each grain color channels had become the single averaged value. All data were standardized and 60% of the data was reserved as train and the rest as test data.

We employed six different machine learning algorithms to reach the highest classification accuracy. In addition, we chance the hyper-parameters of some algorithms to increase the classification success. Five algorithms gave the accuracy results approximately 99% (Table 1). The highest accuracies were obtained more than 99% from both ANN, SVM and DT classifier models. The lowest accuracy determined with NB algorithm is 96%.

For the ANN algorithm, we used for loop and if statements to find the neuron numbers, had the highest accuracy, instead of randomly choosing [37], [38]. The highest neurons numbers were taken from the user for all the hidden layers. Then the neurons numbers were decreased one

by one for each epoch. We decreased the neurons because we observed that the same accuracy can obtain with different neurons number of the hidden layers. Therefore, the architecture of the ANN fine-tuned with the least neurons. The program started with 10 neurons for each hidden layer.

Table 1 Machine learning algorithm performance results.

Algorithms	Accuracy(%)	Precision	F1 Score	Recall	Specificity	AUC
ANN	**99.29**	**0.9929**	**0.9929**	**0.9929**	**0.9353**	**0.9971**
SVM (linear kernel, c=3)	**99.11**	**0.9911**	**0.9911**	**0.9911**	**0.9347**	**0.9984**
SVM (polynomial kernel, c=3)	98.93	0.9828	0.9821	0.9821	0.9330	0.9976
SVM(Rbf kernel, c=4)	98.93	0.9859	0.9857	0.9857	0.9332	0.9998
k-NN (k=5)	98.57	0.9849	0.9857	0.9857	0.9330	0.9882
DT (Entropy)	**99.11**	**0.9912**	**0.9911**	**0.9911**	**0.9349**	**0.9843**
DT (Gini index)	98.57	0.9860	0.9857	0.9857	0.9349	0.9803
RF (Entropy)	98.75	0.9877	0.9875	0.9875	0.9337	0.9960
RF (Gini index)	98.75	0.9877	0.9875	0.9875	0.9337	0.9960
NB	95.71	0.9602	0.9568	0.9571	0.9247	0.9924
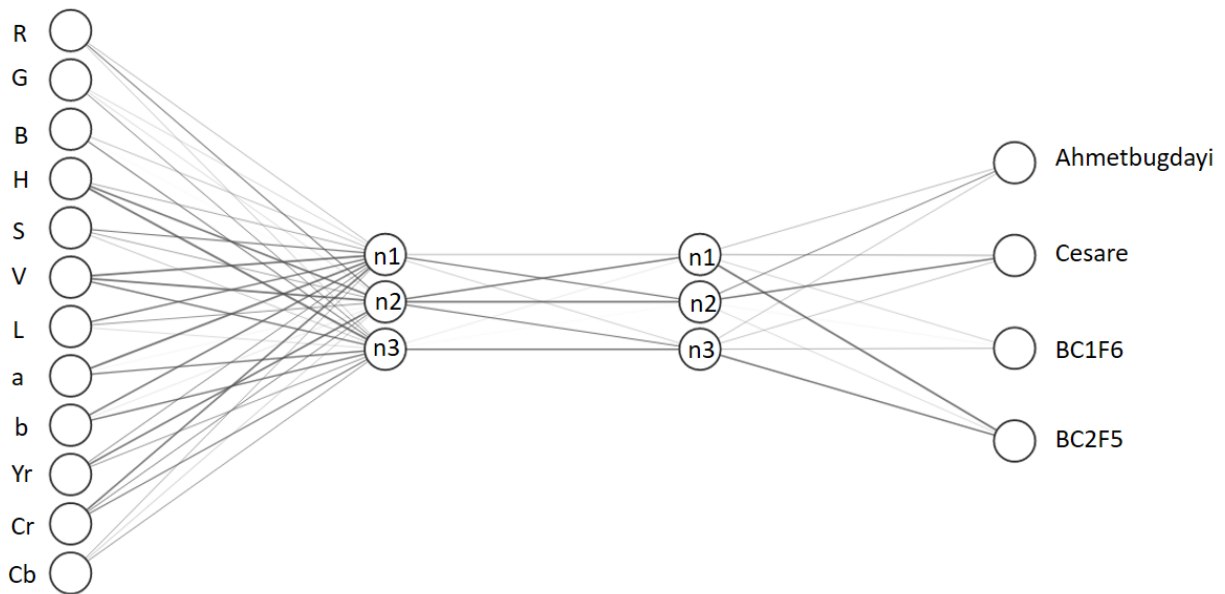
Fig.5 ANN architecture

The highest accuracy results were 99.29%, obtained by the network consisting of 3 neurons in the first and second hidden layers. Achieved the best architecture of the network was demonstrated in Fig.5. In the SVM model, Linear, Polynomial and Rbf kernel functions were used and their accuracy results were compared. For loop used to determine the optimum c values for all SVM models [39]. C value increased by 0.1 per epoch. The for loop was started with c value equal to 0.1 and continued until 10, which was set by the user.

As shown in Table 1, we obtained 99.11%, the highest accuracy in SVM models, with the model tuned with linear kernel and c value equal to 3. The second model tuned with polynomial kernel and c=3 had 98.93% accuracy as the same as third model, had Rbf kernel and c=4. As seen Table 1. All SVM models gave very closed accuracies.

In DT model, by using Gini index and Entropy criteria for node distribution optimization, we determined each contribution to accuracy. As seen in Table 1, the model that includes Gini index had a little bit lower accuracy than the other model. The obtained accuracy with model fine-tuned with Entropy was determined 99.11% and for the model with Gini index was 98.57%.

Confusion matrices were shown in Fig.6, represent the classification of 560 test grains for the most successful

three algorithms.   Just four grains placed wrong classes for the ANN model and 5 grains were placed wrongly for SVM and DT.
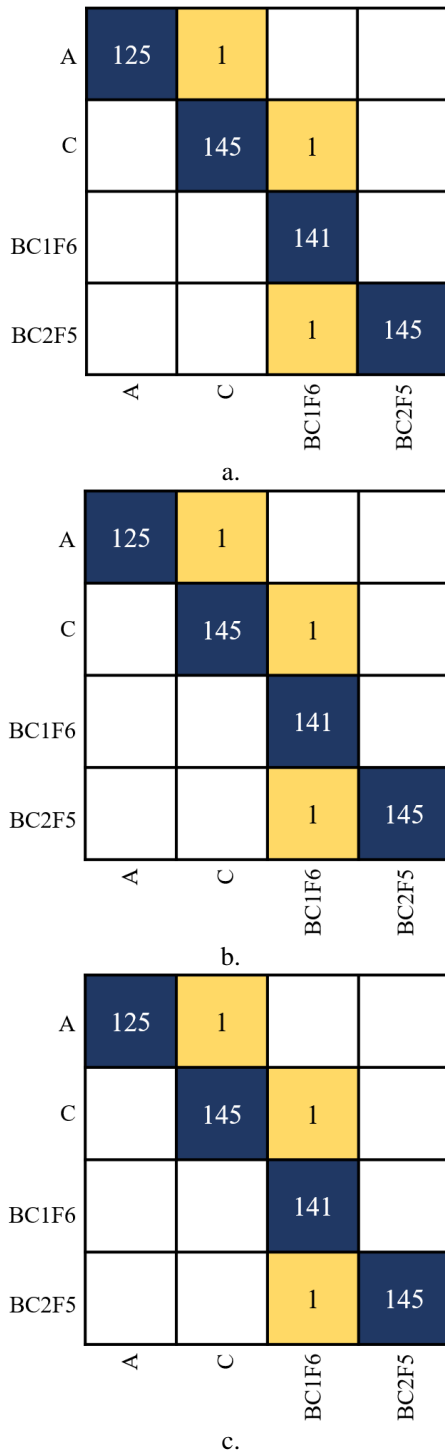

a.

b.

c.

Fig.6 Confusion matrices. a. ANN, b. SVM, c. DT

In Fig.7, we showed the overlaps of tests and predict results for each class.

ROC is another way to present the classifier model's performance. ROC graph gives details on true positive rate and false positive rates on Y and X axis.
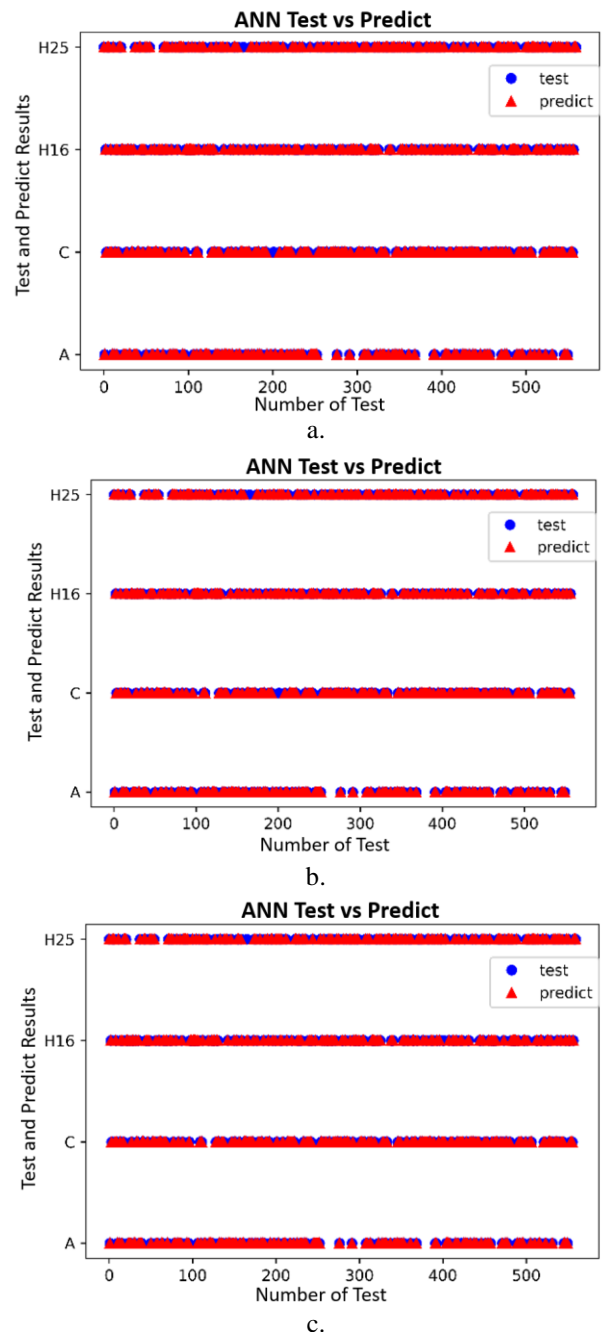

a.

b.

c.

Fig.7 Test – Predict blasted results. a. ANN, b. SVM, c. DT

Point (0,1), means false positive rate is 0 and true positive rate is 1, specifies that the classifier reaches the best classification success and point (1,0) classification of all the data were placed wrong classes. Misclassified data in the ROC graph is considerable specifying the dip in the curve [40].ROC curves of three classifier models are shown in the Fig.8. These curves demonstrate that three classifier model performance reached the almost perfect classification.

a.

b.

c.
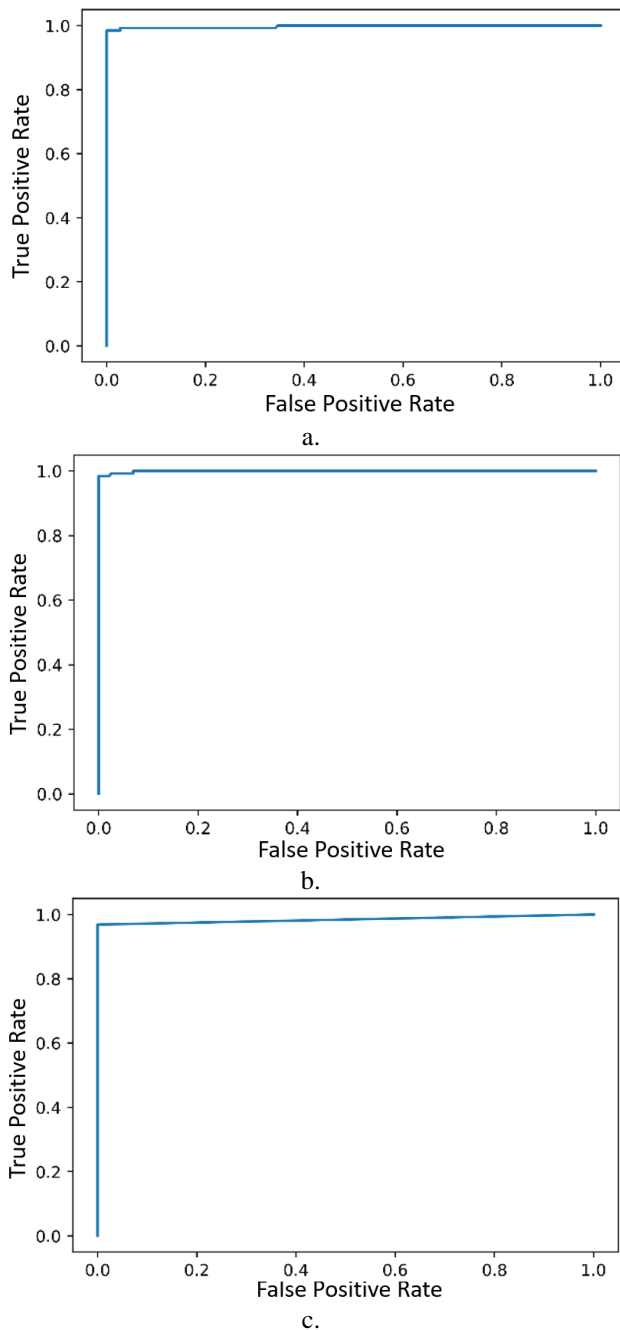
Fig.8 ROC curves. a. ANN, b. SVM, c. DT

The areas under the curve (AUCs) received from the ROC analysis for test showed the highest value was 0.9984 for SVM, the other algorithm had 0.9971 (ANN) and 0.9843 (DT). The AUC results imply that, three algorithms have very good classification capacity for color feature included data. For three algorithms precision, F1 score and recall were calculated approximately 0.99, specificity values were determined about 0.945 for all the algorithms shown in Table 1.

For the k-NN algorithm For loop is again used to find the best number of neighborhoods instead of randomly choosing [41]. The highest accuracy (98.57%) was achieved when k equaled 5. Precision, f1 score and recall were determined roughly 0.986 and specificity of k-NN was calculated 0.933. AUC was computed 0.998

It was used the criterions by one by to find the contribution of Gini index and Entropy to classification accuracy in RF algorithm. But the results of accuracy, precision, F1 score and recall were obtained almost same (0.988) that shown in the Table 1. Also, specificity and AUC of them were determined 0.934, 0.996 respectively. The least accuracy was obtained with NB algorithm that accuracy was approximately 96.0% and precision, F1 score, recall was determined roughly 0.96, specificity and AUC was calculated 0.93, 0.99 respectively.

4. Conclusions

Wheat classification is a very important process for producers and breeders. While breeding species suitable for environmental conditions means more profit for producers and choosing the right parents for breeders is one of the most important criteria for selecting desired traits in future generations. On the other hand, physical appearance of wheat such as color and grain size are also important reasons for preference for producers. Particularly for durum wheat, grain color is one of the important quality criteria. Conventional methods require expensive equipment and samples used during analysis cannot be reused. In this study, we presented a different analysis method for wheat classification, including image processing techniques and machine learning algorithms. In this way, product loss that occurred during the analysis was prevented. For different color spaces were used. Color channels of these spaces were obtained by image processing techniques. The color feature samples, 25x25 pixel squares (ROI), were collected from the center of the grain, instead of taking a single pixel. All the pixels in ROI were listed. The average of the list gave the color channel value for the grain. This method was applied to 1400 wheat grains. 12 different color channel features were determined and used as input data for the classification algorithms. Six machine learning algorithms were employed to classify the wheat grains. Some of the hyperparameters of the algorithms were determined by the program, was coded in python. ANN, SVM and DT algorithms achieved more than 99% accuracies. In addition, k-NN and RF algorithms had almost 99% accuracy. NB algorithm also had higher accuracy, was approximately 96%. This study suggests that wheat grains can be classified using machine learning algorithms according to the color channels of different color spaces. In addition, Determining the color values of the pixels by drawing a square in the center of the grain can be used in the selection of wheat varieties with the desired color characteristics.

Author contributions

The authors declare that they have contributed equally to the article.

Funding

The authors have no relevant financial or non-financial interests to disclose.

Conflict of interest

The article authors declare that there is no conflict of interest between them.

Ethical approval

Ethical approval is not required for this research.

References

[1] FAO, "FAOSTAT," 2020. https://www.fao.org/faostat/en/#data/QCL (accessed Dec. 05, 2021).

[2] M. K. van Aalst, "The impacts of climate change on the risk of natural disasters," *Disasters*, vol. 30, no. 1, pp. 5–18, Mar. 2006, doi: 10.1111/J.1467-9523.2006.00303.X/FORMAT/PDF.

[3] N. Arunrat, S. Sereenonchai, W. Chaowiwat, and C. Wang, "Climate change impact on major crop yield and water footprint under CMIP6 climate projections in repeated drought and flood areas in Thailand," *Sci. Total Environ.*, vol. 807, Feb. 2022, doi: 10.1016/J.SCITOTENV.2021.150741.

[4] S. Mehryar and S. Surminski, "National laws for enhancing flood resilience in the context of climate change: potential and shortcomings," *Clim. Policy*, vol. 21, no. 2, pp. 133–151, 2021, doi: 10.1080/14693062.2020.1808439/SUPPL_FILE/TCPO_A_1808439_SM1165.ZIP.

[5] S. Zhang, S. Wang, L. Yuan, and X. Liu, "The impact of epidemics on agricultural production and forecast of COVID-19," 2020, doi: 10.1108/CAER-04-2020-0055.

[6] D. Bochtis, L. Benos, M. Lampridi, V. Marinoudi, S. Pearson, and C. G. Sørensen, "Agricultural Workforce Crisis in Light of the COVID-19 Pandemic," *Sustain. 2020, Vol. 12, Page 8212*, vol. 12, no. 19, p. 8212, Oct. 2020, doi: 10.3390/SU12198212.

[7] P. Bahadur Poudel *et al.*, "COVID-19 and its Global Impact on Food and Agriculture," *J Biol Today's World*, vol. 9, no. 5, p. 221, 2020.

[8] WHO, "The state of food security and nutrition in the world 2020: transforming food systems for affordable healthy diets," 2020.

[9] A. I. Olagunju, T. D. Oluwajuyitan, and S. I. Oyeleye, "Multigrain bread: dough rheology, quality characteristics, in vitro antioxidant and antidiabetic properties," *J. Food Meas. Charact.*, vol. 15, no. 2, pp. 1851–1864, 2021, doi: 10.1007/s11694-020-00670-3.

[10] B. Svihus, *Nutritive and Digestive Effects of Starch and Fiber in Whole Wheat.* Elsevier, 2014. doi: 10.1016/B978-0-12-401716-0.00007-6.

[11] P. R. Shewry and S. J. Hey, "The contribution of wheat to human diet and health," *Food Energy Secur.*, vol. 4, no. 3, pp. 178–202, 2015, doi: 10.1002/FES3.64.

[12] FAO, "FAOSTAT," 2020. https://www.fao.org/faostat/en/#data/QCL/visualize (accessed Dec. 29, 2021).

[13] C. Miralbés, "Discrimination of European wheat varieties using near infrared reflectance spectroscopy," *Food Chem.*, vol. 106, no. 1, pp. 386–389, Jan. 2008, doi: 10.1016/J.FOODCHEM.2007.05.090.

[14] R. Tkachuk and V. J. Metlish, "Wheat cultivar identification by high voltage gel electrophoresis.," *Ann. Technol. Agric.*, vol. 29, no. 2, pp. 207–212, 1980.

[15] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," *Sensors 2018, Vol. 18, Page 2674*, vol. 18, no. 8, p. 2674, Aug. 2018, doi: 10.3390/S18082674.

[16] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artif.*

[17] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Comput. Electron. Agric.*, vol. 153, pp. 69–81, Oct. 2018, doi: 10.1016/J.COMPAG.2018.08.001.

[18] R. V. Ronge and M. M. Sardeshmukh, "Comparative analysis of Indian wheat seed classification," *Proc. 2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2014*, pp. 937–942, Nov. 2014, doi: 10.1109/ICACCI.2014.6968483.

[19] K. Sabanci and M. Akkaya, "International Journal of Intelligent Systems and Applications in Engineering Classification of Different Wheat Varieties by Using Data Mining Algorithms," *IJISAE*, vol. 4, no. 2, 2016, doi: 10.18201/ijisae.62843.

[20] A. Kayabasi, "An Application of ANN Trained by ABC Algorithm for Classification of Wheat Grains," *Int. J. Intell. Syst. Appl. Eng.*, vol. 6, no. 1, pp. 85–91, Mar. 2018, doi: 10.18201/IJISAE.2018637936.

[21] J. J. Martín-Gómez, A. Rewicz, K. Goriewa-Duba, M. Wiwart, Á. Tocino, and E. Cervantes, "Morphological Description and Classification of Wheat Kernels Based on Geometric Models," *Agron. 2019, Vol. 9, Page 399*, vol. 9, no. 7, p. 399, Jul. 2019, doi: 10.3390/AGRONOMY9070399.

[22] K. Laabassi, M. A. Belarbi, S. Mahmoudi, S. A. Mahmoudi, and K. Ferhat, "Wheat varieties identification based on a deep learning approach," *J. Saudi Soc. Agric. Sci.*, vol. 20, no. 5, pp. 281–289, Jul. 2021, doi: 10.1016/J.JSSAS.2021.02.008.

[23] M. E. Sonmez, N. Eczacıoglu, N. E. Gumuş, M. F. Aslan, K. Sabanci, and B. Aşikkutlu, "Convolutional neural network - Support vector machine based approach for classification of cyanobacteria and chlorophyta microalgae groups," *Algal Res.*, vol. 61, no. November 2021, 2022, doi: 10.1016/j.algal.2021.102568.

[24] S. Kanaan, S. Al-Doori, Y. Selim Taspinar, and M. Koklu, "Distracted Driving Detection with Machine Learning Methods by CNN Based Feature Extraction," *Int. J. Appl. Math. Electron. Comput.*, vol. 9, no. 4, pp. 116–121, Dec. 2021, doi: 10.18100/IJAMEC.1035749.

[25] Y. S. Taspinar, M. Koklu, and M. Altin, "Fire Detection in Images Using Framework Based on Image Processing, Motion Detection and Convolutional Neural Network," *Int. J. Intell. Syst. Appl. Eng.*, vol. 9, no. 4, pp. 171–177, Nov. 2021, doi: 10.18201/ijisae.2021473636.

[26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/J.PATREC.2005.10.010.

[27] K. S. Narendra and K. Parthasarathy, "Neural Networks In Dynamical Systems," in *Intelligent Control and Adaptive Systems*, 1990, vol. 1196, p. 230. doi: 10.1117/12.969922.

[28] S. Haykin, J. Nie, and B. Currie, "Neural network-based receiver for wireless communications," *Electron. Lett.*, vol. 35, no. 3, pp. 203–205, 1999, doi: 10.1049/el:19990177.

[29] K. Sabanci *et al.*, "Wheat flour milling yield estimation based on wheat kernel physical properties using artificial neural networks," *Int. J. Intell. Syst. Appl. Eng.*, vol. 8, no. 2, pp. 78–83, 2020, doi: 10.18201/ijisae.2020261588.

[30] S. R. Gunn, "Support Vector Machines for Classification and Regression," 1998.

[31] D. W. Aha, D. Kibler, M. K. Albert, and J. R. Quinian, "Instance-based learning algorithms," *Mach. Learn. 1991 61*, vol. 6, no. 1, pp. 37–66, Jan. 1991, doi: 10.1007/BF00153759.

[32] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Arch. Psychiatry*, vol. 27, no. 2, p. 130, Apr. 2015, doi: 10.11919/J.ISSN.1002-0829.215044.

[33] X. Mei *et al.*, "A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat," *Remote Sens. 2019, Vol. 11, Page 920*, vol. 11, no. 8, p. 920, Apr. 2019, doi: 10.3390/RS11080920.

[34] L. Guo, N. Chehata, C. Mallet, and S. Boukir, "Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 1, pp. 56–66, Jan. 2011, doi: 10.1016/J.ISPRSJPRS.2010.08.007.

[35] D. L. Naik and R. Kiran, "Naïve Bayes classifier, multivariate linear regression and experimental testing for classification and characterization of wheat straw based on mechanical properties," *Ind. Crops Prod.*, vol. 112, pp. 434–448, Feb. 2018, doi: 10.1016/J.INDCROP.2017.12.034.

[36] D. Zhang *et al.*, "Integration of spectroscopy and image for identifying fusarium damage in wheat kernels," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 236, p. 118344, 2020, doi: 10.1016/j.saa.2020.118344.

[37] K. Sabanci, A. Kayabasi, and A. Toktas, "Computer vision-based method for classification of wheat grains using artificial neural network," *J. Sci. Food Agric.*, vol. 97, no. 8, pp. 2588–2593, Jun. 2017, doi: 10.1002/JSFA.8080.

[38] Y. Cakir, M. Kirci, and E. O. Gunes, "Yield prediction of wheat in south-east region of Turkey by using artificial neural networks," *2014 3rd Int. Conf. Agro-Geoinformatics, Agro-Geoinformatics 2014*, 2014, doi: 10.1109/Agro-Geoinformatics.2014.6910609.

[39] L. Ravikanth, C. B. Singh, D. S. Jayas, and N. D. G. White, "Classification of contaminants from wheat using near-infrared hyperspectral imaging," *Biosyst. Eng.*, vol. 135, pp. 73–86, Jul. 2015, doi: 10.1016/J.BIOSYSTEMSENG.2015.04.007.

[40] C. E. Metz, "Basic principles of ROC analysis," *Semin. Nucl. Med.*, vol. 8, no. 4, pp. 283–298, Oct. 1978, doi: 10.1016/S0001-2998(78)80014-2.

[41] Z. Cataltepe, E. Cetin, and T. Pearson, "Identification of insect damaged wheat kernels using transmittance images," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2, pp. 2917–2920, 2004, doi: 10.1109/ICIP.2004.1421723.