Journal of Scientific & Industrial Research
Vol. 80, November 2021, pp. 985-991

Classifying Wheat Genotypes using Machine Learning Models for Single Kernel Characterization System Measurements

Imren Kutlu*, Mehmet Bilginer Gulmezoglu and Yasar Karaduman
Eskişehir Osmangazi University, 26480, Eskişehir, Turkey

Received 16 September 2020; revised 06 May 2021; accepted 22 October 2021

The properties related to market value, milling, classification, storage, and transportation of bread wheat are determined by using some important physical quality characteristics such as weight, shape, dimensions, and hardness of wheat kernels. It is possible to measure all these features using single kernel characterization system (SKCS). Classification of wheat genotypes using computer-based algorithms is crucial to determine the most accurate physical quality classification for breeding studies. In this paper, four commercial wheat cultivars (Altay-2000, Bezostaja-1, Harmankaya-99, and Kate A-1) and six doubled haploid (DH) wheat genotypes are studied to classify wheat cultivars and DH wheat genotypes separately. In the classification stage, feature vectors constructed from measured characters namely, kernel weight, diameter, hardness, and moisture are applied to well-known classifiers such as Common Vector Approach (CVA), Support Vector Machines (SVM) and K-Nearest Neighbor (KNN). Satisfactory results especially for the training set are obtained from the experimental studies. Classification results are compared with single linkage hierarchical cluster (SLHC) analysis, which is the most widely used in breeding studies. Recognition of clustered genotypes in all three classification methods and dendrograms present similar results. The SVM model is found to be outperformed over other methods for studied characters and could therefore effectively be utilized for characterizing, classifying and/or identifying the wheat genotypes.

Keywords: Common vector approach, Kernel hardness, KNN method, SVM classification, Wheat quality

Introduction

Wheat is the main substance of most widespread food products in daily lives and acquisition quality of wheat grains is a considerable issue for improving food supplies. Wheat's physical quality characteristics including grain weight, grain density, kernel sizing, and hardness index are important in terms of commercial classification. Wheat grain size and weight are significant quality components for valuation of milling characters. Hardness, which has important relations with the end-use traits of wheat, is a major index for discrimination of wheat classes. Wheat grains with optimum hardness are preferred to milling flour considering wheat class.[1] There are numerous methods to measure wheat grain hardness. Two grain hardness measurements approved by American Association of Cereal Chemists (AACC) are the particle size index (PSI; 55-30) and the Single Kernel Characterization System 4100 (SKCS; 55-31). The SKCS 4100 is a popular measurement technique because it uses a small number of sample for single kernel hardness measurement, rapid and reliable. The SKCS 4100 has been verified as a beneficial device in the evaluation of wheat grain hardness based on the benefit of the estimation of hardness index that it generates.[2] Mis *et al.*[1] reported that the SKCS parameters can be used to develop evaluation of wheat quality for flour milling and breadmaking besides wheat classification. There is a little complexity in classifying wheat cultivars according to the hardness index due to the particular genetic difference between hard and soft wheats. Thus, characterization of the varieties in order to distinguish each other is rather important.

In recent years, interest in correct classification of wheat varieties has greatly increased because misclasification may result in choosing of nonsuitable wheats in bread and pasta production. Certain approaches namely morphological[3], biochemical[4], molecular[5] are being used. Machine learning is a trending technology nowadays and it can be used in modern agriculture to improve the productivity and quality of the crops.

New progresses in information technology and in digital image processing led to advances and practical use of computer-aided techniques in data analysis. Many researchers, through the last years, have been

*Author for Correspondence
E-mail: ikutlu@ogu.edu.tr

endeavored to achieve the best evaluation results by using image-processing techniques and trying to improve preferable algorithms for many aims.[6–8] Machine vision systems and image processing techniques render possible to describe morphological, color and texture characters of wheat grains quantitatively. The identification of wheat grains and high accuracy of classification requires some knowledge of kernel characteristics such as color, length, shape, and texture.[9] These characters are regarded as major differences and can be combined to obtain the feature vector, which represents wheat grain. The majority of dissimilar properties have involved the recognition of grain varieties.[9] Proposed the feature extraction method requiring the usage of X-ray images to measure few grain traits and classify them. They indicated the usage of grain geometric properties in recognition of wheat variety and determined a principal set of these parameters according to morphology of wheat grain. In another research, Cesit-1252 durum wheat grains, which have wide harvest areas in Turkey and are the main ingredients of pasta and semolina products were investigated and classified to attain best quality wheat grains according to their vitreousness. In their study, researchers classified vitreous and starchy durum wheat kernels by training several Artificial Neural Networks (ANNs) with different amount of traits such as colour, gaborlet, morphological, and wavelet to classify vitreous and starchy durum wheat kernels.[10]

Generally, studies about the classification of various plants and their characteristics or samples have been used the computer-based algorithms in agriculture. Image-based results of some studies have been promising; however, it requires high-end imaging processing techniques and the respective test dataset which makes the method too costly and not available by the consumer frequently.[11–13] Different multivariate calibration techniques, such as K-nearest neighbors (KNN), partial least square regression and principal components analysis (PCA) models have been represented to be practical tools for relating with spectral and image analysis about wheat.[6,14,15] Recently,[7,16] distinguished and identified the promising rice and wheat varieties based on the self-collected dataset using KNN classifier.

The main purpose of this paper is to propose a simple and cost-effective approach to classifying wheat genotypes for various breeding aims. In the first step, the parents were clustered according to single linkage hierarchical clustering (SLHC) method used commonly in breeding studies. Then, three well known classifiers which are Common Vector Approach (CVA), Support Vector Machines (SVM) and KNN technique were applied to four commercial wheat cultivars (Altay-2000, Bezostaja-1, Harmankaya-99, and Kate A-1) and six doubled haploid (DH) wheat genotypes separately. Finally, the best model was selected.

Materials and Methods

In this study, four characters, which are kernel diameter, weight, hardness and moisture, are measured from four commercial wheat cultivars (Altay-2000, Bezostaja-1, Harmankaya-99, and Kate A-1) and six doubled haploid (DH) wheat genotypes (DH6, DH16, DH18, DH19, DH20, and DH21) with SKCS 4100 (Perten Instruments, North America, Inc., Springfield, IL, USA). The measurements were carried out on 300 grains of each sample according to AACC Method 55–31.[(17)] SKCS analyses were repeated 3 times for each genotype. Therefore, each genotype has 900 samples and there are 3600 features for each genotype. Finally, 45 feature vectors each of which has 80 features are constructed by using 3600 features.

In this paper, four commercial cultivars and six DH genotypes are classified separately because it gives better results when evaluated like this. That is, four-class problem and six-class problem are examined separately. In the classification stage, three well known classifiers which are CVA, SVM and KNN are used. Satisfactory results are obtained especially for training set. In addition, the genotypes were evaluated according to SLHC method for same characters.

The machine learning algorithms used in the study are summarized below:

CVA Method

CVA is a succesfull subspace-based pattern classification method.[18,19] CVA is used as a classification algorithm in many different areas such as speech recognition[18], speaker recognition[20], motor fault diagnosis[19,21], plant classification[22] and image classification.[23]

CVA is most useful especially when the dimension (n) of feature vectors in the training set of each class is greater than or equal to the number (m) of feature vectors, that is, when $\boldsymbol{n} \geq \boldsymbol{m}$.[18] This is named as the insufficient data case. After the feature vectors belonging to each class are constructed, first of all, the covariance matrix is calculated for each class. The formula of the covariance matrix is given as:

$$\Phi^c = \sum_{i=1}^{m} (\mathbf{a}_i^c - \mathbf{a}_{ave}^c)(\mathbf{a}_i^c - \mathbf{a}_{ave}^c)^T \quad \ldots(1)$$

where, c is the class index, m is the number of feature vectors, $\mathbf{a}_i^c$ is the feature vector belonging to c^{th} class and $\mathbf{a}_{ave}^c$ is the average of the feature vectors in c^{th} class.

After the covariance matrix for each class is calculated, the eigenvalue-eigenvector decomposition is applied to these covariance matrices and the eigenvalues (λ_j^c) and corresponding eigenvectors $\mathbf{u}_j^c$ are obtained. These eigenvalues are sorted in descending order. In the insufficient data case, the number of nonzero eigenvalues is equal to (m−1) and the number of zero eigenvalues is equal to (n−m+1). The eigenvectors corresponding to non-zero (*m-1*) eigenvalues span the difference subspace and the eigenvectors corresponding to zero (n−m+1) eigenvalues span the indifference subspace. Union of these two subspaces covers whole feature space.[18]

The common vector for c^{th} class is obtained by taking the projection of any feature vector of that class onto the indifference subspace of the same class[18]:

$$\mathbf{a}_{common}^c = \sum_{j=m}^{n} \left[(\mathbf{a}_i^c)^T \mathbf{u}_j^c \right] \mathbf{u}_j^c \quad \ldots(2)$$

After the common vector is obtained, the classification process is implemented by using the following decision criterion:

$$K = \operatorname{argmin}_{1 \le c \le S} \left\| \sum_{j=m}^{n} \{ [(\mathbf{a}_{test} - \mathbf{a}_i^c)^T \mathbf{u}_j^c] \mathbf{u}_j^c \} \right\|^2 \quad \ldots(3)$$

where, $\mathbf{a}_{test}$ is unknown or test vector, c is the class index and S is the total number of classes. The value of K is found for each class index and test vector is assigned to the class which has minimum value of K.

SVM Method

SVM is a binary classification algorithm developed by Vapnik and widely used in pattern recognition.[24] SVM algorithm intends to find a hyperplane which maximizes the distance between the hyperplane and each class.[25] Support vectors correspond to the data samples which are nearest to the hyperplane.

For dealing with multi-class problems (with S classes), it is possible to construct "$S(S-1)/2$" classifiers.[26] In this paper, the linear SVM classifier is used.

For multi-class problems, the strategies such as "one against one" and "one against all" are applied.[27] In this paper, the classification process is realized for multi-class problem by using "one against all" strategy.

KNN Method

KNN model is one of the frequently used methods in various areas because it is an uncomplicated classification algorithm and therefore easy to use.[28] KNN algorithm compares test data with nearest k points in the training set and find correct class by comparing similarity and dissimilarity values with data in training set. Samples in training data set are related to classes on coordinate plane in the training stage of KNN algorithm. Later, the samples to be tested are placed on coordinate plane created. Class of test data is assigned by considering the test data at the nearest neighbors to samples belonging to each class.

To run the KNN algorithm, the user must enter both a K value externally and distance criteria to be determined. K value was decided with the best k selection model as 10 in this study and closeness among the samples was evaluated by Euclidean neighbor distance.

SLHC Method

The distance between closest neighbors of two groups is determined by the single-link hierarchical clustering method.[29] The two closest groups are combined to present a new cluster which depends on the distance for each level.[30, 31] Therefore, this algorithm is named as nearest neighbor or minimum algorithm. Algorithm will stop when set threshold value is reached and will result in a tree. This algorithm will be called the single-link algorithm.[29–31]

Data processing was carried out by an IBM SPSS 20 for KNN and SLHC. The block diagram of wheat grain classification including all approaches is shown in Fig. 1.

Datasets

The datasets collected from the SKCS include single kernel weight, hardness, diameter, moisture and their standard deviations. The SKCS parameters are used to wheat classification for market value.Various datasets related with hardness classification reported by the SKCS include the hardness index.[32–35] The SKCS 4100 hardness classification logic diagram used for classifying soft, hard, and mixed wheat as officially approved by the Grain Inspectors, Packers and Stockyard Association, United States Department of Agriculture (GIPSA-USDA).[32] According to this official classification, wheats with a value of 75 and above in kernel hardness measured with SKCS are classified as hard grained, while those with a value of 33 or less are included in the soft grained wheat class.The grain dataset in the study was originally explored by Kutlu[35] and was derived from harvested wheat grains in a breeding study for aiming wheat end-use quality. The dataset contains four SKCS data mentioned above and is presented as mean values of them in Table 1. The highest average values of kernel weight were observed at cv. Bezostaja-1 and DH20 genotype. The kernel diameter values belonging to DH wheat genotypes were higher than commercial cultivars. No significant differences were observed in kernel moisture values. The kernel hardness differentiated only the cv. Altay-2000 from the others with significantly lower values. According to SKCS hardness index clasifications, cv. Altay-2000 is a soft grained while the all DH genotypes are hard grained.

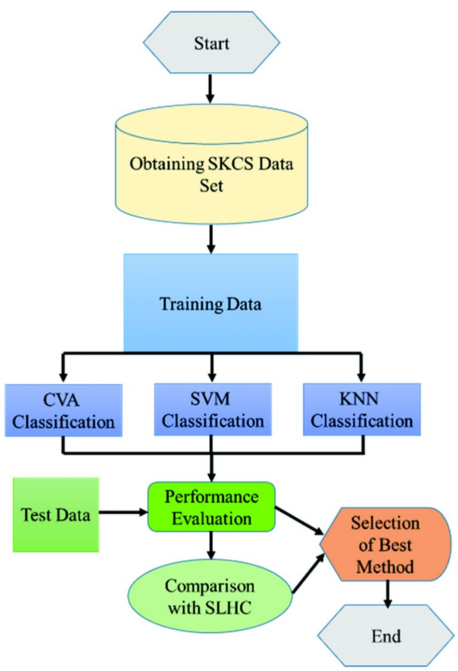

Fig. 1 — Block diagram of wheat grain classification

Results and Discussion

In this study, SKCS measurements were used to acquire the data. The results presented in the datasets section were intended to confirm via machine learning clasification methods. Since the average values presented in Table 1 are obtained from a highly variable data set, it is thought that it would be appropriate to select a classification model in which the raw data will be used rather than the average values. Therefore, the SLHC model is aimed to find the most accurate approach in evaluating the data obtained through SKCS, and compared with CVA, SVM and KNN methods which are widely used in the classification of genotypes in breeding studies. Genotypes had close values in terms of the properties studied.

According to Cluster analysis of the genotypes used in the research, four different groups were determined, commercial cultivars formed two groups and DH genotypes also formed two groups. Altay-2000, which is one of the commercial cultivar, showed that it is different from the other three cultivars as a group alone. DH6, DH16 and DH19 were collected in a group while DH18, DH20 and DH21 were classified as genotypes with high similarity (Fig. 2). Kernel hardness has been a distinctive feature for this classification.

The proposed plant classification system was tested on the dataset of ten different wheat genotypes each of which has 900 grain samples. Each sample is represented with four features. These data were trained and tested with three different classifiers (KNN, SVM and CVA).

In this study, first, four commercial cultivars (Altay-2000, Bezostaja-1, Harmankaya-99, and Kate

Table 1 — The average SKCS values of cultivars and DH genotypes

Cultivars	Kernel weight	Kernel diameter	Kernel hardness	Kernel moisture
Altay-2000	32.70	2.64	26.48	9.85
Bezostaja-1	37.24	2.78	60.44	8.70
Harmankaya-99	34.93	2.72	67.68	8.84
Kate A-1	29.25	2.56	70.77	9.07
DH genotypes				
DH16	39.78	2.91	83.56	9.47
DH18	37.28	2.83	76.18	8.98
DH19	33.49	2.82	86.83	9.70
DH20	42.25	3.02	75.25	9.79
DH21	38.69	2.89	75.73	8.63
DH6	36.78	2.85	86.49	8.27

A-1) are studied. Each of four cultivars forms one class in the classification methods. Four characters (diameter, weight, hardness and moisture) are taken from 900 grains samples of each cultivar. Therefore, there are 3600 characters or features for each cultivar. The 45 feature vectors each of which has 80 features (dimension) are constructed by using 3600 features. The "leave-five-out" strategy is used in the testing stage, that is, 40 feature vectors are used in the training stage and remaining five feature vectors are tested. Thus, testing stage has nine steps because each class has 45 feature vectors. Confusion matrices and correct recognition rates obtained by using CVA and SVM are given in Table 2 as an average of nine "leave-five-out" steps. KNN classifier is trained on 70% of the collected dataset and tested on the remaining 30%. Confusion matrix obtained by using KNN is also given in Table 2.

The correct recognition rates obtained from three classifiers for commercial cultivars are given in Fig. 3. As seen from Fig. 3, the results obtained in the test set are very low because values of the characters belonging to different genotypes are very close to each other. Also standard deviation of values in each character is very high. When the feature vectors in the training set are tested by using CVA, all classes (cultivars) are correctly classified, i.e., correct recognition rate of 100% is obtained. When the feature vectors in the training set are tested by using SVM, correct recognition rate of 85% is achieved. Correctness scores of the KNN are of 59.8% for commercial cultivars on the training dataset.

Secondly, six DH wheat genotypes (DH6, DH16, DH18, DH19, DH20, and DH21) are classified. In this case, there are six classes and each class has 45 feature vectors each of which has the dimension of 80. The "leave-five-out" strategy is also used in the testing stage, that is, 45 feature vectors are used in the training stage and remaining five feature vectors are tested. Confusion matrices and correct recognition rates obtained by using CVA, SVM are given in Table 3 as an average of nine "leave-five-out" steps. KNN classifier is trained on 70% of the collected dataset and tested on the remaining 30%. Confusion matrix obtained by using KNN is also given in Table 3.

The correct recognition rates obtained from three classifiers for DH genotypes are given in Fig. 4. The results found from the test set are appreciably low because the characters belonging to different DH wheat genotypes have very close or even same values. Meanwhile, when the number of classes increases recognition rate decreases. Also standard deviation of values in each character is very high. When the feature vectors in the training set are tested by using CVA, all classes (genotypes) are correctly classified, i.e., the correct recognition rate of 100% is attained. When the feature vectors in the training set are tested by using SVM, correct recognition rate of 69.03% is

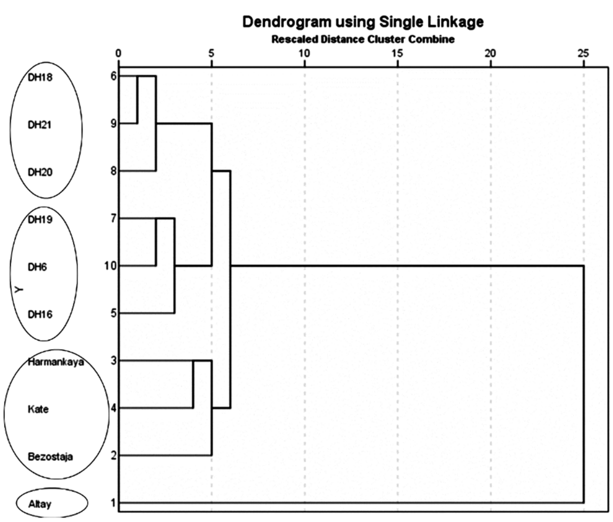

Fig. 2 — Dendrogram of all genotypes created according to single linkage hierarchical cluster (SLHC) analysis

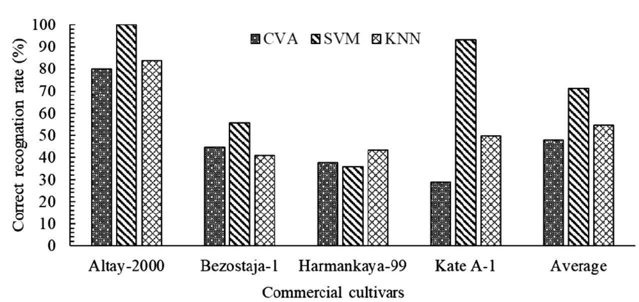

Fig. 3 — Correct recognition rates (%) obtained by using CVA, SVM and KNN for commercial cultivars

Table 2 — Confusion matrices obtained by using CVA, SVM and KNN for four commercial cultivars

Cultivars	CVA				SVM				KNN			
Altay-2000	36	7	0	2	45	0	0	0	38	2	3	3
Bezostaja-1	0	20	16	9	0	25	17	3	3	18	15	8
Harmankaya-99	1	23	17	4	0	14	16	15	1	15	19	9
Kate A-1	0	17	15	13	0	2	1	42	3	6	14	22

Table 3 — Confusion matrices obtained by using CVA, SVM and KNN for six DH genotypes

DH genotypes	CVA						SVM						KNN					
DH6	17	6	5	5	5	7	20	5	1	14	0	5	14	5	9	6	4	8
DH16	1	16	5	10	6	7	6	14	2	10	7	6	1	20	4	2	9	8
DH18	3	4	19	5	7	7	4	8	10	5	8	10	8	6	22	5	1	2
DH19	4	6	2	25	5	3	5	1	3	33	1	2	9	6	8	14	3	5
DH20	0	6	7	11	15	6	1	3	4	2	32	5	2	13	2	1	16	11
DH21	2	3	11	5	5	19	4	4	3	3	9	22	2	9	3	1	10	22

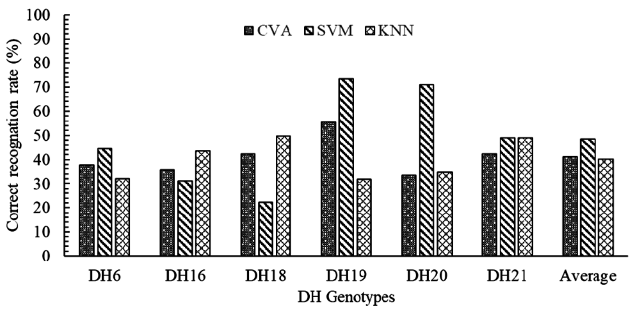

Fig. 4 — Correct recognition rates (%) obtained by using CVA, SVM and KNN for DH genotypes

achieved. Correctness scores are of 42.8% for DH genotypes on the training dataset.

Although the above-mentioned classifiers show different accuracies in comparison with each other, SVM outperforms all others for wheat classification. After experimenting with the proposed system, we conclude that SVM performs better than other classifiers for classification of these genotypes. Other two classification methods showed similar results to the clustering method and SKCS hardness index classification based on average values. In terms of ease of application, SVM can be chosen as the most suitable method for the studied properties. However, methods should be studied for more features.

Conclusions

In this paper, four class problem (classification of four commercial cultivars) and six-class problem (classification of six DH genotypes) are studied separately. Totally 900 samples for each genotype are sufficient but four characters taken from each sample are not sufficient, that is, they do not represent each genotype or class efficiently. In addition, in some cases, these characters are much different from each other for each genotype. This is another reason for representing each genotype ineffectually. Furthermore, characters in different genotypes are generally close to each other. These three adverse situations lead to confusion of classes.The main distinction for all classifiers has been the kernel hardness character, which is evident from the marked separation of the cv. Altay-2000.

The present results obtained based on employing feature normalization revealed that SVM model is quite promising for classification of wheat genotypes. The recognition scores of collected datasets are higher than the scores of other methods. Thus, this model can provide an accurate solution to the wheat genotypes for their classification and/or identification problem as alternative to sophisticated image segmentation techniques. Furthermore, the proposed method can also be used for mobile application, where even an occupational worker on the field can take a measurement from the required features of wheat varieties to find the specific category.

References

1 Miś A & Klockiewicz-Kamińska E, Technologicalsignificance of evaluation of physicalproperties of wheatsingle-kernelsandtheirheterogeneity, *ActaAgrophys*, **2002(78)** (2021) 185–197.

2 Osborne B G & Anderssen R S, Single kernel characterization principles and applications, *Cereal Chem*, **80(5)** (2003) 613–622.

3 Malik Rekha, Sharma H, Sharma I, Kundu S, Verma A, Sheoran S, Kumar R & Chatrath R, Genetic diversity of agro-morphological characters in Indian wheat varieties using GT biplot, *Aust J Crop Sci*, **8(9)** (2014) 1266–1271.

4 Sharma D, Saharan V, Joshi A & Jain D, Biochemical characterization of bread wheat (*Tritticumaestivum* L) genotypes based on SDS-PAGE, *Triticeae Genom Genet*, **6(2)** (2015) 1–7.

5 Yadawad A, Gadpale A, Hanchinal R R, Nadaf H L, Desai S A, Biradar S & Naik V R, Pyramiding of leaf rust resistance genes in bread wheat variety DWR 162 through marker assisted backcrossing. *Indian J Genet*, **77(2)** (2017) 251–257.

6 Sabanci K, Kayabasi A & Toktas A, Computer vision-based method for classification of wheat grains using artificial neural network, *J Sci Food Agric*, **97(8)** (2017) 2588–2593.

7 Dheer P & Singh P V, Classifying wheat varieties using machine learning model, *J Pharmacogn Phytochem*, **8(3)** (2019) 47–49.

8 Zhang L & Ji H, Identification of wheat grain in different states based on hyperspectral imaging technology, *Spectrosc Lett*, **52(6)** (2019) 356–366.

9 Charytanowicz M, Kulczycki P, Kowalski PA, Lukasik S & Garbacz R C, An evaluation of utilizing geometric features for wheat grain classification using X-ray images, *Comput Electron Agric*, **144** (2018) 260–268.

10 Kaya E & Saritas İ, Towards a real-time sorting system: identification of vitreous durum wheat kernels using ANN based on their morphological, colour, wavelet and gaborlet features, *Comput Electron Agric*, **166** (2019) 1–9.

11 Camelo M G A, Camacho D B H, del Villar M A A, Arenas O M L, Bello P L A & Jiménez A A R, Digital image analysis of diverse Mexican rice cultivars, *J Sci Food Agric*, **92** (2012) 2709–2714.

12 Hobson D M, Carter R M & Yan Y, Characterization and identification of rice grains through digital image analysis, *IEEE Instrumentat Measur Technol Conf*, *IMTC* (Warsaw) 2007 1–5.

13 Kong W, Zhang C, Liu F, Nie P & He Y, Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis, *Sensors*, **13(7)** (2013) 8916–8927.

14 Arazuri S, Ignacio Arana J, Arias N, Arregui L M, Gonzalez-Torralba J & Jaren C, Rheological parameters determination using near infrared technology in whole wheat grain, *J Food Eng*, **111**(2012) 115–121

15 Li Vigni M & Cocchi M, Near infrared spectroscopy and multivariate analysis to evaluate wheat flour doughs leavening and bread properties, *Anal Chim Acta*, **764** (2013) 17–23.

16 Dheer P, Distinguishing of rice varieties by using machine learning models, *Int J Adv Res Comput Commun Eng*, **8(1)** (2019) 55–57.

17 AACC, Method 55–31. *Single-Kernel Characterization System for Wheat Kernel Texture, Official Methods of AACC International* (St. Paul MN, USA), 2002.

18 Gülmezoğlu M B, Dzhafarov V & Barkana A, The common vector approach and its relation to principal component analysis, *IEEE Trans Speech Audio Process*, **9(6)** (2001) 655–662.

19 Gülmezoğlu M B & Ergin S, An approach for bearing fault detection in electrical motors, *Int Trans Electr Energy Syst*, **17(6)** (2007) 628–641.

20 Sadıç S & Gülmezoğlu M B, Common vector approach and its combination with GMM for text-independent speaker recognition, *Expert Syst Appl*, **38(9)** (2011) 11394–11400.

21 Ergin S, Uzuntas A & Gülmezoğlu M B, Detection of stator, bearing and rotor faults in induction motors, *Procedia Eng*, **30** (2012) 1103–1109.

22 Aytaç Z, Gülmezoğlu N & Gülmezoğlu M B, Application of the common vector approach on identification of winter rapeseed (*Brassica napus* L.) cultivars and their yield characters, *Int J Sustain Agric Manag Inform*, **2(1)** (2016) 66–78.

23 Kalyoncu H B, Ergin S & Gülmezoğlu M B, Block-based noisy/clean classification of images using the common vector approach, *Circuits Syst Signal Process*, **39** (2020) 1387–1418.

24 Vapnik V, *Statistical learning theory* (John Wiley & Sons, New York) 1998.

25 Guo G, Li S Z & Chan K, Face recognition by support vector machines, in *Proceedings of Fourth IEEE International Conference*, (2000) 196–201.

26 Ruiz-Pinales J, Acosta-Reyes J J, Salazar-Garibay A & Jaime-Rivas R, Shift invariant Support Vector Machines face recognition system, in *Proc World Academy Sci, Eng Technol*, (2006) 167–171.

27 Hsu C W & Lin C J, A comparison of methods for multiclass support vector machines, *IEEE Trans Neural Netw*, **13(2)** (2002) 415–425.

28 Cover T & Hart P, Nearest neighbor pattern classification, *IEEE Trans Inf Theory*, **13(1)** (1967) 21–27.

29 Webb A, *Statistical pattern recognition* (Wiley, Oxford) 2002.

30 Alpaydin E, *Introduction to machine learning-adaptive computation and machine learning series* (MIT Press, Cambridge) 2010.

31 Murphy K, *Machine learning a probabilistic perspective* (MIT Press, Cambridge) 2012.

32 Maghirang E B & Dowell F E, Hardnessmeasurement of bulk wheat by single-kernel visible and near-infrared reflectance spectroscopy, *Cereal Chem*, **80(3)** (2003) 316–322.

33 Morris C F, Bettge A D, Pitts M J, King G E, Pecka K & McCluskey P J, Compressive strength of wheatendosperm: comparison of endosperm bricks to the single kernel characterization system, *Cerea lChem*, **85(3)** (2008) 359–365.

34 Przyborowski M, Gasparis S, Kała M, Orczyk W & Nadolska-Orczyk A, Thevariability of puroindoline-encodingalleles and their ınfluence on grain hardness in modern wheat cultivarscultivated in poland, breedinglinesandpolisholdlandraces (*Triticumaestivum* L.), *Agron*, **10(8)** (2020) 1075.

35 Kutlu I, Heritability of end-use quality and biofortification characteristics in line×tester bread wheat (*Triticumaestivum* L.) crosses, *Appl Ecol Environ Res*, **16(5)** (2018) 7305–7326.