CROP BREEDING AND APPLIED BIOTECHNOLOGY

ARTICLE

Use of machine learning models-based image analysis for classification of haploid and diploid maize

Crop Breeding and Applied Biotechnology
23(4): e45322349, 2023
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332023v23n4a44

Fatih Kahrıman[1*], Abdurrahman Muhammed Güz[1] and İpek Pehlivan[1]

Abstract: *Image analysis is a straightforward and non-destructive technique used to identify haploids/diploids in maize. This study was carried out to characterize haploid/diploid maize kernels based on color space data and to compare the success of classification models developed using different machine learning techniques in maize. In this study, haploid (n=390) and diploid (n=495) kernels obtained by crossing five different donors with a Navajo inducer were used. Kernel images were collected using a standard desktop scanner. After extracting the RGB color space data, it was converted to hue-saturation-value (HSV) and Lab color spaces. Seven combinations of color space datasets were used as predictor variables. Support vector machines (SVM-C), random forest (RF), classification and regression tree (CART) methods were used to develop ML models. The classification success of the models was found between 0.74 and 0.86. The Support Vector Machines model (Accuracy = 0.86) created with RGB+Lab input data was the best.*

Keywords: *Kernel classification, image analysis, doubled haploid, machine learning*

INTRODUCTION

The in-vivo doubled haploid technique is one of the techniques proven to develop 100% homozygous lines in maize breeding. There are two types of use in practice for doubling of chromosome numbers of the plants in the in vivo conditions, which are named as “in vivo maternal” and “in vivo paternal” methods (Chaikam et al. 2019). Both relied on using special genotypes called “inducer” lines. Since the discovery that haploid seeds are formed in donor materials hybridizing with these inducers, numerous inducer lines with different features have been developed (Choe 1959, Chalky 1994, Prigge et al. 2012, Kalinowska et al. 2019, Uliana Trentin et al. 2020). There are three main steps of in-vivo maternal or paternal doubled haploid techniques in practice. The first step is the hybridization of the donor material with inducer line, and the second step is selection of haploid samples based on the color changes in the seeds or the root of seedlings, and the last step is chemical treatment for chromosome doubling after the selfing plants by growing them under field or greenhouse conditions (Röber et al. 2005, Chidzanga et al. 2017). Within these steps, separation of haploid samples from the others is one of the most important processes for the success of the technique. In the classical method,

***Corresponding author:**
E-mail: fkahriman@hotmail.com
ORCID: 0000-0001-6944-0512

Received: 10 May 2023
Accepted: 05 October 2023
Published: 20 October 2023

[1] Çanakkale Onsekiz Mart University, Faculty of Agriculture, Barbaros District, Terzioğlu Campus, Prof. Dr. Sevim Buluç St. No:62, 17100 Merkez/Çanakkale, Türkiye

seed classification is performed using the method based on visual phenotypic markers distinguishable by human eyes. However, it takes a long time and its margin of error for classification is high. For this reason, alternative methods that can perform seed discrimination in a more practical way are emphasized. There are numerous approaches to separate haploid and diploid seeds such as stomata measurements (Ribeiro et al. 2022), flow cytometry (Baleroni et al. 2021), near infrared reflectance (NIR) spectroscopy (Jones et al. 2012, Liu et al. 2017, Cui et al. 2019), near infrared transmittance (NIT) spectroscopy (Lin et al. 2017) and image analysis (Veeramani et al. 2018). Among these methods, image analysis is one of the alternatives with significant advantages.

Image analysis is practically used in solving classification, detection or characterization problems encountered in many different fields today. The use of this method in the discrimination of haploid/diploid maize kernels has also been the subject of different scientific studies. There are a considerable number of papers showing that separation of haploid/diploid kernels can be done through image data analysis, especially samples obtained using inducer lines carrying Navajo genes (Altuntaş et al. 2018, Veeramani et al. 2018, Altuntaş et al. 2019, Altuntaş and Kocamaz 2019). The input data used in these studies are data from different color spaces obtained from maize haploid and diploid kernel samples. Color spaces can be used for different purposes, such as sample characterization and feature extraction based on image analysis. The RGB color space, which is accepted as basic color space today, covers three basic channels (Red, Green, Blue); apart from it, there are also color spaces such as hue-saturation-value (HSV) and luminance-red/green axis-blue/yellow axis (Lab). Studies on the discrimination of haploid/diploid maize kernels based on color space data are also found in the literature (Altuntaş and Kocamaz 2019). However, in our best knowledge, there is no study on the selection of haploid samples using different color spaces.

Among the techniques used to develop sample classification models, machine learning methods are among the ones that give successful results. Different studies have been conducted using machine learning techniques to distinguish haploid/diploid maize seeds, and model success depends not only on the algorithm of the technique used, but also on the size of the input data, its qualities, and its discrimination-enhancing properties (Zhang et al. 2013). In this respect, it is useful to investigate the color space effect in detail in models created with different machine learning methods.

This study was carried out to i) investigate the differences in haploid and diploid maize kernels based on the mean values of the channels of the color spaces ii) evaluate relationships between features extracted from color spaces, and iii) to examine the classification success of different machine learning methods when the color space data are used separately or together.

MATERIAL AND METHODS

In this study, a total of 885 seed samples obtained from crossing five donors with an inducer line were used as material. Donors are F2 materials obtained from maize breeding studies carried out in Çanakkale Onsekiz Mart University, Faculty of Agriculture, Department of Field Crops. All donors have the dent type of kernel morphology. Inducer line (CIM2GTAIL-P2) was used with permission from The International Maize and Wheat Improvement Center (CIMMYT) for scientific studies in the Field Crops Department of Çanakkale Onsekiz Mart University (Turkey), Faculty of Agriculture (Table 1).

Induction crossing was performed under field conditions in the year of 2020 in Çanakkale, Turkey. Each donor was planted to four-rowed plots with a 70x20 cm plant density. When the plants reached the flowering stage, pollination was carried out with at least 10 plants belonging to each donor genotype by using the CIM2GTAIL-P2 inducer line as the pollen source (male parent). The controlled pollination method suggested by Kahrıman (2016) was used in the induction crossing. In particular, the bulk pollination method was used. In this context, the ears of the donor plants were protected with shoot bags before showing the silks. When protected ears form the silk, they are pollinated using the collected pollen from 5-10

Table 1. Donor and inducer materials used in this study

Code	General Features	Source
DON1	Donor material, Experimental F2	COMU
DON2	Donor material, Experimental F2	COMU
DON3	Donor material, Experimental F2	COMU
DON4	Donor material, Experimental F2	COMU
DON5	Donor material, Experimental F2	COMU
CIM2GTAIL-P2	Second generation inducer line	CIMMYT

COMU: Çanakkale Onsekiz Mart University, CIMMYT: International Maize and Wheat Improvement Center.

plants of inducer line. The harvest was carried out by hand and DH_0 seeds obtained from induction crosses were kept at +4 °C for use in the further steps.

The classification of putative haploid kernels was made according to the coloration of the embryo and crown area in DH_0 kernels. If there was coloration in the crown area and without coloration in the embryo of the seed examined, this kernel was considered as haploid, and if there was coloration in both the crown area and the embryo region, this kernel was evaluated as diploid.

Image classification at the single seed level was performed by making some revisions to the method suggested by Altuntaş et al. (2018). The images of the kernels, which were visually separated by the eye, were recorded by the embryo side in the desktop scanner (HP Scanjet 3970, USA) with jpeg extension. Digital images were saved with 300 dpi resolution. Also, a specific black background was used to obtain clean kernel images, and this background was removed before the extraction of image features. A total of 5 different moments (mean, minimum, maximum, standard deviation, and median) were extracted to be used in modeling studies for each channel of color spaces. Thus, we obtained 45 different features for all color spaces. Segmentation and feature extraction operations were carried out with the R program using *EBImage* and *colorspace* packages (R Core Team 2019). These data were kept as an excel file to be used in model development studies.

To develop classification models, six different features (mean, standard deviation, median, mode, skewness, kurtosis) extracted from the images were taken as predictive variables. Support vector machines (SVM), the Classification and Regression Tree (CRT) and Random Forest (RF) methods were used. SVM identifies critical support vectors, data points that define the edge between different classes in space. It maximizes the margin between these classes by finding a hyperplane or line that separates them. SVM's advantage lies in focusing on support vectors, not the entire dataset, making large training sets manageable. SVM performs well in non-linear, sparse, and high-dimensional problems. However, its sensitivity to variable settings poses a challenge, requiring careful configuration for optimal results. Decision trees, also known as CRT, have been fundamental in data mining and are a cornerstone of classic machine learning algorithms. Since their inception in the 1980s, they have enjoyed widespread usage as a model-building tool for data mining based on machine learning. Their appeal lies in the straightforwardness of the resulting model, particularly smaller decision trees which are easily comprehensible, interpretable, and can be effectively communicated to management. The structure of decision trees is versatile, capable of representing both classification and regression models. RF is an ensemble of unpruned decision trees used for large datasets with many input variables. Each tree is built from a random subset of the training data, and they collectively vote on outcomes. This approach offers robustness against noise and overfitting. Randomness in dataset and variable selection enhances resilience and computational efficiency. Random forests require minimal data preprocessing, adapt well to variable selection, and excel in reducing overfitting risks through ensemble techniques. All models were developed in the R program using the *rattle* package (R Core Team 2019). Default parameters are assigned in the rattle package to develop classification models.

Boxplot plots were used to compare haploid diploid seed groups based on the data of the channels belonging to the color spaces. While creating these graphs, the averages of each channel of three different color spaces were used. The differences between the mean values of each channel were compared with the t test. The *ggstatsplot* package in the R package program was used to create the graphs.

The Spearman correlation test was used to examine the relationships between the features extracted from the color spaces. In order to examine these relationships according to haploid and diploid seed groups, analyses were performed separately for two seed classes and the results of the correlation analysis were shown as network graphs. Only significant relationships (r>30) were shown in these graphs. In these graphs, blue color is used for positive correlation and red color is used for negative correlation.

For comparison of the created classification models, the performance metrics predicted over the confusion matrix in Table 2. The performance metrics of the classification models

Table 2. Confusion matrix template used in calculations for classification of haploid and diploid seeds

	Actual	
Prediction	Haploid	Diploid
Haploid	TP	FN
Diploid	FP	TN

FP: False positive, FN: False negative, TP: True positive, TN: True negative.

were calculated according to the equations 1-6. These metrics are calculated for training and external validation sets separately. In these equations, TP, FP, TN, FN are true positive, false positive, true negative and false negative, respectively.

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$Specificity = \frac{TN}{FP + TN} \quad (2)$$

$$Pos.Pred.Val. = \frac{Sensitivity \times Prevalence}{((Sensitivity) \times Prevalence) + ((1 - Specificity) \times (1 - Prevalence))} \quad (3)$$

$$Neg.Pred.Val = \frac{Specificity \times (1 - Prevalence)}{((1 - Sensitivity) \times Prevalence) + ((Specificity) \times (1 - Prevalence))} \quad (4)$$

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \quad (5)$$

$$F1\ Score = \frac{(1 + Beta2) \times Precision \times Recall}{(Beta2 \times Precision) + Recall} \quad (6)$$

RESULTS AND DISCUSSION

The average values of color channels by haploid and diploid kernels are shown in Figure 1. Haploid and diploid kernels showed considerable differences according to different color spaces. Haploid samples had a higher average than the diploid group for all channels of RGB and HSV color spaces. Contrary to this situation, the average of the diploid group in the a channel in Lab color space was higher than the haploid group. Lab color-space consists of three channels, L, a and b; the L channel is related to lightness, and it takes values from 0 (black) to 100 (white). The a channel is related to red/green colors; positive and negative a* describe red and green values, respectively. Also, it is related to saturation. The b channel indicates yellowness or blueness; positive and negative b* describe yellow and blue values (Ly et al. 2020). Results revealed that haploid samples had low saturation value because they had high value of channel a. The absence of anthocyanin formation in haploid samples, especially in the embryo section, may be a result of this situation.

To the best of our knowledge, there is no study comparing the characteristics of haploid and diploid maize kernels in different color spaces. Therefore, there is a limitation for the comparison of the results for the characteristics of haploid and diploid maize kernels of the current study with the previous findings. However, it was understood that there may be distinctive features among haploid and diploid kernels in terms of the channels of color-spaces. For example, haploids have lower average values for color channels than diploids, except the a* channel in the Lab color-space. This may be used as a distinguishing criterion for haploid and diploid maize kernels. As a matter of fact, the results of previous studies in which color space features were used to distinguish normal maize genotypes confirm this situation. Beyaz and Koç (2021) compared the maize genotypes having different endosperm characteristics, such as hybrid, sugar and dent maize in terms of average values of RGB color space. In this study, the order of average values of the color channels in all maize genotypes is R>G>B. On the other hand, genotypes rank as hybrid > sugar > dent in terms of channel averages. Another study focused on separating different maize genotypes using HSV color space and seed morphological characteristics, and its results revealed that color space data gave better classification results than morphological trait data (Yafie et al. 2020).

The correlation network graphs created to compare the relationships between the features of the color spaces based on the data obtained from the diploid and haploid kernels are presented in Figure 2. It was observed that the mean and median values of the green, red, L and V channels were positively correlated with the mean and median values of the blue channel for the color space data obtained from diploid kernels. It is observed that the A and H channels are clearly congregated, and their minimum values coexist on the network plot of diploid samples. The other group on the network was mostly related to the maximum, median and standard deviations (Figure 2a). Average and median values of green, blue, red, V and L channels in the network of haploid seed samples formed a cluster in the network graph for haploid kernels (Figure 2b). The relations and distributions of other moments for color channels showed differences. It is not easy to clarify all relationships between moments on the network plots due to the high number of features (n= 45) extracted from color space datasets. However, we could say that some moments have characteristic relationships according to haploid and diploid sample sets.

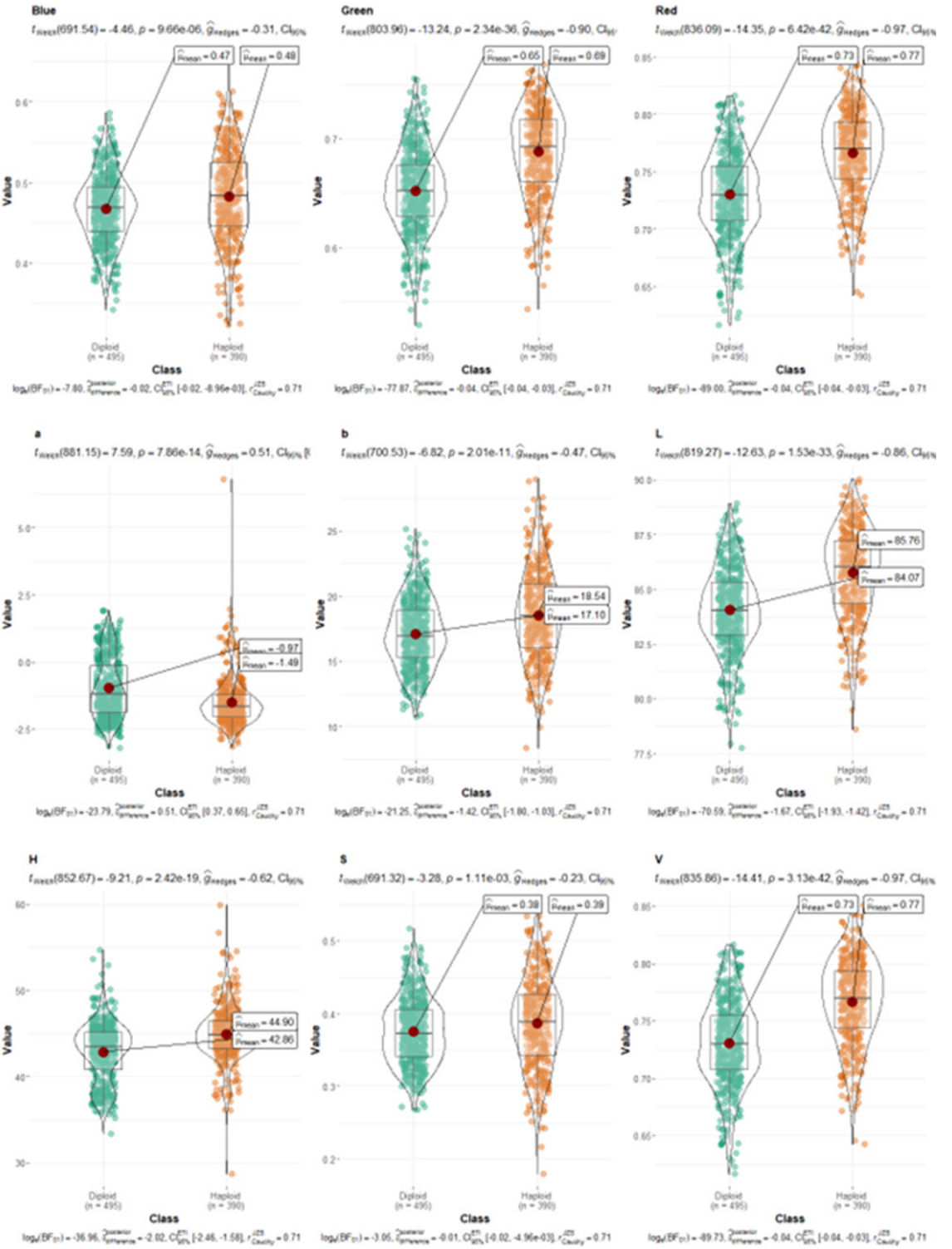

Figure 1. Plots of mean pixel values for RGB, HSV and Lab color spaces of haploid and diploid seed samples.

The relationships between color space data obtained from haploid and diploid maize kernels were not discussed in current literature. On the other hand, the relationships between color characteristics of seeds and endosperms of cereals have been the subject of some studies. Zapotoczny and Majewska (2010) calculated correlations between RGB, HSV, HSL and Lab color-spaces data in wheat using different imaging devices. They found positive correlations in RGB and Lab color spaces, while negative correlations were observed for HSV and HSL color spaces. It was observed that

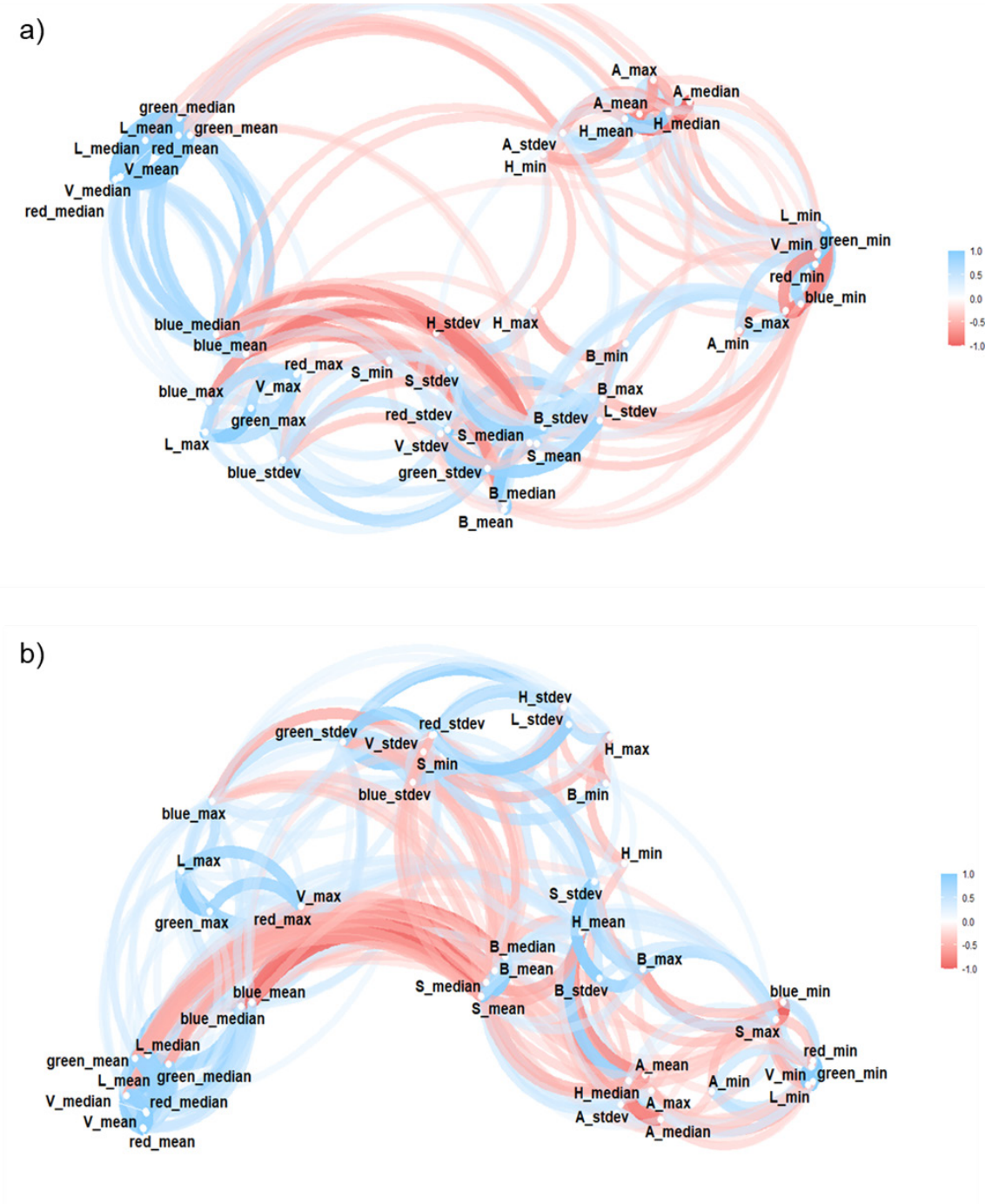

Figure 2. Network plot showing the correlation between features of color spaces in diploid (a) and haploid (b) kernel samples.

there were positive correlations between the parameters of RGB data in our study. This result is in agreement with the previous studies.

The numbers for TP, TN, FP and FN in the training and validation sets are shown in Table 3. In the training set, it was determined that the RF model was able to correctly classify haploid seeds for all data sets (Table 3). On the other hand, this classification success of RF models was not achieved in the validation set. Classification success was found to be close to each other in models created with training data with CART and SVM methods. In the training set, effects of using different color-spaces together in the prediction models showed differences according to the modeling technique used. It has been observed that in the use of color spaces in combinations of 2 and 3 the SVM model improves the classification success of TPs. However, this effect disappeared in the validation set and the classification success increased if the color spaces were used separately for the SVM method (Table 3).

The performance metrics allowed us to make more detailed evaluations about the robustness of the developed models. Considering the averages of the performance metrics (Table 4.), it was observed that all statistics of the RF method in the learning set were equal to 1. After the RF method, the most successful results were obtained from the SVM models for the learning set. Accuracy values were found between 0.82 and 0.85 in the CART models. The fact that the sensitivity value is lower than the specificity value and the NPV value is below the PPV value in the CART technique indicates that the classification success of diploid seeds is low in this method. In other words, it can be said that the probability of classifying seeds that are actually diploid as haploid is high in the CART modeling. Evaluation model performances based only on sensitivity and specificity values could be biased if the numbers of samples are different in haploid and diploid classes (Altuntaş and Kocamaz 2019). Quality index or F1-Score should be considered for true decision on the robustness of the developed models.

Table 3. Evaluation of model performances for training and validation sets according to color spaces

		Training Set				Validation Set			
Model	**Color Space**	**TP**	**FN**	**FP**	**TN**	**TP**	**FN**	**FP**	**TN**
CART	HSV	245	79	31	264	124	47	13	82
	Lab	272	52	39	256	132	39	26	69
	RGB	273	51	43	252	140	31	22	73
	HSV+Lab	284	40	45	250	137	34	37	58
	RGB+HSV	240	84	19	276	124	47	10	85
	RGB+Lab	282	42	37	258	131	40	26	69
	RGB+HSV+Lab	248	76	27	268	126	45	18	77
	Mean	263	61	34	261	131	40	22	73
RF	HSV	324	0	0	295	149	22	17	78
	Lab	324	0	0	295	145	26	15	80
	RGB	324	0	0	295	137	34	17	78
	HSV+Lab	324	0	0	295	150	21	17	78
	RGB+HSV	324	0	0	295	143	28	17	78
	RGB+Lab	324	0	0	295	144	27	15	80
	RGB+HSV+Lab	324	0	0	295	143	28	15	80
	Mean	324	0	0	295	144	27	16	79
SVM	HSV	276	48	36	259	137	34	16	79
	Lab	283	41	42	253	147	24	17	78
	RGB	280	44	50	245	147	24	16	79
	HSV+Lab	298	26	17	278	142	29	18	77
	RGB+HSV	290	34	28	267	142	29	17	78
	RGB+Lab	291	33	18	277	150	21	15	80
	RGB+HSV+Lab	309	15	11	284	143	28	21	74
	Mean	290	34	29	266	144	27	17	78

FP: False positive, FN: False negative, TP: True positive, TN: True negative.

Table 4. The performance metrics of classification models according to color spaces

		Training Set						Validation Set					
Modeling Tec.	**Color Space**	**Sens.**	**Spec.**	**PPV**	**NPV**	**Acc.**	**F1**	**Sens.**	**Spec.**	**PPV**	**NPV**	**Acc.**	**F1**
CART	HSV	0.756	0.895	0.888	0.770	0.826	0.817	0.725	0.863	0.905	0.636	0.794	0.805
	Lab	0.840	0.868	0.875	0.831	0.854	0.857	0.772	0.726	0.835	0.639	0.749	0.802
	RGB	0.843	0.854	0.864	0.832	0.848	0.853	0.819	0.768	0.864	0.702	0.794	0.841
	HSV+Lab	0.877	0.847	0.863	0.862	0.862	0.870	0.801	0.611	0.787	0.630	0.706	0.794
	RGB+HSV	0.741	0.936	0.927	0.767	0.838	0.823	0.725	0.895	0.925	0.644	0.810	0.813
	RGB+Lab	0.870	0.875	0.884	0.860	0.872	0.877	0.766	0.726	0.834	0.633	0.746	0.799
	RGB+HSV+Lab	0.765	0.908	0.902	0.779	0.837	0.828	0.737	0.811	0.875	0.631	0.774	0.800
	Mean	0.813	0.883	0.886	0.814	0.848	0.846	0.764	0.771	0.861	0.645	0.768	0.808
RF	HSV	1.000	1.000	1.000	1.000	1.000	1.000	0.871	0.821	0.898	0.780	0.846	0.884
	Lab	1.000	1.000	1.000	1.000	1.000	1.000	0.848	0.842	0.906	0.755	0.845	0.876
	RGB	1.000	1.000	1.000	1.000	1.000	1.000	0.801	0.821	0.890	0.696	0.811	0.843
	HSV+Lab	1.000	1.000	1.000	1.000	1.000	1.000	0.877	0.821	0.898	0.788	0.849	0.888
	RGB+HSV	1.000	1.000	1.000	1.000	1.000	1.000	0.836	0.821	0.894	0.736	0.829	0.864
	RGB+Lab	1.000	1.000	1.000	1.000	1.000	1.000	0.842	0.842	0.906	0.748	0.842	0.873
	RGB+HSV+Lab	1.000	1.000	1.000	1.000	1.000	1.000	0.836	0.842	0.905	0.741	0.839	0.869
	Mean	1.000	1.000	1.000	1.000	1.000	1.000	0.845	0.830	0.899	0.749	0.837	0.871
SVM	HSV	0.852	0.878	0.885	0.844	0.865	0.868	0.801	0.832	0.895	0.699	0.816	0.846
	Lab	0.873	0.858	0.871	0.861	0.866	0.872	0.860	0.821	0.896	0.765	0.840	0.878
	RGB	0.864	0.831	0.848	0.848	0.847	0.856	0.860	0.832	0.902	0.767	0.846	0.880
	HSV+Lab	0.920	0.942	0.946	0.914	0.931	0.933	0.830	0.811	0.888	0.726	0.820	0.858
	RGB+HSV	0.895	0.905	0.912	0.887	0.900	0.903	0.830	0.821	0.893	0.729	0.826	0.861
	RGB+Lab	0.898	0.939	0.942	0.894	0.919	0.919	0.877	0.842	0.909	0.792	0.860	0.893
	RGB+HSV+Lab	0.954	0.963	0.966	0.950	0.958	0.960	0.836	0.779	0.872	0.725	0.808	0.854
	Mean	0.894	0.902	0.910	0.885	0.898	0.902	0.842	0.820	0.894	0.743	0.831	0.867

The success of the models varied according to color space data used for model development based on the results obtained from the external validation set. For the CART modeling technique, the use of RGB+HSV data gave a more successful classification result than other data sets. In the RF technique, it has been observed that the best classification results were obtained from the prediction model where HSV and Lab color spaces are used together. It also noted that the averages of the performance metrics of the RF model were higher than those of the other modeling methods. However, the best prediction performance was obtained from SVM model in which the RGB+Lab data was used as the estimator (Sens=0.877, Spec=0.842, PPV=0.909, NPV=0.792, Accuracy=0.860, F1 Score=0.893) (Table 4).

There are several studies in which image processing has given successful results for separating haploid and diploid maize samples. SVM, Random Forest (RF) and Logistic Regression methods provided a classification success of 87.6% in a previous study where kernel images were taken by the embryo side (Veeramani et al. 2018). Zhang et al (2013) used the RGB data to separate haploid and cross kernels and they achieved the accuracy of 98.04% and 94.4% for haploids and diploids, respectively. Our results were in agreement with these findings. However, it should not be ignored that differences in imaging devices or image resolutions were used in the mentioned studies. Undoubtedly, the quality of the input data used in the studies also affected the model success.

Also, we could compare our results with other techniques such as stomata traits and NIR measurements. Jones et al. (2012) achieved haploid discrimination with a success rate of 87.5% using NIR spectroscopy. Ribeiro et al. (2022) suggested that stomata density is an efficient parameter for distinguishing doubled haploids from false positives in in vivo haploid technique. The prediction model created in our study achieved similar success to these methods.

CONCLUSION

In this study, haploid and diploid maize kernels were differentiated based on their distinct features in various color spaces. Mean values of haploid kernels consistently exceeded those of diploids in most color channels, demonstrating the potential of mean color values for effective classification. The choice of modeling technique was crucial, with SVM applied to RGB+Lab data yielding the best results. This suggests that combining color space moments and selecting the right modeling approach improves separation accuracy. Also, using data from color spaces together eliminates the difficulties in implementing separate models based on different color space datasets. Future work could explore diverse color space combinations and modeling techniques for even better results. Using higher-resolution imaging devices and expanding the range of materials could also enhance classification. These models have promising applications in software and device development.

REFERENCES

Altuntaş Y and Kocamaz AF (2019) Renk momentleri ve destek vektör makineleri kullanarak haploid mısır tohumlarının tanımlanmasında renk uzaylarının sınıflandırma performansına etkisinin karşılaştırılması. **Fırat Üniversitesi Mühendislik Bilimleri Dergisi 31**: 551-560.

Altuntaş Y, Cömert Z and Kocamaz AF (2019) Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. **Computers and Electronics in Agriculture 163**: 1-11.

Altuntaş Y, Kocamaz AF, Cengiz R and Esmeray M (2018) Classification of haploid and diploid maize seeds by using image processing techniques and support vector machines. In IEEE 2018 26th Signal processing and communications applications conference (SIU). IEEE, İzmir, p.1-4.

Baleroni AG, Ré F, Pelozo A, Kamphorst SH, Carneiro JWP, Rossi RM and Scapim CA (2021) Identification of haploids and diploids in maize using seedling traits and flow cytometry. **Crop Breeding and Applied Biotechnology 21**: e38422145.

Beyaz A and Koç D (2021) Meta-learning-based prediction of different corn cultivars from color feature extraction. **Journal of Agricultural Sciences 27**: 32-41.

Chaikam V, Molenaar W, Melchinger AE and Boddupalli PM (2019) Doubled haploid technology for line development in maize: technical advances and prospects. **Theoretical and Applied Genetics 132**: 3227-3243.

Chalky ST (1994) Properties of maternal haploid maize plants and potential application to maize breeding. **Euphytica 79**: 13-18.

Chidzanga C, Muzawazi F, Midzi J and Hove T (2017) Production and use of haploids and doubled haploid in maize breeding: A review. **African Journal of Plant Breeding 4**: 201-213.

Choe EH (1959) A line of maize with high haploid frequency. **American Naturalist 93**: 381-382.

Cui Y, Ge W, Li J, Zhang J, An D and Wei Y (2019) Screening of maize haploid kernels based on near infrared spectroscopy quantitative analysis. **Computers and Electronics in Agriculture 158**: 358-368.

Jones RW, Reniot T, Frei UK, Tseng Y, Lubberstedt T and Mcclelland JF (2012) Selection of haploid maize kernels from hybrid kernels for plant breeding using near infrared spectroscopy and SIMCA analysis. **Applied Spectroscopy 66**: 447-450.

Kahrıman F (2016) **Mısırda polen etkisi ve bu etkinin kontrolünde uygulanan yöntemler**. Lambert Academic Publishing Saarbrücken, Berlin, 52p.

Kalinowska K, Chamas S, Unkel K, Demidov D, Lermontova I, Dresselhaus T, Kumlehn J, Dunemann F and Houben A (2019) State-of-the-art and novel developments of in vivo haploid technologies. **Theoretical and Applied Genetics 132**: 593-605.

Lin J, Yu L, Li W and Qin H (2017) Method for identifying maize haploid seeds by applying diffuse transmission near-infrared spectroscopy. **Applied Spectroscopy 72**: 611-617.

Liu WJ, Li WJ, Li HG, Qin H and Ning X (2017) Research on the method of identifying maize haploid based on KPCA and near infrared spectrum. **Spectroscopy and Spectral Analysis 37**: 2024-2027.

Ly BCK, Dyer EB, Feig JL, Chien AL and Bino SD (2020) Research techniques made simple: cutaneous colorimetry: a reliable technique for objective skin color measurement. **The Journal of Investigative Dermatology 140**: 3-12.

Prigge V, Schipprack W, Mahuku G, Atlin GN and Melchinger AE (2012) Development of in vivo haploid inducers for tropical maize breeding programs. **Euphytica 185**: 481-490.

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for statistical computing. Available at <https://www.R-project.org>. Accessed on December 18, 2022.

Ribeiro MR, Trindade RDS, Souza IRPD, Carneiro AA, Azevedo TCD, Guimarães LJM, Chaves SFS and Dias LAS (2022) Are stomatal area and stomatal density reliable traits for identification of doubled haploids in maize? **Crop Breeding and Applied Biotechnology 22**: e42222226

Röber FK, Gordillo GA and Geiger HH (2005) In vivo haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding. **Maydica 50**: 275-283.

Uliana Trentin H, Frei UK and Lübberstedt T (2020) Breeding maize maternal haploid inducers. **Plants (Basel) 9**: 614.

Veeramani B, Raymond JW and Chanda P (2018) DeepSort: deep convolutional networks for sorting haploid maize seeds. **BMC Bioinformatics 19**: 289.

Yafie HA, Rachmawati E, Prakasa E and Nur A (2020) Corn seeds identification based on shape and colour features. **Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika 6**: 66-72.

Zapotoczny P and Majewska K (2010) A comparative analysis of color measurements of the seed coat and endosperm of wheat kernels performed by various techniques. **International Journal of Food Properties 13**: 75-89.

Zhang J, Wu Z, Song P, Li W, Chen S and Liu J (2013) Embryo feature extraction and dynamic recognition method for maize haploid seeds. **Transactions of the Chinese Society of Agricultural Engineering 29**: 199-203.