Selcuk Journal of Agriculture and Food Sciences

http://sjafs.selcuk.edu.tr/sjafs/index

Research Article

SJAFS

(2021) 35 (3), 229-243

e-ISSN: 2458-8377

DOI:10.15316/SJAFS.2021.252

Determination of Effective and Specific Physical Features of Rice Varieties by Computer Vision In Exterior Quality Inspection

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

[1]Selçuk University, Faculty of Technology, Department of Computer Engineering, Konya, Turkey

ARTICLE INFO

Article history:
Received date: 14.10.2021
Accepted date: 26.10.2021

Keywords:
Color spaces
Feature extraction
Feature selection
Image processing
Quality control

ABSTRACT

In this study, feature extraction processes were performed based on the image processing techniques using morphological, shape and color features for five different rice varieties of the same brand. A total of 75 thousand pieces of rice grain were obtained, including 15 thousand pieces of each variety of rice. Pre-processing operations were applied to the images and made available for feature extraction. A total of 106 features were inferred from the images; 12 morphological features and 4 shape features obtained using morphological features and 90 color features obtained from five different color spaces (RGB, HSV, L*a*b*, YCbCr, XYZ). In addition, for the 106 features obtained, features were selected by ANOVA, X2 and Gain Ratio tests and useful features were determined. In all tests, out of 106 features, the 5 most effective and specific features were obtained roundness, compactness, shape factor 3, aspect ratio and eccentricity. The color features were listed in different order following these features.

1. Introduction

Rice is the most important product after wheat and corn when the production values of grain products worldwide are considered. Rice is a grain product that is quite rich in carbohydrates and starch. In addition, it also has great significance in human nutrition due to its economical and nutritious value. In the same time, it is also widely used in many fields of industry (Juliano 1993).

There are different quality criteria for rice varieties produced in the world. These criteria include properties of rice such as cooking properties, physical appearance, aroma and taste properties as well as productivity. From the point of view of the ultimate consumer, the first criterion that comes to mind for rice varieties sold in packages on market shelves is physical appearance (Webb 1991, Hua, Xu et al. 2021). Therefore, more technological and effective methods are needed. It is especially inefficient and take excessive time to calibrate rice and separate them within various quality criteria during production especially one considers the high production volume.

Recent studies using image processing and machine learning methods have been studied in the literature. In one of these studies, in which the projection areas of some grain products such as wheat, barley, corn, chickpeas, lentils, beans, kidney beans and soy were determined by image processing technique, the projection areas of the products in three different locations were determined. With the UTHSCSA (University of Texas Health Science Center, San Antonio) image processing program, the relationships between the projection areas were analyzed by regression analysis by obtaining feature values such as length, width and thickness of the products used in the study. As a result of the research, it was concluded that the image processing technique is sufficient for the precise determination of the projection areas of small grain products. Morphological features of 13 different wheat varieties from bread and durum wheat type with image processing technique have been evaluated using the UTHSCSA Image Tool Version 3.0 program. As a result, it has been concluded that the results of measurements obtained by hand and image processing are close and that the image processing technique can be used to determine some of the morphological features of wheat grains (Demirbas and Dursun 2007). In another study using dried tobacco leaves, a system based on machine vision techniques was developed for automatic examination of the leaves. In this system, it is aimed to analyze tobacco leaves using color, size, shape and surface texture features. Based on experimental results, it is stated that this system is a suitable route for dried tobacco leaves. Besides, it is also stated that these features of tobacco leaves can be used for automatic classification as the purpose of the next studies (Zhang and Zhang 2008).

* Corresponding author email: *mkoklu@selcuk.edu.tr*

Aggarwal and Mohan analyzed the aspect ratio using the image processing technique for the grain quality of the rice. Rice samples from three different classes (full, semi-and broken) sold in grocery stores and priced according to their size were taken. It is aimed to examine the mixtures of these samples and determine the reference aspect ratio on the market (Aggarwal and Mohan 2010). Another study aimed to extract the morphological features of 5 pieces pasta wheat varieties called Showa, Altar 84, Altar 84-3, Dipper and Bushen with image processing technique. Five sets of features were used for linear discrimination analysis. 67.66% classification accuracy was obtained as the best result of the discrimination analysis using 11 morphological features (Farahani 2012). One of the studies was carried out using Gujarat-17 rice seeds with image analysis. It was aimed to perform quality analysis using the area, major axis length, minor axis length and eccentricity features on a certain number of samples. It is stated that the traditional quality assessment made by humans can be time consuming and expensive, and that quality analysis can be done without destruction with image analysis (Maheshwari, Jain et al. 2012). In a study on defining barley varieties, the effectiveness of determining the varieties based on the shape, color and texture features obtained from the seed image was evaluated. In the study, success rates ranging from 67% to 86% were achieved in linear distinctive analysis and artificial neural networks classifications. It has been stated that the classification results can be improved by standardizing seed images in terms of front and rear orientations and additional analyses that can be applied to wrinkled regions on the seed (Szczypiński, Klepaczko et al. 2015).

In another study on rice grains, a method for quality analysis using image processing techniques and geometric features of grains was proposed. Using MATLAB technology, the seven geometric features of rice grains (major axis length, minor axis length, eccentricity, area, orientation, perimeter, aspect ratio) were extracted from digital images and then grains belonging to certain varieties are divided into three different classes. The error ratio measuring different geometric features between the recommended method and the experimental analysis was achieved between -1.39% and 1.40% (Kaur and Singh 2015). Tin at all carried out using image processing techniques on 5 different varieties of rice specific to Myanmar, 5 morphological features were extracted for each variety. The study is stated to be realized out to develop computer-based systems in order to automatically classify rice varieties (Tin, Mon et al. 2018).

When the studies conducted in recent years are examined, it can be seen that image processing and machine vision systems are used on various grain products. In these studies, products were examined in terms of various morphological features such as quality, texture, color and size. In addition to morphological features, it also can be seen that various studies have been done using shape and color features.

2. Materials and Methods

In the study, a total of 75 thousand rice grain images were obtained, primarily 15 thousand for each variety (Available from: www.muratkoklu.com/datasets/Rice_Image_Dataset.rar). The resulting images have been prepared for the feature extraction phase by undergoing various pre-operations in MATLAB software. For images passed from pre-processing, 12 morphological features, 4 shape features obtained using morphological features, and with 90 color features obtained from five different color spaces, a total of 106 features were extracted (Available from: www.muratkoklu.com/datasets/Rice_MSC_Dateset.rar). Density distribution graphs of morphological and shape features were given and the distributions of rice varieties on features were examined. Finally, 106 features obtained to reduce the data size were feature selection made by ANOVA, X2 and Gain Ratio tests and the effective features were determined. The process stages of the study are given in Figure 1.

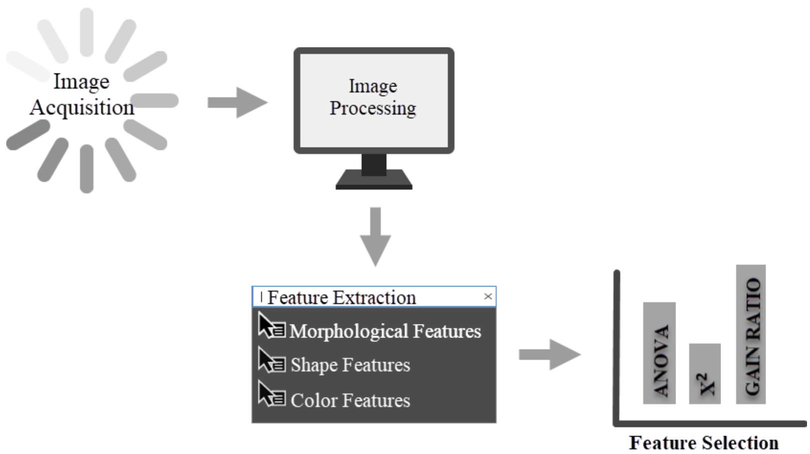


Figure 1
Process steps for the study

2.1. Image Acquisition

In image processing, firstly, the image is obtained with the help of sensors or cameras and digitization process is performed. In Figure 2 shows the image capture and digitization processes.

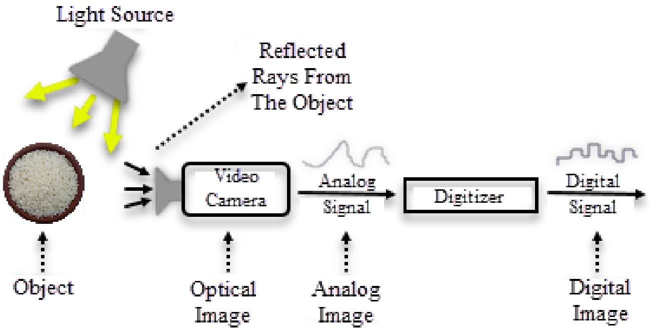


Figure 2
Capture and digitization of images

In Figure 2, reflected rays from the object illuminated by the light source are transferred to the camera. With the help of the camera, the object image is converted into analog form. The digitizer converts analog signals to digital signals, allowing the image to be processed and analyzed in a computer environment. In Figure 3, it is shown the post-digitization status of an image taken with the help of the camera.

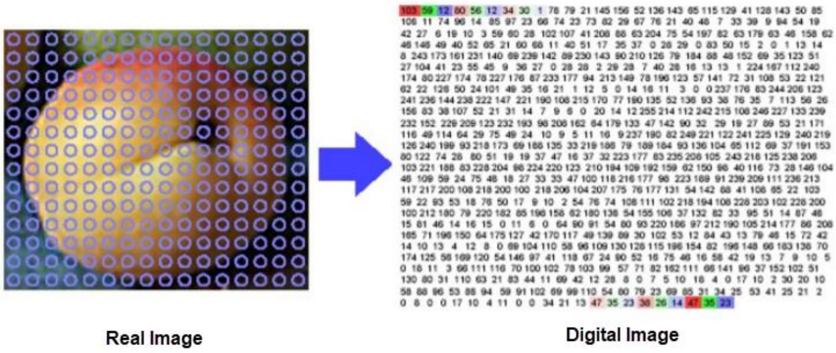


Figure 3
Real and digital image

In order to obtain images of rice varieties, the equipments given in Figure 4 was used. A camera with Ikegami brand CCD imaging sensor was used to acquire images. The camera has advanced sensitivity and high image quality. It displays at PAL resolution of 752(H) X582 (V). Features such as white balance and backlight correction are available. It is powered by 12V DC voltage and has a power consumption below 4.5W (Ikegami 2019).

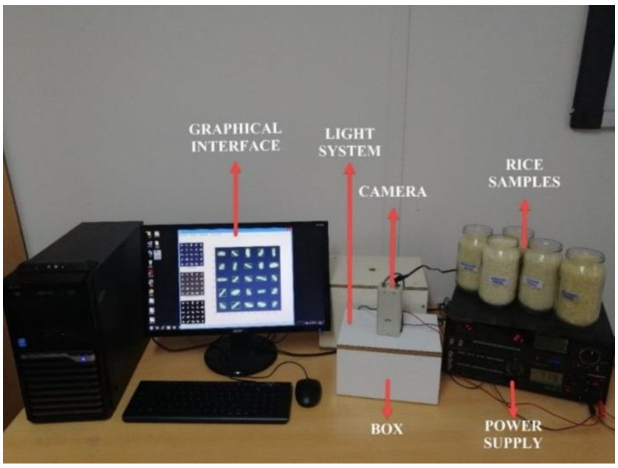


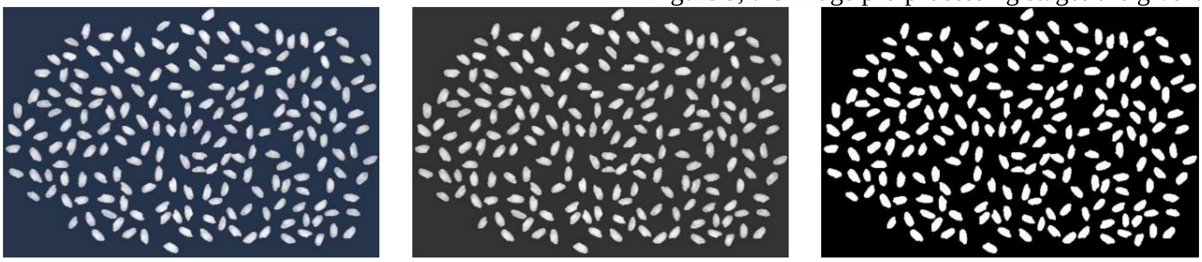
Figure 4
Equipment used to obtain images

The camera used in the study was placed on a closed box with a lighting mechanism in it and a structure to prevent taking light from the outside environment. The box floor background is chosen black for easy processing of the image. The box dimensions are designed so that an image can be taken from an area of 14 cm wide and 18 cm length. The height of the camera to the box floor is set to 15 cm. The images obtained from the camera were transferred to the computer and recorded (Cinar 2019).

2.2. Image Processing

Image processing is the process of transferring digitally acquired images to the computer environment and processing them and then transmitting them to the output unit (Kwan, Mora et al. 1999). In the image processing phase, pre-processes related to images are explained in order to perform feature extraction operations in the most accurate way.

Image processing was carried out with the help of MATLAB software. Images taken from the camera were primarily converted to grayscale images. Later, with the help of otsu method, the grayscale image was converted to a binary image using the global threshold level (Otsu 1979). Unwanted objects in the obtained binary images have been removed and prepared for the feature extraction stage by applying imopen process. In Figure 5, the image pre-processing stages are given.



a b c

Figure 5
Image pre-processing stages ((a) Color Image (b) Grayscale image (c) Post-preprocessing binary image)

2.3. Feature Extraction

Feature extraction is the process of obtaining characteristics such as shape, texture, color and contrast from images as numerical information (Shree and Kumar 2018). In the study, 12 morphological features were extracted and 4 shape features obtained by using these morphological features were extracted. In addition color images, from RGB (red, green, blue) color space to HSV (hue, saturation, value), L*a*b* (L*: lightness, a*: red/green value, b*: blue/yellow value.), YCbCr (y: luminance, cb: chroma blue, cr: chroma red) and XYZ color spaces by performing conversion operations, a total of 90 color features were extracted from five different color spaces (Cinar 2019). In Figure 6, the process stages of feature extraction are given.

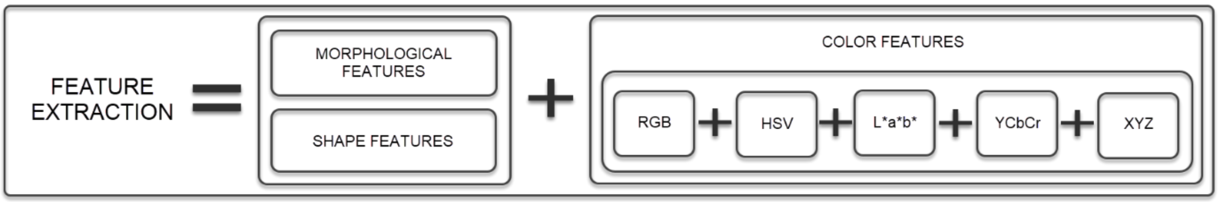


Figure 6
Process stages of feature extraction

2.3.1. Morphological and shape features

Morphological and shape features have been obtained using MATLAB regionprops function components (Buksh, Routh et al. 2014). The feature values obtained represent the number of pixels of each grain of rice. In Table 1, A list of morphological and shape features is given.

Table 1
List of morphological and shape features

Morphological Features						Shape Features	
1	Area	5	Eccentricity	9	Extent	1	Shape_Factor_1
2	Perimeter	6	Equivalent Diameter	10	Aspect Ratio	2	Shape_Factor_2
3	Major Axis Lenght	7	Solidity	11	Roundness	3	Shape_Factor_3
4	Minor Axis Lenght	8	Convex Area	12	Compactness	4	Shape_Factor_4

Explanations of morphological and shape features components are given below (Pazoki, Farokhi et al. 2014);

Area (A): It is the number of pixels within the boundaries of the rice grain area.

Perimeter (P): The rice grain gives the perimeter boundary length of.

Major Axis Lenght (L): It is the longest line that can be drawn on a grain of rice.

Minor Axis Lenght (l): It is the longest line on a grain of rice that can be drawn perpendicular to the major axis.

Eccentricity (Ec): Gives the eccentricity of the circle, which has the same moments as the region.

Equivalent Diameter (ED): It is the diameter of a circle with the same area as the area of the rice grain. The equivalent diameter was calculated according to Equation 1.

$$ED=\sqrt{\frac{4xA}{\pi}} \tag{1}$$

Solidity (S): It is the ratio of pixels in the convex stem to pixels in the rice grain region. Calculated according to Equation 2.

$$S=\frac{A}{CA} \tag{2}$$

Convex Area (CA): It is the number of pixels in the smallest convex Polygon that can accommodate the rice grain area.

Extent (Ex): Bounding is the ratio of pixels in the box to pixels in the rice grain region.

Aspect Ratio (AR): It is calculated by dividing the major axis length by the minor axis length. Calculated according to Equation 3.

$$AR=\frac{L}{l} \tag{3}$$

Roundness (Ro): It is calculated by making use of the area and the perimeter. The Roundness was calculated according to Equation 4.

$$Ro=\frac{4xAx\pi}{P^2} \tag{4}$$

Compactness (Co): It is calculated by dividing the equivalent diameter by the length of the major axis. Calculated according to Equation 5.

$$Co=\frac{ED}{L} \tag{5}$$

Shape Factor: Shape features are calculated using area, major axis and minor axis lengths from morphological features. Calculation formulas for shape factors are given below (Pazoki, Farokhi et al. 2014, Martínez, Gila et al. 2018).

Shape_Factor_1 (SF1): It is calculated by dividing the major axis length by the area. The calculation was made according to Equation 6.

$$SF1=\frac{L}{A} \tag{6}$$

Shape_Factor_2 (SF2): It is calculated by dividing the minor axis length by the area. Calculated according to Equation 7.

$$SF2=\frac{l}{A} \tag{7}$$

Shape_Factor_3 (SF3): The calculation was made according to Equation 8.

$$SF3=\frac{A}{\left(\frac{L}{2}\right)^2 x\ \pi} \tag{8}$$

Shape_Factor_4 (SF4): Calculated according to Equation 9.

$$SF4=\frac{A}{\frac{L}{2} x \frac{l}{2} x\ \pi} \tag{9}$$

2.3.2. Color features

Color images of rice grains used in the study were converted from RGB color space to HSV, L*a*b*, YCbCr and XYZ color spaces with the help of MATLAB software. Information about color spaces and conversion formulas is given below.

Color spaces: It is a mathematical representation used in defining color. Some color fields are formulated to allow people to choose colors, while others are formulated to facilitate the processing of data on machines. Color spaces are designed in three dimensions. Each pixel in the color image consists of 3 color channels (Wu and Sun 2013).

RGB Color space: RGB represents the primary colors red (R), green (G), and blue (B). All other colors are represented by a linear combination of these three main colors (Koschan and Abidi 2008). RGB color values are expressed in a range of values from 0 to 255. It is represented by black color values (0, 0, 0) and white color values (255, 255, 255). The values on the main diagonal represent the gray color values. Since RBG is considered the basic color model for image applications, it is possible to see the image on the screen without requiring any conversion (Ibraheem, Hasan et al. 2012).

HSV Color space: It consists of three parameters: hue (H), saturation (S), and value (V). The hue refers to the dominant wavelength of the color and takes a value between 0-360 degrees. Saturation refers to the vitality of color. A high saturation value causes colors to be vivid, while a low saturation value causes the color to approach tones of gray. The value refers to the brightness of the color, that is, the ratio of white in it. As the brightness value gets closer to zero, the color gets closer to black tones, and otherwise to white tones (García-Mateos, Hernández-Hernández et al. 2015).

RGB-HSV Conversion; RGB-HSV conversion formulas are given between Equation 10 and Equation 14 (Chaudhary, Chaudhari et al. 2012, Pazoki, Farokhi et al. 2014).

$$Max = Max(R,G,B) \tag{10}$$

$$Min = Min(R,G,B) \tag{11}$$

$$V = Max \tag{12}$$

$$S=\frac{Max - Min}{Max} \tag{13}$$

$$H = \begin{cases} \frac{1}{6}\frac{G - B}{Max - Min}, & V=R \\ \frac{1}{6}\frac{B - R}{Max - Min}+\frac{1}{3}, & V=G \\ \frac{1}{6}\frac{R - G}{Max - Min}+\frac{2}{3}, & V=B \\ (\text{Eğer } H<0 \rightarrow H = H+1) & \end{cases} \quad (14)$$

L*a*b* Color space: It is a color space defined by the CIE (International Commission on Illumination) and is also known as the CIELAB. It is often used for color control on objects. In this color space, colors are expressed as three values. The value L * refers to the paleness (0 Black, 100 white). a * value refers to red and green, b * value refers to yellow and blue, and axis values range from -128 to +128 (Beyaz, Ozturk et al. 2010, McGrath, Beck et al. 2017). While these values in L*a*b* space are being designed, it is designed to be perceived by the human eye (McGrath, Beck et al. 2017).

RGB-L*a*b* Conversion; RGB-L*a*b* conversion formulas are given between Equation 15 and Equation 17 (Chaudhary, Chaudhari et al. 2012, Pazoki, Farokhi et al. 2014).

$$L = 0.2126 \text{ x } R + 0.7152 \text{ x } G + 0.0722 \text{ x } B \quad (15)$$

$$A = 1.4749 \text{ x } (0.2213 \text{ x } R - 0.3390 \text{ x } G + 0.1177 \text{ x } B) + 128 \quad (16)$$

$$B = 0.6245 \text{ x } (0.1949 \text{ x } R + 0.6057 \text{ x } G - 0.8006 \text{ x } B) + 128 \quad (17)$$

YCbCr Color space: The YCbCr components are the brightness (Y), blue difference chroma (Cb), and red difference chroma (Cr) components (Ibraheem, Hasan et al. 2012). YCbCr distinguishes color space, color and brightness information. Brightness actually gives information about the amount of light on the image, and colorfulness gives information about the amount of hue ratio on the image (Chaudhary, Chaudhari et al. 2012, Pazoki, Farokhi et al. 2014).

RGB-YCbCr Conversion; RGB-YCbCr conversion formulas are given between Equation 18 and Equation 20 (Chaudhary, Chaudhari et al. 2012, Pazoki, Farokhi et al. 2014).

$$Y = 0.299 \text{ x } R + 0.587 \text{ x } G + 0.114 \text{ x } B \quad (18)$$

$$Cb = - 0.168 \text{ x } R - 0.331 \text{ x } G + 0.500 \text{ x } B \quad (19)$$

$$Cr = 0.500 \text{ x } R - 0.418 \text{ x } G - 0.081 \text{ x } B \quad (20)$$

XYZ Color space: It is the first color space defined by the CIE. Since the tristimulus values required to obtain color matching are in some cases negative, an artificial coordinate system has been developed to convert these values into positive. Thus, the tristimulus value and the Y color brightness value are chosen to be equivalent. XYZ consists of three components. X denotes red, Z denotes blue, and the Y component denotes brightness (Pratt 2001, Ibraheem, Hasan et al. 2012). The XYZ color space covers all the color values visible to humans (Mendoza, Dejmek et al. 2006).

RGB-XYZ Conversion; RGB-XYZ conversion formulas are given between Equation 21 and Equation 23 (Chaudhary, Chaudhari et al. 2012, Pazoki, Farokhi et al. 2014).

$$X = 0.4124 \text{ x } R + 0.3576 \text{ x } G + 0.1805 \text{ x } B \quad (21)$$

$$Y = 0.2126 \text{ x } R + 0.7152 \text{ x } G + 0.722 \text{ x } B \quad (22)$$

$$Z = 0.0193 \text{ x } R + 0.1192 \text{ x } G + 0.9505 \text{ x } B \quad (23)$$

After the conversion process, the features of these color spaces were replicated using the regionprops function in MATLAB with data from the MeanIntensity (mean intensity value of each an image) and PixelValue (number of pixels in the each an image region) components. Using RGB, HSV, L*a*b*, YCbCr and XYZ color spaces, a total of 90 color features were obtained with the components of average, standard deviation, skewness, roundness, entropy and wavelet decomposition for each color channel of the color features of rice images (Arefi, Motlagh et al. 2011). The list of color features is given in Table 2.

The components applied to the color features are described below, respectively (Arefi, Motlagh et al. 2011);

Mean (Me): It is the mean density value. (N variable vector, represents X input data). It is calculated according to Equation 24.

$$Me = \frac{1}{N}\sum_{i=1}^{N} Xi \quad (24)$$

Standard Deviation (SD): Returns the standard deviation of pixel values. The standard deviation is the square root of the variance (V). Calculation equations are given below.

$$V = \frac{1}{N-1}\sum_{i=1}^{N} (Xi - Me)^2 \quad (25)$$

$$SD = \sqrt{V} \quad (26)$$

Skewness (Skw): Returns the skewness value of pixel values. The calculation formula for Skewness is calculated according to Equation 27.

$$Skw = \frac{\frac{1}{N - 1}\sum_{i=1}^{N} (Xi - Me)^3}{SD^3} \quad (27)$$

Kurtosis (Ku): Returns the kurtosis value of pixel values. Kurtosis is calculated according to Equation 28.

$$Ku = \frac{\frac{1}{N - 1}\sum_{i=1}^{N} (Xi - Me)^4}{SD^4} - 3 \quad (28)$$

Entropy (E): Returns the entropy of the pixel values. Entropy is a statistical measurement used to characterize image texture. Given in Equation 29, pi represents the probability of the state i, while m represents the number of states.

$$E = - \sum_{i=1}^{m} p_i \log_2 p_i \quad (29)$$

Wavelet Decomposition: Returns the wavelet decomposition level of the matrix from pixel value using two-dimensional wavelet. Wavedec2 function was used and wavelet order db4 was selected.

Table 2
List of features obtained from color spaces

Color Space	Mean	Standard Deviation	Skewness	Kurtosis	Entropy	Wavelet Decomposition
RGB	Mean_RGB_R	StdDev_RGB_R	Skewness_RGB_R	Kurtosis_RGB_R	Entropy_RGB_R	Daub4_RGB_R
	Mean_RGB_G	StdDev_RGB_G	Skewness_RGB_G	Kurtosis_RGB_G	Entropy_RGB_G	Daub4_RGB_G
	Mean_RGB_B	StdDev_RGB_B	Skewness_RGB_B	Kurtosis_RGB_B	Entropy_RGB_B	Daub4_RGB_B
HSV	Mean_HSV_H	StdDev_HSV_H	Skewness_HSV_H	Kurtosis_HSV_H	Entropy_HSV_H	Daub4_HSV_H
	Mean_HSV_S	StdDev_HSV_S	Skewness_HSV_S	Kurtosis_HSV_S	Entropy_HSV_S	Daub4_HSV_S
	Mean_HSV_V	StdDev_HSV_V	Skewness_HSV_V	Kurtosis_HSV_V	Entropy_HSV_V	Daub4_HSV_V
L*a*b*	Mean_LAB_L	StdDev_LAB_L	Skewness_LAB_L	Kurtosis_LAB_L	Entropy_LAB_L	Daub4_LAB_L
	Mean_LAB_A	StdDev_LAB_A	Skewness_LAB_A	Kurtosis_LAB_A	Entropy_LAB_A	Daub4_LAB_A
	Mean_LAB_B	StdDev_LAB_B	Skewness_LAB_B	Kurtosis_LAB_B	Entropy_LAB_B	Daub4_LAB_B
YCbCr	Mean_ YCbCr_Y	StdDev_ YCbCr_Y	Skewness_ YCbCr_Y	Kurtosis_ YCbCr_Y	Entropy_ YCbCr_Y	Daub4_ YCbCr_Y
	Mean_YCbCr_Cb	StdDev_ YCbCr _Cb	Skewness_ YCbCr_Cb	Kurtosis_ YCbCr_Cb	Entropy_ YCbCr_Cb	Daub4_ YCbCr_Cb
	Mean_ YCbCr_Cr	StdDev_ YCbCr_Cr	Skewness_ YCbCr_Cr	Kurtosis_ YCbCr_Cr	Entropy_ YCbCr_Cr	Daub4_ YCbCr_Cr
XYZ	Mean_XYZ_X	StdDev_XYZ_X	Skewness_XYZ_X	Kurtosis_XYZ_X	Entropy_XYZ_X	Daub4_XYZ_X
	Mean_XYZ_Y	StdDev_XYZ_Y	Skewness_XYZ_Y	Kurtosis_XYZ_Y	Entropy_XYZ_Y	Daub4_XYZ_Y
	Mean_XYZ_Z	StdDev_XYZ_Z	Skewness_XYZ_Z	Kurtosis_XYZ_Z	Entropy_XYZ_Z	Daub4_XYZ_Z

2.4. Feature Selection

Feature selection is defined as the selection of the best subset that can represent the existing dataset. Feature selection is the process of selecting the best K pieces feature out of the N pieces features in the dataset by evaluating the features according to the method used (Forman 2003, Peralta, Del Río et al. 2015). In feature selection, ANOVA (Analysis of variance), X2 (Chi square) and Gain Ratio were used from commonly used tests.

ANOVA: It is a technique used to determine whether data from a different feature group has a common average. It is a feature selection technique used to determine whether differences in two or more sets of data are statistically significant (Gelman 2006, Beyaz and Ozturk 2016).

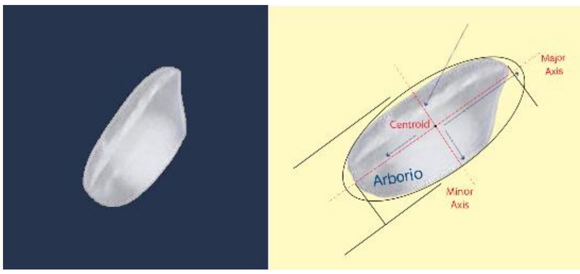
Chi square (X2): It is based on whether the difference between observed and expected frequencies is significant. It is used to test whether there is a relationship between features. As a result of the test, the features that are found to be unrelated are removed from the dataset (Liu and Setiono 1995).

Gain Ratio: The gain ratio is a non-symmetric criterion. Selects the highest rated features from among the features in the dataset with an average or better earnings (Quinlan 1986).

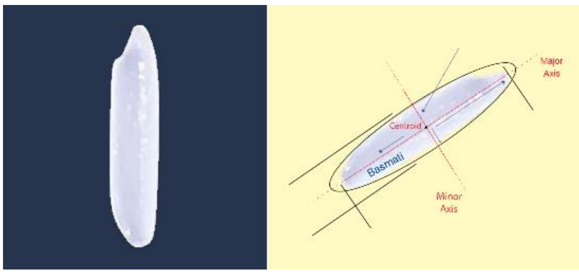
3. Results and Discussion

15 thousand pieces of rice grain image of each rice variety were obtained. In total, studies were carried out on the image of 75 thousand pieces rice grains. These images are pre-processed, free from unwanted substances that can be found on the image and prepared for feature extraction. Image examples of rice varieties are given in Figure 7.

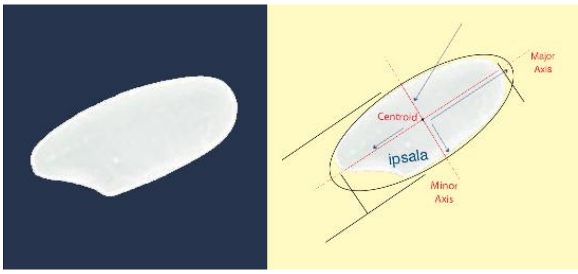
A total of 106 features, including 12 morphological features, 4 shape features and 90 color features obtained from 5 different color spaces were extracted on the pre-processed images. Morphological features have been obtained using regionprops components in MATLAB software. Shape features were also obtained using these morphological features. The statistical information of the morphological and shape features, where each rice variety is evaluated separately, is given in Table 3. The values given in Table 3 are expressed as the number of pixels. The density distribution graphs for the features used to extract morphological and shape features for rice varieties are given in Figure 8.

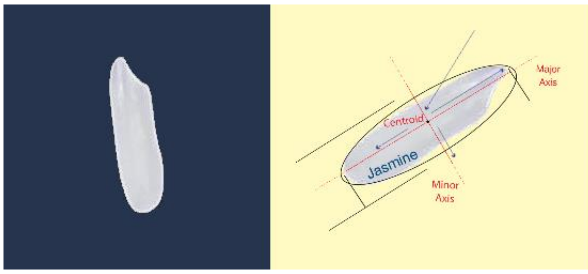


a) Arborio

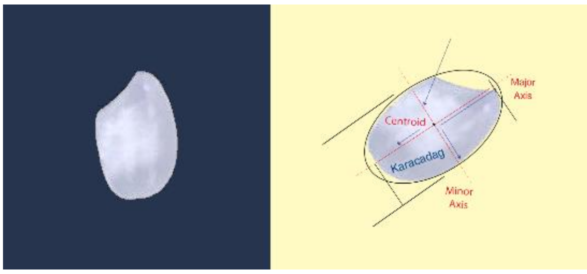


b) Basmati





c) Ipsala d) Jasmine



e) Karacadag

Figure 7
Image examples of rice varieties

When the density distribution graphs obtained for the 16 features of morphological and shape features are examined, it is observed that Ipsala variety differs from other rice varieties in area, minor axis length, equivalent diameter, convex area and shape factor 1 features. In the density distribution graphs of major axis length, roundness, compactness, shape factor 2, and shape factor 3, it is seen that the Karacadag variety differs from other varieties. When we look at the graphics of solidity, extent, shape factor 4, it is seen that all varieties are mixed with each other. In addition, when the graphics of the area, perimeter, major axis length, minor axis length, equivalent diameter, convex area, shape factor 1, and shape factor 2 features are examined, it is seen that Jasmine variety is divided into two regions in terms of these features.

RGB, HSV, L*a*b, YCbCr and XYZ color spaces were used for color features. Conversion operations to other color spaces were performed by using pixel values for each RGB image. After the color conversion process, a total of 90 color features were obtained using 5 pieces color space data, mean, standard deviation, skewness, roundness, entropy and wavelet decomposition components.

In addition to the statistical information of morphological and shape features given for each of the rice varieties used in the study, the statistical information regarding the morphological, shape, and color features for a total of 75,000 rice grains for all varieties are given in Table 4.

Table 3
Statistical information about morphological and shape features for rice varieties

Rice Variety	Statistics	Area	Perimeter	Major Axis Lenght	Minor Axis Lenght	Eccentricity	Equivalent Diameter	Solidity	Convex Area	Extent	Aspect Ratio	Roundness	Compactness	Shape Factor_1	Shape Factor_2	Shape Factor_3	Shape Factor_4
Arborio	Minimum	4,666	269.03	97.8148	45.6124	0.6639	77.0775	0.9394	4,758	0.4628	1.3373	0.637	0.5551	0.0148	0.0077	0.3081	0.9585
	Mean	7,531.72	339.8524	137.5845	70.4594	0.8574	97.7905	0.9766	7,712.89	0.6833	1.9581	0.8177	0.7114	0.0184	0.0094	0.5066	0.9876
	Maximum	10,336	395.894	166.4845	86.7822	0.9494	114.7179	0.9912	10,537	0.8482	3.1827	0.9364	0.8559	0.0284	0.0135	0.7326	0.9987
	Standard Deviation	781.1792	18.4993	8.0226	4.6157	0.0214	5.1676	0.0068	807.1189	0.0609	0.1336	0.0261	0.0238	0.0013	0.0006	0.034	0.005
	Skewness	-0.4525	-0.5754	-0.5999	-0.6583	-0.9719	-0.631	-0.8178	-0.4404	0.4171	0.4896	-0.0553	0.103	1.2194	1.102	0.3006	-0.6146
	Kurtosis	0.3685	0.5491	0.8186	0.8279	4.8166	0.6368	0.5135	0.3354	-0.8458	2.6305	0.9233	1.9298	2.7985	2.4745	2.1372	0.7341
Basmati	Minimum	4,996	315.324	136.7606	34.673	0.9173	79.7565	0.8775	5,142	0.2788	2.5114	0.3925	0.4006	0.0203	0.0051	0.1605	0.8962
	Mean	7,563.94	426.9058	202.3362	48.4943	0.9702	97.9748	0.9701	7,797.52	0.5038	4.1942	0.5216	0.4853	0.0269	0.0065	0.2361	0.9808
	Maximum	10,835	515.519	255.6472	63.7871	0.9868	117.4545	0.9907	11,198	0.8888	6.1795	0.7253	0.625	0.0369	0.0095	0.3907	0.999
	Standard Deviation	862.1709	28.0444	14.8203	3.7793	0.0064	5.6257	0.0074	891.1227	0.1438	0.4117	0.036	0.0242	0.0021	0.0005	0.0239	0.0084
	Skewness	-0.0173	-0.2519	-0.3145	-0.0928	-1.3871	-0.1713	-1.5679	-0.0211	0.8361	0.0411	0.5374	0.6129	0.5254	0.8159	0.8458	-0.9985
	Kurtosis	-0.2244	-0.1373	0.0635	0.0478	4.6278	-0.1848	6.4033	-0.2273	-0.4332	0.586	1.384	1.4635	0.4738	1.2391	2.1578	2.8281
Ipsala	Minimum	8,277	361.258	150.1548	62.105	0.7882	102.6577	0.9391	8,422	0.4811	1.6249	0.6188	0.5481	0.0113	0.0052	0.3004	0.9528
	Mean	14,048.65	476.4978	197.0714	91.8166	0.8838	133.549	0.9775	14,373.35	0.6629	2.153	0.7757	0.6781	0.0141	0.0066	0.4603	0.9868
	Maximum	21,019	593.698	247.1038	113.4411	0.9537	163.5916	0.9911	21,633	0.8485	3.3246	0.8948	0.7786	0.0205	0.0085	0.6063	0.9987
	Standard Deviation	1,493.60	25.6224	11.1519	6.2955	0.0159	7.2079	0.0066	1,531.84	0.0747	0.148	0.0256	0.0221	0.001	0.0004	0.0296	0.005
	Skewness	-0.3	-0.3858	-0.2676	-0.6003	0.082	-0.4836	-0.4647	-0.2938	0.4447	1.0685	-0.5835	-0.5966	1.096	0.7094	-0.4498	-0.3798
	Kurtosis	0.2751	0.5899	0.5783	0.5467	1.0509	0.499	-0.0272	0.2712	-0.9838	2.6398	1.5165	1.355	2.009	1.2554	1.1494	0.5116
Jasmine	Minimum	3,929	261.04	106.0622	38.8685	0.8124	70.7288	0.9047	4,032	0.3628	1.715	0.5061	0.47	0.0165	0.0053	0.2209	0.9123
	Mean	6,267.31	347.7815	157.0765	50.9507	0.945	88.5446	0.9725	6,442.76	0.5898	3.0892	0.6419	0.565	0.0259	0.0084	0.3197	0.9813
	Maximum	12,43	523.891	243.8224	79.2672	0.9749	125.8029	0.9901	12,809	0.9017	4.4926	0.8725	0.7591	0.0329	0.0121	0.5762	0.9988
	Standard Deviation	1,800.34	49.1774	22.7591	6.849	0.0093	11.8175	0.0068	1,841.54	0.121	0.2519	0.034	0.023	0.003	0.001	0.0261	0.0069
	Skewness	1.4812	1.3798	1.3403	1.3173	-0.73	1.4185	-0.974	1.481	0.5599	0.3896	0.1278	0.1113	-0.9978	-1.0536	0.2801	-0.8945
	Kurtosis	0.5651	0.4025	0.5038	0.5542	3.7978	0.3484	2.6273	0.5717	-0.9152	0.8178	0.3963	0.718	-0.1046	-0.0962	1.0934	2.2343
Karacadag	Minimum	5,014	262.372	96.9683	54.0399	0.6277	79,9001	0,9474	5,11	0.5809	1.2845	0.7521	0.6513	0.0147	0.0087	0.4242	0.9675
	Mean	6,484.38	299.8098	114.9591	72.4257	0.7739	90,7974	0,9828	6,597.79	0.7263	1.5908	0.9056	0.7906	0.0178	0.0112	0.6256	0.9912
	Maximum	9,053	354.371	146.3295	87.4875	0.9041	107,3622	0,9921	9,188	0.8301	2.3401	0.98	0.8799	0.0236	0.0133	0.7743	0.9988
	Standard Deviation	495.5896	11.8513	5.5351	3.6547	0.0324	3,464	0,0055	500.7093	0.0365	0.103	0.023	0.0249	0.0009	0.0005	0.0392	0.0045
	Skewness	0.2597	0.2572	0.2885	-0.2869	-0.2001	0,1437	-1,3887	0.2635	0.2243	0.6858	-0.4829	-0.2903	0.6736	0.0292	-0.172	-0.9543
	Kurtosis	0.1275	0.0976	0.2136	0.6037	0.4063	0,0497	1,7242	0.1269	-0.5409	1.5149	0.58	0.6061	1.3193	-0.0139	0.4605	0.7453

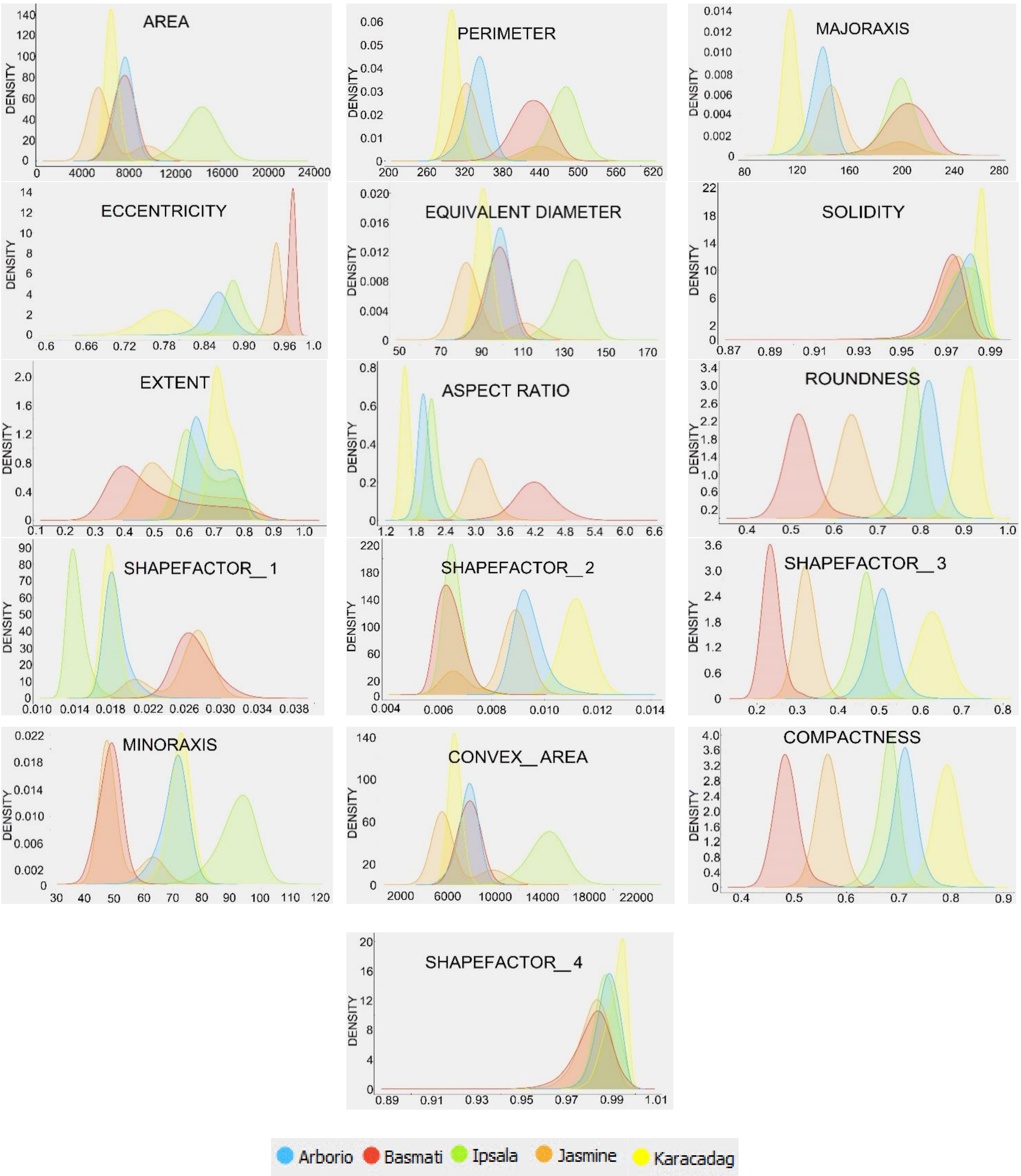


Figure 8
Density distribution graphs for morphological and shape features of rice varieties

Table 4
Statistical information on morphological, shape and color features for all varieties

	Feature	Minimum	Mean	Maximum	Standard Deviation	Skewness	Kurtosis	Feature	Minimum	Mean	Maximum	Standard Deviation	Skewness	Kurtosis
Morphological and Shape	Area	3,929	8,379.20	21,019	3,119.21	1.2189	0.2882	Extent	0.2788	0.6332	0.9017	0.1238	-0.6361	-0.4527
	Perimeter	261.04	378.1695	593.698	70.597	0.3784	-1.2991	Aspect Ratio	1.2845	2.5971	6.1795	0.969	0.8026	-0.5591
	Major Axis Lenght	96.9683	161.8055	255.6472	36.461	0.1279	-1.4311	Roundness	0.3925	0.7325	0.98	0.1386	-0.3623	-1.0926
	Minor Axis Lenght	34.673	66.8293	113.4411	16.6893	0.2691	-0.9539	Compactness	0.4006	0.6461	0.8799	0.1108	-0.1976	-1.1811
	Eccentricity	0.6277	0.8861	0.9868	0.0719	-0.5488	-0.636	Shape_Factor_1	0.0113	0.0206	0.0369	0.0053	0.3516	-1.2157
	Equivalent Diameter	70.7288	101.7313	163.5916	17.8741	1.0127	-0.0988	Shape_Factor_2	0.0051	0.0084	0.0135	0.0019	0.277	-1.2157
	Solidity	0.8775	0.9759	0.9921	0.008	-0.7854	1.8152	Shape_Factor_3	0.1605	0.4297	0.7743	0.1411	0.0087	-1.1657
	Convex Area	4,032	8,584.86	21,633	3,189.30	1.2145	0.2798	Shape_Factor_4	0.8962	0.9855	0.999	0.0073	-1.0799	2.7359
RGB	Mean_RGB_R	153.8	216.398	252.1837	13.3083	0.2342	-0.5234	Kurtosis_RGB_R	1.8413	11.9555	75.2016	7.4795	1.5759	2.9093
	Mean_RGB_G	157.2499	218.2058	252.3231	13.6464	0.2281	-0.6209	Kurtosis_RGB_G	1.8781	12.9443	89.3631	9.303	1.9358	4.5437
	Mean_RGB_B	160.1584	227.9184	252.1085	10.6825	-0.5402	0.2319	Kurtosis_RGB_B	1.8852	14.4673	71.9804	7.7546	1.299	2.6695
	StdDev_RGB_R	6.8171	15.3428	29.9674	3.4542	0.1086	-0.5677	Entropy_RGB_R	-1,36E+10	-4,43E+09	-1,47E+09	2,24E+09	-1.3145	0.3964
	StdDev_RGB_G	6.4117	15.4498	30.7654	3.5626	0.0513	-0.5205	Entropy_RGB_G	-1,38E+10	-4,51E+09	-1,55E+09	2,27E+09	-1.2476	0.2445
	StdDev_RGB_B	6.4175	15.4778	30.858	3.4686	0.0932	-0.3951	Entropy_RGB_B	-1,32E+10	-4,82E+09	-1,61E+09	1,99E+09	-1.1093	0.1689
	Skewness_RGB_R	-6.9388	-1.7785	0.9179	0.9487	-0.7312	0.4464	Daub4_RGB_R	76.8436	108.1788	126.1056	6.658	0.2351	-0.5238
	Skewness_RGB_G	-7.9118	-1.9385	0.7719	1.1119	-1.0563	1.0603	Daub4_RGB_G	78.5723	109.0821	126.1697	6.8276	0.2288	-0.6214
	Skewness_RGB_B	-6.9382	-2.3601	1.1624	0.951	-0.39	-0.0583	Daub4_RGB_B	80.0278	113.9363	126.0672	5.3433	-0.5392	0.229
HSV	Mean_HSV_H	0.0341	0.5477	0.8171	0.1856	-1.3173	0.0519	Kurtosis_HSV_H	1.0093	131.8412	5.504.56	261.9851	5.5243	46.9331
	Mean_HSV_S	0.0014	0.0606	0.2417	0.0367	0.4267	-0.7421	Kurtosis_HSV_S	1.2751	4.6014	76.7594	2.8191	4.1803	35.0858
	Mean_HSV_V	0.6281	0.8981	0.9906	0.0434	-0.5804	0.106	Kurtosis_HSV_V	1.885	15.4024	89.2129	9.05	1.5768	3.6972
	StdDev_HSV_H	0.0021	0.0642	0.4103	0.0614	2.078	5.1374	Entropy_HSV_H	262.2016	2,309.98	4,868.36	602.8766	-0.3309	0.4003
	StdDev_HSV_S	0.003	0.0191	0.0939	0.0105	1.0525	0.5996	Entropy_HSV_S	1.6196	184.7795	975.8339	159.4672	1.1109	0.9359
	StdDev_HSV_V	0.0251	0.0603	0.1184	0.0136	0.0581	-0.503	Entropy_HSV_V	137.2792	1,246.24	4,814.08	468.0433	0.662	0.5961
	Skewness_HSV_H	-70.8665	-4.7977	25.0218	7.1947	-1.3219	4.0856	Daub4_HSV_H	0.0172	0.2738	0.4087	0.0928	-1.3173	0.0518
	Skewness_HSV_S	-2.7134	0.0194	6.9277	1.0434	0.8717	0.8201	Daub4_HSV_S	0.0007	0.0303	0.1208	0.0183	0.4267	-0.7418
	Skewness_HSV_V	-7.9104	-2.4893	0.7774	1.0516	-0.5561	0.2247	Daub4_HSV_V	0.3139	0.449	0.4951	0.0217	-0.5788	0.1021

Table 4 (Continued)
Statistical information on morphological, shape and color features for all varieties

	Feature	Minimum	Mean	Maximum	Standard Deviation	Skewness	Kurtosis	Feature	Minimum	Mean	Maximum	Standard Deviation	Skewness	Kurtosis
LAB	Mean_LAB_L	164.7042	222.2155	252.5057	11.801	0.1584	-0.574	Kurtosis_LAB_L	1.9187	13.8507	89.816	9.3466	1.8296	4.0838
	Mean_LAB_A	118.2685	128.7591	134.9238	2.3522	-1.4086	3.4522	Kurtosis_LAB_A	1	4.2824	85.7886	2.3616	4.5529	65.7742
	Mean_LAB_B	107.3037	122.9208	140.5676	4.8736	0.0887	-0.8352	Kurtosis_LAB_B	0.9999	4.7306	73.882	2.9661	3.694	29.8098
	StdDev_LAB_L	5.8407	14.1272	27.4407	3.2473	0.0105	-0.5068	Entropy_LAB_L	-1,39E+10	-4,65E+09	-1,69E+09	2,25E+09	-12.554	0.2758
	StdDev_LAB_A	0.1061	0.9399	3.2582	0.396	0.7919	-0.06	Entropy_LAB_A	-3,27E+09	-1,34E+09	-6,29E+08	4,75E+08	-12.301	0.3996
	StdDev_LAB_B	0.0	2.2151	10.8211	1.327	1.0337	0.3072	Entropy_LAB_B	-3,47E+09	-1,25E+09	-5,47E+08	5,70E+08	-13.045	0.322
	Skewness_LAB_L	-7.9113	-2.0838	0.6713	1.0723	-0.9676	0.8708	Daub4_LAB_L	82.3006	111.0883	126.2651	5.9049	0.1592	-0.5748
	Skewness_LAB_A	-8.2952	0.1146	9.2085	0.9142	-0.0629	0.7586	Daub4_LAB_A	59.1379	64.3794	67.459	1.1756	-1.409	3.4536
	Skewness_LAB_B	-3.1682	0.5295	8.5405	0.9971	0.2693	0.895	Daub4_LAB_B	53.6538	61.4615	70.284	2.4356	0.0891	-0.8349
YCbCr	Mean_YCbCr_Y	150.4745	203.8867	232.55	10.8661	0.1382	-0.5441	Kurtosis_YCbCr_Y	1.8717	12.8945	83.3924	8.5364	1.7961	3.9484
	Mean_YCbCr_Cb	116.642	132.4831	146.8554	4.3203	-0.1078	-0.8464	Kurtosis_YCbCr_Cb	1	5.1287	14,559.16	58.3584	220.1377	52,636.73
	Mean_YCbCr_Cr	114.724	126.4034	133.0762	2.3508	-2.5254	7.2388	Kurtosis_YCbCr_Cr	0.9999	67.69	13,529.95	560.1777	11.444	145.0333
	StdDev_YCbCr_Y	5.5892	13.1638	25.4601	2.9794	0.0621	-0.4935	Entropy_YCbCr_Y	-1,15E+10	-3,86E+09	-1,38E+09	1,86E+09	-1.2598	0.2959
	StdDev_YCbCr_Cb	0.0	1.969	9.4741	1.1542	1.026	0.3329	Entropy_YCbCr_Cb	-3,22E+09	-1,41E+09	-6,65E+08	4,39E+08	-0.9979	0.0922
	StdDev_YCbCr_Cr	0.0	0.7544	4.1673	0.4073	1.2608	2.3222	Entropy_YCbCr_Cr	-3,29E+09	-1,30E+09	-6,00E+08	5,04E+08	-13.003	0.4479
	Skewness_YCbCr_Y	-7.5318	-1.9555	0.8661	1.0143	-0.9465	0.8552	Daub4_YCbCr_Y	75.1918	101.9254	116.2873	5.4369	0.1391	-0.5449
	Skewness_YCbCr_Cb	-7.5012	-0.4714	120.6575	1.2269	17.507	1,526.02	Daub4_YCbCr_Cb	58.3238	66.2405	73.4247	2.1591	-0.1081	-0.8461
	Skewness_YCbCr_Cr	-9.5813	1.7344	116.3182	7.8688	8.1093	7.25775	Daub4_YCbCr_Cr	57.3634	63.2021	66.5391	1.175	-2.5259	7.2413
XYZ	Mean_XYZ_X	0.3198	0.6841	0.9276	0.0838	0.1899	-0.5239	Kurtosis_XYZ_X	1.6937	8.2797	53.9629	5.2984	1.7544	3.8905
	Mean_XYZ_Y	0.3384	0.7144	0.977	0.0942	0.293	-0.5787	Kurtosis_XYZ_Y	1.698	8.3067	58.7763	5.7698	1.9101	4.5628
	Mean_XYZ_Z	0.3842	0.8427	1.0613	0.0863	-0.3714	-0.1616	Kurtosis_XYZ_Z	1.691	9.274	49.4251	5.1514	1.436	3.3613
	StdDev_XYZ_X	0.0483	0.0979	0.1949	0.0213	0.3779	-0.3846	Entropy_XYZ_X	764.2551	2,588.90	6,322.97	743.4779	1.309	1.3118
	StdDev_XYZ_Y	0.0497	0.1033	0.2119	0.0231	0.3855	-0.4311	Entropy_XYZ_Y	343.7069	2,367.18	5,835.09	596.1722	1.3253	2.5313
	StdDev_XYZ_Z	0.0533	0.1163	0.2262	0.0252	0.2634	-0.3409	Entropy_XYZ_Z	-2,074.59	1,489.69	5,615.51	828.0021	0.3848	1.1058
	Skewness_XYZ_X	-5.7217	-1.1577	1.6876	0.8228	-0.9473	1.1369	Daub4_XYZ_X	0.1597	0.3419	0.4639	0.0419	0.1908	-0.5244
	Skewness_XYZ_Y	-6.1703	-1.1312	1.6338	0.9002	-1.0683	1.3254	Daub4_XYZ_Y	0.169	0.3571	0.4886	0.0471	0.2937	-0.579
	Skewness_XYZ_Z	-5.9538	-1.504	1.8644	0.8295	-0.6169	0.57	Daub4_XYZ_Z	0.1918	0.4212	0.5302	0.0431	-0.3704	-0.1635

Finally, among a total of 106 features of morphological, shape, and color features, the importance order of the effective features was determined by selecting the features with ANOVA, X2, and Gain Ratio tests. The importance order of the effective features obtained from the tests is given in Table 5.

Table 5
Importance order of effective features obtained by ANOVA, X2 and Gain Ratio tests

Features	ANOVA	X^2	Gain Ratio	Features	ANOVA	X^2	Gain Ratio	Features	ANOVA	X^2	Gain Ratio	Features	ANOVA	X^2	Gain Ratio
Roundness	1	1	1	Entropy_ YCbCr_Cb	28	54	25	Daub4_RGB_G	55	57	50	Extent	82	92	72
Compactness	2	2	2	StdDev_HSV_V	29	15	27	Daub4_ YCbCr_Y	56	61	56	Entropy_XYZ_Y	83	93	78
Shape_Factor_3	3	3	3	StdDev_LAB_L	30	19	30	Mean_RGB_G	57	58	51	Skewness_XYZ_X	84	96	96
Aspect Ratio	4	4	4	StdDev_ YCbCr_Y	31	21	32	Mean_ YCbCr_Y	58	62	57	Skewness_ YCbCr_Cb	85	75	70
Eccentricity	5	5	5	StdDev_RGB_G	32	22	33	Daub4_XYZ_X	59	65	60	Daub4_XYZ_Z	86	84	90
Minor Axis Lenght	6	7	9	StdDev_RGB_B	33	20	34	Mean_XYZ_X	60	66	61	Mean_XYZ_Z	87	86	91
Entropy_LAB_B	7	31	13	StdDev_RGB_R	34	16	31	Kurtosis_HSV_V	61	25	42	Daub4_RGB_B	88	87	92
Entropy_RGB_R	8	30	24	StdDev_LAB_B	35	29	40	StdDev_XYZ_X	62	47	63	Mean_RGB_B	89	88	93
Entropy_ YCbCr_Cr	9	24	23	Entropy_XYZ_X	36	68	38	Skewness_RGB_G	63	69	75	Entropy_XYZ_Z	90	95	94
Shape_Factor_2	10	9	7	Kurtosis_RGB_R	37	35	55	StdDev_XYZ_Y	64	49	64	Solidity	91	97	100
Daub4_HSV_H	11	71	66	StdDev_ YCbCr _Cb	38	28	37	Skewness_LAB_L	65	72	76	Shape_Factor_4	92	99	102
Mean_HSV_H	12	70	65	Daub4_RGB_R	39	51	45	Skewness_LAB_A	66	56	69	Mean_ YCbCr_Cr	93	81	84
Shape_Factor_1	13	6	8	Mean_RGB_R	40	52	47	Skewness_RGB_R	67	80	79	Daub4_ YCbCr_Cr	94	82	85
Entropy_LAB_A	14	38	26	Kurtosis_RGB_G	41	26	41	Skewness_ YCbCr_Y	68	78	80	Skewness_XYZ_Z	95	94	98
Entropy_ YCbCr_Y	15	34	21	Kurtosis_LAB_L	42	32	46	Skewness_LAB_B	69	74	71	StdDev_ YCbCr_Cr	96	90	87
Entropy_LAB_L	16	36	20	Kurtosis_ YCbCr_Y	43	33	48	Skewness_HSV_V	70	67	73	Entropy_HSV_H	97	101	77
Entropy_RGB_G	17	37	19	Mean_HSV_S	44	18	36	Kurtosis_XYZ_Z	71	50	62	Skewness_HSV_H	98	102	99
Major Axis Lenght	18	8	6	Daub4_HSV_S	45	17	35	Kurtosis_RGB_B	72	53	68	Kurtosis_HSV_S	99	89	89
Area	19	45	17	Kurtosis_XYZ_Y	46	23	39	Skewness_HSV_S	73	73	74	Kurtosis_LAB_B	100	91	97
Convex Area	20	42	16	Kurtosis_XYZ_X	47	27	43	Skewness_XYZ_Y	74	85	88	StdDev_HSV_H	101	100	95
Equivalent Diameter	21	44	18	StdDev_HSV_S	48	48	44	Daub4_HSV_V	75	76	81	Kurtosis_LAB_A	102	103	103
Perimeter	22	12	10	StdDev_LAB_A	49	43	49	Mean_HSV_V	76	77	82	Kurtosis_HSV_H	103	104	105
Daub4_LAB_B	23	14	15	Daub4_LAB_L	50	59	52	Entropy_HSV_S	77	55	67	Skewness_YCbCr_Cr	104	105	104
Mean_LAB_B	24	13	14	Mean_LAB_L	51	60	53	Entropy_HSV_V	78	83	86	Kurtosis_YCbCr_Cr	105	106	106
Daub4_ YCbCr_Cb	25	11	12	Daub4_XYZ_Y	52	63	58	Mean_LAB_A	79	39	28	Kurtosis_YCbCr_Cb	106	98	101
Mean_YCbCr_Cb	26	10	11	Mean_XYZ_Y	53	64	59	Daub4_LAB_A	80	40	29				
Entropy_RGB_B	27	41	22	StdDev_XYZ_Z	54	46	54	Skewness_RGB_B	81	79	83				

The results from the ANOVA test show that morphological and shape features are the majority among the 10 most effective features. The color features contained within the first 30 features are components derived from L*a*b*, HSV, RGB, and YCbCr color spaces. It is also observed that components derived using the XYZ color space do not exist among the first 35 effective features.

In the X2 test, it is observed that all but one of the 10 most effective features are morphological and shape features. The color features contained within the first 30 features are components derived from L*a*b*, HSV, RGB, and YCbCr color spaces. Only two of the components derived using the XYZ color space are among the first 30 effective features.

When looking at the Gain Ratio test results, it is seen that all 10 most effective features consist of morphological and shape features. The color features contained within the first 30 features are components derived from L*a*b*, HSV, RGB, and YCbCr color spaces. Components derived using the XYZ color space are not included among the first 37 effective features.

Finally, when the ANOVA, X2 and Gain Ratio test results are evaluated together, it is observed that the ranking has not changed in terms of the top 5 effective features. In all tests, roundness, compactness, shape factor 3, aspect ratio and eccentricity features were obtained as the 5 most effective features.

4. Conclusions

A database can be created for the features of using the rice varieties mentioned in the paper. This database can be made available to the relevant sector in the field of agriculture. In addition, information such as the determination of rice varieties, morphological features etc. can be accessed instantly. Furthermore, by increasing the number of rice varieties, database can be extended.

Using the data obtained, an automated and moving image-capturing system can be designed to distinguish rice species, and a machine can be designed to perform operations such as calibration or separation of undesirable substances from varieties.

With the 106 features used in the study, feature extractions can be performed on other varieties of rice. Automatic classification processes of varieties can be performed by using these features. In addition, the effects of effective features on the success of classification can be compared with the ANOVA, X2 and Gain Ratio tests, where effective features are determined.

Classification studies can be done with artificial intelligence methods and new algorithms. The data obtained can be applied to other agricultural products that have not been studied in the literature. Furthermore, by increasing the number of features, a greater number of feature extractions can be realized.

5. Acknowledgements

We would like to thank Selcuk University Scientific Research Coordinator for their support.

This research was produced from Ilkay CINAR's published Master thesis (Cinar 2019).

6. References

Aggarwal AK and Mohan R (2010). "Aspect ratio analysis using image processing for rice grain quality." International Journal of Food Engineering 6(5).DOI: 10.2202/1556-3758.1788.

Arefi A, Motlagh AM and Teimourlou RF (2011). "Wheat class identification using computer vision system and artificial neural networks." International Agrophysics 25(4): 319-325.

Beyaz A. and Ozturk R (2016). "Identification of olive cultivars using image processing techniques." Turkish Journal of Agriculture and Forestry 40(5): 671-683.DOI: 10.3906/tar-1504-95.

Beyaz A, Ozturk R and Turker U (2010). "Assessment of mechanical damage on apples with image analysis." Journal: Food, Agriculture & Environment (JFAE) 8(3&4): 476-480.

Buksh R, Routh S, Mitra P, Banik S, Mallik A and Gupta SD (2014). "MATLAB based image editing and color detection." International Journal of Scientific Research Publications 4(1): 1-6.

Chaudhary P, Chaudhari AK, Cheeran A and Godara S (2012). "Color transform based approach for disease spot detection on plant leaf." International journal of computer science telecommunications 3(6): 65-70.

Cinar I. (2019). Classification of rice varieties using artificial intelligence techniques. Master Master Selcuk University.

Demirbas H and Dursun I (2007). "Determination of some physical properties of wheat grains by using image analysis." Journal of Agricultural Sciences.

Farahani L (2012). "Discrimination of some cultivars of durum wheat (Triticum durum Desf.) using image analysis." International Research Journal of Applied Basic Sciences 3(7): 1375-1380.

Forman G (2003). "An extensive empirical study of feature selection metrics for text classification." Journal of machine learning research 3(Mar): 1289-1305.

García-Mateos G, Hernández-Hernández J, Escarabajal-Henarejos D, Jaén-Terrones S band Molina-Martínez J (2015). "Study and comparison of color models for automatic image analysis in irrigation management applications." Agricultural Water Management 151: 158-166.DOI: 10.1016/j.agwat.2014.08.010.

Gelman A (2006). "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)." Bayesian analysis 1(3): 515-534.

Hua S, Xu M, Xu Z and Ye H (2021). "Quantitative Evaluation of Leaf Morphology with Different Rice Genotypes Based on Image Processing." Mathematical Problems in Engineering 2021.DOI: 10.1155/2021/6620636.

Ibraheem NA, Hasan MM, Khan RZ and Mishra PK (2012). "Understanding color models: a review." ARPN Journal of science technology 2(3): 265-275.

Ikegami (2019). Retrieved 14 March 2019, from https://www.ikegami.com.

Juliano BO (1993). Rice in human nutrition, Int. Rice Res. Inst.

Kaur S and Singh D (2015). "Geometric feature extraction of selected rice grains using image processing techniques." International Journal of Computer Applications 124(8).

Koschan A and Abidi M (2008). Color spaces and color distances. Digital Color Image Processing, Wiley. 1: 37-44.

Kwan AK, Mora C and Chan H (1999). "Particle shape analysis of coarse aggregate using digital image processing." Cement Concrete Research 29(9): 1403-1410.DOI: 10.1016/S0008-8846(99)00105-2.

Liu H and Setiono R (1995). Chi2: Feature selection and discretization of numeric attributes. Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence, IEEE.DOI: 10.1109/TAI.1995.479783.

Maheshwari CV, Jain KR and Modi CK (2012). "Non-destructive quality analysis of Indian Gujarat-17 Oryza sativa SSP Indica (Rice) using image processing." International Journal of Computational Engineering Science 2: 48-54.

Martínez SS, Gila DM, Beyaz A, Ortega JG and García JG (2018). "A computer vision approach based on endocarp features for the identification of olive cultivars." Computers Electronics in Agriculture 154: 341-346. DOI: 10.1016/j.compag.2018.09.017.

McGrath JR, Beck M and Hill Jr ME (2017). "Replicating Red: Analysis of ceramic slip color with CIELAB color data." Journal of Archaeological Science: Reports 14: 432-438.DOI: 10.1016/j.jasrep.2017.06.020.

Mendoza F, Dejmek P and Aguilera JM (2006). "Calibrated color measurements of agricultural foods using image analysis." Postharvest Biology Technology 41(3): 285-295. DOI: 10.1016/j.postharvbio.2006.04.004.

Otsu N (1979). "A threshold selection method from gray-level histograms." IEEE transactions on systems, man, cybernetics 9(1): 62-66.

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

Peralta D, Del Río S, Ramírez-Gallego S, Triguero I, Benitez JM and Herrera F (2015). "Evolutionary feature selection for big data classification: A mapreduce approach." Mathematical Problems in Engineering 2015.DOI: 10.1155/2015/246139.

Pratt WK (2001). "Digital Image Processing: PIKS Inside, John Wiley and Sons." Inc., New York.

Quinlan JR (1986). "Induction of decision trees." Machine learning 1(1): 81-106.DOI: 10.1007/BF00116251.

Shree NV and Kumar T (2018). "Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network." Brain informatics 5(1): 23-30.DOI: 10.1007/s40708-017-0075-5.

Szczypiński PM, Klepaczko A and Zapotoczny P (2015). "Identifying barley varieties by computer vision." Computers Electronics in Agriculture 110: 1-8.DOI: 10.1016/j.compag.2014.09.016.

Tin MM, Mon KL, Win EP and Hlaing SS (2018). Myanmar Rice Grain Classification Using Image Processing Techniques. International Conference on Big Data Analysis and Deep Learning Applications, Springer.DOI: 10.1007/978-981-13-0869-7_36.

Webb B (1991). Rice quality and grades. Rice, Springer: 508-538.DOI: 10.1007/978-1-4899-3754-4_16.

Wu D and Sun DW (2013). "Colour measurements by computer vision for food quality control–A review." Trends in Food Science & Technology 29(1): 5-20.DOI: 10.1016/j.tifs.2012.08.004.

Zhang X and Zhang F (2008). Images features extraction of tobacco leaves. 2008 Congress on Image and Signal Processing, IEEE.DOI: 10.1109/CISP.2008.88.