

UDK: 640.412

Originalni naučni rad
Original scientific paper

MACHINE VISION BASED CLASSIFICATION OF RICE (*Oryza sativa* L.) CULTIVARS USING MORPHOLOGICAL, CHROMATIC AND TEXTURAL FEATURES OF SEED IMAGES

**V. Bhushana Babu^{*}, Madhvi Tiwari, Nachiket Kotwaliwale,
Karan Singh, Rajendra Hamad**

ICAR-Central Institute of Agricultural Engineering, Bhopal, India

Abstract: Variety identification is an important task for plant breeders, farmers and traders. DUS (Distinctness, Uniformity, Stability) protocol is generally carried out for identification of plant variety which is time consuming and laborious. An attempt was made to quantify 28 rice varieties based on seed images by digital image analysis. Rice seed images were captured using Canon-LiDE110 flatbed scanner at 600 dpi resolutions. An algorithm was developed using Matlab 2012B to capture and extract seven morphological, 18 textural features and seven chromatic features. Discriminant analysis was carried out to identify critical parameters and classified them into similar groups. The study identified 14 best features out of 32 features that has capability to discriminate between rice cultivars. Eccentricity, awn length, major axis, equivalent diameter, kernel area, kernel perimeter and minor axis were found to be most critical among morphological features while standard deviation (STD) and Energy were found to be most critical among textural features while Hue, Red and Green were found to be most critical among chromatic features. Thus the present study indicated that morphological, chromatic as well as textural features play a vital role in identification of new varieties and distinguishing them to classify into similar groups.

Key words: *color features, discriminant analysis, morphological features, rice seed image analysis, textural features*

^{*}Corresponding author. E-mail: bhushanciae@gmail.com

Acknowledgements: The authors are grateful to National Agricultural Science Foundation (NASF), Indian council of Agricultural Research, India and Director, ICAR-CIAE, Bhopal for their financial and moral support.

INTRODUCTION

Rice is the most important food crop in India. It plays a vital role in primary food source for millions of people in the country. A rich and wide range of genetic wealth of rice exists in India. Further in various surveys, it is estimated that around 50,000 rice cultivars are still being grown in the country. The varietal identification is becoming a very difficult task for breeders, seed quality testing experts, farmers and traders as the number of new varieties released every year is increasing. Though crops are classified based on morphology, color and texture of various plant parts, identification and classification of varieties is difficult and complex task. Thus, identification of varieties based on morphological, chromatic and textural features is essential for improving the quality of seed production. According to the International Union for the Protection of New Varieties of Plants (UPOV), several characteristics are described for different crop varieties.

In DUS protocol, color and morphological features play a vital role in classification of cultivars. Morphological or chromatic or textural features alone may not be sufficient to identify new cultivars or help in classification of cultivars into similar groups. Thus machine vision plays a key role in using morphological, chromatic and textural features in combination to identification and classification of varieties instantly with more accuracy and at low cost by extracting more number of parameters that couldn't be visible by naked eye. The machine vision is a powerful tool of automation and they highlighted its potential for the inspection and evaluation of grain quality and food products [1]. Texture is the most efficient feature utilized in machine vision to distribute images into groups [2]. Several researchers have used flatbed scanner for classification of Indian wheat varieties using kernel and shape features. Vision system composed of two stereoscopic cameras and a matrix of laser diodes were used to distinguish fruits that grow in bunches [3]. Color and textural features were used for distinguish among quality categories of potato chips [4]. Barley kernel images were used to evaluate cereal grain quality and perform grain classification [5]. Seed shape and seed color were extracted from digital flax seed images to categorize similar flax cultivars into clusters [6]. Three grasses (wheat, ryegrass and brome grass) were classified using combination of texture, color and shape features [7]. Wheat grain kernels were classified using discriminant analysis to extract features and obtained 100 % accuracy [8]. Four paddy varieties were classified (Karjat-6, Ratnagiri-2, Ratnagiri-4 and Ratnagiri-24) using texture, shape and texture-shape features and found that 82.61%, 88% and 87.27% accuracy [9]. Engineering properties of rice namely, size, shape, thousand grains mass, aspect ratio, surface area, volume, bulk density, true density and porosity were also used for development of storage bin studies [10]. Therefore, the present study is carried out to identify best morphological, chromatic and textural features of seeds that critically contribute to classification of rice cultivars. Discriminant analysis was carried out to identify critical parameters and classified them into similar groups.

MATERIAL AND METHODS

Twenty eight rice varieties (Tab. 1) were grown at ICAR-Central Institute of Agricultural Engineering, Bhopal in RBD experimental layout to ensure authenticity of

the varieties. After harvesting, rice seeds were threshed manually. Seed samples were obtained after harvesting which were used as input for image analysis.

Table 1. Varieties of rice (*Oryza Sativa L.*) cultivars

S. No.	Variety	Code	S. No.	Variety	Code
1	Basmati 370	B370	15	PR 113	P113
2	CSR 27	CS27	16	Pusa Basmati 1	PBT1
3	Improved PBI	IPBI	17	Pusa Basmati 6	PBT6
4	Jaya	JAYA	18	PusaSugandh 2	PS02
5	Jyothi	JYTH	19	PusaSugandh 3	PS03
6	Kasturi	KSTR	20	PusaSugandh 4	PS04
7	Kranthi	KRNT	21	PusaSugandh 5	PS05
8	Makom	MKOM	22	Ravi	RAVI
9	MandyaVijaya	MDVJ	23	Tulasi	TLSI
10	NDR 359	ND59	24	Vasumati	VSMT
11	Pant Dhan 11	PD11	25	Vijetha	VJTH
12	Pant Dhan 12	PD12	26	Vikash	VKSH
13	Pant Dhan 4	PD04	27	Vikarmarya	VKMY
14	Phalguna	PHGN	28	VL Dhan 81	VL81

Image Acquisition. A flatbed scanner (LiDe110 Canon Make) with Canon solution menu GUI- based software was used for image acquisition. Seeds of each variety were scanned at 600 dpi resolutions. The images were stored in JPEG format for further analysis. All images were analyzed by Matlab 2012 B software. Seven morphological, seven chromatic and 18 textural features were extracted from segmented images. The flowchart of image acquisition and features analysis algorithm sequence for classification of rice cultivars were shown in Fig. 1.

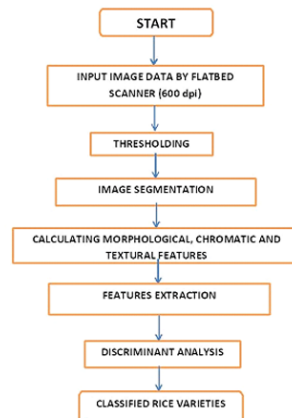


Figure 1. Flow diagram of image acquisition and procedure for classification of rice cultivars

Extraction Morphological features. Morphological features play a vital role in image segmentation. These features were extracted by developing an algorithm to

generate values corresponding to awn length, kernel area, kernel perimeter, major axis, minor axis, eccentricity and equivalent diameter.

ExtractionChromatic features. Chromatic features were extracted to generate values corresponding to red, green, blue, saturation, hue, huestd and saturation. Hue, saturation and value of an object is classified into red, blue, and green. Saturation refers to how much of the light is concentrated on its wavelength and is independent of intensity and value represents the brightness of a reflecting object.

ExtractionTextural features. Textural features are the most significant feature for distinguishing among images. The values of contrast, correlation, energy and homogeneity were extracted using grey level co-occurrence matrix (GLCM). Short run emphasis (SRE), long run emphasis (LRE), grey level non uniformity (GLN), run percentage (RP), run length non uniformity (RLN), low grey level run emphasis(LGRE), high grey level run emphasis (HGRE) were extracted and defined as:

$$\text{Contrast} = \sum_{i,j} |i - j|^2 p(i, j) \quad (1)$$

$$\text{Correlation} = \sum_{i,j} [(i - \mu_i)(j - \mu_j)p(i, j) / \sigma_i \sigma_j] \quad (2)$$

$$\text{Energy} = \sum_{i,j} p(i, j)^2 \quad (3)$$

$$\text{Homogeheity} = \sum_{i,j} [p(i, j)^2 / (1 + |i - j|)] \quad (4)$$

Where μ_i, μ_j, σ_i and σ_j are the mean and standard deviation of p_i and p_j .

$$\text{ShortRunEmphasis(SRE)} = \frac{1}{n_r} \sum_{j=1}^N \frac{p_r(j)}{j^2} \quad (5)$$

$$\text{Long Run Emphasis(LRE)} = \frac{1}{n_r} \sum_{j=1}^N p_r(j) \cdot j^2 \quad (6)$$

$$\text{Gray - Lavel Nonuniformity(GLN)} = \frac{1}{n_r} \sum_{i=1}^M p_g(i)^2 \quad (7)$$

$$\text{Run Percentage(RP)} = \frac{n_r}{n_p} \quad (8)$$

$$\text{Run Length Non - uniformity (RLN)} = \frac{1}{n_r} \sum_{j=1}^N \left(\sum_{i=1}^M p(i, j) \right)^2 \quad (9)$$

$$\text{Low Gray - levelRunEmphasis(LGRE)} = \frac{1}{n_r} \sum_{i=1}^M \frac{p_g(i)}{i^2} \quad (10)$$

$$\text{High Gray - levelRunEmphasis(HGRE)} = \frac{1}{n_r} \sum_{i=1}^M p_r(j) \cdot i^2 \quad (11)$$

where $p_{(i,j)}$ is a run length metric, i is pixel of grey level and run length j , n_r and n_p are total numbers of runs and total numbers of pixels in the image respectively.

Statistical analysis. Discriminant analysis was carried out using SAS STEPDISC procedure to reduce the number of variables. The procedure evaluated the seven morphological features, seven chromatic and 18 textural features using the stepwise test procedure for entering and removing variables from the model. The stepwise procedure begins with no entries in the model. At each step of the process, if the variable within the model, which contributes least to the models determined by the Wilk's lambda method, does not pass the test to stay, it is removed from the model. The variable outside the model which contributes most to the model and passes the test to be admitted is added. When no more steps can be taken the model is reduced to its final form.

Five reduced models were created using STEPDISC on various combination of the morphological, chromatic and textural features of seeds. The training data set consisted of 28 classes with 20 replications per class. Each of the unreduced models used different combination of the original 20 features per image. The unreduced variables consisted of seven morphological, seven chromatic and 18 textural features. The aim of the model was to minimize the computational requirement, while maintaining high classification accuracy of varieties based on these properties.

The SAS DISCRIM procedure evaluated the ability of the four models to determine classification capabilities between the rice cultivars. Four models were used from the previous STEPDISC variables reduction study, and a fifth model was added which was an unreduced model containing all 32 features. The discriminant function is established using a measure of the generalized squared distance between a specific test image texture variable input set and the class texture variable means, with an additional criteria being the posterior probability of the classification groups (SAS 9.3 version). Each test observation is placed in the class for which it has the smallest generalized square distance between the test observation and the selected class, or the largest posterior probability of being in the selected class. The DISCRIM procedure utilized a likelihood ratio test for homogeneity of the within-group covariance matrices, at 0.1 test significance level. A training data set consisting of 28 varieties and 32 properties replicated 20 times were used in analysis and a random test data set consisting of few classes with 20 replications per class were created for each of the STEPDISC reduced model described above. The training set was used to train DISCRIM function for classification of rice cultivars and the test data set was used to evaluate different model's classification accuracy.

RESULTS AND DISCUSSION

The data from the cultivars were collected and analysed. The replicated data of morphological, chromatic and texture features of all the cultivars were divided into training and test data sets. These features provided the data to generate five models viz. model1, 2, 3, 4 and 5. Stepwise discriminant analysis was carried out that selects a subset of the quantitative variables for use in discriminating among the classes.

Morphological features. The results indicated that all the seven morphological features having discriminatory power used in the classification are shown in Tab. 2. The probability of F value indicates all the parameters are significant at 1% level of

significance. None of the features were dropped in the model. The partial R square indicates the amount of variation caused by each parameter indicating the level of discriminatory power. i.e. eccentricity (0.806), awn length (0.619), major axis (0.439) and equivalent diameter (0.447) have more discriminatory power than minor axis (0.161) which had the least partial R square value.

Table 2. Summary of morphological features

Code	Features	Partial R-Square	Pr> F	Wilks' Lambda	Pr<Lambda	Average Squared Canonical Correlation	Pr> ASCC
M6	Eccentricity	0.806	<.0001	0.194	<.0001	0.03	<.0001
M1	Awn length	0.619	<.0001	0.074	<.0001	0.05	<.0001
M4	Major axis	0.439	<.0001	0.041	<.0001	0.07	<.0001
M7	Equivalent diameter	0.447	<.0001	0.023	<.0001	0.08	<.0001
M2	Kernel area	0.223	<.0001	0.010	<.0001	0.10	<.0001
M3	Kernel perimeter	0.197	<.0001	0.006	<.0001	0.11	<.0001
M5	Minor axis	0.161	<.0001	0.003	<.0001	0.13	<.0001

Chromatic features. The output of the chromatic properties indicated that all the seven chromatic features are having discriminatory power to classify rice cultivars (Table3).The probability of F value indicated that all the features are significant at 1% level of significance (<0.01) except Hue std value which was found to be non-significant, therefore, it was dropped. The partial R square values of Blue (0.450), Red (0.190) and Green (0.281) were found to be having more discriminatory power than Hue (0.063) which had the least partial R square value. The value of Wilk's Lambda (close to zero) indicates that the two groups are well separated. Wilk's Lambda indicated that all the parameters were significant inferring that all the cultivars were well separated for parameter under study.

Table 3. Summary of chromatic features

Code	Features	Partial R-Square	Pr> F	Wilks' Lambda	Pr<Lambda	Average Squared Canonical Correlation	Pr> ASCC
C3	Blue	0.450	<.0001	0.550	<.0001	0.02	<.0001
C1	Red	0.190	<.0001	0.446	<.0001	0.02	<.0001
C2	Green	0.281	<.0001	0.321	<.0001	0.03	<.0001
C5	Saturation	0.156	<.0001	0.271	<.0001	0.04	<.0001
C7	Hue Std	0.078	0.065	0.249	<.0001	0.04	<.0001
C4	Hue	0.063	<.0001	0.234	<.0001	0.04	<.0001

Textural features. The analysis indicated that as shown in Table 4, all the eighteen textural features were significant at 1% level of significance (<0.01). The partial R square indicates that Offset0 (0.625), Offset90 (0.468), Entropy (0.295), GLN (0.249), LRE (0.189) and Contrast (0.161) have more discriminatory power than Homogeneity (0.047) which had the least partial R square value. This showed that all the varieties are well separated for the parameters under study.

Table 4. Summary of textural features

Code	Features	Partial R-Square	Pr> F	Wilks' Lambda	Pr<Lambda	Average Squared Canonical Correlation	Pr> ASCC
T8	Offset0	0.625	<.0001	0.375	<.0001	0.03	<.0001
T10	Offset90	0.468	<.0001	0.200	<.0001	0.03	<.0001
T3	Entropy	0.295	<.0001	0.141	<.0001	0.04	<.0001
T14	GLN	0.249	<.0001	0.106	<.0001	0.05	<.0001
T13	LRE	0.189	<.0001	0.086	<.0001	0.06	<.0001
T5	Contrast	0.161	<.0001	0.072	<.0001	0.06	<.0001
T4	Correlation	0.120	<.0001	0.063	<.0001	0.05	<.0001
T9	Offset45	0.112	0.0013	0.056	<.0001	0.07	<.0001
T2	STD	0.089	<.0001	0.512	<.0001	0.07	<.0001
T1	Range	0.163	<.0001	0.043	<.0001	0.08	<.0001
T11	Offset135	0.096	0.0001	0.039	<.0001	0.08	<.0001
T18	HGRE	0.090	<.0001	0.035	<.0001	0.08	<.0001
T16	RLN	0.107	<.0001	0.032	<.0001	0.09	<.0001
T12	SRE	0.096	0.0024	0.028	<.0001	0.09	<.0001
T17	LGRE	0.068	0.0537	0.027	<.0001	0.09	<.0001
T15	RP	0.065	<.0001	0.025	<.0001	0.09	<.0001
T6	Energy	0.065	<.0001	0.023	<.0001	0.09	<.0001
T7	Homogeneity	0.047	0.0066	0.022	<.0001	0.10	<.0001

Five statistical models were analysed using the SAS STEPDISC procedure to determine a reduced variable set with the greatest discriminant power. The STEPDISC results for each model are shown in Tab. 5. The number of variables in the original model is dependent upon the properties considered. For instance, model1 had 32 variables in the original model. The STEPDISC selected variables are listed in the order in which they were added to the model, and are therefore, listed from the highest to the lowest significance based on partial R square. In the case of model 2, the number of textural properties selected to classify the varieties were reduced from the original 18 properties to eight properties. All the variable names are coded using a single letter and a numeric extension. Table 5 describes the interpretation of these statistics. Almost the all variables were found to be significant except value and entropy were found to be non-significant therefore, they were dropped from the model 1 for discrimination. However, based on the partial R-square value (> 0.10), the variables were arranged in the order of importance as shown below.

The training data set for each model was presented to SAS DISCRIM to train the data and then the test data set for the corresponding model was evaluated using

DISCRIM classification test. The procedure was repeated for all the five models. The results presented in Table 6 illustrate the classification accuracy of the models. Model 5 had an average classification accuracy of 80%, model 2 had shown the poorest average classification accuracy of 0 percent indicating that classification based only on textural properties has poor classification ability. The results indicated that model 5 with 14 parameters had shown better classification accuracy with 80% while model 1 with 31 parameters had shown classification accuracy of 40%, model 3 and 4 had shown an accuracy of 20% each. The models which used only one set of parameters showed a loss of classification accuracy as might be expected. This demonstrates the superior discrimination capability of mixture of parameters to either morphological or chromatic or textural features alone.

Table-5 Identified best features using Discriminant analysis by STEPDISC procedure

Model	Features	STEPDISC reduced variable list
1	All features	M6, M1, M4, M7, C3, M2, T2, M3, C4, C1, T1, M5, T6, C2, C5, T8, T3, T4, T10, T16, T5, T15, T14, T17, T13, T9, T11, T12, T7, T17, C7.
2	Texture	T8, T10, T3, T14, T13, T5, T4, T9, T2, T1, T11, T18, T16, T12, T17, T15, T6, T7.
3	Chromatic	C3, C1, C2, C5, C7, C4.
4	Morphological	M6, M1, M4, M7, M2, M3, M5.
5	Best features	M6, M1, M4, M7, C3, M2, T2, M3, C4, C1, T1, M5, T6, C2.

Table 6 PROC DISCRIM classification accuracy in percentage

Model	Features	Classification Accuracy percentage
1	All features	40
2	Texture	0
3	Chromatic	20
4	Morphological	20
5	Best features	80

CONCLUSIONS

A set of 1109 data from 28 rice cultivars based on seed images were analysed. Rice seed images were captured using flatbed scanner at 600 dpi resolutions. An algorithm was developed using Matlab2012B to capture and extract seven morphological features, 18 textural features and seven chromatic features. Discriminant analysis were carried out to identify critical parameters and classified them into similar groups. It was found that mixture of parameters classifier was capable of classifying rice varieties with high degree of accuracy (80 %) compared to only one set of parameters i.e. either morphological or chromatic or texture alone. The use of SAS procedure STEPDISC proved beneficial for reducing the number of variables in the data models. Model 1 was reduced from 32 to 14 while all the other models were unreduced. It can be concluded eccentricity, awn length, major axis, equivalent diameter, blueness, kernel area, STD, kernel perimeter, hue, red color, range, minor axis, energy and green color in this order

are the major factor that contribute to discrimination. Also morphological features along with chromatic and textural features will improve the classification accuracy.

BIBLIOGRAPHY

- [1] Patel, K.K., Kar, A., Jha, S.N., Khan, M.A. 2012. Machine vision system: a tool for quality inspection of food and agricultural products. *Journal Food Science Technology* 49(2): 123-141.
- [2] Selvarajah, S., Kodituwakku, S.R. 2011. Analysis and Comparison of Texture Features for Content Based Image Retrieval. *International Journal of Latest Trends in Computing* 2:108-113.
- [3] Tarrio, P., Bernardos, A., Casar, J., Portillo, J. 2006. A Vision System to Efficiently Harvest Fruits Which Grow in Clusters. *Agricultural Engineering (Serbia)*, 2006 (3), 43-48.
- [4] Mendoza, F., Dejmek, P., Aguilera, J.M. 2007. Colour and image texture analysis in classification of commercial potato chips. *Food Research International* 40: 1146-1154.
- [5] Szczypinski, P.M., Zapotoczny, P. 2012. Computer vision algorithm for barley kernel identification, orientation estimation and surface structure assessment. *Computers and Electronics in Agriculture* 87: 32-38.
- [6] Wiesnerova, D.W., Wiesner, I. 2008. Computer image analysis of seed shape and seed color for flax cultivar description. *Computers and Electronics in Agriculture* 61:126-135.
- [7] Golzarian, M.R., Frick, R.A. 2011. Classification of images of wheat, ryegrass and brome grass species at early growth stages using principal component analysis. *Plant Methods* 7 (28): 2-11.
- [8] Choudhary, R., Paliwal, J., Jayas, D.S. 2008. Classification of cereals grains using wavelet, morphological, colour, and textural features of non-touching kernel image. *Biosystem Engineering* 99: 330-337.
- [9] Chaugle, A., Mali, S.N. 2014. Evaluation of texture and shape features for classification of four paddy varieties. *Journal of Agricultural Engineering*, 1-8.
- [10] Ravi, P., Venkatachalam, T., Mohan, S. 2015. Engineering Properties of Rice. *Agricultural Engineering (Serbia)*, 2015 (3), 69-78.

MAŠINSKA VIZUELNA KLASIFIKACIJA SORTI PIRINČA (*Oryza sativa* L.) UPOTREBOM MORFOLOŠKIH, HROMATSKIH I TEKSTURALNIH OSOBI NA SLIKA SEMENA

V. Bhushana Babu, Madhvi Tiwari, Nachiket Kotwaliwale,
Karan Singh, Rajendra Hamad

ICAR-Centralni Institut za Poljoprivrednu tehniku, Bhopal, India

Sažetak: Identifikacija vrste je važna za odgajivače, farmere i trgovce. DUS (Rastojanja, Ujednačenost, Stabilnost) protokol je generalno izveden za identifikaciju biljne vrste koja zahteva mnogo vremena i rada. Pokušali smo da kvantifikujemo 28 sorti

pirinča na osnovu slika semena digitalnom analizom slike. Semena pirinča su snimljena skenerom Canon-LiDE110 u rezoluciji 600 dpi. Razvijen je algoritam upotrebom Matlab 2012B za hvatanje i izvođenje 7 morfoloških, 18 teksturalnih i 7 hromatskih osobina. Diskriminaciona analiza je izvedena radi identifikacije kritičnih parametara i njihove klasifikacije u slične grupe. Studija je identifikovala 14 najboljih od 32 osobine koje imaju mogućnost razlikovanja sorti pirinča. Ekscentričnost, dužina pleve, glavna osa, ekvivalentni prečnik, površina zrna, obim zrna i mala osa su definisane kao kritične među morfološkim osobinama dok su standardna devijacija i energija izdvojene kao najkritičnije među osobinama teksture. Crvena i zelena boja su bile najkritičnije među hromatskim osobinama. Tako je ova studija pokazala da morfološke, hromatske i osobine teksture igraju odlučujuću ulogu u identifikaciji novih sorti i njihovog razlikovanja radi klasifikacije u slične grupe.

Ključne reči: boja, diskriminaciona analiza, morfološke osobine, analiza slike semena, tekstura

Prijavljen: 13.12.2016.
Submitted:
Ispravljen:
Revised:
Prihvaćen: 22.07.2017.
Accepted: