

# Evaluation of Shape and Color Features for Classification of Four Paddy Varieties

**Archana A. Chaugule**

Pad. Dr. DYPIET/Computer Department, Pune, 411017, India  
Email: archna.isha@gmail.com

**Dr. Suresh N. Mali**

Pad. Dr. DYPIET/Computer Department, Pune, 411017, India  
Email: snmali@rediffmail.com

**Abstract**—This research is aimed at evaluating the shape and color features using the most commonly used neural network architectures for cereal grain classification. An evaluation of the classification accuracy of shape and color features and neural network was done to classify four Paddy (Rice) grains, viz. Karjat-6, Ratnagiri-2, Ratnagiri-4 and Ratnagiri-24. Algorithms were written to extract the features from the high-resolution images of kernels of four grain types and use them as input features for classification. Different feature models were tested for their ability to classify these cereal grains. Effect of using different parameters on the accuracy of classification was studied. The most suitable feature set from the features was identified for accurate classification. The Shape-n-Color feature set outperformed in almost all the instances of classification.

**Index Terms**—Color, Mean, Moments, Standard deviation, Variance, Shape features.

## I. INTRODUCTION

THE application of computer vision technologies along with capturing, processing and analyzing images can be effectively used in non-destructive assessment of visual quality characteristics in agricultural and food products [1]. Seed Image Analysis includes the techniques of acquisition, pre-processing and interpretation, resulting in quantification and classification of seeds. The objective of this research is to identify and classify grain kernels of four paddy types to minimize the losses incurred during harvesting, production and marketing by sowing proper type of seeds in the farm.

Grain kernels considered as agricultural objects are of variable sizes, shapes, colors and textures. The accuracy of the classifying algorithm depends on the extracted features which can be further processed to identify the class of the seed. The precision of computer vision can further be exploited to detect the seeds infected by insects or to detect damaged grain kernels. This process of perfectly classified seeds without any damage or infections is essential to increase the productivity of specific grains. Substantial work in seed shape and color

feature extraction using image processing has been carried out under the application of computer vision in agriculture.

Xiao Chena et al. [2] found a total 58 features to be extracted for identifying corn varieties, including 30 morphological features and 28 color features. Color features have been widely used to classify grain varieties. But different from most grains, colors of a corn kernel are not quite uniform. Both the germ and tipcap of corn are typically white, whether the whole kernel is white or yellow. To study the effect of color features on the identification performance of corn varieties, four transformations of RGB color space were evaluated, i.e., rgb (normalized RGB values),  $Y_CbCr$ ,  $I_1I_2I_3$ , and HSV. Furthermore, 28 features such as mean and standard deviation of these color components were calculated. From these the 18 color features (Mean and standard deviation of B, g, G, R, S, H,  $I_3$ ,  $I_2$ , r) were selected using stepwise selection for recognizing three types (white, yellow and mixed) corn kernels and finally six color features (standard deviation of g, r,  $I_1$ , mean of r, g,) were selected for recognizing the three varieties of yellow corn kernels using stepwise discrimination.

Min zhao et al. [3] extracted the 11 geometric features of corn kernels based on binary image including contour points, perimeter, area, circular degrees, equivalent diameter, major length, minor length, stretching the length of the rectangle, maximum inscribed circle, the smallest excircle also twelve features of corn seed's color mean and standard deviation of red, green and blue, the mean and standard deviation of hue, saturation and intensity from the acquired images were obtained.

S.P. Shouche et al. [4] extracted geometric features and shape related features described by moments for feature analyses of the cereal grains. Geometry related features including area, perimeter, major and minor axis lengths, compactness, axis ratio, Shape factor 1 to Shape factor 5, Spread and slenderness were measured from the binary images. Standard or raw, central, normalized central and invariant moments were computed from the digital images of each grain and mean, standard deviation (S.D.) and standard error (S.E.) were calculated. Standard moments  $m00$ ,  $m10$ ,  $m20$ ,  $m02$ ,  $m30$ ,  $m12$ ; central moments like  $\mu00$ ,  $\mu20$ ,  $\mu02$ ,  $\mu30$ ; normalized

central moments  $\mu_{20}$ ,  $\mu_{02}$ ,  $\mu_{30}$ ; and invariant moments  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$  all showed coefficients of variation below 5%. The other moments showed higher coefficients of variation. Chandra B. Singh et al. [5] extracted the morphological features from grain kernels.

Marian Wiwart et al. [6] determined the following descriptors for the image of each wheat kernel represented by a single blob (ROI – region of interest): Area, Perimeter, Circularity, Feret Diameter, Minimal Feret Diameter, Aspect Ratio, Roundness, and Solidity and performed the color analysis based on the average values of variables R G B for every ROI, which were then used to calculate the values of H S I and L a b.

Kuo-Yi Huang [7] calculated a pair of orthogonal eigenvectors of the covariance matrix. The geometric features, the principle axis length ( $L_p$ ), secondary axis ( $L_s$ ), the centroid, axis number ( $L_p/L_s$ ), area (A), perimeter (P), compactness ( $4\pi A/P^2$ ) were computed using eigenvectors for areca nuts. The color features— $R_m$ ,  $G_m$ , and  $B_m$  (i.e., the mean gray level of areca nut on the R, G, and B bands) of the entire areca nut were also computed using eigenvectors.

Li Jingbin et al. [8] extracted fourteen shape characteristic parameters of cottonseeds the Area, Perimeter, NCI ratio, Circular degree, Center of gravity X, Center of gravity Y, Major diameter, Short diameter, Second moment X ( $M_x^2$ ), Second moment Y ( $M_y^2$ ), Second moment XY ( $M_{xy}$ ), Major axis of oval, Short axis of oval, Shape coefficient of oval. The 12 color characteristic parameters, the mean and standard deviation of the color characteristics, including R (red), G (green), B (blue), H (hue), S (saturation), and I (luminance) were also defined for cotton seeds.

H.K. Mebatsion et al. [9] determined three geometric features, namely; aspect ratio (AR), major diameter (MD) and roundness (Ceq) using ellipse fitting and Green's transformation of curve integrals, respectively. The geometric characteristics of individual, non-touching grain kernels were estimated after the transformation of digital images to representative polygons defined by coordinates on the natural boundary of grain kernels. The morphological classification model was defined using SFX, AR, MD, and Ceq.

Image retrieval system proposed by Kohei Arai et al. [10] utilizes HSV-based color histogram, Zernike complex moments, and Dyadic wavelet transformation from leaf images as features.

The objective of this study is:

- (1) To extract fifty three shape-n-color features and form three different feature-sets viz. shape, color and shape-n-color feature set.
- (2) To compare the performance of three feature-sets for classification of K6, R2, R4 and R24 paddy types.

- (3) To find the most suitable feature-set from the three feature-sets for accurate classification.

This paper is organized as follows. In Section II, the proposed methodology is explained that consists of Materials and grain samples, Image capturing, Image preprocessing, Shape and Color features. Section III describes artificial neural network architecture. Section IV gives the mathematical model using set theory. Section V discusses the result. Conclusion and future work will be the end section of this paper.

## II. PROPOSED METHODOLOGY

The block diagram for the proposed methodology is as shown in Fig.1. The steps are as follows:

### A. Material and Grain samples

Sony Make 18.9 Megapixels Digital camera, Black cloth Sheet for background, Photographs of the Paddy seeds. The Seed Testing Laboratory-pune, India, provided the grain samples used in this study. The unclean commercial samples of four paddy grains, K6, R2, R4 and R24 were collected.

### B. Image capturing

The images were acquired using above specified digital camera from different distances. Randomly any numbers of seeds were placed on the black background with the kernels not touching each other. The sample images and number of seeds for the specific type is shown in the Table 1.

### C. Image pre-processing

The image analysis software was developed in Matlab version 7.12.0.635 (R2011a). In order to extract object features, image segmentation and any necessary morphological filtering was done. A total of fifty-three features which included, forty-four shape and nine color features, were extracted by the algorithm and then used for evaluating the performance and the obtained statistics were fed to ANN classifier.

### D. Feature extraction

#### 1. Shape features extraction

For a set of complete and compact descriptors to characterize the seeds, algorithms were written to extract region descriptors. Region descriptors characterize an arrangement of pixels within the area. There are actually many techniques that can be used to obtain region descriptors of an object. Here, we have extracted two forms of region descriptors: basic descriptors (geometric features) and statistical descriptors (shape related features) defined by moments.

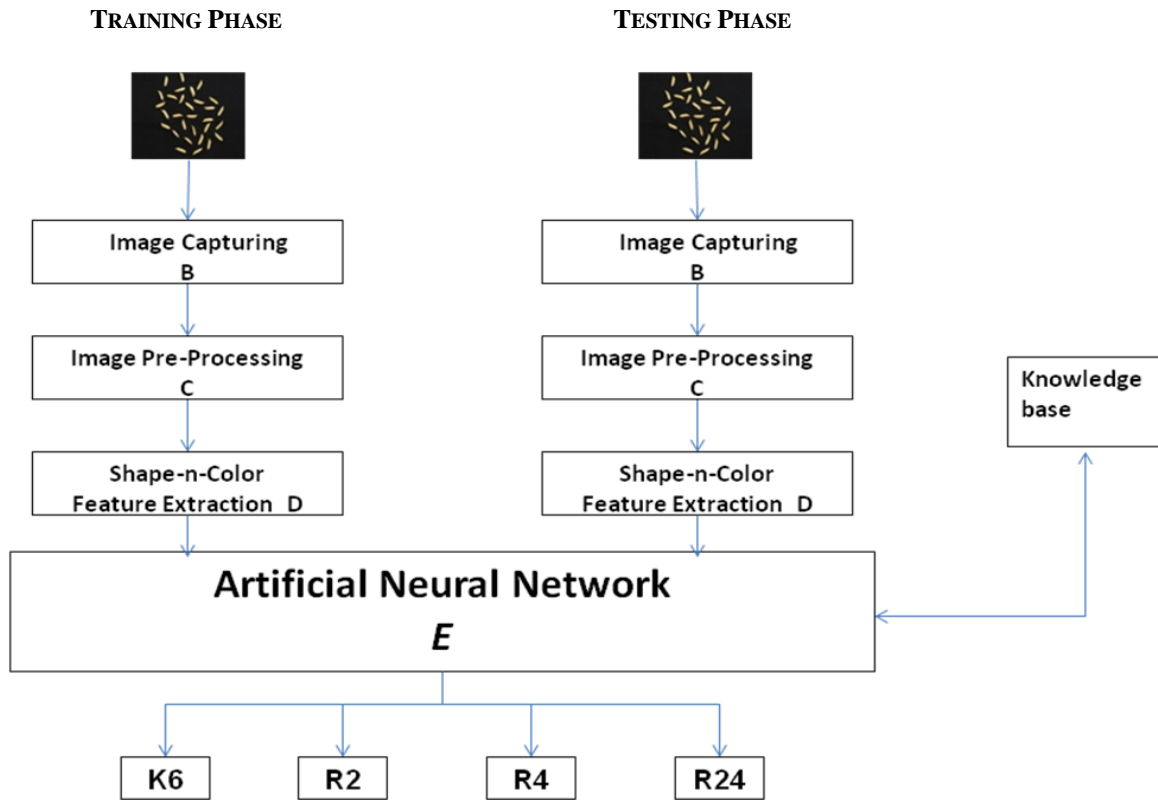






Fig.1. Block Diagram

Table 1. Type Of Different Variety Of Grains And Number Of Seeds

Type of Seed	No. of seeds	Type of Seed	No. of seeds
 K-6	3417	 R-2	1863
 R-4	1780	 R-24	1919

Feature analyses of grains include extraction of a total of eight geometric features, five shape factors, ten standard moments, seven central moments, seven invariant moments, and seven normalized central moments from the high-resolution images of kernels of paddy grains. Geometric descriptors include Area, Perimeter, Major-Axis length, Minor-axis length, Axis Ratio, Compactness. From the values of axis length, perimeter and area, shape factors (1-5) were derived, as described in [4, 11].

$$ShapeFactor1 = \frac{1}{compactness} \tag{1}$$

$$ShapeFactor2 = \frac{MAL}{Area} \tag{2}$$

$$ShapeFactor3 = \frac{Area}{MAL^3} \tag{3}$$

$$ShapeFactor4 = \frac{Area}{(MAL/2)(MAL/2)\Pi} \tag{4}$$

$$ShapeFactor5 = \frac{Area}{(MAL/2)(MiAL/2)\Pi} \tag{5}$$

Where,  
MAL: Major axis length, MiAL : Minor axis length

Moment include standard, central, normalized central and invariant moments. Moments were also used to find the spread and slenderness of individual grains according to Hu [12]. However, moments include ten standard moments  $m_{00}, m_{01}, m_{10}, m_{02}, m_{20}, m_{03}, m_{30}, m_{11}, m_{12}$  and  $m_{21}$ , seven central moments  $\mu_{11}, \mu_{02}, \mu_{20}, \mu_{03}, \mu_{30}, \mu_{12}$  and  $\mu_{21}$ , seven normalized central moments  $\eta_{11}, \eta_{02}, \eta_{20}, \eta_{03}, \eta_{30}, \eta_{12}$  and  $\eta_{21}$ , and seven invariant moments  $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6$  and  $\phi_7$ . Standard formulae from [13, 14] were used to calculate the moments. Moments describe a shape's layout (the arrangement of its pixels), a bit like combining area, compactness, irregularity and higher order descriptions together [15].

2. Color features extraction

The feature extraction algorithm extracted nine color features of individual kernels for each grain type. Color features of kernels included Mean ( $\mu$ ), Standard deviation ( $\sigma$ ), and Variance ( $\text{var}$ ), of the red (R), green (G), and blue (B) color primary values. They are calculated as follows:

1. Mean ( $\mu$ ):

Mean is the average color value in the image. Every color is defined by 3 or more values or 3 channels.

$$\mu = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m x_{ij} \tag{6}$$

Where,  $x_{ij}$ : pixel value of  $i^{\text{th}}$  row and  $j^{\text{th}}$  column and Image is of size  $m \times n$ ,

2. Standard deviation ( $\sigma$ ):

Deviation means how far from normal. Standard deviation is a measure of how spread-out numbers are.

$$\sigma = \sqrt{\text{var}} \tag{7}$$

3. Variance ( $\text{var}$ ):

It is average of squared difference from mean. It is calculated as follows:

$$\text{var} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \mu)^2 \tag{8}$$

III. ARTIFICIAL NEURAL NETWORK ARCHITECTURES

Artificial Neural Network (ANN) is the best suited classifiers for pattern recognition. They are based on the concept of biological nervous system. Pattern classification was done using a Two-layer (i.e. one-hidden-layer) back propagation supervised neural networks with a single hidden layer of 20 neurons with LM training functions. The choice of the BPN classifier was based on previous research conducted in [4].

A back propagation network (BPN) consists of an input layer, one or more hidden layers, and an output layer and has ability to generalize. The number of nodes as 20 was used to train the network. The trained neural network was tested with the testing samples to find how well the network will do when applied to data from the real world. One measure to find how well the neural network has fit the data, the confusion matrix was plotted across all samples. The fifty-three features were used as inputs to a neural network and the type of the seed as target. Given an input, which constitutes the features of a seed, the neural network is expected to identify the type of the seed which is achieved by neural network training. The ANN topology is shown in Fig.2.

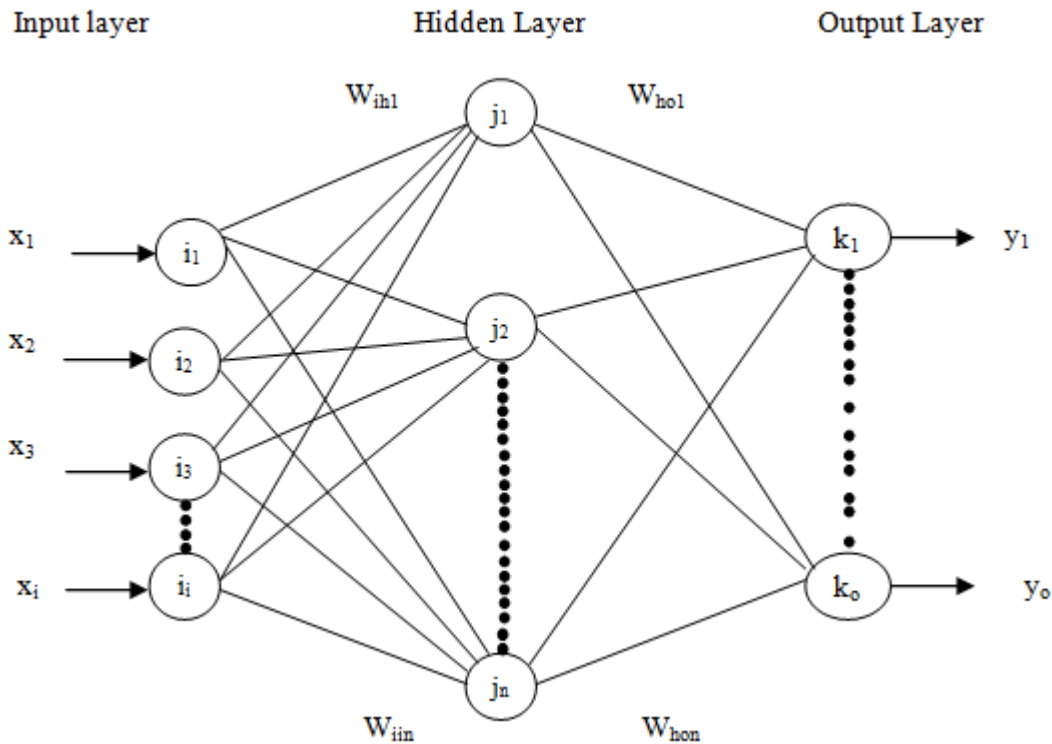


Fig.2. ANN topology,  $x_i$  is the  $x^{\text{th}}$  input feature,  $i_i$  is the  $i^{\text{th}}$  input node,  $j_j$  is the  $j^{\text{th}}$  hidden layer;  $k_k$  is the  $k^{\text{th}}$  output layer neuron and  $y_i$  is the  $y^{\text{th}}$  output,  $w_{ij}$  is weight between input and hidden and  $w_{ho}$  is weight between hidden and output

## IV. MATHEMATICAL MODEL USING SET THEORY

1. Let  $S$  be system that classifies seeds from the four varieties

$$S = \{I, O, P1, P2, F, G\} \quad (9)$$

2. Identify I/P as I

$$S = \{I\}, \text{ Where,} \\ I = \left\{ \begin{array}{l} i_1, i_2 \mid i_1 \in \text{imageroclassify,} \\ i_2 \in \text{threefeature models} \end{array} \right\} \quad (10)$$

3. Identify O/P as O

$$S = \{I, O\} \\ O = \{T \mid T \text{ is seedtype}\} \quad (11)$$

4. Identify Process as P1

$$S = \{I, O, P1\} \\ P1 = \{f1, f2, f3\} \quad (12)$$

Where,  $f1 = \{FE()\}$

Where, FE() extracts features and returns feature vector

$$f1: I \rightarrow F$$

F is the feature vector = { area, perimtr, maxlen, minlen, axisratio, compactness, spread, slender, shapefactor1, shapefactor2, shapefactor3, shapefactor4, shapefactor5, stdmom1, stdmom2, stdmom3, stdmom4, stdmom5, stdmom6, stdmom7, stdmom8, stdmom9, stdmom10, mu1, mu2, mu3, mu4, mu5, mu6, mu7, phi1, phi2, phi3, phi4, phi5, phi6, phi7, centralmom1, centralmom2, centralmom3, centralmom4, centralmom5, centralmom6, centralmom7, uncentralmom1, uncentralmom2, uncentralmom3, uncentralmom4, uncentralmom5, uncentralmom6, uncentralmom7,  $\mu_R, \mu_G, \mu_B, \sigma_R, \sigma_G, \sigma_B, \text{var}_R, \text{var}_G, \text{var}_B$  }

Three feature set models are derived from above set as follows

a. First feature set:

$Shape \subseteq F$ , which are first forty-four features

b. Second feature set:

$Color \subseteq F$ , which are next nine features

c. Third feature set:

$$Shape - n - Color = F$$

$$f2 = \{net()\}$$

Where, net() returns a pattern recognition neural network to classify inputs according to target classes

$$f3 = \{train()\}$$

Where, train() trains the neural network and returns the new network

$$f3: net, F \rightarrow N$$

Identify Process as P2

$$S = \{I, O, P1, P2\} \\ P2 = \{f4()\} \quad (13)$$

Where, f4() performs match of seed with the features from feature vector

$$f4: N \rightarrow O$$

Identify failure cases as F

$$S = \{I, O, P1, P2, F\}$$

Failure occurs when,

The training is not properly done.

The images are taken in variable illumination.

The different types of seeds have same features.

Identify success cases as G

$$S = \{I, O, P1, P2, F, G\}$$

Success is defined as for given set of images, system gives output with exact match.

## V. RESULTS AND DISCUSSIONS

The effect of shape, color and shape-n-color features on the sensitivity of class was studied using the selected ANN configuration described above. Table 2 gives the Sensitivity and Total accuracy for different feature set and different variety of grains and Fig.3. show the plot of sensitivity and Total accuracy for different variety of grains. The sensitivity and accuracy is calculated as follows.

Sensitivity:

The sensitivity for  $i^{\text{th}}$  class is defined as the number of patterns correctly predicted to be in class i with respect to the total number of patterns in class i

$$\text{Sensitivity} = \frac{n_{ii}}{f_i} \quad (14)$$

Where,

$n_{ij}$ : number of times the patterns predicted to be in class  $i$ , when they really are in same class, i.e. number of correctly classified patterns.

$f_i$  is Number of patterns associated with class  $i$  and is given by

$$f_i = \sum_{j=1}^C n_{ij}, \tag{15}$$

$n_{ij}$ : Number of times the patterns predicted to be in class  $j$ , when they really are in class  $i$ .

$C$ : Number of classes

Accuracy:

The accuracy is defined as the sum of number of correctly classified patterns upon the number of testing patterns.

$$Accuracy = (1/N) \sum_{i=1}^C n_{ii} \tag{16}$$

Where,

$C$ : number of classes, here four

$N$ : testing patterns,

$n_{ii}$ : same as in (14).

The results showed that the accuracy of classification of the ANN was best most of the time when shape-n-color features were used. This proves that color of the seed have less discriminating power than shape.

Table 2 Sensitivity for Class And Total Accuracy%

Seeds	Sensitivity for class (%)	Total accuracy%
<b>Shape Features</b>		
K6	82.0	88.00
R2	96.3	
R4	90.4	
R24	93.6	
<b>Color Features</b>		
K6	63.1	74.02
R2	81.1	
R4	90.9	
R24	72.2	
<b>Shape-n-Color Features</b>		
K6	83.2	89.00
R2	98.9	
R4	94.5	
R24	87.6	

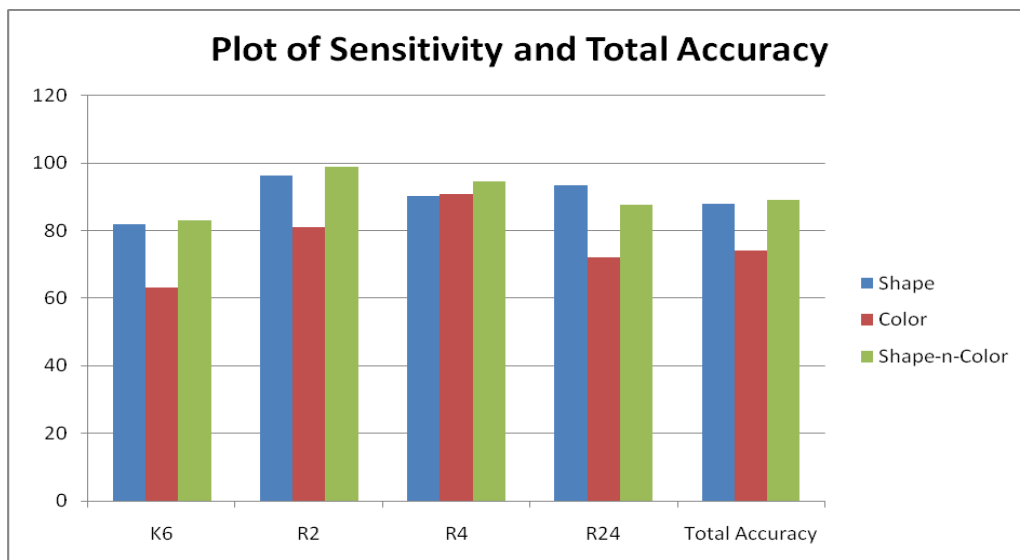


Fig.3. Plot of Sensitivity class wise

## VI. CONCLUSIONS

The shape, color and shape-n-color features were extracted from images of individual grains and the same were assessed for classification of grains. The accuracy shown is 88.00%, 74.02% and 89.00% with shape, color and shape-n-color features respectively. The most satisfactory results were delivered by the shape-n-color feature set. Only, Color feature set gave lower accuracy than all other sets because the difference between the features of different varieties is negligible. It can be concluded that invariant moments, standard moments and central moments of shape along with color mean, standard deviation and variation have a significant role in discriminating the paddy varieties.

## ACKNOWLEDGEMENT

We thank Seed Testing Laboratory, Pune for providing the help required for this study.

## REFERENCES

- [1] Archana Chaugule and Dr. Suresh Mali, "Seed Technological Development – A Survey," ACEEE, Proc. of International Conference on Information Technology in Signal and Image Processing, doi:03.LSCS.2013.6.528, pp: 71-78, 2013.
- [2] Xiao Chena, Yi Xunb, Wei Li, Junxiong Zhang, "Combining Discriminant Analysis and Neural Networks for Corn Variety Identification," Elsevier, Computers and Electronics in Agriculture 71S, pp: S48-S53, 2010.
- [3] Min zhao, Wenfu Wu, Ya qiu Zhang and Xing Li, "Combining genetic algorithm and SVM for corn variety identification," 2011 International Conference on Mechatronic Science, Electric Engineering and Computer, Jilin, China, 978-1-61284-722-1/11/\$26.00 ©2011 IEEE, pp:990-993, August 19-22, 2011.
- [4] S.P. Shouche, R. Rastogi, S.G. Bhagwat, Jayashree Krishna Sainis, "Shape analysis of grains of Indian wheat varieties," Elsevier, Computers and Electronics in Agriculture 33, pp: 55–76, 2001.
- [5] Chandra B. Singh, Digvir S. Jayas, Jitendra Paliwal, Noel D.G. White, " Identification of insect-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging," Elsevier, Computers and Electronics in Agriculture, 73, pp: 118–125, 2010.
- [6] Marian Wiwart, Elzbieta Suchowilska, Waldemar Lajszner, ukasz Graban, "Identification of Hybrids of Spelt and Wheat and their Parental Forms Using Shape and Color Descriptors," Elsevier, Computers and Electronics in Agriculture, 83, pp: 68-76, 2012.
- [7] Kuo-Yi Huang, "Detection and Classification of Areca Nuts with Machine Vision," Elsevier, Computers and Mathematics with Applications, 64, pp: 739-746, 2012.
- [8] Li Jingbin, Chen Bingqi, ,Shao Luhao,Tian Xushun, Kan Za, "Variety Identification of Delinted Cottonseeds Based on BP Neural Network," Transactions of the Chinese society of Agricultural Engineering, Vol. 28, pp: 265-269, Oct.2012.
- [9] H.K. Mebatsion, J. Paliwal, D.S. Jayas , "Automatic classification of non-touching cereal grains in digital images using limited morphological and color features," Elsevier, Computers and Electronics in Agriculture 90, pp: 99–105, 2013.
- [10] Kohei Arai, Indra Nugraha Abdullah, Hiroshi Okumura, " Image Retrieval Based on Color, Shape, and Texture for Ornamental Leaf with Medicinal Functionality Images," I.J. Image, Graphics and Signal Processing, 2014, 7, 10-18 Published Online June 2014 in MECS (<http://www.mecspress.org/>) DOI: 10.5815/ijigsp.2014.07.02.
- [11] Symons, S.J., Fulcher, R.G., "Determination of wheat kernel morphological variation by digital image analysis, I Variation in eastern Canadian milling quality wheats," J. Cereal Sci. 8, pp: 211–218, 1988.
- [12] Hu M K, "Visual pattern recognition by moment invariant," IRE Transactions of Information Theory, 8, pp: 179–187, 1962.
- [13] Gonzalez R C; Woods R E; Steven L. Eddins, "Digital Image Processing Using MATLAB," Second Edition, Tata McGraw Hill Pvt. Ltd., New Delhi, 2010.
- [14] Jain A K, Moment representation. In: Fundamentals of Digital Image Processing (Jain A K, ed), Prentice Hall of India Private Limited, New Delhi, 1995.
- [15] Mark S. Nixon, Alberto S. Aguado, Newnes, "Feature Extraction and Image Processing," Oxford Auckland Boston Johannesburg Melbourne New Delhi, 2002.

## Authors' Profiles

**Archana Chaugule** has completed her M.E. (Computer) from University of Pune, Maharashtra, India. She is presently pursuing Ph.D. She is having 16 Years of teaching experience in Computer Engineering and is having 12 international and national journals and conferences publications. She is member of ACM and life member of ISTE. Her areas of interest include Image Processing and Machine Learning.

**Dr. Suresh N. Mali** is presently working as Principal Sinhgad Institute of Technology and Science, Narhe, Pune, India. He has written 3 technical books and published 25 papers in various national and international journals and conferences. He is member of 'Board of Studies' for Computer Engineering in various universities like University of Pune, Shivaji University, Kolhapur, MS, India. He has also worked as Member of 'Local Inquiry Committee' on behalf of University of Pune. He is member of IEEE, life member of ISTE and his research interests are information security, data hiding, signal processing, digital multimedia communications and Steganography.

**How to cite this paper:** Archana A. Chaugule, Suresh N. Mali, "Evaluation of Shape and Color Features for Classification of Four Paddy Varieties", IJIGSP, vol.6, no.12, pp.32-38, 2014.DOI: 10.5815/ijigsp.2014.12.05