Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE xxx (xxxx) xxx

journal homepage: www.elsevier.com/locate/inpa

Automated recognition and classification of adulteration levels from bulk paddy grain samples

Basavaraj S. Anami [a], *Naveen N. Malvade* [b,*], *Surendra Palaiah* [c]

[a] *Department of Computer Science and Engineering, K L E Institute of Technology, Hubballi 580030, India*
[b] *Department of Information Science and Engineering, K L E Institute of Technology, Hubballi 580030, India*
[c] *Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwar, 580002, India*

A R T I C L E I N F O

Article history:
Received 8 February 2018
Received in revised form
30 August 2018
Accepted 5 September 2018
Available online xxxx

Keywords:
Bulk paddy grains
Adulteration quantification
Feature selection
GLCM
LBP
BPNN
PCA
SVM
k-NN

A B S T R A C T

Fraudulent labeling and adulteration are the major concerns in the global rice industry. Almost all the paddy varieties being sold in the market are prone to adulteration. It is very difficult to differentiate paddy grains of various varieties in the mixed bulk sample based on visual observation. Currently, there is no sophisticated appearance-based commercial scale technology to reliably detect and quantify adulteration in bulk paddy grain samples. The paper presents a cost-effective image processing technique for the recognition of adulteration and classification of adulteration levels (%) from the images of adulterated bulk paddy samples using state-of-the-art color and texture features. In this work, seven adulterated bulk paddy samples are considered and each of the samples is prepared by mixing a premium paddy variety with the identical looking and commercially inferior paddy variety at five different adulteration levels (weight ratios) of 10%, 15%, 20%, 25% and 30%. The study compares the performances of three different classification models, namely, Multilayer Back Propagation Neural Network (BPNN), Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN). The Principal Component Analysis (PCA) and Sequential Forward Floating Selection (SFFS) methods have been employed separately for the automatic selection of optimal feature subsets from the combined color and texture features. The maximum average adulteration level classification accuracy of 93.31% is obtained using the BPNN classification model trained with PCA-based reduced features. The proposed technique can be used as an economic, rapid, non-destructive and quantitative technique for testing adulteration, authenticity, and quality of bulk paddy grain samples.

1. Introduction

The paddy trade is the biggest trade in India and plays an important role in the economy. Indian rice is highly appreciated by all the export destinations for its taste and suitability. The huge demand for rice in the global market has created an excellent environment for the export of paddy and rice. In India, the export quality of paddy and rice has been moved from free to a restricted category in order to protect the grain quality and integrity. There are about thousands of predominantly grown paddy varieties around the country and out of which only a few varieties are preferred varieties for exports.

* Corresponding author.
E-mail address: naveen.malvade@gmail.com (N.N. Malvade).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2018.09.001
2214-3173

Consumer preference brings higher returns for different varieties, leading to the generation of brand equity. Therefore, exporting firms in the country would be cautious to maintain the purity of the samples and keen on protecting consumer interests.

The varietal purity is mandatory for international paddy trade. Now-a-days, the exporting firms in the country are facing major hindrance due to grain adulteration which resulted in a great economic loss. The adulteration of premium and commercially leading paddy varieties with cheaper look-alike varieties is abundantly widespread because of inter-varietal homogeneity. The images of some freshly harvested premium and local paddy varieties are shown in Fig. 1. Many rice importing countries allow certain level i.e. 5% to 8% of paddy adulteration considering non-deliberate admixture of grains during harvest and post-harvest processing. The lot exceeding this permitted limit will be considered as the deliberate adulteration. Considering the paddy trade volume and economic returns, an accurate method is needed to detect and determine the extent of adulteration. There are several methods being employed over the past decades for unambiguous detection and precise quantification of the adulterant. These methods are classified as non-DNA based and DNA based methods [15]. The non-DNA methods use morphological features, physiochemical properties and protein profile to detect and quantify adulteration. The use of DNA markers has been suggested for precise and reliable characterization and discrimination of rice varieties because they are not affected by environmental factors [1,12]. The DNA based method is the most genuine and widely accepted one due to their simplicity, accessibility, repeatability and rapidness. Though the physicochemical methods are simple and cost-effective, the results are highly influenced by the environment. The visual inspection method of bulk sample is popularly used in the grain markets, but it is prone to human sensory errors and time-consuming. Lack of an image data set for a large number of paddy varieties and the expensive protein and DNA based methods are the hindrances to the commercial scale implementation of adulteration detection and quantification technique. These limitations can be overcome by employing a digital inspection technique using computer vision. The need for computer vision systems to automate this task of inspection is felt that would benefit the potential farmers, grain inspectors and traders. To know

Fig. 1 – Images of freshly harvested premium and local paddy varieties.

the state-of-the-art in automation of such activities in the agriculture field, a survey is made and the following papers have been cited during the literature survey to understand the different applications of computer vision in allied areas of the present work carried out.

A novel computer vision based neural network approach was proposed for detecting fraudulent labels for rice samples using the fuzzy class knowledge database. A total of 34 morphological, color, and texture features were extracted from each rice grain kernel image and the STEPDISC procedure was used for optimal features selection. A precision of more than 90% was achieved in identifying fraudulent labels of rice [2]. A machine vision technique was developed to classify three Iranian rice varieties in mixed bulks of three and two varieties using multilayer perceptron (MLP) neural network. 17 morphological and 41 textural features are extracted from the singleton rice seed images and principal component analysis method was employed to select and rank the most significant features for the classification [6]. A low-cost E-nose based on metal oxide gas sensors were presented for monitoring the adulteration of uncooked Jasmine rice. The PCA with simple Euclidean plane calculation, SVM and BPNN were used separately for pattern recognition and evaluation of adulteration percentages. The BPNN classifier revealed a good actual adulteration prediction percentage of 99.30% corresponding to the PCA and SVM analysis [13]. A machine vision technique was proposed to classify wheat seeds belong to nine varieties using 131 textural features extracted from GLCM, LBP, LSN, LSP and GLRM matrices. A total of 50 significant textural fea-

Table 1 – Paddy varieties considered for the adulteration study.

Sl. No.	Premium variety	Adulterant variety
1	Jaya	Abhilasha
2	Jaya	Mugad101
3	Jaya	Thousand One
4	Mugad Siri	Thousand Ten
5	Mugad Sughand	Thousand Ten
6	PSB68	Abhilasha
7	PSB68	Budda

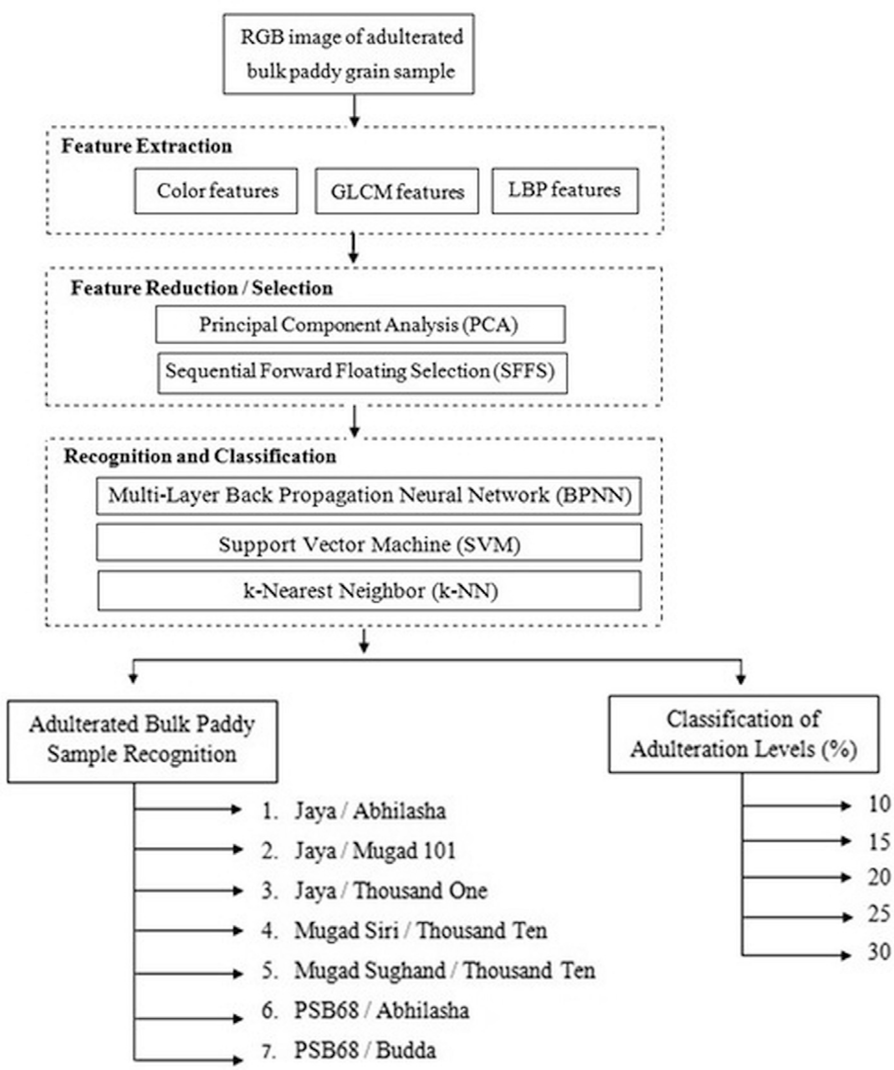

Fig. 2 – Block diagram of the proposed methodology.

Please cite this article as: B. S. Anami, N. N. Malvade and S. Palaiah, Automated recognition and classification of adulteration levels from bulk paddy grain samples, INFORMATION PROCESSING IN AGRICULTURE, https://doi.org/10.1016/j.inpa.2018.09.001

tures were selected by employing stepwise discrimination method. The average classification accuracy of 98.15% was obtained using the LDA (linear discriminate analysis) classifier [10]. A cost-effective digital image processing technique was presented for rice authenticity test based on kernel morphology and fuzzy logic. The outer characteristic of the singleton rice grain, namely the aspect ratio was used in the authenticity test [5]. The computer vision based automated tools were developed to identify seed types from their bulk samples [7,8,16].

Most of the published research works have focused on identification and quantification of adulteration by considering singleton rice grain images. Several researchers have demonstrated that the size, shape and color features extracted from the singleton rice grain images are effective in distinguishing different rice kernels [17,18,20]. The rice kernel separation process and the single-kernel-feature based classification techniques are usually too slow for practical use. In addition, most of the previous research work has been focused on identifying and classifying foreign bodies in bulk food grains using image analysis [19]. To the best of the author's knowledge, no work has been cited in the literature survey with respect to the detection and quantification of adulteration in bulk paddy grain samples using computer vision techniques. This is the motivation for the present work.

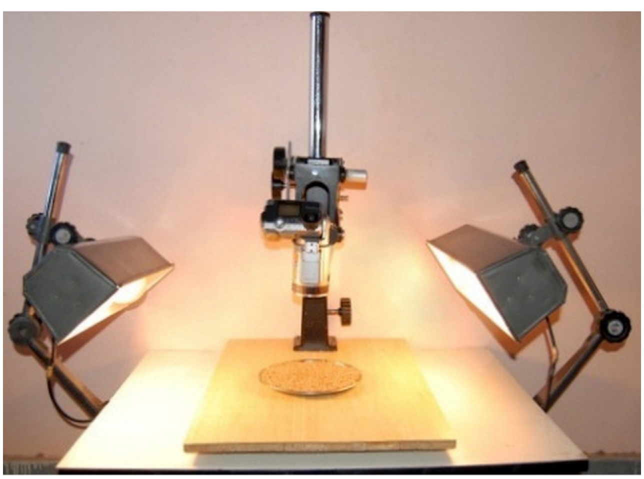

Fig. 3 – Image capturing setup.

The paper is organized into four sections. Section 2 presents the proposed methodology. Section 3 describes the classification results of using color and texture features. Section 4 gives the conclusion of the work.

2. Proposed methodology

The proposed methodology consists of five stages, namely image acquisition, features extraction, feature selection, recognition of adulterant paddy varieties and classification of adulteration levels (%). The block diagram illustrating the stages involved in the proposed methodology is shown in Fig. 2.

2.1. *Grain samples*

The present work considers freshly harvested and naturally dried 4 premium paddy varieties, namely Jaya, Mugad Siri, Mugad Sughand and PSB 68 and 5 commercially inferior paddy varieties, namely Abhilasha, Budda, Mugad 101, Thousand One and Thousand Ten. All the paddy samples are obtained from the University of Agricultural Sciences (UAS), Dharwad, Karnataka State, India. One-hundred grams of grain samples from each of the premium paddy varieties are drawn randomly from the lots and the grain samples are cleaned from impurities, damaged, discolored and immature grains. For adulteration study, the local adulterant paddy varieties are mixed with the similarly looking premium paddy varieties at five weight ratios of 10%, 15%, 20%, 25%, and 30% according to commercial practice. The list of premium paddy varieties and their corresponding adulterant paddy varieties considered in the present work are given in Table 1.

2.2. *Image acquisition*

The seven main paddy samples in mixed bulks of two varieties and five different levels of adulteration percentages are considered. A total of 35 adulterated bulk paddy samples is obtained after the adulteration procedure. From each of the adulterated samples, a total of 200 images is captured using a digital camera (PENTAX MX-1, USA) having a resolution of 12 megapixels. The images are taken keeping approximately the object distance of 0.5 m. The visible light range of wavelength 400–700 nm is used. Four circular fluorescent bulbs (100 Watts) are used to illuminate the grain samples as uni-

(a) Original image of size 1920 X 1080 pixels (b) Cropped image of size 400 X400 pixels (c) Enhanced image of size 400 X400 pixels

Fig. 4 – Image preprocessing.

Please cite this article as: B. S. Anami, N. N. Malvade and S. Palaiah, Automated recognition and classification of adulteration levels from bulk paddy grain samples, INFORMATION PROCESSING IN AGRICULTURE, https://doi.org/10.1016/j.inpa.2018.09.001

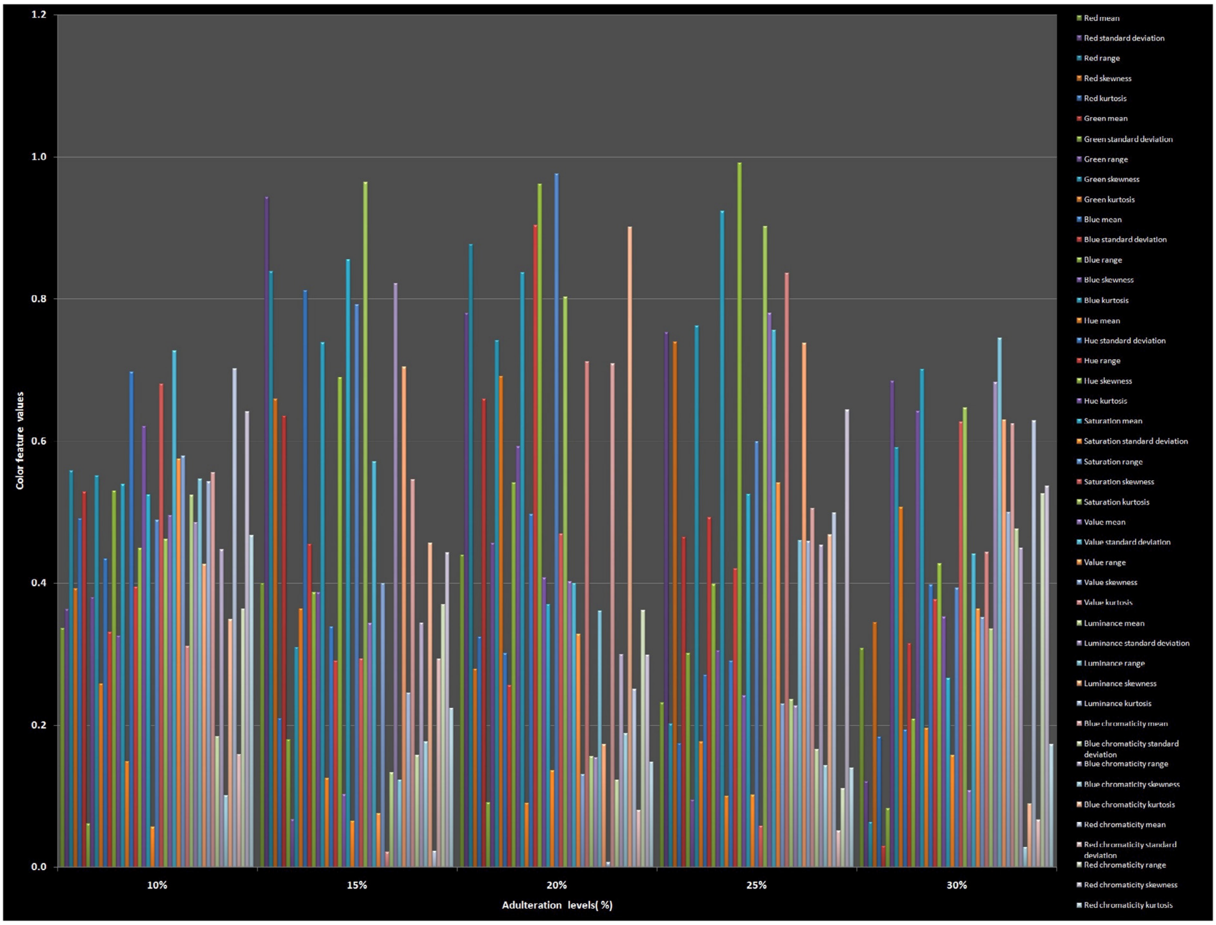

Fig. 5 – Graph showing the color feature values of the paddy variety Jaya adulterated with Abhilasha at five different adulteration levels (%).

formly as possible. The angle between the camera lens and the lighting source axis is maintained at approximately 45^0. In order to provide a rigid and stable support and easy vertical movement, the camera is mounted on a metal rod as shown in Fig. 3.

The experimental dataset contains a total of 7000 images from all the seven paddy samples in mixed bulks of two varieties at five different adulteration levels. The images are captured under standard lighting conditions. As a part of the dataset preparation, the acquired images of size 1920 × 1080 pixels are manually cropped to size 400 × 400 pixels with the assistance of grain experts. The cropped images are subjected for image preprocessing to eliminate the influence of illumination changes and noise. The removal of shading and correction of color changes in the images is accomplished through the histogram equalization technique. The preprocessed sample image is shown in Fig. 4.

2.3. *Feature extraction*

Human beings identify paddy varieties from their bulk sample by their color and textural features [3,4]. The adulteration among different paddy varieties will cause the significant change in chrominance and texture features of bulk sample because of the different shapes and colors of individual grains. Hence, the evaluation of statistical chrominance features and textural features is performed to realize the comprehensive adulteration level classification process [14]. In the present work, the feature extraction deals with the color and texture features of the images. The choice of the color models can be a very important decision which can dramatically influence the results of the classification. The color features are explored using three different color models, namely, RGB, HSI, and YC_bC_r. The RGB color model is very common and being used in almost every computer system. The HSI and YC_bC_r color models are used to eliminate the mixing of chrominance and luminance information of the RGB color model, thereby providing a way to get intensity-invariant chromaticity measures. The texture features are explored by constructing GLCM and LBP. The LBP is one of the best performing and highly discriminative texture descriptors. The computational efficiency and invariance to monotonic gray level changes have made LBP features suitable for demanding image analysis tasks [9].

2.3.1. *Color feature extraction*

The color feature extraction starts with the separation of the color channels of the work images. In this, the images in the

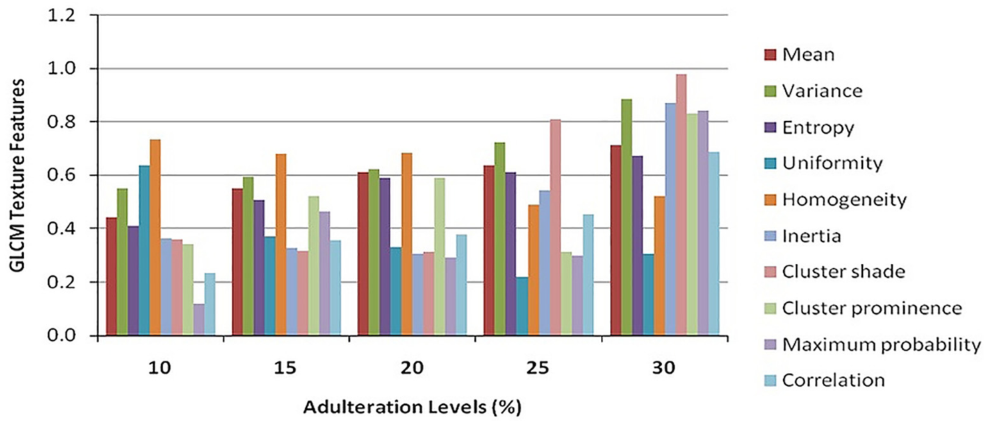

Fig. 6 – Graph showing the GLCM feature values of the paddy variety Jaya adulterated with Abhilasha at five different adulteration levels (%).

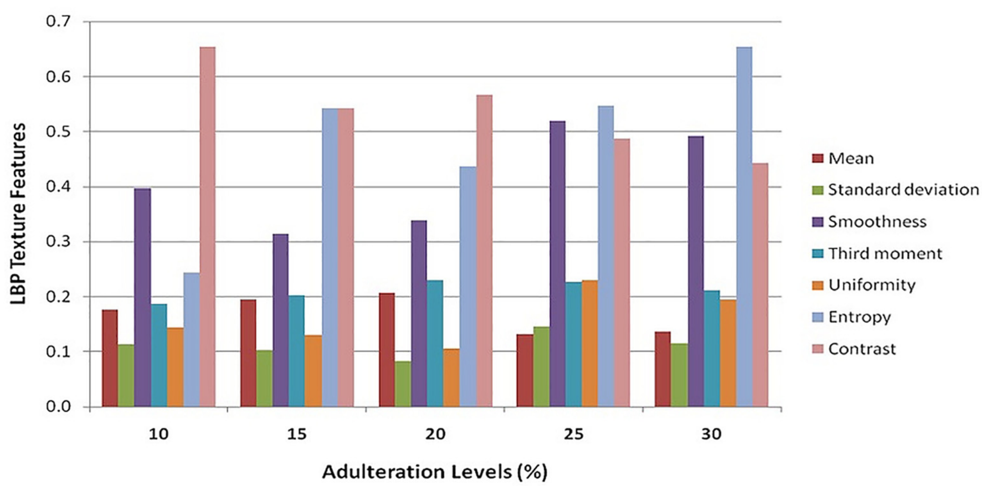

Fig. 7 – Graph showing the LBP feature values of the paddy variety Jaya adulterated with Abhilasha at five different adulteration levels (%). (HG1 – HG25 features are not shown in the graph).

RGB color model are separated into R, G, and B components respectively. The Hue (H), Saturation (S), Intensity (I), Luminance (Y), two chrominance difference channels, namely, Blue Chromaticity (C_b) and Red Chromaticity (C_r) are obtained from the R, G, and B components using Eqs. (1) through (6). From each of the nine color channels (R, G, B, H, S, I, Y, C_b, and C_r) five statistical color features namely, mean, standard deviation, range, skewness, and kurtosis are extracted using Eqs. (7) through (11). The color feature values of the variety mixture Jaya and Abhilasha at five different levels of adulteration percentages are shown graphically in Fig. 5.

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{\left[(R-G)^2+(R-B)(G-B)\right]^{\frac{1}{2}}}\right\} \quad (1)$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \quad (2)$$

$$I = \frac{1}{3}(R+G+B) \quad (3)$$

$$Y = 0.299 * R + 0.587 * G + 0.114 * B \quad (4)$$

$$C_b = B - Y \quad (5)$$

$$C_r = R - Y \quad (6)$$

Eqs. (7) through (11) are used for calculating the mean, standard deviation, range, kurtosis and skewness of all the color channels, where M, N, and P(i, j) denote the dimension of the image matrix, the total number of pixels in the image and the color value of i^{th} column and j^{th} row respectively.

$$Mean(\mu)\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}P(i,j) \quad (7)$$

$$Standard\ Deviation(\sigma) = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\{(P(i,\ j)\ -\mu)\}^2} \quad (8)$$

$$Range\ (r) = max\ \{(P(i,\ j)\} - min\ \{P(i,\ j)\} \quad (9)$$

$$Skewness(\theta) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\{(P(i,j)-\mu)\}^3}{MN\sigma^3} \quad (10)$$

Table 2 – Classification results of using separate color, GLCM and LBP texture features with BPNN, SVM and K-NN classifiers.

Sl. No.	Premium paddy variety	Adulterant paddy variety	Adulteration level (%)	Average adulteration level classification accuracy (%) of classifiers								
				BPNN			SVM			k-NN (For k = 3)		
				Color	GLCM	LBP	Color	GLCM	LBP	Color	GLCM	LBP
1	Jaya	Abhilasha	**10**	39	45	35	30	39	45	37	39	44
			15	39	48	38	29	44	36	45	34	30
			20	50	49	37	41	39	38	37	38	29
			25	43	41	41	38	36	36	33	25	29
			30	43	44	37	33	26	46	26	31	26
2	Jaya	Mugad 101	**10**	35	46	36	41	31	27	27	35	37
			15	50	49	50	32	46	32	40	31	29
			20	43	51	45	27	29	36	42	26	28
			25	31	42	39	41	40	30	41	36	29
			30	42	47	35	38	44	27	33	38	43
3	Jaya	Thousand One	**10**	43	42	41	42	38	26	34	27	39
			15	45	45	49	39	44	35	29	38	36
			20	30	50	30	31	41	38	29	31	34
			25	50	52	49	42	39	42	44	29	27
			30	42	49	31	37	33	36	32	46	27
4	Mugad Siri	Thousand Ten	**10**	50	44	47	38	33	42	43	26	26
			15	47	40	43	38	45	45	31	38	44
			20	35	39	48	29	33	25	37	37	31
			25	39	43	37	30	33	38	40	44	41
			30	42	46	42	44	45	28	41	28	42
5	Mugad Sughand	Thousand Ten	**10**	49	43	30	35	29	45	38	34	31
			15	32	48	35	27	38	25	31	44	30
			20	47	47	31	41	37	46	44	34	46
			25	32	40	31	31	31	30	42	30	43
			30	39	49	43	44	30	25	46	42	43
6	PSB68	Abhilasha	**10**	46	38	37	31	34	42	40	33	34
			15	39	40	32	33	29	26	31	44	33
			20	39	42	31	43	46	35	37	31	44
			25	49	41	42	25	38	46	25	37	39
			30	32	44	49	25	38	38	33	26	33
7	PSB68	Budda	**10**	48	43	46	32	38	29	36	29	36
			15	40	46	45	41	43	33	40	25	29
			20	32	42	39	44	41	40	33	37	39
			25	46	51	34	44	28	25	38	42	35
			30	38	40	31	37	37	32	39	39	29
Overall Average Adulteration Level Classification Accuracy (%)				**41.31**	**44.74**	**39.03**	**35.80**	**37.00**	**35.00**	**36.40**	**34.40**	**34.71**

Please cite this article as: B. S. Anami, N. N. Malvade and S. Palaiah, Automated recognition and classification of adulteration levels from bulk paddy grain samples, INFORMATION PROCESSING IN AGRICULTURE, https://doi.org/10.1016/j.inpa.2018.09.001

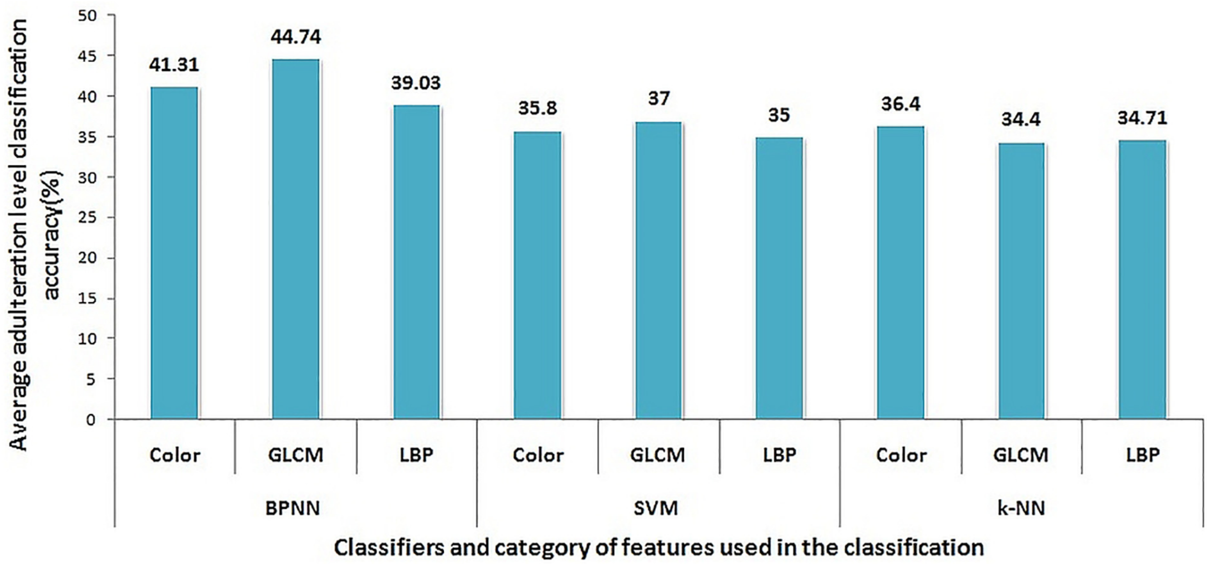

Fig. 8 – Classification results using separate Color, GLCM, and LBP features independently with BPNN, SVM, and k-NN classifiers.

$$\text{Kurtosis}(\gamma) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\{(P(i,j)-\mu)\}^4}{MN\sigma^4} \quad (11)$$

2.3.2. *Texture feature extraction*

The textures of bulk paddy grain samples are not often uniform due to variations in orientation, scale, uneven illumination and other visual appearance of individual grains. This has influenced to incorporate invariance with respect to spatial scale, orientation and grayscale while extracting texture features from the images of adulterated bulk paddy grain sample. Two texture analysis methods, namely GLCM and basic LBP with 8-neighborhood have been employed to extract and characterize texture features from the grayscale images of the adulterated bulk paddy sample. A total of forty-two features considering 10 GLCM features (mean, variance, entropy, uniformity, homogeneity, inertia, cluster shade, cluster prominence, maximum probability, correlation) and 32 LBP features (mean, standard deviation, smoothness, third moment, uniformity, entropy, gray level range and HG1 thru HG25 histogram bands) are extracted [10]. In order to extract the LBP histogram group (HG1 thru HG25) features, 256 Gy values from gray matrix are grouped into 25 histogram bands and the number of pixels in each band is counted. The GLCM and LBP feature values of the variety mixture Jaya and Abhilasha at five different levels of adulteration percentage are shown graphically in Figs. 6 and 7.

2.4. *Classifiers*

2.4.1. *Back Propagation Neural Network (BPNN)*

Multilayer back propagation neural network has been used as one of the classifiers in the present work because of its ease and strength in execution for large training data set. Levenberg-Marquardt (LM) back-propagation algorithm is used for the training. The termination error (TE) is set to 0.01, the learning rate (η) is set to 0.05 and the momentum coefficient (μ) is set to 0.6. The sigmoid activation functions are used in the hidden layers. The color and texture features are used to train and test the neural network model. The number of neurons in the input layer is set to the number of chosen color or/and texture features. The number of output neurons is set to 35 (7 mixed bulk paddy samples at 5 different adulteration levels). The network is trained and tested for 1000 epochs.

2.4.2. *Support Vector Machine (SVM)*

Multi-class Support vector machine (SVM) is a potential linear classifier based on the concept of decision planes that defines decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. It builds a hyperplane from the training data which separates pixels with different class memberships. In the proposed methodology, the preprocessed images are classified using SVM with Gaussian Radial Basis Function (RBF) kernel function. The optimal sigma parameter value of RBF has been sampled over the range 1.0 to 2.0.

2.4.3. *k-Nearest Neighbor (k-NN)*

The k-nearest neighbor (k-NN) algorithm is an ad-hoc classifier used to classify test data based on a distance metric. In the present work, Euclidean distance with a desired range of values for the neighborhood parameter 'k' (k = 1, 2, 3...) is used for the image classification. The choice of 'k' value is driven by the end application as well as the dataset and plays an important role in the classifier performance.

2.5. *Feature selection/reduction*

Two feature selection techniques, namely principal component analysis and sequential forward floating selection

Table 3 – Classification results of using combined color-texture features with BPNN, SVM and K-NN classifiers.

Sl. No.	Premium paddy variety	Adulterant paddy variety	Adulteration level (%)	Average adulteration level classification accuracy (%) of classifiers		
				BPNN	SVM	k-NN (For k = 3)
1	Jaya	Abhilasha	**10**	72	65	60
			15	78	75	63
			20	77	78	79
			25	68	66	79
			30	74	58	67
2	Jaya	Mugad 101	**10**	78	79	58
			15	69	68	68
			20	73	64	76
			25	72	75	71
			30	80	69	55
3	Jaya	Thousand One	**10**	66	58	74
			15	79	58	60
			20	77	56	67
			25	65	62	59
			30	69	73	58
4	Mugad Siri	Thousand Ten	**10**	70	65	61
			15	80	67	66
			20	67	61	55
			25	55	79	63
			30	76	60	58
5	Mugad Sughand	Thousand Ten	**10**	69	65	64
			15	68	75	57
			20	55	78	56
			25	71	73	65
			30	66	67	65
6	PSB68	Abhilasha	**10**	72	58	69
			15	74	77	57
			20	72	70	44
			25	79	56	75
			30	68	57	75
7	PSB68	Budda	**10**	77	58	74
			15	66	75	65
			20	64	62	74
			25	72	65	61
			30	77	80	56
Overall Average Adulteration Level Classification Accuracy (%)				**71.29**	**67.20**	**64.40**

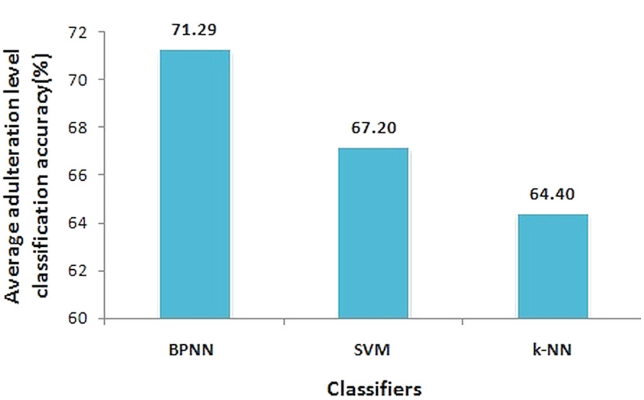

Fig. 9 – Classification efficiencies using combined Color, GLCM, and LBP features with BPNN, SVM, and k-NN classifiers.

algorithm have been employed to decrease the computational overhead and increase the average bulk paddy adulteration level recognition accuracy by selecting significant and non-overlapping color and texture features.

2.5.1. *Principal component analysis (PCA)*

The PCA is a powerful tool for analyzing patterns in high dimensional data, which can be compressed by reducing the number of dimensions without losing abundant information. In the present work, the PCA is employed to find nonlinear relationships of the extracted color and texture features. The Algorithm 1 gives the steps involved in the reduction of color-texture features using PCA.

Table 4 – Fifty-one selected color and texture features obtained by SFFS algorithm.

Sl. No.	Features	Feature type	Sl. No.	Features	Feature type
1	Red mean	**Color**	22	Mean	**GLCM**
2	Red standard deviation		23	Variance	
3	Red skewness		24	Uniformity	
4	Red kurtosis		25	Homogeneity	
5	Green mean		26	Inertia	
6	Green standard deviation		27	Cluster shade	
7	Green skewness		28	Cluster prominence	
8	Green kurtosis		29	Maximum probability	
9	Blue mean		30	Mean	**LBP**
10	Hue mean		31	Standard deviation	
11	Hue standard deviation		32	Third moment	
12	Hue skewness		33	Entropy	
13	Hue kurtosis		34	Contrast	
14	Saturation mean		35	Third moment	
15	Intensity mean		36–51	H7–H25 (Histogram groups)	
16	Luminance mean				
17	Blue chromaticity mean				
18	Blue chromaticity standard deviation				
19	Red chromaticity mean				
20	Red chromaticity standard deviation				
21	Red chromaticity kurtosis				

Algorithm 1: Color-texture feature reduction using PCA.

Input: Adulterated bulk paddy grain images (RGB).
Output: Lower dimensional feature vector.
Start
Step 1: Extract and load the color and texture features from the images of adulterated bulk paddy grains
Step 2: Compute the d-dimensional mean vectors for the different adulterated bulk paddy grain sample image data set.
Step 3: Compute the scatter or co-variance matrices (among adulterated paddy variety samples).
Step 4: Compute the eigenvectors (e1, e2,…, en) and corresponding eigenvalues (λ1, λ2, …, λn) for the scatter matrices.
Step 5: Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a $n \times k$-dimensional matrix W (where every column represents an eigenvector).
Step 6: Derive the new fused color and texture feature set. Use $n \times k$ eigenvector matrix to transform the samples onto the new subspace.
$y = W^T \times x$
Where $\times$ is a $d \times 1$-dimensional vector representing one sample and y is the transformed $k \times 1$-dimensional sample in the new subspace.
Stop.

2.5.2. *Sequential forward floating selection algorithm (SFFS)*

The sequential forward floating selection algorithm is employed in the present work to determine the features with the significant contribution to the classification accuracy [11]. The SFFS algorithm begins the search with an empty feature set and uses the basic sequential forward selection algorithm to add one feature at a time to the selected feature subset. Every time a new feature is added to the current feature set, the algorithm attempts to backtrack by using the sequential backward selection algorithm to remove one feature at a time to locate a better subset. The algorithm stops when the size of the current feature set is larger than the required number of features.

3. Experimental results and discussion

In the present work, the overall bulk paddy adulteration level classification performances of the individual and combined color-texture features are evaluated through the use of BPNN, SVM and k-NN classifiers. The PCA and SFFS feature selection methods have been employed separately to optimize the classification results. The dataset consisting of a total of 7000 sample images are partitioned into two equal halves and one half is used for training and other is used for testing. The percentage accuracy of adulteration level classification is defined as the ratio of correctly classified sample images to the total number of sample images considered.

3.1. *Classification based on separate color and texture features*

The training and testing processes have been carried out using all the separate 45 color, 10 GLCM and 30 LBP texture features with the BPNN, SVM and k-NN classifiers independently. The classification results are summarized in Table 2 and graphically represented in Fig. 8. From Fig. 8, the maxi-

Table 5 – Classification results of using combined selected color-texture features with BPNN, SVM, and K-NN classifiers.

Sl. No.	Premium paddy variety	Adulterant paddy variety	Adulteration Level (%)	SFFS-based selected color-texture features			PCA-based features (40 principal component coefficients)		
				BPNN	SVM	k-NN	BPNN	SVM	k-NN (For k = 3)
1	Jaya	Abhilasha	**10**	94	83	82	98	88	82
			15	93	92	89	92	82	87
			20	94	93	83	90	93	77
			25	87	92	89	93	92	74
			30	91	86	76	89	87	91
2	Jaya	Mugad 101	**10**	92	90	90	88	88	83
			15	92	83	78	96	90	95
			20	93	86	85	91	93	90
			25	93	90	90	93	91	94
			30	89	87	86	88	85	88
3	Jaya	Thousand One	**10**	92	81	77	98	90	87
			15	87	93	77	98	94	88
			20	90	84	79	93	96	83
			25	90	80	80	95	93	82
			30	88	88	83	88	91	91
4	Mugad Siri	Thousand Ten	**10**	91	93	79	93	88	81
			15	87	88	76	93	84	77
			20	95	92	84	97	88	79
			25	91	92	78	96	90	83
			30	94	87	90	96	93	88
5	Mugad Sughand	Thousand Ten	**10**	88	81	82	95	91	81
			15	95	87	78	92	94	80
			20	91	87	75	96	88	75
			25	94	94	80	92	89	81
			30	90	83	79	95	81	84
6	PSB68	Abhilasha	**10**	88	87	82	98	82	75
			15	94	80	80	92	89	77
			20	87	81	88	94	79	83
			25	88	89	78	89	94	88
			30	92	82	88	90	97	82
7	PSB68	Budda	**10**	88	83	84	89	88	86
			15	94	82	77	98	89	79
			20	91	92	85	94	89	82
			25	87	89	78	92	88	88
			30	95	85	90	95	91	87
Overall Average Adulteration Level Classification Accuracy (%)				**91.00**	**86.91**	**82.14**	**93.31**	**89.29**	**83.66**

mum average adulteration level classification accuracy of 44.74% is obtained from the 10 GLCM features with the BPNN classifier and the lowest average adulteration level classification accuracy of 34.40% is obtained from the 10 GLCM features with the k-NN classifier.

3.2. *Classification based on combined color and texture features*

It is observed through the experimentation that neither the color feature nor the texture features are successful in yielding the intended results. So, an attempt has been made to combine the color and texture features in a suitable way to realize comprehensive image classification process. 45 color, 10 GLCM and 30 LBP texture features are simply concatenated into a single feature vector and the feature vector is used for training and testing the BPNN, SVM and k-NN classifiers independently. The classification results are summarized in Table 3 and graphically represented in Fig. 9. From Fig. 9, the average adulteration level classification accuracies of 71.29%, 67.20% and 64.40% are obtained from the combined color-texture features with BPNN, SVM and k-NN classifiers respectively. From the experimental results, it is observed that the combined color-texture features provide improved classification accuracy results over using them independently for all the considered classifiers.

3.3. *Classification based on combined selected color and texture features*

The high dimensionality of the extracted image features led to improper and poor predictive performance of the classi-

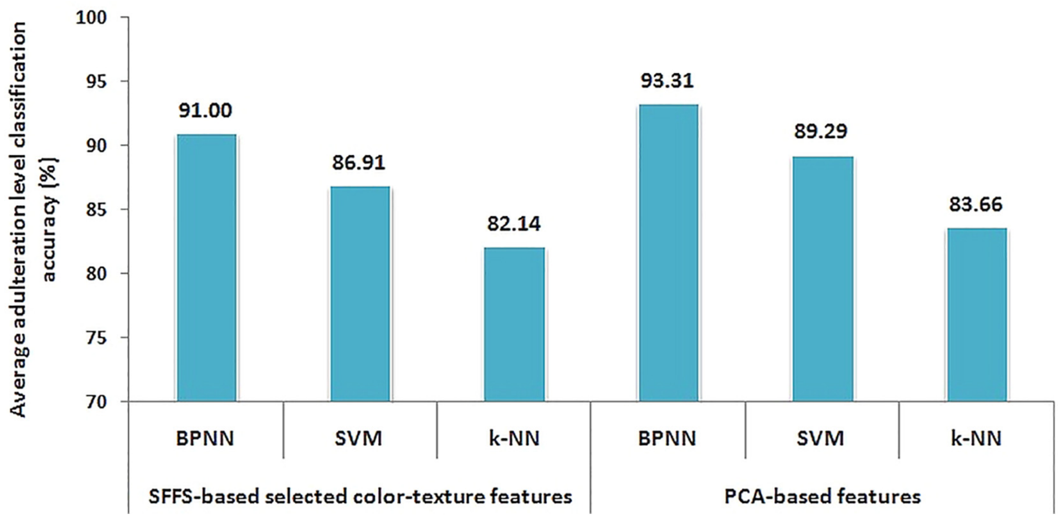

Fig. 10 – Classification efficiency results of selected color-texture features with different classification methods.

fiers. This problem is rectified by eliminating highly correlated features. The dimensionality of the extracted 45 color, 10 GLCM and 30 LBP features is reduced by employing PCA and SFFS methods separately. Five sets of principal components, considering 10, 20, 30, 40, 50 and 60 principal component coefficients in each set are extracted using the PCA method. Through experimentation, it has been observed that all the considered classifiers achieve approximately similar performance up to around 40 principal component coefficients. The SFFS method selects a total of 51 color-texture features, considering 21 color, 8 GLCM, and 22 LBP texture features and the selected features are listed in Table 4. Six classification models have been developed, considering three models for the PCA-based features and three models for the SFFS-based selected color-texture features. The experimental results of all the classification models are summarized in Table 5 and graphically represented in Fig. 10. From Table 5, it is observed that the overall average adulteration level classification accuracies of 91%, 86.91% and 82.14% are obtained from the BPNN, SVM and k-NN classifiers when trained using SFFS-based color-texture features, and the average classification accuracies of 93.31%, 89.29% and 83.66% are obtained from BPNN, SVM and k-NN classifiers when trained using PCA-based features. From the graph shown in Fig. 10, it is observed that the classification models based on the PCA-based features outperform other classification models based on SFFS-based selected color-texture features. It is also observed from the results that the classification performance of the BPNN classifier is better than the performances of SVM and k-NN classifiers. The classification results of the BPNN classifier using the PCA-based features are given in Table 6. From Table 6, only 96 images out of 3500 adulterated bulk paddy grain sample test images have been misclassified by the BPNN classifier thereby yielding an average adulterated bulk paddy sample recognition accuracy of 97.26%.

4. Conclusion

The potential of color and texture features is demonstrated using different classification methods to classify adulteration levels (%) from the images of mixed bulk paddy sample consists of two varieties. The study compares three different classifiers (BPNN, SVM, and k-NN) and describes how the choice of image features affects the classification performances. The experimental results have shown that the combined color and texture features are suitable for recognition and classification of adulteration levels in bulk paddy grain samples. The BPNN classification model using PCA-based features has given the good average adulteration classification accuracy of 93.31% compared to SVM and k-NN classifiers. The proposed work is also effective in recognizing paddy varieties in mixed bulk grain samples and has a number of advantages over traditional DNA approaches including rapidness and lower cost. These characteristics may make it suitable for efficient bulk paddy grain sample screening, thereby providing an alternative to pre-existing approaches. The results of the proposed work are promising and still, there is scope for classification accuracy optimization. Therefore, in the light of the above analysis, the further focus will be on using the variety of wavelet transforms for texture analysis and different classification techniques (decision tree, random forest) for adulteration study. The work finds application in rice production industries, import and export trading firms, agricultural produce market committee (APMC) yards and in agricultural research universities.

Table 6 – Classification results of using PCA-based features (40 principal component coefficients) with BPNN classifier.

Sl. No.	Premium paddy variety	Adulterant paddy variety	Number of test images	Actual Adulteration Level (%)	Predicted Adulteration Level (%)					MC (%)	CA (%)
					10	15	20	25	30		
1	Jaya	Abhilasha	100	**10**	**98**	1	0	0	0	1	98
			100	**15**	1	**92**	1	1	0	5	92
			100	**20**	1	2	**90**	1	2	4	90
			100	**25**	0	1	3	**93**	2	1	93
			100	**30**	1	1	1	4	**89**	4	89
2	Jaya	Mugad 101	100	**10**	**88**	2	1	0	0	9	88
			100	**15**	0	**96**	2	1	1	0	96
			100	**20**	1	1	**91**	3	1	3	91
			100	**25**	0	1	1	**93**	1	4	93
			100	**30**	0	1	2	2	**88**	7	88
3	Jaya	Thousand One	100	**10**	**98**	0	1	0	0	1	98
			100	**15**	1	**98**	1	0	0	0	98
			100	**20**	1	1	**93**	3	0	2	93
			100	**25**	0	1	1	**95**	1	2	95
			100	**30**	0	1	3	1	**88**	7	88
4	Mugad Siri	Thousand Ten	100	**10**	**93**	1	1	1	3	1	93
			100	**15**	2	**93**	0	1	1	3	93
			100	**20**	1	1	**97**	0	1	0	97
			100	**25**	0	2	1	**96**	1	0	96
			100	**30**	1	0	0	0	**96**	3	96
5	Mugad Sughand	Thousand Ten	100	**10**	**95**	1	0	0	1	3	95
			100	**15**	1	**92**	3	2	1	1	92
			100	**20**	0	1	**96**	1	0	2	96
			100	**25**	1	2	3	**92**	0	2	92
			100	**30**	0	0	2	1	**95**	2	95
6	PSB68	Abhilasha	100	**10**	**98**	1	0	0	1	0	98
			100	**15**	2	**92**	3	0	1	2	92
			100	**20**	2	0	**94**	3	0	1	94
			100	**25**	0	0	3	**89**	2	6	89
			100	**30**	0	0	1	3	**90**	6	90
7	PSB68	Budda	100	**10**	**89**	3	0	2	1	5	89
			100	**15**	0	**98**	1	0	0	1	98
			100	**20**	0	1	**94**	1	2	2	94
			100	**25**	0	0	2	**92**	2	4	92
			100	**30**	0	0	0	3	**95**	2	95
Overall Average Adulteration Level Classification Accuracy (%)											**93.31**

Note: MC – Misclassification with respect to recognition of adulterated bulk paddy sample, CA – Classification Accuracy with respect to adulteration level.

REFERENCES

[1] Archak S, Reddy LV, Nagaraju J. High-throughput multiplex microsatellite marker assay for detection and quantification of adulteration in basmati rice (Oryza sativa). Electrophoresis 2007;28(14):2396–405.

[2] Ali T, Jhandhir Z, Ahmad A, Khan M, Khan AA, Choi GS. Detecting fraudulent labeling of rice samples using computer vision and fuzzy knowledge. Multimedia Tools Appl 2017;76(23):24675–704.

[3] Anami BS, Naveen NM, Hanamaratti NG. A colour features-based methodology for variety recognition from bulk paddy images. Int J Adv Intell Paradigms 2015;7(2):187–205.

[4] Anami BS, Naveen NM, Hanamaratti NG. Behavior of HSI color co-occurrence features in variety recognition from bulk paddy grain image samples. Int J Signal Process, Image Process Pattern Recogn 2015;8(4):19–30.

[5] Carter RM, Yan Y, Tomlins K. Digital imaging based classification and authentication of granular food products. Meas Sci Technol 2006;17:235–40.

[6] Fayyazi S, Abbaspour-Fard MH, Rohani A, Monadjemi SA, Sadrnia H. Identification and classification of three iranian rice varieties in mixed bulks using image processing and MLP neural network. Int J Food Eng 2017;13(5).

[7] Majumdar S, Jayas DS. Advanced Computer Vision capabilities provide an alternative to manual inspection. J Agric Eng Res 1999;73(1):35–47.

[8] Majumdar S, Jayas DS. Classification of cereal grains using machine vision. II. Colour models. Trans ASAE 2000;43(6):1677–80.

[9] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24(7):971–87.

[10] Pourreza A, Pourreza H, Abbaspour-Fard MH, Sadrnia H. Identification of nine Iranian wheat seed varieties by textural

Please cite this article as: B. S. Anami, N. N. Malvade and S. Palaiah, Automated recognition and classification of adulteration levels from bulk paddy grain samples, INFORMATION PROCESSING IN AGRICULTURE, https://doi.org/10.1016/j.inpa.2018.09.001

analysis with image processing. Comput Electron Agric 2012;83:102–8.
[11] Porebski A, Vandenbroucke N, Macaire L. Comparison of feature selection schemes for color texture classification. In: Image Processing Theory Tools and Applications (IPTA), 2010 2nd International Conference (IPTA'10), IEEE, Paris. p. 32–7.
[12] Rahman MM, Rasaul MG, Hossain MA, Iftekharuddaula KM, Hasegawa H. Molecular characterization and genetic diversity analysis of rice (Oryza sativa L.) using SSR markers. J Crop Improv 2012;26(2):244–57.
[13] Timsorn K, Lorjaroenphon Y, Wongchoosuk C. Identification of adulteration in uncooked Jasmine rice by a portable low-cost artificial olfactory system. Measurement 2017;108:67–76.
[14] Mäenpää T, Pietikäinen M. Classification with color and texture: jointly or separately? Pattern Recogn 2004;37 (8):1629–40.
[15] Vemireddy LR, Satyavathi VV, Siddiq EA, Nagaraju J. Review of methods for the detection and quantification of adulteration of rice: Basmati as a case study. J Food Sci Technol 2015;52 (6):3187–202.
[16] Visen NS, Paliwal J, Jayas D, White ND. Image analysis of bulk grain samples using neural networks. 2003 ASAE Annual Meeting. American Society of Agricultural and Biological Engineers, 2003.
[17] Pazoki AR, Farokhi F, Pazoki Z. Classification of rice grain varieties using two artificial neural networks (MLP and Neuro-Fuzzy). J Anim Plant Sci 2014;24(1):336–43.
[18] Mousavi Rad S, Rezaee J, Nasri K. A new method for identification of Iranian rice kernel varieties using optimal morphological features and an ensemble classifier by image processing. Majlesi J Multimedia Process 2012;1(3):1–8.
[19] Savakar DG, Anami BS. Effect of foreign bodies on recognition and classification of bulk food grains image samples. J Appl Comput Sci Math 2009;3(6):77–83.
[20] Khunkhet Somthawin, Remsungnen Tawun. Non-destructive identification of breeding rice seed by using image processing and fuzzy logic. J Sci Eng Res 2018;5(3):108–21.

Please cite this article as: B. S. Anami, N. N. Malvade and S. Palaiah, Automated recognition and classification of adulteration levels from bulk paddy grain samples, INFORMATION PROCESSING IN AGRICULTURE, https://doi.org/10.1016/j.inpa.2018.09.001