Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 7 (2020) 272–285

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

Classification of yield affecting biotic and abiotic paddy crop stresses using field images

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

[a] Department of Computer Science and Engineering, K. L. E. Institute of Technology, Hubballi 580030, India
[b] Visvesvaraya Technological University, Belagavi, 590018 Karnataka, India
[c] Department of Information Science and Engineering, K. L. E. Institute of Technology, Hubballi 580030, India
[d] Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwar, 580002 Karnataka, India

A R T I C L E I N F O

Article history:
Received 22 November 2018
Received in revised form
29 April 2019
Accepted 29 August 2019
Available online 29 August 2019

Keywords:
Paddy crop stress
Color moments
MPEG-7 color descriptors
BPNN
SVM
k-NN

A B S T R A C T

On-time recognition and early control of the stresses in the paddy crops at the booting growth stage is the key to prevent qualitative and quantitative loss of agricultural yield. The conventional paddy crop stress identification and classification activities invariably rely on human experts to identify visual symptoms as a means of categorization. This process is admittedly subjective and error-prone, which in turn may lead to incorrect action in stress management decisions. The proposed work presented in this paper aims to develop an automated computer vision system for the recognition and classification of paddy crop stress types from the field images using the state-of-the-art color features. The work examines the impact of eleven stress types, two biotic and nine abiotic stresses, on five different paddy crop varieties during the booting growth stage using field images and analyzes the stress responses in terms of color variations using lower-order color moments and two visual color descriptors defined by the MPEG-7 standard, the Dominant Color Descriptor (DCD) and Color Layout Descriptor (CLD). The Sequential Forward Floating Selection (SFFS) algorithm has been employed to reduce the overlapping between the features. Three different classifiers, the Back Propagation Neural Network (BPNN), the Support Vector Machine (SVM), and the k-Nearest Neighbor (k-NN) have been deployed to distinguish among stress types. The average stress classification accuracies of 89.12%, 84.44% and 76.34% have been achieved using the BPNN, SVM, and k-NN classifiers, respectively. The proposed work finds application in the development of decision support systems and mobile apps for the automation of crop and resource management practices in the field of agricultural science.

1. Introduction

Rice is life, for most of the people living in Asia, a region with higher population density and more than three billion people around the globe eat rice every day. It has shaped the cultures, diets, and economies of thousands of millions of people. Rice makes up 20% of the world's dietary energy supply. Saying

* Corresponding author at: Department of Information Science and Engineering, K. L. E. Institute of Technology, Hubballi, 580030 Karnataka, India.
E-mail addresses: anami_basu@hotmail.com (B.S. Anami), naveen.malvade@gmail.com (N.N. Malvade), surendrap@uasd.in (S. Palaiah).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2019.08.005

that rice production is critical to global food security is an understatement. The paddy crop cultivation is an integral part of Indian agricultural economy, ranking first in the cultivation area with 43.92 million hectares with the production of 111.50 million tonnes [3]. Although the overall figures are impressive, the productivity of Indian paddy cultivation is low. According to the Food and Agricultural Organization (FAO), the average yield of paddy in India is 4.50 tonnes per hectare. The comparative figure for China, by contrast, is more than 6.2 tonnes per hectare. This difference in the yield is due to the differences in the management practices adopted by the farmers. Rice production is coming under increasing pressure in Asia due to population growth and changing socio-economic factors. The major rice producing belts in Asia regions achieve only 40% of total production efficiency due to damage caused by drought, diseases, and pests. A more integrated approach involving optimal crop improvement and resource management practices such as nutrient management, irrigation regime, and other agronomic management factors for paddy crops with existing farmers' practices is a potential option for minimizing the crop yield gap [1].

Paddy crop is cultivated in a wide range of environments characterized by different temperatures, climates, and soil-water conditions. The crops are therefore exposed to various types of stress, both biotic and abiotic, whose combined effect can adversely affect crop performance and survival. Biotic stresses include insect pests, fungus, bacteria, viruses, and herbicide toxicity. Abiotic stresses include such as drought, high salinity, high or low temperatures, flooding, high light, ozone, low nutrient availability, mineral deficiency, heavy metals, pollutants, wind, and mechanical injury. It is generally believed that all these stresses are considered as a serious threat to sustainable paddy production. Among these stresses, pathogen, drought, over-irrigation or submergence, nutrient deficiency, toxicity due to over-feeding of fertilizers, and high salinity stress factors have a huge impact on the world agriculture and they reduce average yields by more than 50%. The paddy plants have evolved and developed specific mechanisms to respond to complex stress conditions. The plant responses to the individual stress vary depending on the nature and severity of the stress involved, the age of the plant at which the stress is encountered, and inherent stress tolerant nature of the plants. The on-time monitoring and recognition of these stresses, the supply of adequate farm inputs and rapid morphological diagnoses can reduce the adverse effects of stresses on the crops. The stressed paddy crops show obvious symptoms in the color, shape, and texture of the leaves. It is difficult to capture and quantify these micro-symptoms by manual visual observation. The sample field images of normal and stressed paddy crops are shown in Fig. 1.

One of the most sensitive indices of the paddy crop under stress is the change in the color of leaves or a yellowing of leaves caused by stress. This forms the basis for the present work. Two simple decision tools, namely, leaf color chart (LCC) and chlorophyll meter (SPAD) are used to recognize plant nutrient deficiency in Nitrogen [2]. The LCC measures the greenness of a paddy crop leaf as a plant Nitrogen status indicator. However, critical LCC values vary considerably among different paddy varieties and critical color shades on the LCC (seven green shades ranging from yellowish green to dark green) need to be determined accurately to guide fertilizer application. In LCC, the accuracy is not guaranteed, especially for different lighting conditions, as the approach is based on visual inspection of the leaf color.

The leaf injury and leaf color variation are the two visual scoring symptoms used in the identification and classification of different stress categories. The trained personnel visually observe these symptoms as a qualitative assessment to determine the stress types. In most of the cases, interpreting visual nutrient deficiency, toxicity and disease symptoms in paddy plants is difficult, even for the most experienced eyes. Hence, the visual inspection is considered as subjective, time-consuming, sometimes destructive and tedious by nature. An objective and rapid stress identification system would be beneficial to the potential farmers and researchers for timely intervention and mitigation of the problems by applying the proper crop management strategies that can effectively boost the crop yields. To know the state-of-the-art automation of crop stress identification and classification in the field of cereal crops, a survey has been carried out and the gist of a survey is given as under.

A non-destructive nutrient stress diagnosis method has been proposed to identify NPK (Nitrogen, Phosphorus, Potassium) deficiencies using the paddy plant leaf color indices

Fig. 1 – Sample field images of (a) Normal and (b) Stressed paddy crops.

and morphology. In the work, 9 morphological and 13 color indices are extracted from the RGB images of leaves. A region of interest (ROI) has been selected to extract the green part from the entire leaf and the extracted part has been used in the quantification of individual RGB channels. The stepwise discriminant analysis (SDA) with leave-one-out-cross-validation (LOOCV) algorithm has been used to identify NPK deficiencies [17]. A rule-based approach has been developed to determine NPK nutrient deficiencies based on the HSV color model features such as mean, minimum, maximum, and deviation features extracted from the paddy leaf images and achieved the average classification accuracy of 95.39% [8]. A principle component analysis plus support vector machine (PCA-SVM) model has been developed to identify rice leaf blast and nitrogen stress using the hyperspectral rice leaf images. The average identification accuracy of 97.5% is obtained from the two paddy varieties [7]. A disease detection technique has been developed to recognize three types of diseases, namely, brown spot, leaf blast, and bacterial blight using paddy plant leaf images. The Scale Invariant Feature Transform (SIFT) features are used to train and test the SVM and k-NN classifiers. The average recognition accuracies of 91.10% and 93.33% are obtained for the SVM and k-NN classifiers respectively [10]. The Laser Induced Fluorescence (LIF) technique combined with the Support Vector Machine (SVM) and Principal Component Analysis (PCA) has been proposed to identify Nitrogen stress in paddy crops. Two crop growth stages, namely, tillering and shooting have been considered in the work and obtained an average recognition accuracy of 95% [20]. A hyperspectral image analysis based bag of spectra words (BoSW) model has been presented for rice panicle blast grading using a chi-square kernel support vector machine (chi-SVM) classifier. The procedure of k-means clustering algorithm with Euclidian distance, quantization, and histogram statistics transform each hyperspectral image with millions of dimensions into a BoSW model. An average test accuracy of 81.41% has been achieved for grading six levels of panicle blasts [18]. A diagnosis model has been developed to identify NPK nutrient deficiencies in the paddy crops using the shape and color characteristics acquired from the scanned images of paddy plant leaves and sheaths. The optimal features required for the discrimination have been obtained using the Fisher Discriminant Analysis (FDA) and the support vector feature selection methods. The overall identification accuracies obtained for N, P, and K deficiencies are 86.15%, 87.69% and 90% respectively [5]. A correlation study has been presented among SPAD readings, Leaf Nitrogen Concentration (LNC) of paddy crop leaves, and the color indices calculated from the images of paddy crop leaves using RGB, HSV, and L*a*b* color models. The regression analysis showed that the significant linear relationships among index b*, SPAD meter readings, and LNC parameters [19]. An automated system has been developed to classify the brown spot and blast diseases of rice plant based on the morphological changes extracted from the segmented leaf images. The radial hue distribution feature vectors are used separately with the Bayes' and SVM classifiers and obtained the average classification accuracies of 68.1% and 79.5% from the classifiers, respectively [12]. An automated color image analysis system has been developed to detect two types of diseases present in paddy plant leaves. The hue and saturation color features are extracted from the segmented individual leaf images and the k-means technique is applied to cluster the pixels into a number of groups. These clusters are subjected to further analysis to determine the suspected diseases of the rice leaf [14]. A color texture analysis of paddy plant leaves approach has been presented for recognizing six mineral deficiencies such as Nitrogen, Iron, Magnesium, Potassium, Boron, and Manganese using Multi-Layer Perceptron (MLP) classifier. The color and texture based segmentation approaches are combined to obtain individual paddy plant images for color and texture features extraction. The overall average classification accuracy of 88.56% is obtained using 70 color and 40 texture features [16].

The literature survey has revealed that the response mechanisms to abiotic and biotic stresses in paddy crop have been extensively investigated at physiological, biochemical, genetic and molecular levels. In this respect, thermal, reflectance and fluorescence images have proven their potential by detecting the physiological state or stress related changes in individual plant leaves nondestructively [4]. Several research works have been published on the recognition, classification and quantification of paddy plant stresses using the shape, color and texture features extracted from the images of single leaflets. It is also understood from the field experts and the literature that the paddy crops are more susceptible to different stresses during the booting growth stage and the severe stress at this growth stage can cause irrecoverable damage to plants, resulting in reduced yields. However, no referable results and sophisticated works have been cited with regard to the recognition of paddy crop stresses from the field images. This brings the desire of developing a comprehensive paddy crop stress recognition system, which could be used as an effective tool to construct crop management strategies.

In consultation with the agricultural experts, the present study considers eleven different types of stresses (two biotic and nine abiotic stresses) in paddy crops, namely, bacterial blight, fungal blast, drought, submerged, nitrogen deficiency, phosphorus deficiency, potassium deficiency, boron deficiency, zinc deficiency, iron deficiency, and chemical injury. The paddy crop stress classification tree is shown in Fig. 2.

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

2. Proposed methodology

To deal with the stated challenges, a crop field image based approach has been introduced to recognize and classify the paddy crop stresses. The proposed methodology consists of five stages, namely, image acquisition, feature extraction, feature selection, and identification and classification of paddy crop stresses. The block diagram illustrating the stages involved in the proposed methodology is shown in Fig. 3.

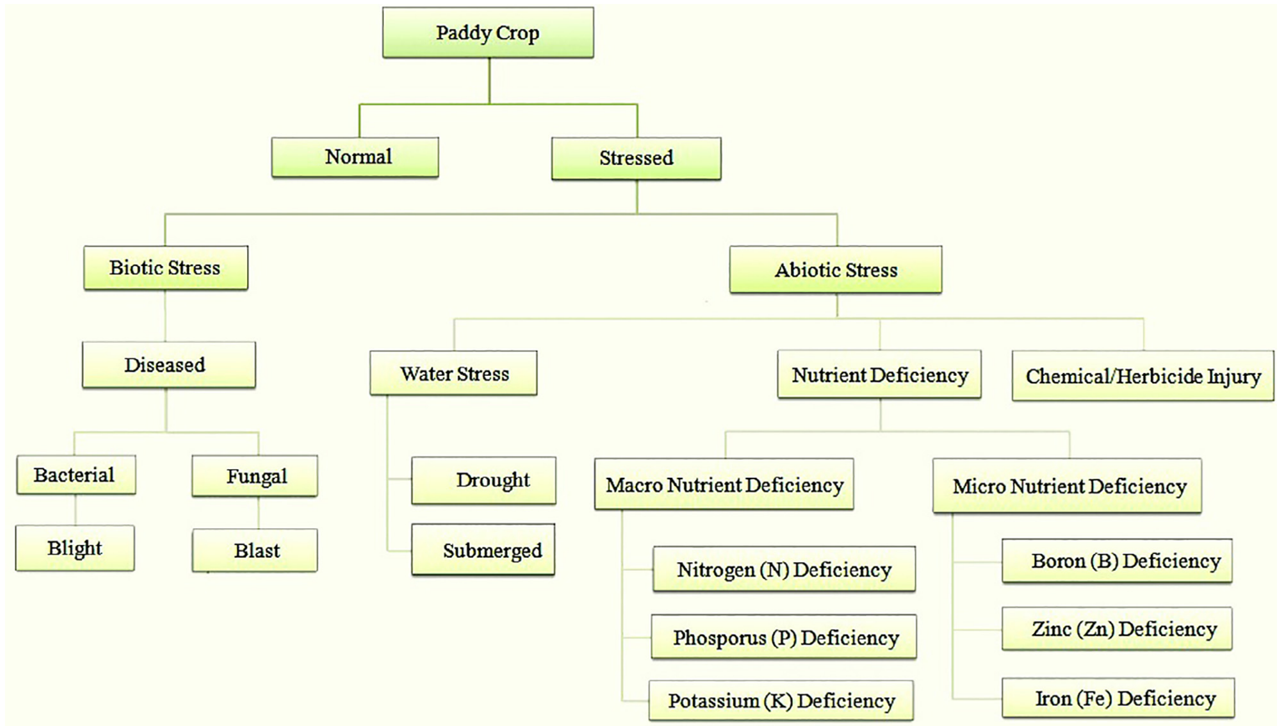

Fig. 2 – Paddy crop stress classification tree.

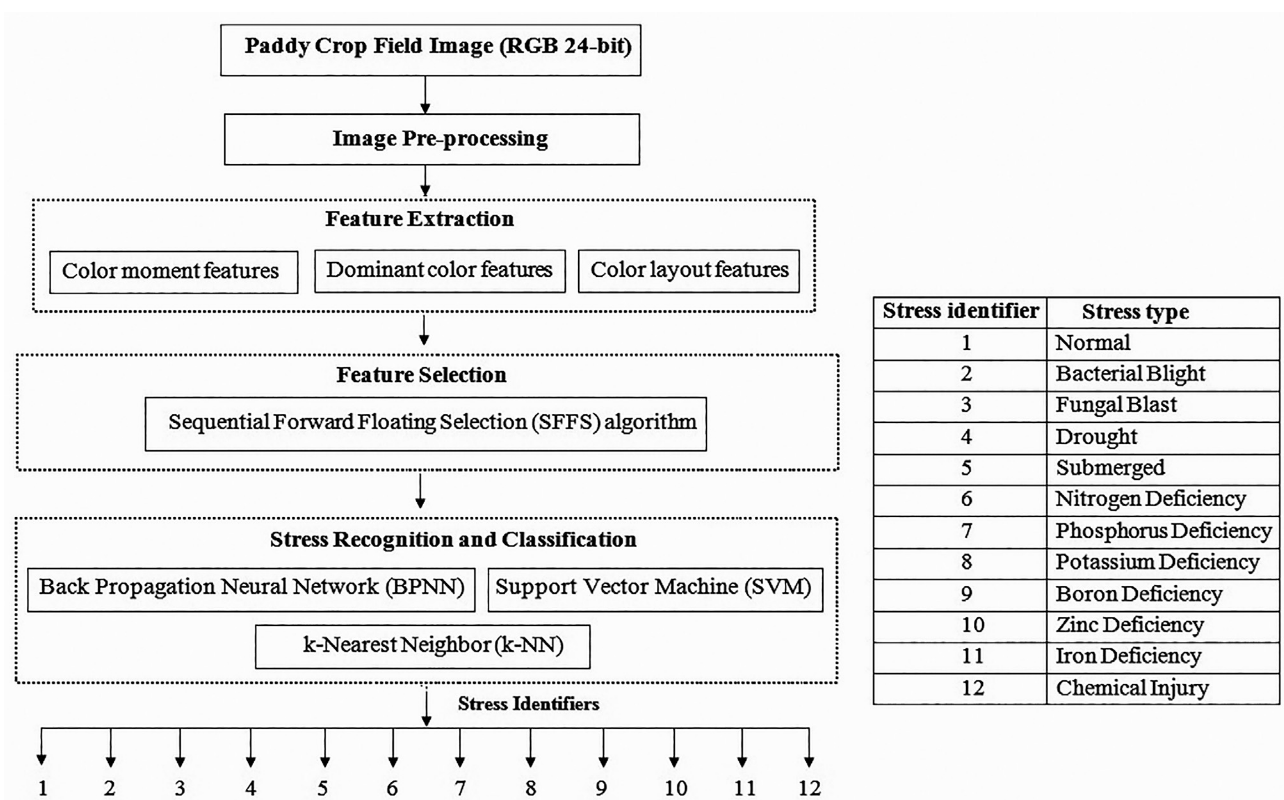

Fig. 3 – Block diagram of the proposed methodology.

2.1. Study area and site description

The experimental paddy fields are situated at All India Coordinated Rice Improvement Project (AICRIP), University of Agricultural Sciences, Mugad, Dharwad, India. The study area is categorized as the dry zone with the annual rainfall ranges between 464.5 and 785.7 mm. The soil is medium to deep black clay in larger areas and sandy loam in a small portion with high organic matter content. For the individual paddy crop stress analysis, 60 experimental fields have been considered.

2.2. Crop sample preparation

In consultation with the University of Agricultural Sciences, Dharwad, Karnataka State, India, five certified and popular paddy varieties, namely, Jaya, Abhilasha, Mugad Suganda, Mugad 101, and Mugad Siri are selected as experimental grain samples. The grain samples are having 100% physical and genetic purity. The collected paddy grains were sown separately in a raised bed nursery (direct seeding) as per the standard guidelines. All the necessary precautions were taken to maintain a uniform plant population of each variety and the plants were grown under controlled nutrient conditions. These paddy crop varieties have been challenged by all the considered eleven stress types and the observations have been carried out under conditions favoring normal growth and expression of all the crop stress symptoms in order to fulfill the objectives of the study. The field tests were conducted during the Kharif season of 2018.

2.3. Image acquisition

The work considers eleven stress types, two biotic and nine abiotic stresses, on five different paddy crop varieties. A total of 2200 images of each paddy crop variety are acquired, taking into account 200 images per stress type. A total of 200 healthy field images per paddy crop variety along with the stressed paddy field images have been included in the dataset. The stress identities have been confirmed by the experts. Stresses such as bacterial blight and fungal blast are present at low to a medium severity. Each image in this dataset is associated with an expert marked label indicating a stress class. The images were captured in the field under natural light condition near solar noon, which is the period with the most stable illumination at the top of the atmosphere, using a Nikon D3300 Digital SLR camera having a resolution of 24 megapixels. The camera parameters, such as the ISO, shutter speed, and aperture have been set to 400, 1/500, and f/32, respectively. In order to provide a rigid and stable support and easy movement, the camera is mounted on a tripod stand. The angle between the camera lens and the object axis is maintained at approximately 45° as shown in Fig. 4. The images are taken keeping the object distance of 1.5 m. A white foam board is used as a reflector to bounce and intensify natural light in the opposite direction to fill the shadows.

Over 12,000 labeled images are collected to create the experimental dataset of paddy field images from healthy crops and crops exhibiting different stresses. 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 agricultural experts and rice scientists. The cropped images are subjected for image preprocessing to eliminate the influence of illumination changes and noise. The removal of shading and contrast correction in the images is accomplished through the general histogram equalization technique. The technique is applied in such a way that the intensity values are equalized without disturbing the color balance of the image and this is achieved by converting the color space of an image from RGB to HSV, then performing histogram equalization on the intensity plane (V), and finally converting the resultant HSV image back

Fig. 4 – Image acquisition system (Image courtesy: Paul W. N. et al., [11]).

(a) (b)

Fig. 5 – Image pre-processing (a) Original paddy crop field image (b) Histogram equalized image.

to RGB [6]. The median filter is used to remove the noise in the images. A sample pre-processed paddy crop field image is shown in Fig. 5. The sample images of paddy fields under various stress conditions are shown in Fig. 6.

2.4. Feature extraction

The present work employs three lower-order color moments, namely, mean, variance and skewness, and two color descriptors defined by the MPEG-7 standard, namely, Dominant Color Descriptor (DCD) and Color Layout Descriptor (CLD) to describe the color variations and distribution in the stressed paddy crop field images [9]. The DCD captures a set of dominant or salient colors in an image, and considers the percentage of image pixels used in each color, as well as the variance and spatial coherence of the colors. The CLD represents the spatial distribution of colors. One of the main aspects of color feature extraction is the choice of a color model. Four different color models, namely, RGB, HSV, YC_bC_r, and L*a*b* have been used to explore the required color features.

2.4.1. Color moment features extraction

The color moment feature extraction starts with the separation of the color channels of the stressed paddy crop field images. The images in the RGB color model are separated into R, G, and B components respectively. The components Hue (H), Saturation (S), and Value (V) are obtained from the R, G, and B components using Eqs. (1)–(3). Three color moments, namely, mean (μ), variance (s), and skewness (θ) are extracted per color component using the Eqs. (4)–(7). A total of 18 color moments is extracted from the six color components (R, G, B, H, S, and V) and the color moment features for Jaya paddy crop field images under various stress conditions are listed in Table 1.

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

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

$$V = max\ (R,\ G,\ B) \tag{3}$$

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}P(i,j) \tag{4}$$

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\{(P(i,\ j)\ -\mu)\}^2} \tag{5}$$

$$s = \sigma\ X\ \sigma \tag{6}$$

$$\theta = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\{(P(i,j)-\mu)\}^3}{MN\sigma^3} \tag{7}$$

where *M* is the image matrix dimension, *N* is the total number of pixels in the image and *P(i, j)* is the color pixel value at i^{th} row and j^{th} column.

2.4.2. Dominant color features extraction

The DCD allows a specification of a small number of representatives or principal color values in an image as well as their statistical properties, such as distribution and variance. The DCD is defined by the Eq. (8).

$$F = \{(ci, pi, vi), s\},\quad (i = 1, 2, \ldots, N) \tag{8}$$

where 'c_i' is i^{th} dominant color value, 'p_i' is a percentage value of the i^{th} dominant color, 'v_i' is variance, which describes the variation of the color values of the pixels in a cluster around the corresponding i^{th} dominant color, 's' is the spatial coher-

Fig. 6 – Sample field images of Jaya paddy crop under various stress conditions. (a) Bacterial Blight (b) Fungal Blast (c) Drought (d) Submergence (e) Nitrogen Deficit (f) Phosphorus Deficit (g) Potassium Deficit (h) Boron Deficit (i) Zink Deficit (j) Iron Deficit (k) Chemical Injury.

ency, which is a single number that represents the overall spatial homogeneity of the dominant colors in the image and 'N' is the number of dominant colors. In the present work, only the dominant colors and their percentage values are considered. The procedure involved in the DCD feature extraction is summarized in Algorithm 1. The extracted mean DCD features for Jaya paddy crop field images under various stress conditions are listed in Table 2.

Table 1 – The color moment features of normal and stressed Jaya paddy variety crop field images.

Sl. no	Paddy crop stress	Color moments features in color planes								
		R			G			B		
		μ	s	θ	μ	s	θ	μ	s	θ
1	Normal	0.0386	0.5926	0.8166	0.796	0.5506	0.7152	0.2963	0.6518	0.5783
2	Bacterial Blight	0.8049	0.3382	0.6477	0.2067	0.0776	0.4282	0.9789	0.4479	0.3322
3	Fungal Blast	0.8629	0.8618	0.0581	0.4353	0.492	0.673	0.5738	0.7932	0.2689
4	Drought	0.6837	0.5915	0.287	0.5273	0.4601	0.5465	0.119	0.4879	0.2503
5	Submergence	0.6864	0.245	0.422	0.493	0.1459	0.1814	0.2455	0.1301	0.3127
6	Nitrogen Deficiency	0.4563	0.5284	0.356	0.8582	0.228	0.7511	0.2431	0.9105	0.6464
7	Phosphorus Deficiency	0.287	0.4189	0.5295	0.4478	0.3193	0.6191	0.9671	0.849	0.5743
8	Potassium Deficiency	0.893	0.7675	0.0872	0.9705	0.4612	0.843	0.6769	0.5979	0.6401
9	Boron Deficiency	0.6362	0.629	0.6202	0.0079	0.8253	0.2079	0.7203	0.8752	0.1574
10	Zinc Deficiency	0.7568	0.4766	0.5698	0.9174	0.2989	0.6931	0.7714	0.4112	0.5771
11	Iron Deficiency	0.4563	0.5284	0.356	0.8582	0.1128	0.7511	0.2431	0.9105	0.6464
12	Chemical Injury	0.287	0.4189	0.5295	0.4478	0.2293	0.6191	0.9671	0.849	0.5711
Sl. no	Paddy crop stress	Color moments features in color planes								
		H			S			V		
		μ	s	θ	μ	s	θ	μ	s	θ
1	Normal	0.9053	0.2375	0.1495	0.0061	0.1542	0.9709	0.2081	0.3305	0.2979
2	Bacterial Blight	0.9188	0.9863	0.5629	0.0665	0.8455	0.7254	0.4321	0.8716	0.4051
3	Fungal Blast	0.2364	0.8361	0.1166	0.7082	0.5426	0.3317	0.6888	0.4495	0.1722
4	Drought	0.3888	0.2077	0.6624	0.3558	0.4757	0.4788	0.5571	0.0377	0.5951
5	Submergence	0.9260	0.9649	0.1704	0.7975	0.9044	0.7066	0.7934	0.5883	0.6633
6	Nitrogen Deficiency	0.6375	0.9362	0.9812	0.3767	0.7074	0.9849	0.7042	0.0265	0.2830
7	Phosphorus Deficiency	0.4005	0.5444	0.5131	0.8674	0.1675	0.7450	0.5976	0.7473	0.3985
8	Potassium Deficiency	0.6950	0.0635	0.0843	0.5951	0.0234	0.2077	0.4194	0.7089	0.6831
9	Boron Deficiency	0.4323	0.8279	0.2133	0.3449	0.0702	0.8547	0.8520	0.1382	0.8856
10	Zinc Deficiency	0.3827	0.6869	0.4583	0.2699	0.0849	0.7577	0.6311	0.8907	0.1911
11	Iron Deficiency	0.9347	0.5813	0.9387	0.3601	0.4639	0.3894	0.6822	0.0302	0.4823
12	Chemical Injury	0.2937	0.9977	0.8658	0.3997	0.4829	0.0492	0.9541	0.0875	0.5417

Algorithm 1 (*DCD feature extraction*).

Input: RGB images of stressed paddy crop field image
Output: Dominant colors and their percentage values
Description: Start with initializing number of dominant colors 'N' = 3 and the number of color clusters 'k' = 3
Step 1: Convert RGB image to CIE L*a*b* image [15].
Step 2: Calculate the centroids for each of the color components of the image.
Step 3: Use the k-means clustering method to segment the dominant colors of the image.
(K-means clustering is a method of cluster analysis which aims to divide 'n' observations into 'k' clusters in which each observation belongs to the cluster with the nearest mean. This algorithm follows a sequence of centroid calculations and clustering steps until the number of desired clusters is met)
Step 4: Calculate the percentage of pixels belonging to each of the color clusters or segmented dominant color. The centroids (*c*) and their corresponding percentage of pixels (*p*) are the DCD features of the image.
Stop.

2.4.3. *Color layout features extraction*

The CLD is used to represent a compact and resolution-invariant spatial distribution of colors in images. The feature extraction using a CLD includes two stages; grid-based representative color selection and Discrete Cosine Transform (DCT) with quantization. The procedure involved in extracting the CLD features is summarized in Algorithm 2. The extracted mean CLD features for Jaya paddy crop field images under various stress conditions are listed in Table 3.

Algorithm 2 (*CLD feature extraction*).

Input: RGB images of stressed paddy crop field image
Output: Mean value from Y, C_b, and C_r components
Description: Start with initializing number of dominant colors 'N' = 3 and the number of color clusters 'k' = 3
Step 1: Convert RGB image color space into YC_bC_r image.
Step 2: Divide each color component into $8 \times 8 = 64$ blocks and derive the average color of each block.
Step 3: Transform the derived average colors into a series of coefficients by performing the 8×8 Discrete Cosine Transform (DCT).
Step 4: Calculate the average vertical (C_v), horizontal (C_h) and diagonal (C_d) values from all the color components.
Stop

2.5. *Feature selection*

The Sequential Forward Floating Selection (SFFS) algorithm has been employed to decrease the computational overhead

Table 2 – The dominant color (DCD) features of normal and stressed Jaya paddy variety crop field images.

Sl. no	Paddy crop stress	Dominant color values or centroids (c) and their percentages (p) in L*a*b* color space																	
		L*						a*						b*					
		c_1	p_1	c_2	p_2	c_3	p_3	c_1	p	c_2	p_2	c_3	p_3	c_1	p_1	c_2	p_2	c_3	p_3
1	Normal	26.72	0.29	75.81	0.59	10.14	0.1	20.06	0.17	53.72	0.31	50.28	0.49	72.12	0.19	47.81	0.54	18.7	0.15
2	Bacterial Blight	58.54	0.27	73.95	0.29	61.9	0.31	53.15	0.47	27.67	0.12	17.47	0.30	38.62	0.33	94.16	0.24	89.06	0.29
3	Fungal Blast	79.97	0.13	8.02	0.27	35.2	0.57	1.95	0.48	75.22	0.3	46.42	0.18	21.92	0.18	32.42	0.45	67.54	0.25
4	Drought	96.82	0.29	15.09	0.43	43.22	0.2	66.17	0.59	93.01	0.22	79.27	0.13	52.61	0.44	31.95	0.26	0.59	0.29
5	Submergence	11.60	0.23	11.11	0.33	21.51	0.37	9.66	0.42	21.08	0.17	29.05	0.37	65.24	0.27	11.27	0.45	63.64	0.26
6	Nitrogen Deficiency	81.32	0.42	93.81	0.12	64.52	0.45	44.32	0.43	41.62	0.13	78.69	0.39	89.73	0.28	70.1	0.47	28.5	0.15
7	Phosphorus Deficiency	93.77	0.22	66.37	0.43	16.74	0.28	1.29	0.27	61.1	0.34	54.38	0.3	80.24	0.33	68.72	0.41	95.08	0.18
8	Potassium Deficiency	56.14	0.29	12.89	0.23	35.79	0.42	81.74	0.13	75.23	0.42	44.3	0.44	96.05	0.28	11.18	0.22	0.74	0.45
9	Boron Deficiency	77.01	0.44	73.16	0.11	80.92	0.39	7.15	0.26	19.11	0.37	68.19	0.28	78.5	0.14	2.21	0.29	61.78	0.43
10	Zinc Deficiency	27.55	0.42	1.72	0.25	71.35	0.22	40.03	0.33	90.33	0.38	13.23	0.22	44.69	0.31	10.54	0.27	2.24	0.34
11	Iron Deficiency	12.43	0.22	37.4	0.26	99.87	0.35	49.07	0.26	87.46	0.29	13.33	0.39	91.62	0.26	62.93	0.48	48.58	0.23
12	Chemical Injury	35.88	0.42	65.52	0.32	34.28	0.23	39.3	0.24	6.19	0.48	99.96	0.22	88.58	0.29	37.76	0.34	26.14	0.27

and increase the average paddy crop stress recognition and classification accuracy by selecting significant and non-overlapping color features [13]. 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.

2.6. *Classifiers*

The proposed work has adopted three classifiers, namely, Multi-layer Back Propagation Neural Network (BPNN), Multi-class Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN) for classification of paddy crop field images. The experiment has been conducted to study their suitability in the process of classification.

2.6.1. *Back propagation neural network*

Multi-layer Back Propagation Neural Network (BPNN) 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 training the network. The termination error (TE) is set to 0.01, the learning rate is set to 0.05, and the momentum coefficient is set to 0.6. The sigmoid activation function is used in the hidden layers. The lower-order color moments, dominant colors, and spatial color distribution 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 the chosen color features. The network has been trained and tested for 800 epochs.

2.6.2. *Support vector machine*

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 hyper-plane from the training data which separates pixels with different class memberships. In the proposed methodology, the stressed paddy crop field 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 of 1.0–2.0.

2.6.3. *k-Nearest neighbor*

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. In the present work, the value of k has been set to 5.

Table 3 – The color layout (CLD) features of normal and stressed Jaya paddy variety crop field images.

Sl. no	Paddy crop stress	Mean of (C_v, C_h, C_d)
1	Normal	47.42
2	Bacterial Blight	27.35
3	Fungal Blast	24.14
4	Drought	20.93
5	Submergence	22.50
6	Nitrogen Deficiency	25.42
7	Phosphorus Deficiency	31.09
8	Potassium Deficiency	19.02
9	Boron Deficiency	27.17
10	Zinc Deficiency	15.43
11	Iron Deficiency	30.60
12	Chemical Injury	35.02

Table 4 – Stress recognition and classification performances of BPNN, SVM and k-NN classifiers across the five paddy crop varieties using the color moments, DCD, and CLD feature vectors separately.

Sl. no.	Paddy crop stress	Stress classification accuracies (%) across the paddy varieties using separate color feature vectors								
		Color moment features			DCD features			CLD features		
		BPNN	SVM	k-NN	BPNN	SVM	k-NN	BPNN	SVM	k-NN
1	Normal	69.28	79.01	58.52	70.65	66.24	55.93	33.01	45.64	41.47
2	Bacterial Blight	67.11	75.06	52.35	54.57	58.39	52.68	47.44	54.27	53.70
3	Fungal Blast	65.85	74.84	60.50	43.25	63.30	57.45	48.07	40.46	51.12
4	Drought	71.43	77.90	58.86	53.23	60.87	41.88	43.04	31.15	39.73
5	Submergence	69.49	70.77	60.02	57.29	69.05	43.55	39.38	38.04	47.17
6	Nitrogen Deficiency	60.61	67.59	56.13	55.69	63.24	58.03	45.31	60.58	64.79
7	Phosphorus Deficiency	58.19	66.39	55.77	60.26	60.34	50.06	47.72	61.63	59.86
8	Potassium Deficiency	57.23	69.47	46.55	67.58	57.24	52.81	48.89	58.36	53.87
9	Boron Deficiency	61.02	61.68	53.93	61.48	68.44	51.50	53.97	47.01	46.80
10	Zinc Deficiency	63.57	66.52	64.20	49.43	63.43	45.35	58.19	54.94	55.17
11	Iron Deficiency	66.89	67.30	53.65	55.62	57.11	48.22	56.89	58.93	63.89
12	Chemical Injury	68.96	74.11	64.81	58.56	65.11	57.55	46.59	49.86	52.75
Average stress classification accuracy (%) across the paddy crop varieties and the stress types		64.97	**70.89**	57.11	57.30	62.73	51.25	**47.38**	50.07	52.53

3. Experimental results and discussion

In the present work, the overall paddy crop stress identification and classification performances of the individual and combined color feature categories are evaluated through the use of BPNN, SVM, and k-NN classifiers. The SFFS feature selection method is used to optimize the results of the classification. The dataset consisting of a total of 12,000 sample images are partitioned into two equal halves and one half is used for training and the other half is used for testing. The percentage accuracy of stress classification is defined as the ratio of correctly classified sample images to the total number of sample images considered.

3.1. *Classification results based on the individual color feature vector*

Three separate color feature vectors are constructed from 18 color moment features, 18 DCD features, and 1 CLD feature, respectively. The feature vectors are used to train and test

Table 5 – Stress recognition and classification performances of BPNN, SVM and k-NN classifiers across the five paddy crop varieties using the combined color feature vector.

Sl. no.	Paddy crop stress	Stress classification accuracies (%) across the paddy varieties using combined color feature vector (18 color moment features + 18 DCD features + 1 CLD feature)		
		BPNN	SVM	k-NN
1	Normal	80.11	75.21	66.15
2	Bacterial Blight	77.17	79.37	75.53
3	Fungal Blast	74.87	85.44	70.23
4	Drought	86.43	82.52	73.56
5	Submergence	79.22	76.68	71.59
6	Nitrogen Deficiency	75.55	80.92	64.51
7	Phosphorus Deficiency	85.32	73.32	73.34
8	Potassium Deficiency	79.16	87.65	75.37
9	Boron Deficiency	74.77	71.39	64.49
10	Zinc Deficiency	87.34	81.19	72.28
11	Iron Deficiency	85.18	75.96	67.11
12	Chemical Injury	84.28	70.85	73.19
Average stress classification accuracy (%) across the paddy crop varieties and the stress types		**80.78**	78.38	70.61

Table 6 – Selected color features obtained for the BPNN, SVM, and k-NN classifiers using the SFFS algorithm.

Sl. no.	Classifier	Color feature category	Selected color features	Feature vector size
1	BPNN	Color moments	Red mean, Red variance, Red skewness, Green mean, Green variance, Green skewness, Blue mean, Blue skewness, Hue mean, Hue variance, Hue skewness, Saturation mean, Saturation variance, Value mean, Value variance	30
		DCD	L* $\{c_1, p_1, c_2, p_2, c_3, p_3\}$, a* $\{c_2, p_2, c_3, p_3\}$, b* $\{c_1, p_1, c_2, p_2\}$	
		CLD	Mean of (C_v, C_h, C_d)	
2	SVM	Color moments	Red mean, Red variance, Red skewness, Green mean, Green variance, Blue mean, Blue skewness, Hue mean, Hue variance, Hue skewness, Saturation mean, Saturation skewness	25
		DCD	L* $\{c_1, p_1, c_2, p_2\}$, a* $\{c_2, p_2, c_3, p_3\}$, b* $\{c_2, p_2, c_3, p_3\}$	
		CLD	Mean of (C_v, C_h, C_d)	
3	k-NN	Color moments	Red mean, Red variance, Red skewness, Green mean, Green variance, Green skewness, Blue mean, Blue variance, Hue mean, Hue variance, Saturation mean, Value variance	23
		DCD	L* $\{c_1, p_1, c_2, p_2, c_3, p_3\}$, a* $\{c_1, p_1, c_2, p_2\}$ b* $\{c_2, p_2, c_3, p_3\}$	
		CLD	Mean of (C_v, C_h, C_d)	

Table 7 – Stress recognition and classification performances of the BPNN, SVM and k-NN classifiers across the five paddy crop varieties using the selected color feature vectors obtained using SFFS algorithm.

Sl. no.	Paddy crop stress	Stress classification accuracies (%) across the paddy varieties using selected color feature vectors		
		Selected 30 color features	Selected 25 color features	Selected 23 color features
		BPNN	SVM	k-NN
1	Normal	88.80	84.20	72.18
2	Bacterial Blight	89.80	81.33	77.15
3	Fungal Blast	90.80	85.45	70.95
4	Drought	89.80	82.17	75.67
5	Submergence	84.60	85.65	80.54
6	Nitrogen Deficiency	91.40	80.26	73.97
7	Phosphorus Deficiency	90.20	83.39	79.13
8	Potassium Deficiency	89.40	87.14	86.17
9	Boron Deficiency	90.80	89.12	70.74
10	Zinc Deficiency	87.80	82.34	75.63
11	Iron Deficiency	88.20	87.53	70.87
12	Chemical Injury	87.80	84.66	83.12
Average stress classification accuracy (%) across the paddy crop varieties and the stress types		**89.12**	84.44	76.34

the BPNN, SVM, and k-NN classifiers and the classification results are summarized in Table 4. From Table 4, the SVM classifier using color moment features has given a maximum average classification accuracy of 70.89% and the minimum average classification accuracy of 47.38% has been achieved for CLD feature with the BPNN classifier.

3.2. *Classification results based on the combined color features*

A combined color feature vector comprising a total of 37 color features (18 color moment features + 18 DCD features + 1 CLD feature) is constructed to train and test the BPNN, SVM, and k-NN classifiers. The classification results are summarized in Table 5. From Table 5, the BPNN classifier has given the maximum average classification accuracy of 80.78% and the minimum average classification accuracy of 70.61% has been achieved by the k-NN classifier.

3.3. *Classification results based on the selected color features*

In order to improve the classification performances of the combined color feature vector, a feature selection operation is carried out using the SFFS algorithm and the feedback from the classifiers is used in selecting the significant color features. The SFFS algorithm returns three separate color feature vectors of sizes 30, 25, and 23 for the BPNN, SVM, and k-NN classifiers, respectively and the feature vectors are listed in Table 6. The training and testing processes have been carried

Table 8 – Stress recognition and classification performance of the BPNN classifier using selected 30 color features.

Sl. no.	Paddy crop stress type	Stress recognition and classification accuracy (%) per paddy crop variety					Average stress classification accuracy (%) across paddy crop varieties
		Jaya	Abhilasha	Mugad Suganda	Mugad 101	Mugad Siri	
1	Normal	90	88	90	83	93	88.80
2	Bacterial Blight	91	87	88	89	94	89.80
3	Fungal Blast	82	93	94	89	96	90.80
4	Drought	91	89	89	92	88	89.80
5	Submerged	91	83	81	84	84	84.60
6	Nitrogen Deficiency	92	93	93	90	89	91.40
7	Phosphorus Deficiency	88	93	91	85	94	90.20
8	Potassium Deficiency	85	89	96	91	86	89.40
9	Boron Deficiency	86	91	96	92	89	90.80
10	Zinc Deficiency	83	94	88	84	90	87.80
11	Iron Deficiency	89	83	88	94	87	88.20
12	Chemical Injury	88	93	94	82	82	87.80
Average stress recognition and classification accuracy (%) across the paddy crop varieties and the stress types							**89.12**

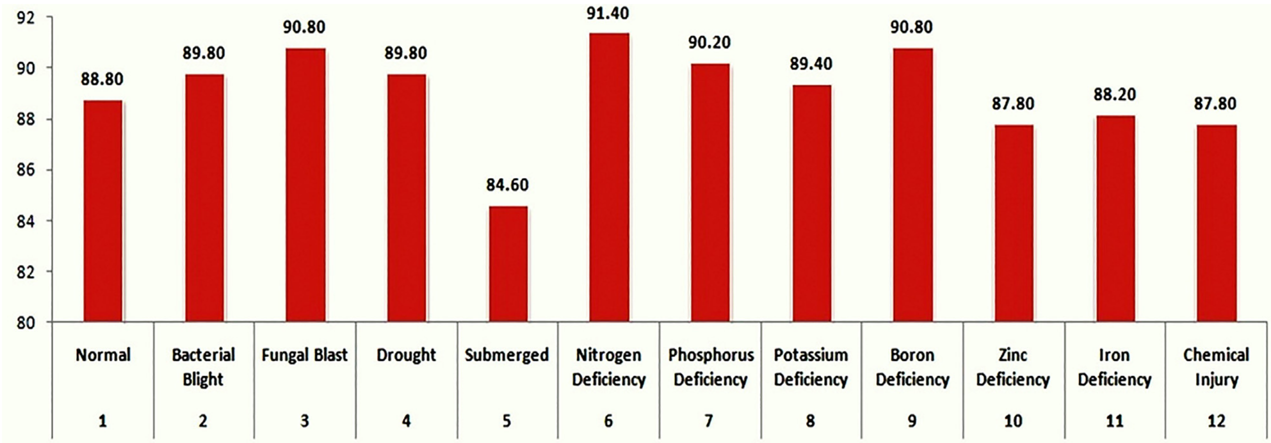

Fig. 7 – Graph showing the average stress classification accuracy results of the BPNN classifier across the five paddy crop varieties.

out using the feature vectors with the BPNN, SVM, and k-NN classifiers and the classification results are summarized in Table 7. From Table 7, the average classification accuracies of 89.12%, 84.44%, and 76.34% have been obtained for the BPNN, SVM, and k-NN classifiers, respectively. The BPNN classifier trained using 30 selected color features has yielded a maximum average classification accuracy and the detailed stress classification results with respect to all the paddy crop varieties considered in the work are given in Table 8. From Table 8, the maximum and minimum average stress classification accuracies of 91.40% and 84.60% are obtained for the Nitrogen deficiency and submerged stress types, respectively. The BPNN classifier based average stress classification results across the paddy crop varieties are graphically shown in Fig. 7.

4. Conclusion

The work has successfully used field images to classify different categories of paddy crop stress. The color-based features are used to describe the human-interpretable visual crop stress symptoms. The color models RGB, HSV, YC_bC_r, and L*a*b* are used to analyze the color variation in stressed paddy crop field images. The combination of color features consisting of color moments, dominant colors, and spatial distribution of colors is used to train the BPNN, SVM, and k-NN classifiers. The BPNN-based approach outperforms the two other methods, resulting in a maximum average stress classification accuracy of 89.12% across the twelve stress types on five different paddy crop varieties considered. The results obtained are encouraging as the work considers more number of stress types than in the reported works. There is a tremendous opportunity for plant scientists and breeders to quantify paddy crop stresses. But the quantification methods are an extension of classification methods in which each stress class is quantified on a scale of 0 to100% based on stress severity. This stress quantification has been considered as the next frontier for our future research efforts. The work is complex and challenging in terms of high irregularity in an outdoor environment. The presented approach is reasonably robust against outdoor illumination changes. The generality of the proposed methodology can make it applicable to a wide variety of field crops such as wheat, maize, barley, soybean, etc. The effects of environmental conditions such as extreme temperatures and soil factors, the presence of combinatorial stresses, the quantification of stresses, and the prediction of the gap between yield potential and yield under stress are the factors for further studies.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

REFERENCES

[1] Alam MM, Karim MR, Ladha JK. Integrating best management practices for rice with farmers' crop management techniques: a potential option for minimizing rice yield gap. Field Crops Res 2013;144:62–8.
[2] Ali MM, Al-Ani A, Eamus D, Tan DK. Leaf nitrogen determination using non-destructive techniques – a review. J Plant Nutr 2017;40(7):928–53.
[3] Anonymous, USDA Foreign Agriculture Service, India-Grain and Feed Annual, GAIN Report Number IN8027, link: <http://www.gain.fas.usda.gov>; 2018.
[4] Bock CH, Poole GH, Parker PE, Gottwald TR. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Crit Rev Plant Sci 2010;29(2):59–107.
[5] Chen L, Lin L, Cai G, Sun Y, Huang T, Wang K, et al. Identification of nitrogen, phosphorus, and potassium deficiencies in rice based on static scanning technology and hierarchical identification method. PLoS ONE 2014;9(11) e113200.
[6] Jeon Gwanggil, Young-Sup Lee. Histogram equalization-based color image processing in different color model. Adv Sci Technol Lett 2013;28:54–7.
[7] Yuan Jianqing, Su Zhongbin, Jia Yinjiang, Zhang Yu, Zhang Zongxin. Identification of rice leaf blast and nitrogen deficiency in cold region using hyperspectral imaging. Trans Chinese Soc Agric Eng 2016;32(13):155–60.
[8] Latte MV, Shidnal S, Anami BS. Rule based approach to determine nutrient deficiency in paddy leaf images. Int J Agric Technol 2017;13(2):227–45.

[9] Manjunath BS, Salembier P, Sikora T. Introduction to MPEG- 7: multimedia content description interface. John Wiley & Sons; 2002. p. 1.
[10] Mohan KJ, Balasubramanian M, Palanivel S. Detection and recognition of diseases from paddy plant leaf images. Int J Comput Appl 2016;144(12).
[11] Paul WN, Joseph A Shaw, Jha Prashant, Scherrer Bryan, Donelick Andrew, Kumar Vipan. Discrimination of herbicide-resistant kochia with hyperspectral imaging. J Appl Remote Sens 2018;12(1) 016037.
[12] Phadikar S, Sil J, Das AK. Classification of rice leaf diseases based on morphological changes. Int J Inform Electron Eng 2012;2(3):460–3.
[13] Porebski A, Vandenbroucke N, Macaire L. Comparison of feature selection schemes for color texture classification. In: Image processing theory tools and applications (IPTA), 2nd International conference (IPTA'10), IEEE, Paris; 2010. p. 32–7.
[14] Pugoy RAD, Mariano VY. Automated rice leaf disease detection using color image analysis. In: Third international conference on digital image processing (ICDIP 2011), International society for optics and photonics; 2011; 8009. p. 80090F.
[15] Robertson AR. The CIE 1976 color-difference formulae. Color Res Appl 1977;2(1):7–11.
[16] Sanyal P, Bhattacharya U, Parui SK, Bandyopadhyay SK, Patel S. Color texture analysis of rice leaves diagnosing deficiency in the balance of mineral levels towards improvement of crop productivity. In: Information technology, (ICIT 2007), 10th International conference, IEEE. p. 85–90.
[17] Sun Y, Tong C, He S, Wang K, Chen L. Identification of nitrogen, phosphorus, and potassium deficiencies based on temporal dynamics of leaf morphology and color. Sustainability 2018;10(3):762.
[18] Huang Shuangping, Qi Long, Ma Xu, Xue Kunnan, Wang Wenjuan, Zhu Xiaoyuan. Hyperspectral image analysis based on BoSW model for rice panicle blast grading. Comput Electron Agric 2015;118:167–78.
[19] Wang Y, Wang D, Shi P, Omasa K. Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. Plant Methods 2014;10(1):36.
[20] Yang J, Gong W, Shi S, Du L, Sun J, Ma YY, et al. Accurate identification of nitrogen fertilizer application of paddy rice using laser-induced fluorescence combined with support vector machine. Plant, Soil Environ 2015;61(11):501–6.