

REGULARIZED ISOTONIC REGRESSION BASED JENSEN-SHANNON BOOSTING CLASSIFICATION FOR PLANT LEAF DISEASE IDENTIFICATION

Sheela Newsheeba* and Dr. UmaDevi

Received: March 29, 2021; Revised: October 17, 2021; Accepted: November 07, 2021

Abstract

Automatic plant leaf disease identification and classification systems are important for precision agriculture. Fast and accurate identification of leaf diseases is a vital problem solved in agriculture and prevents losses. To enhance the accuracy of plant leaf disease identification, Regularized Perona-Malik Segmentive Isotonic Regression-based Jaccard Jensen-Shannon Boosting Classification (RPMSIR-JJSBC) technique is introduced. The proposed technique includes five processing steps. Initially, the image acquisition process is performed to collect multiple rice plant leaf images with IoT. Secondly, input rice plant leaf images are preprocessed to improve quality by removing noises. Thirdly, region-based segmentation is performed to partition images into number of segments. Next, weighted Isotonic Regression is applied to minimize disease identification time. Finally, extracted features are analyzed and classify the image into normal or disease with higher accuracy. Thus, RPMSIR-JJSBC technique performs disease identification. The simulation result demonstrates different parameters namely peak signal to noise ratio, disease identification accuracy, false-positive rate, and disease identification time.

Keywords: Plant leaf disease identification, preprocessing, segmentation, feature extraction classification

Introduction

To improve the quality of agricultural products, the development of agricultural domain image processing technology is applied. Accurate and timely identification of diseases in plants helps the farmers for treatment on the plants and decreases the economic losses. Conventionally, several methods have been developed to diagnose plant leaf diseases. A hybrid Bacterial Foraging Optimization Algorithm and Deep Neural Network (BFOA-DNN) was

developed in (Aakrati *et al.*, 2020) for identifying the paddy leaf diseases with unique features.

To enhance detection rate and minimize entropy loss, a classification method was employed.

However, the designed (BFOA-DNN) model failed to identify and detect plant leaf diseases with minimum time consumption. However, the designed simple CNN model failed to perform the segmentation for detecting and classifying rice

¹ Sri Jayendra Saraswathy Maha Vidyalaya College of Arts & Science, Coimbatore, India. E-mail: sheela79.rincy@gmail.com

² Department of Computer Science & IT, Sri Jayendra Saraswathy Maha Vidyalaya College of Arts & Science, Coimbatore, India. E-mail: umadevigayathri@gmail.com

* Corresponding author

diseases and pests more effectively. In order to overcome these issues, we have used the proposed RPMSIR-JSBC technique which is used for exact disease identification with lesser time.

RPMSIR-JSBC technique provides enhanced performance for achieving high accuracy than the existing methods. In order to overcome these issues, we have used the proposed RPMSIR-JSBC technique which is used for exact disease identification with lesser time. RPMSIR-JSBC technique provides enhanced performance for achieving high accuracy than the existing methods.

A sine cosine algorithm-based rider neural network was presented in (Monalisa *et al.*, 2021) for disease identification. In order to create appropriate plant disease detection, images were pre-processed by the median filter. But, it failed to detect which type of disease has been occurred in plant leaves. A convolutional neural network (CNN) was developed in (Sachin *et al.*, 2020) for soybean disease identification. Plant pathologists need a correct and consistent soybean plant disease diagnosis system. But it failed to increase the performance rate of the disease identification. A Bayes' classifier and minimum distance classifier were developed in (Vikas *et al.*, 2020) for recognition of rice disease with higher accuracy. In order to enhance the development and construction of rice crops, the ground-breaking method for the identification of disease that takes place in rice is identified. Thus, rice crops provide an essential part. However, the time consumption was not reduced.

A Fuzzy-Filtered Neural Network was designed in Toran *et al.* (2019) for diagnosing rice disease. However, the optimal feature selection was not carried out to reduce the complexity of the algorithm.

A Faster region-based CNN (R-CNN) was developed in Guoxiong *et al.* (2019) for rapid recognition of Rice disease. Faster R-CNN fusion is developed for tackling different issues by rice disease images. R-CNN algorithm is employed for the detection of rice diseases. But the accuracy of disease identification was not improved. In David *et al.* (2019) various machine learning algorithms were developed for identifying the rice blast disease. Rice blast disease is a major biotic limitation of rice farming which causes every year millions dollars of victims. Rice blast prediction is a crucial device to maintain rice growers in scheming disease. But the presence of false positives was not minimized. An Enhanced Weighted Sum Back Propagation (EWSBP) Algorithm was designed in Lohith *et al.* (2020) for leaf disease classification. Image processing provides an essential part for identifying plant disease. The designed algorithm offers an important and well-organized scheme for leaf

diseases. But the performance of time consumption was not analyzed. A generative adversarial network was introduced in Qiang *et al.* (2020) for crop leaf disease prediction. The designed model significantly minimizes the number of metrics. However, the designed network failed to perform the image denoising for accurate prediction.

Contribution

The major contribution of the proposed RPMSIR-JSBC technique on the contrary to the existing literature is given as follows, to increase the disease identification accuracy, a novel RPMSIR-JSBC technique is introduced based on four different processes namely preprocessing, segmentation, feature extraction, and classification.

To enhance the peak signal to noise ratio, the RPMSIR-JSBC uses the Regularized Perona-Malik filtering technique to preprocess the input rice plant leaf images to improve the image quality by removing the noises.

To minimize the disease identification time, the RPMSIR-JSBC technique performs the region-based segmentation using Hartigan-Wong clustering technique and feature extraction using weighted Isotonic Regression.

To enhance the accuracy and reduce the false positive rate, the RPMSIR-JSBC technique uses Jaccardized Rocchio Jensen-Shannon Boosting to consider the extracted features and classifies the image into the normal or disease.

Extensive experimental evaluation is carried out to evaluate the quantitative and qualitative analysis of the RPMSIR-JSBC technique with the various algorithms and different performance metrics.

Organization of the Work

The rest of the paper is organized as follows. Section 2 contains the related works. Section 3 contains the methodology (method). Section 4 contains the experimental results. Section 5 contains the conclusions and policy implications.

Related Works

A deep transfer learning model was developed in Junde *et al.* (2020a) for the identification of rice plant diseases. Though, harvester and gardening professionals are facing numerous agricultural challenges for the era. But, the deep transfer learning model failed to use the IoT devices to monitor and detect the wide range of plant disease information automatically. Mobile Net with squeeze-and excitation (SE) was designed in Junde *et al.* (2020b) for the successful classification of plant diseases.

SE-MobileNet is employed for detecting plant diseases. Deep learning enhanced the accuracy of image detection. But it failed to automatically recognize the wide-ranging information regarding plant diseases.

A new automatic detection and classification technique was developed in Chen *et al.* (2020) for plant leaf diseases. Plant disease identification is helpful to observe plants and plant leaves are the initial basis of plant disease identification. But, the designed technique was not efficient to perform accurate disease classification with minimum time consumption. An Improved Convolutional Neural Network was developed in Jie *et al.* (2020) for the classification of Plant Leaf Diseases. In order to detect and categorize plant leaf diseases, the deep learning-based method is used. Plant leaf diseases represent that the designed method is available and efficient. But, the image preprocessing remained unsolved to minimize the time consumption.

A high-quality generative adversarial network was designed in Bin *et al.* (2020) for the classification of grape leaf diseases. But, the accuracy of disease classification was not enhanced. The designed method offers sufficient results for data growth of grape leaf disease images. In order to produce grape leaf disease, a generator model by digressive channels is developed. An improved artificial plant optimization (IAPO) algorithm was developed in Deepak *et al.* (2020) for identifying plant diseases as healthy or infected. By employing a histogram of oriented gradients, Features are removed by images. But the designed algorithm failed to perform accurate disease identification.

A new maize leaf disease recognition technique was introduced in Mingjie *et al.* (2020) with higher accuracy. The designed technique produces a maize leaf feature improvement framework by improving the features of maize in complex surroundings. The designed technique failed to reduce the time consumption. An Advanced Segmented Dimension Extraction was developed in Sanjeevi *et al.* (2020) for the identification of plant leaf disease. The designed method is employed for detecting the effect on gesture leaf image, which is not important to identify time. However, the disease identification performance was not improved. A tomato leaf disease recognition method was introduced in Xiao *et al.* (2020) with higher accuracy and minimum loss. The designed model is simple to produce noise through image acquisition, and characteristics of dissimilar diseases are comparable, which makes image disease recognition is tricky. But the filtering-based leaf disease recognition was not performed. A Feed-Forward Neural Network was developed in Tan *et al.* (2020) for early disease identification of

mango leaves. In order to secure crops in time, there is necessitates for early disease detection. But the efficient feature selection algorithm was not used.

Proposal Methodology

Based on motivation, a novel proposed technique is developed for accurate disease identification with minimum time. RPMSIR-JJSBC technique consists of different image processing steps for recognition of diseases comprise acquisition of images, Regularized Perona-Malik filter-based image preprocessing, Hartigan-Wong clustered region-based Segmentation, Weighted isotonic regression-based feature extraction, and finally classification. The basic architecture of the plant leaf disease detection is shown in Figure 1.

Figure 1, describes the basic architecture of the proposed technique to identify the disease at earlier stages with the help of four different processes. In the acquisition process, numerous plant leaf images $I_1, I_2, I_3, \dots, I_n$ are captured through IoT devices and stored into the image dataset 'D'. Then the other processes are explained in detail in the following subsections.

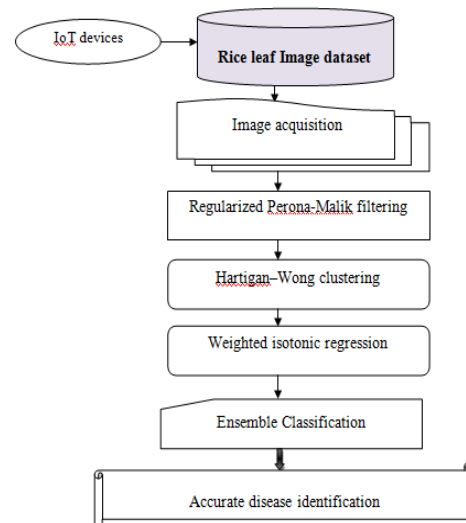


Figure 1. Architecture of proposed RPMSIR-JJSBC technique

Regularized Perona-Malik Filter-Based Image Preprocessing

Image pre-processing is the initial step of operations that are normally required before the image analysis. Preprocessing is the process of enhancing the image quality for the needs of disease identification. The main aim of pre-processing is noise suppression and removal. The proposed

RPMSIR-JJSBC technique uses the Regularized Perona-Malik filtering technique with the aiming at minimizing the image noise without removing significant parts of the image like edges, lines, or other details.

Let us consider, image dataset 'D' with total number of images $I_1, I_2, I_3, \dots, I_n$ where 'n' denotes the total number of images. The total number of pixels in each image is denoted by $p_1, p_2, p_3, \dots, p_m$. Then the pixels are arranged in the kernel window based on rows and columns to find the noisy pixels.

$$w = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & p_9 \end{bmatrix} \quad (1)$$

where w indicates a 3×3 kernel window. Then the pixels are sorted into ascending order and take the center pixel value is taken as new pixels. The Regularized Perona-Malik filtering convolved with a Gaussian inside the non-linearity to obtain the denoised image.

$$\frac{\partial I}{\partial t} = \text{div}(a(|D\beta * I|)\nabla I) \quad (2)$$

$$\text{where, } \beta = \frac{1}{2\pi d^2} \exp\left(-\frac{\|p_i - p_c\|^2}{2d^2}\right) \quad (3)$$

From (2), ∇I denotes the gradient version of an image to preserve the edges in the image, I denotes a sample original input image, div be a divergence operator, a diffusion coefficient to preserve edges in the input image, D denotes a conductance function which manages the diffusion strength, β Gaussian distribution. From (3), d denotes a deviation, p_i denotes a pixel in kernel window, p_c denotes a center pixel in the kernel window. The maximum likelihood between the center and neighboring pixels is denoted as normal pixels and the pixel that diverged from the center pixel is known as noisy pixels. The noisy pixels are removed from the filtering window. As a result, the quality of input images gets enhanced. Hence, it increases the peak signals to noise ratio. The algorithmic process of Regularized Perona-Malik filter-based image preprocessing is explained as given below,

//Algorithm 1: Regularized Perona-Malik filter-based image preprocessing

Input: Image dataset 'D' with number of images $I_1, I_2, I_3, \dots, I_n$, pixels $p_1, p_2, p_3, \dots, p_m$

Output: Enhance the image quality

Begin

1. Collect plant leaf images $I_1, I_2, I_3, \dots, I_n$ as input
2. For each plant leaf images I_i

3. Sort the pixels $p_1, p_2, p_3, \dots, p_m$ in the filtering window
 4. Take the center value p_c
 5. Measure the likelihood between the center and neighboring pixels ' β '
 6. Find the noisy pixels
 7. Remove noisy pixels from the filtering window
 8. Obtain quality enhanced image
 9. End for
- End**

Algorithm 1 depicts the step-by-step process of Regularized Perona-Malik filter to remove the noisy pixels from the input image. The plant leaf images are collected from the dataset. The pixels in the images are arranged and take the center pixels into the filtering window. The Gaussian distribution is applied to measure the likelihood among the center pixel and neighboring pixels in the window. Based on the likelihood measure, the noisy pixels are removed from the window. Therefore, the quality of images is improved for accurate plant leaf disease prediction.

Hartigan-Wong Clustered Region-Based Segmentation

Image segmentation is the significant process in the proposed technique for partitioning a whole image into parts, called segments. The main aim of the segmentation is to minimize the time consumption of plant disease identification. The proposed image Hartigan-Wong clustering-based segmentation techniques partition the image into several parts depends on certain picture features like pixel intensity value. The Hartigan-Wong clustering is a machine learning technique to partition the image into different regions based on the similarity of pixels characteristics.

Figure 2 illustrates the block diagram of the Hartigan-Wong clustering based on image segmentation. First of all the preprocessed plant leaf

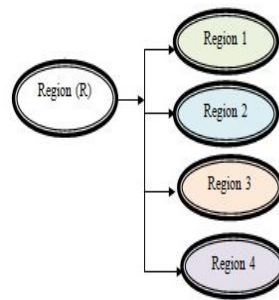


Figure 2. Segmented regions based on Hartigan-Wong clustering

image regions (q) are partitioned into several regions $q_k \in \{q_1, q_2, q_3, q_4\}$ based on the pixel similarity. The Hartigan–Wong clustering technique aims to partition image regions (R) observations into clusters based on the nearest mean. Let us consider the number of pixels in an image $p_1, p_2, p_3, \dots, p_m$. The similarity between the pixels is measured in terms of Euclidean distance.

$$Z_{ij} = \|p_i - p_j\|^2 \quad (3)$$

$$Y = \arg \min Z_{ij} \quad (4)$$

where, Y specifies a clustering output, p_i denotes a pixel in the preprocessed image, p_j denotes an adjacent pixel. From (4), $\arg \min$ represents an argument of the smallest function, $Z_{ij} = \|p_i - p_j\|^2$ designates a distance between the pixels which denotes the two pixels are statistically dependent or independent. Similarly, the distances between the pixels in the input image are calculated. Based on a distance measure, similar pixels are grouped into a particular region.

//Algorithm 2: Hartigan-Wong clustered region-based Segmentation

Input: Preprocessed image, pixels $p_1, p_2, p_3, \dots, p_m$

Output: Obtain the multiple regions

Begin

1. For each preprocessed plant leaf images I_i
2. For each pixel p_i
3. For each neighboring pixel p_j
4. Measure the distance between the pixels Z_{ij}
5. Find minimum distance $Y = \arg \min Z_{ij}$
6. Group the similar pixels
7. Obtain different regions of the image
8. End for
9. End for

End

Algorithm 2 provides the step-by-step process of segmentation to obtain the segmented regions. For each preprocessed image and image pixels, the distance is measured. The minimum distance between the pixels is grouped into particular regions. As a result, multiple regions are obtained and find the region of interest part i.e. infected area.

Weighted Isotonic Regression-Based Feature Extraction

The feature extraction process is carried out after the segmentation phase in the proposed technique. The feature plays a very important role in image processing for detecting the disease with minimum time and maximum accuracy. In general, input images consist of two types of feature vectors

such as local and global. The local features of the images are point, edge, and small image patch. Whereas the global features give visual content of the image by a distinct vector. These features are texture, color, shape information. Processing the multiple features causes more complexity in disease identification. Therefore, the proposed weighted isotonic regression function finds the best features for accurate disease identification with minimum time.

The weighted isotonic regression is a machine learning technique used for finding weighted least-squares fitted features for disease identification. Let us consider the ‘b’ feature vector $F_i \in [f_1, f_2, f_3, f_b]$. The regression function analyzes the feature vectors and finds the fittest feature vectors from the original feature vector ‘ F ’. The weighted isotonic regression function solves the following constraints.

$$R = \min \sum_{i=1}^n \vartheta_i [F_i - o]^2 \quad (5)$$

where, R denotes a regression output, ϑ_i denotes a weight vector, F_i denotes a feature vector, ‘ o ’ denotes an objective function (i.e. disease identification). From (5), $\min$ denotes a minimum function used for finding the minimum weighted least-squares features. In the proposed technique, color, texture, shape that features are selected and extracted.

The color features are visual features of the human eye. There are three orders are used for representing the color distributions such as first-order (mean i.e. hue), second (standard deviation i.e. saturation), and third-order (skewness). The first order of the color distributions is given below,

$$\mu = \frac{\sum p_{iq}}{m} \quad (6)$$

where μ designates the mean, p_{iq} symbolizes the pixel intensity, m designates a total number of pixels. The pixel intensity is given as follows,

$$p_{iq} = \sum_i \sum_j \|p_i - p_j\|^2 \quad (7)$$

From (7), Where p_i, p_j denotes two pixels. The first order of the color distribution i.e. standard deviation is given by,

$$\sigma = \sqrt{\frac{\sum [(p_{iq} - \mu)^2]}{m}} \quad (8)$$

From (8), σ represents the standard deviation, p_{iq} denotes the pixel intensity, μ specifies a mean. The third order i.e. skewness is given below,

$$\varphi = \frac{1}{m} \sum \left[\frac{(p_{iq} - \mu)}{m} \right]^3 \quad (9)$$

In (9), φ denotes the skewness. With the estimated mean, standard deviation, and skewness, the color features are extracted.

The texture features are extracted and it provides information about the spatial representation of colors or intensities.

$$Tx = \frac{\sum_i \sum_j [p_i - \mu_i][p_j - \mu_j]}{\sigma_i * \sigma_j} \quad (10)$$

From (10), Tx designates the texture feature, μ_i and μ_j indicate a mean of the pixels p_i, p_j and deviations are σ_i and σ_j .

The other visual feature of the object is called shapes that are extracted based on the contour representation. The image region is transformed into a two-dimensional space. After the transformation, the center location of the leaf is the origin $(0,0) = (a_1, b_1)$ and the boundary of the object is denoted as (a_2, b_2) . The distance between these two points is measured as given below,

$$Dist = \sqrt{(a_2 - a_1)^2 - (b_2 - a_1)^2} \quad (11)$$

Similarly, all the distance is measured for all the boundary points from the origin. As a result, the perfect outline of the object is determined.

//Algorithm 3: Weighted isotonic regression based feature extraction

Input: Segmented regions $q_k \in \{q_1, q_2, q_3, \dots, q_n\}$

Output: Extract the global features

Begin

1. For each segmented region
2. Construct the feature vectors $F_i \in [f_1, f_2, f_3, \dots, f_b]$
3. For each feature in F_i
4. Perform regression analysis $R = \min \sum_{i=1}^n \vartheta_i [F_i - o]^2$
5. Find global features
6. Extract the features color, texture, shape
7. End for
8. End for

End

Algorithm 3, illustrates the step-by-step process of feature selection and extraction from the segmented images. The Weighted isotonic regression is used to analyze the feature vector and find the fittest features related to the objective from the original feature vector. Followed by, the color, shape, and texture features are extracted through the statistical measure. The feature selection and extraction process of the proposed technique reduce the time consumption of the disease identification.

Jaccardized Rocchio Jensen-Shannon Boosting Classification

Finally, the proposed RPMSIR-JJSBC technique performs the classification to detect the disease based on the extracted features. To classify the input leaf images into the normal or abnormal (i.e. disease), the proposed technique uses the Jensen-Shannon Boost technique. The sensor node clustering is performed using Jensen-Shannon Boost Node Clustering technique. The designed Jensen-Shannon Boost technique is an ensemble technique to enhance the performance of the machine learning algorithm by converting the set of the weak learner into strong. The weak learner is a base classifier and it provides accurate classification results. In contrast, an ensemble technique gives accurate classification results by combining the set of weak learners.

Figure 3 depicts the construction of a Jensen-Shannon clustering technique to classify the images with higher accuracy. The ensemble technique uses a training set $\{(F_i, Y_i)\}$ where ' S_i ' denotes the extracted feature vector $F_E \in [f_1, f_2, f_3, \dots, f_b]$ ' Y ' denotes ensemble clustering results. As shown in Figure 2, the ensemble technique initially constructs ' v ' weak learners as a Jaccardized Rocchio classifier to classify the given input samples as disease or not based on a similarity measure between the extracted features and leaf disease features. The classifier uses the Jaccard similarity as give below,

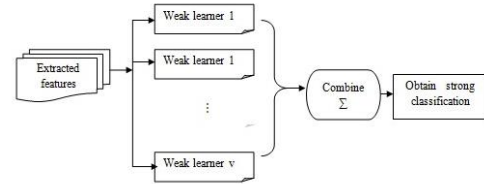


Figure 3. Construction of Jensen-Shannon Boost

$$\rho = \frac{F_E \cap F_d}{\sum F_E + \sum F_d - F_E \cap F_d} \quad (12)$$

From (12), ρ denotes a Jaccard similarity coefficient, F_E represents extracted feature, F_d denotes a leaf disease features. $F_E \cap F_d$ denotes a mutual dependence between the features. $\sum F_E$ is the sum of the F_E score, $\sum F_d$ is the sum of the F_d score. The Jaccard similarity coefficient (ρ) provides the similarity value between 0 and 1 ($0 \leq \rho \leq 1$). The higher similarity is classified as disease-affected. Otherwise, it is classified as a normal leaf. In this way, the weak learner identifies the disease-affected leaf. But, all the weak learners lack providing the exact classification. In order to solve

these issues, the ensemble technique combines all the weak learners.

$$Y = \sum_{i=1}^v z_i \quad (13)$$

where ‘Y’ denotes strong classification results, z_i indicates an output of the weak learner. For each weak learner, the influence is initialized

$$Y = \sum_{i=1}^v \omega z_i \quad (14)$$

From (14), ‘ ω ’ indicates the weight in terms of a random integer number. After that, the ensemble technique measures Jensen-Shannon divergence for each weak learner based on the squared difference among actual and predicted results.

$$Div = [act_{res} - pred_{res}]^2 \quad (15)$$

where, Div symbolizes the divergence, act_{res} indicates actual results of the weak learner, $pred_{res}$ symbolizes classification results. The minimum divergence has lesser error whereas the higher divergence represents the weak learner output has the high error. Based on the error value, influence gets updated for each weak learner. Finally, output with minimum error is taken as strong classification results. The algorithmic process of Jaccardized Rocchio Jensen-Shannon Boosting classification is described as given below.

//Algorithm 4: JaccardizedRocchio Jensen-Shannon Boosting classification

Input: Extracted features $F_E \in [f_1, f_2, f_3, \dots, f_b]$

Output: Improve the disease identification accuracy

Begin

1. for the extracted features F_E
2. Construct the ‘v’ weak learners
3. Measure similarity ‘ ρ ’
4. If (arg max ρ) then
5. Image is classified as ‘abnormal’
6. else
7. Image is classified as ‘normal’
8. end if
9. Combine all weak classifier $Y = \sum_{i=1}^v z_i$
10. Assign weights ‘ ω ’
11. For each z_i
12. Measure Jensen-Shannon divergence
13. End for
14. Update the weight to weak learners
15. Find weak classifier with minimum error
16. Obtain strong classification results
17. End for

End

Algorithm 4 illustrates the step-by-step process of disease classification using the Jaccardized Rocchio Jensen-Shannon Boosting technique. The ensemble technique constructs ‘v’ weak learners as a Jaccardized Rocchio classifier. The weak learner measures the similarity among the extracted features and disease features using the Jaccard index. The high similarity between the features is classified as ‘abnormal’ or disease-affected. Otherwise, the image is classified as normal. The set of weak learner results are transformed and assign a similar weight. For each weak learner, the divergence is measured among the actual and predicted results. Followed by, the initial weight gets updated. The ensemble learning classifier finds the weak learner with minimum error. This helps to enhance the disease identification accuracy and reduces the false positive rate.

Experimental Settings

Experimental evaluation of the proposed technique and existing methods namely hybrid BFOA-DNN (Aakrati *et al.*, 2020) Simple CNN (Rahman *et al.*, 2020) are implemented using MATLAB with the help of the Rice Diseases Image Dataset taken from the <https://www.kaggle.com/minhhuy2810/rice-diseases-image-dataset>. The dataset comprises the 5447 rice leaf image dataset. Here, 500 leaf image datasets are evaluated. The dataset comprises four different files are Labeled as BrownSpot, Healthy (i.e. normal), Hispa, LeafBlast. From the dataset, the leaf images are pre-processed using Regularized Perona-Malik filtering technique by reducing image noise without removing important parts of the image. Then the Hartigan-Wong clustering method is applied to perform the region-based segmentation to partition images into various regions on the similarity of pixels characteristics. Weighted isotonic regression is applied to perform the feature extraction which plays a significant role in image processing to identify disease by lesser time and highest accuracy and finally, the Jaccardized Rocchio Jensen-Shannon Boosting technique is applied to perform classification to find healthy or disease-affected images. The statistical test is conducted by various parameters such as peak to noise signal ratio, disease detection accuracy, disease detection time, false-positive rate, etc.

Results and Discussion

In this section, the performance of the proposed technique and existing methods namely hybrid BFOA-DNN (Aakrati *et al.*, 2020), Simple CNN

(Rahman *et al.*, 2020) are discussed based on a qualitative and quantitative manner. The qualitative performance analysis is performed to show the visual representation of the four different processes namely preprocessing, segmentation, feature extraction, and classification. The quantitative performance analysis is done with the help of different parameters.

Qualitative Performance Analyses

In this subsection, the qualitative analysis of the RPMSIR-JSBC is discussed with various processes.

Figure 4 given above depicts the qualitative analysis of the RPMSIR-JSBC technique. Firstly, the input rice leaf image is collected from the Rice Diseases Image Dataset. The Regularized Perona-Malik filtering technique is applied to enhance the image contrast. Then the Hartigan–Wong clustering method is applied for region-based segmentation. Thirdly, the weighted isotonic regression is applied

to perform the feature extraction and lastly, the Jaccardized Rocchio Jensen-Shannon Boosting technique is applied for classifying the disease with greater accuracy.

Quantitative Performance Analyses

Quantitative analysis of proposed and existing methods are discussed with different parameters such as Peak Signal to noise ratio, disease identification accuracy, false-positive rate, and disease identification time with respect to variant numbers of plant leaf images and variant sizes.

Impact of Peak Signal to noise ratio

Peak signal to noise ratio is measured on the mean square error and it is referred to as the difference among original leaf images and quality enhanced images after preprocessing. The peak signal to noise ratio is given as follows,

$$PSNR = 10 * \log_{10} \left[\frac{p^2}{E_{MS}} \right] \quad (16)$$

$$E_{MS} = [I_o - I_p]^2 \quad (17)$$

where, $PSNR$ represents a peak signal-to-noise ratio, p represents the highest possible pixel value (i.e. 255), E_{MS} is a mean square error, ' I_o ' represents an original image size, I_p represents a pre-processed image size. Therefore, the PSNR is measured in decibel (dB).

Sample Calculation

Proposed RPMSIR-JSBC

Original leaf images 0.98 and the preprocessed image size is 0.9. Therefore mean square error is calculated as follows,

$$E_{MS} = [I_o - I_p]^2 = [0.98 - 0.9]^2 = 0.0064$$

Hence, peak to noise is measured by,

$$PSNR = 10 * \log_{10} \left[\frac{p^2}{E_{MS}} \right] = 10 * \log_{10} \left[\frac{255^2}{0.0064} \right] = 70.06$$

Existing hybrid BFOA-DNN

Original leaf images 0.98 and the preprocessed image size is 0.87. Therefore mean square error is calculated as follows,

$$E_{MS} = [I_o - I_p]^2 = [0.98 - 0.87]^2 = 0.0121$$

Hence, peak to noise is measured by,

$$PSNR = 10 * \log_{10} \left[\frac{p^2}{E_{MS}} \right] = 67.30$$

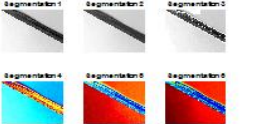
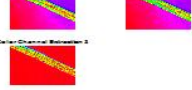
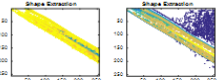
Processes	Results
1. Input rice leaf image	
2. Regularized Perona-Malik filtering technique	Image Quality Improve 
3. Hartigan-Wong clustered region-based segmentation	Segmentation 1 Segmentation 2 Segmentation 3 Segmentation 4 Segmentation 5 Segmentation 6 
4. Weighted isotonic regression	Color-Channel Extraction 1 Color-Channel Extraction 2 Color-Channel Extraction 3  (a) Color feature extraction
	Texture Extraction Texture Extraction Texture Extraction Texture Extraction  (b) Texture feature extraction
	Shape Extraction Shape Extraction  (c) shape feature extraction
5. Jaccardized Rocchio Jensen-Shannon Boosting classification based disease identification	Classification Detection Area 

Figure 4. Qualitative results of RPMSIR-JJBC

Existing Simple CNN

Original leaf images 0.98 and the preprocessed image size is 0.85. Therefore mean square error is calculated as follows,

$$E_{MS} = [I_o - I_p]^2 = [0.98 - 0.85]^2 = 3847633$$

Hence, peak to noise is measured by,

$$PSNR = 10 * \log_{10} \left[\frac{p^2}{E_{MS}} \right] = 65.85$$

Table 1 illustrate the simulation results of peak signal to noise ratio with sizes of the various leaf images. The observed results illustrate that an RPMSIR-JJSBC technique provides improved performance in achieving higher accuracy than the existing methods. From the observed simulation results, the proposed technique achieves a higher peak signal-to-noise ratio than the existing hybrid BFOA-DNN (Aakrati *et al.*, 2020 and Simple CNN (Rahman *et al.*, 2020). This is proved through statistical calculation. Let us consider the 0.98MB sizes of the plant leaf images. The peak signal to noise ratio of the proposed technique is 70.06 dB and the peak signal to noise ratio of existing hybrid BFOA-DNN (Aakrati *et al.*, 2020) Simple CNN (Rahman *et al.*, 2020) are 67.30 dB and 65.85 dB respectively. For each method, ten runs are carried out with variant sizes of the input. The obtained results of the RPMSIR-JJSBC technique are compared to the existing methods. The average of comparison results represents that the overall peak signal to noise ratio is minimized using the proposed technique by 4% when compared to hybrid BFOA-DNN (Aakrati *et al.*, 2020) and 6% when compared to Simple CNN (Rahman *et al.*, 2020).

Figure 5 describes the comparative analysis of peak signal to noise ratio based on the number of variant sizes of leaf images using three different classification techniques. As shown in the graphical plot, the variant image sizes are taken as input in the horizontal axis and the Peak signal to noise ratio results are taken as output in the vertical axis. The peak signal to noise ratio of the three methods namely the proposed RPMSIR-JJSBC technique and existing BFOA-DNN (Aakrati *et al.*, 2020) and Simple CNN (Rahman *et al.*, 2020) are represented by three different colors of lines such as violet, red and green respectively. Among three methods, the RPMSIR-JJSBC technique provides enhanced performance. The reason behinds this improvement is the application of Regularized Perona-Malik filter to find and eliminate the noisy pixels from the input rice leaf image. The accurate plant leaf disease prediction is enhanced for image

quality. By applying the Gaussian distribution, the noisy pixels in the window are identified and removed. As a result, the image quality gets improved and minimizes the mean square error, and maximizes the peak signal to noise ratio.

Impact of Disease identification accuracy

Disease identification accuracy is defined as the ratio of the number of plant images that are correctly classified into a normal or disease to the total number of leaf images. Therefore, the overall accuracy is as given below,

$$DIA = \left[\frac{NCC}{n} \right] * 100 \quad (18)$$

where *DIA* denotes a Disease identification accuracy, '*n*' represents the number of plant leaf images, *NCC* indicates the number of images the overall accuracy is calculated in percentage (%).

Table 1. Peak signal to noise ratio

Image Size (MB)	Peak signal to noise ratio (dB)		
	RPMSIR	Hybrid BFOA-DNN	Simple CNN
0.98	70.06	67.30	65.85
1.11	69.04	65.20	64.04
1.24	72.56	70.06	68.13
1.43	74.15	71.22	69.04
1.46	71.22	68.13	67.30
1.54	76.08	74.15	72.56
1.79	74.15	71.22	69.04
2.40	68.13	66.54	65.20
3.01	67.30	65.85	63.52
3.50	62.11	60.17	59.18

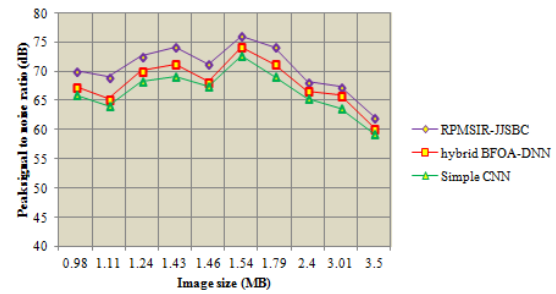


Figure 5. Comparison analysis of peak signal to noise ratio

Sample Calculation

Proposed RPMSIR-JJSBC

Number of plant images that are correctly classified by 25 and the total number of leaf images are 21. Therefore disease identification accuracy is calculated as follows,

$$DIA = \left[\frac{NCC}{n} \right] * 100 = \left[\frac{21}{25} \right] * 100 = 84$$

Existing hybrid BFOA-DNN

Number of plant images that are correctly classified by 25 and the total number of leaf images are 20. Therefore disease identification accuracy is calculated as follows

$$DIA = \left[\frac{NCC}{n} \right] * 100 = \left[\frac{20}{25} \right] * 100 = 80$$

Existing hybrid BFOA-DNN

Number of plant images that are correctly classified by 25 and the total number of leaf images are 19. Therefore disease identification accuracy is calculated as follows

$$DIA = \left[\frac{NCC}{n} \right] * 100 = \left[\frac{19}{25} \right] * 100 = 76$$

Table 2 reports the experimental results of disease identification accuracy with a number of rice leaf images taken in the ranges from 25 to 250. Accuracy is measured as the algorithm correctly identified the disease from the rice leaf images. The compared results indicate that the RPMSIR-JJSBC technique provides better performance in terms of achieving higher accuracy of disease identification. As shown in Table 2, '25' numbers of rice leaf images are considered for simulation and the number of images correctly classified is 21 using RPMSIR-JJSBC technique and accuracy is 84%. Whereas, the accuracy of conventional technique BFOA-DNN (Aakrati *et al.*, 2020) and Simple CNN (Rahman *et al.*, 2020) are 80% and 76% with the application of similar input i.e. 25 images. Similarly, other runs are performed with the number of images. From the analysis, the disease identification accuracy is found to be minimum using RPMSIR-JJSBC technique than BFOA-DNN (Aakrati *et al.*, 2020) and Simple CNN (Rahman *et al.*, 2020). The average of ten comparison results of the RPMSIR-JJSBC technique indicates that the disease identification accuracy is found to be increased by 4% when compared to (Aakrati *et al.*, 2020) and 8% when compared to (Rahman *et al.*, 2020) respectively.

Figure 6 illustrates the Disease identification accuracy of three different techniques with respect to 250 different images with 10 runs. Ten different runs are performed to measure the disease identification accuracy. This is because of the Jaccardized Rocchio Jensen-Shannon Boosting technique for analyzing the extracted features and disease features based on the Jaccard similarity index. Based on the similarity value, the disease is identified. Disease identification accuracy is improved and the false positive rate is lesser. The ensemble technique combines the weak learner and

finds the best classification results based on divergence measures.

Impact of False-positive rate

The false-positive rate is defined as the ratio of the number of plant leaf images that are wrongly classified to the total number of images. The false-positive rate is given as follows,

$$R_{fp} = \left[\frac{NWC}{n} \right] * 100 \quad (14)$$

where R_{fp} indicates the false positive rate, 'n' represents the number of images, NWC denotes the number of images wrongly classified. The false-positive rate is measured in percentage (%).

Table 2. Disease identification accuracy

Number of images	Disease identification accuracy (%)		
	RPMSIR-JJSBC	Hybrid BFOA-DNN	Simple CNN
25	84	80	76
50	88	84	80
75	91	87	83
100	90	86	84
125	91	88	85
150	90	85	83
175	89	86	82
200	90	87	84
225	91	88	85
250	89	86	84

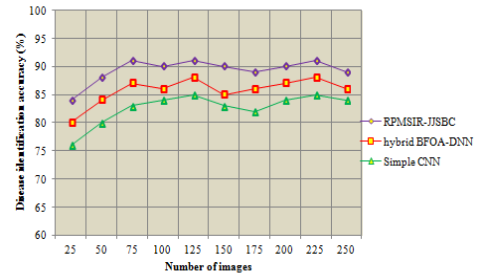


Figure 6. Comparison analysis of disease identification accuracy

Sample Calculation

Proposed RPMSIR-JJSBC

Number of plant images that are correctly classified by 25 and the total number of leaf images are 4. Therefore false positive rate is calculated as follows,

$$R_{fp} = \left[\frac{NWC}{n} \right] * 100 = \left[\frac{4}{25} \right] * 100 = 16$$

Existing hybrid BFOA-DNN

Number of plant images that are correctly classified by 25 and the total number of leaf images

are 5. Therefore false positive rate is calculated as follows

$$R_{fp} = \left[\frac{NWC}{n} \right] * 100 = \left[\frac{5}{25} \right] * 100 = 20$$

Existing hybrid BFOA-DNN

Number of plant images that are correctly classified by 25 and the total number of leaf images are 6. Therefore false positive rate is calculated as follows

$$R_{fp} = \left[\frac{NWC}{n} \right] * 100 = \left[\frac{6}{25} \right] * 100 = 24$$

Table 3 provides the experimental results of the false-positive rate with respect to the total number of rice leaf images. The tabulated results demonstrate that the RPMSIR-JJSBC technique provides a lesser false positive rate than the existing methods. In the first run, the experiment is conducted with 25 images, the false-positive rate of RPMSIR-JJSBC is 16%. Besides, the false-positive rate of BFOA-DNN (Aakrati *et al.*, 2020 and Simple CNN (Rahman *et al.*, 2020) are 20% and 24% respectively. Identically, the other runs are performed with different counts of input. The obtained results of the RPMSIR-JJSBC are compared to the results of existing methods. The average results noticed that the false-positive rate is considerably reduced by 25% and 39% when compared to the existing BFOA-DNN (Aakrati *et al.*, 2020) and Simple CNN (Rahman *et al.*, 2020) respectively.

Figure 7 illustrates the false positive rate for different numbers of leaf images in the range of 25 to 250. From the graphical plot, it is inferred that the false positive rate is significantly reduced using the RPMSIR-JJSBC technique than the existing methods. This is due to the reason that the ensemble technique uses the Rocchio classifier as a weak learner. The set of weak learner results are transformed and assign a similar influence. For each weak learner, the Jensen-Shannon divergence is measured between the actual and predicted disease identification results. The original influence gets updated. Ensemble learning classifier discovers the weak learner with minimum error. This helps to enhance the accuracy as well as minimize the false positive rate.

Impact of Disease identification time

Disease identification time is referred to as the amount of time taken by an algorithm for identifying the disease through classification. The time consumption is mathematically calculated as given below,

$$DIT = [n] * T (\text{classify one image}) \quad (15)$$

where, *DIT* denotes a Disease identification time, '*n*' represents the number of leaf images, *T* denotes time consumption to classify a single image. The computational time is measured in milliseconds (ms).

Sample Calculation

Proposed RPMSIR-JJSBC

The amount of time taken by an algorithm is 25 and detects the disease through classification is 0.8. Therefore disease identification time is calculated as follows,

$$DIT = [n] * T (\text{classify one image}) = [25] * 0.8 = 20$$

Table 3. False positive rate

Number of images	False positive rate (%)		
	RPMSIR-JJSBC	Hybrid BFOA-DNN	Simple CNN
25	16	20	24
50	12	16	20
75	09	13	17
100	10	14	16
125	09	12	15
150	10	15	17
175	11	14	18
200	10	13	16
225	09	12	15
250	11	14	16

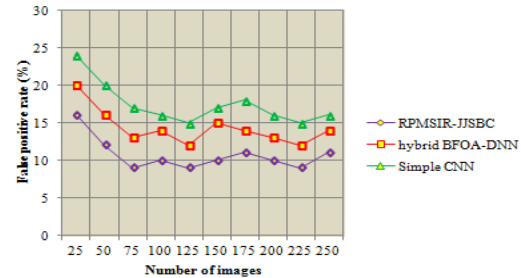


Figure 7. Comparison analysis of the false-positive rate

Existing hybrid BFOA-DNN

The amount of time taken by an algorithm is 25 and detects the disease through classification are 0.88. Therefore disease identification time is calculated as follows,

$$DIT = [n] * T (\text{classify one image}) = [25] * 0.88 = 22$$

Existing hybrid BFOA-DNN

The amount of time taken by an algorithm is 25 and detects the disease through classification are

0.97. Therefore disease identification time is calculated as follows,

$$DIT = [n] * T (\text{classify one image}) = [25] * 0.97 = 24$$

Table 4 describes the disease identification time based on the number of images. The numbers of images are taken as input in the ranges 25 to 250. The RPMSIR-JJSBC technique reduces the identification time in the disease diagnosis compared to existing classification techniques. For example, with 25 images, the disease identification time of a single image is '0.8 ms' using the RPMSIR-JJSBC technique, whereas the disease identification is '0.88 ms' using hybrid BFOA-DNN (Aakrati *et al.*, 2020) and time for identifying the single image being '0.97 ms' using Simple CNN (Rahman *et al.*, 2020). Therefore, the overall disease identification time was found to be '20 ms', '22 ms', and '24 ms' using RPMSIR-JJSBC technique, hybrid BFOA-DNN (Aakrati *et al.*, 2020), and Simple CNN (Rahman *et al.*, 2020) respectively. The above statistical analysis proves that the disease identification time is minimized using the RPMSIR-JJSBC technique. The ten different results reveal that the RPMSIR-JJSBC technique minimizes the disease identification time by 8% when compared to hybrid BFOA-DNN (Aakrati *et al.*, 2020) and 14% as compared to Simple CNN (Rahman *et al.*, 2020). The ten different results are plotted in the graph.

Figure 8 illustrates the disease identification time for three different techniques with respect to 250 different numbers of rice leaf images. The plot inferred that increasing number of leaf images causes, therefore, results in the increasing disease identification time. From the evaluation, it is found to the disease identification time is minimum using the RPMSIR-JJSBC technique than (Aakrati *et al.*, 2020; Rahman *et al.*, 2020). The reason behind the improvement is due to the incorporation of segmentation and feature extraction processes. To partition input images into different regions, Hartigan-Wong clustering technique is applied. Plant disease identification of time consumption is lesser. To find the best features and extracted from the segmented infected region, the weighted Isotonic Regression is applied. The color, shape, and texture features are extracted to reduce the time consumption of the disease identification. Lastly, the Jaccardized Rocchio Jensen-Shannon Boosting technique is applied for analyzing the extracted features with the disease features. As a result, the time consumption for disease identification from the rice leaf image is minimized.

Conclusions

In the agriculture field, the loss of yield generally happened due to the rising speed of the leaf disease and it also cause the loss of yield, time, and wealth. Therefore, early detection of the disease is essential to improve crop yield. An efficient technique called RPMSIR-JJSBC is introduced for accurate disease identification. In the proposed RPMSIR-JJSBC, the classification and identification of rice plant leaf diseases are carried out with variant processing methods. To get a quality enhanced image using Regularized Perona-Malik filtering technique, the rice leaves are captured and preprocessed. After the preprocessing, the region-based segmentation is carried out to partition the images into various segments using Hartigan-Wong clustering technique. Then the weighted Isotonic Regression is applied to perform the feature selection and extraction to minimize the disease identification time. Finally, the Jaccardized Rocchio Jensen-Shannon Boosting technique is applied for analyzing the extracted features with higher accuracy. The comprehensive experimental evaluation is performed with different rice leaf images dataset. The qualitative and quantitative analysis confirms that the proposed technique has higher disease identification accuracy and lesser time as well as false-positive rate when compared to other conventional methods.

Table 4. Disease identification time

Number of images	Disease identification time (ms)		
	RPMSIR-JJSBC	Hybrid BFOA-DNN	Simple CNN
25	20	22	24
50	24	28	31
75	30	32	35
100	34	36	40
125	38	41	44
150	41	44	47
175	44	47	49
200	46	49	51
225	48	52	54
250	50	53	55

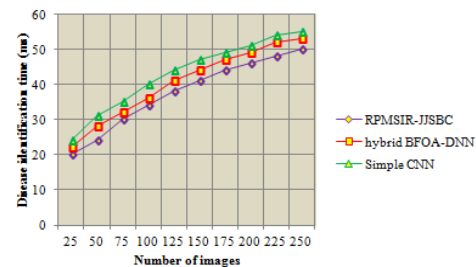


Figure 8. Comparison analysis of disease identification time

References

- Aakrati, N., Avdhes, K.T., and Akhilesh P. (2020). Paddy leaf diseases recognition and classification using PCA and BFODNN algorithm by image processing. In: International Conference on Nanotechnology: Ideas, Innovation, and Industries, Netherlands (NLD), Materials Today: Proceedings, p. 4,856-4,862.
- Bin, L., Cheng, T., Shuqin L., Jinrong H., and Hongyan W. (2020). A data augmentation method based on generative adversarial networks for grape leaf disease identification. IEEE Access, 8:102,188-102,198.
- Chen, J., Yin H., and Zhang, D. (2020). A self-adaptive classification method for plant disease detection using GMDH-Logistic model. Sustain. Comp. Inform. Sys., 28:100,415
- David, F.N., Dimitrios K., Argyris K., Natasa S., Pau. P., and Roberto, C. (2019). Predicting rice blast disease: machine learning versus process-based models. BMC Bioinform., 20: 1-16.
- Deepak, G., Purna, S., Krishna, C., Kshitij, G., Rahul, C., Ashish, K., Victor H.C. d. A. (2020). Artificial plant optimization algorithm to detect infected leaves using machine learning Expert Systems, (Special Issue Paper): e12490.
- Guoxiong, Z., Wenzhuo, Z., Aibin, C., Mingfang, H., and Xueshuo, M. (2019). Rapid detection of rice disease based on FCM-KM and faster R-CNN fusion. IEEE Access., 7:143,190-143,206.
- Jie, H., Dexiang, Z., Peng, C., Jun, Z., and Bing, W. (2019). Classification of plant leaf diseases based on improved convolutional neural network. Sensors, 19(19):4,161.
- Junde, C., Defu Z., Md, S., Yaser, A.N., and Yuandong, S. (2020a). Identification of plant disease images via a squeeze-and-excitation MobileNet model and twice transfer learning. IET Image Processing, 15(5):1,115-1,127.
- Junde, C., Defu, Z., Yaser, A.N., and Dele, L. (2020b). Detection of rice plant diseases based on deep transfer learning, J. Sci. Food Agri., 100(7):3,246-3,256.
- Lohith, K.P., Vinay K.K.G., Vasanth K.G., and Shijin K.P.S. (2020). Enhanced weighted sum back propagation neural network for leaf disease classification. In: Materials Today: Proceedings, p. 1-8.
- Mingjie, L., Guoxiong, Z., Mingfang, H., Aibin, C., Wenzhuo, Z., and Yahui, H. (2020). Maize Leaf disease identification based on feature enhancement and DMS-robust alexnet. IEEE Access, 8:57,952-57,966.
- Monalisa, M., Prasenjit, C., and Bibudhendu P. (2021). Modified ride NN optimizer for IoT-based plant disease detection. J. Ambi. Intellig. Human. Comp., 12(6):691-703.
- Qiang, D., Xi, C., Yan, Q., and Youhua, Z. (2020). Crop leaf disease image superresolution and identification with dual attention and topology fusion generative adversarial network. IEEE Access, 8:55,724-55,735.
- Rahman, C.R., Arko, P.S., Ali, M.E., Khan, M.A.I., Apon, S.H., Nowrin, F., and Wasif, A. (2020). Identification and recognition of rice diseases and pests using convolutional neural networks. Biosys. Eng., 194:112-120.
- Sachin, B.J. Vishwanath R.U., and Sanjay B.P. (2020). Identification of plant diseases using convolutional neural networks. Int. J. Inform. Technol., p. 1-10.
- Sanjeevi, P., Ashwin, M., Manikandan, R., Karthick, R.K.M., and Anantha, R.G.R. (2020). Heterogeneous internet of things organization predictive analysis platform for apple leaf diseases recognition. Comp. Commun., 154:99-110.
- Tan, N.P., Ly, V.T., and Son, V.T.D. (2020). Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection. IEEE Access, 8:189,960-189,973.
- Toran, V. and Sipi D. (2019). Fuzzy-filtered neural network for rice disease diagnosis using image analysis. Int. J. Inno. Technol. Explo. Eng. (IJITEE), 8(8S3):437-446.
- Vikas, S., Aftab, A.M., and Dr. Abid, S. (2020). Detection of rice disease using bayes' classifier and minimum distance classifier. J. Multimed. Inform. Sys., 7(1):17-24.
- Xiao, C., Guoxiong Z., Aibin, C., Jizheng, Y., Wenzhuo, Z., Yahui, H. (2020). Identification of tomato leaf diseases based on the combination of ABCK-BWTR and B-ARNet. Comp. Elect. Agri., 178:105,730.