A survey on recognition and classification of paddy leaf diseases using image processing and machine learning techniques

Cite as: AIP Conference Proceedings **2463**, 020009 (2022); https://doi.org/10.1063/5.0080336
Published Online: 02 May 2022

P. Kalpana and S. Sridevi

A Survey on Recognition and Classification of Paddy leaf Diseases Using Image Processing and Machine Learning Techniques

P. Kalpana[a)], S. Sridevi

Department of CSE, VELS Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, Tamil Nadu, India.

[a)]Correspondingauthor:gonakalpanarao@gmail.com

Abstract. From the past few years spreading of the pests and diseases in plants have been increasing significantly. Paddy plants are the most important crop used in our country for the food, so it is most necessary to detect any disease of them within a short period of time for ensuring a proper and healthy growth of paddy plants. The process of manual disease detection requires labor and a large amount of time. Then utilizing the leaf images of plants for recognizing and classifying of diseases is the more focused research topic in the agriculture field. A survey on recognition and classification of paddy leaf diseases using image processing and machine learning techniques is presented in this paper based on the disease infected leaf images of paddy plants. Firstly the concept of various plant diseases and the standard process of plant disease detection are discussed in this paper. A study and survey on the totally 5 papers of work is carried out in detail by covering a work on leaf diseases of the paddy plants based on the certain criteria. Such criteria are different preprocessing techniques, various diseases/ classes, different segmentation methods, various classifiers and accuracy of employed techniques. The experimental results of these various techniques are compared and evaluated to design a best on detecting and classifying of paddy plant leaf diseases.

Keywords: Leaf disease detection, recognition, image processing, machine learning techniques and classification.

INTRODUCTION

The major source of income for most of the people in India is agriculture because mostly 70% of population depends only on farming. Since the farming is an essential way of living in rural area, about 58% of the people are depending on it. In India one of most important food is the rice or paddy.

Between 10% -15% of production has been destroyed in Asia due to the effect of diseases on rice plants i.e. paddy plants. Brown spot, Leaf blast, Leaf scald and Sheath blight are some of the different types of leaf diseases that occur in rice plants. Sometimes farmers are not able to search for diseases or have a difficulty to identify the diseases resulting to a loss in crop production. Different means of therapy is there for every disease. Currently, majority of the farmers generally rely on guides or using their own knowledge for the identification of diseases. This is a manual disease detection approach which has less accuracy and takes more time. There are different development stages in any disease of a plant so it is necessary to monitor the plants towards the infection if a disease occurred in a plant, which is a slow process and require safety measures when selecting pesticides.

In this survey, image processing and machine learning techniques used to detect diseases in paddy plants are focused [1]. Various paddy plant disease detection approaches and techniques used are studied in order to gain a comparative understanding of the different works. Different types of diseases in paddy plants are studied in prior for better understanding of the present survey. Various image processing methods as well as machine learning techniques are reviewed in this paper from the total of five papers that are used for the detection of leaf diseases in paddy plants and comparative analysis is derived from the survey. In the first paper, image segmentation and pre-processing techniques in combination with various types of classification features for example SVM are used to identify and diagnose diseases of paddy leaves. Then the second paper presented a machine learning techniques

International Conference on Recent Innovations in Science and Technology (RIST 2021)
AIP Conf. Proc. 2463, 020009-1–020009-8; https://doi.org/10.1063/5.0080336
Published by AIP Publishing. 978-0-7354-4192-7/$30.00

based paddy leaf disease detection system. In this three most frequently occurring paddy plant diseases especially brown spot, Bacterial Leaf Blight (BLB) and leaf smut are analyzed. In third paper a Mobile App is provided which aggregates the ML algorithms for disease identification in the paddy crops. In fourth paper, a Deep Neural Network (DNN) with Jaya Optimized Algorithm (JOA) for recognizing and classifying the leaf diseases in paddy plants is proposed. In this, preprocessing is carried out by converting the Red Green Blue (RGB) color images into Hue Saturation Value (HSV) color images to remove the background. Then the diseased and non-diseased parts are divided by performing the extraction of binary images based on the hue and saturation parts. A clustering technique is used to segment the diseased part, the normal part, and the background part. In the fifth paper image segmentation and a technique of feature extraction were used to analyze the plant diseases. Here, the blast and brown spot diseases of paddy plants are mainly discussed. Scale Invariant Feature Transform (SIFT) is used in this to extract the shape and color features.

DIFFERENT TYPES OF PADDYPLANT DISEASES

Various types of diseases that most widely occurred in paddy plants are explained briefly in this. Understanding the process of image processing types which would require and feature types that would necessary to consider for preparing the disease detection systems is the main intention of explaining the types of diseases in paddy plants. Five most widely occurred diseases are described briefly. Moreover, all of the paddy plant disease types are explained in [9]

1. Brown spot: The paddy plant leaves are affected by this disease which was occurred on leaves. Round to oval shape with a dark brown lesion is the symptom of this disease.
2. Leaf blast: Dark spot to an oval shaped spot ranging with narrow red-brown colored edges and a gray or white colored center is the symptom of this disease.
3. Sheath blight: Leaves and also stems of the plants are affected by this disease which occurred on both of them. The symptoms are white or straw color areas with oval shape and with red-brown spots in the center.
4. Bacterial leaf blight: Lengthen lesions on tip of leaves are the symptoms of this disease. In general lesions are of several inches long and are turn from white to yellow color due to the bacterial effect.
5. Leaf scald: Narrow broad band colored with reddish brown are the symptoms of this disease. Lesions sometimes are on the leaf margins with yellow or gold edges.

STANDARD PROCESS OF PLANT DISEASE DETECTION

Figure (1) shows the standard process of plant disease detection from the images. Following sections explains about the main steps of this process. The whole process includes capturing an image from the fields, preprocessing it, extracting the part of leaf image that is infected and performing the classification of diseases with the machine learning methods by extracting the features from segmented images [2].

Image Acquisition

The International Rice Research Institute is the one which is an image database especially available forpaddy plant disease images but not in a simple way. As a result, it is necessary to prepare animage database which needs the imageacquisition from real-time agriculture lands on our own self.In order to apply the digital image processing for the disease detection process, images are captured in this process using the digital camera which acquires the images in a digital form directly from the real-time agriculture lands with numerical values.Canon Power shot, Canon 450, Nikon Coolpix and Nikon are some of the most commonly utilized digital cameras for this purpose.

Image Pre-Processing

Since there is presence of noise in images, this image pre-processing step is essential in order to get the best results on further processing stages. Generally, dewdrops, dust and excrements of insects present on the plants are considered as the image noise. In this image pre-processing stage different types of techniques such as background removal techniques, images contrast enhancement technique, median filters, weiner filters, image size reduction, etc. are utilized based on the specific application [9].

Image Segmentation

In a plant disease detection process, this Image segmentation stage plays a significant role. After pre-processing, images are divided into particular objects or regions called as segments for better analyzing of data in the images and then allowed to easily extract the useful features from the data. This segmentation can be carried out according to the similarity based, done by using the thresholding (such as Ostu's method) and discontinuity based, done by using the edge detection method.

Feature Extraction

Identification of the features or intrinsic characteristics of objects that are present in an image is the main objective of this feature extraction method. Objects are described by using these identified features or characteristics. In general, the features are extracted from the color, shape and texture of three categories. In the case of paddy plant diseases each disease symptom is described by using the color feature, so it is very essential to trait to distinguish one disease from another. Moreover, disease symptom is described with the shape then by using the shape feature also diseases can be distinguished.

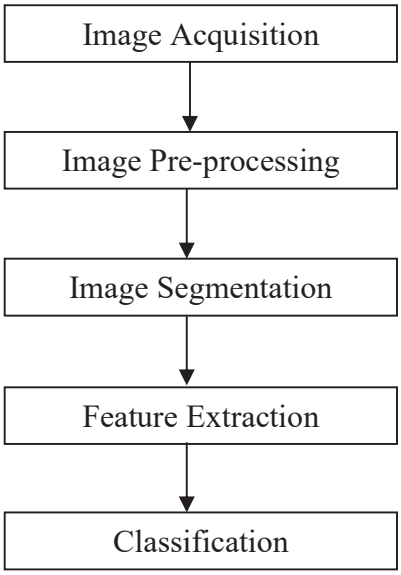

FIGURE 1. Standard Process of Plant Disease Detection from Images

Classification

The extracted data obtained from the feature extraction is then mapped to the various specific classes are groups in this image classification stage which is also called as supervised learning approach. This image classification can be carried out in two steps. First step is the training of learning step in which a classifier model is generated by developing the classification algorithms on learning the data and described a predefined set of classes. Second step is the testing step in which the classification of the trained data is performed using the generated classifier model in the first step to calculate the accuracy of the generated model by evaluating its performance against the test data. Rule based classifier, Neural Network (NN), Support Vector Machine (SVM) and nearest neighbor classifier are the various commonly used classification models.

As shown in figure (1), this was focused as the main research topic for the many of researches since the plant disease detection process has a possibility of different alternatives in various states.

SURVEY ON PADDY LEAF DISEASE DETECTION METHODS

Image Processing and Machine Learning Methods are used for the paddy leaf disease detection in the paper [3]. First the leaf images to be examined are gone under pre-processing stage which enhanced their quality and then image segmentation process is performed. In segmentation process, thresholding or the Otsu's method are used to find the threshold value. Here, Ostu's is the one of the most popular binarization algorithm used and is named after the Nobuyuki Otsu. Then the founded threshold value is set to the mean pixel intensity with linear combination of RGB planes to discriminate plant leaves pixels with the coefficients of r = -0.884, g = 1.262, b = 0.311 that are determined using genetic optimization algorithm. Since the shape is a key element by which people can mostly understand and differentiate any object, feature extraction based on the shape elements is focused in this paper. The most significant features which describe the shape of objects are the general descriptors such as length, width and number of objects. The lesion feature of a leaf is extracted by using these features in image processing, minimum colored background and foreground spread sum plays as a more significant one for all the time to recognize the different classes.

One of the most widely used machine learning algorithm for the classification problem is the Support Vector Machine (SVM). SVM is an algorithm in which 'n' number of features or object data are plotted as points in a n-dimensional coordinate space. The each coordinate value is represented by the corresponding feature. Then the hyper plane is found for well differentiating of two classes to carry out the classification process. A decision based hyper plane boundary is used between the two classes in the SVM which is a binary classifier. The data points or features relayed on either of two hyper plane sides are recognized as different classes. The number of features 'n' decides the hyper plane space dimension. Training vectors are used in the SVM to find the hyper-plane with maximum separation margin between two classes.

Different Machine Learning Techniques are used in the paper [4] for the detection of paddy leaf disease. This disease detection model using different machine learning algorithms is developed by using open source software which trains the model with different classifiers. A dataset containing the 40 number of images each regarding to the three different types of diseases in a form of .jpg is created manually for the paddy leaf disease detection by classifying the infected leaves. Such three types of diseases are namely Brown spot, Leaf smut and Bacterial leaf blight. There is a possibility to increase the dataset size using data augmentation up to 480images [9]. Then the additional 35 number of attributes are added to this dataset by converting the images into features using image filtering such as Color Layout Filter. After that, significant features are determined by using the correlation based attributes selection technique. Dataset is separated into training and testing datasets as two parts. Lastly four different machine learning classifiers such as logistic regression, k-nearest neighbor, decision tree, and naïve bayes are used based on different input image characteristics and generated the different results.

One of the foremost suitable models that can train the dataset in a case of target class having the categorical values is the Logistic regression since the objective of this paper is to calculate and classify the paddy leaf diseases from the affected leaves. However, a multiclass logistic regression was used in this paper as it is necessary to predict the three different types of diseases. Similar to this logistic regression model, K-Nearest Neighbor model was also suitable in the case of target classes having discrete values. The K nearest neighbors are determined by finding the k minimum distances from the calculation of query point distances to predict the class of any query point.

Another suitable classifier mostly used for the disease detection in plants is the Decision tree. This decision tree algorithm splits the dataset into divisions and most suitable attribute is taken at root. Separating the data is the purpose of this dataset splitting. Dataset is repeatedly splitted until the partitions finally group the data so that it is homogeneous. The main algorithm of a decision tree that uses a greedy approach is the Iterative dichotomizer 3 (ID3). The tree is constructed under this approach by following the information theory concept from which information gain and entropy are determined. For this purpose of attributes selection, the attribute with the high information gain has been selected. Then the entropy measured the contamination of attributes. Lastly, the probabilistic algorithm developed derived from Baye's theorem is Naïve Bayes algorithm from which a best hypothesis was chosen.

Optimized DNN with Jaya Algorithm is discussed in the paper [5] for detecting and classifying the paddy leaf diseases. First a dataset is created by capturing the leaf images of paddy plants from the agriculture fields. Image pre-processing is performed first which reduces the image sizes as well as performs the background removal operation. Next a k-means clustering segmentation approach was used which segmented the diseased and normal portions of the leaf images in the image segmentation step. The diseases are then classified using the DNN_JOA classification method. The main significance of this DNN_JOA algorithm is that stability of this method is précised more by sending a feedback to the segmentation stage after the classification is performed when it results are not satisfied. Best weights selected by the JOA are used to classify the diseases in this DNN_JOA method.

DNN with SVM Classifier is used in the paper [7] to manage and identify the paddy crop disease in smart way. The Smart Paddy Pest Management model is based sensor network incorporated to a mobile application. The disease identification and disease management are the two modules include in developing of this approach. The type of infection that is occurred in the paddy plant is detected in this Disease Identification module. Then the resulted detected infection type in a disease identification module is intimated to the farmer through mobile app in the disease management module. The disease identification process is implemented with the help of mobile application. It is a four step process namely: 1) Image capture & selection 2) Image zoom and crop, 3) Upload image and 4) Receive notification. Image capture & selection: Diseased affected paddy image is captured through a clear camera. Multiple snapshots are to be taken for choosing the appropriate affected area. A clear image is chosen such that the disease affected areas are clearly visible. In case of same crop problem, choose images from the database which was created earlier; Image Zoom and crop: Choose the best portion of the disease affected image and crop it; Upload Image: Cropped image is to be uploaded in the remote server using the mobile app; Receive notification: Once image has been uploaded in the remote server, pattern matching is performed with the available datasets using pattern matching algorithm, and the precaution is send to the farmer via mobile app by the expert. For detecting various paddy crop diseases of Nematode, Blast, Smut, Spots the image processing techniques are introduced.

Support Vector Machine is utilized to develop the ideal isolating hyper plane for different paddy crop disease structures. For distinguishing an infection, paddy plant ailments highlights are extricated utilizing SIFT from the info pictures for the three sickness classes. In the preparation stage, seven dimensional component vectors are removed from each unhealthy picture and is given as contribution to the SVM classifiers. The seven highlights are x position, y position, scale(sub-level), size of highlight on picture, edge hail, edge introduction, ebb and flow of reaction through scale space. For classification, seven infection highlights nourished into the SVM demonstrate and the separation between each of the element vectors and the SVM hyper plane is determined. The normal separation is computed for each model. The normal separation gives preferred come about over utilizing separation for each component vector.

SVM Classifier method is used in the paper [10] for the leaf disease detection in paddy plants from the leaf images. Improving the data present in an image is the main aim of this paper. In this, mainly image resizing and image segmentation processes are concentrated. Image resizing is needed for the different applications such as storing, displaying and transmitting of images whereas image segmentation process is required to petition the image in to several small regions or segments for changing of the image representation to simply understand and analyze the segment or region of interest in the image. First an input image is considered in a (Red Green Blue) RGB form then a structure called color transformation is created for the RGB image of paddy leaf. The before created color transformation structure is applied with the device-in dependent color space transformation. The resulted image is preprocessed by considering the techniques of image clipping, cropping, smoothing and enhancement to remove noise, increasing the contrast and better acquiring the interest of image region and enhance the performance of image in the image pre-processing stage. Next, preprocessed image is segmented in to same features of several partitions based on some similarities. K-means clustering and Ostu's methods are used in this paper for the segmentation process, converting the RGB image into the (Hue Intensity Saturation) HIS color model image. A particular object of an image is extracted to identify the particular object of diseased leaf image. This process is the image extraction which is most widely used in various image processing applications. This leaf disease detection process in paddy plants uses the morphology, color, edges, and textures etc. as feature components. Lastly, SVM classifier model is used to classify the features extracted in the feature extraction process which was trained by the Artificial Neural Networks (ANN) on database images.

RESULTS

Database creation is the most basic and tedious undertaking in paddy crop disease administration. On the premise of a vigilant examination of the contemporary resources, including distributed and unpublished writing, of crop diseases and populace flow of specialists 'conclusions, applicable data for the paddy crop infections was gathered from various sources. In this dataset 80% of images are applied to the training set and the remaining 20% of images are applied to the training set. Then the training and testing accuracy results for different algorithms used in the survey papers are compared based on the four types of disease classes such as blast, brown spot, bacterial blight and sheath rot disease as shown in figure (2) and figure (3).

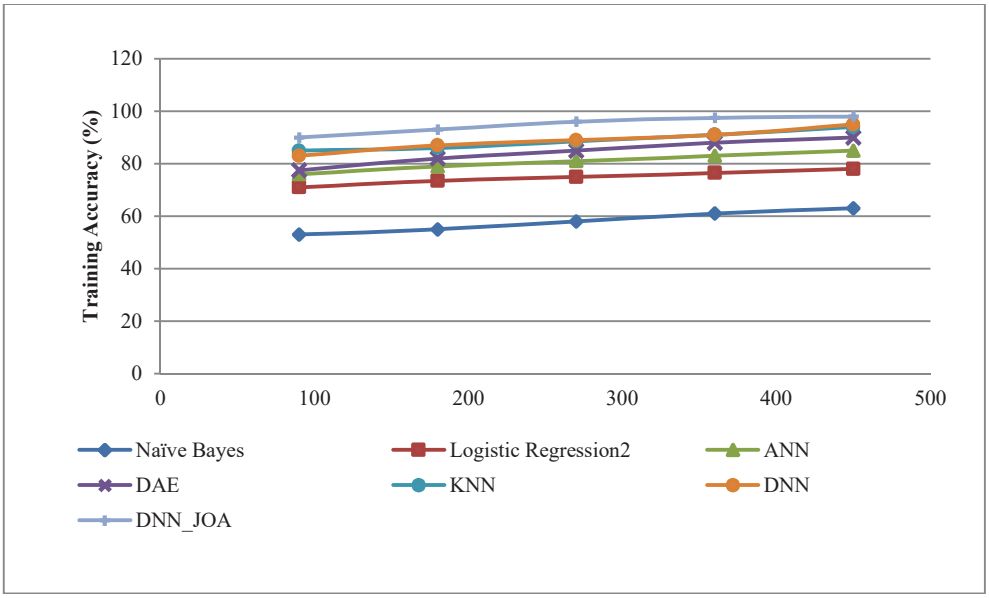

FIGURE 2. Training Accuracy of Different Classifiers

The definition for the accuracy measures is given by,

$$Traing\ Accuracy\ (\%) = \frac{\text{Correctly Recognized Images in Training dataset}}{\text{Total Number of images in Training dataset}} \times 100$$

$$Testing\ Accuracy\ (\%) = \frac{\text{Correctly Recognized Images in Testing dataset}}{\text{Total Number of images in Testing dataset}} \times 100$$

$$Accuracy\ (\%) = \frac{\text{Correctly Recognized Images in a dataset}}{\text{Total Number of images in dataset}} \times 100$$

Training accuracy comparison of seven classification methods such as Naïve bayes, logic regression, Artificial Neural Networks (ANN), k-Nearest Neighbor (KNN), Denoising Auto Encoder (DAE), Deep Neural Network (DNN), Deep Neural Network with Jaya Optimized Algorithm (DNN_JOA) with respect to the four types of disease classes is shown in figure (2). Here, the 500 number of samples are considered to train the classifier models. The increase in the number of considered samples increases training accuracy as shown from the figure (2). Training accuracy is obtained high for the DNN_JOA algorithm while Naïve Bayes has a lowest accuracy than all of the six algorithms.

Similarly testing accuracy comparison also carried out among seven types of classifiers on four types of disease classes as shown in figure (3). Same as the training accuracy, DNN_JOA classifier has higher accuracy and naïve bayes classifier has lower accuracy over 500number of considered samples. Then by observing both the training and testing accuracy of classifiers it can be concluded that training accuracy is attained high values than the testing accuracy.

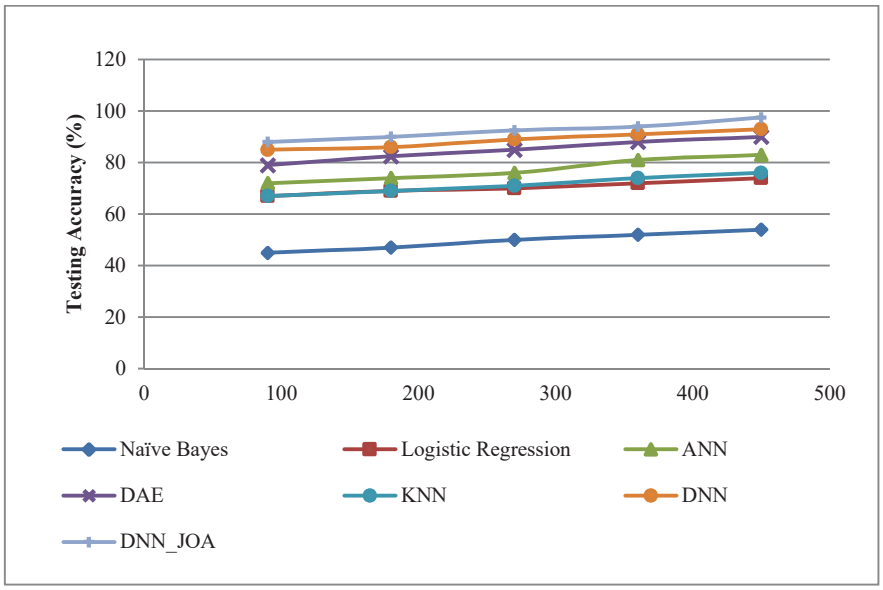

FIGURE 3. Testing Accuracy of Different Classifiers

The performance of the ANN, k-means & Fuzzy logic, KNN & SVM classifier, DAE, DNN, Deep CNN & SVM and DNN-JOA classification methods are compared here with respect to the four different diseases in terms of accuracy. Form the table (1), it can be observe that the high accuracy of blast, brown spot, BLB and sheath blight are obtained as 98%, 94%, 96% and 92% respectively for a DNN_JOA method where as low accuracy of 88%, 82%, 78% and 78% respectively for the blast, brown spot, BLB and sheath blight diseases is obtained when using ANN classifier.

TABLE 1. Disease Detection Accuracy Comparison of Different Classification Algorithms

Disease / **Classifier**	Blast	Brown spot	Bacterial leaf blight (BLB)	Sheath blight
ANN	88	82	78	78
K-mean & Fuzzy logic	80	80	76	76
KNN & SVM Classifier	84	84	84	80
DAE	86	78	76	81
DNN	90	79	80	86
Deep CNN & SVM Classifier	92	98	92	88
DNN_JOA	98	94	96	92

CONCLUSION

Appropriate measures that have not taken about the diseases of paddy plants can lead to a large amount of production loss in agriculture. Then various aspects of image processing methods along with the machine learning technique of existing woks are studied. Image processing and machine learning techniques based paddy leaf disease

detection papers are studied and evaluated in this paper according to the accuracy of different image segmentation algorithms along with accuracy of training and testing classifier algorithms. Variety of segmentation techniques that extract the region of image from a diseased leaf are studied first and then compared. In the same way different types of machine learning techniques of classifiers are studied which classifies the various types of diseases in the plants and then compared in the results. This survey acquired significance in designing a best method to effectively detect the leaf diseases in the paddy plants.

REFERENCES

1. Yong Ai, Chong Sun, Jun Tie, Xiantao Cai, "Research on Recognition Model of Crop Diseases and Insect Pests Based on Deep Learning in Harsh Environment", IEEE Access **8** (2020).
2. Guoxiong Zhou, Wenzhuo Zhang, Aibin Chen, Mingfang He and Xueshuo Ma, "Rapid detection of Rice Disease Based on FCM-KM and Faster R-CNN Fusion", IEEE Access **7** (2019)
3. Priyanka B Raj,Pooja RDr. Neha Mangla, and Soumya G Hegd , "Paddy Leaf Disease Detection Using Image Processing and Machine Learning techniques", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJREEICE) **7** (2019)
4.]Syed. Md. Irfanul Alam, Sifat Momen, Kawcher Ahmed andTasmia Rahman Shahidi, "Rice Leaf Disease Detection Using MachineLearning Techniques", Inter. Confe. on Sustainable Technologies forIndustry 4.0 (STI), 24-28 (2019)
5. Dr.D.Vydeki and S.Rames, "Recognition and Classification of Paddy LeafDiseases using Deep Neural Networkwith Jaya Optimized Algorithm", Information Processing in Agriculture (2019).
6. X. Zhang, Y. Qiao, F. Meng, C. Fan and M. Zhang, "Identification of maize leaf diseases using improved deep convolutional neural networks*," IEEE Access **6**,* 30370-30377 (2018).
7. D.Raghu Raman, R.Rajmohan, U. Prabu, R.Rajesh, and M.Pajany, "Smart Paddy Crop Disease Identification and Management Using Deep Convolution Neural Network and SVM Classifier", Inter., Jour., of Pure and Applied Mathematics **118**, 255-264 (2018).
8. G. K. Birajdar and B. S. Ghyar, "Computer vision based approach to detect rice leaf diseases using texture and color descriptors", *International Conference on Inventive Computing and Informatics(ICICI),* 1074-1078 (2017).
9. J. Ding, B. Chen, H. Liu and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition", *IEEE Geosci Remote Sens. Lett **13***, 364-368 (2016).
10. V.Praveena,T.Monika,S.Pavithra andA.Priyadharshini, "Paddy Leaf Disease Detection Using SVM Classifier", International Journal of communication and computer Technologies **3**, (2015).
11. A. Kumar, M. Jhuria and R. Borse, "Image processing for smart farming: Detection of disease and fruit grading", *IEEE Second International Conference on Image Information Processing (ICIIP),* 521-526 (2013).
12. A Nasir, A Rahman, A Fakhri, A Rasid Mamat and M Nordin, "A studyof image processing in agriculture application under high performancecomputing environment," Inter. Jour. Comp. Scie.&Tele comm **3**,16-24 (2012).
13. A.-R. Mohamed, G. E. Dahl and G. Hinton, "Acoustic modeling using deep belief networks", *IEEE Trans. Audio Speech Lang. Process **20***,14-22 (2012).
14. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and F.-F. Li, "ImageNet: a Large-Scale Hierarchical Image Database," *IEEE Conference on Computer Vision and Pattern Recognition* (2009).
15. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," Int. J. Comput. Vis **88** ,303–338 (2009).
16. Z. Peng, G. Liu, M. Z. Li, and D. L. Li, "Management information systemfor apple diseases and insect pests based on GIS," Transactions of the CSAE, **12** (2006).
17. E.-C. Oerke, U. Steiner, H.-W. Dehne, and M. Lindenthal, "Thermalimaging of cucumber leaves affected by downy mildew and environmental conditions," J. Exp. Botany **57**, 2121–2132.(2006).
18. A. J. Smola and B. Scholkopf, "A tutorial on support vector regression,"Statist. Comput.**14** ,199–222 (2004).
19. S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vectormachines with Gaussian kernel," Neural Comput. **15** ,1667–1689 (2003).
20. B. Bhanu and J. Peng, "Adaptive integrated image segmentation and object recognition", *IEEE Trans., on Systems Man and Cybernetics Part C **30*** ,427-441 (2000).