REVIEW ARTICLE Indian Journal of Agricultural Research, Volume 58 Issue 2: 183-187 (April 2024)

Computer Vision and Machine Learning in Paddy Diseases Identification and Classification: A Review

Hemanthakumar R. Kappali[1], K.M. Sadyojatha[1], S.K. Prashanthi[1]

10.18805/IJARe.A-6061

ABSTRACT

Food security of the Indian population depends upon the agriculture. The annual crop loss due to the pestsand diseases is a severe problem which is addressed by the researcher time and again Paddy is one of the common crops found in most of the agriculture areas in India. Paddy plants exhibit health condition mainly in stem and leaf. Farmers are still using the traditional method and experts suggestions in identifying the diseases. Research in disease identification and classification process is going on for the past three decades but its automation is still an open challenge. Image segmentation based computer vision techniques, supervised and unsupervised classification algorithms of machine learning are widely used in the analysis of vegetation health analysis. Pattern recognition is playing a major role in identifying the features causing the diseases. The focus of the work is on investigating the performance of various segmentation techniques with support vector machine and K-Neural Network classification algorithms using statistical features. The hybrid IP-PR (Image Processing-Pattern Recognition)techniques are used for paddy disease classification. The performance of the classification algorithm is primarily depending upon the intelligence of the segmentation techniques. The properly segmented outputs give a better accuracy.

Key words: Hybrid IP_PR (Image processing - Pattern recognition), Vegetation.

Agriculture has become a means to feed ever-growing populations. The plant has become a primary source of energy for humans as well as animals. It is also an important piece in the puzzle to solve the problem of global warming. Leaves of the various plants and herbs are helpful to mankind due to their medicinal attributes. In Asia and Africa where over 50% of population depends on agriculture production for employment, export earnings and food security (Anthonys and Wickramarachchi, 2009; Singh and Kaur, 2018).

It is assessed that 30% to 40% of crops are lost each year from production chain and in India, the annual estimated losses due to nematodes have been assessed to be about a 242.1 billion (Dhingra and Joshi, 2018) which is having an impact on economical, social and ecological conditions. There are several diseases which affect plants These diseases can be broadly classified under parasitic or non parasitic causes as in Fig 1.

Most diseases however generate some kind of manifestation in the visible spectrum. In most ofthe cases, the diagnosis about the disease is performed visually by humans. However this method is having disadvantages in accuracy, substantial inter and intra variability prone to various illusions (Anthonys and Wickramarachchi, 2009). The foundation of quality assessment is basically dependent upon features of leaves such as its appearance, cracks, textures and surface where human alert can be easily fooled. In fact 60% to 70% of diseases appear on leaf only. So we have interest in the plant leaf rather than whole plant (Dhingra *et al*., 2018). The review is restricted to the rice plant and its diseases and the considerable research is also done on rice plant diseases. The commonly occurring rice diseases and their symptoms and causes are as in Table 1.

The important doctrine and standard studies show that various image processing based machine learning and deep learning techniques are effective tools for identification and classification of plant diseases. The underlying principle and standard remain same, but computation and testing become accurate with the application of computer vision techniques.

The image analysis technique is used to distinguish the object from the background, there by isolating quantifiable information which is used in various decisions. Machine learning based, location and acknowledgment of

[1]Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari-583 104, Karnataka, India.

Corresponding Author: Hemanthakumar R. Kappali, Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari-583 104, Karnataka, India. Email: hemanthakumar@bitm.edu.in

How to cite this article: Kappali, H.R., Sadyojatha, K.M. and Prashanthi, S.K. (2024). Computer Vision and Machine Learning in Paddy Diseases Identification and Classification: A Review. Indian Journal of Agricultural Research. 58(2): 183-187. doi: 10.18805/IJARe.A-6061.

Submitted: 06-10-2022 **Accepted:** 08-02-2023 **Online:** 27-03-2023

plant infections give extensive information to recognize and treat the diseases (Mallikarjuna *et al*., 2011; Singh and Kaur, 2018).

Many researchers have worked on the improvement strategies for the programmed recognition and classification of leaf infections, in light of high determination multi spectral, hyper spectral and stereo pictures. Disclosure of plant illnesses based on the symptoms of diseases showed on the plant leaves is the topic of present focus. The authors are searching for quick, affordable and precise technique to recognize plant diseases (Singh and Kaur, 2018). The solution will help the agriculture domain through the automation of the plant disease detection and classification.

Classification is playing a key role in replacing the manual processes by the automatic process. The classification of the desired classes is automated using the various classification algorithms. The algorithm takes the physical features as input and gives the labeled object as output. Soil related features like, moisture, temperature, electrical conductivity pH, organic carbon, available nitrogen are considered as input data and soil type is considered as class label. Features are fed to the Linear Discriminate Analysis algorithm and are selected using the Boruta algorithm and its performance is found to be 96.3% indicated by scatter plot (Radhika and Madhavi Latha, 2019).

Remote sensing is a new technology to sense things that are far away, from the ground, air and satellite. This Technique uses different imaging sensors and also a few non-imaging sensors to collect the data. Plant diseases reflect the various symptoms on various parts of the plants such as the leaf, stem, flower, *etc*. These parameters can be sensed remotely to detect and monitor the health of the plant. Rice disease sheath blight was sensed remotely using Air Borne Data Registration in 2005. This has become the motivation for automating the identification and classification of plant diseases using technology like remote sensing, computer vision and machine learning. These techniques are helpful in large-scale farming to detect diseases on time. They also provide non-invasive, rapid, reliable, precise and accurate estimates of diseases which place a very crucial role in the monitoring of epidemics (Gogoi *et al*., 2018).

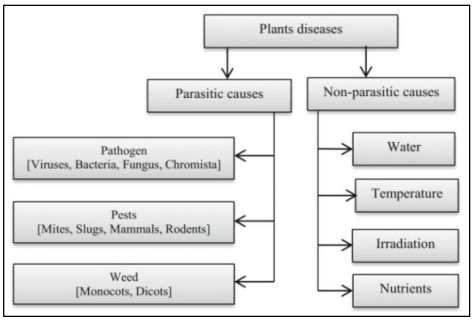

Fig 1: General structure of types of plant diseases (Dhingra *et al*., 2018).

Automation of agricultural processes is the recent advancements using computer vision and Neural Network based machine learning techniques. The dry red chilly images are acquired using image sensors and segmented using the Otsu thresholding algorithm to separate the chilly from its background. Morphological and color-based features are extracted. The dry chilly grading system is automated using self organizing map clustering algorithm (Shetty *et al*., 2020).

The affected part and the healthy part of the plant leaf is separated using thresholding segmentation techniques. This techniques will consider the task as two class problem and the threshold is calculated using local entropy, Otsu method (Anthonys and Wickramarachchi, 2009; Mallikarjuna *et al*., 2011; Phadikar and Goswami, 2016), variable and global threshold values are calculated to segment an image into affected and non affected part (Guru *et al*., 2011; Orillo *et al*., 2014). For better segmentation results K-means clustering, region based segmentation are used in one or two stages (Arnal, 2013; Guru *et al*., 2011; Hasan *et al*., 2019; Qiangqiang *et al*., 2015; Ramesh and Vydeki, 2020; Singh and Kaur, 2018).

The performance of the segmentation technique is evaluated by few researchers using error rate, over lapping, under and over segmentation, Dice similarity, Precision and Recall (Anthonys and Wickramarachchi 2009; Qiangqiang *et al*., 2015). The classification methods used by many researchers takes the features of the image and label of the classes as input and get training further for

Table 1: Rice diseases, symptoms and cause.

Disease	Symptoms	Pathogen
Brown spot	Dark irregular brown spot	Fungi
Blast leaf	Lesions on all parts. White to gray or diamond shaped lesions with green border	Fungi
Bacterial leaf streak	Leaf turns to brown and then gray white in color	Bacterium
Bacterial blight	Leaves turning yellow	Bacterium
Stem root	Black lesions on outer leaf sheath	Fungus
Grassy stunt	Narrow pale green or yellow leaves, dark brown spots on leaves	Virus
Tungru	Yellow orange discoloration, rust colored spots on leaves	Virus
Sheath blight	Pale green or white center and purple brown margin	Fungi
Bakanae	Seedlings are elongated, slender and pale. Abnormal elongation of old plant	Fungi
False smut	Chalkiness of the green individual grain is covered with white fungal mass	Fungi

[Rice | Diseases and Pests, Description, Uses, Propagation (psu.edu)].

testing phase. The following are the features extracted in many researches. Shape features are extracted from the segmented image using neighborhood technique and Blob analysis (Guru *et al*., 2011; Mallikarjuna *et al*., 2011; Orillo *et al*. 2014; Pinki and Islam, 2017). Texture features of first and second order are extracted by contrast stretching transformation (Anthonys and Wickramarachchi, 2009; Mallikarjuna *et al*., 2011; Pinki *et al*., 2017) and gray level co-occurrence method (Arnal, 2013; Hasan *et al*., 2019). Fractal descriptors are derived from fractal dimension. (Asfarian *et al*., 2013; Devi and Muthukannan, 2014) Color features are extracted using color components and by their histograms (Arnal, 2013; Khirade and Patil, 2015; Ramesh and Vydeki, 2020).The contrast energy, energy relation and homogeneity are also used for classification (Phadikar and Goswami, 2016).

Automation is achieved by many researchers using the machine learning for classification. Considerable accuracy is achieved using probabilistic neural network, fuzzy classification model, back propagation neural network, random forest algorithm, support vector machine, decision tree and convolution neural network. Performance of the classification models are evaluated using confusion matrix, performance plot, training state plot, accuracy, cross validation are addressed by most of the researchers (Devi and Muthukannan, 2014; Dhingra *et al*., 2018; Guru *et al*., 2011; Hasan *et al*., 2019; Khirade and Patil, 2015; Singh and Kaur, 2018). The correlation between the performance of classification method and precision of the segmentation technique is mostly unaddressed by the researchers.

The proposed work is carried out in the Department of Electronics and Communication research center at Ballari Institute of Technology and Management, Ballari, Karnataka during the period from June 2021 to August 2022 in collaboration with the expert from the department of plant pathology, university of Agricultural Science, Dharwad, Karnataka. The samples required for the work are taken from machine learning repository [https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases].

Methodology

The methodology used by most of the researchers includes the following steps in identifying and classifying the plant diseases. Image acquisition Image pre-processing, Segmentation, Feature extraction, classification and identification and Performance evaluation are as per the sequence given in below Fig 2.

Image acquisition

The input data required are images of different diseases, collected from the paddy fields using the digital camera and digitized into [M] rows and [N] columns (Singh and Singh 2015) directly in the RGB color model using digital cameras and stored in the computer memory for further processing (Mallikarjuna *et al*., 2011; Orillo *et al*., 2014). These images are transformed into smaller dimensions (Arnal, 2013; Ramesh and Vydeki, 2020). Images are collected in controlled environmental factors by using a controlled light module box (Khirade and Patil, 2015).

Image pre-processing and image segmentation

Making an image ready for segmentation by applying image enhancement, blob analysis for noise reduction, contrast adjustment and image intensity adjustment (Khirade and Patil, 2015) Converting images from RGB color models into HSV, Gray images (Mallikarjuna *et al*., 2011; Singh and Singh, 2015). Image smoothing and linear filtering (Dhingra *et al*., 2018) are performed as a pre-processing steps.

The next step is to segment an image into a region of interest and background or diseased part and healthy part of the leaf. It is found in literature that, segmentation is performed by Local entropy, Otsu thresholding methods (Mallikarjuna *et al*., 2011), Region growing segmentation and Mean shift segmentation (Pinki *et al*., 2017). Histogram equalization using two bins and masking with the original image and segmented image. Segmentation using color model boundary and edge detection using the Sobel operator. (Arnal, 2013) K-means clustering (Phadikar and Goswami, 2016; Ramesh and Vydeki, 2020; Singh and Kaur, 2018). Active contour models and super pixels segmentation (Qiangqiang *et al*., 2015) are also used for segmentation.

Features extraction in image

The next step is to extract the characteristics from the segmented image. Features extracted from acquired images are, statistical features using, color co-occurrence matrix or gray image, shape features using blob analysis (Arnal, 2013; Hasan *et al*., 2019; Khirade and Patil, 2015; Phadikar and Goswami, 2016) and Fourier dimension and descriptors (Anthonys and Wickramarachchi, 2009; Narmadha and Arulvadivu, 2017).

Detection and classification of plant disease

These features are fed to the classification model and the methods used in classification are, the production rule based method through serial interviews (Mallikarjuna *et al*., 2011; Orillo *et al*., 2014). Features mapped to the disease will decide the membership function (Singh and

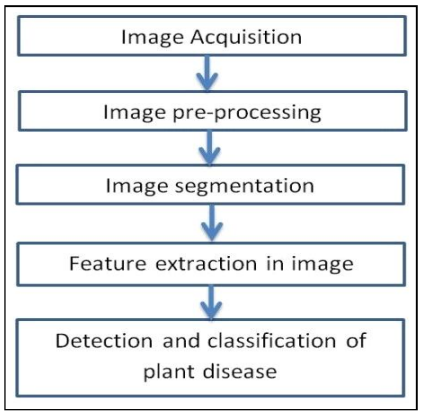

Fig 2: Methodology for plant disease detection (Khirade and Patil 2015).

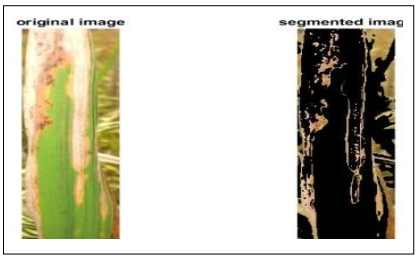

Fig 3(a): Properly segmented image.



Fig 3(b): Poorly segmented image.

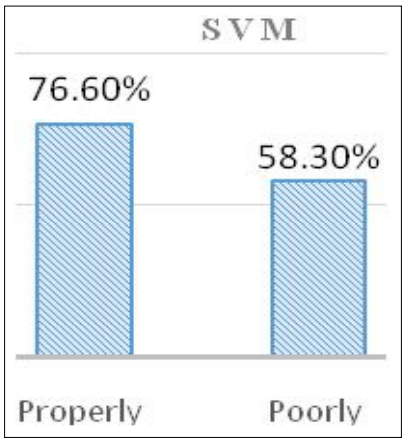

Fig 4: Classification accuracy.

Singh, 2015) and the machine learning-based supervised and unsupervised algorithms are used to identify the disease type. The model preparation is done in two phases. one is training phase and the other is testing phase. In the training phase, the extracted features are prepared with the name of the disease as the label. These are given to the models like Support vector machine, Random forest classifier (Hasan *et al*., 2019), probabilistic neural network (Devi and Muthukannan, 2014; Khirade and Patil, 2015; Narmadha and Arulvadivu, 2017), linear discriminate analysis and Back propagation algorithm (Arnal, 2013; Ramesh and Vydeki, 2020) and neural network models. The crucial steps in the above methodology are segmentation and classification. Accuracy of segmentation is measured using Jaccard similarity. Over and under segmentation is measured through (Mallikarjuna *et al*., 2011; Pinki *et al*., 2017; Singh and Singh, 2015) structural components, normalized cross-correlation, peak signal to noise ratio based on difference in the image pixel value (Guru *et al*., 2011). Confusion matrix, accuracy, sensitivity, specificity, Dice's coefficient, precision using true/false positive and negative rates are used as performance evaluating metrics for classification (Dhingra *et al*., 2018; Kaur and Oberoi, 2020; Khirade and Patil, 2015; Zhang, Yan and Hou, 2018).

The literature review has revealed that the automation of Rice plant health monitoring and classification of diseases based on the features of the affected part is an open-ended problem. The techniques supporting the development of solutions are from image processing, pattern recognition and machine learning domains.

Various segmentation techniques of image processing are used to separate the healthy part and affected parts. Based on the precision of segmentation; the techniques are classified as primary, secondary and higher segmentation for this work. Thresholding, local entropy and Otsu segmentations are treated as primary techniques. Region growing clustering, mean shifting and super pixel segmentation as secondary techniques. K means, Active contour are considered as higher level techniques.

The segmentation output of disease affected paddy leaves are obtained using K-means segmentation and Otsu thresholding method then the disease part and the normal parts of the leaves are visually inspected and termed as properly and poorly segmented outputs as shown in Fig 3(a) and 3(b).

The statistical features say contrast, energy, homogeneity, mean, standard deviation, entropy, root mean square(RMS), variance, smoothness and inverse difference moment (IDM) are extracted from segmented images and used as input for multi class support vector machine classification algorithm.

The classification algorithm support vector machine(SVM) results were analyzed and the classification accuracy was found to be 76.60%, for properly segmented images against the 58.30% accuracy for the poorly segmented images (Fig 4).

CONCLUSION

The hybrid IP-PR (Image Processing-Pattern Recognition segmentation technique is used for paddy disease classification. Properly segmented outputs using k-means segmentation techniques gave the better accuracy in classification using support vector machine. However, instead of visual assessment of segmentation a suitable metric may be used and correlated with performance metric of classification algorithm.

The authors have also experienced the data acquisition as a challenging issue because of varying environmental factors like light, distance, location of the symptoms which affect the performance of image segmentation techniques.

The techniques discussed in the present work are may be solution for monitoring, controlling and minimizing the annual crop loss, provided real time data set of the disease affected leaf images is readily available.

Conflict of interest

All authors declare that they have no conflict of interest.

REFERENCES

Anthonys, G. and Wickramarachchi, N. (2009). An Image Recognition System for Crop Disease Identification of Paddy Fields in Sri Lanka in 2009 International Conference on Industrial and Information Systems (ICIIS). IEEE. pp: 403-407.

Arnal, B., Garcia, J. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. Springer Plus. 2(1): 1-12.

Asfarian, A., Herdiyeni, Y., Rauf, A. and Mutaqin, K.H. (2013). Paddy Diseases Identification with Texture Analysis Using Fractal Descriptors Based on Fourier Spectrum in 2013 International Conference on Computer, Control, Informatics and Its Applications (IC3INA). IEEE. pp: 77-81.

Devi, D.A. and Muthukannan, K. (2014). Analysis of Segmentation Scheme for Diseased Rice Leaves. In IEEE International Conference on Advanced Communications, Control and Computing Technologies. IEEE. pp: 1374-1378.

Dhingra, Gittaly, Kumar, V. and Joshi, H.D. (2018). Study of digital image processing techniques for leaf disease detection and classification. Multimedia Tools and Applications. 77(2): 1-50.

Gogoi, N.K., Deka, B. and Bora, L.C. (2018). Remote sensing and its use in detection and monitoring plant diseases: A review. Agricultural Reviews. 39(4). doi: 10.18805/ag.R-1835.

Guru, D.S., Mallikarjuna, P.B. and Manjunath, S. (2011). Segmentation and Classification of Tobacco Seedling Diseases in Proceedings of the Fourth Annual ACM Bangalore Conference. pp: 1-5.

Hasan, M.d.J., Mahbub, S., Alom, M.S. and Nasim, M.A. (2019). Rice Disease Identification and Classification by Integrating Support Vector Machine with Deep Convolutional Neural Network in 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE. pp: 1-6.

Kaur, V. and Oberoi, A. (2020). Novel Approach for Plant Disease Detection Based on Textural Feature Analysis. In Data Management, Analytics and Innovation. Springer. pp: 439-449.

Khirade, Sachin D. and Patil, A.B. (2015). Plant Disease Detection Using Image Processing in 2015 International Conference on Computing Communication Control and Automation. IEEE. pp: 768-771.

Mallikarjuna, P.B., Guru, D.S. *et al.*, (2011). Performance evaluation of segmentation and classification of tobacco seedling diseases. International Journal of Machine Intelligence. 3(4): 204-11.

Narmadha, R.P. and Arulvadivu, G. (2017). Detection and Measurement of Paddy Leaf Disease Symptoms Using Image Processing in 2017 International Conference on Computer Communication and Informatics (ICCCI). IEEE. pp: 1-4.

Orillo, William, J., Cruz, J.D., Agapito, L., Satimbre, P.J. and Valenzuela, I. (2014). Identification of diseases in rice plant (*Oryza sativa*) using back propagation artificial neural network. in 2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE. pp: 1-6.

Phadikar, S. and Goswami, J. (2016). Vegetation indices based segmentation for automatic classification of brown spot and blast diseases of rice. in 2016 3rd International Conference on Recent Advances in Information Technology (RAIT). IEEE. pp: 284-289.

Pinki, Tazmim, F., Khatun, N. and Islam, S.M.M. (2017). Content Based Paddy Leaf Disease Recognition and Remedy Prediction using Support Vector Machine in 2017 20th International Conference of Computer and Information Technology (ICCIT). IEEE. pp: 1-5

Qiangqiang, Z., Zhicheng, W., Weidong, Z. and Yufei, C. (2015). Contour-Based Plant Leaf Image Segmentation Using Visual Saliency in International Conference on Image and Graphics. Springer. pp: 48-59.

Radhika, K. and Latha, D.M. (2019). Machine learning model for automation of soil texture classification. Indian Journal of Agricultural Research. 53(1): 78-82. doi: 10.18805/IJARe.A-5053.

Ramesh, S. and Vydeki, D. (2020). Recognition and classification of paddy leaf diseases using optimized deep neural network with jaya algorithm. Information Processing in Agriculture. 7(2): 249-260.

Shetty, Dasharathraj, K., Acharya, U.D., Narendra, V.G. and Prajual, P.J. (2020). Intelligent system to evaluate the guality of drc using image processing and then categorize using artificial neural network (ANN). Indian Journal of Agricultural Research. 54(6): 716-723. doi: 10.18805/IJARe.A-5374.

Singh, A. and Singh, M.L. (2015). Automated Color Prediction of Paddy Crop Leaf Using Image Processing in 2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR). IEEE. pp: 24-32.

Singh, J. and Kaur, H. (2018). A Review on: Various Techniques of Plant Leaf Disease Detection in 2018 2nd International Conference on Inventive Systems and Control (ICISC). IEEE. pp: 232-238.

Zhang, J., Yan, L. and Hou, J. (2018). Recognition of Rice Leaf Diseases Based on Salient Characteristics in 2018 13th World Congress on Intelligent Control and Automation (WCICA). IEEE. pp: 801-806.