International Journal of Nanoelectronics and Materials
Volume 14 (Special Issue) December 2021 [253-263]

Image Processing for Paddy Disease Detection Using K-Means Clustering and GLCM Algorithm

A. F. A. Ahmad Effendi[1], M. N. Md Isa[1*], M. I. Ahmad[1], M. F. Che Husin[1] and S. Z. Md Naziri[1]

[1]Faculty of Electronic Engineering Technology Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia.

ABSTRACT

The traditional human-based visual quality inspection approach in agriculture is unreliable and uneven due to various variables, including human errors. In addition to the lengthy processing durations, the traditional method necessitates plant disease diagnostic experts. On the other hand, existing image processing approaches in agriculture produce low-quality output images despite having a faster computation time. As a result, a more comprehensive set of image processing algorithms was used to improve plant disease detection. This research aims to develop an efficient method for detecting leaf diseases using image processing techniques. In this work, identifying paddy diseases based on their leaves involved a number of image-processing stages, including image pre-processing, image segmentation, feature extraction, and eventually paddy leaf disease classification. The proposed work targeted the segmentation step, whereby an input image is segmented using the K-Means clustering with image scaling and colour conversion technique in the pre-processing stage. In addition, the Gray Level Co-occurrence Matrix technique (GLCM) is used to extract the features of the segmented images, which are used to compare the images for classification. The experiment is implemented in MATLAB software and PC hardware to process infected paddy leaf images. Results have shown that K-Means Clustering and GLCM are capable without using the hybrid algorithm on each image processing phase and are suitable for paddy disease detection.

Keywords: GLCM, Image processing, K-Mean clustering, MATLAB, Paddy disease.

1. INTRODUCTION

Paddy is the main source of calories, carbohydrates, and protein in the diet. It is the world's second most extensively farmed cereal crop, with more than half of its population, particularly in developing nations, consuming it [1]–[4]. Paddy illnesses can harm rice production, and one of the most extensively used methods for detecting diseases in paddy is to perform image processing techniques, especially on the leaves [1], [4]. The traditional method employs a manual eye-based inspection technique, in which paddy illnesses are manually monitored during all stages of growth, including vegetative, reproductive, and ripening. Before using image processing in agriculture, manual visual quality inspection was one of the preferred methods for detecting the presence of illnesses in rice plants.

Plant diseases are visually inspected using a manual inspection approach. However, this technique becomes unstable and inconsistent when dealing with several inspection duties in large agricultural areas. This is because the process is repetitive, time-consuming, and involves the expertise of professionals in the field of plant disease diagnostics [2], [5], [6]. Therefore, image processing is one of the most excellent techniques to improve traditional manual inspection methods in today's technological era. This research aims to develop an efficient method for detecting leaf diseases using image processing techniques.

*nazrin@unimap.edu.my

Furthermore, today's advances in machinery/automation and electronic technology have made these complex and time-consuming jobs easier. For example, image processing is well-known as a non-invasive approach. Furthermore, with the advent of the Internet of Things (IoT) technology, inspection and monitoring of plants over long distances and across large areas has become one of the most desirable platforms for implementing innovative farming technologies [7]–[9].

2. RELATED WORKS

2.1 Image Pre-processing

Image pre-processing is an essential part of the process that can improve segmentation accuracy. Image pre-processing improves the contrast of an input image while also enhancing the visibility of a specified region [10]. Image improvement, colour-based transformation, noise reduction, and scaling are only a few of the procedures included. They all have strengths and downsides that can be compared. Furthermore, it will place a greater emphasis on segmentation until illness categorisation and detection [2].

2.2 Image Segmentation

The technique of segmenting a pre-processed image into many segments in order to discover the image's regions of interest more easily is known as image segmentation. Image segmentation seeks to represent an image in a more meaningful way so that it can be processed/analysed further. In segmentation, K-means are used by most previous researchers [2], [5], [11], [12] due to the simple step and being widely compatible with most types of plant detection. In addition, the researchers state that K-Means clustering can support a large number of datasets images [7], [9], [13], [14]. However, performing image segmentation on input photos with low light intensity issues presents a number of challenges, lowering the quality of segmented images. The texture of the disease zone on the leaf, on the other hand, is another challenge to jump in order to improve the quality of the segmented image and the disease detection accuracy [2], [5].

2.3 Features Extraction

Feature extraction is a dimensionality reduction procedure that reduces a large set of raw data into smaller groupings for processing. The texture, shape, and colour characteristics of the images will be retrieved. Objects in input images typically come in a variety of shapes, motions, sizes, colours, and textures. In order to recognise illnesses in plant leaves, features extraction is also necessary. The feature extraction approach is used in agriculture to extract relevant information from input images, such as shape, colour, and texture. In a survey of previous works, GLCM and feature extraction (colour, shape, texture) is the most common method used in the extraction of process images [6], [11], [15]. Previous work used a hybrid method for feature extraction by combining the discrete wavelet transform (DWT), scale-invariant feature transform (SIFT), and GLCM [6]. The purpose is that more features can be extracted reliably.

Nevertheless, a poor feature extraction technique worsens image processing processes overall. This contains irrelevant features in the high dimension of retrieved features, requiring more processing time and energy. The consistency of retrieved features, lighting conditions, image scale, and rotation are all elements that affect the quality of the feature extraction operation.

3. MATERIAL AND METHODS

3.1 Image Processing Method

In most cases, image acquisition, image pre-processing, image segmentation, image features extraction, and image classification are the five main phases. Each step is significant since it contributes to improving the accuracy of the output images. Furthermore, image processing methods/algorithms and the platform on which they are implemented are critical in determining the accuracy of the processed input image and the computation outcomes. Image processing stages consist of 5 main stages as shown in Figure 1. For this research, the process will be highlight on 3 stages which are image preprocessing, segmentation, and features extraction.

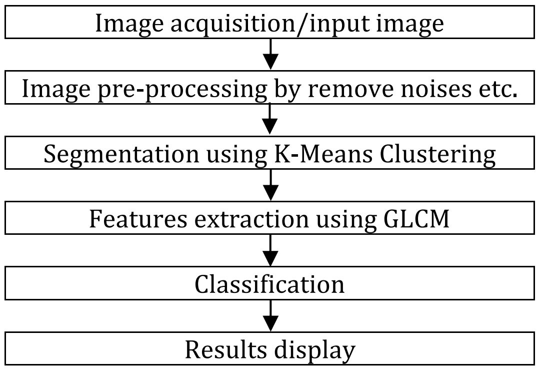


Figure 1. The process flow of proposed image processing methods/techniques.

3.1.1 Image Pre-processing

The image captured by the camera or input images will be resized to the desired size during the pre-processing process. This is because the input images are usually large in size and resolution. Thus, resized images can contribute to faster image processing and still providing good results. Then, the input image is prepared for the next stage, which is the segmentation of the input image.

3.1.2 Image Segmentation Process

For the proposed method, the K-Means clustering is chosen as the image segmentation algorithm. The reason is that the algorithm has been used commonly on plants compared to another algorithm. For example, Gayathri et al. [6] has implemented this image segmentation algorithm for paddy plant leaf, and the authors claimed that the segmented image is accurate by using the unsupervised machine learning algorithm. Therefore, the K-Means clustering will be used in the image segmentation process for this research.

3.1.3 Features Extraction Process

For the case of leaf-based disease detections, it could be detected through changes in colours on the leaves. Other notable features include variations in the leaf's texture and shape, which will also provide important clues for detecting plant diseases. The Gray Level Co-occurrence Matrix (GLCM) technique will be used in the image feature extraction process. Even previous work mainly combines more than two algorithms, and it is still possible to use only one algorithm because the paddy leaf structure is not too complicated compared to other types of object detection. The main changes for the paddy leaf are the colour structure of its leaf, which helps to detect the disease.

3.2 Datasets, Software, and Hardware

Image processing requires a hardware setup to process the infected leaves' images. In experimental result, the proposed work is implemented in MATLAB software and PC hardware with a specification of Intel Core i5 5200U (2 Cores 4 Threads, 2.20GHz) CPU, 10GB of RAM, and AMD Radeon R5 (M330, 2GB) GPU. The input images in this experiment are from the IRRI datasets, which are the collection of captured paddy disease images. Previous research principally collects images from PlantVillage datasets and IRRI (International Rice Research Institute) collections.

4. RESULTS AND DISCUSSION

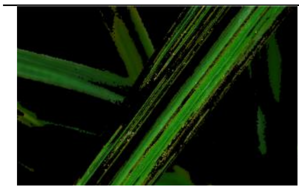
This research will be focusing on the image pre-processing, segmentation, and features extraction phase. Gayathri et al. [6] stated that leaf blast, leaf blight, false smut, brown spot, and leaf streak are the most common diseases found in paddy leaves. Therefore, this experiment will focus on three significant diseases: leaf blast, leaf blight, and leaf streak, which will go through the process of pre-processing, image segmentation, and features extraction of images.

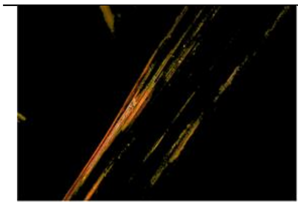
4.1 Image Pre- processing and Segmentation

K-means clustering is a method of separating groups of things in which each object is assigned a spatial position. It identifies divisions that keep things inside each cluster as close together as possible while keeping them as far apart as possible from objects in other clusters. Thus, it requires a distance metric to measure how close two items are to each other, as well as the number of clusters to be partitioned. The events of the image segmentation process using the L*a*b* color space and K-means clustering in MATLAB were explained in this section.

During the pre-processing, the input images will perform image adjustments such as contrast improvement, image resizes, and others. In this testing, the input images will be resized to 250x250 pixels to speed up the process for the next stage to be executed, and the input images will convert to Red, Green, Blue (RGB) color space then convert to L*a*b color space. The segmentation process for three forms of paddy leaf diseases, bacterial leaf streak, bacterial leaf blight, and bacterial leaf blast, is shown in Table 1, Table 2, and Table 4.

Table 1 Bacterial leaf streak using the proposed method

Process	Details
	- Input image of the diseased leaf - The input image will be resized to the desire resolution
	- Input image convert from RGB colour space to L*a*b colour space - Classify the colours of the input image in 'a*b' space using K-Means clustering
	- Create images that segment the input image by colour - Separate the objects in the input image by colour - Results will divide into three sections of images - Object in cluster 1

	- Object in cluster 2
	- Object in cluster 3 - The final segmented image is created and displayed

Based on the previous research by Manu [16], the process to classify the unwanted parts and diseases parts uses Otsu's Threshold segmentation and K-Means clustering in which each object is assigned a spatial position. Figure 2 shows Manu's method for segmentation and pre-processing of the input image, while Figure 3 shows the process of Manu's segmentation method. This can be compared to the proposed experiment shown in Figure 4 and Figure 5.

Figure 2. Manu's method for segmentation and pre-processing of the input image [16].

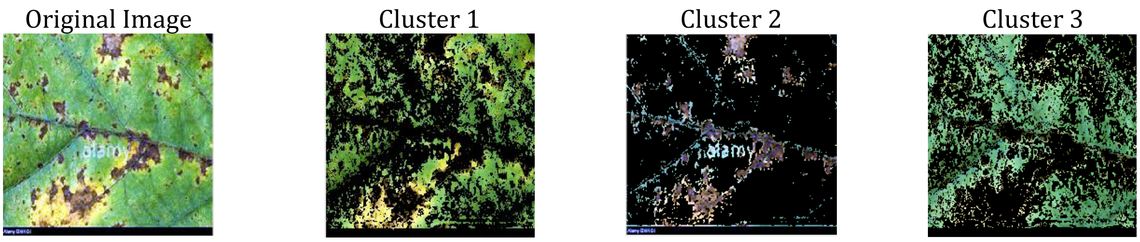


Figure 3. Process of Manu's segmentation method [16].

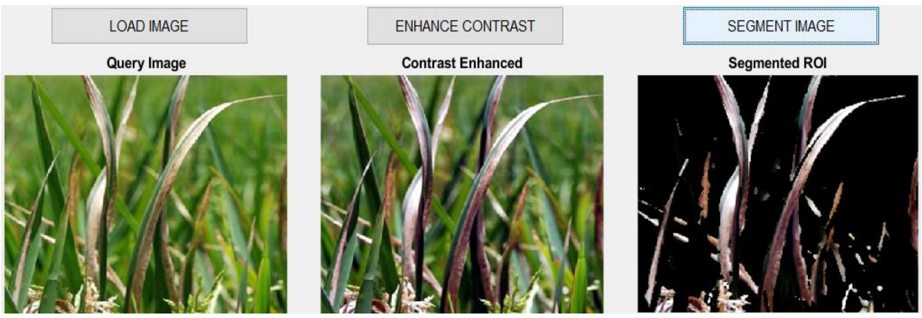


Figure 4. Own input image using Manu's method.

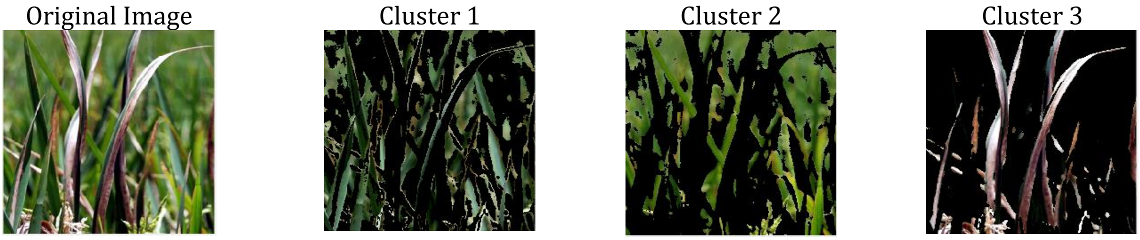


Figure 5. Own input image using Manu's segmentation method.

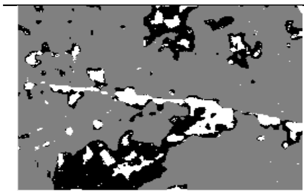
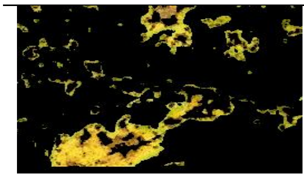
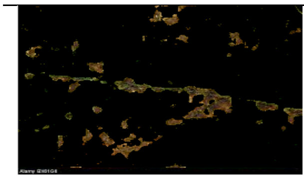
The experiment was carried out by comparing the results of Manu's input image [16] with the proposed input image using the proposed method, which only used one algorithm compared to Manu's segmentation method. Table 2 and Table 3 show the comparison between the proposed input image against Manu's input image using the proposed method.

Table 2 Bacterial leaf blight using the proposed method

Process	Details
	- Input image of the diseased leaf - The input image will be resized to the desire resolution
	- Input image convert from RGB colour space to L*a*b colour space - Classify the colours of the input image in 'a*b' space using K-Means clustering
	- Create images that segment the input image by colour - Separate the objects in the input image by colour - Results will divide into three sections of images - Object in cluster 1
	- Object in cluster 2

	- Object in cluster 3 - The final segmented image is created and displayed

Table 3 Bacterial leaf blight of Manu's input image using the proposed method

Process	**Details**
	- Input image of the diseased leaf - The input image will be resized to the desire resolution
	- Input image convert from RGB colour space to L*a*b colour space - Classify the colours of the input image in 'a*b' space using K-Means clustering
	- Create images that segment the input image by colour - Separate the objects in the input image by colour - Results will divide into three sections of images - Object in cluster 1
	- Object in cluster 2
	- Object in cluster 3 - The final segmented image is created and displayed

Refer to previous research by Gayathri et al. [6], using K-Means clustering can classify the unwanted parts and diseases parts in which each object is assigned a spatial position. Figure 6 shows the pre-processing image of the input image done by Gayathri et al. This can be compared to the proposed experiment shown in Table 4.

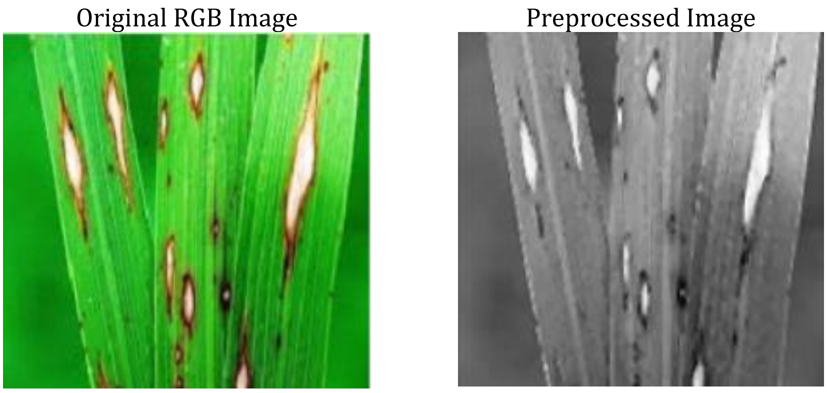


Figure 6. Pre-processing of input image [6].

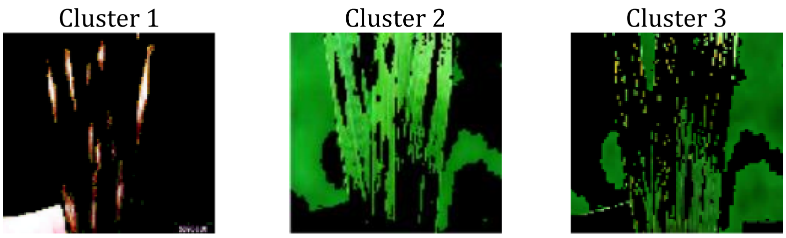
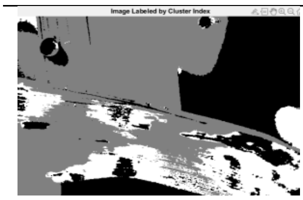
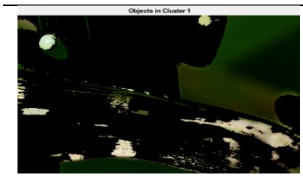
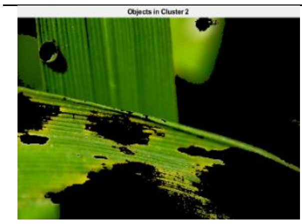
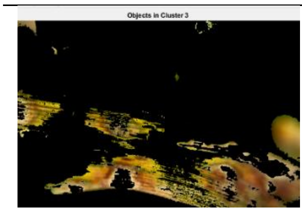


Figure 7. Segmentation of K-Means clustering [6].

After the pre-processing part, the experiment continues with the segmentation process done by Gayatri et al. [6] in Figure 7 compared to Table 4 of the proposed experiment. The segmented images will be used in the following process, which is features extraction of images.

Table 4 Bacterial leaf blast using the proposed method

Process	**Details**
	- Input image of the diseased leaf - The input image will be resized to the desire resolution
	- Input image convert from RGB colour space to L*a*b colour space - Classify the colours of the input image in 'a*b' space using K-Means clustering
	- Create images that segment the input image by colour - Separate the objects in the input image by colour - Results will divide into three sections of images - Object in cluster 1

	- Object in cluster 2
	- Object in cluster 3 - The final segmented image is created and displayed

4.2 Features Extraction

The GLCM algorithm is used to extract features from the input photos in this part. Each input image's data is gathered based on contrast, correlation, energy, and homogeneity fields. These data will be used in the classification process to ensure that the input photos can be compared to the train images and that the detection can be established.

The features extraction data from the image of three types of paddy leaf diseases, bacterial leaf streak, bacterial leaf blight, and bacterial leaf blast, are shown in Figure 8 (a), Figure 8 (b), and Figure 8 (c). This step starts once the segmentation phase is finished and the segmented image is used. Based on the results, the GLCM itself can be used to collect the features of the segmented image compared to a previous work that extracted the segmented images using hybrid methods of SIFT, DWT, and GLCM algorithm [6].

Another previous work extracted the segmented images using the combination algorithm of the colour co-occurrence method, SGDM, HIS, and GLCM [16], [17]. Even though the hybrid methods can increase the accuracy of data, this also will take time to process the segmented images to collect the features of the images.

```
stats =

  struct with fields:

       Contrast: 0.1041
    Correlation: 0.5998
         Energy: 0.8794
    Homogeneity: 0.9760


features =

    0.1041    0.5998    0.8794    0.9760
```

(a)

```
stats =

  struct with fields:

       Contrast: 0.3467
    Correlation: 0.7129
         Energy: 0.7884
    Homogeneity: 0.9617


features =

    0.3467    0.7129    0.7884    0.9617
```

(b)

```
stats =

  struct with fields:

       Contrast: 0.2959
    Correlation: 0.9219
         Energy: 0.6425
    Homogeneity: 0.9716


features =

    0.2959    0.9219    0.6425    0.9716
```

(c)

Figure 8. Data for bacterial (a) leaf streak, (b) leaf blight, and (c) leaf blast.

5. CONCLUSION

In conclusion, different image processing algorithms for the detection of plant diseases are explored. Each step of their procedures is compared to the previous one. Pre-processing techniques are used throughout the survey to improve segmentation accuracy. K-Means clustering in segmentation and GLCM in texture features are suitable algorithms to use for paddy disease detection. These methods are frequently utilised because they can identify various plant diseases, including paddy/rice disease, which is now under investigation. It is suitable for paddy disease detection to use multiclass SVM in classification, which will be used in future experiments to compare the accuracy and efficiency of the detection.

The detection's accuracy can be enhanced by tweaking key stages, including segmentation, feature extraction, and classification. Since this experiment was carried out, it has been proved that using a single algorithm of each stage can manage the detection process with almost similar output. Each technique can be hybridised or modified with other processes in the future, resulting in more characteristics and more consistent output.

In the future, IoT applications are possible, and they can also be integrated into image processing to detect diseases in real-time. The system can be measuring the efficiency of the detection by testing the classification process using the percentage of accuracy between training and real-time data or images.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Ministry of Higher Education Malaysia (MOHE) for the financial support of Modelling of a Soil Nutrients-EC Correlationship for Paddy Field Nutrient Treatment Map using the Internet of Things (IoT) connectivity research under the Fundamental Research Grant Scheme (FRGS) with grant number of FRGS/1/2018/TK04/UNIMAP/02/6.

REFERENCES

[1] S. Kumar.S and B.K.Raghavendra, "Diseases detection of various plant leaf using image processing techniques : a review," in 5th International Conference on Advanced Computing & Communication Systems, (2019) pp. 313–316.

[2] G. Dhingra, V. Kumar, and H. D. Joshi, Study of digital image processing techniques for leaf disease detection and classification, Multimed Tools Appl. vol. **77**, (2018) pp. 19951–20000.

[3] M. A. A. Kahar, S. Mutalib, and S. Abdul-Rahman, Early detection and classification of paddy

diseases with neural networks and fuzzy logic, Recent Adv. Math. Comput. Methods vol. **2**, no. 1 (2010) pp. 248–257.

[4] S. Ramesh and D. Vydeki, Recognition and classification of paddy leaf diseases using optimized deep neural network with jaya algorithm, Inf. Process. Agric., (2019) pp. 1–12.

[5] Z. Iqbal, M. Attique, M. Sharif, J. Hussain, M. Habib, and K. Javed, An automated detection and classification of citrus plant diseases using image processing techniques : A review, Comput. Electron. Agric. vol. **153**, (2018) pp. 12–32.

[6] T. Gayathri Devi and P. Neelamegam, Image processing based rice plant leaves diseases in Thanjavur, Tamilnadu, Cluster Comput. vol. **22**, no. s6 (2019) pp. 13415–13428.

[7] P. G. Shinde, A. K. Shinde, A. A. Shinde, and S. P. Borate, Plant disease detection using raspberry pi by k-means clustering algorithm, Int. J. Electr. Electron. Comput. Syst. vol. **5**, no. 1 (2017) pp. 92–95.

[8] A. Kapoor, S. I. Bhat, S. Shidnal, and A. Mehra, "Implementation of IoT (Internet of Things) and image processing in smart agriculture," in International Conference on Computation System and Information Technology for Sustainable Solutions, (2016) pp. 21–26.

[9] R. Kamath and M. Balachandra, Raspberry pi as visual sensor nodes in precision agriculture : a study, IEEE Access vol. **7**, (2019) pp. 45110–45122.

[10] M. Grand-Brochier, A. Vacavant, G. Cerutti, C. Kurtz, J. Weber, and L. Tougne, Tree leaves extraction in natural images: Comparative study of preprocessing tools and segmentation methods, IEEE Trans. Image Process. vol. **24**, no. 5 (2015) pp. 1549–1560.

[11] M. Sankar, D. N. Mudgal, T. V. Jagdish, N. W. G. Laxman, and M. M. Jalinder, "Green leaf disease detection using raspberry pi," in 1st International Conference on Innovations in Information and Communication Technology, (2019) pp. 1–6.

[12] S. Ramesh and D. Vydeki, "Rice blast disease detection and classification using machine learning algorithm," in Proceedings - 2nd International Conference on Micro-Electronics and Telecommunication Engineering, (2018) pp. 255–259.

[13] S. Mutalib, M. H. Abdullah, S. Abdul-Rahman, and Z. A. Aziz, "A brief study on paddy applications with image processing and proposed architecture," in IEEE Conference on Systems, Process and Control, (2016) pp. 124–129.

[14] V. P. Kour and S. Arora, Particle swarm optimization based support vector machine (P-SVM) for the segmentation and classification of plants, IEEE Access vol. **7**, (2019) pp. 29374–29385.

[15] K. J. Shalini, M. D. Hema, J. P. Kavya, B. N. Meghana, and M. C. Jeevitha, Leaf disease detection by using image processing, Int. J. Innov. Res. Sci. Eng. Technol. vol. **7**, no. 5 (2018) pp. 4462–4508.

[16] M. B.N, Plant Leaf Disease Detection and Classification using Multiclass SVM Classifier, MATLAB Central File Exchange, (2016) [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/55098-plant-leaf-disease-detection-and-classification-using-multiclass-svm-classifier. [Accessed: 01-Jul-2021].

[17] S. D. Khirade and A. B. Patil, "Plant disease detection using image processing," in Proceedings - 1st International Conference on Computing, Communication, Control and Automation, (2015) pp. 768–771.