Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

AN ANALYSIS OF PLANT DISEASES USING NEURAL NETWORK FOR INCREASING ECONOMIC GROWTH OF FARMERS

Dr. PRASAD M R
Associate Professor, Department of CSE, Vidyavardhaka College of Engineering, Mysuru.

Dr. KAVITHA K J
Associate Professor, ECE Department, GM Institute of Technology, Davangere.
Email: kavithakj219@gmail.com

RASHMI M HULLAMANI
Assistant Professor, ETE Dept., JNNCE, Shimoga, Karnataka.

RACHANA M HULLAMANI
Assistant Professor, EEE Department, DSCE, Bengaluru, Karnataka.

SUPRITH P G
Assistant Professor, ECE Department, GM Institute of Technology, Davangere.

RAVIRAYAPPA
Assistant Professor, Jain Institute of Technology, ECE Department, Davangere.

HANUMANTHRAJU R K
Assistant Professor, ECE Department, GM Institute of Technology, Davangere.

Abstract

India is an agricultural country wherein most of the population depends on agriculture. About 70% of the India economy depends on agriculture. Due to environmental changes the crops get heavily affected and characteristics symptoms such as leaf spot, dryness, and colour change and defoliation occurs. This paper discusses the development of automatic detection system using advanced computer technology such as image processing and neural network help to support the farmers in the identification of diseases at an early or initial stage and provide useful information for its control. Automatic detection of plant diseases is a very important research topic as it may prove the benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves. Machine learning based detection and recognition of plant diseases can provide extensive clues to treat the diseases in its very early stages. Comparatively, visually or naked eye identification of plant diseases is quite expensive, inefficient, inaccurate and difficult. Also, it requires the expertise of a well-trained botanist. In this paper, an approach is proposed to find the interactions between the disease causing agents and host plant in relation to overall environment; to identify various diseases in plants and to implement a method for preventing the diseases and preventing management for reducing the losses/ damages caused by diseases so as to prevent diseases on plants for the farmers benefits and to help out pesticide company in predicting the new pesticide solutions.

Keywords: Neural Network, Plant Diseases, Agriculture, Machine Learning

1. INTRODUCTION

The most common plant diseases occur in India is classified as per their features available on the leaf or fruit. The color, the texture and the morphological features are used to train a neural network model. Diseases occur in plants under the influence of various factors—

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

pathogens and unfavorable environmental conditions; and are manifested in the disturbance of function (photosynthesis, respiration, synthesis of tissue and growth substances, and the flow of water and nutritive substances), and the structure of the organism, causing premature destruction of the plant or affecting some of its organs. There is not yet a precise and comprehensive definition of plant diseases. In the early stages of the development of Phyto-pathology, any deviation from the normal condition of a plant was considered plant disease. The inadequacy of this definition lay in the difficulty of distinguishing a normal (healthy plant) from an abnormal (diseased plant) condition. The determination of the presence of a pathological process in a plant organism made it possible to redefine plant disease in a new way and to conceive it not as a static condition but as a dynamic process that arises and develops as a result of interaction of the plant with its environment.

Plant diseases diminish yields and impair the quality of plant production. For instance, in years favourable to the spread of phytophthora infection of potatoes, the yield of tubers is decreased by 15 to 20 percent and, in some regions, by 50 percent or more. More than 30,000 separate plant diseases are known. They are classified by symptoms or types (patho-graphic classification), by the plants affected (plant-growing classification), and by causes, or causative agents, of the disease (etiological classification). The last, according to which plant diseases are divided into non-infectious and infectious, plays a leading role. Non-infectious plant diseases are caused mainly by abiotic factors in the environment: disruptions in the regime of mineral feeding, most often by a deficiency (rarely, a unilateral excess) of macro elements (nitrogen, phosphorus, potassium, and magnesium) or a deficiency of microelements, especially boron, zinc, iron, copper, and molybdenum; an unfavourable water regime (deficiency or excess of water in the soil, prolonged rains, or high relative humidity of the air), causing “bleeding” of plants, premature drying up, premature withering of plants, or leaves falling under conditions of water deficiency; or the effects of high or low temperatures on plants, abrupt changes in air and soil temperatures (freezing of shoots, frost cracks, chilling of heat-loving plants in greenhouses and hotbeds or during irrigation of the soil with cold water, and so forth). Causes of non-infectious plant diseases may be harmful impurities in the air and soil (blight and falling of leaves from the effects of sulphur dioxide gas, for example, in the vicinity of metallurgical and chemical plants); residual effects of certain herbicides carried into the soil; an unfavourable light regime, mainly a deficiency of light in greenhouses and hothouses (chlorosis and lodgement or dwarfing with a shortened day); ionizing radiation (alpha, beta, and gamma rays, X rays, and neutrons); or toxins excreted into the soil by certain fungi (species of Fusarium, Botrytis, and so forth) and some higher plants.

Plant pathology (also phytopathology) is the scientific study of plant diseases caused by pathogens (infectious diseases) and environmental conditions (physiological factors). Plant diseases may be broadly classified into three types. They are bacterial, fungal and viral diseases. Plant diseases have turned into a dilemma as it can cause significant reduction in both quality and quantity of agricultural products. It is estimated that 2007 plant disease losses in Georgia (USA) is approximately $539.74 million. Of this amount, around 185 million USD was spent on controlling the diseases, and the rest is the value

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

of damage caused by the diseases. The naked eye observation of experts is the main approach adopted in practice for detection and identification of plant diseases .However; this requires continuous monitoring of experts which might be prohibitively expensive in large farms. Further, in some developing countries, farmers may have to go long distances to contact experts, this makes consulting experts too expensive and time consuming Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves. Therefore; looking for fast, automatic, less expensive and accurate method to detect plant disease cases is of great realistic significance. Machine learning based detection and recognition of plant diseases can provide clues to identify and treat the diseases in its early stages. Comparatively, visually identifying Using this new tool helps to improve the images from microscopic to telescopic range and also offers a scope for their analysis. It, therefore, has many applications in biology. However, as is the case with any new technology, imaging technology also has to be optimised for each application, since what each user is looking for in an image is quite unique. Plant diseases is expensive, inefficient, and difficult. Also, it requires the expertise of trained botanist Images of the leaves, captured by a camera or a scanner for Colour image analysis for estimation of normal leaf, infected leaf and chlorophyll. Many times a viral or a fungal attack on plants results in degradation of chlorophyll pigments in leaves. Such infected leaves have patches of green and yellow colour. In plant breeding, it is important to quantify the leaf infection. Thus the extent of infection can be quantified without much efforts. Plant leaf colour is also commonly used as an indication of health status of plants. The loss of chlorophyll content of leaves occurs due to nutrient imbalance, excessive use of pesticides, environmental changes and ageing.

2. LITERATURE SURVEY

[1] Proposed an algorithm to classify the diseases that appear in apple fruit the work concentrated mainly on three diseases that appear in apple namely apple scab, apple rot and apple blotch. The work proposed a hybrid feature extraction technique. Colour features were obtained by making use of global colour histogram and colour coherence vector and texture feature were obtained using local binary pattern and complete local binary pattern. The features obtained by both methods were fused together. Random forest algorithm was used to classify the disease. Further K-means algorithm was used to identify the infected area in the fruit. The proposed work was trained only for 70 images and efficiency obtained was 80%.

[2] Grapes are one of the commercial crops in India and this crop is subjected to diseases like anthracnose, powdery mildew and downy mildew. The proposed work helps to identify and classify the disease using image pre-processing and machine learning. Different images were captured under uncontrolled environments. The captured images were subjected to image pre-processing and image enhancement was done to remove the noises present the image. As the images were captured under different environment conditions, sun spot was removed using threshold values greater than

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

150. Texture feature extraction was done using GLCM techniques. Feature extracted were trained for classification using SVM, RF and BPN techniques. The paper concluded that RF gave more accuracy of 86% compared to rest of the image classification algorithms.

[3] Proposed an algorithm to identify three fruits namely apple, strawberry and oranges. The captured images were pre-processed by resizing the images. Two feature extraction techniques were used, first the shape and colour were considered and second involved scale invariant feature transformation (SIFT). The features extracted were used to train RF, SVM and KNN algorithm to identify the fruit images. The work concluded the combination of feature extraction with Random Forest algorithm gives highest accuracy of 94% compared to other algorithms.

[4] Potato is one of the prominent crop and is subjected to disease like early blight and late blight. The proposed work uses machine learning to identify the diseases in potato. Different images pertaining to two diseases are captured and pre-processed by resizing and converting RGB to HSV. Image segmentation was done by generating mask by using colour information, intensity of colour and brightness of HSV colour space. Feature extraction was done based on colour, texture and shape features. Colour histogram was used to find colour intensity of image. GLCM was used for extracting texture feature and Hu moment was used for extracting shape features. Different classifiers were used to identify and classify the diseased leaf. Random forest gave high accuracy of 97%. The work was limited to identify only two diseases and lesser number of input images was used to train the algorithm.

[5] Plant diseases are significant reason for low profitability. Proposed work consists of an algorithm to identify diseases like common rust, grey hair and leaf damage that occur in maize. Total 3823 images were collected from internet to form the data set. Image pre-processing was done to reduce the noises in the image by converting RGB image to grey. Feature extraction was done based on shape colour and texture. These extracted features used to train different algorithms to identify and classify the diseases. Random forest gave highest accuracy of 80.68.

[6] Tomato plants provide a lot of health benefits and hence require monitoring the health of the plant. The proposed works uses IOT and Deep learning techniques to monitor health of the plant. A neural network model was developed and was trained over 200 images of tomato. The leaf images captured using camera was taken as test images. The trained model outputs a prediction and class with greatest probability is considered as predicted disease. The proposed work provided an overall efficiency of 92%. The proposed work limited to identify only two diseases namely tomato leaf curl and tomato septorial spot.

[7] Plant diseases often appear on the leaf and it becomes difficult to distinguish and identify the particular disease manually. Hence author proposed CNN based squeeze net architecture to identify the disease in tomato leaf. Nearly 1400 images pertaining to seven different diseases in tomato were collected to farm data set. These images were pre-processed by resizing the images in order to eliminate the noise from the

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

image. Then the images were trained using CNN model based on squeeze net architecture. The model developed was able to identify and classify the diseases and overall efficiency of 87% was obtained.

[8] Proposed an algorithm using Random forest to identify and classify healthy and diseased leaf in papaya plant. 160 images of papaya leaf were collected for training the algorithm. The collected images were subjected to pre-processing to reduce the noise. After pre-processing the feature extraction was done using HOG feature extraction technique. The extracted features were then used to train Random Forest model to identify and classify the healthy and diseased leaf. The algorithm provided an accuracy of only 70%. Some more literature on various plant diseases is shown in table 1 and in the next section; we discuss the proposed algorithm for plant disease detection.

Table1: Summarization of different techniques of leaf disease detection

Work	Type of Plant	Type of Disease	No. of images	Accuracy	Features	Advantages	Limitations
Samajpati, B. J[1] et al.	Apple fruit	Apple scab, apple rot, apple blotch	70	80%	Colour features, texture feature	Multiple feature extraction are combine together to obtain better results	Results obtained are for small database.
Sandika, B[2]	Grapes	Anthracnose, powdery mildew, downy mildew	350	86%	Texture feature	Work is computationally efficient	Classification degrades for images captured with high brightness
Zawbaa, H. M[3]	Apple, strawberry and oranges	Classify Apple, strawberry and oranges	178	94%	Shape and colour	Work is computationally efficient	The algorithm is more time consuming
Iqbal, M. A.[4]	Potato	Early and late blight	450	97%	Shape, colour and texture	Multiple classifiers algorithm used to identify disease	Results obtained are for small database.
Chauhan, M[5]	Maize	Common rust, grey hair and leaf damage	3823	80.68	Shape, colour and texture	Approach requires less time	The algorithm cannot be used for all data sets
Deepak, A. H.[6]	Tomato	Leaf curl and tomato septorial spot.	200	92%	Texture	Also Monitors temperature, soil moisture, humidity	Algorithm suffers from oversized data set
Hidayatuloh, A [7]	Tomato	Early and late blight, leaf miner, calcium and magnesium deficiency	1400	87%	Shape and texture	Compatible for mobile phones	Approach is slow and time consuming
Hatuwal, B. K [9]	Plants	Early and late scorch, yellow and brown spot	30000	87%	Texture	Multiple classifiers algorithm used to identify disease	Classification method can be further improved

3. PROPOSED METHODOLOGY

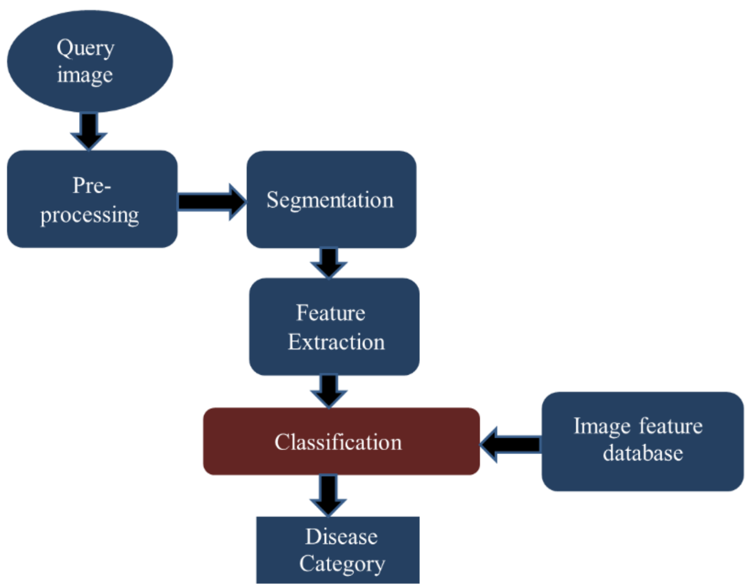
The underlying approach for all of the existing techniques of image classification is almost the same. First, digital images are acquired from environment around the sensor using a digital image. Then image-processing techniques are applied to extract useful features that are necessary for further analysis of these images.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

After that, several analytical discriminating techniques are used to classify the images according to the specific problem at hand. This constitutes the overall concept that is the framework for any vision related algorithm. The figure 1 depicts the basic procedure of the proposed vision-based detection algorithm in this research.

First, the images of various leaves are acquired using a digital camera. Then image processing techniques are applied to the acquired images to extract useful features that are necessary for further analysis. After that, several analytical techniques are used to classify the images according to the specific problem at hand.

Figure 1: Basic procedure of the proposed vision-based detection algorithm



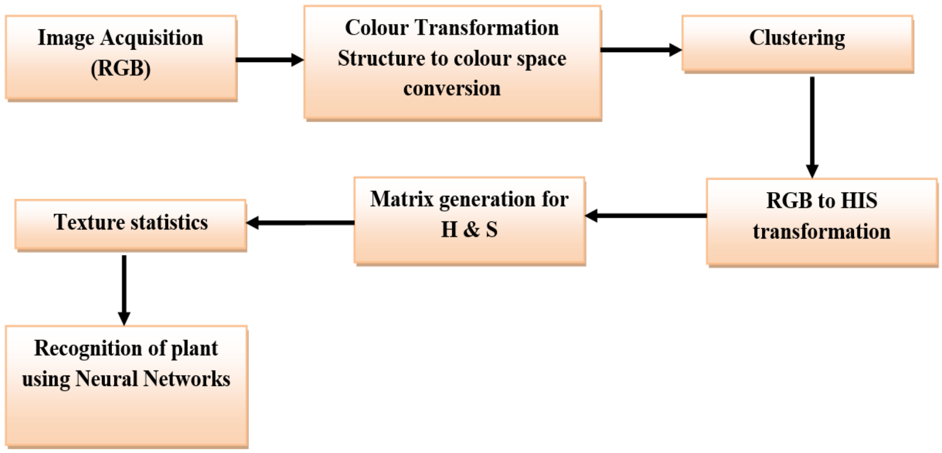
The first phase is the image acquisition phase. In this step, the images of the various leaves that are to be classifies are taken using a digital camera. In the second phase image pre-processing is completed. In the third phase, segmentation using K-means clustering is performed to discover the actual segments of the leaf in the image. Later on, feature extraction for the infected part of the leaf is completing based on the specific properties among pixels in the image or their texture.

After this step, certain statistical analysis tasks are completed to choose the best features that represent the given image, thus minimizing feature redundancy. Finally, classification is completed using neural network based algorithm.

The detail step-by-step account of the image acquisition and classification process is shown in figure 2. In the initial step, the RGB images of all the leaf samples were obtained. Some samples of diseased leaf images are taken.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

Figure 2: Image Acquisition and Classification Process



For each image in the data set the subsequent steps were repeated. Image segmentation of the leaf is done on each image of leaf sample using K-means clustering. A sample clustered image with four clusters of the leaf sample image is shown in figure 4.2. In this experiment multiple values of number of clusters are considered. Best results can obtain when the number of clusters are 4.

Once the infected objects are determined, the image is then converted from RGB format to HSI format. The SGDM matrices are then generated for each pixel map of the image for only H and S images. The SGDM is a measure of the probability that a given pixel at one particular gray-level will occur at a distinct distance and orientation angle from another pixel, given that pixel has a second particular gray-level. From the SGDM matrices, the texture statistics for each image were generated.

We propose neural network - based software for the automatic leaf diseases identification and classification. The proposed algorithm is tested on five different diseases which effect on the plants. They are cassava bacterial blight, cassava leaf spot, phoma blight, and bacterial canker, red rust and sooty mould. Identification and recognition of leaves diseases are likely to give better performance and provide solutions to treat the diseases in its early stages. Visual interpretation of plant diseases manually is both inefficient and difficult; also it requires a trained botanist.

A closer inspection of the plant diseases images reveals several difficulties for the possible leaves diseases identification. The developed system classifies the leaves into infected and non-infected classes. The overall concept that is the framework for any vision related algorithm of image classification is almost the same. First, the digital images are acquired from the environment using a digital camera. Then image-processing techniques are applied to the acquired images to extract useful features that are necessary for further analysis. After that, several analytical discriminating techniques are used to classify the images according to the specific problem at hand.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

The proposed system can:

- Identify disease type
- Deal with other diseases
- Identify and classify diseases that infect plant leaves.
- Provide advice to treat the diseases in its early stages

First, the images of various leaves are acquired using a digital camera. Then image processing techniques are applied to the acquired images to extract useful features that are necessary for further analysis. The steps involved in recognition and classification are image acquisition, pre-processing, feature extraction, segmentation and classification. Algorithm: Basic steps describing the proposed algorithm:

- RGB image acquisition.
- Create the colour transformation structure.
- Convert the colour values in RGB to the space specified in the colour transformation structure.
- Apply K-means clustering
- Masking green-pixels.
- Remove the masked cells inside the boundaries of the infected clusters.
- Convert the infected (cluster / clusters) form RGB to HSI Translation.
- SGDM Matrix Generation for H and S. (Another name for gray-level co-occurrence.
- Matrix is gray-level spatial dependence matrix.
- Calling the GLCM function to calculate the features.
- Texture Statistics Computation.
- Configuring Neural Networks for Recognition.

In details, in step 2 a colour transformation structure for the RGB leaf image is created, and then, a device-independent colour space transformation for the colour transformation structure is applied in step 3. Steps 2 and 3 are inevitable for carrying out step 4. In this step the images at hand are segmented using the K-Means clustering technique [10; 7; 3; 9]. These four steps constitute phase 1 whereas, the infected object (s) is/are determined. In step 5, we identify the mostly green coloured pixels. After that, based on specified and varying threshold value that is computed for these pixels using Otsu's method [12; 13], these mostly green pixels are masked as follows: if the green component of pixel intensities is less than the pre-computed threshold value, the red, green and blue components of the this pixel is assigned to a value of zero. This is done in sense that these pixels have no valuable weight to the disease identification and classification steps, and most probably those pixels represent healthy areas in the leave. Furthermore, the image processing time should become significantly reduced. In step 6 the pixels with zeros red, green and blue values and the pixels on the boundaries of the infected cluster (object) were completely removed. Steps 5 and 6 form phase 2, and this phase is helpful as it gives more accurate disease classification and identification results with satisfied performance and the overall computation time should become significantly less. The observations behind steps 5 and 6 were experimentally validated. Next, in step 7 the infected cluster was then converted from RGB format to HSI format. In the next step, the SGDM matrices were then generated for each pixel map of the image for only H and S images. The SGDM is a measure of the probability that a given pixel at one particular gray-level will occur at a distinct distance and orientation angle from another pixel, given that pixel has a second particular gray-level. From the SGDM matrices, the texture statistics for each image were generated. Concisely, the features set were computed only to pixels inside the boundary of the infected areas of the leaf. In other words, healthy areas inside the infected areas were also removed. Steps 7 – 10 form phase 3 in which the texture features for the segmented infected objects in this phase are calculated. Finally, the recognition process in the fourth phase was performed to the extracted features through a pre-trained neural network. For each image in the data set the subsequent steps in Algorithm 1 were repeated. Image segmentation process do partitioning of an image into segments/parts to better understand the image, it is commonly performed to determine boundaries or objects in an image. Segmentation can be performed by various methods namely; segmentation based on region, based on edges, and based on clustering. K-means clustering is one of the mostly used clustering techniques employed for segmentation task, it divides the data into k number of clusters, where each data value belongs to the cluster having the closest mean [3, 10]. Feature extraction is very important for differentiating objects of one class from another class; it is the process of transforming input data into useful set of features. FE is an essential step to extract the features of interested region of an image, there are three most common methods used for feature extraction are Texture-based, Colour-based, and Shape-based. How the colour distribution over the entire image, hardness, roughness, all these indicates texture features of the image.

4. RESULTS AND DISCUSSION

In our work a database of 800 images of diseased sunflower were collected as an input to the system. Data set consisted of 200 images of leaf blight, 200 images of powdery mildew, 200 images of leaf spot and 200 images of downy mildew. In the proposed work, the dataset was split into training phase, consisting of 60% image and rest of the images being used test images. For the classification of the diseases Random Forest was used as a classifier. The results obtained during this process were tabulated as shown in figure 3 to figure 7 and statistical results in table 2. Noise gets added during acquisition of leaf images. So we use different types of filtering techniques to remove noise. We create device independent colour space transformation structure. Thus we create the colour transformation structure that defines the colour space conversion. The next step is that we apply device-independent colour space transformation, which converts the colour values in the image to colour space specified in the colour transformation structure. The

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

colour transformation structure specifies various parameters of transformation. A device independent colour space is the one where the resultant colour depends on the equipment used to produce it. For example the colour produced using pixel with a given RGB values will be altered as brightness and contrast on display device used as shown in figure 3.

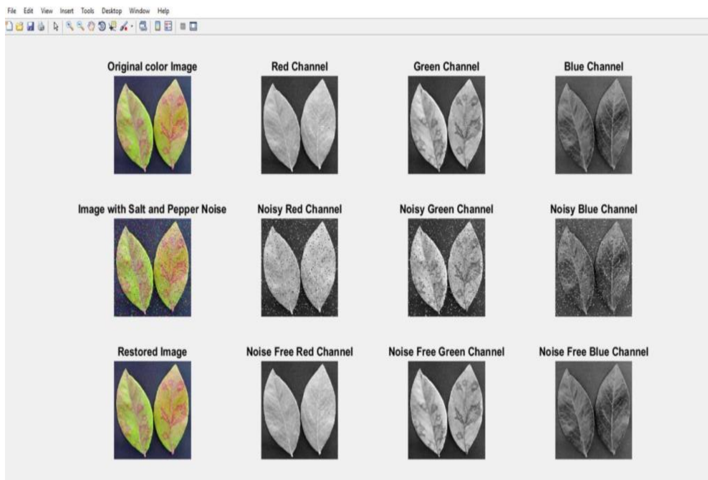


Figure 3: Pre-Processed Image

First, the RGB images of leaves are converted into Hue Saturation Intensity (HSI) colour space representation as shown in figure 4. The purpose of the colour space is to facilitate the specification of colours in some standard, generally accepted way. HSI (hue, saturation, intensity) colour model is a popular colour model because it is based on human perception [10]. Hue is a colour attribute that refers to the dominant colour as perceived by an observer. Saturation refers to the relative purity or the amount of white light added to hue and intensity refers to the amplitude of the light. Colour spaces can be converted from one space to another easily. After the transformation process, the H component is taken into account for further analysis. S and I are dropped since it does not give extra information. Figure 3 shows the H, S and I components.

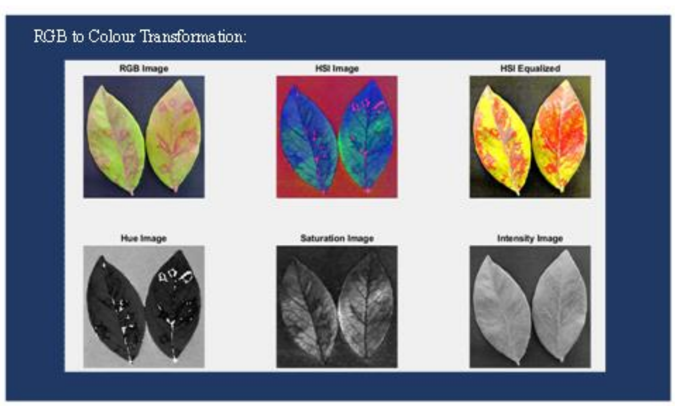


Figure 4: Colour Transformation structure and HSI images

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

In this step, we identify the mostly green colored pixels. After that, based on specified threshold value that is computed for these pixels, the mostly green pixels are masked as follows: if the green component of the pixel intensity is less than the pre-computed threshold value, the red, green and blue components of the this pixel is assigned to a value of zero. This is done in sense that the green colored pixels mostly represent the healthy areas of the leaf and they do not add any valuable weight to disease identification and furthermore this significantly reduces the processing time. With the help of texture features, plant diseases are classified into different types The pixels with zeros red, green, blue components as well as pixels on the boundaries of infected cluster are completely removed as shown in figure 5. This is helpful as it gives more accurate disease classification and significantly reduces the processing time. Infected cluster is converted from RGB to HSI color format.

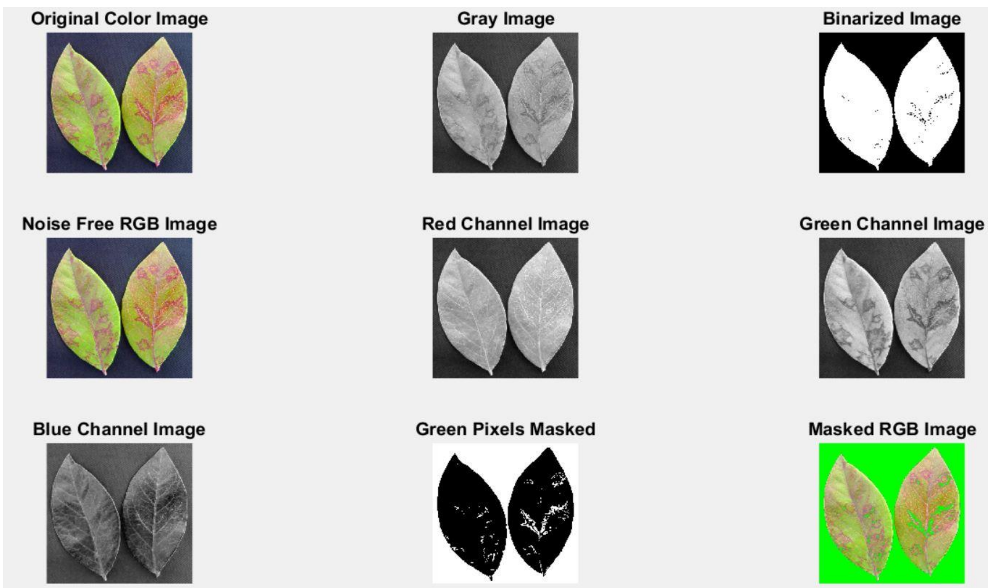


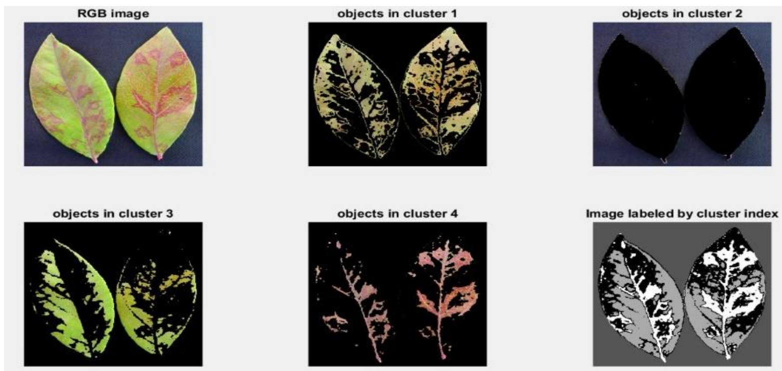
Figure 5: Masking green pixels

K-means clustering is used to partition the leaf into four clusters in which one or more clusters contain the disease in case when the leaf is infected by more than one disease. The K means clustering algorithms tries to classify objects (pixels in our case) based on a set of features into K number of classes. The classification is done by minimizing the sum of squares of distances between the objects and the corresponding cluster or class centroid. In this step, segmentation of images is done in order to separate the leaves from the background. Segmentation is performed using K-means clustering with 2 cluster centers, one for background and one for foreground. K-means clustering is unsupervised learning technique that is used to segregate the data points in the predefined number (k) of clusters or groups on the basis of their similarities. K –Means algorithm. The size of the patch is chosen in such a way that the significant information is not lost. In this approach patch size of 32×32 pixels is taken. The next step is to extract the useful segments. Not all segments contain significant amount of information. So the patches which are having

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

more than fifty percent of the information are taken into account for the further analysis and the results are shown in figure 6 & 7.

1) Pick center of K cluster, either randomly or based on some heuristic.
2) Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center.
3) Again compute the cluster centers by averaging all of the pixels in the cluster. Repeat steps 2 and 3 until convergence is attained.



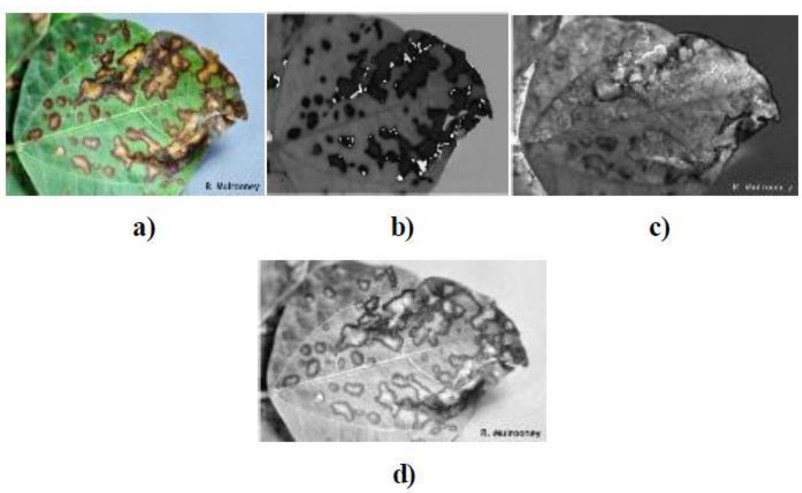


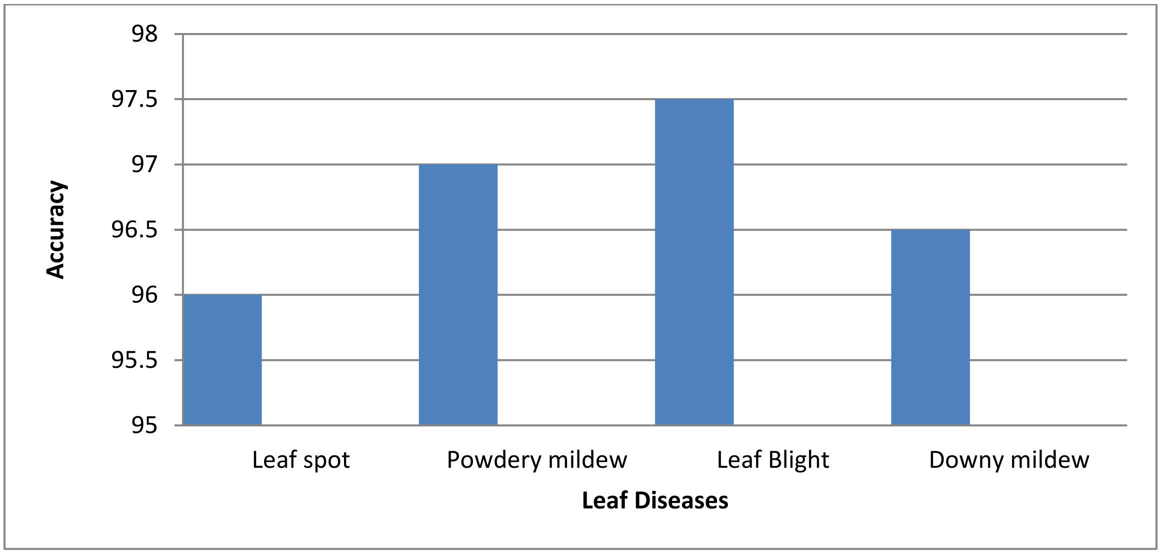
Figure 6: a) Input image infected by Bacterial Brown Spot b) Hue Component c) Saturation Component d) Intensity Component

Table 2: Classification results of Random Forest

Data set	Total test images	Classified images	Accuracy
Leaf spot	200	192	96%
Powdery mildew	200	194	97%
Leaf blight	200	195	97.5%
Downy mildew	200	193	96.5%

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

Figure 7: Accuracy results of the diseases classified using RF



Colour Co-occurrence Method (CCM) for texture analysis is also done and the use of colour image features in the visible light spectrum provides additional image characteristics features over the traditional gray scale representation.

- **Colour co-occurrence Method (CCM)**

In this method both colour and texture are taken into account to get a unique features for that image. For that the RGB image is converted into the HSI translation. For the texture statistics computation the SGDM matrix is generated and using GLCM function the feature is calculated.

The input image is enhanced by using anisotropic diffusion technique to preserve the information of the affected pixels before separating the colour from the background [8]. To distinguish between grape leaf and the non-grape leaf part, H and B components from HIS and LAB colour space is considered. A SOFM with back propagation neural network is implemented to recognize colours of disease leaf. The CCM methodology consists of three major mathematical processes.

First, the RGB images of leaves are converted into HSI colour space representation. Once this process is completed, each pixel map is used to generate a colour co-occurrence matrix, resulting in three CCM matrices, one for each H, S and I pixel maps. Hue Saturation Intensity (HSI) space is also a popular colour space because it is based on human colour perception. Hue is generally related to the wavelength of a light and intensity shows the amplitude of the light.

And saturation measures the colorfulness in HSI space. Colour spaces can easily be transformed from one to another. Following equations can be used to transform the images from RGB to HSI.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

$$Intensity\ (I) = \frac{R + G + B}{3}$$

$$Saturation\ (S) = 1 - \frac{3\ min\ (R, G, B)}{(R + G + B)}$$

$$Hue\ (H) = 2 - ACOS\left\{\frac{[(R-G)+(R-B)]}{\sqrt[2]{(R-G)^2 + (R-G)(G-B)}}\right\}, B > G$$

$$Hue\ (H) = ACOS\left\{\frac{[(R-G)+(R-B)]}{\sqrt[2]{(R-G)^2 + (R-G)(G-B)}}\right\}, B \leq G$$

The colour co-occurrence texture analysis method is developed through the Spatial Gray-level Dependence Matrices (SGDM). The gray level co-occurrence methodology is a statistical way to describe shape by statistically sampling the way certain gray-levels occur in relation to other gray levels. These matrices measure the probability that a pixel at one particular gray level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray level. For the position operator p, we can define a matrix Pij that counts the number of times a pixel with gray-level i occurs at position p from a pixel with gray-level j. SGDM's are generated for H image. The SGDM's are represented by the function P (i, j, d, θ) where i represent the gray level of the location (x, y) in the image I(x, y), and j represents the gray level of the pixel at a distance d from location (x, y) at an orientation angle of θ. The reference pixel at image position (x, y) is shown as an asterix. All the neighbors from 1 to 8 are numbered in a clockwise direction. Neighbors 1 and 5 are located on the same plane at a distance of 1 and an orientation of 0 degrees. In this research one pixel offset distance and a zero degree orientation angle is used. The RGB image is converted to HIS and then calculate the feature set for H and S, we drop the intensity (I) since it does not give extra information. However, we use GLCM function in MatLab to create Gray-Level Co-Occurrence Matrix. The number of gray-levels is set to 8 and the symmetric value is set to true and finally offset is given a 0 value.

➢ Otsu Threshold Algorithm

Thresholding creates binary images from grey-level images by setting all pixels below some threshold to zero and all pixels above that threshold to one. The Otsu algorithm
defined in

i. According to the threshold, Separate pixels into two clusters
ii. Then find the mean of each cluster.
iii. Square the difference between the means.
iv. Multiply the number of pixels in one cluster times the number in the other.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

- **Normalizing the CCM matrices**

The CCM matrices are then normalized using the equation below, where, (i, j, 1, 0) represents the intensity co-occurrence matrix:

- **Neural networks for recognition**

In this paper, neural networks are used in the automatic detection of leaves diseases. Neural network is chosen as a classification tool due to its well-known technique as a successful classifier for many real applications. The training and validation processes are among the important steps in developing an accurate process model using NNs. The dataset for training and validation processes consists of two parts; the training feature set which are used to train the NN model; whilst a testing features sets are used to verify the accuracy of the trained NN model. Before the data can be fed to the ANN model, the proper network design must be set up, including type of the network and method of training. Classification Using ANN after feature extraction is done, the learning database images are classified by using neural network. These feature vectors are considered as neurons in ANN The output of the neuron is the function of weighted sum of the inputs. The back propagation algorithm, modified SOM this was followed by the optimal parameter selection phase. However, this phase was carried out simultaneously with the network training phase, in which the network was trained using the feed-forward back propagation network. In the training phase, connection weights were always updated until they reached the defined iteration number or acceptable error. Hence, the capability of ANN model to respond accurately was assured using the Mean Square Error (MSE) criterion to emphasis the model validity between the target and the network output.

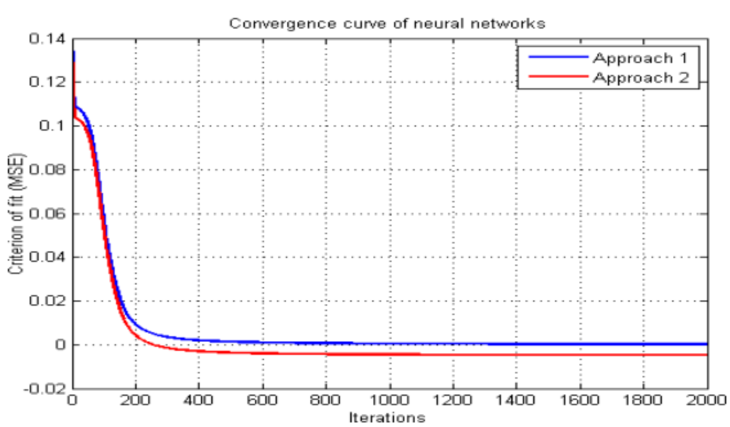


Figure 8: The convergence comparison curve proposed method with the existing method

The convergence curve for the learning step of the neural network in the proposed study in this work is better than that of [1] as shown in Figure 8. In Figure 8 Approach 1 represents the study presented in [1], while Approach 2 represents our proposed study. It can be concluded from the above tables and figures that the obtained results achieved an acceptable level of optimal results which can also add more weighting rate to the

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

proposed study. The presented approaches in this paper and in [1] have been implemented in MATLAB under Windows XP environment. All the experiments are carried out on a desktop computer with Intel (R) Pentium (R) CPU 2.20GHZ 645 MHZ and 256MB of RAM. The average computation time for the two proposed approaches was computed in seconds for the models M1, M2, M3, M4 and M5 as shown in Table 5. The data in Table 5 was obtained for the two approaches using the same neural network structure and under the same machine. Figure 9 shows that our proposed approach has 19% speedup over the approach of [1].

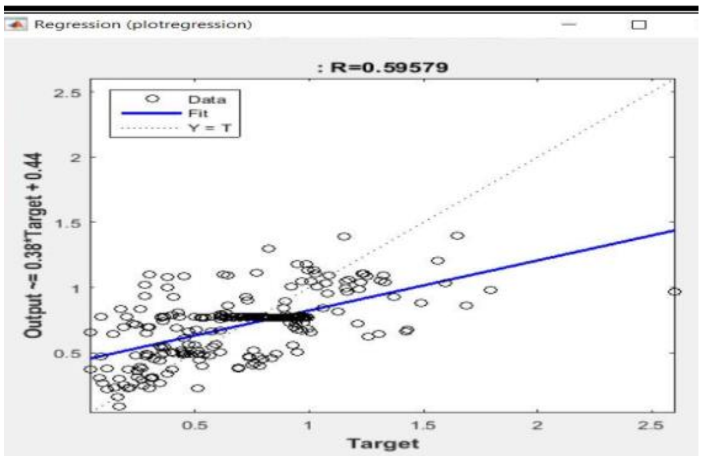


Figure 8: The convergence comparison curve proposed method with the existing method (Contd.)

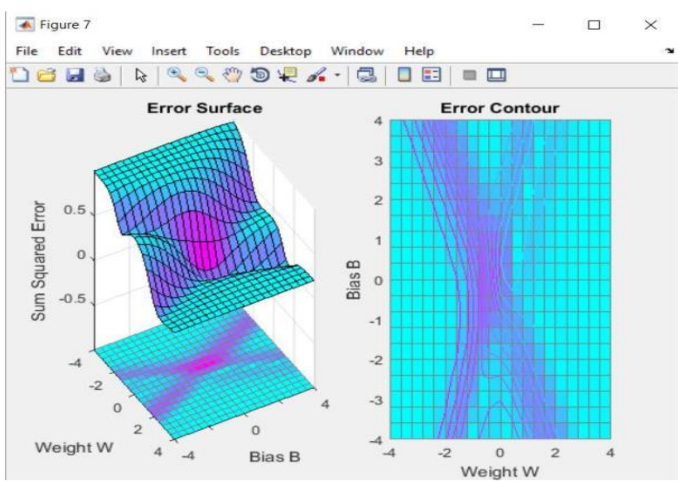


Figure 9: The speed up comparison between proposed methods with the existing method

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

5. CONCLUSION AND FUTURE ENHANCEMENT

In this paper, respectively, the applications of K-means clustering and Neural Networks (NNs) have been formulated for clustering and classification of diseases that effect on plant leaves. Recognizing the disease is mainly the purpose of the proposed approach. Thus, the proposed Algorithm was tested on five diseases which influence on the plants; they are: Early scorch, Cottony mold, ashen mold, late scorch, tiny whiteness. The experimental results indicate that the proposed approach is a valuable approach, which can significantly support an accurate detection of leaf diseases in a little computational effort. Recognizing the disease is mainly the purpose of the proposed approach which can recognize the leaf diseases with little computational effort. This approach can be used for the agricultural applications like detection & classification of diseases of plant parts like leaf with suitable classifier .This project will describes a possible approach for extraction of low level image feature like colour. This paper addresses how the disease analysis is possible for the cotton leaf diseases detection, the analysis of the various diseases present on the cotton leaves can be effectively detected in the early stage before it will damage the whole plant. The efficiency of the proposed work is about 80% and hence the model presented can able to detect the disease more accurately compare to the other classifiers. Future expansion of this work will be focused on following points:

1) To develop combinations of more algorithms by using fusion classification technique, so as to improve the detection rate of the classification process. 2) On the basis of detection of disease the proper mixture of fungicides will be provided to the farmer for further use in their farms. 3) To design an automated system with the help of embedded system so that this fungicide mixture will be automatically sprayed using spraying mechanism.

An extension of this work will focus on developing hybrid algorithms such as genetic algorithms and Neural networks in order to increase the recognition rate of the final classification process underscoring the advantages of hybrid algorithms; also, we will dedicate our future works on automatically estimating the early detection of the disease and sucrose level detection especially in sugarcane plants as India is the 2nd largest sugarcane growing country in the world.

References

[1] Samajpati, B. J., & Degadwala, S. D. (2016, April). Hybrid approach for apple fruit diseases detection and classification using random forest classifier. In 2016 International conference on communication and signal processing (ICCSP) (pp. 1015-1019).IEEE.

[2] Sandika, B., Avil, S., Sanat, S., & Srinivasu, P. (2016, November). Random forest based classification of diseases in grapes from images captured in uncontrolled environments. In 2016 IEEE 13th international conference on signal processing (ICSP) (pp. 1775-1780).IEEE.

[3] Zawbaa, H. M., Hazman, M., Abbass, M., & Hassanien, A. E. (2014, December). Automatic fruit classification using random forest algorithm. In 2014 14th International Conference on Hybrid Intelligent Systems (pp. 164-168).IEEE.

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/CKM9V

[4] Iqbal, M. A., & Talukder, K. H. (2020, August). Detection of potato disease using image segmentation and machine learning. In 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET) (pp. 43-47). IEEE.

[5] Chauhan, M. D. (2021). Detection of maize disease using random forest classification algorithm. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(9), 715-720.

[6] Deepak, A. H., Gupta, A., Choudhary, M., &Meghana, S. (2019, December). Disease Detection in Tomato plants and Remote Monitoring of agricultural parameters. In 2019 11th International Conference on Advanced Computing (ICoAC) (pp. 28-33).IEEE.

[7] Hidayatuloh, A., Nursalman, M., & Nugraha, E. (2018, October). Identification of tomato plant diseases by Leaf image using squeezenet model. In 2018 International Conference on Information Technology Systems and Innovation (ICITSI) (pp. 199-204).IEEE.

[8] Sarangdhar, A. A., &Pawar, V. R. (2017, April). Machine learning regression technique for cotton leaf disease detection and controlling using IoT. In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) (Vol. 2, pp. 449-454).IEEE.

[9] Hatuwal, B. K., Shakya, A., & Joshi, B. (2020).Plant leaf disease recognition using random forest, KNN, SVM and CNN. Polibits, 62, 13-19.

[10] Dhingra, G., Kumar, V., & Joshi, H. D. (2018). Study of digital image processing techniques for leaf disease detection and classification. Multimedia Tools and Applications, 77(15), 19951-20000.

[11] Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2022). A novel deep learning method for detection and classification of plant diseases. Complex & Intelligent Systems, 8(1), 507-524.

[12] Hungilo, G. G., Emmanuel, G., & Emanuel, A. W. (2019, April). Image processing techniques for detecting and classification of plant disease: a review. In Proceedings of the 2019 international conference on intelligent medicine and image processing (pp. 48-52).

[13] Khirade, S. D., & Patil, A. B. (2015, February). Plant disease detection using image processing. In 2015 International conference on computing communication control and automation (pp. 768-771).IEEE.

[14] Ramesh, S., & 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.

[15] Das, D., Singh, M., Mohanty, S. S., & Chakravarty, S. (2020, July). Leaf Disease Detection using Support Vector Machine. In 2020 International Conference on Communication and Signal Processing (ICCSP) (pp. 1036-1040).IEEE.