

Investigation on plant leaf diseases using machine learning

Srimaya Mohapatra

Abstract

Disease diagnosis display an important character in better understanding the Indian financial system in terms of agricultural production. Many machine learning processes such as SVM (support vector machine), informal forest, KNN, Naïve Bayes, pruning trees, etc., are used to detect and distinguish plant illness. However, the development of machine learning with deep learning (DL) is intended to have a significant impact on improving accuracy. Deep learning may be a branch of computer science. The benefits of automatic learning and having a domain, are closely related to theoretical and industry company It is widely used in video and image enhancing, voice upgrading, and dialogue processing. The use of in-depth learning in the diagnosis of disease can intercept the deterioration generated by genetic alternatives, make the absorption of the disease a major goal, and improve the effectiveness of analysis and the pace of mechanical change. We hope that this work will be very helpful to researchers conducting research on plant and insect diseases.

Keywords

Image processing, Deep learning, KNN,SVM.

2. Introduction

We humans use more than 2000 kinds of plants to make a variety of delicious foods in our diet. But the plants are affecting some of the diseases which make unhealthy to plants [1]. Disease of plant harms agriculture production, the rapid spread of germs, fungi, insects, and microorganisms on plants and plant areas can greatly affect their survival [2]. If the plant disease does not dissolve in time, malnutrition will increase. In general, plant leaves are the main origin for identifying plant problem and most diseases can start appearing on the leaves [3]. In the agricultural sector, plant diseases, pests and weeds lead to about 36% loss

of crops. [4]. Crop yields are sensitive to a variety of biological stresses that cause significant damage in the form of reduced yields. Food dependability, nourishing and the agricultural wealth are entangled in a harsh cycle. It brings about profound changes in developing and underdeveloped nations and leads to well-being and wealth problems [5]. It is often difficult to distinguish the above symptoms, so early detection and further treatment to avoid economic loss are problematic. Early detection is essential for effective avoidance and manage of plant diseases, it is necessary to find leaves that need to be removed or treated in a timely manner to mitigate the subsequent spread of pests, we will identify the plant disease and infected leaves by using machine learning Algorithm and by deep learning [6].

Features were trained using SVM, KNN, RNN, decision trees, and disease detection using wavelet-based plant diseases. Diseases are easily detected in leaves using hyperspectral measurements [7]. In this model, we used texture and color features and SVM for classification [8]. We can use weighted KNNs to characterize Gabor wavelets and gray-level correspondence matrices. Using texture and color characteristics, Kmeans can be used to detect diseases in which areas of infection are segmented. Kmeans can be used to regulate the existence of fungal contamination on leaves [9]. This model identifies disease using scale invariance that characterizes SIFT. For efficient classification, the following is a SIFT function with SB distribution in plant disease classification for disease detection in tomato leaves. There is image fractionalization, which is defined as a method of dividing an image into individual sets using the Kmeans clustering method [10]. This model uses a categorization method based on a backpropagation neural network that describes defects, color, and surface areas of segments [11]. We present a pattern identification method using snake categorization. In this model, artificial neural networks were used for classification to help determine the stiffness of diseased leaves [12]. Using deep CNN, we count the accuracy. The nature of handcrafted features purely needs accuracy. Using this model, the shortcomings can be overcome for plant diseases discovery.

Although these diseases only affect the leaves, some can also affect the stems and/or fruits [13]. Leaf disease is the most usual disease of many plants. They are usually regulated by fungicides, viruses and resistant varieties. Leaf disease is defined by a variety of symptoms, but be aware that the distinction is not always clear and can be confusing as there are many different names than those listed for leaf disease [14].

2. COMMON PLANT DISEASES

2.1. Leaves Dots

Leaf spots (other names: anthrax, scabies, leaf color, buds) often represent specific areas of varying size, shape, and color [15]. There are always certain limits. Sometimes areas that can be caused by bacteria or fungi are surrounded by a yellow halo. When referring to fungi, fungi generally grow in that area, especially in humid climates [16]. These fungal growths are also small pimple-like structures, usually black or soft growths of seeds. It is often mandatory to use a manual lens or a microscope to understand these compositions [17]. If there are too many spots or rashes, the affected areas come together to form abnormal

areas called "spots". The common name for leaf spot disease may be as common as bacterial leaf spot. Description, such as the color of a frog's leaf; Or named after a fungus. For example, the colour of the leaves of Septaria [18].

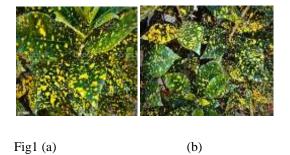


Figure 1 shows leaves dots diseases in two different leaves

2.2. Leaf Blights

Leaf diseases are usually areas with more disease than leaves and are not uncommon. Perhaps the arrival of "dangerous" leaves is the outcome of a combination of various small spots [19]. In general, the general term involves the word "blight" namely Southern Corn Blight or Corn Blight



Fig2 (a) (b)

- a) shows starting stage of Leaf Blights disease spread in a leave
- b) figure shows when leaf blight diseases spreads more in a leave

2.3. Rust

Rust often causes spots like leaf spots, still these marks are called "pustules." [20]". Pseudomonas pustules are shining t yellow, orange-red, maroon, or black. Pustules normally rise on the petiole, and rubbing with a whitematerial often results in a pustular coating on the tissue. In severe cases, the leaves wither and dry quickly. It is found in other types of rusted stems. Common in melted grains and herbs.



Fig3 (a) (b)

- a) shows rust disease spread in multiple leaves
- b) detailed version of rust disease in a single leaf

2.4. Powdery Mildew

Powdery mildew also causes white or light gray surfaces on the surface and powdery mildew or powdery mildew on the leaves, but can also take place on the stems and flowers. Overdone leaves generally turn yellow, dry out and die quickly. This problem is common with vegetables such as pumpkin and small grains.



Fig-4 (a)

- (b)
- Powdery Mildew in single leaf a)
- Powdery Mildew in multiple leaf b)

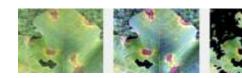
2.5. Downy Mildew

Signs of Downy mildew green or yellow part on the upper side of the leaf; Light gray to purple mist growing under the leaves. Green tobacco fungus can be a fungal disease. Impaired plant growth ("crazy top") can be caused by mold as it does in corn or sorghum fever.



Fig -5 (a)

- (b)
- Zoomed image of Downy Mildew in a leave a)
- Multiple disease of Downy Mildew in leave b)



(a)









(h)

Fig6 – Some examples of diseased leaf split images where the leaves are contrasted and then segmentation done to determine these diseases (a)Cercospora Leaf Spot (b)Alternaria Alternata c) Bacterial Blight d) Bacterial Blight e) Anthracnose f) Healthy Leaves g) Anthracnose h) Bacterial Blight

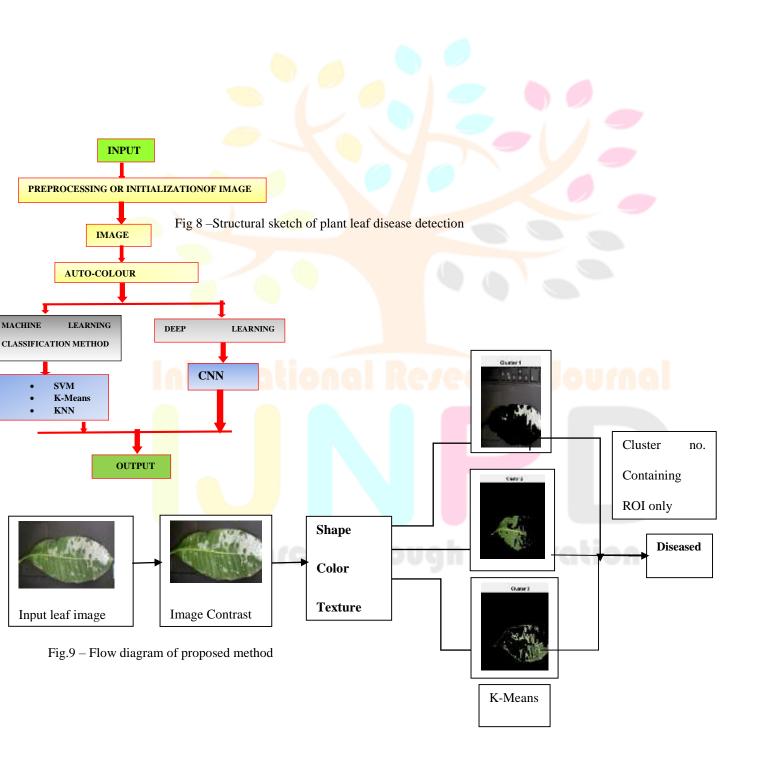
3. PROPOSED METHODOLOGY

The main goal of this effort is to enable digital automation to help solve usual problems that arise mainly in the classification of plant diseases. It consists of collecting an image database by collecting multiple leaf samples and using kmeans clustering, a GLCM method for feature extraction, a step-by-step method to preprocess and segment these images using CNN for disease classification, and finally machine learning. Using the SVM algorithm.



Fig 7 -Plant Disease Detection

In this figure, plant disease detection is carried out on MATLAB software, where a leaf is loaded, then image reprocessing is done where enhance contrast is applied to the image then segmentation is applied for differentiating the images into clusters, then the image shown according to the cluster chosen. Then the disease with accuracy is determine as result.



A Matlab code is written to classify leaves into one of the following types: 'Alternaria Alternata', 'Anthracnose', 'Bacterial Blight', 'Cercospora Leaf Spot' and 'Healthy Leaves'. Classification is done with Multiclass SVM

3.1. Matlab Processing

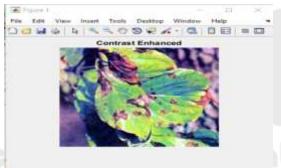
3.1.1. Image Acquisition

This is an early digital imaging process in which images of plant leaves are captured by a digital camera and stored on digital media for use in MATLAB. Here an image is being processed from Hardware, then it will be passed for further process. The images are in RGB (Red, Green, And Blue) formin the image a diseased image of the plant leaf is captured.



Fig 10 - Preprocessing or initialization of image

Here image preprocessing is done to improve the image properties and data through enhancing features of the image for further processing. Initializing an image is done by various methods and techniques such as changing the size and shape of the image, filtering noise, transforming the image, enhancing the image or creating contrast. As shown in the figure, certain MATLAB



codes are used to resize and enhance the image.

Fig11-Image Contrast

3.1.2. Image Segmentation

It is used to detect and locate a part that is different from other parts. For leaf disease detection color and texture plays an important vital role. Basically, for differentiating the images into clusters, K – Means Cluster one part of the cluster contains images with large areas of the affected part, the mean cluster is used. KMEANS CLUSTER classifies objects into K classes. Categorization is done by reducing the sum of squared distances in the middle of data objects and their clusters. From the evaluation of K Means, each component in the image is labeled and a split image containing the disease is also generated. The image segmentation process technique is used for partitioning images into three clusters for better segmentation results. Here the figure shows leaf image Segmentation.

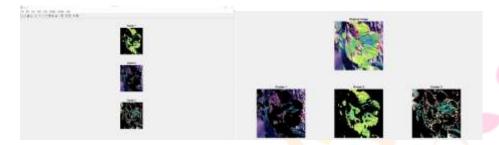


Fig 12 Image Segmentation

Fig-n contains 3 clusters formed by the K-means clustering method

3.1.3 Colour and Texture Detection Method

- I. $H = \begin{cases} \emptyset, if B < G \\ 360 \emptyset, if B > G \end{cases},$
- II. $S = 1 \frac{3}{R+G+B} [min\{R,G,B\}]$
- III. $I = \frac{1}{3} (R + G + B)$

The RGB (Red, Green, Blue) image is converted into the HIS (Hue, Saturation, Intensity) translation.

The RGB (Red, Green, Blue) image is converted into the HIS (Hue, Saturation, Intensity) translation.

- I. Hue value is multiplied by 60 to convert it to degrees on the color circle.
- II. equation which is used to get S component of every RGB pixel in RGB color format
- III. equation which is used to obtain I(Intensity) component of every RGB pixel in RGB color format

3.1.4. Feature Extraction

For identification of the object, color, texture, morphology, edges feature extraction plays an important role. It can describe a large set of datasets and resources very accurately

3.1.5. Training & Classification

K-means and SVM clustering classification are used to separate infected and healthy parts of leaf areas, performed by extracting texture and color features.

3.1.6. EXPERIMENTAL RESULTS

Here, the results determine the proportion of areas affected by plant diseases and classify plant-borne diseases like Alternaria Alternate, Anthracnose, Fire blight, Cercospora leaf spot and healthy leaf.

3.2. RESULT AND DISCUSSION

3.2.1Image Acquisition



3.2.2Preprocessing or initialization of image



Fig.13 – Results obtained in image processing

3.2.3 Image Segmentation



3.2.4 Disease



3.3. Algorithms Used

3.3.1 CNN Algorithm

ConvNet/CNN (Convolutional Neural Network) is a deep learning design or algorithm that can take a captured image, assign values (human readable weights and offsets) to different image elements(s) and distinguish them from each other. Convolution formula:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau$$

Convolution: A convolution layer includes a square grid of neurons. current layers are required. A rectangular grid of neurons. Every neuron absorbs enter from the square part of the previous layer. The weights of this rectangular topology are the same for all neurons within the convolution layer, this is, a convolution layer is a convolution of the photo of the previous layer, and the weights determine the convolution clear out.

3.3.2. SVM Algorithm

Support Vector Machine (SVM) is the most famous popular supervised learning algorithms used to time table and update tasks. However, it is mainly used to solve deployment problems in machine learning.

The purpose of the SVM algorithm is to create better decision lines that can divide a multidimensional space into classes., making it easier to place future data points into appropriate categories. This limit of the best solution is called the hyperplane.

SVM picks additional points/vectors to create the hyperplane. This extreme case is called a support vector, which is why the algorithm is called a vector support machine. Consider the diagram below with two categories separated by a boundary or hyperplane

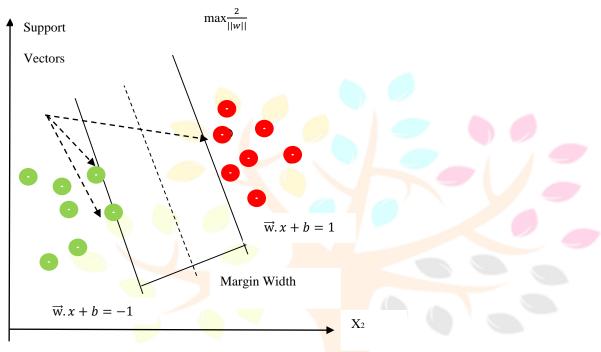


Fig-14 Optical hyperplane –maximum width pf the margin(w)

3.3.3. K-Means Algorithm

The K-means algorithm includes inches and iterates until an appropriate center of gravity is found, you can inform how many collections there are, it's also called flat clustering algorithm. The wide variety of clusters at the records is decided through the letter "k" through the K technique, on this approach, the data points are grouped so that the full variety of distances among the records points and the center is as small as feasible, it's miles essential to note that decreased variety among organizations ended in very comparable facts points inside the same collection.



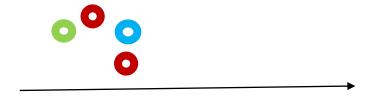
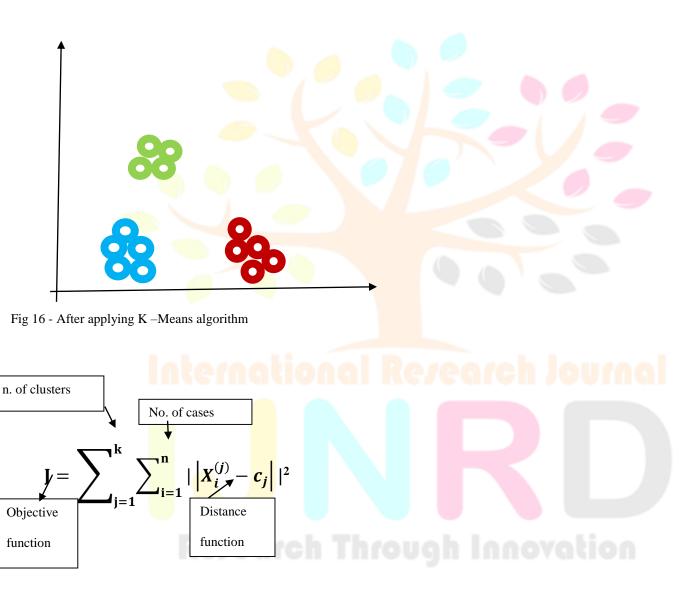


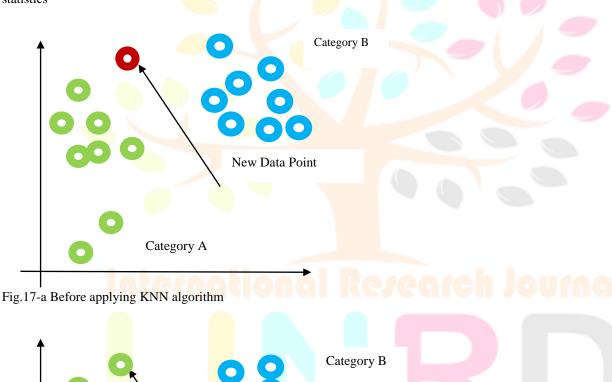
Fig- 15 before applying K –Means algorithm



- i. Clusters the data into k groups where k is predefined.
- ii. Choose random k points as cluster centers.
- iii. Assign objects to the nearest cluster center according to the Euclidean distance function.
- iv. Calculate the centroid or mean of all objects in each cluster.
- v. Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds.

3.3.4 KNN Algorithm

KNearest Neighbour is one of the only systems studying algorithms based on supervised mastering techniques. The KNN algorithm captures similarities among new cases/records and available cases and places new cases in classes which can be very much like the available ones. The KNN set of rules shops all available records and classifies new information points primarily based on similarity, which means that where the brand new facts come from, we can effortlessly classify it into true sections of the set the use of the KNN set of rules. The KNN algorithm may be used for undo and edit, however is in general used for edit operations, on the grounds that KNN is a parameter less algorithm, it makes no assumptions approximately the underlying statistics



New Data Point assigned to category

Category A

Fig-b After applying KNN Algorithm

Euclidean -
$$\sqrt{\sum_{i=1}^{n}(\mathbf{x_i}-\mathbf{y_i})^2}$$
,

Given two vectors x and y, we take the square root of the sum of the squares of the differences in their elements

Manhattan -
$$\sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{y}_i|$$
,

Rectilinear distance refers to the sum of the absolute differences of the Cartesian differences

Minikowski -
$$\left(\sum_{i=1}^{n}(|\mathbf{x}_i|-\mathbf{y}_i|)^q\right)$$
14q,

Minkowski Distance is used for nearness variable distance to find the similarity of distances between vectors given two or

more vectors -

	Contrast					
	Segmented ROI					
Features	Mean	128.906	19.741	41.6235	76.1852	51.2905
	S.D.	118.697	49.3344	63.6705	86.6728	70.0072
	Entropy	3.6237	1.95386	4.10045	4.35589	5.0792
	RMS	12.5744	5.62567	9.52182	10.8753	10.4595
	Variance	10666.3	2162.61	3237.63	6910.07	2936.17
	Smoothness	1	1	1	1	1
	Kurtosis	1.09883	9.92354	3.77506	1.41485	3.1091
	Skewness	0.0159212	2.69714	1.39154	0.419311	1.17011
	IDM	255	255	255	255	255
	Contrast	0.41152	1.00107	1.1544	1.98655	0.38796
	Correlation	0.981417	0.748981	0.77305	0.854609	0.950036
	Energy	0.350451	0.644755	0.408599	0.283333	0.227298

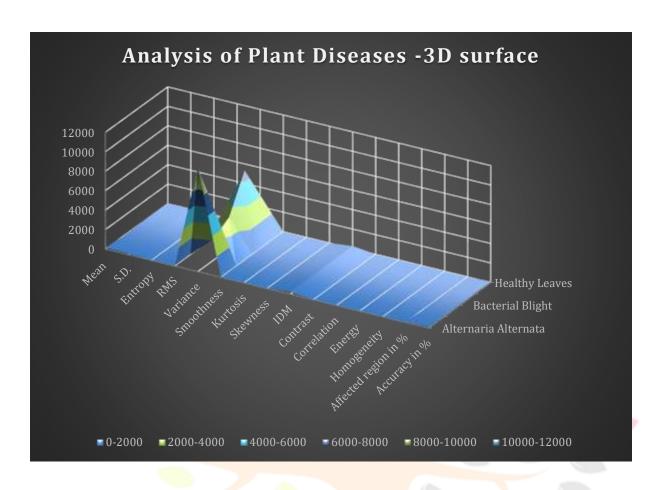
	Homogeneity	0.94482	0.919102	0.893283	0.89868	0.929342
Percent	Affected region	59.9961	11.9856	15.0062	15.0015	None
age	in %					
	Accuracy in %	96.7742	98.3871	96.7742	98.3871	96.7742

Table 1. Brief Comparisons of Features of Plant Diseases

The table showing different types of plant leaves where images are loaded, then the plant leaves is enhanced with contrast, and then the leave is segmented showing 3 types of clusters. Different types of features of plant diseases is updated where it shows some features like Mean, S.D., Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM, Contrast, Correlation, Energy, Homogeneity and showing percentage of Affected region in %, Accuracy in %

Our gadget detects the location of the affected leaf and the leaf sickness, this is finished thru photograph Processing; there are applications that predict leaf diseases. Our device uses a aggregate of the k-Medoid and the Random forest set of rules to supply greater precision in leaf detection. The image is first processed and a composite method is used to decide the affected place of the leaf, that is then processed to down load 13 characters like Mean, SD, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM, Contrast, Correlation, Energy and Homogeneity for which we can degree accuracy and locate ailment.





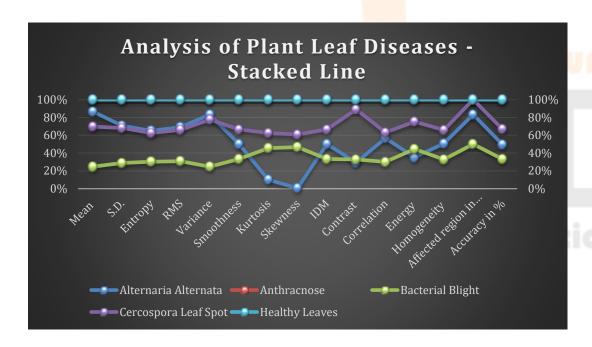


Fig – 19 Analysis of Plant Diseases

The graph characters like Mean, SD, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM, Contrast, Correlation, Energy and Homogeneity and their value.

5. APPLICATION



Fig – 20 Plant Leaf disease detection application. 'DetectDisease GUI'

A window application on plant leaf disease detection which helps to shows diseases affected in leaves of the plants through image processing technology.

5. CONCLUSION

However, there are many problems associated with accurately diagnosing plant illnesses in hard situations. accurate identification and class of plant diseases is critical for higher planting, which can be achieved by picture processing. The significance of gathering large amounts of information through dynamic development, information enlargement, transfer studying and show of CNN impulsion maps to improve class accuracy, and the importance of small leaf seek and the significance of hyperspectral pictures for fast detection of plant sicknesses. although a plant-primarily based diagnostic version based at the in-depth studies proposed on this paper may additionally be triumphant the complexity of nature and enhance diagnostic accuracy, there are still some troubles that need to be targeted. As we recognize, the list of inequality level evaluation facts in the first stage is an obstacle to obtaining a high quality segregation end result thorough evaluation of your site performs a crucial function in improving machine overall performance via segmentation. Step-with the aid of-step transfer research helped lessen CNN communique time. now not most effective did the proposed version perform well, however the different models were also skilled to study with

fewer gears, showing large improvements. studying to slow down will no longer only accelerate your assembly however additionally make mastering hard. Proposed a plant disorder diagnostic plan considering the troubles dealing with plant sicknesses, in particular by looking, the ability to apply a separator on a portable machine to create a piece solution. Such devices do no longer have excessive hardware, therefore, it's miles crucial to layout effective solutions. The proposed department machine is examined in the Plant Database and the publicly available records set of PlantVillage. This function can be extended to different flora and illnesses and extra superior learning techniques may be implemented to programs. CNN-based pc paintings principle has certainly accomplished a milestone with excessive accuracy however, it's far important for focused paintings in both industries and purchasers so as to gain from it with the very best satisfactory studies, when used successfully, this application can play a vital role in decreasing losses on small farms and ultimately increasing crop yields, eventually, this procedure is likewise relevant to many industrial programs wherein the improvement of device learning algorithms is urgent

6. Future Enhancement

We have made an image processing technique for Plant disease detection by using Matlab where a PC or laptop is required. Image processing techniques can be developed by app making where small and marginal farmers can simply scan the leave by using their smartphone. Rich farmers can use advanced techniques like using drones for image processing techniques. It is important to note that the dataset used for this function is very limited due to images from Plant Village and crowdAI, or images generated from pre-rendered DCGANs. In an active environment, results are affected by many factors such as differences in light, differences in sensors, scores, images, and more. Therefore, our work must be tested in many real-world situations. In future work, we are able to create a internet site with a actual photograph surroundings and test what different results the deep studying version may have, additionally, many conventional strategies such as scrolling, translating, rotating, measuring, cropping, and so forth, are not famous with taxpayers. Used to amplify facts, therefore, to compare DCGAN information supplementation with conventional records talking can be certainly one of our future duties

Research Through Innovation

7. Reference

[1] Plant Disease Detection Using Image Processing Sachin D. Khirade, M.E Student (Electronics & Telecommunication Engg.) Pune, India.B. Pati Professor (Electronics & Telecommunication Engg.) Pimpri Chinchwad College of Engg.

[2] V. Singh and A. K. Misra, "Detection of unhealthy region of plant leaves using processing and genetic algorithm," in Proc. Int. Conf. Adv. Comput. Eng. Appl., Mar. 2015, pp. 1028–1032.

[3] Plant Disease Detection and Classificationby Deep Learning_A Review LILI LI1, SHUJUAN ZHANG 2, AND BIN WANG 2

1College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong 030800, China 2College of Agricultural Engineering, Shanxi Agricultural University, Jinzhong 030800, China Corresponding author: Shujuan Zhang (zsujuan1@163.com)

[4] Plant Disease Detection in Imbalanced Datasets Using Efficient Convolutional Neural Networks With Stepwise Transfer Learning MOBEEN AHMAD, MUHAMMAD ABDULLAH, HYEONJOON MOON, AND DONGIL HAN, (Member, IEEE)

Department of Computer Engineering, Sejong University, Seoul 05006, South Korea

[5] IoT Enabled, Leaf Wetness Sensor on the Flexible Substrates for In-Situ Plant Disease Management Kamlesh S. Patel, Riya Saini, Ahlad Kumar, Sandeep G. Surya, Member, IEEE, Vinay S. Palaparthy, and Khaled N. Salama, Senior Member, IEEE

[6] An Investigation into Machine Learning RegressionTechniques for the Leaf Rust Disease DetectionUsing Hyperspectral Measurement

Davoud Ashourloo, Hossein Aghighi, Ali Akbar Matkan, Mohammad Reza Mobasheri, and Amir Moeini Rad

- [7] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," Biosyst. Eng., vol. 172, pp. 84–91, Aug. 2018.
- [8] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers Plant Sci., vol. 7, p. 1419, Sep. 2016.

- [9] M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant disease detection and classification by deep learning," Plants, vol. 8, no. 11, pp. 468–489, Oct. 2019.
- [10] V. Singh, N. Sharma, and S. Singh, "A review of imaging techniques for plant disease detection," Artif. Intell. Agricult., vol. 4, pp. 229–242, Oct. 2020
- [11]] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, "Recent advances in image processing techniques for automated leaf pest and disease recognition—A review," Inf. Process. Agricult., vol. 180, pp. 26–50, Apr. 2020.
- [12]] A. C. Cruz, A. Luvisi, L. De Bellis, and Y. Ampatzidis, "Vision-based plant disease detection system using transfer and deep learning," in Proc. Asabe Annu. Int. Meeting, Spokane, WA, USA, Jul. 2017, pp. 16–19.
- [13]] J. G. A. Barbedo, "Plant disease identification from individual lesions and spots using deep learning," Biosyst. Eng., vol. 180, pp. 96–107, Apr. 2019.
- [14] A. Johannes, A. Picon, A. Alvarez-Gila, J. Echazarra, S. Rodriguez-Vaamonde, A. D. Navajas, and A. Ortiz-Barredo, "Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case," Comput. Electron. Agricult., vol. 138, pp. 200–209, Jun. 2017.
- [15] S. D. Khirade and A. B. Patil, "Plant disease detection using image processing," in Proc. Int. Conf. Comput. Commun. Control Autom., Feb. 2015, pp. 768–771.
- [16]] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," Frontiers Plant Sci., vol. 8, p. 1852, Oct. 2017
- [17]] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers Plant Sci., vol. 7, p. 1419, Sep. 2016.
- [18]] M. Arsenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "Solving current limitations of deep learning based approaches for plant disease detection," Symmetry, vol. 11, no. 7, p. 939, Jul. 2019.
- [19] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," Inf. Process. Agricult., vol. 4, pp. 41–49, Mar. 2017.
- [20] X. Zhang, L. Han, Y. Dong, Y. Shi, W. Huang, L. Han, P. González-Moreno, H. Ma, H. Ye, and T. Sobeih, Remote Sens., vol. 11, no. 13, Jun. 2019, Art. no. 1554.