Evolution in Electrical and Electronic Engineering Vol. 2 No. 2 (2021) 667-676
© Universiti Tun Hussein Onn Malaysia Publisher's Office

EEEE

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/eeee
e-ISSN : 2756-8458

Tomato Leaf Disease Detection using Convolution Neural Network (CNN)

Lawra Norria Samba Ricky[1], Ain Nazari[1]*,

[1]Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/eeee.2021.02.02.080
Received 04 July 2021; Accepted 03 August 2021; Available online 30 October 2021

Abstract: Leaf diseases are the major problem in the agricultural sector, which affects crop production as well as economic profit. In certain cases, these diseases may not destroy the plant, but they do reduce yield and quality significantly. Therefore, the main purpose of this research is to identify the tomato leaf disease using image processing technique, to develop an automatic disease detection system for tomato leaf and to classify tomato leaf disease using Convolution Neural Network (CNN). The method to be used in this research consist of three-part which are pre-processing, segmentation, feature extraction and classification using Convolutional Neural Network (CNN). The experimental results have shown that the proposed method is able to successfully detect the leaf disease of affected tomato leaf. The average accuracy for Healthy Leaf is 93.6769 %, sensitivity is 93.3591 % and Fl-Score is 93.0463%. While, for Late Blight Leaf the accuracy is 92.0857%, sensitivity is 92.4698% and Fl-Score is 92.1244%. Meantime, for Septoria Spot Leaf the accuracy 91.3381%, sensitivity is 91.1303% and Fl-Score is 91.0449%. As the result show the healthy leaf has high accuracy, sensitivity and f1-score because healthy leaf has least affected area among three datasets. In future, it is fair to propose expanding the present model to identify other diseases in different plants. It may be possible to investigate the severity of the disease's spread using an automated approach, which could provide additional assistance to persons dealing with the diseases.

Keywords: Tomato Leaf Disease, Convolutional Neural Network, Image Processing Technique

1. Introduction

Tomato (Lycopersicon esculentum Mill) broadly known as a vegetable rich in medicinal and nutritional values. Tomatoes are among the few local crops that can be sold both domestically and internationally. The most important field for tomato cultivation is Cameron Highland in Pahang State, Peninsular Malaysia. Leaf tomato issues may be caused by nutrient deficiencies, infection, fungi or insects. Leaf spot and fruit rot are caused by a variety of tomato diseases and disorders. The most

*Corresponding author: ain@uthm.edu.my

common conditions include a bacterial, virus, fungal infections, among others. The tomato leaf disease characteristics that affected the tomato leaf are stage (early and last), type of symptom, front and back sides leaf, type of fungus, colour and shape and size. The type of symptom diseases portrays different symptoms in the leaf and in the stem or fruits, affecting unlike parts of the plant. The indication of the same infection may look different on each side of the leaf. The variety of fungus of some diseases produces a kind of fungus that is a proper characteristic of them. The color and shape of each leaf may be different for each condition. It depends on the stage of the disease. For the size relay on the grade of the picture, some images are taken, including the leave or stem in the foreground, but others may include various part of the background with small leaves [1]. Fungi can cause the most common types of infectious diseases that the plant may encounter. Pests make the ground unproductive and make it impossible for the soil to be used as a parent plant for other plants. It is possible to classify the diseased region, type, and severity using image processing [2]

Rupali Patil (2016) had convert the input RGB image into HSV (Hue Saturation Value) format using rgb2hsv command. HSV color space is preferred manipulation of Hue and saturation (to shift color or adjust amount of color) Fuzzy C-Mean clustering techniques are mainly based fuzzy behavior and they provide a technique which is producing a clustering [3]. S. Prakanshu, M. Kritika, A. Vibhav, K. Vivek and Pawan (2021) improve image data suppress unwanted distortion, Image reshaping, resizing and conversion to an array form. Picture pre-processing is the strategy of improving information. Using CNN, P. Srivastava, K. Mishra, V. Awasthi, V. Kumar Sahu, and P. Kumar Pal (2021) hoped to enable the model to differentiate between unhealthy and healthy leaves, as well as the environment [4]. Deepa (2018) devised that the pre-processing the image is to obtain a high-quality image by eliminating any undesired details. Sharpening of an image enhances the contrast between dark and bright regions that helps to bring out features. Filters are generally used for elimination of noise, from images, deblurring and edge detection [5]. Researcher use method image reshaping, resizing in pre-processing method so the quality is not compromised to a great extent. A. Mushtaq, A. Karol, and G. Drushti (2019) describe use Convolution Neural Network (CNN) which comprises of different layers which are used for prediction [6]. In 2017, N. Jayamoorthy and Palanivel used clipping of the leaf image is performed to get the interested image region and then image smoothing is done using the smoothing filter. The unsupervised clustering algorithm that is applied to wide range of problems connected with feature analysis, clustering and classifier design [7]. In research of S. Rajbir and S. Bhupinder (2019) used preprocessing of the image which reduces the noisy data from it by used noise removal for the images. The fuzzy c - mean method research by S. Rajbir and S. Bhupinder in 2019 as a classifier which classifies the texture of the images. Image segmentation is performed which divides the image into the small segments. The leaf spot disease from the complex background in a more efficient manner using the enhanced fuzzy c-means technique. The affected area is identified using Fuzzy C-means grouping segmentation, and then features are removed. CNNs have proven to be a good alternative to other deep learning models in classification and detection [8].

Recently, A. K. Ghorai et al published a paper in which that the image containing unwanted noise is filtered prior further processing. In bilateral filter the noise removal can be done by replacement of intensity of each pixel with weighted average of intensity values of neighbour pixels. CNN's design includes two unique features: local connections and shared weights. CNN uses local connection between neurons in nearby layers to take advantage of local correlation. The data constraint as well as a significant quantity of noise in the plants data were successfully managed. The Convolutional neural network (CNN) is a type of machine learning technique or classifier applied in the prediction of diseases in important crops, according to the authors [9]. It has conclusively been shown that, M. Chowdhury et al. (2016) uses CNN-based architectures to classify tomato leaf images. To classify tomato diseases, researchers used convolutional neural networks (CNNs) classification network architecture such ResNet18, MobileNet, DenseNet201, and InceptionV3 on 18,162 plain tomato leaf photos [10]. In their recent studies, M. Agarwal, S. K. Gupta, and K. K. Biswas suggested CNN model avoids the usage of a complicated network structure of pre-trained models with a high number of hidden layers and parameters, resulting in lower storage capacity and faster response while keeping the same level of accuracy[11]. The authors discussed the CNNs are recognized for their robustness in the case of little input variation and their execution requires little preprocessing. They're also capable of extracting relevant features while discriminating at the same time [12].

The production of tomato crops is increasing due to the high demand from supermarkets. When a plant becomes infected with a pathogen, the disease may spread across the crop, resulting in several casualties. But in producing tomatoes the quality, farmers have to deal with the infection that plagues the leaves of the crops. Leaf diseases can be into different shapes, forms, colors on the tomato leaf. This disease may affect the quality of the tomatoes and can cause the tomatoes to be rotten quickly.

The purpose of this research is to identify the tomato leaf disease using image processing technique and develop an automatic disease detection system for tomato leaf. The diseases of the tomato leaf classify using Convolution Neural Network (CNN) in MATLAB Software. The system that uses advanced technologies related to image processing to classify tomato diseases early or early and provide helpful information for their management. From a quantitative standpoint, the images of leaf disease will analyze using digital image processing techniques and derive disease spot features associated with color, texture, and other characteristics through the purpose method. In this research, tomato leaves will be used as a sample for disease detection. In this project, images of dataset are acquired from the Kaggle website [5] that contains healthy leaf, late blight leaf and septoria leaf. The total of the sample images will be used are 60 images.

2. Materials and Methods

The purpose of this method is to identify disease on a tomato leaf based on visual symptoms. The experimental works are conducted on MATLAB R2019b software and run-on Intel® Core ™ i5-10300 CPU @ 2.50GHz processor and 8 GB RAM used as the primary medium for the tomato leaf disease detection system. The image processing of a tomato leaf involves five stages, as indicated in the block diagram. The steps for detecting and classifying plant leaves diseases using image processing are as shown in Figure 1.

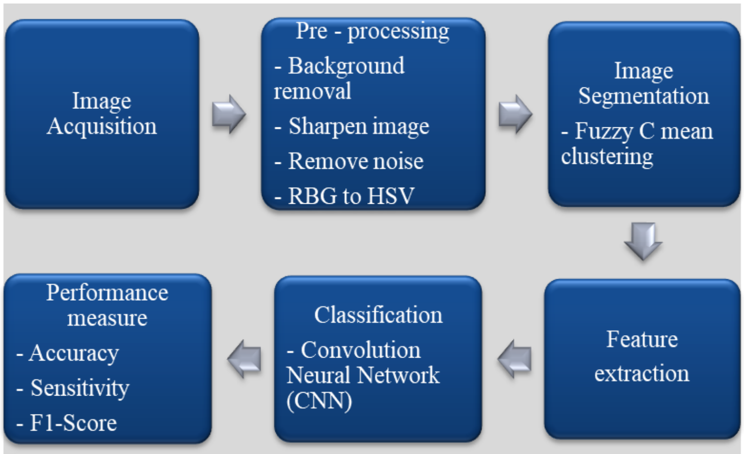

Figure 1. Block diagram for the proposed system

2.1 Image acquisition

Image acquisition is the first step of any image processing process, and it involves scanning and storing images for further use in MATLAB. This system's image was taken from an online archive. For tomato trees, there are 3 types of leaves used like Septoria leaf, and Late Blight. The image of dataset tomato leaf had obtained from online database Kaggle website [13]. For each leaf disease, 20 photos will be used in this research. The size of each image is 256 X 256 pixels.

2.2 Pre - processing

The image was preprocessed to improve its quality and remove any unwanted noise before being cropped and smoothed. To increase the image quality and make the feature extraction phase more trustworthy, image-processing techniques such as background removal, sharpening, RGB to HSV conversion, and noise removal are used as illustrates in Figure 2.

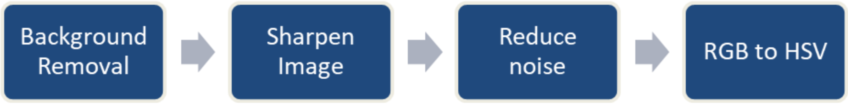

Figure 2. Block diagram for the pre-processing process

2.2.1 Background removal

The light intensity on each object causes the background object to be nearly identical to the target object. The background has been eliminated by finding the current pixel's RGB value equal to the default value of leaf disease and removing all other associated colors [14]. The processing of colors and the removal of the unrelated background will occur separately and only the leaf will remain as shown in Figure 3.

Figure 3. Before (Left) and after (right) leaf background removal [13]

2.2.2 Sharpen image

Sharpening an image is an effective way to emphasize texture and attract the viewer's attention to it. The "unsharp mask" is used to sharpen a picture. While an unsharp mask added detail, it can significantly improve the appearance of detail by increasing a slightly complex mechanism [15]. Sharpening can help to focus on texture and detail of the leaf as in Figure 4.

Figure 4. Before (Left) and after (right) leaf sharpen image [13]

2.2.3 Remove noise

The means variance of brightness or color information in images is known as image noise. Noise is frequently described as the unpredictability of a signal caused by random fluctuations [15]. The Figure 5 shown below, the left if the image of leaf before noise removal and right is after removal noise. The image of leaf clearer and smoother.

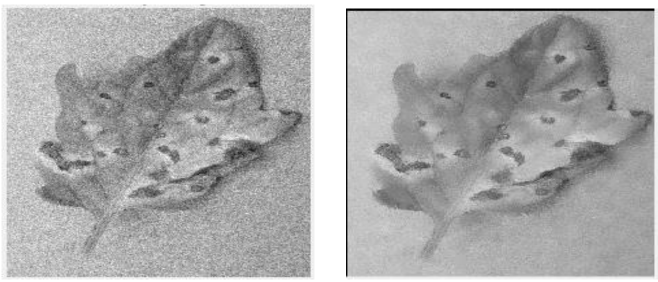

Figure 5. Before (Left) and after (right) leaf reduce noise [13]

2.2.4 Color space conversion RGB to HSV

RGB images are as Hue, Saturation Value (HSV). The HSV model provides the best starting point for creating optimal color combinations. Hue is a color definition, as perceived by an observer. The saturation called relative purity refers to the level of white light applied to the hue and value, and white light as the full saturation [16]. The leaf colors processed through the colour space transformation, the color of leaf and disease area on leaf clearer as illustrate in Figure 6.

Figure 6. Before (Left) and after (right) RGB disease leaf convert to HSV [13]

2.2.5 PSNR and MSE

The MSE and PSNR is calculated to measure the effectiveness of the method used. Mean Square Error (MSE) indicate the cumulative squared error between the pre-processed image and the original image, while PSNR represents the measure of the peak error. The image's MAXI (maximum possible pixel value) is 255 [17].

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \qquad Eq.1$$

$$PSNR = 10\ log_{10}(\frac{MAX_1^2}{MSE}) \qquad Eq.2$$

$$= 20\ log_{10}(\frac{MAX_1}{\sqrt{MSE}})$$

2.3 Fuzzy C – Mean Segmentation

Fuzzy C-Mean clustering is a data clustering method that works well. The algorithm is an unsupervised clustering method that has had considerable success in medical diagnosis, image analysis, and target recognition. In segmentation technique based on fuzzy clustering is performed on the enhanced picture. With the fuzzy clustering and cluster base attribute values in FCM are continuously modified, resulting in high-precision image segmentation. The area of the leaf disease has been clustering or grouping in image as in shown in Figure 7.

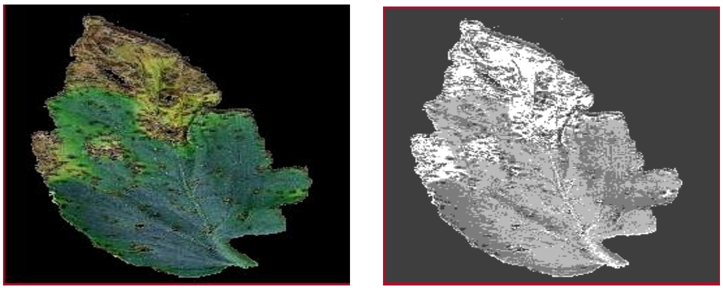

Figure 7. Before (Left) and after (right) Fuzzy C-Mean clustering [13]

2.4 Feature extraction

ResNet50 was created by Kaiming He et al. and is also known as Residual Neural Network (ResNet). The dataset needs a more complex neural network to learn, so the weights (accuracy) begin to saturate and then degrade [18]. Batch normalization is a network layer that allows each layer to develop separately of the others. It's used to normalize the output of the previous layers [19]. ResNet-50" is one such model and can be loaded using the resnet50 function from Neural Network Toolbox™. Figure 8 illustrate the block diagram CNN of ResNet50 to improve accuracy while do the complex task for classification.

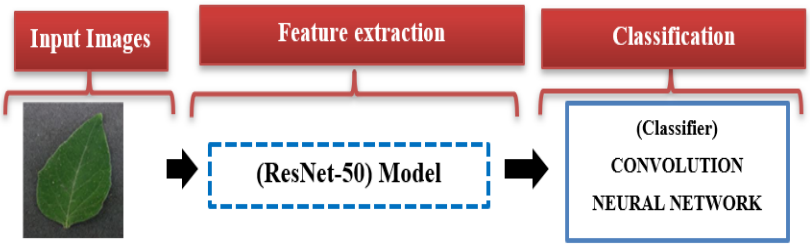

Figure 8. Block Diagram Convolution Neural Network (CNN) of ResNet50

2.5 Convolution Neural Network (CNN) Classification

CNN is used to create attributes dynamically and merge them with the classifier. The number of layers that convert input volume to output layer is the easiest of all the classification methods in this scheme, one of the advantages of a CNN classifier. There are only a few materials, and each layer uses a differentiable function to translate the input to output. Convolution is a much slower mechanism than the max pool, both forward and backward. Each training phase will take a long time if the network is large [20].

The CNN parameters are usually composed of multiple specific elements that represent the differences in different architectures. Figure 9 shows the general architecture, including its core aspects such as the input layer, convolutional layer, pooling layer, and a flattening process. Data entered into a set of dense layers, which symbolize the neural network result.

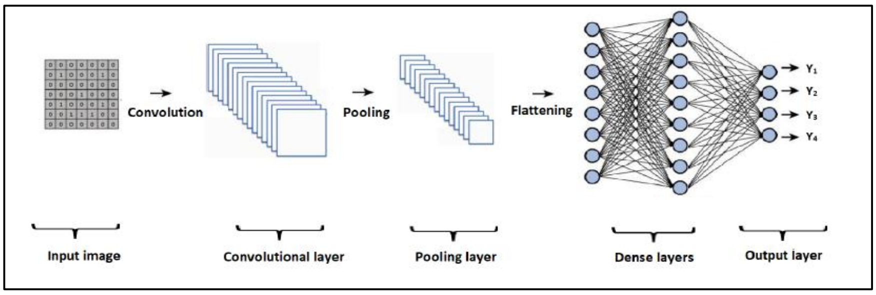

Figure 9. Architecture of a convolutional neural network (CNN) [28]

As the efficiency of the classifier has to assess performance parameters accuracy, Sensitivity, and f1-score to evaluate the algorithm's performance, this score takes both false positives and false negatives into account. Terms that used in the calculation is true positive (TP), true negative (TN), false positive (FP) and false negative (FN) [21]. The formula calculation for the performance measure is as below:

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \qquad Eq.3$$

$$Precision = \frac{TP}{(TN+FP)} \qquad Eq.4$$

$$\text{Sensitivity} = \frac{TP}{(TP+FP)} \qquad Eq.5$$

$$F1-Score = 2 * \frac{Precision * \text{Sensitivity}}{Precision + \text{Sensitivity}} \qquad Eq.6$$

3. Results and Discussion

The pre-processing stage is the initial phase in the system of tomato leaf disease. Background removal, image sharpening, RGB to HSV conversion, and noise reduction are all pre-processing parts. The images were then segmented using the Fuzzy C-Mean approach, which segmented the diseased portion of the leaf. Finally, to diagnose leaf disease, the feature extraction and classification of an image using the Convolution Neural Network (CNN) approach. The purpose system used healthy, late blight and septoria leaf disease with 60 total of images.

Fuzzy C - mean calculates the cluster centers and updates the fuzzy partition matrix with the determined center positions in each clustering iteration. The objective function value is then computed. After each iteration, cluster the data and display the objective function result.

A response, or activation, is generated by each layer of a CNN in response to an input image. After that, train a multiclass SVM classifier using the CNN image features CNN is used to create attributes dynamically and merge them with the classifier. The number of layers that convert input volume to output layer is the easiest of all the classification methods in this scheme, one of the advantages of a CNN classifier. When working with high-dimensional CNN feature vectors, this speeds up the training process. The trained classifier is then used to identify the image's disease.

The table 1 below convey the average of PSNR and MSE value of three datasets. The closer the MSE number is to zero, the better the algorithm's outcome. The greater the PSNR number, the greater the image quality. The PSNR value obtained after pre-processing is clearly greater than the original image, according to the results. Overall show that the healthy leaf had acquire the highest PSNR average compare to the late blight and septoria spot leaf as illustrate in Table 1.

Table 1. Average MSE and PSNR of database

Datasets	Average	
	MSE	**PSNR**
Healthy Leaf	0.0015	76.8333
Late Blight Leaf	0.0020	74.5070
Septoria Spot Leaf	0.0085	71.1935

For the performance measure stage, the proposed algorithm is applied on three databases to the system. The testing dataset consists of 60 leaf images of three disease classes that are healthy, blight late and Septoria leaf images. Figure 10 illustrate a bar chart of the value of accuracy, sensitivity, and F1-score of Healthy, Late Blight and Septoria Spot leaves in this proposed algorithm. The bar chart show value of accuracy for Healthy Leaf is 93.6769 %, sensitivity is 93.3591 % and Fl-Score is 93.0463%. While, for Late Blight Leaf the accuracy is 92.0857%, sensitivity is 92.4698% and Fl-Score is 92.1244%. Meantime, for Septoria Spot Leaf the accuracy 91.3381%, sensitivity is 91.1303% and Fl-Score is 91.0449%. Based on the result above, the value of Accuracy, Sensitivity and F1-socre of the healthy leaf more high compare to Late Blight leaf and Septoria leaf. This is because the healthy leaf has less noise and does not have any affected disease. Based on the bar chart above, the healthy leaf has the highest of accuracy, sensitivity and f1-score among the three datasets because healthy leaf has the least affected disease area compare to the Septoria leaf and late blight leaf.

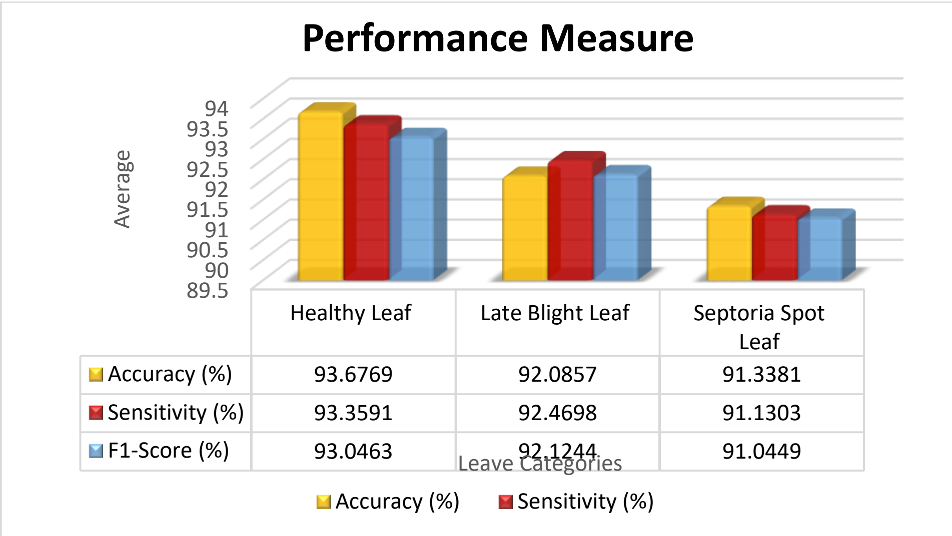

Figure 10. Performance Measure of healthy, late blight and Septoria leaves

4. Conclusion

The purpose system able to identify the tomato leaf disease using image processing technique and developed an automated disease detection system for tomato leaf. Aside that, the system for tomato leaf disease detection able to classify tomato leaf disease using Convolution Neural Network (CNN). In addition, the performance of the system had evaluated by terms of accuracy, sensitivity and F1-score to see the effectiveness of the applied method. It is concluded that the proposed method be able to recognize three different types of tomato leaf diseases and the result for performance of the system also good as well. Without proper identification of the disease and the disease-causing agent, disease control measures can be a waste of time and money and can lead plant losses in future. So, there are numerous problems that can be improved for future use.

In future, implementing better methods and techniques to the algorithm can result in a better algorithm. To improve the algorithm's accuracy, methods for creating an algorithm for database extraction and classification can be added. Moreover, carried out for improvement of replicability and accuracy of the quantification of the diseases.

Acknowledgement

The authors would like to thank the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia for its support.

References

[1] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 293–301, 2020, doi: 10.1016/j.procs.2020.03.225.

[2] J. K. Patil, "Color Feature Extraction of Tomato Leaf Diseases," *Int. J. Eng. Trends Technol.*, vol. 2, no. 2, pp. 72–74, 2011.

[3] R. Patil, S. Udgave, S. More, D. Nemishte, and M. Kasture, "Grape Leaf Disease Detection Using K-means Clustering Algorithm," *Int. Res. J. Eng. Technol.*, vol. 3, no. 4, pp. 2330–2333, 2016, [Online]. Available: https://www.irjet.net/archives/V3/i4/IRJET-V3I4461.pdf.

[4] P. Srivastava, K. Mishra, V. Awasthi, V. Kumar Sahu, and P. Kumar Pal, "Plant Disease Detection Using Convolutional Neural Network," *Int. J. Adv. Res.*, vol. 9, no. 01, pp. 691–698, 2021, doi: 10.21474/ijar01/12346.

[5] Deepa, "A Pre Processing Approach for Accurate Identification of Plant Diseases in leaves," *3rd Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2018*, no. December, pp. 249–252, 2018, doi: 10.1109/ICEECCOT43722.2018.9001616.

[6] Mushtaq Adnan, Karol Ali, and G. Drushti, "Plant Disease Detection using CNN & Remedy," pp. 622–626, 2019, doi: 10.15662/IJAREEIE.2019.0803014.

[7] N. Jayamoorthy, S., & Palanivel, "Identification of Leaf Disease Using Fuzzy C-MEAN and Kernal Fuzzy C-MEAN and Suggesting the Pesticides," vol. 4, no. 5, pp. 3852–3855, 2017.

[8] R. Singh and B. Singh, "Improved Leaf Disease Detection and Classification Convolution base Particle Swarm Optimization," vol. 7, no. 2, pp. 3124–3132, 2019.

[9] A. K. Ghorai *et al.*, "Image Processing Based Detection of Diseases and Nutrient Deficiencies in Plants," no. March, 2021.

[10] T. R. E. T Muhammad. Chowdhury *et al.*, "Tomato Leaf Diseases Detection Using Deep Learning Technique," *Intech*, no. tourism, p. 13, 2016, [Online]. Available: https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics.

[11] M. Agarwal, S. K. Gupta, and K. K. Biswas, "Development of Efficient CNN model for Tomato crop disease identification," *Sustain. Comput. Informatics Syst.*, vol. 28, p. 100407, 2020, doi: 10.1016/j.suscom.2020.100407.

[12] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," *Lect. Notes Informatics (LNI), Proc. - Ser. Gesellschaft fur Inform.*, vol. 266, pp. 79–88, 2017.

[13] 2019 and N. Lamrahi, "Kaggle online database Tomato Leaf Disease Detection," 2019. https://www.kaggle.com/noulam/tomato (accessed Nov. 25, 2020).

[14] R. Hussin, M. R. Juhari, N. W. Kang, R. C. Ismail, and A. Kamarudin, "Digital image processing techniques for object detection from complex background image," *Procedia Eng.*, vol. 41, no. Iris, pp. 340–344, 2012, doi: 10.1016/j.proeng.2012.07.182.

[15] P.Sathya, M.Bhuvaneswari, G.Kesavaraj, and K.S.Saravanan, "A REVIEW ON IMAGE ENHANCEMENT USING SHARPENING TECHNIQUES," *Int. J. Res. Comput. Appl. Robot.*, vol. 2, no. 1, pp. 74–78, 2014.

[16] S. Perumal and T. Velmurugan, "Preprocessing by Contrast Enhancement Techniques for Medical Images," *Int. J. Pure Appl. Math.*, vol. 118, no. 18, pp. 3681–3688, 2018.

[17] R. Mehra, "Estimation of the Image Quality under Different Distortions," *Int. J. Eng. Comput. Sci.*, vol. 5, no. 17291, pp. 17291–17296, 2016, doi: 10.18535/ijecs/v5i7.20.

[18] A. Kumar and M. Vani, "Image Based Tomato Leaf Disease Detection," *2019 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019*, pp. 6–11, 2019, doi: 10.1109/ICCCNT45670.2019.8944692.

[19] J. M. Celaya-padilla, J. I. Galván-tejada, H. Gamboa-rosales, and C. A. Olvera-olvera, "applied sciences Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases," doi: 10.3390/app10041245.

[20] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," *2018 3rd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. RTEICT 2018 - Proc.*, no. November, pp. 2319–2323, 2018, doi: 10.1109/RTEICT42901.2018.9012507.

[21] E. Hossain, M. F. Hossain, and M. A. Rahaman, "A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier," *2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019*, pp. 7–9, 2019, doi: 10.1109/ECACE.2019.8679247.