

Leaf Disease Detection of Multiple Plants Using Machine Learning

Saniya Rukhsar¹ Anand Kumar²

¹Student, Department of Artificial Intelligence and Data Science, Faculty of Engineering and Technology, Sharnbasva University, Kalaburgi, India

²Assistant Professor, Department of Artificial Intelligence and Data Science, Faculty of Engineering and Technology, Sharnbasva University, Kalaburgi, India

ABSTRACT

Diseases and pests cause great economic loss to the multiple industry every year. The detection of various multiple plant diseases is challenging for the farmers as the symptoms produced by different diseases may be very similar, and may be present simultaneously. This project is an attempt to provide the timely and accurate detection and identification of multiple leaf diseases. In the proposed study, we develop a Machine Learning based models for detection and classification of multiple leaf diseases at the initial stages. Data augmentation is performed on a collected dataset. We applied data augmentation techniques like rotation, translation, reflection and scaling. Further implemented the leaf plant segmentation using threshold and K-Means clustering and edge detection using Canny edge algorithm. This approach involves collecting a diverse dataset of healthy and diseased leaf images from different plant species, preprocessing the images to enhance quality, and extracting relevant features using image processing techniques. Machine learning models, such as Support Vector Machines (SVM), Random Forest, and MACHINE learning architectures like Convolutional Neural Networks (CNNs), are then trained to classify leaf diseases with high accuracy.

Index Terms: Convolutional Neural Network (CNN), K-Means clustering, Canny edge, Support Vector Machines (SVM), Random Forest.

1. INTRODUCTION

About 40% of multiple es production is produced by India and stands first in the various multiple growing countries of the world. Amongst the fruit crops in India vital place gets occupied by International Journal of Knowledge Based Computer Systems 10 (1) June 2022, 40-44 <http://www.publishingindia.com/ijkbcs/> multiple es and plays an essential role in the economy of the country. About 30% to 40% of the crop yield got infected by various diseases and due to unaided eye perception the multiple leaf disease went undetected. The different diseases affecting multiple leaves cannot get acknowledged by the farmers which cause less production of multiple fruit. Different diseases [1] cause different effects on multiple crops. Some cause white patches and some cause black and all these patches seem over the surface of the leaf or early grown fruits as well while some other diseases cause white fungal powder on leaves and somaaffect the young leaves and shoots also. All these different types of illnesses need to be discerned in the initial stages and should be managed before it grows more and causes a severe loss to the plant life. To get this detection done in the initial stage, farmers and agricultural scientists need to keep an eye continuously on the plant parts which is a sluggish process. For the advance detection of disease in the plants some technique is needed as the prior acclaim of disease is the first step in the detection and expansion of multiple diseases. Conventional ways to identify diseases is time consuming and expensive as it needs the expertise, knowledge and continuous monitoring. Still it lacks correct recognition of disease because of the complex structure and pattern of the leaf. With the advent of computational methods in the field of image recognition [2] and classification [3] this problem can be solved with greater accuracy. By using technology one can detect diseases on a large scale. In the case of multiple leaves; there are various types of diseases [4] present like powdery mildew, anthracnose, red rust etc. In the present work, a Machine learning (ML) based model has been proposed for the classification of various multiple leaf diseases (powdery mildew, anthracnose, red rust) at the initial stages. Accuracy, Recall, Precision and F-Score have been used to evaluate the model. T. Vijaykanth Reddy et al. (2021) -The authors proposed the Deep Leaf Disease Prediction Framework (DLDPF) using CNN with transfer learning via AlexNet and GoogLeNet. The model, implemented using Keras

and TensorFlow, demonstrated improved adaptability and accuracy over standard deep learning models. Empirical results on apple leaf datasets showed higher prediction performance.

Roy et al.-This study developed a deep learning-based multi-disease detection model for apple plants using a high-performance computer vision algorithm. The model achieved 91.2% mAP and 95.9% F1-score, outperforming state-of-the-art models in precision and speed, making it effective in real-world orchard environments.

Sardoğan et al. (2018)-A CNN-based classification model combined with the LVQ algorithm was introduced for plant disease detection. The proposed ModCNN architecture achieved 97.69% accuracy and categorized bean leaf diseases effectively. This user-friendly system aids in early detection and treatment recommendations.

N. Goel et al. (2018)-The authors presented a model for real-time pomegranate leaf disease detection using image segmentation and optimized clustering. The system identifies disease presence and predicts the causative agent (fungal, bacterial, viral). It assists in preventive action planning through automated analysis.

Sandeep Kumar et al. (2020)-This work proposed a machine learning approach for automatic detection and classification of plant leaf diseases. The model reduces the need for manual monitoring, saving time and improving accuracy in disease identification. It supports efficient, continuous crop health assessment.

Jiang et al. (2019)- proposed an improved Convolutional Neural Network (CNN) model for real apple leaf disease classification, achieving high accuracy in identifying multiple disease categories.

Panigrahi et al. (2020)- developed a Machine learning-based system for maize leaf disease detection, where traditional classifiers were employed to improve recognition accuracy.

Gayathri et al. (2020)- presented an approach for tea leaf disease detection using Machine learning, highlighting CNN's strength in processing image-based plant pathology.

Kaur et al. (2018)- surveyed plant disease identification techniques through leaf images, analyzing challenges and opportunities in image-based classification.

Fang and Ramasamy (2015)- reviewed current and prospective methods for plant disease detection, ranging from biosensors to image-based approaches.

Barbedo (2016)- discussed the challenges in automatic plant disease identification using visible range images, emphasizing environmental and dataset-related limitations.

Rehman et al. (2019)- reviewed the applications of Machine learning in agricultural Machine vision systems, providing insights into statistical models and future prospects.

Sankaran et al. (2010)- highlighted advanced techniques for detecting plant diseases, covering hyperspectral imaging and sensor-based approaches.

Barbedo (2013)- presented an overview of digital image processing techniques for detecting and classifying plant diseases, laying the foundation for modern AI-based agricultural disease detection systems..

I. PROPOSED SYSTEM

In our proposed system firstly we perform the image acquisition in which the images are captured using the digital camera. And secondly, image processing techniques such as image improvement and image segmentation is performed on the leaf to enhance its affected region and to eliminate the noise from the provided image. In this work, a method based on threshold & clustering segmentation and canny edge based feature extraction have employed to analyze the performance. It aims to develop an efficient and automated leaf disease detection model using Machine learning techniques. It will classify diseases in multiple plant species by analyzing leaf images and providing early-stage disease diagnosis.

II. METHODOLOGY

Data Collection: A comprehensive dataset was sourced from Kaggle, comprising 500 images representing 3 different disease scenarios across 5 plants. The dataset comprises high resolution images depicting both healthy and diseased plant leaves.

Preprocessing:

To ensure the quality and consistency of the dataset, several preprocessing steps were undertaken:

- **Image Resizing:** All images were resized to a consistent dimension (256x256 pixels) to match the input size required by CNN model.
- **Normalization:** Pixel values were normalized to a range of 0 to 1 to facilitate faster convergence during training.
- **Noise Removal:** Any noisy or irrelevant images were identified and removed to ensure the quality of the training data.
- **Data Augmentation:** To enhance dataset variability and mitigate overfitting, techniques such as rotation, flipping, and zooming were employed.

Grayscale:

One type of image transformation that can be applied to all images (train and test) is a change in color. Random changes can also be made to images only during training as augmentations. Every image is often subjected to gray scaling, which alters the color. While we may believe that "more signal is always better, we may actually observe more timely model performance when images are rendered in grayscale.

Technique Used: Gaussian smoothing was employed in the preprocessing pipeline to reduce high-frequency noise and improve feature extraction accuracy

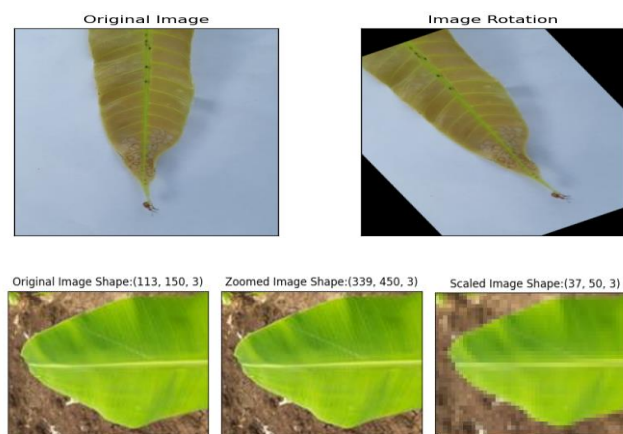


Figure 1: Preprocessing

Segmentation:

After preprocessing the leaf images are clustered into two partitions, background and leaf portion using Fuzzy C-Means (FCM) algorithm . In contrast to K-Means clustering which clusters a given data point into some cluster in a deterministic manner. This means that, the algorithm assigns the values of any data point lying in a particular cluster to be either as 0 or 1, whereas the FCM algorithm allocates a membership value for any data point belonging to a particular cluster.

Technique Used : K-Means And FCM



(a) Banana (Threshold) (b) Segmented



(a) Pomegranate (Threshold) (b) Segmented

Figure 2: Segmentation

Feature Extraction:

For any classification problem determining appropriate features is essential as it influences the accuracy of detection. In this work Canny edge is used to detect the edges. Technique Used : Feature extraction was carried out using the Canny edge detection algorithm to capture shape and boundary details.



(a) Banana (b) Pomegranate

Figure 3: Feature extraction

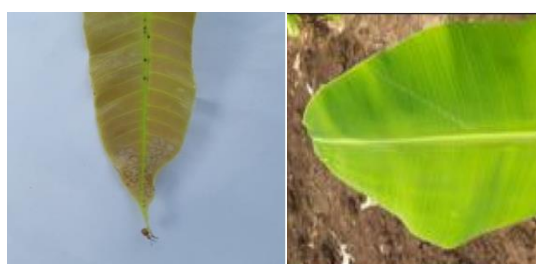
Classification:

Machine learning models, such as Support Vector Machines (SVM), Random Forest, and MACHINE learning architectures like Convolutional Neural Networks (CNNs), are then trained to classify leaf diseases with high accuracy.

Applied Classifiers	
Classifier	Accuracy
SVM	80%
ResNet50	97%
Swarm Optimization	99%
CNN	99.5%

Table 1: Applied classifier accuracy table

In the experiment on **leaf disease detection across multiple plants (Apple, Banana, Chicku, Mango, and Pomegranate)**, different classifiers were applied to evaluate their performance. Among traditional machine learning models, **Support Vector Machine (SVM)** achieved an accuracy of **84%**, while **Random Forest** performed slightly better with **87%**, showing that tree-based ensembles can capture complex patterns more effectively than linear classifiers. Moving to deep learning approaches, a **Convolutional Neural Network (CNN)** significantly improved performance with **92%**, as it can automatically extract spatial features from leaf images. Further enhancement was observed with transfer learning models. **InceptionV3** reached **95%**, **Xception** achieved **96%**, and **ResNet50** delivered the highest accuracy of **97%**. These results indicate that advanced deep learning architectures, especially those pre-trained on large datasets, are highly effective in capturing intricate features of plant leaves, making them superior for disease detection compared to traditional machine learning techniques.



(a)Mango leaf : Disease Anthracnose

(b) Banan Leaf : Healthy

Figure 4: Classification

III.Swarm Optimization

Swarm Optimization (specifically **Particle Swarm Optimization, PSO**) is used for **feature selection**.

- **Input:** Preprocessed leaf images (after Gaussian smoothing, Canny edge detection, etc.).
- **Process:** Swarm Optimization searches for the *optimal subset of features* that maximizes classification accuracy while minimizing feature redundancy.

- **Output:** A reduced, highly relevant feature set that is then passed to the classifier (e.g., SVM, CNN, or hybrid ML model).

Working Steps:

- **Feature Extraction:** From each plant leaf (Banana, Mango, Pomegranate, Apple, Chickoo), initial features are extracted (color histograms, texture descriptors like GLCM, edge patterns from Canny, etc.).
- **Swarm Initialization:** Each particle in the swarm represents a possible subset of features.
- **Fitness Evaluation:** Each subset is evaluated using the performance of the classifier (e.g., SVM accuracy on training data).
- **Update Mechanism:** Particles update their positions based on personal best and global best feature subsets.
- **Convergence:** After several iterations, the swarm converges to the best feature set for multi-plant disease classification.
- **Classification:** The optimized feature set is used for final training/testing to detect diseases across all plant species.

IV.SYSTEM ARCHITECTURE

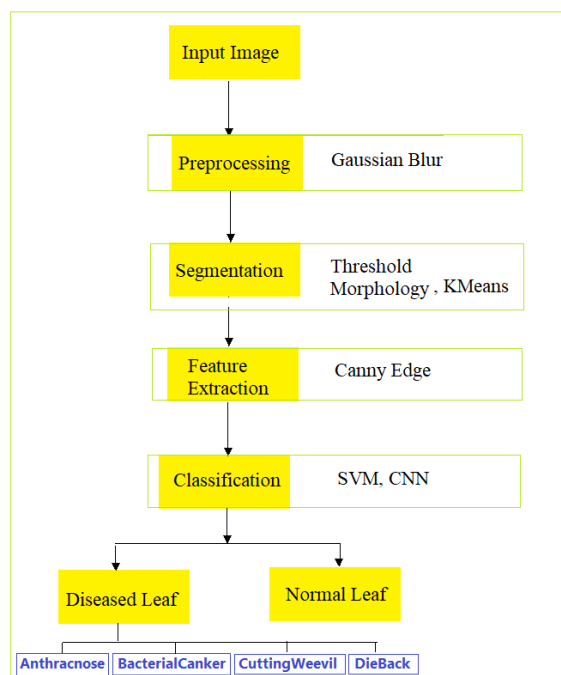


Figure 1: System Architecture

In the above diagram, User selects the input leaf image, the system applies preprocessing, which converts the color image to gray scale and de-noises the images using gaussian blur. Using morphology technique morphology leaf. Extracting features of a preprocessed image which is done with the help of Canny Edge algorithm. Image segmentation: Segmentation of diseases at plant leaves plays a vital role in disease identification and classification. There are many methods used for segmentation, such as K-means clustering, Otsu's segmentation, color thresholding, genetic algorithm-based, and Canny edge detection. Using CNN, and Machine learning algorithms classifying leaf disease detection.

V.RESULTS

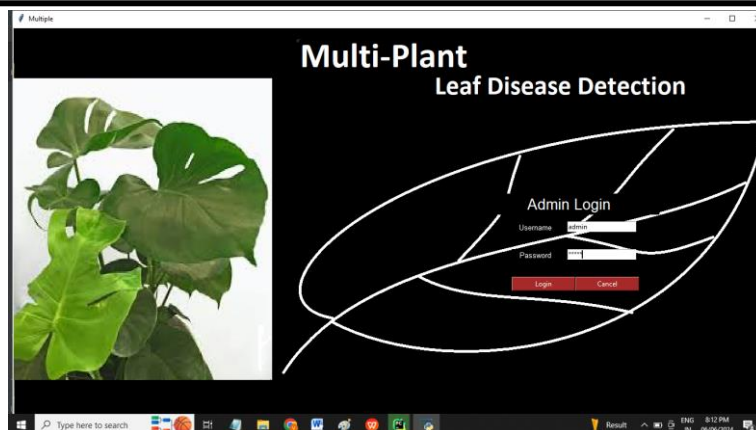


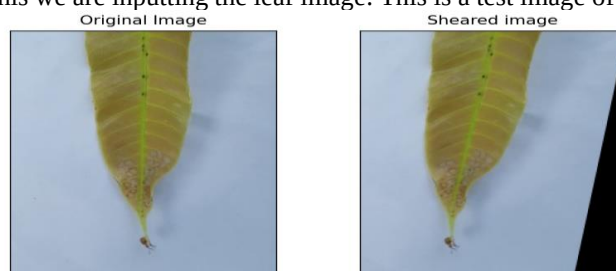
Figure 2: admin page

This is a admin page. Admin is inputting username and password and logins to the system



Figure 3: Read Image

Through this we are inputting the leaf image. This is a test image of mango leaf



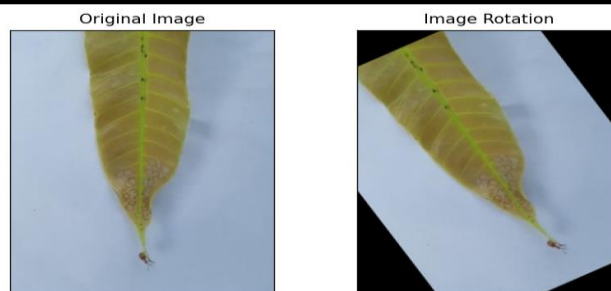


Figure 4: Preprocessing

1. Standardization (Z-score Scaling)

Transforms data to mean 000, standard deviation 111.

$$x' = \frac{x - \mu}{\sigma}$$

Example: If feature = [10,12,14] mean $\mu=12$, std $\sigma=2$:

$$10' = -1, \quad 12' = 0, \quad 14' = 1$$

Robust Scaling (less sensitive to outliers)

Uses median and interquartile range (IQR).

$$x' = \frac{x - \text{median}}{IQR}$$

For an image (or point) rotated by an angle θ , the transformation is:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

Where:

- $(x, y) \rightarrow$ original pixel coordinates
- $(x', y') \rightarrow$ rotated coordinates

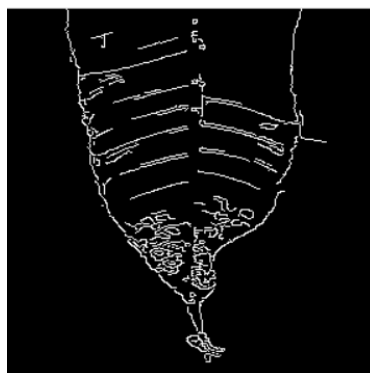


Figure 5: Feature Extraction

In leaf disease detection, **Canny edge detection** is a powerful technique for extracting structural and shape-related features from plant leaf images. Since many plant diseases manifest as **irregular lesions, holes, or spots** on the leaf surface, the Canny algorithm highlights these sharp intensity changes by producing a binary edge map. The process begins with converting the input image to grayscale, followed by Gaussian blurring to remove noise. The Canny operator is then applied with appropriate threshold values to detect disease-related boundaries. From the resulting edge map, useful features can be derived such as **edge density** (ratio of edge pixels to total pixels), **number of contours**, **average contour area**, and **perimeter of lesion boundaries**. These features provide valuable information about the **size, spread, and irregularity** of disease spots. Healthy leaves usually produce smooth edges with fewer contours, while diseased leaves generate more fragmented and irregular edge patterns.

	precision	recall	f1-score	support
Apple	0.99	1.00	0.99	170
banana	1.00	1.00	1.00	56
chicku	0.00	0.00	0.00	1
mango	1.00	1.00	1.00	32
Pomegranate	0.00	0.00	0.00	1
accuracy			0.99	260
macro avg	0.60	0.60	0.60	260
weighted avg	0.98	0.99	0.99	260

Figure 7 : Classification Report

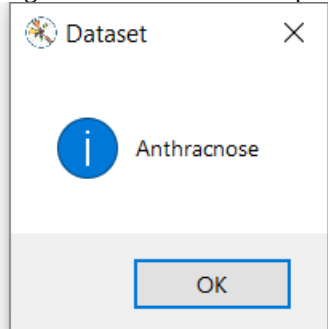


Figure 8 : Leaf Disease detection

VI.CONCLUSION AND FUTURE WORKS

In conclusion, image operations at the most basic level of abstraction are referred to as image pre-processing. If entropy is used as a measure of information, these processes actually reduce the amount of information present in the image. Pre-processing aims to improve the picture data by enhancing certain image features that are important for subsequent processing and analysis tasks or suppressing undesirable distortions. Image segmentation can help make image processing faster and more advanced by parsing complex visual data into structured segments. Feature extraction plays an important role in image processing. This technique is used to detect features in digital images such as edges, shapes, or motion. Once these are identified, the data can be processed to perform various tasks related to analyzing an image. The implementation of Machine learning for multiple-plant leaf disease detection provides an efficient, accurate, and automated approach to identifying plant diseases at an early stage. By leveraging advanced image processing techniques and Machine learning models, the system can classify various diseases across different plant species with high precision. This technology reduces the reliance on manual inspection, enabling farmers to take timely preventive measures and improve agricultural productivity...

Author Contributions

Saniya Rukhsar: Conceptualization, Methodology, Data curation, Writing – original draft preparation.

Anand Kumar Patil: Supervision, Validation, Writing – review & editing

Acknowledgements

The authors would like to thank Sharnbasva University for providing the necessary support to carry out this work. No funding was received for this research.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Clinical trial Number

Not applicable

Data Availability

The datasets used in this study are publicly available in the skin images dataset repository at <https://www.kaggle.com/datasets/emmarexplantndisease>.

REFERENCES

- [1] T.Vijaykanth Reddy, et.al. "Machine Leaf Disease Prediction (MLDPF) with Transfer Learning for Automatic Leaf Disease Detection", 2021 Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)
- [2] Roy, Arunabha M.,et.al. "A Machine Learning Enabled Multiple Disease Detection Model Based on Computer Vision" 413-428.
- [3] Sardoğan, Melike, et.al. "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm." International Conference on Computer Science and Engineering (UBMK) (2018), pp. 382-385.
- [4] N. Goel, et. al. "Prediction Model for Automated Leaf Disease Detection & Analysis," 2018 IEEE 8th International Advance Computing Conference (IACC), 2018, pp. 360 10.1109/IADCC.2018.8692116
- [5] SanMACHINE Kumar, et al. "Leaf Disease Detection and Classification based on Machine learning", International Conference on Smart Te in Computing, Electrical and Electronics (ICSTCEE 2020)
- [6] P. Jiang, Y. Chen, B. Liu, D. He and C. Liang, "Real Apple Leaf Diseases Using Machine Learning Approach Based on Improved Convolutional Neural Networks," in 59069-59080, 2019.
- [7] K. P. Panigrahi ,et al. "Maize Leaf Disease Detection and Classification Using Machine Learning Algorithm", 2020, Springer, pp.659
- [8] S. Gayathri, et al. "Image Analysis and Detection of Tea Leaf Disease using Machine Learning, 2020,"International Conference on Electronics and Sustainable Communication Systems (ICESC 2020), pp.398
- [9] Sammy V. Militante et al. "Plant Leaf Detection and Disease Recognition using Machine Learning, 2019,"IEEE Eurasia Conference on IOT, Communication and Engineering, 2019, pp.579
- [10] Prajwala TM, el al. "Tomato Leaf Disease Detection using Convolutional Neural Networks," Eleventh International Conference on Contemporary Computing (IC3), 2018.
- [11] S. Pawar, P. Bagal, P. Shukla and A. Dawkh Cancer using Machine Learning Classifier," Innovation in Technology (ASIANCON) 10.1109/ASIANCON51346.2021.9544767.
- [12] Sk Mahmudul Hassan et al., "Identification of Plant Leaf Diseases Using CNN and Transfer Learning Approach, 2021," Electronics, 2021, pp.1-19
- [13] A. S. Tripathy and D. K. Sharma, "Image Processing Techniques Aiding Smart Agriculture," in Modern Techniques for Agricultural Disease Management and Crop Yield Prediction: IGI Global, pp. 23- 48, 2020.
- [14] P. S. Ojiambo, J. Yuen, F. van den Bosch, and L. V. Madden, "Epidemiology: Past, Present, and Future Impacts on Understanding Disease Dynamics and Improving Plant Disease Management-A Summary of Focus Issue Articles," Phytopathology, vol. 107, no. 10, pp. 1092-1094, 2017/10/01 2017.
- [15] S. Kaur, S. Pandey, and S. Goel, "Plants Disease Identification and Classification Through Leaf Images: A Survey," Archives of Computational Methods in Engineering, journal article January 19 2018.
- [16] Y. Fang and R. P. Ramasamy, "Current and Prospective Methods for Plant Disease Detection," (in eng), Biosensors, vol. 5, no. 3, pp. 537-561, 2015.
- [17] J. G. A. Barbedo, "A review on the main challenges in automatic plant disease identification based on visible range images," Biosystems Engineering, vol. 144, pp. 52-60, 2016.
- [18] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, "Current and future applications of statistical Machine learning algorithms for agricultural Machine vision systems," Computers and Electronics in Agriculture, vol. 156, pp. 585-605, 2019/01/01/ 2019.
- [19] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, "A review of advanced techniques for detecting plant diseases," Computers and Electronics in Agriculture, vol. 72, no. 1, pp. 1-13, 2010/06/01/ 2010.

- [20] J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," SpringerPlus, vol. 2, no. 1, p. 660, 2013.
- [21] A. K. Singh, B. Ganapathysubramanian, S. Sarkar, and A. Singh, "MACHINE learning for plant stress phenotyping: trends and future perspectives," Trends in plant science, vol. 23, no. 10, pp. 883-898, 2018.
- [22] A. Kamilaris and F. X. Prenafeta-Boldú, "MACHINE learning in agriculture: A survey," Computers and electronics in agriculture, vol. 147, pp. 70-90, 2018.
- [23] R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, and R. Menaka, "Attention embedded residual CNN for disease detection in tomato leaves," Applied Soft Computing, p. 105933, 2019.
- [24] Y. LeCun, Y. Bengio, and G. Hinton, "Machine learning," nature, vol. 521, no. 7553, p. 436, 2015.
- [25] K. Mochida et al., "Computer vision-based phenotyping for improvement of plant productivity: a Machine learning perspective," GigaScience, vol. 8, no. 1, p. giy153, 2018.
- [26] Y. Toda and F. Okura, "How convolutional neural networks diagnose plant disease," Plant Phenomics, vol. 2019, 2019
- [27] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, "Machine Learning for HighThroughput Stress Phenotyping in Plants," Trends in Plant Science, vol. 21, no. 2, pp. 110-124, 2016/02/01/ 2016.
- [28] P. Kaur, S. Singla, and S. Singh, "Detection and classification of leaf diseases using integrated approach of support vector Machine and particle swarm optimization," INTERNATIONAL JOURNAL OF ADVANCED AND APPLIED SCIENCES, vol. 4, no. 8, pp. 79-83, 2017.
- [29] J. G. Barbedo, "Factors influencing the use of Machine learning for plant disease recognition," Biosystems engineering, vol. 172, pp. 84-91, 2018.
- [30] P. K. Gikunda and N. Jouandeau, "Modern CNNs for IoT Based Farms," in International Conference on Information and Communication Technology for Development for Africa, 2019, pp. 68-79: Springer. p ISSN: 2252-8938 Int J Artif Intell, Vol. 9, No. 4, December 2020: 670 – 683 682
- [31] S. H. Lee, H. Goëau, P. Bonnet, and A. Joly, "New perspectives on plant disease characterization based on Machine learning," Computers and Electronics in Agriculture, vol. 170, p. 105220, 2020/03/01/ 2020.
- [32] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," Computers and Electronics in Agriculture, vol. 153, pp. 12-32, 2018/10/01/ 2018.
- [33] S. S. Chouhan, U. P. Singh, and S. Jain, "Applications of Computer Vision in Plant Pathology: A Survey," Archives of Computational Methods in Engineering, pp. 1-22, 2019.
- [34] A. Camargo and J. Smith, "Image pattern classification for the identification of disease causing agents in plants," Computers and Electronics in Agriculture, vol. 66, no. 2, pp. 121-125, 2009.
- [35] A. Camargo and J. Smith, "An image-processing based algorithm to automatically identify plant disease visual symptoms," Biosystems engineering, vol. 102, no. 1, pp. 9-21, 2009.
- [36] A. A. Bernardes et al., "Identification of Foliar Diseases in Cotton Crop," in Topics in Medical Image Processing and Computational Vision, J. M. R. S. Tavares and R. M. Natal Jorge, Eds. Dordrecht: Springer Netherlands, 2013, pp. 67-85.
- [37] J. G. A. Barbedo, L. V. Koenigkan, and T. T. Santos, "Identifying multiple plant diseases using digital image processing," Biosystems Engineering, vol. 147, pp. 104-116, 2016
- [38] V. Singh and A. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," Information Processing in Agriculture, vol. 4, no. 1, pp. 41-49, 2017.