Plant Disease Detection: A Comprehensive Survey

Radhika Chapaneri
Computer Engineering
MPSTME
Mumbai, India
radhika.chapaneri@nmims.com

Maithili Desai [1], Anmolika Goyal [2]
Shreya Ghose [3], Sheona Das [4]
Computer Engineering
MPSTME
Mumbai, India
[1] maithili.desai20@nmims.edu.in
[2] anmolika.goyal31@nmims.edu.in
[3] shreya.ghose29@nmims.edu.in
[4] sheona.das18@nmims.edu.in

Abstract— **Plant diseases can affect vast produce of crops posing a major menace to food security. To avoid this risk, an approach is needed which performs early diagnosis, lacking in copious parts of the globe due to the dearth of essential infrastructure. This paper discusses several experiments and techniques performed for plant disease detection. Each set of methods has its own advantages, limitations and the parameters affecting the results. In this paper, we have shown a general flow observed in most of the Plant Disease Detection techniques and have given a detailed overview and comparison at stages such as selected dataset, pre-processing methods, feature selection and extraction, classification and performance metrics utilized. This paper aims to get an in-depth understanding of algorithm selection and key challenges faced in adopted approaches. Using this analysis, we have identified different techniques that can be used in different stages of a plant disease detection system to give the best results at each stage and identified key challenges that can be faced during detection.**

Keywords— Deep Learning, Neural Networks, Image processing, CNN, Plant Disease, SVM

I. INTRODUCTION

Plant diseases can affect the whole produce which makes early diagnosis and classification vital. This helps save resources, time, money and adopt timely preventive measures to avoid the affliction in the future. To detect affected areas on plant leaves, there are two major approaches that can be followed. The first approach is Image Processing by applying various techniques: filtering, clustering, histogram analysis among many others, to find the region of interest. The region of interest can effectively constitute the damaged portion of the leaf and the shape and size analysis can detect the disease that is afflicting the plant. Image processing algorithms are easy to understand and the use of Python libraries, such as OpenCV, make applications of these techniques simpler. However, in these algorithms, the average accuracy of disease detection is not very high [1]. The other approach is Deep Learning where a deep neural network architecture can be applied to train and test on the database. The most common neural network architecture to be used for images is the convolutional neural network (CNN). A CNN consists of convolutional layers, often layered with a max pooling layer, which are followed by one or more fully connected layers. The weights used for feature selection as well as classification is determined during the training period which helps to save memory requirements and computational complexity. The CNN takes in input images and based on the weights assigned to its nodes can classify the type of disease that is present on the leaf. The accuracy achieved can be high but there are a number of cons to this approach: the computational complexity, even with CNN, can be quite high. Knowing the region of interest may cut down considerable computation time and increase the accuracy of the system. Hence, we found that the approaches worked the best when used together. The image processing part of the algorithms limits the portion that needs to be analyzed, in many cases enhances the images for better classification and then passes it on the deep learning architecture.

Table I gives a brief overview of the techniques discussed and crop coverage in the plant disease detection domain and their availability in currently available review papers. In our paper, we have incorporated a comparative study to understand different techniques used for each step of a general plant disease detection system.

II. GENERAL FLOW

This section provides a comparative in-depth study of algorithms utilized in each step of the general flow. This general flow (Fig 1) is designed with seven steps most commonly used in plant disease detection.

A. Input Leaf Image

We review the available databases of disease afflicted plant crops. The datasets include both diseased and healthy leaves of crop. We divided the available datasets on the basis of the number of crops, namely, single crop and multiple crops. Table II is a tabular comparison between the datasets [2, 3, 4] based on the crops available, number of crop images, disease covered, method of plant leaf image capture and limitations of the corresponding dataset. We have also studied the common leaf diseases in the surveyed papers to understand their symptoms and the dataset including them. Table III categorizes the diseases broadly according to the main disease class such as fungal, bacterial, mold, viral and mite-affected and then discusses the sub-categories.

Table I. Overview of Techniques discussed and crops coverage in papers

Techniques discussed/Crops Covered		[1]	[5]	[6]	[7]	[8]	[9]	[10]	Our Survey
	K-Means	✔	✔	✔	✔	✔	✔	✔	✔

Techniques Discussed	Otsu's method	✔	✔	✔	✔	✔	✔	✔	✔
	K-Medoids								✔
	Support Vector Machine	✔	✔	✔	✔	✔	✔		✔
	Neural Networks	✔	✔	✔	✔	✔	✔	✔	✔
	Deep Convolutional Neural Networks				✔				✔
	Learning Vector Quantization								✔
Crops Covered	Rice	✔	✔			✔		✔	✔
	Grape								✔
	Orange								✔
	Soybean						✔		✔
	Apple						✔		✔
	Blueberry								✔
	Cherry								✔
	Corn					✔			✔
	Peach								✔
	Bell Pepper					✔			✔
	Potato					✔			✔
	Raspberry								✔
	Squash								✔
	Strawberry					✔			✔
	Tomato				✔	✔			✔
	Cotton			✔		✔		✔	✔

Table II. Feature-wise comparison of dataset

Dataset	Crops available	Number of images	Disease covered	Method of capture	Limitations
Maize Dataset [2]	Maize	18,222	Northern leaf blight	hand-held camera; with a camera mounted on a boom; with a camera mounted on a small unmanned aircraft system	All photographs were taken in a single field which limits the generalizability of the data, as symptoms of the same disease in other regions may present or develop differently
Rice Dataset [3]	Rice	40	Bacterial leaf blight, Brown spot, and Leaf smut	Hand-held camera	Small Dataset
PlantVillage Dataset [4]	Grape, Orange, Soybean, Apple, Blueberry, Cherry, Corn, Peach, Bell Pepper, Potato, Raspberry, Squash, Strawberry Tomato	54,309	Fungal, bacterial, mold, viral, mite-affected	Hand-held camera	Extracted leaves on plain background are CNN trained which did not perform well on field images

Table III. Different category of diseases and symptoms available in different datasets

Disease Categories		Diseases (Agent)	Symptoms	Dataset 1 [2]	Dataset 2 [3]	Dataset 3 [4]
Fungal	Apple	Cedar Apple Rust	Pale yellow pinhead sized spots on the upper surface of the leaves which gradually enlarge to bright orange-yellow spots.			✔
		Apple Scab	Pale yellow or olive-green spots on the upper surface of leaves. Dark, velvety spots may appear on the lower surface.			✔
		Botryosphaeria obtusa	Slightly sunken reddish/brown spots			✔
	Cherry	Podosphaera spp	Light-green, circular lesion on either leaf surface, a subtle white cotton-like growth develops in the infected area.			✔
	Corn	Cercospora zeae	Grey leaf spots surrounded by a yellow halo which become irregular or extended pale-brown streaks.			✔
		Maydis Puccinia sorghi	Circular pustules, powdery, brown becoming brown-black as the plant matures.			✔
		Exserohilum turcicum	Cigar-shaped or elliptical necrotic grey-green lesions on the leaves.	✔		✔

	Grape	Guignardia bidwellii	Tan, circular spots with dark brown margin.			✔
		Phaeomoniella spp	Superficial brown to purple spots scattered over the leaf surface.			✔
		Pseudocercospora vitis	Spots on the leaf			✔
	Potato	Alternaria solani	Dark lesions with yellow border which may form concentric rings of raised and sunken tissue on the leaves.			✔
	Squash	Erysipha cichoracearum/Sphaerotheca fuliginea	White powdery growth on the upper surfaces of leaves.			✔
	Strawberry	Diplocarpon earlianum	The leaves turn brown and dry up.			✔
	Tomato	Alternaria solani	Oval shaped lesions with a yellow chlorotic region across the lesion.			✔
		Septoria lycopersici	Water-soaked spots or circular greyish-white spots on the underside of older leaves.			✔
		Corynespora cassiicola	Small, pinpoint, water-soaked spots on the leaf.			✔
		Fulvia fulva	Randomly spaced, diffuse pale-green or yellowish spots.			✔
Bacterial	Orange	Candidatus Liberibacter	Yellowing of one limb or one area of canopy, yellowing of leaf veins; blotchy mottling and/or green islands (spots) surrounded by completely yellow leaf tissue.			✔
	Peach	Xanthomonas campestris	Spots on the leaf			✔
	Bell pepper	Xanthomonas campestris	Small water-soaked areas			✔
	Tomato	Xanthomonas campestris pv. Vesicatoria	Small water-soaked spots.			✔
	Rice	Bacterial Leaf Blight	Yellow-orange stripes on leaf blade.		✔	✔
		Brown Spot	Spots on the leaf		✔	✔
		Leaf Smut	Black Spores, heavily infected leaves turn yellow, and the leaf tips die and turn grey.		✔	✔
Mold	Potato	Phytophthora Infestans	Sporangiophores on the surface of the leaves.			✔
	Tomato	Phytophthora Infestans	Leaves turn yellow and may become curled and deformed			✔
Viral	Tomato	Tomato Yellow Leaf Curl Virus	Leaves become reduced in size, curl upward, appear crumpled and show yellowing of veins and leaf margins.			✔
		Tomato Mosaic Virus	Dark green mottling or mosaic.			✔
Mite-Affected	Tomato	Tetranychus urticae	Pale Spots, in severe cases leaves appear bronzed or silvery.			✔

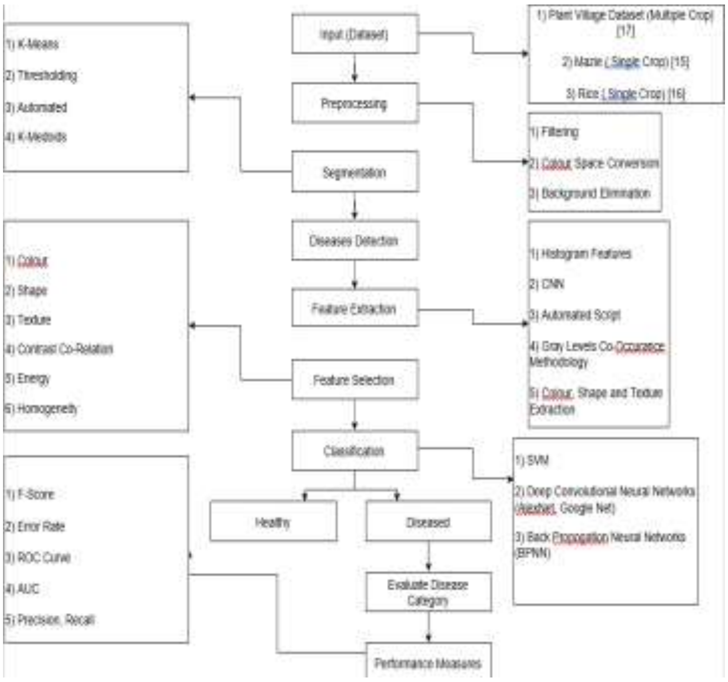

Fig 1: General Process Flow

B. Preprocessing

This is an important step before analyzing the image as it helps remove any noise present in the input image by carrying out operations at the lowest abstraction level. If advanced algorithms are applied before removing noise, then it may lead to unwanted results. Table IV gives an overview of the different preprocessing techniques, their advantages and disadvantages.

C. Segmentation

This technique is useful in dividing the images into several segments or sets of pixels. The idea is that at least one segment will have the region of interest (ROI) which can be utilized for further sophisticated algorithms. Here our ROI is the diseased part. Table V gives a brief overview of the different segmentation techniques like K-Means, Thresholding, Automated and K-Medoids.

D. Disease Detection

After receiving the segmented art from the earlier step, the next step is to analyze the diseased part and to detect the most probable disease.

Table IV. Comparison of Preprocessing Techniques

Techniques	Background Elimination	Color space conversion	Filters
Different sub-techniques	Computation of binary mask for manual selection of region of interest followed by addition of the image obtained in the previous step and original image to remove background. Finally, eliminating additional white spaces by cropping. [11]	• RGB to L*a*b* color space conversion [11]. • RGB to gray-scale conversion [12] • Resizing to standard resolution and then converting RGB image to gray-scale image. [13]	Replacement of every pixel value with the median of the neighborhood intensity
Advantages	Closely resembles human perception and has a strong inclination towards good segmentation results. [11]	This gives better splitting of information about chrominance and a color space independent of device model. [11]	Filters like median filter help with salt-pepper noise removal. [14]
Disadvantages	Difficult to deal with sudden lighting changes.	RGB color space is sensitive to non-uniform illumination; L*a*b* color space faces singularity problem.	The type of filter to be used is specific to the type of noise in the image. With the incorrect filter, image details could be degraded.

Table V. Comparison of Segmentation techniques

Techniques	K-Means	Thresholding	K-Medoids	Automated
Brief Description	Cluster centroids are recalculated after every sum of distance calculation. Three partitions generated corresponding to different portion of leaf. To obtain clusters corresponding to different color, texture and shape [13].	Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels each side of the threshold [7] [10].	Random select for initial medoids which will randomly select non-medoids object. Cost is computed for swapping and if the cost is negative then perform the swap operation [14].	Segmentation is automated with the help of a script.
Advantages	Iterative K-Means reduces the sum of distance value by reassigning every pixel to its closest neighbor [11].	Helps detect diseases having a particular color symptom [11].	More robust as compared with K-Means. Less affected by outlines and other extreme values	The script is tuned to perform well for Plant Village dataset [13].
Disadvantages	Choosing a k value manually for the image dataset is difficult.	It may not work for diseases showing large color invariability.	The first k medoids are chosen randomly. Hence, it may obtain different results for different runs on the same dataset.	Low Robustness and high complexity of implementation

Table VI. Different Feature Extraction techniques available

Paper	[11]	[12]	[13]	[14]	[15]	[16]
Techniques:	Histogram: Features extracted are color, texture	CNN: Contextual information such as color of leaves is extracted	Color, shape and texture features extracted	GLCM*: Shape feature is described by gray level sampling. SGDM**: Contrast, energy, local homogeneity and correlation are extracted from the HUE component.	CNN: Performed feature extraction of color for both full-color and grey-scale approach	Automated Script: Color, lightness and saturation values extracted

GLCM*: Grey Level Co-Occurrence Methodology; SGDM**: Spatial Grey Level Dependence Matrices

E. Feature Extraction

To detect the probability of disease it is important to analyze certain features. The most common features extracted from a leaf image are color, texture, shape, size of spots on them, etc. The techniques used are: Histogram preprocessing, CNN, Automated Script, Gray Level Co-occurrence Methodology or Spatial Gray Level Dependence Matrix. In Histogram techniques, the common features extracted are color and texture. When grayscale images are used to extract features, it is fairly convenient to use CNN as opposed to histogram extraction as histogram approaches tend to generalize pixel values to threshold values. Automated Script is similar to both histogram and CNN approaches in dealing with color and extraction values. It was noted that the Histogram and Automated Script techniques are exclusively used to extract the color feature. Grey Level Co-occurrence Methodology (GLCM), on the other hand incorporates a statistical approach to describe the shape feature by gray level sampling whereas Spatial Gray-level Dependence Matrices (SGDM) extracts features like contrast, energy, local homogeneity and correlation from the HUE component. The usage of GLCM and SGDM together is advantageous for grayscale as well as colored images and covers most features. Table VI illustrates the usage of these techniques in various papers.

F. Feature Selection

Once the features are extracted, not all features contribute equally to disease detection. To remove data redundancy, it is vital to select only those features which can alone contribute to the feature vector and still give a good accuracy. Each feature has certain advantages and disadvantages with respect to disease detection on leaves. Table VII gives an overview of the features selected in the previous section. From the table it is clear that the more the number of features, the more the complexity increases and better the diseased portions of a leaf are targeted for the next section. A few common features selected include color and texture.

Table VII. Summary of selected features in referred papers

Paper	Features Selected	Advantages	Disadvantages
[11]	Color features (Color moments, color auto correlogram, HSV histogram), texture features (Haralick, Gabor features, 2D DWT)	In depth analysis of features are done before selection. Maximum good quality features are chosen to give maximum output.	Shape based feature extraction is left out and will return error values for grayscale images. Harder to implement.
[12]	Contextual information like color of the leaf.	Simple to implement.	Does not take into account grayscale images.
[13]	Color, shape and texture	Covers the three important feature analysis for disease detection. Efficiently detects most patterns that indicate that plant is diseased.	Hard to implement and complexity is increased.
[14]	Contrast, Correlation, Energy, Homogeneity	Uses pixel-based comparison methods to detect abnormalities. Covers newer and more efficient methods. Also incorporates grayscale images.	Difficult to implement. Requires more storage overhead.
[15]	Features selected for both full-color and grey-scale approach	Works for both color and grayscale images. Comparatively simpler to implement.	Shape and texture-based extraction is left out.
[16]	Color, lightness and saturation values	Adds to basic color methods to give better results while detection of leaf surface anomalies.	Even though it can accurately detect larger area of irregularities on a leaf, it fails to do so for smaller spots.

Table VIII. Feature-wise comparison of various classifiers

Classifier	Advantages	Disadvantages	Accuracy
SVM	Can handle leaf images in complex background with different lighting conditions. Gives high accuracy Overfitting not usually experienced.	High Complexity Extensive memory requirement for multiclass classification of plant diseases. Longer training phases	85 % – 95 %
Deep Convolutional Neural Networks	With a good training dataset, complex leaf images, different background conditions, noisy data can be handled. Gives high accuracy	High chances of overfitting. High processing time	>92%
Back Propagation Neural Networks (BPNN)	Applicable to a wide range of leaf images taken from different angles under different conditions. Easy to implement.	Slow learning rate. Hard to know how many neurons as well as layers are required.	>90%

G. Classification and Performance Metrics

In the papers reviewed, classification is performed for majorly two classes, namely, healthy or diseased. In some papers, they have classified the diseases into 6 classes. In the general algorithm, if the input leaf is classified as healthy then the algorithm halts. On the other hand, if the algorithm classifies leaf image as diseased then we evaluate the type of disease and performance metrics like F-score and accuracy. Table VIII describes the classifiers used by the research papers and the accuracy range obtained by them. We have also inferred the advantages and disadvantages when these classifiers which will help in making decision about which classifier would be suited given the other parameters like dataset, computation time and memory constraints. To validate the accuracies achieved by different plant disease detection algorithms, it is vital to include performance metrics. This comparative study can help identify the appropriate metrics to be incorporated. The average accuracy range observed in majority papers spanned between 87% to 96%. The best-case accuracy, 99.84% is achieved by using F1 score on Full Color and Grey Scale Model. Other performance metrics include Error Rate, ROC curve, AUC, Precision, Recall etc. However, these performance metrics were observed on dataset taken in a controlled environment which do not indicate the performance of the system on real time condition.

III. KEY CHALLENGES AND FUTURE DIRECTIONS

We have identified some of the important challenges – both extrinsic like capture conditions and intrinsic like disease variations, to explore in depth their causes and their impact on the performance of the techniques discussed so far. Our aim is to overcome these challenges and eliminate their adverse effects in our proposed approach in the future. The challenges are as follows:

a) The resolution of camera must be appropriate so that the quality of leaf image is not compromised.

b) Influence of climate while taking pictures can change the illumination spots which might affect the detection algorithm.

c) The symptoms produced by different diseases may be very similar and they may be present simultaneously, leading to fuzzy classification instead of crisp.

d) Size of the dataset should be large enough to efficiently train the model for testing dynamically without increasing computing time.

e) Appropriate pre-processing techniques to avoid noise interference at disease detection stage.

f) Deciding the division of training and testing data to avoid overfitting and underfitting.

g) Most of the datasets available are country or region specific. To work with plants from other countries or regions, raw images from the fields have to be taken, labelled and then processed. Labelling the different diseases is possible only under expert supervision.

h) Different filters work best for different diseases, Ex: Otsu filter gives best results for Bacterial diseases. So, choosing the right set of filters is very important.

IV. CONCLUSION

In this paper we have discussed the need for a system to detect and classify plant disease supported by statistical parameters. Furthermore, we have reviewed various research papers based on the general flow followed by plant disease detection systems like input leaf, pre-processing, segmentation, disease detection, feature extraction, feature selection, classification as healthy or diseased, performance measures and accuracy involved. We also compared our survey paper with existing survey papers to help identify the research gaps and work on them further. Through this survey, we have identified key techniques involved at each stage and methods that are most used within each stage. The different image processing techniques can be combined with different classifiers to produce optimal results, especially when the data is noisy.

REFERENCES.

[1] Sachin D. Khirade and A. B. Patil. “Plant Disease Detection Using Image Processing”, International Conference on Computing Communication Control and Automation (ICCUBEA), 2015, pp. 768-771. IEEE, 2015.

[2] Wiesner-Hanks, Tyr & Stewart, Ethan & Kaczmar, Nicholas & DeChant, Chad & Wu, Harvey & Nelson, Rebecca & Lipson, Hod & Gore, Michael. (2018). “Image set for deep learning: Field images of maize annotated with disease symptoms” BMC Research Notes. 11. 10.1186/s13104-018-3548-6.

[3] Jitesh P Shah, ‘Rice Leaf Disease Dataset’, 2014. Available: https://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases.

[4] Mohanty Sharada P., Hughes David P., Salathé Marcel, “Using Deep Learning for Image-Based Plant Disease Detection”, Frontiers in Plant Science Vol. 7 ISSN: 1664-462X, 2016.

[5] Jitesh P. Shah, Harshadkumar B. Prajapati, Vipul K. Dabhi, “A Survey on Detection and Classification of Rice Plant Diseases” ,978-1-5090-1936-6/16/$31.00 ©2016 IEEE.

[6] Arpita Patel, Barkha Joshi, “A Survey on the Plant Leaf Disease Detection Techniques”, International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 1, January 2017.

[7] Santhosh Kumar.S, B.K. Raghavendra, “Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review”, 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS).

[8] Sagar Patil, Anjali Chandavale, “A Survey on Methods of Plant Disease Detection”, International Journal of Science and Research (IJSR)ISSN (Online): 2319-7064.

[9] Gurleen Kaur Sandhu, Rajbir Kaur, “Plant Disease Detection Techniques: A Review”, 2019 International Conference on Automation, Computational and Technology Management (ICACTM) Amity University.

[10] M. Malathi, K. Aruli, S. Mohamed Nizar, A. Sagaya Selvaraj, “A Survey on Plant Leaf Disease Detection Using Image Processing Techniques”, International Research Journal of Engineering and Technology (IRJET) Volume: 02 Issue: 09 Dec-2015 e-ISSN: 2395-0056 p-ISSN: 2395-0072.

[11] Sukhvir Kaur, Shreelekha Pandey and Shivani Goel. “Semi-automatic leaf disease detection and classification system for soybean culture”, IET Image Process., 2018, Vol. 12 Issue. 6, pp. 1038-1048.

[12] Melike Sardogan, Adem Tuncer and Yunus Ozen, “Plant Leaf Disease Detection and Classification based on CNN with LVQ Algorithm”, 3rd International Conference on Computer Science and Engineering, UMBK, 2018

[13] Sushma S. Patil, Suhas K. C, “Identification and Classification of Cotton Leaf Spot Diseases using SVM Classifier”, International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 4, April - 2014, ISSN: 2278-0181.

[14] Arti N. Rathod, Bhavesh A. Tanawala, Vatsal H. Shah, “Leaf Disease Detection Using Image Processing and Neural Network,” International Journal of Advance Engineering and Research Development (IJAERD), Volume 1, Issue 6, June 2014, e-ISSN: 2348 - 4470, print-ISSN:2348-6406.

[15] Belal A. M. Ashqar, Samy S. Abu-Naser, “Image-Based Tomato Leaves Diseases Detection Using Deep Learning”, International Journal of Academic Engineering Research (IJAER) ISSN: 2000-001X Vol. 2 Issue 12, December – 2018.

[16] Sharada P. Mohanty, David P. Hughes and Marcel Salathé, “Using Deep Learning for Image-Based Plant Disease Detection”, Front. Plant Sci., 22 September, 2016.

[17] Hilman F. Pardede, Endang Suryawati, Rika Sustika, and Vicky Zilvan, “Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases”, 8 International Conference on Computer, Control, Informatics and its Applications, 2018.