International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 03 October 2023 **Accepted:** 22 October 2023*

Analysis on Leaf Disease Identification using Classification Models

R. Salini[1], S. Derisha Mahil[2], J. Chenni Kumaran[3], G. Bhuvaneswari[4], C. Vijiayalakshmi[5], J. Savija[6]
[1]Department of Computer Science and Engineering
Panimalar Engineering College
Chennai,India
shalinirajendran@gmail.com
[2]Department of Artificial Intelligence and Machine Learning
S. A. Engineering College
Chennai,India
derisha@saec.ac.in
[3]Division of Spatial Informatics ,Department of Computer Science and Engineering
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Science, Chennai, Tamilnadu, India.
drchennikumaran@gmail.com
[4]Department of Computer Science and Engineering
Saveetha Engineering College
Chennai,India
bhuvankerani@gmail.com
[5]Research Scholar
SRM institute of science and technology
Chennai,India
vijidurai88@gmail.com
[6]Department of Artificial Intelligence and Data Science
Velammal Engineering College,Chennai,India
savija01@gmail.com

Abstract—The Researchers have all been aware of the rising food demand brought on by the population's rapid growth and the high mortality rates caused by medical developments. One of the many farming practises where computerization in agriculture has made significant progress is the identification of numerous plant diseases. The focus of almost every nation has shifted towards mechanising agriculture in order to achieve accuracy and precision and to meet the continually increasing demand for food. Identification of plant diseases is one of the most difficult tasks in agriculture and has a significant effect on crop yield. Artificial intelligence has recently begun to concentrate on smart agriculture science.Ground-breaking methods in plant science through deep learning and hyperspectral imaging to locate and recognise plant diseases has been addressed in this study.

Keywords-Plant disease; Machine learning algorithm; Leaf Disease; Image Processing;Feature Extraction.

I. INTRODUCTION

The agricultural cycle starts with the planting of seeds and ends with the harvesting of the crop. During this crucial process, diseases, pest infestations, weed growth, an unsuitable soil type, a shortage of water, and other factors might have an impact on the overall yield. Artificial intelligence and machine learning speed up the resolution of the challenges mentioned. By tackling the underlying assumptions of prior learning experiences, artificial intelligence progresses.

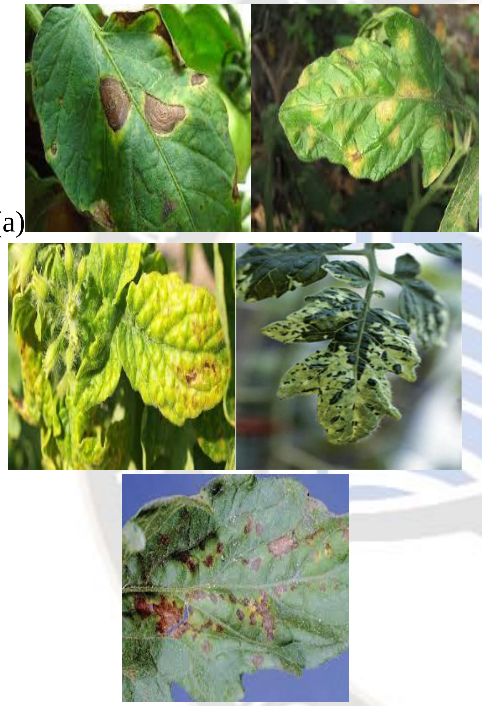
Historically, detection and cure for plant disease was provided by organizations such as local plant clinics. For successful crop production in the farm, there was a need to monitor the disease and health of plants. In earlier days, experienced agriculturists performed analysis and monitoring of leaf diseases manually, which is a time consuming process.

Convolutional neural networks, artificial neural networks, and other Machine Learning (ML) techniques have shown to be efficient at resolving the challenges mentioned. Therefore, a more accurate diagnosis and prompt proper crop care are required to shield them from severe losses. Finding plant diseases early on is crucial since they might have an impact on the yield's overall quality and quantity [1]. Therefore, it is crucial to identify and classify plant diseases. Plants like the tomato, potato, pepper, and rice can get diseases in their roots, stems, leaves, and fruits, among other plant components. Target spot,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 03 October 2023 **Accepted:** 22 October 2023*

leaf mold, mosaic virus, yellow leaf curl virus, bacterial spot, early blight, healthy, late blight, septoria leaf spot, spider mites, and two spotted spider mite are some of the major diseases that can harm tomatoes. Diseases in the rice crop typically include leaf smut, brown spot, and bacterial leaf. In the case of potatoes, diseases include early blight and late blight; in peppers, diseases can be bacterial [2]. In many instances, the disease's symptoms can be seen on the plant's fruit, stem, and leaves. In this article, we explain the image processing method used for plant leaf disease identification. In addition to this, we discuss the percentage of accuracy for detecting and classifying the leaf diseases at the early stage of disease by applying Machine Learning (ML) models, and expedite various solutions to stray away from further infestations [3].

(a)



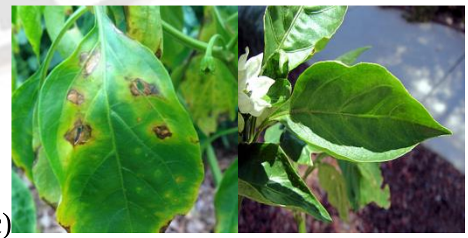
Target spot Leaf mold Yellow Leaf Curl Virus
Mosaic virus Bacterial spot

Early blight Late blight Healthy
Septoria Leaf Spot Spider mites

(b)

Brown spot leaf smut Bacterial leaf smut

(c)



Bacterial Healthy

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 03 October 2023 **Accepted:** 22 October 2023*

(d) 

Early blight [6] healthy [6] late blight [6]

Figure 1. Examples of leaf photos displaying several disease types in (a)Tomato (b)Rice (c) Pepper (d) Potato

II. LITERATURE SURVEY

A significant amount of research is being performed in the field of automated detection and classification of plant disease using machine learning algorithms, and this article includes an overview of several useful ML algorithms developed by various researchers, as well as the accuracy of the algorithms in terms of classification and forecasting of the detected result.

TABLE 1 SURVEY OF LITERATURE

.

References	*Algorithm Used*	*Observations*	*Accuracy*
[2]	Hyper spectral measurement	Techniques to identify diseases in leaves by imaging them in non destructive and non contact way.	97.82%
[3]	Moving Center Hyperspace Classifier(MCH) and Locally Linear Embedding Classifier(LLE).	Technique for plant leaf recognition.	97.47%
[4]	Back propagation Neural network	Technique for Identifying rice plant diseases.	100%
[5]	Scale-Invariant Feature Transforms	Use of SIFT as an approach to find the presence of the disease in plants.	98.35%
[6]	Convolutional Neural Network And Image processing	The application distinguishes rice irritations and infections was produced using CNN and image processing to assist farmers.	99.6%
[7]	SVM model using a thermal imaging camera and a pair of visible light imaging cameras	Involves classification temperature and colour for easy detection of infected plants from clusters caused by the causal pathogen Oidium neolycopersici.	98.7%
[8]	Multi-class Support Vector machine [16]	Technique is used for the detection and classification of disease in apple fruit.	93%
[9]	Nearest neighbours, RF, DT and Naive bayes	Technique helped in diagnosis of diseases in plants.	96.3%
[10]	Backpropagation neural network	A classification technique that explains the afflicted area's colour, texture and uses these as features.	97.3%

Also this article includes an overview of several useful ML algorithms developed by various researchers, as well as the accuracy of the algorithms in terms of classification and forecasting of the detected result

III. DISEASE DETECTION AND CLASSIFICATION PROCEDURE

This section provides a detailed explanation of the classification model's working process and overall architecture. The block diagram and fundamental procedures for image processing-based plant disease detection and classification are shown in Figure 2. Picture acquisition, image preprocessing, image segmentation, feature extraction, classification, and performance evaluation are some of the phases in this method. In order to detect the disease, we first acquire photographs of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *18 August 2023* ***Revised:*** *03 October 2023* ***Accepted:*** *22 October 2023*

leaves and then preprocess the images in accordance with the dataset using image processing algorithms. After completing the aforementioned tasks, we categorize the illness in accordance with the pertinent data set [8]. The following is a detailed explanation of each of these processes.

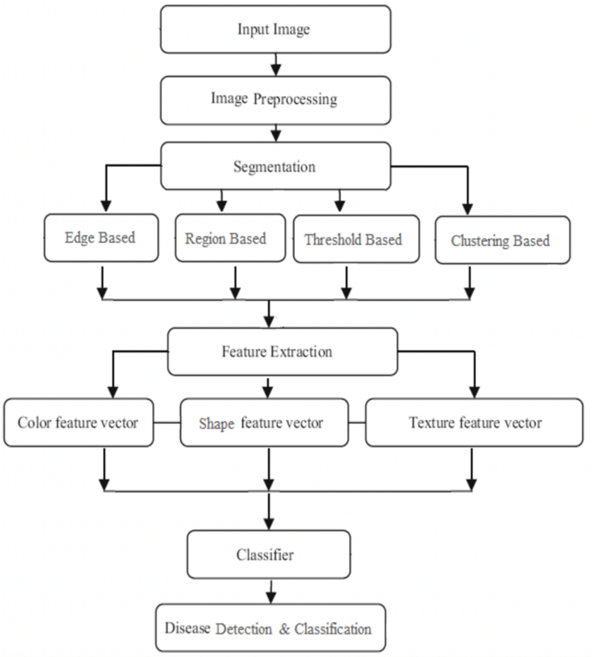


Figure 2. Steps of Plant Leaf Disease Detection and Classification

A. Input Image

This step is also known as image acquisition step, this is the first step in all machine learning classifier methodology. Image acquisition is the process in which we acquire the input image and convert it into the desired output format. We can either acquire these images from genuine websites specifically for plant images which are verified and approved for research purposes or manually take sample images using digital cameras. All the pictures of healthy and diseased plants are stored in RGB colour form with unique names. Three classes make up the rice dataset: leaf smut (40), brown spot (40), and bacterial leaf blight (40). (40). The pepper datasets are divided into two categories: bacterial (300) and healthy (300). (300). The potato datasets are divided into three classes: healthy (100), late blight (500), and early blight (500). (500). Ten classes make up the tomato datasets: spider mites with two spots (500), target spot (500), mosaic virus (200), yellow leaf curl virus (1000), bacterial spot (500), septorial leaf spot (500), early blight (500), healthy (200), late blight (500), and leaf mould (500). The sample count for the plant disease category is shown by the numbers in parenthesis [2].

B. Image pre-processing

The second phase in the classification and identification of leaf disease is image pre-processing [4]. It is used to take the necessary data out of the photos. It entails the removal of background and noise as well as the correction of distortion and enhancement of key features needed for image processing and analysis. It aids in minimising image redundancy. Using different colour models like CIELAB, YCbCr, and HSV, the RGB image is turned into grayscale during colour space conversion. After scaling, converting to a different colour space, and enhancing the image, Histogram equalisation techniques are employed to indicate intensities.

C. Image Segmentation Techniques

Plant leaf disease detection and classification heavily rely on image segmentation. The input image is segmented into different sections in this step. In essence, it divides the dataset's differences into discrete clusters [2]. These clusters can be further examined and divided into other groupings. To further process these categories, pertinent data from the picture data is extracted. There are two approaches to segment an image; one by being dependent on the discontinuities and two being dependents on similarities. In discontinuities, the images are divided depending on factors such as unexpected changes like in edge detection. whereas in similarities, the image is divided depending on predefined circumstances.

Generally, the segmentation is achieved with varied technologies like conversion of RGB image to HSI(hue (H), saturation (S), intensity (I)) model, k-means clustering, otsu' methodology etc[4].We employ the k-means cluster in this instance since it is the most effective at identifying the leaf disease and is widely utilised by researchers. The image of the sick leaf is segmented into three groups using the k-means cluster algorithm.

D. Feature Extraction

The process of acquiring the features and details from the input image for classification is called feature extraction. The entire set of retrieved features is represented as an entity and subdivided into three categories based on texture, colour, and shape[3]. Some datasets contain combinations of features - these are heterogeneous datasets. Texture is one of the best features for detection of plant diseases, similarly shaped features such roundness, area, concavity and eccentricity. Whereas colour features generally include mean, variance and skewness of grey value of R G B and H S V components. Colour features also include colour ratio according to the details extracted from colour segmented images.

E. Feature Extraction

This is the final step in detection and classification of plant diseases [10]. Classifiers are employed in the classification process.[2]; here we discuss popular classification models which are used by numerous researchers, like Naive Bayes (NB), K-Nearest Neighbour(KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Minimum Distance Classifier (MDC) and Artificial Neural

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 03 October 2023 **Accepted:** 22 October 2023*

Network (ANN). The following section will contain the brief explanation for the above mentioned classifier models.

IV. MACHINE LEARNING (ML) CLASSIFICATION MODELS

Machine learning classification models are of four types; Supervised classification, Unsupervised classification, Semi-Supervised classification and Reinforcement classification. While the set of classes for supervised classification [12][13] are known in advance, the set of classes for unsupervised learning are unknown. Reinforcement model continues to improve with criticism by learning examples and conducts.

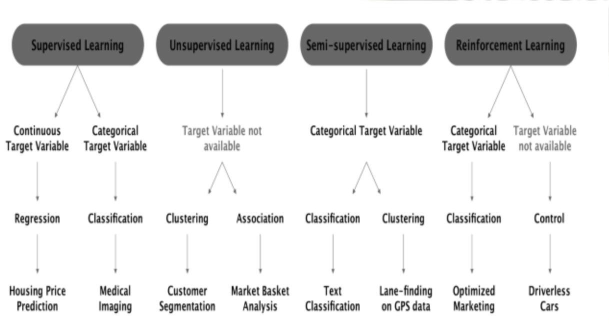


Figure 3. Types of Machine Learning Models

A. Naive Bayes (NB)

Naive Bayes classifier is a "probabilistic classifier". According to Bayes theorem, Naive Bayes classifier is implemented by independent presumptions among the features. It assumes that the class labels have been given posterior probabilities and that the prior probabilities of the patterns are well understood. The posterior probability for these hypotheses is estimated with the highest likelihood and assigned to a particular class label. Despite the fact that these theories typically do not perform well in the real world, they do well in many classification applications [3].

B. K-Nearest Neighbors (KNN)

Naive B KNN classifier makes the categorization of unidentified instances based on a similarity measure or distance function. It is a nonparametric, lazy learning, supervised machine learning model. It is typically utilised in pattern recognition. It is based on the nearest neighbour theory. There is no training pattern necessary for this classifier. In the testing phase, all training patterns are used to categorise the test pattern based on similarity function. It acts like a particular form of instance-based learning where the functions are locally approximated and all calculations have changed up until the classification method's conclusion.[8].

C. Decision Tree (DT)

This approach combines supervised classification with regression. In supervised learning, which builds classifiers, the data is divided into a variety of smaller trees and sub-tree structures depending on the division to build the higher inconsistency. The attribute selection measures such as Gini index, entropy are usually employed as disparity measures. For implementers, evaluation of the results using this model will be easy. If the tree had learned with no restriction of tree depth then DT would have generated minimum training error. Several types of decision trees like CART, C4.5 and ID3 are most commonly used in ML and data mining applications [7].

D. Random Forest (RF)

It is a collection of randomly generated DT (Decision Tree) classifiers. At the training time, multiple DTs are constructed. The testing dataset's class labels are determined by the voting of all classification trees, which becomes the result of this classifier. While building each individual tree, this classifier model uses bagging and random features. This model endeavours to make an unrelated forest of trees. The prediction of forest of tree's performance will be more accurate than the individual tree [3].

E. Support Vector Machine (SVM)

SVM is a supervised machine learning classifier model. It is classified by the separating hyperplane. This model gets the best hyperplane, which maximises the distance among the nearest data points. This distance is referred to as a margin. There are two varieties of SVM: linear and nonlinear. The uniform distributions of data are permitted to draw a straight hyperplane amongst the classes in a linear SVM classifier. In comparison, non-linear SVM classifier use data that are dispersed and possess high dimensions. Non-linear SVM classifiers are used to solve the majority of real-world applications.

SVM has a feature called kernel tricks that is useful for nonlinear classification. Various generic functions, including radial basis, polynomial, and linear functions, are used to convert the features. The transformation of characteristics increases the training phase of the classification process.

F. Minimum Distance Classifier (MDC)

MDC is a classifier method to classify the images with Manhattan distance. The advantage of MDC is ease of implementation and also takes less time for classification. Manhattan distance is also termed as "cityblock" distance. A (x1, y1) and B (x2, y2) are the two points and the distance between these two points [12] are termed as Manhattan distance (d). It can be computed by using the equation 1.

$$d = |x_1 - x_2| - |y_1 - y_2| \qquad (1)$$

G. Artificial Neural Network (ANN)

The neural architecture of the human brain serves as the foundation for ANN [9]. Techniques like ANN, which are modelled after the analytical skills of the human brain, can be used to solve problems that are beyond the capacity of contemporary computers. As a result, a neural network has several hidden layers and is created based on the inputs or outputs that are needed. The weights of the neurons determine the network's individuality or intelligence. In order to obtain the precise output, neural network weights must be changed for learning. The back propagation neural network is quite important in this situation. The goal of the back propagation algorithm, which employs the gradient descent method, is to minimise the error function. To reduce the error value, the weights of each layer in the network are adjusted one at a time at each iteration.

V. CONCLUSION

This article describes a detailed survey on the prediction of plant leaf disease detection and classification. The literature survey concludes better results for plant leaf disease detection and classification with several ML classifiers. Performance improvement for the detection and classification of plant leaf disease is said to be a complicated task[8]. The review work of this article helps in doing that.

The following conclusion has come up from study of mentioned ML classification models. In more than five articles SVM and Neural Network were found above 90% accuracy, which compete with the best ML classification models available for classifying high-dimensional data sets. The optimal outcomes were obtained in low computing endeavours that prove the efficacy of all ML models in early recognition and classification of the leaf disease. Hybrid algorithms can even be used to improve the recognition rate in the process of classification. In the future we are going to implement some new and hybrid models for our research work.

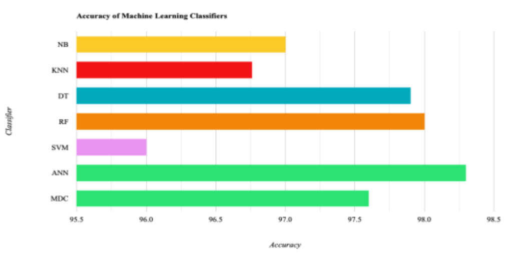


Figure 4. Accuracy of classifiers

REFERENCES

[1] Monika Lamba, Yogita Gigras, and Anuradha Dhull ,Classification of plant diseases using machine and deep learning ; https://doi.org/10.1515/comp-2020-0122 .

[2] Panigrahi K.P., Das H., Sahoo A.K., Moharana S.C. (2020) Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms. In: Das H., Pattnaik P., Rautaray S., Li KC. (eds) Progress in Computing, Analytics and Networking. Advances in Intelligent Systems and Computing, vol 1119

[3] Husna Tabassum, Prasannavenkatesan Theerthagiri. "Performance Analysis of AI- based Learning Models on Leaf Disease Prediction", 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), 2022 ; publication

[4] Monika Lamba, Yogita Gigras, Anuradha 1% Dhull. "Classification of plant diseases using machine and deep learning", Open ComputerScience, 2021 Publication

[5] Rakesh Meena, Sunil Joshi, Sandeep Raghuwanshi. "VGG19-based Transfer learning for Rice Plant Illness Detection", 2022 6th International Conference on Electronics, Communication and Aerospace Technology, 2022.

[6] Aashish Jha, Madhavi Purohit, Vivek Maurya,Amiya Kumar Tripathy. "Plant Leaf Disease Detection And Classification Based On Machine Learning Model", 2022 IEEE Bombay Section Signature Conference (IBSSC), 2022 Publication.

[7] Sakshi Sharma, Vatsala Anand, Swati Singh. "Classification of Diseased Potato Leaves Using Machine Learning", 2021 10th IEEE International Conference on CommunicationSystems and Network Technologies (CSNT), 2021.

[8] Vishakha A. Metre, Sudhir D. Sawarkar. "Reviewing Important Aspects of Plant Leaf Disease Detection and Classification", 2022 International Conference for Advancement in Technology (ICONAT), 2022.

[9] Eftekhar Hossain, Md. Farhad Hossain,3Mohammad Anisur Rahaman. "A Color andTexture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier", 2019 International Conference on Electrical, Computer andCommunication Engineering (ECCE), 2019; Publication

[10] Khirade, Sachin D., and A.B. Patil. "Plant Disease Detection Using Image Processing",2015 International Conference on Computing Communication Control and Automation, 2015.

[11] Rose, J. D. ., R, V. R. ., Lakshmi, D., Saranya, S. ., & Mohanaprakash, T. A. . (2023). Privacy Preserving and Time Series Analysis of Medical Dataset using Deep Feature Selection. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 51–57. https://doi.org/10.17762/ijritcc.v11i3.6201

[12] Ramachandran Veerachamy, Ramalakshmi Ramar, S. Balaji, L. Sharmila, Autonomous Application Controls on Smart Irrigation, Computers and Electrical Engineering, Volume 100,2022, 107855,7906,https://doi.org/10.1016/j.compeleceng.2022.107855.

[13] G Soniya Priyatharsini, A. Jyothi Babu, M. Gnana Kiran, P.J. Sathish Kumar, C. Nelson Kennedy Babu, Aleem Ali,Self secured model for cloud based IOT systems, Measurement: Sensors, Volume 24,2022,100490,ISSN 2665-9174, https://doi.org/10.1016/j.measen.2022.1004