International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com
doi : https://doi.org/10.32628/IJSRCSEIT

Machine Learning-Based Approaches for Plant Leaf Disease Identification : A Survey

Nandha Kumar G, Dr. V. Vijayakumar

Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India

Professor, Computer Science and Controller of Examinations, Sri Ramakrishna College of Arts and Science-Coimbatore, India

ARTICLEINFO

Article History:

Accepted: 05 Sep 2023

Published: 16 Sep 2023

Publication Issue

Volume 9, Issue 5

September-October-2023

Page Number

118-124

ABSTRACT

Plant leaf diseases can significantly impact agricultural yields and food security. Machine learning-based approaches have emerged as a promising solution for the rapid and accurate identification of plant leaf diseases, aiding farmers in timely disease management and crop protection. This survey provides a comprehensive overview of the state-of-the-art in machine learning techniques for plant leaf disease identification. We explore various aspects of this domain, including the types of machine learning algorithms commonly employed, the diverse datasets used for training and evaluation, and the challenges associated with real-world deployment. Additionally, we discuss the potential impact of machine learning in agriculture and propose future research directions. By synthesizing existing knowledge and highlighting key trends, this survey serves as a valuable resource for researchers, practitioners, and stakeholders interested in leveraging machine learning to combat plant leaf diseases, thereby contributing to sustainable and resilient agricultural practices.

Keywords: Plant Leaf Diseases, Machine Learning, Agriculture, Classification, Crop Disease.

I. INTRODUCTION

Plant diseases pose a substantial threat to global agriculture, jeopardizing food production, economic stability, and food security. The timely and accurate identification of these diseases is pivotal for effective disease management and the preservation of crop yields. In recent years, machine learning-based approaches have emerged as a powerful tool in the realm of plant pathology, offering a promising solution to address these challenges. Machine learning leverages computational algorithms and statistical models to automate the process of disease detection by analyzing plant leaf images, sensor data, and other relevant information. These approaches have demonstrated the ability to provide rapid and precise diagnoses, aiding farmers and agricultural experts in making informed decisions about disease control measures, including targeted pesticide application and crop rotation.

This survey delves into the landscape of machine learning-based approaches for plant leaf disease identification, providing a comprehensive and up-to-date examination of the field. It explores the diverse techniques and methodologies used to develop and deploy machine learning models in the context of agriculture. By summarizing the current state of the art, this survey aims to elucidate the strengths and limitations of these approaches, highlight successful applications, and identify areas for further improvement and research.

As agriculture faces the dual challenges of meeting the demands of a growing global population while mitigating the environmental impact of farming practices, the integration of machine learning in plant disease identification offers a ray of hope. This survey seeks to empower researchers, practitioners, and stakeholders in the agriculture sector by offering insights into the latest advancements and future prospects of machine learning-based plant leaf disease identification, ultimately contributing to more resilient and sustainable agricultural systems. In recent years, the agricultural sector has witnessed a growing concern regarding the prevalence and impact of plant diseases. These diseases not only lead to reduced crop yields but also pose environmental challenges due to the excessive use of pesticides and fungicides.

The timely detection and management of plant diseases are critical for minimizing losses and ensuring food security. Conventional methods of disease identification often involve visual inspection by agricultural experts, which can be time-consuming and subject to human error. Machine learning techniques, fueled by advances in computer vision, pattern recognition, and data analysis, have become increasingly important in automating the process of plant leaf disease identification. These methods are capable of processing vast amounts of data quickly and accurately. While machine learning has shown promise, it also brings forth challenges such as dataset annotation, model generalization, and scalability. Addressing these challenges opens up numerous opportunities for innovation and impact.

1.1 Fundamentals of Plant Diseases

Plant disease is an important element in agricultural output that contributes to a decrease in both plant quality and plant quantity. In plant diseases, the most typical strategy is to use a classification and detection model.

i. Diseases Caused By Bacteria

The bacterial illness is commonly referred to as "bacterial leaf spot." To begin, it appears as little, golden blemishes on young leaves, which are regularly warped and twisted. Alternatively, it appears as dark, liquid, greasy blisters on older foliage. Both forms are commonly deformed and twisted.

ii. Diseases Caused By Viruses

All viral illnesses result in a decrease in output, and plants that have been infected with viruses often only have a short lifespan. The leaves are the most common location for disease symptoms to appear on plants; however, symptoms may also be caused by different viruses on the fruits and roots of the plant as well as the leaves. The investigation of a disease that is caused by a virus is a very difficult task. As a consequence of the virus, the plant's development may be slowed down, and its leaves may have a wrinkled and curled appearance.

iii. Diseases Caused By Fungus

Fungal infection may influence infected seeds, ground, output, weeds, and the spread of wind and water. It appears as water-soaked, gray-green patches on lower with more seasoned clears in the initial stage. After that, the spots go away, and white fungal growth spreads across the undersides. Yellow to white stripes appear on the top of the board of more weathered clears in wool build-ups. It spreads outward over the leaf surface, turning the leaf yellow.

iv. Detection of Failure

Aerial images, together with digital image processing techniques, have been widely utilised in precision agricultural applications. These photographs may now be readily transformed into a variety of deliverables, including ortho shots, water stress maps, plant disease infestation indices, and more. Crop failures detection is a tool that precision agricultural firms are eager to have, since it allows them to make early choices and save financial losses. The fundamental concept is to apply several mathematical operators in a sequential manner in order to expose failures that might otherwise go unnoticed.

II. MATERIALS AND METHODS

For plant disease identification based on plant leaves, the basis of plant classification mainly includes three: leaf shape, color and texture. Because the color of most leaves will change gradually over time, the classification of leaf color will be subject to time constraints. However, the shape characteristics of leaves in different periods are similar, generally only size changes, and different leaf contours are different, making it one of the main symbols of plant identification. Compared with color and shape, the texture veins of leaves have more stable characteristics, and the texture veins of different leaves differ greatly. As a plant identification standard, it is easier to distinguish the types of plants.

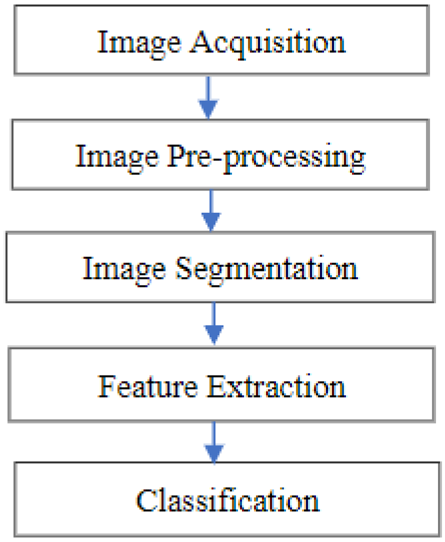

Figure 1. Plant Disease Detection: An Overview

The basis of image recognition is to extract the target features to be recognized in the image. The recognition process is mainly divided into four steps: image pre-processing, image segmentation, image feature extraction and image classification as shown in Figure 1.

III. RELATED WORKS

A plant disease prevents a plant from reaching its maximum potential of production. This definition includes non-infectious and infectious diseases [1] that pose threat to the agriculture industry by causing a decline in production and economic as well as reduction in the quality and the quantity of the plant products. A study conducted to report the effect of the plant disease on global production [2]. The research had shown that plant diseases had caused a high yield loss to subtle crops around the world being the wheat (30%), rice (40%), potato (21%), corn (41%) and soybean (30%). Extreme Learning Machine (ELM) is popular due to its simple design but good generalization. It is formed as single layer feedforward neural network that is used for classification with a minimal amount of modification on the weight the of inputs.

Thus, the complexity of the application of the model is significantly lower. In this work, the image features such as Haralick textures, Hue-Saturation-Value Histogram and colour moments are proposed to identify the plant disease from tomato leaves. The performance based on the accuracy rate of ELM classifier is expected to be on par to the with other classification models such as Decision Tree (DT) and Support Vector Machine (SVM). The complexity will be as well reduced.

ELM is proposed in [3] to detect plant diseases. Their model follows the basic steps of machine learning model such as feature extraction, training the classifier and classification. In their work, the artificial bee colony clustering (ABC) [4] is used isolate infected areas using the collective behaviour of insects such as bees. Second, the feature extraction is completed using transform encoded local patterns (TELP) to compute the texture analysis [5], gradient features and color histogram techniques from the data. For the training and testing, the author tested both classifiers namely Support Vector Machine (SVM) [6] and ELM, which resulted in ELM leading by 2% at the rate of accuracy at 97%. This work proves that ELM can surpass the conventional model, SVM in terms of classification accuracy.

Another approach on ELM is suggested in [7] with Simulated Annealing (SA) to classify the Jatropha Curcas Disease. In this work, the dataset of jatropha diseases is used in the process of model training, testing, and the optimising the output of the weights of SA [8]. The important role of SA is to tune the best weight value of neuron on ELM classifier. The model is further enhanced using the decision tree classifier to determine the jatropha seed disease. The aim of the decision tree classifier is to create a model capable of making a judgement according to the specialised knowledge. The result of this method gives out a much more desirable rate of accuracy at 94.74% compared to pure ELM at 66.67%.

This work [9] proposed a smartphone application to classify the coffee leaf diseases with ELM. The system recognises diseases such as leaf miners and leaf rust and calculates the rate of its severity. The image processing step involves the identification of a leaf from the image by separating the background and foreground of the image to obtain the Hue Saturation Value (HSV) and YCbCr colour space. This method is done via several segmentation methods that include histogram distribution, edge detection, neighbourhood detection, and clustering. Otsu algorithm [10] and the iterative threshold algorithm are used to perform a comparison to k-means in the process to segment the foliar damages. The results show the practicality of the model to identify feeding damage by obtaining the of the severity rate of 99.095%. Thus, the ability of the ELM model is proved for able to produce a significant result with the given image features of the coffee leaves.

In this paper [11], recommends a comparison of various machine learning models such as Support Vector Machine (SVM), K-Nearest Neighbours (KNN), RF and naïve bayes algorithm. It uses the public datasets of diseased leaves for training and testing purposes. Second, the histogram of an orientated gradient (HOG) is used to extract features such as hu moments, Haralick textures and colour histogram. The datasets of diseased and healthy leaves are then collectively trained under the aforementioned models. The comparison resulted with random forests having the highest accuracy at 70.14%. The model can be improved with implementing the ELM model as the random tree does not scale well with the amount features compared to ELM.

The key points are extracted from the database as well as from the test images to register and identify the image using SURF algorithm. It is carried out as local object detector [12]. After the completion of a trained system, this system assists in classifying the unknown examples into a learned class label. Different from the supervised learning, an unlabeled dataset is utilized and items are allocated to specific groups in unsupervised learning [13]. In general, classification task can use voting method, that is, the most visible category label in k samples is chosen as the prediction outcome [14]. With the utilization of various distance calculation techniques, distinct nearest neighbors are obtained due to which the results generated in classification are found diverse. The distance computation is supposed suitable to order to discover the KNNs [15]. The NNR (nearest neighbor rule) is a theoretical foundation of K-NN classifier. In case of enormous number of points, the error rate of the nearest neighbor is found less than twice the Bayes error rate. In case of minor error of the best possible optimal assignment, twice that error is also small [16].

SVM is referred as a second-class classifier model. Its standard model is described as the linear classifier with the highest interval infeature space, which means the learning approach of SVM focusses on to maximizing the interval [17]. Generally, the accuracy of the classification prediction can be defined as a distance from a point to the hyperplane [18]. SVM aims to make this distance maximum. In real time, linear inseparable examples generally occur. Hence, mapping of data features to a high-dimensional space is essential.

The kernel function is measured at lower dimensions before converting to a higher dimensional space, and a substantial classification effect is defined at this location. This approach may prevent complex calculations once this transformation is performed. In contrast to other kernel functions, lesser number of parameters are required by Gaussian kernel function and it also shows more flexibility [19]. ANN is a nearby replica of the natural sensory system [20]. In this model, a neuron multiplies the inputs by loads, computes the total, and implements a threshold. The outcome of this calculation would then be communicated to ensuing neurons.

Identifying and treating these disorders at an early stage can help you save a lot of money and effort. As a resolution for the problem, a platform is developed that uses deep training to investigate, detect, and manage illness that has impacted a plant. It has been noted that leaves data was shown in a staggered manner, with the conversion of characteristics progressing from a lower to a higher simplifies, correlating to plant categories [21]. The Probabilistic Neural Network (PNN) is taught with a level of accuracy that exceeds 90% [22]. The software system is evaluated for its ability to detect and classify plant leaf infections automatically. Green shaded pixels are detected & masked during the segmented stage based on specific threshold estimates that are handled using Otsu's technique [23].

The developed Neural Network classifiers, which are based on statistics categorization, perform well in fully evaluated kinds of leaf illnesses and can accurately detect and classify those disorders with 93 percent reliability [24]. The learned characteristics are not only restricted to form, surfaces, or coloring but also to specific kinds of leaf characteristics, such as fundamental divides, Leaves tips, leaves bottom, border classifications, etc. thus forth are such onwards are all examples from plant classifications. Deep structure has been successfully trained to accept and understand plant features; far larger samples are necessary, preferably with over a million photos and greater class inconsistency to aid further study by the study industry. Leaf categorization combines a Non-Overlap attributes as Local Binary Pattern (LBP) and Grey Level Co-occurrence Matrices (GLCM) inside the study publication, resulting in improved separation capability and classifying reliability [25].

IV. CONCLUSION

In this comprehensive survey, we have explored the rapidly evolving field of plant leaf disease identification using machine learning-based approaches. The integration of machine learning techniques into agriculture has opened new horizons for early disease detection, sustainable farming, and enhanced food security. Machine learning has proven to be a transformative technology in automating the process of plant disease identification. By analyzing plant images and data, these approaches enable rapid, accurate, and scalable detection, reducing the burden on farmers and agricultural experts. Machine learning-based plant disease identification extends beyond conventional methods. It can be applied to various plant species, enabling the identification of a wide range of diseases. Moreover, it is not limited to a specific environment, making it adaptable to both field and controlled-environment agriculture. The journey of machine learning-based plant disease identification is far from over. Future research should focus on improving model interpretability, increasing the resilience of models to adversarial conditions, and ensuring the equitable deployment of these technologies across diverse agricultural communities.

V. REFERENCES

[1]. University of Nabraska -Lincoln, "Plant Disease: Pathogens and Cycles | CropWatch," University of Nabraska, 2020. https://cropwatch.unl.edu/soybean-management/plant-disease (accessed Nov. 12, 2020).

[2]. S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, and A. Nelson, "The global burden of pathogens and pests on major food crops," Nat. Ecol. Evol., vol. 3, no. 3, pp. 430–439, 2019, doi: 10.1038/s41559-018-0793-y.

[3]. S.Vijayalakshmi and D.Murugan, "An Effective Approach for Diagnosis of Plant Disease using ELM," vol. 4, no. 12, pp. 327–335, 2017.

[4]. A. Kumar, D. Kumar, and S. K. Jarial, "A review on artificial bee colony algorithms and their applications to data clustering," Cybern. Inf. Technol., vol. 17, no. 3, pp. 3–28, 2017, doi: 10.1515/cait-2017-0027.

[5]. Z. Pan, Z. Li, H. Fan, and X. Wu, Feature based local binary pattern for rotation invariant texture classification, vol. 88. Elsevier Ltd, 2017.

[6]. B. Kalantar, B. Pradhan, S. Amir Naghibi, A. Motevalli, and S. Mansor, "Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)," Geomatics, Nat. Hazards Risk, vol. 9, no. 1, pp. 49–69, 2018, doi: 10.1080/19475705.2017.1407368.

[7]. T. H. Saragih, W. F. Mahmudy, A. L. Abadi, and Y. P. Anggodo, "Application of extreme learning machine and modified simulated annealing for jatropha curcas disease identification," Int. J. Adv. Soft Comput. its Appl., vol. 10, no. 2, pp. 108–119, 2018.

[8]. L. M. R. Rere, M. I. Fanany, and A. M. Arymurthy, "Simulated Annealing Algorithm for Deep Learning," Procedia Comput. Sci., vol. 72, pp. 137–144, 2015, doi: 10.1016/j.procs.2015.12.114.

[9]. G. L. Manso, H. Knidel, R. A. Krohling, and J. A. Ventura, "A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust," arXiv, 2019.

[10]. C. Huang, X. Li, and Y. Wen, "AN OTSU image segmentation based on fruitfly optimization algorithm," Alexandria Eng. J., vol. 60, no. 1, pp. 183–188, 2020, doi: 10.1016/j.aej.2020.06.054.

[11]. S. R. Maniyath et al., "Plant disease detection using machine learning," Proc. - 2018 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control. ICDI3C 2018, no. April, pp. 41–45, 2018, doi: 10.1109/ICDI3C.2018.00017.

[12]. Saradhambal.G, Dhivya.R, Latha.S, R.Rajesh, "Plant Disease Detection and Its Solution using Image Classification", 2018, International Journal of Pure and Applied Mathematics, Volume 119 No.

[13]. K. Padmavathi, and K. Thangadurai, "Implementation of RGB and Gray scale images in plant leaves disease detection: comparative study," 2016, Indian Journal of Science and Technology, vol. 9, pp. 1-6

[14]. Simranjeetkaur, Geetanjali Babbar, Gagandeep, "Image Processing and Classification, A Method for Plant Disease Detection", 2019, International Journal of Innovative

Technology and Exploring Engineering (IJITEE), Volume 8, Issue 9S
[15]. Rajneet Kaur, Miss. Manjeet Kaur, "A Brief Review on Plant Disease Detection using in Image Processing", 2017, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.2, pg. 101-106
[16]. Shivani K. Tichkule, Dhanashri. H. Gawali, "Plant diseases detection using image processing techniques", 2016, Online International Conference on Green Engineering and Technologies (IC-GET).
[17]. Vijai Singh, Varsha, A K Misra, "Detection of unhealthy region of plant leaves using image processing and genetic algorithm", 2015, International Conference on Advances in Computer Engineering and Application.
[18]. MrunmayeeDhakate, Ingole A. B., "Diagnosis of pomegranate plant diseases using neural network", 2015, Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)
[19]. Rajat Kanti Sarkar, Ankita Pramanik, "Segmentation of plant disease spots using automatic SRG algorithm: a look up table approach", 2015, International Conference on Advances in Computer Engineering and Applications
[20]. Fatma Marzougui, Mohamed Elleuch, MonjiKherallah, "A Deep CNN Approach for Plant Disease Detection", 2020, 21st International Arab Conference on Information Technology (ACIT)
[21]. Vishnoi, V. K., Kumar, K., & Kumar, B. (2021). Plant disease detection using computational intelligence and image processing. Journal of Plant Diseases and Protection, 128(1), 19-53.
[22]. Hernández, S., & Lopez, J. L. (2020). Uncertainty quantification for plant disease detection using Bayesian deep learning. Applied Soft Computing, 96, 106597.
[23]. Nagaraju, M., & Chawla, P. (2020). Systematic review of deep learning techniques in plant disease detection. International Journal of System Assurance Engineering and Management, 11(3), 547-560.
[24]. Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: a dataset for visual plant disease detection. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD (pp. 249-253).
[25]. Shah, D., Trivedi, V., Sheth, V., Shah, A., & Chauhan, U. (2021). ResTS: Residual deep interpretable architecture for plant disease detection. Information Processing in Agriculture.

Cite this article as :

Nandha Kumar G, Dr. V. Vijayakumar, "Machine Learning-Based Approaches for Plant Leaf Disease Identification : A Survey", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 5, pp.118-124, September-October-2023.