Analysis of Various Artificial Intelligence Methods for Plant Disease Detection: A Comprehensive Review

Shivangi Kumari[1], Sandeep Kumar Singh[2]
[1]M.tech Scholar, Dept. of CSE, Saraswati Higher Education & Technical College of Engineering,(AKTU), Varanasi, India
[2]Assistant Professor, Dept. of CSE, Saraswati Higher Education & Technical College of Engineering,(AKTU), Varanasi, India

Abstract - In India, agriculture plays a crucial role in meeting the food demands of the growing population. Improving crop yields is essential due to this. Crop diseases, caused by bacteria, fungi, and viruses, have a significant impact on low yields. Detecting and preventing these diseases can be achieved through the application of plant disease detection methods. The utilization of machine learning techniques for disease identification in plants is crucial as it effectively utilizes data and provides valuable disease detection methods. Employing machine learning methods for disease identification is beneficial as it relies on data quality for specific tasks. The numerous methods used in detecting plant diseases through artificial intelligence (AI) based machine learning and deep learning techniques have been thoroughly examined in this approach. A comparison is conducted regarding the performance and application of machine learning and deep learning methods in various research papers, demonstrating the superior effectiveness of deep learning models over machine learning models. Deep learning techniques can be employed to identify leaf diseases from captured images, aiming to minimize significant crop losses.

Keywords - Plant Disease, Machine Learning, Deep Learning, Textural Features

I. INTRODUCTION

Among the most important factors affecting food output is plant diseases. They cause a severe decline in the economic production of crops and, in certain instances, hinder this activity. To minimize output losses and preserve agricultural sustainability, illness management and control measures must be properly implemented. This emphasizes the significance of ongoing crop monitoring together with early and accurate disease identification. Additionally, an enormous increase in food production is necessary due to the growing global population [1]. This needs to be done in conjunction with using ecologically friendly farming practices to preserve ecological systems [1]. Since the late 1970s, it has become customary to use computer-based image processing technology in agrarian research and engineering. With the emergence of novel techniques and models in machine vision, machine-learning-based artificial intelligence has attracted a lot of attention. Machine learning approaches have the potential to revolutionise the timely eradication of plant-harming organisms while limiting the need for chemical treatment and other forms of intervention to levels that are both ecologically and economically sensible. Particularly DL-based models have been used extensively in plant disease identification [2]. They are state-of-the-art technologies in this area and have solved the issues with conventional categorization methods. The application of DL has demonstrated considerable potential and effectiveness in a number of domains. It is a sophisticated approach. However, it is a collection of machine learning techniques that makes an effort to represent data at a high level of abstraction by articulating the structures of different alterations.

1.1 Basic Steps in Identification of Diseases from Leaf Images

Data collection and pre-processing are the first steps in a series of activities that are necessary for the accurate detection of plant diseases from the leaves of a plant. The extraction of features comes after pre-processing in the identification of illnesses. Finally, various classifiers are fitted with the features to do the classification [3]. The fundamental steps in identifying plant diseases are shown in Figure 1.

Figure 1. Basic Plant Disease Detection Architecture

All steps involved in plant disease detection based on image processing have been described below:

i. Data Collection: The gathering of picture data is the first stage in identifying plant diseases. Online resources include the PlantVillage dataset, the Cassava dataset, and the Hops dataset, among other common plant diseases datasets. Datasets on diseases in cotton and rice. 38 distinct classes of 14 different species of plants make up the PlantVillage dataset (vegetable and fruits) [4]. The Cassava disease dataset includes photos that were captured in the field in real time for five different disease classifications. Cassava mosaic disease, cassava bacterium blight, cassava brown streak disease, cassava green mite, and cassava healthy are among the illnesses included in the dataset for cassava. The Hops dataset includes five distinct illness classifications with irregular background circumstances. Downy, powdery, healthy, nutritional, and insect diseases are among the illnesses. The healthy and diseased cotton leaves and plants make up the Cotton dataset. Four distinct groups of diseases that were seen under field conditions make up the dataset on rice plant. The diseases bacterial blight, blast, brown spot, and tungro are included in the dataset for rice plant [5].

ii. Image Pre-processing: In this process, the photographs' visual appeal is improved. Prior to the computational processing of photographs, the image data is improved by the image pre-processing technique. Various pre-processing techniques are taken into consideration while removing noise from images or other objects. In image clipping, the leaf image is clipped to acquire the desired area of the picture [6]. The smoothing filter is used to smooth images. Image enhancement is used to increase the contrast. During this phase, RGB images are converted into grayscale images.

iii. Image Segmentation: The foundation of image analysis is image segmentation, which is the process of breaking a picture into multiple pieces containing features and extracting areas of interest. The results of the subsequent image processing are directly influenced by the type of picture segmentation used.

- Otsu's segmentation**:** A widely used binarization threshold approach for image segmentation is the OTSU technique (OTSU). According to the threshold selection principle, this technique takes into account the backdrop and the target images biggest interclass difference [7]. The OTSU technique's rule specifies that the biggest between-class variance approach is also its alias. According to the characteristics of the grey scale, it divides the image into foreground and background. The discrepancy between the two portions is greatest when the best threshold is taken into account.
- Region Based Segmentation: The iterative process used by region-based segmentation algorithms involves grouping nearby pixels with similar characteristics and separating groups of pixels with divergent values.

iv. Feature Extraction: Following segmentation, the unhealthy component is successfully removed. In the following stage, the crucial traits are extracted. These attributes are used to confirm a sample's meaning [8]. The characteristics of an image are its colour, form, and texture. The texture of plant leaves is currently the most important factor that researchers frequently take into account when categorizing plant diseases. Using textural traits, pathogens are divided into a number of categories.

- FAST algorithm: Using this method, the interest points in a picture are confirmed. A pixel in an image that has precisely defined placements is the interest point. It may be discovered. Every interest point has informational content, which can be repeated across multiple images [9]. The main goal of the FAST method is to discover corner points because many real-time frames rate applications use them.
- SURF algorithm: To register and identify the image using the SURF method, the key points are collected from the database as well as from the test images. As a local object detector, it operates.

v. Image Classification: The method that involves taking an input, for instance a picture, and producing a class or a probability that the input belongs to a particular class is known as image classification. The major goal of this stage in the identification of plant diseases is to categorise the input plant image as infectious or healthy [10]. If an image is determined to be infectious, then certain methods on the market further categorise it into various infections. Over the years, several academics have developed a variety of classifiers for the classification of photographs. These classifiers include CNN, SVM, ANN, CNN, Random Forest (RF), and support vector machines (SVM). Several of these algorithms are covered below:

- Random Forest: Based on the distinctive qualities and classification outcomes of a given dataset, Random Forest is a great supervised learning method that can train a model to predict which classification results in a particular sample type belong to [11]. The decision tree-based Random Forest uses the Bagging (Bootstrap aggregating) approach to generate various training sample sets. The best attribute from a group of attributes chosen at random is chosen by the random subspace partition approach to divide internal nodes. The voting method is utilized to categorize the input samples, and the many decision trees created are used as weak classifiers. Multiple weak classifiers combine to construct a robust classifier [12]. A fresh set of data is input into a random forest, which has already constructed a huge number of decision trees in accordance with a certain random rule. Each of these decision trees then independently predicts the new set of samples, integrating their individual predictions to produce the final result.
- Support Vector Machine: Support Vector Machine (SVM) was first proposed by Coretes and Vapink in 1995. It exhibits numerous special benefits in the recognition of small sample, nonlinear, and high dimensional patterns and can be used to other tasks including function fitting machine learning issues [13]. Prior to the emergence of deep learning, SVM was regarded as the most effective and productive machine learning technique in recent years. The Vapnik Chervonenkis (VC) dimension theory of statistical learning theory and the idea of structural risk minimization serve as the foundation for the SVM

method. Its main goal is to identify a separation hyperplane between several categories so that they can be more effectively separated. According to the SVM method, the hyperplane can be identified by selecting only the sample point that is closest to the hyperplane and finding the support vector [14].

- Artificial Neural Network: A common ML modelling approach that draws inspiration from the human central nervous system is an ANN. Input layer, hidden layer(s), and output layer are the three layers that commonly make up an ANN model. In order to communicate and derive information, the neurons in various levels are interconnected. The data gathered from the real world are initially supplied into the input layer, which is how a conventional three-layer ANN computes [15]. The net input is then calculated in the hidden layer by adding x multiplications with the appropriate weights and activation functions. The output layer, where the outputs are weighted and then summed in the output neuron, is connected to the hidden neurons last. As a result, the following is given as an ANN's mathematically [10]:

$$y = f(x) = \sum_{j}^{J} \left(w_j \, \delta\left(w_{ij} x_i + b_j \right) + \beta + \varepsilon \right.$$

where w_ij be the weight factor that links the ith input neuron to the jth hidden neuron, x be an I-dimensional input vector, w_j be the weight factor that links the jth hidden neuron to the output neuron. The jth hidden neuron has the bias b_j and random error is ε [16].

II. LITERATURE REVIEW

J. Chen, et.al (2022) suggested a reliable, light-weight network architecture called MobInc-Net to carry out the agricultural pathogen classification and diagnosis [17]. The Inception module was improved in this research by using depth- and point-wise convolutions in place of the original convolutions. The modified Inception (M-Inception) module was then combined with the pre-trained MobileNet as the foundation extractor to retrieve high-contrast features of photos. Following that, the SSD block and the fully linked Softmax layer with the real number of groups were independently inserted behind the core network for categorizing and identifying different sorts of pests and pathogens. The model training process used two-step transfer learning to build a successful system. Using an average detection performance of 99.21% on the real - world dataset and 97.89% on the localized dataset, testing results demonstrate that the suggested strategy could deliver the expected results.

Y. Wu, et.al (2022) discovered a technique for precisely classifying diseases based on attention networks. The "Classification Model" employed an attention technique to enhance identification skills. During training, the "Reconstruction-Generation Model" was added [18], and the "Classification Model" had to focus more on identifying distinctions within specific geographic regions as opposed to focusing more on general attributes. Because "Reconstruction-Generation Model" and "Discrimination Model" were utilised only during training and didn't take part in inference phase operations, the model's intricacy could indeed be increased. The method of generalization ability improvement significantly improves the identification accuracy when compared to the conventional classification network. Additionally, the technique may identify leaf diseases in actual environments while requiring minimal performance and less memory.

Y. Zhao, et.al (2022) projected a DoubleGAN (Double Generative Adversarial Network) for creating images of diseased plant leaves so that these kinds of datasets were balanced [19]. Initially, the normal and infected leaves were utilized for input. WGAN (Wasserstein generative adversarial network) algorithm deployed the normal images for generating a pre-trained framework. Subsequently, the pre-trained model made the implementation of infected leaves for obtaining images of 64*64 pixel. The images of256*256 pixel were generated through SRGAN (Superresolution Generative Adversarial Network) for extending the unbalanced dataset. Eventually, a comparison was done amid these images and images retrieved with DCGAN (Deep convolution generative adversarial network). The projected algorithm offered accuracy of 99.80% to detect the plant species and 99.53% for recognizing the disorders occurred on plants.

F. Li, et.al (2021) established an airborne edge-computing and LDL (Lightweight Deep Learning) based system in order to diagnose PWD (Pine Wilt Disease) via imagery sensors [20]. Afterward, computing potential of edge computing was exploited for filtering these redundant images. For this, a lightweight enhanced YOLOv4-Tiny (You Only Look Once-version 4) based method was put forward to detect the disease quickly at lower missing rate. In the end, the real workstation employed rest of the images to detect the images. The experimental results revealed that the established system diagnosed the diseases quickly and performed better in contrast to other methods for detecting the diseased pine trees.

A. Abbas, et.al (2021) recommended a DL (Deep Learning)-based technique in order to diagnose the tomato disease through C-GAN (Conditional Generative Adversarial Network) for producing the synthetic images of leaves of tomato plant [21]. Later, a training of DenseNet121 algorithm was done on synthetic and real images for classifying the tomato leaves into 10 classes. The method of augmenting the data helped in enhancing the network generalized ability and preventing the over-fitting issue. PlantVillage dataset executed to train and test the recommended technique. The recommended technique yielded the accuracy of 0.9951% to classify 5 classes, 0.9865 for 7 classes, and 0.9711 for 10 classes.

X. Liu, et.al (2021) focused on investigating the issue related to recognizing the visual plant disease for detecting the diseases of plants [22]. A novel dataset was constructed in which 271 plant disease classes comprised for recognizing the disorders of plants. The initial task was executed for computing the weights of all the divided patches from every image on the basis of distributing the cluster of these patches so that the discriminative level of every patch was illustrated. Thereafter, the weight was assigned to every loss for every patch-label pair. At last, the patch attributes were extracted from the network trained with loss reweighting. LSTM (Long Short-Term Memory) algorithm was presented and implemented for encoding the weighed patch feature sequence into a detailed feature representation. The results indicated that the presented technique was effective.

J. Chen, et.al (2020) formulated a new method recognized as GMDH-Logistic technique to diagnose and classify the disorders of plant leaves automatically [23]. The IPT (image process technique) were executed to analyze the feature engineering and for constructing the index system for the predictive models. After that, the formulated method made the deployment of chosen attributes and the verified variables were interpreted for dealing with the limitations of other algorithm. the complicated background scenarios considered to compute the formulated method. The experimental outcomes validated that the formulated method performed effectively for recognizing the plant images as normal and healthy.

G. Yang, et.al (2020) suggested a new framework known as LFC-Net in which three networks: LN (Location network), FN (Feedback network) and CN (Classification network) were deployed [24]. At first, the initial network assisted in detecting the informative areas in the tomato image. The second model was utilized for optimizing the iterations. At last, detected areas were fed in the CN for classifying the full image of the tomato. The suggested framework offered an accuracy of 99.7% on ImageNet dataset. The results indicated the adaptability of the suggested framework on other vegetable and fruit datasets and it was capable of preventing and controlling the infections of tomato leaves.

X. Nie, et.al (2019) developed an approach depending upon Faster R-CNN (Region based Convolutional Neural Network) and MTL (multi-task learning) algorithm for diagnosing SVW (strawberry verticillium wilt) [25]. Afterward, SVWDN (strawberry verticillium wilt detection network) algorithm was introduced in which attention system was implemented for extracting the attributes. This algorithm assisted in diagnosing SVW in accordance with the symptoms of detected plant elements. Moreover, the petioles and young leaves were classified using the introduced algorithm to determine that the plant was infected or not. Around 3, 531 images of four classes were employed to create a dataset so that the introduced algorithm was computed and tested. The developed approach offered mAP (mean average precision) of 77.54% and accuracy of 99.95% to diagnose STW.

Table 1. Comparative analysis of Existing plant disease detection algorithms

Author	Year	Technique Used	Results	Limitations
J. Chen, et.al	2022	MobInc-Net	Using an average detection performance of 99.21% on the real - world dataset and 97.89% on the localized dataset, testing results demonstrate that the suggested strategy could deliver the expected results.	The important characteristics were not extracted in this research, and the information about the features was largely redundant, wasting processing power and recognition time.
Y. Wu, et.al	2022	A technique based on attention networks.	The method of generalization ability improvement significantly improves the identification accuracy when compared to the conventional classification network.	The created model cannot be immediately trained in this manner if there is a lack of substantial amounts of annotated data to use as training samples; instead, the number of samples must be increased.

Y. Zhao, et.al	2022	DoubleGAN (Double Generative Adversarial Network)	The projected algorithm offered accuracy of 99.80% to detect the plant species and 99.53% for recognizing the disorders occurred on plants.	This algorithm was deployed on large number of samples to create the images of higher resolution.
F. Li, et.al	2021	edge-computing and LDL (Lightweight Deep Learning) based system	The experimental results revealed that the established system diagnosed the diseases quickly and performed better in contrast to other methods for detecting the diseased pine trees.	This system was not able to diagnose diverse stages of plant infections.
A. Abbas, et.al	2021	a DL (Deep Learning)-based technique	The recommended technique yielded the accuracy of 0.9951% to classify 5 classes, 0.9865 for 7 classes, and 0.9711 for 10 classes.	This technique was not detected other portions of plants such as fruits, stems, and branches, and distinct stages of plant infection.
X. Liu, et.al	2021	LSTM (Long Short-Term Memory) algorithm	The results indicated that the presented technique was effective.	This algorithm offered lower sensitivity if the images captured in harsh environments.
J. Chen, et.al	2020	GMDH-Logistic technique	The experimental outcomes validated that the formulated method performed effectively for recognizing the plant images as normal and healthy.	The major issue related to improvising the efficacy of diagnosing the plant diseases.
G. Yang, et.al	2020	LFC-Net	The suggested framework offered an accuracy of 99.7% on ImageNet dataset. The results indicated the adaptability of the suggested framework on other vegetable and fruit datasets and it was capable of preventing and controlling the infections of tomato leaves.	This technique was not recognized all the categories of tomato diseases and their different stages.
X. Nie, et.al	2019	Faster R-CNN (Region based Convolutional Neural Network) and MTL (multi-task learning) algorithm	The developed approach offered mAP (mean average precision) of 77.54% and accuracy of 99.95% to diagnose STW.	This technique only determined whether the plant had SVW or not. It was not detected other diseases.

III. CONCLUSION

Extensive research has been conducted on different types of machine and deep learning methods for identifying and categorizing plant diseases. Following this, alternative classification techniques in machine learning may be utilized to detect diseases in plants and assist farmers in automatically detecting various crop diseases. This analysis explores different approaches to using deep learning for detecting plant diseases. Moreover, various methods and mappings were outlined for the detection of disease symptoms. The advancements in deep learning technologies for identifying plant leaf diseases have been significant in recent years. Our expectation is that this research will serve as a valuable resource for scientists studying plant disease detection. In addition, a comparative analysis has been conducted between machine learning and deep learning techniques. Despite the considerable progress observed in recent years, there are still some research gaps that need to be addressed in order to implement effective techniques for detecting plant diseases.

IV. REFERENCES

[1] K. P. Shreyas and G. V. S. Narayana, "Bell Pepper Crop Disease Detection Using Convolution Neural Network," 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), 2022, pp. 1-6

[2] L. S. Puspha Annabel, T. Annapoorani and P. Deepalakshmi, "Machine Learning for Plant Leaf Disease Detection and Classification – A Review," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0538-0542

[3] N. N. Moon, M. S. Sharmin, R. A. Hossain, I. Jahan, F. N. Nur and M. M. Rahman, "Deep Learning Model for Detecting and Diagnosing Plant Disease," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), 2021, pp. 1-8

[4] P. B R, A. Ashok and S. H. A V, "Plant Disease Detection and Classification Using Deep Learning Model," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 1285-1291

[5] R. Kundu, U. Chauhan and S. P. S. Chauhan, "Plant Leaf Disease Detection using Image Processing," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), 2022, pp. 393-396

[6] P. Chaitanya Reddy, R. M. S. Chandra, P. Vadiraj, M. Ayyappa Reddy, T. R. Mahesh and G. Sindhu Madhuri, "Detection of Plant Leaf-based Diseases Using Machine Learning Approach," 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2021, pp. 1-4

[7] Y. Liu, G. Gao and Z. Zhang, "Crop Disease Recognition Based on Modified Light-Weight CNN with Attention Mechanism," in IEEE Access, vol. 10, pp. 112066-112075, 2022

[8] U. P. Singh, S. S. Chouhan, S. Jain and S. Jain, "Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease," in IEEE Access, vol. 7, pp. 43721-43729, 2019

[9] R. Dwivedi, T. Dutta and Y. -C. Hu, "A Leaf Disease Detection Mechanism Based on L1-Norm Minimization Extreme Learning Machine," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 8019905

[10] M. Lv, G. Zhou, M. He, A. Chen, W. Zhang and Y. Hu, "Maize Leaf Disease Identification Based on Feature Enhancement and DMS-Robust Alexnet," in IEEE Access, vol. 8, pp. 57952-57966, 2020

[11] M. A. Khan et al., "An Optimized Method for Segmentation and Classification of Apple Diseases Based on Strong Correlation and Genetic Algorithm Based Feature Selection," in IEEE Access, vol. 7, pp. 46261-46277, 2019

[12] B. Liu, C. Tan, S. Li, J. He and H. Wang, "A Data Augmentation Method Based on Generative Adversarial Networks for Grape Leaf Disease Identification," in IEEE Access, vol. 8, pp. 102188-102198, 2020

[13] C. Zhou, S. Zhou, J. Xing and J. Song, "Tomato Leaf Disease Identification by Restructured Deep Residual Dense Network," in IEEE Access, vol. 9, pp. 28822-28831, 2021

[14] P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network", Artificial Intelligence in Agriculture, vol. 32, no. 4, pp. 1703-1712, 11 May 2021

[15] M. H. Saleem, J. Potgieter and K. M. Arif, "A Performance-Optimized Deep Learning-Based Plant Disease Detection Approach for Horticultural Crops of New Zealand," in IEEE Access, vol. 10, pp. 89798-89822, 2022

[16] M. Ahmad, M. Abdullah, H. Moon and D. Han, "Plant Disease Detection in Imbalanced Datasets Using Efficient Convolutional Neural Networks With Stepwise Transfer Learning," in IEEE Access, vol. 9, pp. 140565-140580, 2021

[17] J. Chen, W. Chen, A. Zeb, S. Yang and D. Zhang, "Lightweight Inception Networks for the Recognition and

Detection of Rice Plant Diseases," in IEEE Sensors Journal, vol. 22, no. 14, pp. 14628-14638, 15 July15, 2022

[18] Y. Wu, X. Feng and G. Chen, "Plant Leaf Diseases Fine-Grained Categorization Using Convolutional Neural Networks," in IEEE Access, vol. 10, pp. 41087-41096, 2022

[19] Y. Zhao et al., "Plant Disease Detection Using Generated Leaves Based on DoubleGAN," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 3, pp. 1817-1826, 1 May-June 2022

[20] F. Li et al., "A Remote Sensing and Airborne Edge-Computing Based Detection System for Pine Wilt Disease," in IEEE Access, vol. 9, pp. 66346-66360, 2021

[21] A. Abbas, S. Jain and S. Vankudothu, "Tomato plant disease detection using transfer learning with C-GAN synthetic images", Computers and Electronics in Agriculture, vol. 6, no. 4, pp. 73191-73199, 29 June 2021

[22] X. Liu, W. Min, S. Mei, L. Wang and S. Jiang, "Plant Disease Recognition: A Large-Scale Benchmark Dataset and a Visual Region and Loss Reweighting Approach," in IEEE Transactions on Image Processing, vol. 30, pp. 2003-2015, 2021

[23] J. Chen, H. Yin and D. Zhang, "A self-adaptive classification method for plant disease detection using GMDH-Logistic model", Sustainable Computing: Informatics and Systems, vol. 28, no. 10, pp. 2325-2332, 26 June 2020

[24] G. Yang, G. Chen, Y. He, Z. Yan, Y. Guo and J. Ding, "Self-Supervised Collaborative Multi-Network for Fine-Grained Visual Categorization of Tomato Diseases," in IEEE Access, vol. 8, pp. 211912-211923, 2020

[25] X. Nie, L. Wang, H. Ding and M. Xu, "Strawberry Verticillium Wilt Detection Network Based on Multi-Task Learning and Attention," in IEEE Access, vol. 7, pp. 170003-170011, 2019