Machine Learning and Deep Learning Techniques for Crop Disease Detection: A Comprehensive Review

[1]Shakiv Pandit, [2]Rajeev Sharma
Scholar, Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]Chandigarh Engineering College, Landran (Mohali), Punjab, India

Abstract: Major challenge in Indian agriculture is that majority of the crops are prone to various diseases throughout its growth. When diagnosing these diseases manually, farmers struggle to make accurate diagnoses due to a lack of training and experience. To ensure that different crops develop normally and healthily, it is essential to identify diseases early and treat affected plants. In the agricultural fields of today, identifying leaf diseases is crucial. This article explains several ML and DL methods that are applied to different crop disease detection. The agriculture industry is in dire need of a computerized framework that can identify one or more crop disease problems. After that, it explored the different method analyses with objectives, advantages, and disadvantages in crop disease detection. The Convolutional Neural Network (CNN) method has attained high accuracy as compared with the existing methods based on ML and DL techniques.

Keywords: *Crop Disease Detection, Machine Learning (ML), Deep Learning (DL), Agriculture.*

I. INTRODUCTION

High infant mortality caused the population to grow slowly until 1700. There will need to be a major increase in agricultural production as the world's population is predicted to reach 7 billion by 2050, 9.7 billion by 2050, and 10.9 billion by 2100.

Crop yield production needs to double by 2050, which calls for a 2.4% yearly increase in key yields. Current yields, however, are only 1.3% annually. Because traditional agricultural production is unsustainable and causes detrimental ecosystem effects like biodiversity loss and increased greenhouse gas emissions, high-yield crops require resource optimization. Furthermore, diseases and insect pests pose a constant threat to crop productivity.

The global economy suffers losses of 20% to 40% because of vegetable diseases and insect attacks. They have a substantial impact on crop production and cost $220 and $70 billion, respectively. These losses vary in magnitude throughout the world and are frequently brought on by transboundary plant diseases and pests. Compared to other parts of the world, crop pests and diseases spread more widely in North America between 1950 and 2000. Climate change-induced increases in global temperatures have an impact on development and pest damage. An increase in temperature causes insects' metabolic rates to rise, which encourages them to eat more and cause more harm. Temperature also affects how quickly some insect species grow. It is estimated that for every degree of average global warming, agricultural losses will rise by 10% to 25% globally.

Farmers are using more and more pesticides and chemical treatments to protect their crops, which harms soil and groundwater and has detrimental health effects. Additionally, this raises the possibility that pests will become resistant to pesticides. Conventional plant disease detection techniques depend on graphical review, which is labor-intensive and impractical for large farms. To get around these restrictions and cut down on the overuse of chemicals and pesticides, which will lower production costs and environmental harm, automated techniques for crop monitoring and forecasting are required [1]. Many nations rely heavily on agriculture as the demand for food rises due to population growth.

Crops are particularly vulnerable to bacterial, fungal, and viral diseases that can drastically lower productivity by 10% to 95%. In order to prevent losses and minimize the use of pesticides, which can be detrimental to both humans as well as the environment, early disease detection is crucial. Small farms and farmers in developing nations frequently identify diseases with the naked eye, which calls for plant pathology knowledge and lengthy treatment times. Accurate diagnosis of rare diseases can result in higher treatment costs, so farmers frequently seek expert advice. For large farms, this approach is impractical, and biased decisions could result in inaccurate forecasts. Early disease detection is essential for averting large losses and protecting the environment and public health. The Indian agriculture industry needs major improvements to adjust to the shifting economic landscape. A significant element influencing the caliber and volume of agricultural output is crop disease. A variety of pesticides are available to control diseases, but it can be difficult to identify the most up-to-date and effective pesticide without professional guidance, which can be costly and time-consuming. For periodic crop monitoring, a highly technical approach is therefore required [2].

Machine Learning (ML) methods are applied in several domains, such as physics, self-driving cars, medical image detection, and image classification. Applications of ML for agriculture are still in their infancy, but they have promise. Long-short-term memory (LSTM) networks predict pests, Convolutional Neural Network (CNN) classifies diseases, and object segmentation and DL methods detect insects on leaves [3]. Learning from training data is a process used in ML approaches to accomplish tasks. Example data in machine learning is characterized by attributes or variables. These characteristics may be ordinal, numeric, binary, or nominal. An improved performance metric is used to gauge an ML model's effectiveness over time. ML models and algorithms are evaluated using a variety of mathematical and statistical models. Following the learning process, new examples can be categorized, predicted, or clustered using the gained experience by the trained model. With this method, ML models and algorithms are shown to be continuously improved over time. ML tasks can be classified as either supervised or unsupervised learning, depending on the learning system's learning signal. To build a general rule by using a supervised process that maps inputs to outputs by presenting data with example inputs and matching outputs. In a dynamic setting, inputs can occasionally be given as feedback or only partially available. Whereas unsupervised learning does not distinguish between training and test sets, and the learner analyses input data to outputs find hidden patterns, supervised learning uses a trained model to predict missing test data [4].

This paper is designed as: Sec 2 represents a detailed analysis of various methods. Sec 3 describes various ML and deep learning (DL) methods and Sec 4 analyzes various DL and Ml method's advantages and drawbacks. Sec 5 describes the conclusion and further enhancement of the crop disease detection system.

II. RELATED WORK

Several articles were analyzed with different research methodologies with outcomes. **Mohit Agarwal et al. (2020) [5]** used a CNN model with eight hidden layers. It performs better in identifying tomato crop diseases than pre-trained models and conventional machine-learning techniques. Utilizing the publicly accessible PlantVillage dataset, the model achieves a better accuracy rate, surpassing both pre-trained models and conventional techniques. The model, with a maximum accuracy of 93.5%, outperforms conventional techniques and incorporates 10 tomato disease classes. After image augmentation, image pre-processing was used to improve CNN performance. Additionally, the model exhibits remarkable performance on other datasets. **Zubair Saeed et al. (2022) [6]** described agricultural yield depends on the health of plants, and diseases. Disease detection requires intelligent techniques, and a reliable strategy utilizing current CNN models was put forth. For disease detection in vital crops like rice and corn, the technique makes use of ResNet-152 and Inception-v3 variants. For corn crops, the method's accuracy using the InceptionV3 and ResNet152. The ResNet152 variant of the method achieves 82.20% accuracy in the classification of rice disease images. The results of the experiment demonstrate that disease detection for several crops was robust. **Anuradha Badage et al. (2018) [7]** suggested system addresses the difficulties farmers encounter in controlling crop diseases brought on by pests, insects, and pathogens by automating the detection of crop diseases using remote sensing images. This allows farmers to take the appropriate action. By training both healthy and diseased data sets, the suggested system seeks to identify diseases early on in the outer layer of leaves. It functions in two stages: training healthy datasets was the first stage, and crop monitoring and disease detection using Canny's edge detection algorithm was the second stage. **Weihui Zeng et al. (2020) [8]** described the Self-Attention CNN (SACNN) to increase the precision of image recognition. The self-attention network was developed to extract local lesion features and a simple network for global features showed 98.0% and 95.33% recognition accuracy, respectively. By concentrating on key areas of the developed model to increase the accuracy of recognition. The study also investigates how network number, channel size setting, and location selection affect recognition performance.**Yong Zhong et al. (2020) [9]** described apple leaf diseases as having a major effect on apple production and resulting in large financial losses each year. Three techniques were utilized to identify these illnesses. 2462 photos of six diseases were used to assess the approach. The suggested techniques demonstrated higher accuracy than the conventional multi-classification method, with a 92.29% accuracy rate. The significance of researching apple leaf diseases for accurate disease detection was highlighted by this study. **Varsha P. Gaikwad et al. (2023) [10]** described agricultural economies were greatly impacted by crop leaf diseases, but manual detection was ineffective and time-consuming. Scientists created automated disease detection techniques. Using a deep KNN classifier optimized via Cetalatran optimization, this study suggests a novel method. Early and effective crop disease detection was made possible by cetalatran optimization, which rapidly finds pertinent features for disease prediction. Accuracy, sensitivity, and specificity metrics were all improved by the Cetalatran-optimized deep KNN, yielding better results of 91.879%. Table I defines the analysis of numerous existing methods based on crop disease detection that includes the proposed method, problem, parameters, and dataset/tools.

TABLE I
ANALYSIS OF VARIOUS EXISTING CROP DISEASE METHODS

Author Name	Implemented Methods	Issues/Problems	Tool/Dataset	Metrics
Mohit Agarwal et al. (2020) [5]	CNN KNN	Inaccurate performance	PlantVillage dataset	Accuracy
Zubair Saeed et al. (2022) [6]	ResNet152 CNN	incorrect classification	Kaggle dataset	Accuracy
Anuradha Badage et al. (2018) [7]	Canny's edge detection algorithm	Limited performance	Image dataset	Accuracy
Weihui Zeng et al. (2020) [8]	Self-Attention CNN	Complex process due to large dataset	MK-D2 and AES-CD9214 dataset	Accuracy
Yong Zhong et al. (2020) [9]	Multi-label classification Regression	Disappearance of gradient	Apple Leaf image dataset	Precision Sensitivity F1 score
Varsha P. Gaikwad et al. (2023) [10]	Dolphin echolocation KNN Coyote optimization method	incorrect classification	MATLAB	Accuracy Sensitivity specificity

III. DIFFERENT DETECTION METHODS USED IN CROP DISEASE DETECTION SYSTEM

This section describes various machine and deep learning methods used in the crop disease detection system. Several methods are detailed and discussed below:

3.1 VGG-16: It has a uniform design and 16 convolutional layers and is frequently used for disease recognition in image classification. Mostly, it shows that system performance can be enhanced by increasing network intensity. The network's spatial dimension was reduced by using a pooling layer and related computations, the convolution layer processes images by applying filters and extracting information with kernel and stride size properties.

3.2 VGG-19: To classify images into 1000 categories, which were initially presented. This model has three dense and sixteen convolution layers.

3.3 RestNet50: It is a subdivision of the ResNet and it consists of MaxPool, a single average pool layer, and forty-eight convolutional layers. Using it to categorize images is very common. The four primary stages of the model are as follows: a triple-replicated first convolution stage with three layers, followed by a 1*1,128 kernel, 3*3,128, and 1*1,512. Four iterations of 12 layers were used, then a 1*1,256 kernel [11].

3.4 CNN: DL is successfully applied to beneficial and informative tasks by CNN, a potent image-processing technique [13][16].

3.5 Recurrent Neural Newtork (RNN): Artificial neural networks (ANN) known as RNN use connections between nodes to create a directed graph. They can learn patterns and relationships over time because they are made to handle sequential data by allowing information to persist. For tasks like language modeling and speech recognition, this architecture is perfect. RNNs can process input sequences by storing and feeding back the output of a layer because they have internal state memory. This feedback loop aids in text generation or sentence prediction [12][13].

3.6 SVM: A machine called a Support Vector Machine (SVM) classifies different classes by drawing a decision boundary. It constructs the hyperplane using extreme points or support vectors, enabling the placement of new data inputs in appropriate classes. In recent years, it has been very useful for disease classification with greater and accuracy fewer errors. Fig 1 represents a block diagram of the SVM method.

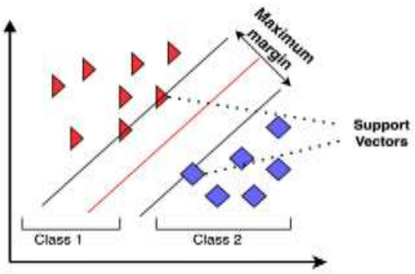

Fig. 1 Linear SVM [13][15]

3.7 KNN: A supervised learning technique called the K-NN algorithm uses the similarity or proximity between new and old values to determine which class to assign to new unknown values. By measuring the separation between two points on a graph, this parameter similarity is ascertained. The distance between each training data point and the new point is calculated using distance matrix techniques, based on the chosen K value. The KNN class is selected based on distance, and the new data point is assigned to the class with the most K neighbors [13] [14].

3.8 M-KNN: The M-KNN technique determines the k parameter in a KNN by weighing all test data. It was put into practice to detect illnesses in soybean plants [14].

3.9 DT: The process of converting vast volumes of data into a decision tree that represents rules in a structured database language like SQL is known as decision tree classification. This approach categorizes and displays attribute relationships, exposing variables that affect the results of alternative decisions. The decision tree enables decision-makers to understand problem solutions by decomposing intricate decision-making procedures into simpler ones. Transforming data into decision-making represented by a tree and rule is the idea behind decision trees [14].

IV. ADVANTAGES AND DISADVANTAGES OF ML AND DL METHODS

Table II describes the analysis of numerous existing approaches based on Different Ml and DL methods with working, aim, advantages, and disadvantages.

TABLE II
ANALYSIS OF VARIOUS ML AND DL METHODS BASED OF DIFFERENT ASPECTS [11-17]

Method's Name	Working	Aim	Advantages	Disadvantages
VGG-16 [11]	It is used for disease detection in image classification.	To demonstrate some issues and improve the intensity as well as performance.	It is widely used for image recognition as well as classification.	It is a huge network and requires more time for processing.
VGG19 [11]	It is also used for disease detection in image classification.	It slightly improves the performance.	It has robust feature extraction capabilities.	It requires a large number of learner parameters.
ResNet50 [11]	This method easily optimizes and reduces the vanishing gradient issues.	To improve the performance by increasing the no. of layers.	It efficiently improves performance and provides adaptability features.	It has a higher failure rate and more time-consuming method.
M-KNN [14]	To analyze computational experimental analysis.	This method provides better validation values for each training data.	Easy to implement and understandable.	It requires more memory and higher computational cost.
KNN [14]	This method is used for testing purposes.	The number of neighbors attracted to the closest one, provides a good accuracy value.	Simple to use and intelligible.	It costs more to compute and uses more memory.
DT [14]	The testing process is done by using 31 testing data.	This method provides better validation values for each training data.	This is a very simple method as compared to another method.	This method obtained low performance as compared to KNN and M-KNN methods.
SVM [15]	This method identifies 19 classes of diseases.	This method improves the classification accuracy.	It is highly accurate and handles many features.	It has a slow speed and requires more time for processing.
CNN [16]	This method simply detects crop disease.	To improve the performance as compared to other methods.	It provides higher performance and it deals with image data.	It requires a high amount of data and more computational cost.
CNN-RNN [17]	This hybrid model detects the disease at an early stage.	This method provides better validation values for each training data.	It requires limited time and reduced computational cost.	It has a slow speed and requires more time for processing.

V. CONCLUSION

The analysis article concluded that India's cultivated crops face numerous diseases, making manual diagnosis difficult due to lack of training and experience. Early detection and treatment are crucial for crop health, and these advanced technologies are very useful for farmers to make timely decisions to prevent disease spread and minimize crop loss. This article provides an outline of different ML and DL methods, their advantages, disadvantages, and operational methods, aiming to improve crop disease detection efficiency in India. This article examines deep learning and machine learning algorithms for detecting and categorizing crop product diseases, such as those affecting fruits, vegetables, and plants. It discusses the classification of diseases in several crops, such as tomato, cucumber, cotton, corn, rice, apples, grapes, potatoes, and rice etc. This review article helps researchers improve the performance and identify multiple ways to increase the accuracy based on many crop disease classification systems.

VI. REFERENCES

[1]. Domingues, T., Brandão, T., & Ferreira, J. C. (2022). Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. *Agriculture*, *12*(9), 1350.
[2]. Badage, A. (2018). Crop disease detection using machine learning: Indian agriculture. *Int. Res. J. Eng. Technol*, *5*(9), 866-869.
[3]. Orchi, H., Sadik, M., & Khaldoun, M. (2021). On using artificial intelligence and the internet of things for crop disease detection: A contemporary survey. *Agriculture*, *12*(1), 9.
[4]. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

[5]. Agarwal, M., Gupta, S. K., & Biswas, K. K. (2020). Development of Efficient CNN model for Tomato crop disease identification. *Sustainable Computing: Informatics and Systems*, *28*, 100407.

[6]. Saeed, Z., Raza, A., Qureshi, A. H., & Yousaf, M. H. (2021, October). A multi-crop disease detection and classification approach using cnn. In *2021 International Conference on Robotics and Automation in Industry (ICRAI)* (pp. 1-6). IEEE.

[7]. Badage, A. (2018). Crop disease detection using machine learning: Indian agriculture. *Int. Res. J. Eng. Technol*, *5*(9), 866-869.

[8]. Zeng, W., & Li, M. (2020). Crop leaf disease recognition based on Self-Attention convolutional neural network. *Computers and Electronics in Agriculture*, *172*, 105341.

[9]. Zhong, Y., & Zhao, M. (2020). Research on deep learning in apple leaf disease recognition. *Computers and electronics in agriculture*, *168*, 105146.

[10]. Gaikwad, V. P., & Musande, V. (2023). Advanced prediction of crop diseases using cetalatran-optimized deep KNN in multispectral imaging. *Traitement Du Signal*, *40*(3), 1093.

[11]. Islam, M. M., Adil, M. A. A., Talukder, M. A., Ahamed, M. K. U., Uddin, M. A., Hasan, M. K., ... & Debnath, S. K. (2023). DeepCrop: Deep learning-based crop disease prediction with web application. *Journal of Agriculture and Food Research*, *14*, 100764.

[12]. Iniyan, S., Jebakumar, R., Mangalraj, P., Mohit, M., & Nanda, A. (2020). Plant disease identification and detection using support vector machines and artificial neural networks. In *Artificial intelligence and evolutionary computations in engineering systems* (pp. 15-27). Singapore: Springer Singapore.

[13]. Goel, L., & Nagpal, J. (2023). A systematic review of recent machine learning techniques for plant disease identification and classification. *IETE Technical Review*, *40*(3), 423-439.

[14]. Saragih, T. H., Fajri, D. M. N., & Rakhmandasari, A. (2020). Comparative study of decision tree, k-nearest neighbor, and modified k-nearest neighbor on jatropha curcas plant disease identification. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, *5*(1), 55-60.

[15]. Thaiyalnayaki, K., & Joseph, C. (2021). Classification of plant disease using SVM and deep learning. *Materials Today: Proceedings*, *47*, 468-470.

[16]. Shobana, M., Vaishnavi S., Gokul Prasad C., Pranava K. S. P., Madhumitha K. P & Nitheesh C., (2022). Plant Disease Detection Using Deep Neural Network. In *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* (pp. 1-6). IEEE.

[17]. Davida, H. E., Ramalakshmib, K., Venkatesanc, R., & Hemalathad, G. (2021). Tomato Leaf Disease Detection Using Hybrid CNN-RNN Model. *Advances in Parallel Computing*, *38*, 593-597.