Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

A Review of AI Methods for Diagnosing Plant and Crop Diseases

Shanmugapriya N.[1], Thalluru Harika[2], Mannem Vijitha Reddy[3], Manchikalapati Subrahmanyam[4], Kolapalli Jaswanth Teja[5]

[1]Associate Professor, Department of CSE, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Trichy

[2-5]UG Student, Department of AIDS, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Trichy

Email: [1]fshanmugapriyan.set@dsuniversity.ac.in, [2]harika20048@gmail.com, [3]vijithareddymannem46@gmail.com, [4]nanisubbu013@gmail.com, [5]Jaswanthk503@gmail.com

Abstract

Plant diseases, caused by infectious organisms and unfavorable environmental conditions, pose a significant threat to agriculture. These diseases lead to notable decreases in crop yields, resulting in substantial economic losses. It is essential to address these challenges and safeguard the food supply and sustain effective farming practices. Though the detection of diseases in plants and crops through the traditional methods has always been a very difficult task to deal with the emergence of the AI (artificial intelligence) technologies has enhanced the efficiency and the accuracy of the diagnosis. This study presents a brief overview of the various machine and deep learning methods used for disease recognition in plants and compares the performance of the machine learning and deep learning algorithms in the detection of diseases in plants. The comparative study demonstrates that the deep learning methods achieve higher accuracy and better performance on complex tasks related to machine learning at the cost of increased computational resources and training time.

Keywords: Disease detection, Machine Learning, Deep Learning, Naïve Bayes, KNN, CNN, RNN.

1. Introduction

The development of a nation depends significantly on its agricultural growth. Though agriculture in India contributes to over 16% of the country's GDP and millions of households

DOI: https://doi.org/10.36548/jtcsst.2024.4.005
Received: 23.10.2024, received in revised form: 17.11.2024, accepted: 01.12.2024, published: 09.12.2024

in India consider agriculture as their primary occupation, satisfying the food requirements of the current population has become increasingly challenging due to the climatic changes and limited resources. A significant challenge in agriculture is the prevalence of plant diseases[1].

Plant diseases are caused by infectious organisms and various environmental factors. These organisms consume the tissues of the plants and affect the plant's health which reduces the crops yield leading to the economic loss of farmers. So, addressing these agricultural challenges becomes very essential to protect the food supply and sustain the effective farming practices.

Though conventional methods such as biological dignositcs, visual observation, and microscopy are available for early diagnosis of plant disease, these methods do not ensure promptness and reliability [2].

To enable fast and reliable identification of plant diseases in their early stages and to increase crop yield, many researchers have proposed the use of AI technologies for plant disease diagnosis. This study also presents a brief overview of machine learning and deep learning methods used in the early diagnosis of plant diseases.

The rest of the manuscript is organized with Section 2 that discusses the details of machine learning and the deep learning process for plant disease detection. Section 3 reviews the existing works on plant disease detection using machine and deep learning. Section 4 compares the performance of ML and DL, and Section 5 presents the conclusion.

2. Artificial Intelligence for Plant Disease Detection

Artificial intelligence (AI) allows machines to learn from experience, adapt to new information, and perform tasks that mimic human abilities. Machine learning is the subset of AI, that uses algorithms to identify patterns in data and make predictions based on structured data. Deep learning, is the subset of machine learning, they are highly complex and mimic the human brain. The deep learning learns complex patterns using a large amount of data. The Figure .1 below shows the overview of artificial intelligence.

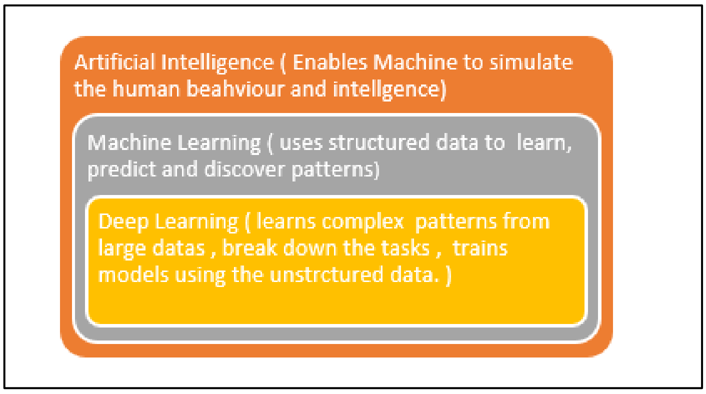

Figure 1. Overview of AI [17]

Machine learning (ML) and deep learning (DL) are widely utilized across various fields, transforming industries by their capacity to analyze complex data and extract meaningful insights. Machine learning finds application in areas like fraud detection, recommendation systems, healthcare diagnostics, and predictive maintenance. Deep learning enhances these capabilities by addressing more intricate tasks, including image recognition, natural language processing, and autonomous systemsIn the agricultural sector, specifically in the detection of plant diseases, these technologies make a significant impact. ML algorithms differentiate between healthy as well as infected plants by evaluating structured datasets, while deep learning models, perform well at analysing large sets of image data to accurately identify diseases. These techniques facilitate early diseases identification, minimalizing the crop loss and boosting productivity by equipping farmers with practical insights. The combination of ML and DL in agriculture thus promotes sustainable farming practices and ensures food security [16].

The development of machine learning and deep learning techniques for disease detection in plants typically follows a standard procedure. This process starts with the dataset collection that includes both healthy and unhealthy plants. Next, preprocessing, feature extraction, and feature selection are performed in the case of machine learning. In contrast, deep learning automatically extracts features during the training phase, eliminating the need for manual feature engineering [17]. Following these steps, model selection, evaluation, and prediction take place. The Figure 2 below demonstrates the general design process for both the approaches.

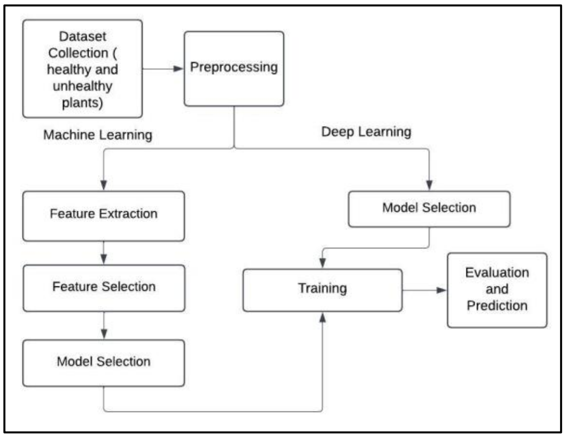

Figure 2. Flowchart for Plant Disease Detection using ML and DL

Table 1 below summarizes the difference in the procedure of the ML and the DL models.

Table 1. Difference between ML and DL

Various Aspects	ML	DL
Data Requirements	Requires a smaller dataset.	Requires a large amount of dataset.
Feature Extraction	Requires manual feature extraction.	Features are automatically extracted by the model.
Training time	Takes a shorter time for small dataset	Requires more time.
Complexity	relies on domain expertise for features.	learns patterns directly from data.
Epochs/Iterations	Few epochs are sufficient	Requires many epochs to converge, depending on the dataset and model size.
Performance	Depends on feature quality.	Achieves higher accuracy for complex data

Deep learning (DL) and Machine learning (ML) are both essential in the detection of plant diseases, each presenting unique benefits. ML algorithms, which depend on manually extracted features such as color, texture, and shape, are quicker to train and effective with smaller datasets, making them ideal for preliminary research or situations with limited resources. Nonetheless, their accuracy is largely influenced by the quality of the features that are used. On the other hand, DL models, are adept at handling extensive image datasets, automatically discovering intricate patterns, and attaining greater accuracy in detecting plant diseases. Although DL demands greater computational resources and takes longer to train, its capability to identify subtle symptoms of diseases is crucial for precision agriculture. ML offers practical and speedy solutions in simpler contexts, while DL tackles more complex issues, providing strong and scalable tools for improving crop health and productivity. The following section presents the review of existing works on the plant disease detection using the machine learning and deep learning algorithms.

3. Existing Work

This study explores the use of a Back Propagation Neural Network (BPNN) classifier to detect plant diseases based on visual leaf symptoms. Specifically, it focuses on two diseases affecting pomegranate plants: Bacterial Blight (BB) and Wilt Complex (WC). The research involves capturing, enhancing, and segmenting images of both healthy and diseased leaves to identify infected areas. Features related to colour and texture are extracted from these images and processed through the BPNN classifier, which accurately classifies the diseases. The proposed classifier achieves a commendable accuracy rate of 97.30%, offering valuable assistance to farmers in making informed decisions.[3].

This study focuses on the early identification of plant diseases, particularly in tomato plants, to help secure food supplies for the increasing populations. Its goal is to inform farmers about advanced technologies that can minimize plant diseases. The research employs ML and image-processing technologies to identify leaf diseases in tomatoes. Tomato leaf samples displaying disorders are resized and enhanced through Histogram Equalization, followed by analysis using K-means clustering, contour tracing, and various feature extraction techniques including DWT, PCA, and GLCM. The identified features are categorized using machine learning algorithms: “Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), and Convolutional Neural Network (CNN)”. The performance of the models is assessed with SVM achieving an accuracy of 88%, K-NN at 97%, and CNN at 99.6% on the tomato leaf samples [4].

This research investigates the use of machine learning (ML) to create early prediction models aimed at detecting plant diseases, with the goal of minimizing plant mortality through timely diagnosis. Given the difficulties in classifying crops during the initial phenological stages, ML techniques are employed for crop identification using high-resolution optical images captured by drones. The greyscale images undergo processing to generate grey level co-occurrence matrices for feature extraction. The research formulates a variety of ML models, such as K-Nearest Neighbors, Random Forest, Neural Networks, Linear Regression Support Vector Machines, and Naive Bayes for the purpose of detecting plant diseases. The findings indicate that the ensemble plant disease model surpasses the other models, providing enhanced early detection for proactive measures and predictive care in plant health management [5].

This study addresses the challenge of rapidly detecting crop diseases, which is crucial for maintaining food security but is difficult in many regions due to inadequate infrastructure. By leveraging the growing availability of smartphones and improvements in computer vision through deep learning, the research creates a smartphone-based system for diagnosing diseases. A publicly available dataset containing 54,306 images of both affected and healthy plant leaves is used to train the model to classify 14 different crop types and 26 distinct diseases. The deep convolutional neural network attains an accuracy rate of 99.35% on the test set, showcasing the possibility of utilizing deep learning models for diagnosing crop diseases. This approach highlights the potential of extensive, publicly accessible image datasets for enabling global smartphone-based disease detection [6].

This research introduces an automated method for identifying plants through leaf images, tackling the issues associated with manual identification, which is often labor-intensive and ineffective. The technique extracts both shape and color characteristics from leaf images and utilizes classification algorithms such as KNN, Naive Bayes, SVM, and Random Forest. This method is assessed on 1,897 leaf images representing 32 different plant species. The findings indicate that the accuracy rate of recognition can reach to 96% when employing both shape and color features, with the Random Forest yielding the highest level of accuracy [7].

This study introduces a DL model called “Plant Disease Detector”, which is aimed at identifying various plant diseases through leaf images. The model employs a “Convolutional Neural Network (CNN)” featuring several convolutional and pooling layers, and it is trained using the PlantVillage dataset. To improve the dataset, augmentation techniques are implemented, and the model is evaluated on 15% of the data, encompassing both healthy and afflicted plant images. The model reaches a testing accuracy of 98.3%. The study emphasizes the application of DL in detecting plant diseases, with intentions to later integrate the model with drones or other technologies for real-time disease detection and location tracking [8].

This study focuses on plant disease recognition using a CNN-based technique, aiming to address the challenges of crop diseases, which significantly impact agricultural productivity in India. The proposed model uses image processing techniques to detect diseases in plant leaves. A total of 15 cases were analyzed, including 12 diseased plant leaves (from crops like Bell Pepper, Potato, and Tomato) and 3 healthy leaves. The model's performance is evaluated based on area of the infected region and the time complexity, achieving a testing accuracy of 88.80%. Various performance metrics are derived to assess the model's effectiveness in detecting plant diseases [9].

This study introduces a method for identifying and classifying diseases in plant leaves using the K-nearest neighbor (KNN) algorithm, focusing on the challenges the modern organic farming faces due to plant disease. The method consists of extracting texture features from images of diseased leaves for the purpose of classification. The KNN algorithm is employed to identify and classify diseases such as “Anthracnose, leaf spot, bacterial blight, Alternaria alternata, and canker” across different plant species. The proposed method achieves an impressive detection and classification accuracy of 96.76% [10].

This research introduces a software-driven method for the automated classification and categorization of groundnut leaf diseases, aiming to enhance crop yields by tackling issues linked to diseases caused by fungi, soil-borne pathogens, and viruses. The procedure includes multiple stages: acquiring images, preprocessing them, segmenting, extracting features, and classifying using the KNN algorithm. To improve the existing algorithm's performance, the SVM classifier is substituted with KNN. The research concentrates on distinguishing four various groundnut diseases with the KNN classifier [11].

The objective is to identify unhealthy plant leaves resulting from diseases such as microbial infections, which adversely affect plant health and yield. This research utilizes MATLAB to assess the impacted area and analyze the coloration of the diseases on plant leaves, facilitating rapid and effective disease identification. A Naïve Bayes classifier is used to recognize and categorize the unhealthy areas of the leaves. First, images of the leaves are gathered, converted into a different color space, segmented, and then analyzed for features prior to classification. This approach achieves an accuracy of approximately 97% [12].

This study focuses on diseases and pests identification in corn plants through the use of statistical machine learning techniques, specifically “Multinomial Naïve Bayes (MNB) and K-Nearest Neighbor (KNN)”. Corn (Zea mays L) serves as an important carbohydrate-producing crop that is susceptible to various diseases and pests, which can adversely affect both the quality and quantity of production. The research utilizes a dataset comprising 761 digital images categorized into 6 classes of corn plant diseases and pests. Findings indicate that the KNN approach surpasses the MNB technique, with KNN achieving an 99.54% accuracy , 88.57% precision , 94.38% recall, and AUC of 95.45%. In contrast, the MNB method reaches an 92.72%, 79.88%, accuracy and precision respectively, and 79.24% recall, and AUC of 71.91% [13].

This study introduces a method for image segmentation aimed at the automated recognition and classification of plant leaf diseases, addressing the substantial consequences of plant disease on agricultural profitability and crop yield. It also examines different classification methods for identifying plant diseases. Given the variability in the size of infected areas, the study utilizes Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNN) to derive feature vectors for disease classification. The LSTM model achieved 95.79% accuracy and 0.9096 a Matthew’s Correlation Coefficient (MCC) . The Random Forest (RF) algorithm demonstrated comparable performance, attaining an accuracy of 94.95% and an MCC of 0.8915 [14].

This study addresses the problem of leaf spot disease, which significantly impacts food production, by detecting and classifying leaf diseases early. The research utilizes deep learning methods, specifically CNN for pre-processing and feature extraction, and LSTM RNN for disease classification. A dataset of 20,000 plant images from Kaggle and 100 locally captured images were used for training and validation. The study compares the performance of various classifiers, including Random Forest (86%), KNN (89%), Artificial Neural Networks (ANN) (96%), SVM (97%), and LSTM (98%). The results show that the LSTM method provides the best classification accuracy of 98% [15].

Table 2. Comparative Study of ML and DL for Plant Disease Detection.

Articles	**Study on**	**Algorithm used**	**Dataset Used**	**Performance and Benefits**	**Limitations**
[3]	Pomegranate Disease Detection	Back Propagation Neural Network (BPNN)	Bacterial Blight and Wilt Complex	Accuracy: 97.30%. Helps in early disease detection	Limited to specific diseases
[4]	Tomato Disease Identification	SVM, KNN, CNN	Tomato leaf images	K-NN: 97%, CNN: 99.6% SVM: 88 %High accuracy with CNN, uses advanced feature extraction techniques	Used only for Tomato disease detection
[5]	Diverse plant disease	Random forest, KNN, LR, NB, SVM	Real-time images captured using drone	Ensemble model shows higher accuracy and early prediction as well as disease detection	Complex model setup and large data requirements
[6]	Diverse plant disease	CNN	Real-time 54,306 images (affected and healthy)	Accuracy: 99.35%	Requires large dataset, limited by image quality
[7]	Plant Recognition using Leaf Images	KNN, SVM, NB, Random forest	1,897 images of 32 plant species	Accuracy: 96 % for random forest, Highly accurate plant recognition	Applicable only for species covered in the dataset
[8]	Plant Disease Detection	CNN	PlantVillage dataset	Accuracy: 98.3%, capable of detecting in real-time	Relies fully on the dataset for training.
[9]	Tomato Plant Disease Detection	CNN	15 disease cases (Bell Pepper, Potato, Tomato)	Accuracy: 88.80% Robust for various plant diseases	Applicable only for species covered in the dataset
[10]	Leaf Disease Classification	KNN	Real-time dataset	Accuracy: 96.76% Efficient for identifying	Applicable only for disease

				various plant diseases	covered in the dataset
[11]	Groundnut Disease Detection	KNN	Real-time dataset of ground nut images	Does early detection and enhances crop yield	Suitable only for the four types of groundnut disease covered in the dataset.
[12]	Leaf Disease Classification	Naïve Bayes (NB)	Real-time dataset	Accuracy: 97% Efficient and rapid disease identification	less accurate for more complex cases
[13]	Corn Disease and Pest Identification	MNB, KNN	761 corn plant images (6 disease and pest classes)	KNN: Accuracy 99.54%, MNB: Accuracy 92.72% KNN performs better with high accuracy and precision	Applicable only for disease covered in the dataset
[14]	Plant Disease Detection	LSTM, Random Forest (RF)	Real-time dataset	LSTM: 95.79% accuracy, RF: 94.95% offers accurate disease classification and LSTM excels at sequence data	Performance depends on feature extraction and segmentation quality
[15]	Leaf Spot Disease Detection	CNN, LSTM	20,000 images from Kaggle and 100 local images	LSTM: Accuracy 98% High accuracy for disease detection by utilizing both CNN and LSTM	Requires large dataset for training and the image quality are affected by the environmental factors

Table.2 above illustrates the comparison of the various ML and the DL methods used in the study.

Machine Learning (ML) and Deep Learning (DL) both provide notable benefits for identifying plant diseases. ML approaches, such as SVM and RF, perform well with smaller datasets and typically require less computational power. They depend on manually crafted features, which makes them interpretable and appropriate for less complex tasks. However, ML models can face challenges with scalability, feature engineering, and managing the complexities of large and varied datasets. On the other hand, DL methods like CNN and LSTM networks thrive in analyzing complex, large-scale data, automatically deriving features from raw images without the need for manual input. DL models achieve greater accuracy and resilience in plant disease detection due to their capacity to learn hierarchical representations directly from image data, with some models reaching accuracies as high as 99.6%. Although they demand significant computational resources and larger amounts of labeled data, DL's capacity to generalize and adapt to differences in plant diseases makes it more advantageous for contemporary, real-time agricultural applications compared to traditional ML techniques.

4. Conclusion

The emergence of Machine and Deep Learning has proven advantageous for early disease identification in plants and crops. This study reviewed various Machine Learning and Deep Learning methods. In summary, all the methods discussed in the document have demonstrated effectiveness in detecting plant diseases. However, the accuracy and other metrics used to evaluate the efficiency of the models depend on the size of the dataset utilized for training. When the dataset is small, all the techniques, including those referenced, perform adequately; however, for larger and more complex datasets, preference should be given to deep learning techniques. Future studies will be focussing on hybridizing ML and DL models to enhance plant disease detection.

References

[1] Balafas, Vasileios, Emmanouil Karantoumanis, Malamati Louta, and Nikolaos Ploskas. "Machine learning and deep learning for plant disease classification and detection." IEEE Access (2023).VOLUME 11, 2023 114352-114377.

[2] Khakimov, Albert, Ilias Salakhutdinov, Almas Omolikov, and Samad Utaganov. "Traditional and current-prospective methods of agricultural plant diseases detection: A review." In IOP Conference series: earth and environmental science, vol. 951, no. 1, IOP Publishing, 2022. 012002.

[3] S. S. Sannakki and V. S. Rajpurohit, "Classification of Pomegranate Diseases Based on Back Propagation Neural Network", International Research Journal of Engineering and Technology (IRJET), vol. 2, no. 02.

[4] Harakannanavar, Sunil S., Jayashri M. Rudagi, Veena I. Puranikmath, Ayesha Siddiqua, and R. Pramodhini. "Plant leaf disease detection using computer vision and machine learning algorithms." Global Transitions Proceedings 3, no. 1 (2022): 305-310.

[5] Ahmed, Imtiaz, and Pramod Kumar Yadav. "Plant disease detection using machine learning approaches." Expert Systems 40, no. 5 (2023): e13136.

[6] Sharada P. Mohanty, P. Hughes David and Marcel Salathe, "Using deep learning for image-based plant disease detection", Frontiers in plant science, vol. 7, 1419, 2016.

[7] Caglayan, Ali, Oguzhan Guclu, and Ahmet Burak Can. "A plant recognition approach using shape and color features in leaf images." In Image Analysis and Processing–ICIAP 2013: 17th International Conference, Naples, Italy, September 9-13, 2013, Proceedings, Part II 17, Springer Berlin Heidelberg, 2013. 161-170.

[8] Chohan, Murk, Adil Khan, Rozina Chohan, Saif Hassan Katpar, and Muhammad Saleem Mahar. "Plant disease detection using deep learning." International Journal of Recent Technology and Engineering 9, no. 1 (2020): 909-914.

[9] Shrestha, Garima, Majolica Das, and Naiwrita Dey. "Plant disease detection using CNN." In 2020 IEEE Applied Signal Processing Conference (ASPCON), IEEE, 2020. 109-113.

[10] Hossain, Eftekhar, Md Farhad Hossain, and Mohammad Anisur Rahaman. "A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier." In 2019 international conference on electrical, computer and communication engineering (ECCE), IEEE, 2019. 1-6.

[11] Vaishnnave, M. P., K. Suganya Devi, P. Srinivasan, and G. Arut Perum Jothi. "Detection and classification of groundnut leaf diseases using KNN classifier." In 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), IEEE, 2019. 1-5.

[12] Mohanapriya, K., and M. Balasubramani. "Recognition of unhealthy plant leaves using Naive Bayes classifier." In IOP Conference Series: Materials Science and Engineering, vol. 561, no. 1, IOP Publishing, 2019. 012094.

[13] Resti, Yulia, Chandra Irsan, Mega Tiara Putri, Irsyadi Yani, Ansyori Ansyori, and Bambang Suprihatin. "Identification of corn plant diseases and pests based on digital images using multinomial naïve bayes and k-nearest neighbor." Science and Technology Indonesia 7, no. 1 (2022): 29-35.

[14] Gnanasaravanan, S., B. Tharani, and Mona Sahu. "Long Short-Term Memory Recurrent Neural Networks for Plant disease Identification." Plant pathology 3, no. 2 (2021).

[15] Sivalingam, Vidya, Rawia Elarabi, J. Bhargavi, M. Sahaya Sheela, R. Padmapriya, and A. N. Arularasan. "A novel approach for plant leaf disease predictions using recurrent neural network RNN classification method." Journal of Advanced Research in Applied Sciences and Engineering Technology 31, no. 2 (2023): 327-338.

[16] https://www.analyticsvidhya.com/articles/machine-learning-vs-artificial-intelligence-vs-deep-learning/#h-what-is-artificial-intelligence

[17] https://www.cengn.ca/information-centre/innovation/difference-between-ai-ml-and-dl/