ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-37799443/IJAEAST.2024.05.0007, Volume-2, Issue-5, pp.187 - 192
International Journal of Advances in Engineering Architecture Science and Technology

# Classification and Prediction of the crop disease based on the Machine Learning Algorithms

Chamarthi Yamini[1], G. Lakshmikanth[2]*
[1]B. Tech. Student, Department of CSE, SREC, Andhra Pradesh, India
[2]Assistant Professor, Department of CSE, SREC, Andhra Pradesh, India
yaminichamarthi321@gmail.com[1], svlakshmikanth21@gmail.com[2]

**Received:** 26-03-2024 **Accepted:** 28-05-2024 **Published:** 31-05-2024

## Abstract

***Background:*** Machine learning is an emerging field, and it has almost applications in all the fields such as agriculture, healthcare, and construction. Here the agriculture application was used the ML techniques t classify and detect the crop disease. Most research has been done in this area where the results still can be improvised in terms of the accuracy.
***Objectives:*** Here the work is to classify and predict the crop diseases based on the machine learning algorithms. Specifically in terms of the accuracy of classification and detection of the crop disease will be improvised.
***Methods:*** Here the machine learning algorithms such as Random Forest Regressor, KNN Regressor, Gradient Boosting Regressor, and Bidirectional Long Short-Term Memory.
***Statistical Analysis:*** Here the MSE (Mean Squared Error) and MAE (Mean Absolute Error) were used for obtaining results.
***Findings:*** Based on the obtained results Gradient boosting algorithm got maximum MSE and MAE.
***Applications:*** Here the crop disease prediction and classification has been done based on the dataset that taken from the UCI repository and machine learning algorithms.
***Improvements:*** In future, different machine learning algorithms and deep learning algorithms will be applied on various crop disease datasets that available in Kaggle and UCI repository for classification, prediction, and detections of the crop diseases.

**Keywords:** Machine Learning, Deep Learning, Plant, diseases, detection, and prediction.

## 1. Introduction

Early and reliable crop yield estimation is crucial for field-level quantitative and financial evaluation of agricultural commodities, defining import-export policies, and tripling farmer income. Machine learning algorithms are used to predict crop yields, addressing a hard issue in agriculture [1]. Globalization, trade, and climate change have increased the prevalence of plant diseases in recent years. These challenges have reached epidemic proportions in numerous nations, increasing crop damage and threatening people's access to basic food and nutrition. Specialists should safeguard agricultural plants. Parasitic species, including bacteria, fungus, viruses, nematodes, and plants, can cause sickness [2]. The lack of sophisticated disease identification methods has significantly impacted crop productivity. Farmers have extensive experience identifying crop diseases and managing them locally.

Lack of knowledge-sharing platforms prevents local expertise from being shared across agricultural regions. Agricultural study indicates that crop productivity has declined due to illnesses, cultivation practices, irrigation, and a loss of local expertise [3]. A secure and sustainable food supply system requires good agriculture management. Technological advancements in agriculture have led to increased productivity and quality. Early diagnosis and characterization of plant leaf diseases is vital for improving agricultural output and overall well-being for farmers and consumers. This research focuses on using machine learning techniques to automatically identify and categorize illnesses affecting apple and maize plants. Deep learning has advanced the automatic identification and categorization of plant diseases in fruits and vegetables. As a result, the quantity and quality of fruits and vegetables have improved [4-7]. As COVID-19 spreads globally, key grain-producing countries restrict exports, raising concerns about food security. Improving grain production is a top priority for all governments. Crop diseases pose a challenge for farmers, therefore it's crucial to precisely assess their severity and take proactive actions to prevent additional infection [9].
The remainder of the paper is, in section 2 discussed about related work, in section 3 discussed about methodology, in section 4 discussed about the results and discussion and finally in section 5 discussed about the conclusion and future work.

## 2. Related Work

A proposed architecture for predicting palm oil yield using machine learning is based on a careful examination of existing works. This technology will achieve its promise by addressing new research difficulties in crop analysis. Here developed a highly successful model for predicting palm oil yields with low computational effort [1]. This research provides a new method for extracting deep characteristics and diagnosing sick plant leaves. Plant diseases have a significant influence on agriculture, leading to crop and economic losses. Accurate and fast diagnosis is essential for treating and controlling plant diseases, as traditional procedures can be expensive and time-consuming. Deep learning technologies may accurately diagnose plant diseases based on derived qualitative data. We present a hybrid model for plant disease classification that combines a Transfer Learning-based model with a vision transformer (TLMViT). TLMViT has four stages: The proposed model is trained and evaluated using the PlantVillage and wheat datasets. Image augmentation is utilized to enhance the number of training examples and address overfitting issues. Leaf features are extracted in two phases: initial features using a pre-trained model and deep features using the ViT model, followed by classification using an MLP classifier. TLMViT is tested using five pre-trained models, followed by ViT individually. TLMViT achieves 98.81% and 99.86% validation accuracy for VGG19, followed by the ViT model on PlantVillage and wheat datasets, respectively. Furthermore, TLMViT is compared to pre-trained-based architecture. TLMViT outperformed the transfer learning-based model in PlantVillage and wheat datasets, improving validation accuracy by 1.11% and 1.099%, and reducing validation loss by 2.576% and 2.92%, respectively [2]. This project uses a crowd-sourced platform to acquire experience from agricultural specialists, farmers, and cultivators. The data is processed to identify various diseases. Early detection of crop diseases allows farmers to use effective management strategies. Researchers have proposed disease management systems that primarily use Machine Learning (ML) algorithms to classify agricultural illnesses. However, these algorithms cannot provide reliable results because to static data provisioning and the changing nature of diseases in diverse agricultural locations. Furthermore, the agricultural expert's experience is not utilized while confirming classification results [3]. This paper provides a complete assessment of leaf disease studies in the literature. It also indicates the gaps that need to be addressed, as well as the

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-37799443/IJAEAST.2024.05.0007, Volume-2, Issue-5, pp.187 - 192
International Journal of Advances in Engineering Architecture Science and Technology

challenges and issues that research initiatives face. A total of 256 papers were collected from five internet databases. We studied and categorized them into seven research questions. The study found that 63% of the publications were journal articles, 35% were conference papers, and 2% were workshop papers [4-6]. This study presents a deep learning-based technique for classifying plant leaf diseases. This work uses two pre-trained deep learning models and proposes a hybrid strategy. Our research proposes a hybrid framework that includes a hybrid preprocessing approach, an ensemble feature engineering phase using texture features, and two types of deep feature extraction methods. The retrieved CNN features are merged with the LBP features. The ensemble feature vector is optimized using three meta-heuristic algorithms: Binary Dragonfly (BDA), Ant Colony, and Moth Flame (MFO). The optimized feature vector is classified utilizing advanced machine learning algorithms [7]. Wheat is a global staple crop with significant nutritional value. It offers over 20% of the daily calorie and protein requirements for humans, highlighting its relevance. Rust disease can lower wheat yield by 30%, posing a significant danger to food security. To reduce loss, it's important to accurately diagnose and localize wheat rust illness and infection types. Various machine/deep learning-based classification and segmentation approaches are utilized [8]. This study analyzes and evaluates several segmentation approaches, including Watershed, Grab Cut, and U2-Net. These algorithms generate numerous datasets from wheat stripe rust data, including Watershed, GrabCut, and U2-Net segments. ResNet-18, a pre-trained deep learning model, is used to evaluate how segmentation affects classification performance on these datasets. The dataset segmented by U2-Net shows the best classification accuracy (96.196%). This study compares the correctness and impact of several segmentation approaches on classification accuracy, providing researchers with a practical guide to selecting the best technique [8]. This paper proposes a restructured residual dense network for identifying tomato leaf diseases. This hybrid deep learning model combines the benefits of deep residual and dense networks, reducing the number of training parameters and improving calculation accuracy and information flow. We need to reconstruct the RDN model for classification tasks by adjusting input image characteristics and hyperparameters, as it was originally designed for image super resolution. The model achieved a top-1 average identification accuracy of 95% on the Tomato test dataset in AI Challenger 2018, demonstrating satisfactory performance [9 -10].

## 3. Methodology

The approach for classifying and predicting agricultural diseases using machine learning algorithms includes numerous essential stages, such as data collection, preprocessing, feature extraction, algorithm selection, model training, evaluation, and deployment. Each stage is methodically planned to enable the creation of a strong and accurate predictive model capable of assisting farmers with timely and effective disease management.

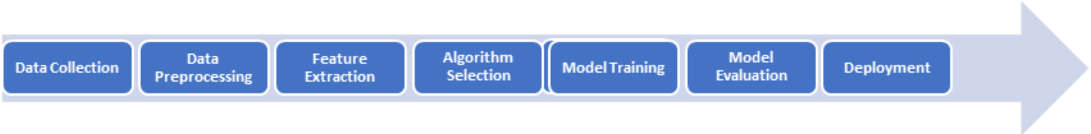


**Figure 1.** Methodology

### Data Collection

The core of our process is rigorous data collection. We collected high-resolution photos of crops in varied health states, from healthy to ill. This visual data was supplemented with sensor data that measured environmental factors such as soil moisture, temperature, and humidity. In addition, historical agronomic data, such as previous disease outbreaks and crop yields, was used to provide context for the predictive models.

**Data Preprocessing**

We performed substantial preprocessing processes on the acquired data before analysing it. This entailed cleaning the data to remove noise and irrelevant information, hence preserving the dataset's integrity. Each image was labelled with the precise type of sickness, resulting in a labelled dataset required for supervised learning. Normalization techniques were used on sensor and historical data to reliably scale the readings, allowing the algorithms to learn more efficiently. Data augmentation techniques like as rotation, flipping, and scaling were used to increase the diversity of the image dataset, which improved the model's generalizability.

**Feature Extraction**

Feature extraction was used to identify and isolate characteristics in the data that were suggestive of disease. Feature extraction for image data included color histograms, texture patterns, and form descriptors, which captured crucial visual cues associated with various disorders. Relevant features found in sensor and environmental data included temperature patterns, humidity levels, and soil conditions. This stage is critical for lowering the dimensionality of the data and directing learning algorithms to the most informative features.

**Algorithm Selection**

The nature of the data and the unique needs of the categorization assignment influenced the choice of suitable machine learning algorithms. Traditional machine learning techniques, such as Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN), were evaluated for robustness and interpretability. However, due to the complexity and high dimensionality of picture data, deep learning approaches, specifically Convolutional Neural Networks (CNNs), were emphasized. Transfer learning with pre-trained models such as ResNet and Inception was also investigated to capitalize on existing knowledge and improve model performance.

**Model Training**

Model training entailed feeding the pre-processed and feature-extracted data into the chosen algorithms and iteratively changing the model parameters to reduce prediction errors. To ensure unbiased evaluation, the dataset was divided into three sets: training, validation, and testing. To maximize the model's performance, hyperparameter tuning was performed, with an emphasis on factors like as learning rate, batch size, and CNN layer count. This iterative procedure intended to create a model capable of reliably classifying and predicting crop diseases.

**Model Evaluation**

Evaluating the trained models was critical for determining their usefulness and reliability. To give a full assessment, performance indicators like as accuracy, precision, recall, and the F1 score were calculated. A confusion matrix was utilized to visually represent the model's performance across multiple classes, showing areas of strength and future improvement. Cross-validation approaches verified that the evaluation was reliable and generalizable, preventing overfitting and ensuring that the model functioned well on previously unseen data.

**Deployment**

Our process concluded with the deployment of the validated model for practical use. This involved transforming the model into a user-friendly interface for farmers and agronomists. The system was created to absorb new input data, such as crop photos or sensor measurements, and deliver real-time illness forecasts. Scalability and maintainability considerations were addressed to ensure that the system could manage vast amounts of data while also adapting to new information over time. Regular updates and retraining of the model with new data were planned to keep it accurate and relevant.

In conclusion, our methodology for classifying and predicting crop diseases using machine learning algorithms is a comprehensive process that combines advanced data collection, sophisticated preprocessing, and cutting-edge algorithmic techniques to provide a dependable and effective solution for agricultural disease management. The results generated by this were shown and discussed in section 4

## 4. Results and Discussion

Here after implementing the algorithms on the crop dataset, that taken from the UCI repository were shown in the figure2.

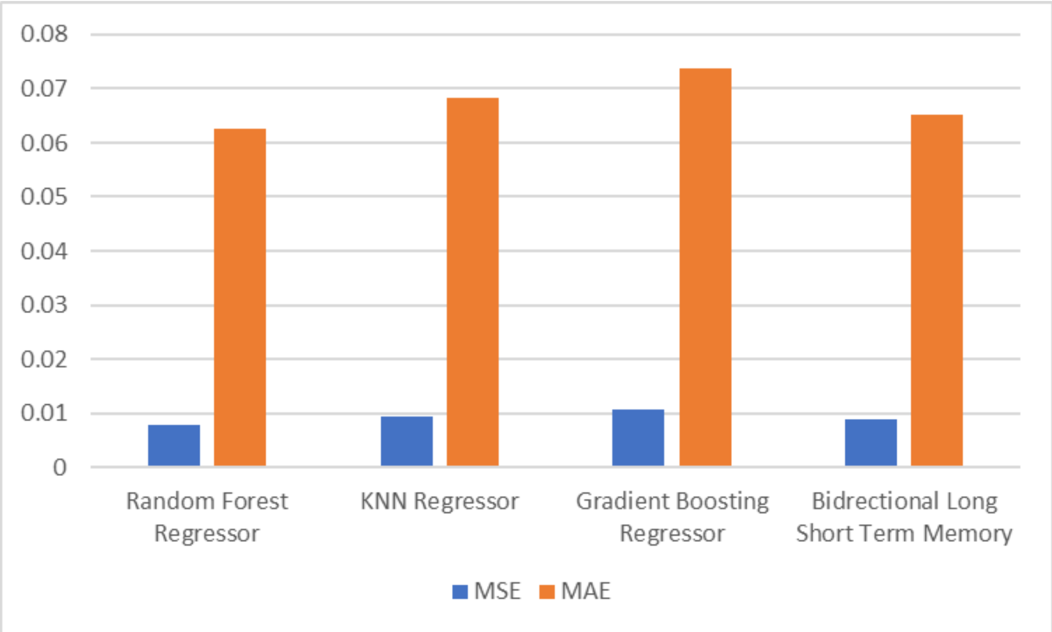


**Figure 2.** Comparison of Machine Learning Algorithms

The random forest, KNN, Gradient boosting and bidirectional machine learning algorithms were implemented on the dataset that taken from the UCI repository, among this algorithm the gradient boosting has maximum MAE and MSE and the random forest has least MAE and MSE.

## 5. Conclusion and Future Work

To summarize, while great progress has been made in the categorization and prediction of crop diseases using machine learning algorithms, further research and development are required to improve these models and broaden their usefulness. By continuing to innovate and integrate new data sources and technology, we can get closer to developing resilient and efficient agricultural systems that can support the world's rising population.

## References

1. Rashid, Mamunur, et al. "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction." IEEE access 9 (2021): 63406-63439.
2. Tabbakh, Amer, and Soubhagya Sankar Barpanda. "A Deep Features extraction model based on the Transfer learning model and vision transformer" TLMViT" for Plant Disease Classification." IEEE Access (2023).
3. Haider, Waleej, et al. "A generic approach for wheat disease classification and verification using expert opinion for knowledge-based decisions." IEEE Access 9 (2021): 31104-31129.
4. Doutoum, Assad S., and Bulent Tugrul. "A Review of Leaf Diseases Detection and Classification by Deep Learning." IEEE Access (2023).
5. Sahu, Priyanka, et al. "A systematic literature review of machine learning techniques deployed in agriculture: A case study of banana crop." IEEE Access 10 (2022): 87333-87360.
6. Shafik, Wasswa, et al. "A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets, Challenges, and Future Trends." IEEE Access 11 (2023): 59174-59203.
7. Shafik, Wasswa, et al. "A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets, Challenges, and Future Trends." IEEE Access 11 (2023): 59174-59203.
8. Bukhari, Hassan Raza, et al. "Assessing the impact of segmentation on wheat stripe rust disease classification using computer vision and deep learning." IEEE Access 9 (2021): 164986-165004.
9. Zhou, Changjian, et al. "Tomato leaf disease identification by restructured deep residual dense network." IEEE Access 9 (2021): 28822-28831.
10. Alharbi, Abdulaziz, Muhammad Usman Ghani Khan, and Bushra Tayyaba. "Wheat disease classification using continual learning." IEEE Access (2023).