IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11547

Machine Learning Models for Detection and Prediction of Crop Diseases : A Review

M. Venkata Ramana

Department of Physics,

Anurag University, Hyderabad - 500088, India.

Abstract: This paper aims to provide a comprehensive overview of machine learning (ML) techniques across various data types, fostering opportunities to address research gaps and advance the field, particularly in the detection and prediction of crop diseases. The survey presents valuable insights into ML-based techniques for forecasting, detecting, and classifying diseases and pests. It highlights the importance of maintaining long-term datasets encompassing weather, disease, and pest data. Time-series ML models, such as recurrent neural networks (RNNs), are shown to be effective tools for accurately predicting disease and pest occurrences based on sequences of meteorological measurements. Additionally, incorporating normalized difference vegetation index (NDVI) measurements can provide supplementary insights into crop development. Leveraging computer vision and deep learning algorithms, particularly convolutional neural network (CNN) models, proves advantageous for detecting and classifying pests and diseases, outperforming traditional approaches that rely on manual feature extraction.

Key words: Machine Learning, Support Vector Machine, Random Forest, Artificial Neural Networks, Plant disease detection.

INTRODUCTION

Global food production systems are grappling with substantial losses in crop yield and financial resources primarily due to plant diseases. Recent reports highlight that these diseases reduce global food production by 20% to 40%, resulting in a 13% decline in global crop yields [1]. Early detection of plant and crop diseases emerges as the most efficient solution, enabling the implementation of proactive measures to enhance expected yields. However, human errors in manual disease identification can compromise the effectiveness of pesticide use, leading to irreversible consequences. Accurate disease identification is pivotal for determining suitable measures to safeguard crop health, thereby boosting yields, which are influenced by factors like seed quality, soil fertility, precipitation, temperature, and the use of natural or commercial fertilizers. While various elements contribute to plant diseases, researchers commonly consider the host, pathogen, and favourable environmental conditions as primary factors, with pesticide resistance, application timing, and pesticide quality playing secondary roles. These three factors are typically regarded as key by researchers and are interconnected, as illustrated in Fig. 1.

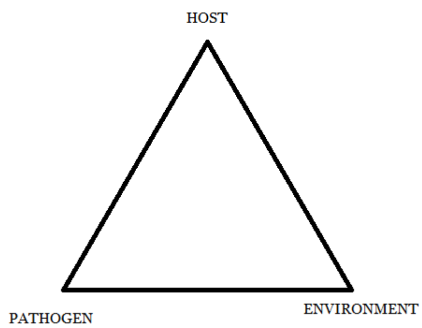

Fig.1. Plant Disease Triangle.

In most cases, plant diseases start showing symptoms at the base of the plant and then spread upwards. This is mainly due to the lower portion's proximity to the soil. The spread of the disease to the entire plant, the whole crop, and neighboring fields largely depends on the speed and direction of the airflow, as well as the mobility of the disease-spreading organisms. Therefore, it is essential to monitor the disease's spread in addition to identifying and controlling it. Early detection and limited spread are crucial for effectively managing the disease and preventing significant yield losses.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⅙ Peer-reviewed & Refereed journal ⅙ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11547

In most instances, plant diseases initially exhibit symptoms at the plant's base before spreading upward. This trend is largely attributed to the proximity of the lower portion to the soil. The spread of the disease to the entire plant, entire crop, and neighbouring fields heavily relies on airflow speed and direction, as well as the mobility of disease-spreading organisms. Thus, it becomes imperative to monitor the spread of the disease alongside its identification and control. Early detection and containment are imperative for effectively managing the disease and mitigating significant yield losses.

Farmers and agricultural scientists commonly rely on visual leaf inspection to identify plant diseases. Plants are susceptible to various diseases, each affecting distinct parts of the plant. Most diseases exhibit symptoms on leaves, referred to as foliar diseases, which undergo discernible changes observable through visual inspection. Fungal diseases, prevalent among plant diseases, often manifest through these leaf alterations. Identification typically entails comparing healthy leaves with diseased ones. Understanding the characteristics and patterns of healthy leaves and their alterations due to disease is essential. Given the voluminous data involved, leveraging plant leaf images coupled with digital vision, machine learning, and deep learning techniques has proven advantageous for continuous monitoring and precise diagnosis. Detecting multiple diseases across various crops, including within the same crop, will significantly benefit the farming community, fostering increased yields.

Manual visual inspection methods for plant disease diagnosis and monitoring are expensive, time-consuming, and dependent on experts, making them ineffective for precision agriculture despite their long-standing use. These traditional methods suffer from low accuracy and are subject to human bias and fatigue [2-4]. By contrast, the use of image processing techniques, machine learning, and deep learning with plant images can address these problems and provide more accurate results.

Although recognizing objects in images is easy for humans, it has traditionally been difficult for automated algorithms. However, recent advancements in deep learning have achieved accuracy levels that surpass human performance in some tasks. Modern deep learning algorithms have outperformed humans in image recognition, thanks to deep learning architectures that mimic biological neural networks. This rapidly evolving field offers various applications in image classification.

Over the past 10 to 12 years, the incorporation of deep learning techniques has enabled automatic feature extraction with high accuracy.

Machine learning models employed in the detection of plant diseases represent a pivotal advancement in addressing the considerable threat these diseases pose to global food security, causing significant economic losses and jeopardizing the livelihoods of agricultural communities. Traditionally, disease detection in plants relied on manual inspection by trained experts, a labor-intensive and time-consuming process. However, the emergence of modern technology has paved the way for machine learning models to automate the detection and diagnosis of plant diseases, offering a promising avenue for revolutionizing the agricultural industry. By harnessing the capabilities of artificial intelligence, these models hold the potential to deliver precise, rapid, and cost-effective solutions for identifying diseases in plants.

This application of machine learning in agriculture aligns with the broader field of precision agriculture, which seeks to optimize farming practices through data-driven decision-making. Leveraging an array of sensors, imaging techniques, and data analysis, machine learning models can discern subtle patterns and symptoms of plant diseases that may elude human detection. Offering real-time disease monitoring, these models facilitate early intervention, thereby mitigating crop losses and ultimately bolstering agricultural productivity.

In this paper, we delve into the current landscape of machine learning models utilized for plant disease detection, shedding light on their capabilities and limitations. Additionally, we explore the underlying technology and data sources, the challenges associated with deploying these models in real-world agricultural settings, and the potential for their widespread adoption. Furthermore, we examine recent breakthroughs and case studies that underscore the effectiveness of machine learning in plant disease detection.

Support Vector Machine (SVM)

Support Vector Machine (SVM) stands out as a favored machine learning algorithm in the realm of plant disease detection due to its adeptness in efficiently and accurately classifying data, particularly in binary classification

DOI: 10.17148/IARJSET.2024.11547

scenarios distinguishing healthy plants from diseased ones. SVMs exhibit versatility, robustness, and suitability for applications in agriculture and plant disease detection. Here, we delve into the application of SVM models in this context:

1. Feature Extraction and Selection: SVMs operate on data-driven principles, wherein their performance hinges upon the quality and relevance of input features. In the realm of plant disease detection, features typically encompass color, texture, and shape attributes extracted from images of plant leaves. The extraction and selection of pertinent features constitute pivotal steps in constructing an effective SVM model, as judiciously chosen features enable SVMs to discern between healthy and diseased plants accurately.

2. Binary Classification: SVM inherently operates as a binary classification algorithm, rendering it well-suited for scenarios framing plant disease detection as a binary problem—distinguishing between healthy and diseased states. Nonetheless, modifications and extensions can be employed to address multi-class problems that may arise with multiple disease classes.

3. Margin Maximization: SVM aims to identify the hyperplane that best separates the two classes—healthy and diseased plants—with the widest margin. This pursuit of margin maximization engenders a robust and generalizable model. SVMs excel particularly in scenarios featuring limited datasets, as they prioritize the most informative data points for delineating decision boundaries.

4. Non-Linearity: SVMs can accommodate non-linear decision boundaries via kernel functions. In plant disease detection, where the relationship between features and disease presence may exhibit non-linearity, kernel SVMs—such as the radial basis function (RBF) kernel—are commonly employed to capture intricate data patterns.

5. Handling Imbalanced Datasets: Imbalanced datasets, where the prevalence of healthy plants significantly outweighs that of diseased ones (or vice versa), are commonplace in plant disease detection. SVMs can be configured to address imbalanced datasets by assigning class-specific weights to instances, thereby circumventing bias toward the majority class.

6. Evaluation and Validation: Rigorous model evaluation is imperative. Cross-validation techniques—such as k-fold cross-validation—facilitate the assessment of SVM's performance and its generalizability to unseen data. Metrics encompassing accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic (ROC) curve typically gauge the model's efficacy.

7. Limitations: SVMs may not deliver optimal performance in scenarios featuring extensive datasets or necessitating real-time processing. Sensitivity to parameter settings—such as the choice of kernel function and regularization parameter—demands meticulous tuning.

Moreover, SVM excels in handling non-linear data through the utilization of the kernel trick technique, transforming low-dimensional input spaces into higher-dimensional spaces conducive to linear separability. Consequently, SVM proves highly effective in high-dimensional spaces. Additionally, SVM finds utility in regression problems, and its hybrid integration—such as combining SVM with the logistic regression algorithm—demonstrates promise in predicting diseases like powdery mildew in tomato plants. A synthesis of agricultural studies employing SVM as the ML model underscores its prevalence, with linear, polynomial, and RBF kernels being the most commonly utilized in SVM-based classification and regression algorithms applied to agricultural contexts. SVM demonstrates superior performance compared to other ML techniques such as ANNs and conventional regression approaches in forecasting plant diseases .

Support Vector Machine models are extensively utilized in plant disease detection due to their proficiency in accurately classifying plants as healthy or diseased based on image and feature data. They excel in handling non-linearity, imbalanced datasets, and contribute to automating disease diagnosis in agriculture. When coupled with appropriate feature engineering and parameter tuning, SVMs serve as a valuable tool for optimizing crop management and mitigating losses attributed to plant diseases.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11547

Random Forest

Random Forest (RF) emerges as a versatile and potent machine learning algorithm well-suited for plant disease detection. It excels in both classification and regression tasks and offers several advantages when applied to plant disease detection scenarios. Here's an overview of how Random Forest can be leveraged in this context:

1. Ensemble Learning: Random Forest operates as an ensemble learning technique, combining the predictions of multiple decision trees to yield a more robust and accurate outcome. In plant disease detection, it amalgamates the decisions of numerous trees to classify plants as healthy or diseased, furnishing a more dependable prediction.

2. Feature Importance: Random Forest models furnish valuable insights into feature importance, particularly crucial in plant disease detection scenarios reliant on features extracted from plant images (e.g., color, texture, shape). By ranking these features, Random Forest aids researchers in discerning the most pertinent characteristics for precise disease detection.

3. Robustness to Noisy Data: Random Forest exhibits resilience to noisy data and outliers, a prized attribute in agricultural settings where environmental conditions can introduce noise and image variations. The ensemble of decision trees adeptly handles noisy data, outperforming individual trees.

4. Handling Imbalanced Datasets: Imbalanced datasets prevalent in plant disease detection, where the healthy plants outnumber the diseased ones (or vice versa), pose no hindrance to Random Forest. It offers mechanisms to accord greater importance to the minority class, ensuring sustained model performance.

5. Non-Linearity: Random Forest adeptly captures intricate, non-linear relationships between features and the presence of plant diseases, a vital asset in agriculture where such relationships abound.

6. Prevention of Overfitting: Random Forest incorporates mechanisms to mitigate overfitting, rendering it more resilient to high-variance models, thereby ensuring robust generalization to unseen data—a critical facet in plant disease detection.

7. Model Interpretability: Despite being an ensemble of decision trees, Random Forest affords insights into the decision-making process, elucidating the most influential features in classifying plant health and facilitating a deeper understanding of disease detection.

8. Model Tuning: Random Forest offers parameters for customization, including the number of trees in the forest and the depth of individual trees, enabling model optimization tailored to the dataset's characteristics and the desired trade-off between bias and variance.

9. Model Validation: Cross-validation techniques, such as k-fold cross-validation, serve to evaluate Random Forest's performance in plant disease detection, with metrics like accuracy, precision, recall, F1-score, and ROC-AUC serving as yardsticks for effectiveness.

Random Forest (RF) stands out as a renowned ensemble comprising decision trees trained on diverse subsets of the training data. During node splitting, RF considers a random set of variables rather than the entire feature set. Each tree contributes its vote during classification, with the most commonly agreed-upon class being returned. Rapid computation owing to the subset-based training, coupled with a diverse array of trees, renders Random Forest robust to noise and outliers. Random Forests (RFs) exhibit superior accuracy even with fewer samples compared to other ML techniques .

Random Forest stands out as a robust and effective machine learning algorithm for plant disease detection. Its ensemble nature, coupled with its ability to handle non-linearity, conduct feature importance analysis, and robustly deal with noisy data, makes it highly suitable for this task. When appropriately configured and trained on high-quality data, Random Forest models can significantly contribute to automating plant disease diagnosis and enhancing crop management practices.

DOI: 10.17148/IARJSET.2024.11547

Artificial Neural Networks

Artificial Neural Networks (ANNs) have garnered significant attention in recent years for their application in plant disease detection and diagnosis. ANNs, a type of deep learning algorithm, excel in modeling complex relationships between input data, such as images of plant leaves, and the presence of diseases. Here's how ANNs can be applied in plant disease detection:

Image-Based Detection: ANNs prove particularly effective in image-based plant disease detection. Trained on extensive datasets of images depicting both healthy and diseased plants, ANNs learn to discern subtle patterns and features distinguishing between the two. Their deep architecture enables them to capture intricate details and relationships in images, essential for accurate diagnosis.

Convolutional Neural Networks (CNNs): CNNs, a specialized type of ANN for image analysis, are commonly employed in plant disease detection. These networks automatically learn hierarchical features from images through convolutional layers, enabling them to identify patterns such as lesions, discolorations, or irregularities on plant leaves.

Transfer Learning: Leveraging pre-trained CNN models (e.g., VGG16, ResNet, Inception) on large image datasets like ImageNet can significantly enhance plant disease detection. Fine-tuning these models for plant disease detection tasks, especially when working with limited datasets, capitalizes on knowledge acquired from general image recognition tasks.

Data Augmentation: Techniques such as data augmentation artificially expand the training dataset, enhancing the generalization capability of ANNs. By introducing variations in lighting, perspective, and background in plant images, ANNs become more robust to real-world conditions.

Multi-Class Classification: ANNs can be trained to perform multi-class classification, enabling the identification of various diseases concurrently. This capability is crucial for real-world applications where plants may suffer from multiple diseases simultaneously.

Hyperparameter Tuning: Fine-tuning ANNs' hyperparameters, including the number of layers, neurons per layer, learning rate, and activation functions, optimizes performance. Grid search or Bayesian optimization methods help identify the most effective hyperparameters.

Model Evaluation: Proper evaluation through techniques like k-fold cross-validation ensures ANNs' performance in plant disease detection. Metrics such as accuracy, precision, recall, F1-score, and ROC-AUC gauge model effectiveness.

Real-Time and On-Device Applications: Lightweight neural network architectures like MobileNet facilitate real-time processing on devices such as smartphones or drones, enabling immediate disease diagnosis and intervention in agricultural settings.

Continuous Learning: ANNs can be continually updated as new data becomes available, enabling adaptation to new disease variants or evolving field conditions.

While ANNs offer significant advantages in plant disease detection, challenges such as the need for large and diverse training datasets, model interpretability, and computational resources for training and deployment must be considered. Nonetheless, ANNs, particularly CNNs, hold the potential to revolutionize plant disease detection by offering accurate, efficient, and scalable solutions for precision agriculture.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have revolutionized computer vision and find extensive use in various domains, including plant disease detection. CNNs excel at learning intricate patterns and features in images, making

DOI: 10.17148/IARJSET.2024.11547

them well-suited for identifying plant diseases based on visual symptoms. Here's how CNN architectures are commonly utilized in plant disease detection:

Image Preprocessing: Before inputting images into a CNN, preprocessing steps are typically undertaken. These may include resizing images to a consistent size, normalizing pixel values, and augmenting the dataset with transformations like rotations, flips, and color adjustments to improve the network's generalization.

Choice of CNN Architecture: Various CNN architectures, such as AlexNet, VGG, ResNet, Inception, and DenseNet, cater to different complexity levels and computational resources, each offering specific advantages in plant disease detection tasks.

Transfer Learning: Pre-trained CNN models, often trained on vast datasets like ImageNet, can be fine-tuned for plant disease detection tasks, leveraging knowledge from general image recognition tasks to enhance model performance, especially when data is limited.

Data Augmentation: Techniques like random rotations, flips, and cropping augment the dataset size, enhancing model robustness and generalization to diverse plant disease variations.

Training and Validation: The dataset is divided into training, validation, and testing sets for CNN training and evaluation. The training set is used to train the CNN, while the validation set aids in hyperparameter tuning and performance monitoring. The testing set evaluates the CNN's generalization to new, unseen data.

Object Localization: Some plant disease detection models incorporate object localization techniques to precisely locate disease symptoms within plant images, employing architectures like Faster R-CNN or YOLO in conjunction with CNNs.

Post-processing: Refinement of disease detection results via post-processing steps such as thresholding, noise reduction, and clustering enhances the model's output accuracy.

Deployment: Trained CNN models can be deployed on various platforms, including mobile apps or embedded systems, enabling real-time or near-real-time disease detection in agricultural settings.

Using CNN architectures for plant disease detection has the potential to transform agriculture by enabling early and precise disease diagnosis, facilitating timely intervention to safeguard crops and enhance yields. The integration of CNNs can notably boost the efficiency and effectiveness of agricultural practices while diminishing the reliance on manual inspection methods.

User-defined Network Architectures

User-defined network architectures provide adaptability and the opportunity to craft tailored neural networks for specific purposes like plant disease detection. When devising custom architectures, it's crucial to account for the distinct characteristics and demands of the task. Here's a guide on crafting user-defined network architectures for plant disease detection:

1. Data Preparation: Initiate the process by gathering and preparing your dataset, typically comprising images of plant leaves depicting both healthy and diseased states. Preprocessing steps involve resizing, standardizing, and augmenting images to enhance model adaptability.

2. Architectural Considerations: Deliberate on architectural elements that align with your objectives. Key considerations include:

- Convolutional Layers: Employ convolutional layers to extract features from images. Customize parameters such as layer count, filter sizes, and activation functions (e.g., ReLU) based on dataset characteristics.

DOI: 10.17148/IARJSET.2024.11547

- Pooling Layers: Integrate pooling layers (e.g., max-pooling) to downscale spatial dimensions and extract pertinent features.
- Fully Connected Layers: Incorporate fully connected layers for classification tasks, with neuron counts tailored to task complexity.
- Dropout: Introduce dropout layers to mitigate overfitting.
- Batch Normalization: Enhance training stability and convergence speed with batch normalization layers.
- Skip Connections: Enhance feature capture with skip connections aiding local and global feature representation.

3. Custom Loss Functions: Design loss functions tailored to specific objectives. For instance, weighted loss functions can prioritize certain diseases over others based on importance.

4. Activation Functions: Experiment with diverse activation functions (e.g., ReLU, Leaky ReLU, Swish) to optimize model performance.

5. Data Augmentation: Implement augmentation techniques like rotation, translation, and color adjustment to augment dataset size and improve model robustness.

6. Regularization Techniques: Incorporate regularization techniques such as L1 and L2 regularization alongside dropout layers for effective regularization.

7. Transfer Learning: Consider leveraging pre-trained features from established models like VGG or ResNet, especially beneficial for smaller datasets.

8. Training Strategy: Define a comprehensive training strategy encompassing optimizer choice (e.g., Adam, SGD), learning rate schedule, and batch size. Continuous monitoring of training progress and validation metrics is essential to ensure model fidelity.

9. Evaluation Metrics: Select relevant evaluation metrics (e.g., accuracy, precision, recall, F1-score, ROC-AUC) aligned with plant disease detection objectives.

10. Hyperparameter Tuning: Iterate through various hyperparameter configurations, including learning rates and batch sizes, optimizing model performance through techniques like grid search or Bayesian optimization.

11. Model Interpretability: Incorporate interpretability techniques such as Grad-CAM or SHAP to gain insights into critical image regions influencing disease predictions.

12. Deployment: Prepare the trained custom architecture for deployment in real-world scenarios, whether on edge devices or cloud platforms.

The CNN, featuring two convolutional layers with 20 filter kernels followed by max-pooling layers, outperformed the MLP. Additionally, LSTM networks were employed for time series data processing, demonstrating superior accuracy compared to RF, SVM, and KNN . Utilization of ANNs to predict rice disease outbreaks based on weather data, has showcased the versatile usage of neural networks in agriculture.

Transfer Learning

Transfer learning emerges as a potent technique in enhancing the performance of plant disease detection models by leveraging pre-trained neural network models. Here's how transfer learning can be effectively applied:Pre-trained Model Selection: Begin by selecting a pre-trained neural network model trained on diverse datasets, such as VGG16 or ResNet, to serve as a base for transfer learning.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11547

1. Layer Modification: Remove the top layers responsible for original categorization while retaining feature extraction layers.

2. Custom Layer Addition: Append new layers to the pre-trained model to suit the plant disease detection task, typically incorporating fully connected layers for classification.

3. Pre-trained Layer Freezing: Freeze pre-trained layer weights to preserve learned features during training, ensuring effective adaptation to the new task.

4. Data Preparation: Organize and preprocess the plant disease dataset, applying suitable transformations and augmentation techniques to enhance model adaptability.

5. Fine-Tuning: Train the modified model on the plant disease dataset while keeping pre-trained layers frozen. Monitor training metrics to gauge model learning effectiveness.

6. Optional Unfreezing and Further Fine-Tuning: Unfreeze pre-trained layers after initial training rounds and continue training with a reduced learning rate to fine-tune feature extraction layers, enhancing model performance.

7. Model Evaluation: Evaluate the fine-tuned model using standard metrics like accuracy, precision, recall, and F1-score, ensuring robust performance assessment.

8. Deployment: Deploy the fine-tuned model in real-world applications, whether on edge devices or cloud platforms, for automated plant disease detection.

Transfer learning presents numerous benefits in plant disease detection:

- It mitigates the necessity for extensive datasets by leveraging knowledge acquired from pre-training.
- This approach often leads to expedited convergence and enhanced generalization, particularly when dealing with limited data.
- Transfer learning is apt for scenarios demanding real-time or near-real-time detection.
- It facilitates the utilization of established, cutting-edge architectures without the need to develop a bespoke network from scratch. Transfer learning stands out as a valuable technique for enhancing the precision and efficacy of plant disease detection models, rendering it a pragmatic option for precision agriculture and crop management.

In this paper SVM, RF, or deep learning-based ML models were discussed. These methodologies have demonstrated promising outcomes, underscoring the potential of ML techniques in disease and pest classification, detection, and prediction. SVMs exhibit robustness and efficacy in high-dimensional spaces owing to their utilization of the kernel trick. RF can forestall overfitting by employing numerous trees trained on diverse data subsets. Deep learning consistently yields superior classification outcomes by generating and extracting hierarchical features from inputs. Notably, deep learning outperforms other ML models, particularly in image classification domains, especially when employing pre-existing CNN architectures like Inception and ResNet. Despite the higher accuracy achieved by deep learning models, SVM and RF can also attain high accuracy rates exceeding 94%, particularly in disease classification using laboratory images . SVM achieves notable accuracy, surpassing 90%, in tomato disease detection. RNNs excel in establishing correlations between weather data and pest occurrence, outperforming RF and SVM .In situations where data acquisition is challenging, models trained with limited data can benefit from transfer learning, rather than being trained from scratch .Most studies pre-train their models on extensive image classification datasets such as ImageNet or COCO. Integrating the PlantVillage dataset with ImageNet for pre-training enhances model accuracy for disease classification in field-acquired images Transfer learning is typically applied by jointly training some of the top layers of the pre-trained model with the new classifier. Alternatively, addressing data scarcity issues can be accomplished through few-shot learning approaches.
C.Jackulin and S. Murugavalli [21] conducted an extensive review on the application of AI-based machine learning and deep learning techniques in plant disease detection. They undertook a comparative analysis of machine and deep learning methods, showcasing their performance and utilization across various research papers to underscore the superior effectiveness of deep learning models over machine learning counterparts. Tiago Domingues, Tomas Brandao, and Juao C. Ferrira [20] conducted a literature review focusing on ML techniques in agriculture, particularly in classifying, detecting, and predicting diseases and pests, with a spotlight on tomato crops. Their survey aimed to foster smart farming and precision agriculture by advocating for techniques that enable reduced pesticide use while enhancing crop quality and yield.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11547

Taranjeet Singh, Krishna Kumar, and SS Bedi [22] explored several techniques based on machine learning and image processing for crop disease recognition, presenting discussions that offer insights for advancements in this field. Tiago Domingues, Tomas Brandao, and Juao C. Ferrira [20] also furnished details on available databases in their work. Kartikeyan, P., and Shrivastava, G. [23] provided an overview of established methods for plant disease detection and classification.

Kamilaris, A., Kartakoullis, A., and Prenafeta-Boldú, F.X. [36] outlined the vast opportunities of big data analysis in agriculture toward smarter farming, highlighting advancements in hardware, software, techniques, and data sources. Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., and Zdeborová [31] discussed the applications of ML methods across various scientific domains, emphasizing recent successes, challenges, and advancements in computing architectures. Gianni Fenu and Francesca Maridina Malloci [24] analyzed and classified research studies focused on forecasting disease onset at pre-symptomatic stages or early stages, examining methodologies, pre-processing techniques, data sources, performance metrics, and encountered challenges. V. Sharma, A. Verma, and N. Goel [30] primarily focused on the most utilized classification mechanisms in plant disease detection, highlighting the superior accuracy of Convolutional Neural Network approach compared to traditional methods.

S. Arivazhagan, S. Newlin, S. Ananthi, and S.V. Varthini [39] proposed a system for automatic plant leaf disease detection and classification, presenting an efficient processing scheme that achieved a 94% accuracy rate in detecting and classifying diseases on a database of 500 plant leaves. Elakeyaa P V, Keerthana A, Bharathi P, Ezhilmani S, and Dr. V. Mohan [26] measured disease detection, infected area, and affected region percentage. Akhtar, Asma, A. Khanum, Shoab A. Khan, and A. Shaukat [40] compared the performance of various ML techniques for identifying and classifying plant disease patterns from leaf images, demonstrating the superiority of Support Vector Machines for disease classification.

Aqeel ur, R., Abbasi, A., Islam, N., and Shaikh, Z. [38] reviewed the use of wireless sensors in agriculture and WSN technology, discussing various system frameworks and ethical considerations of big data in agriculture. Liaghat, S., and Balasundram, S.K. [41] emphasized the utility of remote sensing technology in agriculture. Kondor, R., and S. Trivedi [35] explored the role of convolutional structures in achieving equivariance to the action of a compact group, leveraging concepts from representation theory and noncommutative harmonic analysis. Hopfield [42] discussed the emergence of computational properties useful for biological organisms or computers from systems with a large number of simple components. Noe, F., S. Olsson, J. Köhler, and H. Wu [32] delved into Boltzmann Generators, which employ neural networks to learn a coordinate transformation of the complex configurational equilibrium distribution.

Fenu, G., and Malloci, F.M. [24] presented a case study on forecasting potato late blight, employing models like the Negative Prognosis model and the Fry model. Fenu, G., and Malloci, F.M. [25] utilized regional weather variables to predict potato late blight risk, achieving high prediction accuracy using Machine Learning approaches. Bing Lu and others [27] provided insights into the strengths and limitations of hyperspectral imaging in agriculture applications, aiming to facilitate its adoption in agricultural research and practice.

Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y., and Srinivasan, K. [34] compared supervised and unsupervised machine learning models associated with crop yield, employing error measures like RMSE, RRMSE, MAE, and R2.

Bhatia and others [28] implemented Extreme Learning Machine (ELM) algorithm for plant disease prediction, focusing on real-time scenarios like Tomato Powdery Mildew Disease (TPMD). de Oliveira Aparecido and others [29] employed multiple machine learning algorithms for disease prediction, demonstrating the superiority of Random Forest Regressor in predicting various coffee diseases using weather conditions. Researchers have utilized deep learning-based image classification to determine disease severity based on the area of leaves covered by lesions, considering it as a measure of disease severity.

CONCLUSIONS

The exploration of ML-based techniques for disease and pest forecasting, detection, and classification presented valuable insights in this survey. It emphasized the importance of maintaining long-term records of datasets encompassing weather, diseases, and pests data. Time-series ML models, such as RNNs, emerge as effective tools for accurately predicting disease and pest occurrences based on series of meteorological measurements. Additionally, incorporating NDVI measurements can offer supplementary insights into crop development. Leveraging computer vision and deep-learning algorithms, particularly CNN models, proves advantageous for detecting and classifying pests

DOI: 10.17148/IARJSET.2024.11547

and diseases, outperforming traditional approaches reliant on manual feature extraction. However, the data requirements for deep learning models pose a challenge, necessitating strategies like transfer learning or few-shot learning. Despite their high performance under controlled conditions, further research is needed to analyze images captured in real-life field conditions. Moreover, the absence of substantial work on pest and disease forecasting using diverse data modalities underscores the need for future investigations. This article, by including the general references, aims to provide a comprehensive overview of ML techniques across various data types, fostering opportunities for addressing this research gap and advancing the field.

REFERENCES

1 E.-C. Oerke, H.-W. Dehne, Safeguarding production—losses in major crops and the role of crop protection, Crop Prot. 23 (2004) 275–285.

2. J.G.A. Barbedo, A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses, Drones 3 (2019) 40.

3. C. Bock, G. Poole, P. Parker, T. Gottwald, Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging, Crit. Rev. Plant Sci. 29 (2010) 59–107.

4. W.-S. Kim, D.-H. Lee, Y.-J. Kim, Machine vision-based automatic disease symptom detection of onion downy mildew, Comput. Electron. Agric. 168 (2020), 105099

Shmelkov K, Schmid C, Alahari K. Incremental learning of object detectors without catastrophic forgetting. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 3400-09.

5. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, pp. 1097–1105.

Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:1801 .00631.

6. B. Liu, Y. Zhang, D. He, Y. Li, Identification of apple leaf diseases based on deep convolutional neural networks, Symmetry 10 (2018) 11.

7. Hughes, D., & Salath´ e, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv 2015. arXiv preprint arXiv:1511.08060.

8.S.P. Mohanty, D.P. Hughes, M. Salath´ e, Using deep learning for image-based plant disease detection, Front. Plant Sci. 7 (2016) 1419

9. Q. Liang, S. Xiang, Y. Hu, G. Coppola, D. Zhang, W. Sun, Pd2se-net: computerassisted plant disease diagnosis and severity estimation network, Comput. Electron. Agric. 157 (2019) 518–529.

10. K.K. Singh, An artificial intelligence and cloud based collaborative platform for plant disease identification, tracking and forecasting for farmers. 2018 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), IEEE, 2018, pp. 49–56.

11. J.G.A. Barbedo, L.V. Koenigkan, B.A. Halfeld-Vieira, R.V. Costa, K.L. Nechet, C. V. Godoy, M.L. Junior, F.R.A. Patricio, V. Talamini, L.G. Chitarra, et al., Annotated plant pathology databases for image-based detection and recognition of diseases, IEEE Lat. Am. Trans. 16 (2018) 8431749–8431757.

12. D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, N. Batra, Plantdoc: a dataset for visual plant disease detection, in: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, 2020, pp. 249–253.

13. Wiesner-Hanks, E.L. Stewart, N. Kaczmar, C. DeChant, H. Wu, R.J. Nelson, H. Lipson, M.A. Gore, Image set for deep learning: field images of maize annotated with disease symptoms, BMC Res. Notes 11 (2018) 440.

14. J. Parraga-Alava, K. Cusme, A. Loor, E. Santander, Rocole: a robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition, Data Brief 25 (2019), 104414

15. H.B. Prajapati, J.P. Shah, V.K. Dabhi, Detection and classification of rice plant diseases, Intell. Decis. Technol. 11 (2017) 357–373.

16. D.O. Oyewola, E.G. Dada, S. Misra, R. Damaˇseviˇcius, Detecting cassava mosaic disease using a deep residual convolutional neural network with distinct block processing, PeerJ Comput. Sci. 7 (2021) e352.

17. W.-S. Kim, D.-H. Lee, Y.-J. Kim, Machine vision-based automatic disease symptom detection of onion downy mildew, Comput. Electron. Agric. 168 (2020), 105099.

20.Machine Learning for Detection and Prediction of Crop diseases and Pests: A Comprehensive Survey, Tiago Domingues, Tomas Brandao and Juao C. Ferrira Journal – Agriculture 12 (2022) 1350 Publisher: MDPI

21. A comprehensive review on detection of plant disease using machine learning and deep learning approaches C.Jackulin and S. Murugavalli :Measurement: Sensors 24 (2021) 100441 Publisher: Elsevier

22. A Review on Artificial Intelligence Techniques for Disease Recognition in PlantsTaranjeet Singh, Krishna Kumar and SS Bedi , IOP Conference Series: Materials Science and Engineering 1022 (2021) Publisher : IOP Publishing

23.Kartikeyan, P.; Shrivastava, G. Review on emerging trends in detection of plant diseases using image processing with machine learning. Int. J. Comput. Appl. 2021, 975, 8887.

24.Gianni Fenu and Francesca Maridina Malloci Forecasting Plant and Crop Disease: An explorative Study on Current

DOI: 10.17148/IARJSET.2024.11547

Algorithms Big data and cognitive computing 2021, 5,2

25.Fenu, G.; Malloci, F.M. LANDS DSS: A Decision Support System For Forecasting Crop Disease In Southern Sardinia. Int. J. Decis. Support Syst. Technol. IJDSST 2021, 13, 21–33.

26.Elakeyaa P V1 , Keerthana A2 , Bharathi P3 , Ezhilmani S4 , Dr.V. Mohan5 Plant Leaf Disease Detection Using Image Segmentation 2394-1588 International Advanced Research Journal in Science, Engineering and Technology Vol. 7, Issue 4, 2020

27.Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent advances of hyperspectral imaging technology and applications in agriculture. Remote Sens. 2020, 12, 2659

28.. Bhatia, A.; Chug, A.; Singh, A.P. Application of extreme learning machine in plant disease prediction for highly imbalanced dataset . J. Stat. Manag. Syst. 2020, 23, 1059–1068

29. de Oliveira Aparecido, L.E.; de Souza Rolim, G.; da Silva Cabral De Moraes, J.R.; Costa, C.T.S.; de Souza, P.S. Machine learning algorithms for forecasting the incidence of Coffea Arabica pests and diseases. Int. J. Biometeorol. 2020, 64, 671–688.

30. V. Sharma, A. Verma and N. Goel, “Classification Techniques for Plant Disease Detection”, International Journal of Recent Technology and Engineering (IJRTE) , 2020, vol-8 , Issue-6

31. Carleo, G.; Cirac, I.; Cranmer, K.; Daudet, L.; Schuld, M.; Tishby, N.; Vogt-Maranto, L.; Zdeborová, L. Machine learning and the physical sciences. Rev. Mod. Phys. 2019,

32. Noe, F., S. Olsson, J. Köhler, and H. Wu, 2019, Science 365, 1147.

33.. Fenu, G.; Malloci, F.M. An Application of Machine Learning Technique in Forecasting Crop Disease. In Proceedings of the 2019 3rd International Conference on Big Data Research, Paris, France, 20–22 November 2019; pp. 76–82.

34. Elavarasan, D.; Vincent, D.R.; Sharma, V.; Zomaya, A.Y.; Srinivasan, K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. Comput. Electron. Agric. 2018, 155, 257–282.

35. Kondor, R., and S. Trivedi, 2018, in International Conference on Machine Learning, pp.2747–2755, arXiv:1802.03690.

36 Kamilaris, A.; Kartakoullis, A.; Prenafeta-Boldú, F.X. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 2017, 143, 23–37.

37. Carbonell, I., 2016. The ethics of big data in big agriculture. Internet Policy Review, 5(1), pp. 1-13

38.. Aqeel ur, R., Abbasi, A., Islam, N. & Shaikh, Z., 2014. A review of wireless sensors and networks; applications in agriculture. Computer Standards & Interfaces, 36(2), pp. 263-270.

39. S. Arivazhagan, S. Newlin, S. Ananthi and S.V. Varthini, “Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features”, Agricultural Engineering International: CIGR Journal, 2013, vol.15, no.1,pp.211-217.

40.Akhtar, Asma, A. Khanum, Shoab A. Khan and A. Shaukat, "Automated Plant Disease Analysis (APDA): Performance Comparison of Machine Learning Techniques" , IEEE International Conference on Frontiers of Information Technology (FIT), pp. 60-65, 2013.

41.. Liaghat, S. & Balasundram, S. K., 2010. A review: The role of remote sensing in precision agriculture. American Journal of Agricultural and Biological Sciences, 5(1), pp. 50-55.

42. Hopfield, J. J., 1982, Proc. Natl. Acad. Sci. U.S.A. 79, 2554.

43.Mahlein, A. K. (2016). Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. Plant Disease, 100(10), 241-251.

44.Singh, A., & Ganapathysubramanian, B. (2015). Machine learning for high-throughput stress phenotyping in plants. Trends in Plant Science, 20(9), 522-531.

45.Sladojevic, S., Arsenovic, M., Anderla, A., & Culibrk, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience, 2016, 3289801.

46.Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

47.Bock, C. H., Poole, G. H., & Parker, P. E. (2010). Gottwald. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Critical Reviews in Plant Sciences, 29(2), 59-107.