NEW ERA AGRICULTURE MAGAZINE

Crop Saving with AI: Emerging Trends in Deep Learning for Plant Disease Diagnosis

Imran Khan[1], Narsing Laxmi Prasanna[2], Dharmappa D Chavan[3], Halima Khatoon[3] and Apurba Das[4]

Abstract: -

Plant diseases represent a critical challenge to agricultural productivity and global food security, potentially disrupting entire agricultural ecosystems. The emergence of deep learning technologies has opened new avenues for addressing these challenges, with researchers increasingly exploring machine learning techniques in precision agriculture. While various approaches have been developed for detecting and identifying plant diseases, each method comes with its own set of advantages and limitations. This survey aims to provide a comprehensive examination of the latest advancements in machine learning-based plant disease detection. By systematically analyzing prevalent datasets, emerging techniques, and their respective performance metrics, the research seeks to offer a holistic perspective on the current state of precision agriculture technologies. The paper critically examines the methodological landscape, highlighting not only successful strategies but also the inherent challenges in implementing these advanced detection systems. By presenting a detailed overview of deep learning techniques, their accuracy rates, and associated obstacles, the research aspires to serve as a valuable reference point for agricultural scientists and technology developers. Ultimately, the survey's objective is to stimulate future research and innovation, with the broader goal of developing more sophisticated and effective approaches to crop health management. The insights presented are intended to bridge current technological gaps and inspire more refined solutions in the ongoing battle against plant diseases.

Imran Khan[1], Narsing Laxmi Prasanna[2], Dharmappa D Chavan[3], Halima Khatoon[3] and Apurba Das[4]

[1]Ph.D. Research Scholar, Department of Irrigation and Drainage Engineering, SHUATS, Prayagraj, Uttar Pradesh

[2]Ph.D. Research Scholar, Department of Plant Pathology, Assam Agricultural University

[3]Ph.D. Research Scholar, Advanced Centre for Plant Virology, Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi -110012

[4]Assistant Professor, Department of Plant Pathology, College of Sericulture, AAU, Jorhat

Introduction

Agriculture is a critical sector globally, with over 58% of the world's population employed in agricultural industries. In India, approximately 70% of small households rely on agriculture, making plant health a crucial concern. Plant diseases pose significant threats to both farmers' livelihoods and food consumers. Traditional disease detection methods have been challenging, particularly for small-scale farmers. These approaches typically rely on visual inspection by experts, which can be expensive, time-consuming, and limited by the inspector's knowledge. Climate change has further exacerbated plant disease risks, according to the Food and Agriculture Organization. Historically, scientists used genetic screening and genomic approaches to understand plant-pathogen interactions. Researchers collected extensive data on infected plant behaviors and applied digital image processing to identify disease patterns. Recognizing the limitations of traditional methods, researchers began exploring automatic disease recognition techniques. Technological advancements have revolutionized plant disease detection. Machine Learning (ML) represents a computational approach where systems learn and adapt without explicit programming, using algorithms to analyze data patterns. Traditional ML techniques like Support Vector Machines, Naïve Bayes, random forest, and K-nearest neighbors were initially employed, but these approaches demonstrated limited effectiveness. The emergence of Deep Learning, a sophisticated ML subset, has transformed disease identification and classification. Deep Learning algorithms model complex data abstractions through multi-layered processing networks. Convolutional Neural Networks (CNN) have been particularly promising, demonstrating remarkable capabilities in image classification, object detection, and semantic segmentation. While Deep Learning techniques show significant potential, they require extensive training datasets, typically involving hundreds or thousands of images. This data-intensive nature represents both a strength and a challenge in developing robust plant disease detection systems.

Datasets

Datasets are fundamental to the success of deep learning models in plant disease detection. Comprehensive and diverse image collections are crucial for training models to accurately classify diseases. While the Plant Village dataset stands out as the most widely used publicly accessible resource for researchers, many studies have also developed specialized, smaller datasets targeting specific disease classifications. The significance of datasets lies in their ability to provide the necessary visual information and diversity required for machine learning algorithms to learn and distinguish between various plant disease manifestations. Researchers recognize that the quality and quantity of training images directly impact the performance and reliability of disease detection models. The Plant Village dataset has gained prominence due to its comprehensive collection of plant disease images, serving as a benchmark for many research efforts. However, the limitations of a generalized dataset have prompted researchers to create more focused, custom datasets that address specific regional, crop-specific, or disease-specific research requirements. These tailored datasets allow for more precise and contextually relevant disease classification models, enabling more nuanced and accurate diagnostic capabilities in precision agriculture.

Plant village

The Plant Village dataset, first published in 2015, represents a comprehensive collection of plant disease imagery. Containing 54,306 images, the dataset comprehensively documents both healthy and diseased leaves across various plant species. Each image is meticulously labeled with unique identifiers, specifying the plant species and its health status. Structured to support machine learning research, the dataset is strategically divided into predefined training and testing subsets. It encompasses 14 distinct crop types, further categorized into 38 distinct classes. Tomato emerges as the most prominent species, comprising 43.4% of the total images. The dataset's strength lies in its remarkable diversity of plant diseases, featuring 26 different disease representations. Early and late blight stand out as the most prevalent conditions, accounting for 15.6% and 14.9% of the images, respectively. For researchers and developers in agricultural technology, the Plant Village dataset offers an invaluable resource. Its extensive, well-labeled image collection and carefully structured subsets provide a robust foundation for developing and evaluating plant disease detection models. By facilitating advanced machine learning approaches, this dataset has the potential to significantly enhance agricultural diagnostics and crop management strategies.

PlantDoc

The PlantDoc dataset, developed in 2019, comprises 2,598 images showcasing both healthy and diseased plant leaves. Sourced from diverse platforms including Google and Ecosia, the dataset spans 13 different crop types and encompasses 17 distinct disease classifications. A distinctive feature of the PlantDoc dataset is its emphasis on real-field imaging conditions. By capturing plant images in authentic agricultural environments, the dataset offers a more genuine representation of the challenges encountered in practical plant disease detection scenarios. However, the dataset is not without its limitations. The compilation process revealed potential inconsistencies in image classification, attributed to a lack of specialized domain expertise. These classification inaccuracies could potentially compromise the performance of machine learning models trained using this dataset. The real-world imaging approach provides valuable contextual insights, but the dataset's reliability is somewhat tempered by the potential for misclassification. Researchers utilizing the PlantDoc dataset must exercise caution and potentially implement additional verification mechanisms to ensure the accuracy of their disease detection models.

Digipathos

The Digipathos dataset, created by Barbedo and colleagues, is a comprehensive collection focused on critical cash crops. The dataset contains 3,000 images spanning various economically important crops, including rice, coffee, soybeans, beans, maize, wheat, and other fruits. A key characteristic of this dataset is its extensive disease classification, identifying 171 distinct diseases across these crop types. The image collection methodology is predominantly laboratory-based, with most images captured under controlled environmental conditions. A small subset of the images represents real-field scenarios, providing a limited glimpse of agricultural settings. This approach ensures high-quality, standardized imaging while offering a minimal representation of real-world agricultural variations. By concentrating on cash crops and providing such a detailed disease classification, the Digipathos dataset offers researchers a valuable resource for developing sophisticated plant disease detection models. The comprehensive nature of the dataset allows for in-depth analysis and potential technological interventions in crop health management. However, the dataset's heavy reliance on lab-controlled imaging suggests that additional real-world context might be beneficial for creating more robust and adaptable disease detection algorithms.

PlantCLEF2022

PlantCLEF2022 represents a monumental dataset in plant research, encompassing an impressive collection of over 4 million images that span approximately 80,000 plant species. The dataset's unique composition derives from two primary sources: a meticulously curated set from the Global Biodiversity Information Facility (GBIF) and a supplementary collection gathered from web search engines like Google and Bing. To address potential data imbalance challenges, the dataset implementation includes a strategic limitation, capping the number of images at 100 per class, with an average of 36.1 images per class. This approach ensures a more uniform representation across different plant species. The extensive scale and diversity of PlantCLEF2022 make it an exceptionally valuable resource for machine learning research. By capturing the nuanced variations found in natural plant environments, the dataset provides researchers with a robust foundation for developing and testing advanced species classification models. The inclusion of a trusted GBIF-sourced image set adds a critical layer of reliability and accuracy to the dataset. This feature enhances the dataset's credibility and potential utility for scientific research, offering researchers a comprehensive and dependable resource for exploring plant species classification through advanced machine learning techniques.

Rice Leaf Disease dataset

The Rice Leaf Disease Image dataset represents a concise but thoughtfully structured collection of plant pathology imagery. Consisting of 120 images focused exclusively on infected rice leaves, the dataset demonstrates a carefully maintained balance across its disease classifications. The dataset is systematically organized into three specific disease categories: Bacterial Leaf Disease, Brown Spot Disease, and Leaf Smut. Each category is precisely represented with 40 images, ensuring an equal distribution across the different disease types. This uniform representation is crucial for developing reliable machine learning models. Despite its modest size, the dataset offers researchers a targeted and focused resource for rice disease detection and classification. The collection's strength lies in its real-world image representation, providing a practical foundation for training and evaluating machine learning algorithms in agricultural diagnostic technologies. The balanced nature of the dataset, combined with its representation of actual disease manifestations, makes it a valuable tool for researchers seeking to develop precise and effective plant disease identification models, particularly in rice crop management.

References

1. Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst. Eng. 144, 52–60. doi: 10.1016/ j.biosystemseng.2016.01.017
2. Barbedo, J. G. A., Koenigkan, L. V., Halfeld-Vieira, B. A., Costa, R. V., Nechet, K. L., Godoy, C. V., et al. (2018). Annotated plant pathology databases for image-based detection and recognition of diseases. IEEE Lat. Am. Trans. 16, 1749–1757. doi: 10.1109/TLA.2018.8444395

3. Jackulin, C., and Murugavalli, S. (2022). A comprehensive review on detection of plant disease using machine learning and deep learning approaches. Meas. Sensors 24, 100441. doi: 10.1016/j.measen.2022.100441

4. Jhajharia, K., and Mathur, P. (2022). A comprehensive review on machine learning in agriculture domain. IAES Int. J. Artif. Intell. 11, 753–763. doi: 10.11591/ ijai.v11.i2.pp753-763

5. Karam, C., Awad, M., Abou Jawdah, Y., Ezzeddine, N., and Fardoun, A. (2023). Ganbased semi-automated augmentation online tool for agricultural pest detection: a case study on whiteflies. Water-energy-food-health solutions Innov. low-carbon climateresilient drylands 13, 137. doi: 10.3389/fpls.2022.813050.