GPH-International Journal of Applied Science
(Online ISSN 3050-9653)
Volume 07 | Issue 01 | January 2024
Published by: Global Publication House
DOI: https://doi.org/10.5281/zenodo.20363232
Available online at: www.gphjournal.org
Journal homepage: https://gphjournal.org/index.php/as/index

Plant Disease Detection Using Combined Deep Learning and Machine Learning Techniques

Ge, T.*[1], Adom, I. T.[1], Uhenda, S.[1], Agber, S.[1]

[1]Department of Mathematics and Computer Science, Benue State University, Makurdi, Nigeria.
*Corresponding author: tivlumunge@yahoo.com

Abstract

Effective disease identification in plants is param bount for the success of farming systems. Traditionally, farmers rely on naked-eye observations to recognize disease symptoms in plants, necessitating continuous monitoring. However, this approach becomes cost-prohibitive in large plantations and may be prone to inaccuracies. In certain regions, farmers may need to consult experts by presenting specimens, resulting in time-consuming, less efficient and expensive processes. This study presents a comparative analysis of machine learning and deep learning techniques for automated plant disease detection. Convolutional Neural Network (CNN), Random Forest Algorithm (RFA), and Support Vector Regression (SVR) classification approaches were implemented and evaluated. The proposed system was integrated into a Flask-based web application that enables users to upload plant leaf images for disease classification. Experimental results reveal that the CNN model achieved the highest performance with an accuracy of 82.68%, outperforming Random Forest (74.46%) and SVR (26.84%). Additional evaluation metrics obtained for the CNN model include Precision (0.8610), Recall (0.8442), and F1-Score (0.8417). The findings demonstrate that deep learning techniques, particularly CNNs, provide superior capability for plant disease identification and can contribute significantly to precision agriculture.

Keywords:

Plant Disease Detection, Deep Learning, Machine Learning, Precision Agriculture, Disease Classification.

INTRODUCTION

Plant diseases are influenced by climatic conditions, for example, the appearance of reddish–purple leaf spots on the older leaves of the Ixora plant can be attributed to the entomosporium leaf spot fungus, a perilous condition that thrives in cool and wet weather. Various other diseases, such as rust, kole roga, yellow leaf disease, leaf rot, leaf curl, angular leaf spot, leaf spot, late blight, bacteria wilt, significantly impact plant growth. The combination of global warming, deforestation, pollution, and industrialization poses a threat to many plant species, with plant diseases playing a substantial role in their decline. Anthracnose, characterized by dark, water-soaked lesions on stems, leaves, or fruit, is one such disease (Alagumariappan et al., 2020). Agriculture holds significant importance in the economic development of numerous countries. As highlighted by the Food and Agriculture Organization (FAO), a specialized agency of the United Nations, "a good percentage of the population in most developing countries are involved in the agriculture sector, primarily at a subsistence level." Consequently, harm to crops could result in substantial productivity losses, exerting a direct impact on the economy (Sandhu and Kaur, 2019). There are compelling reasons to measure or evaluate the prevalence of diseases in plants. Gaining insights into the extent of the disease is crucial for timely management decisions, especially when the disease is closely tied to potential yield losses. In the context of plant breeding, it is imperative to assess crops for their resistance and susceptibility to diseases. In the realm of crop protection, accurately gauging plant resistance is essential for the judicious application of pesticides, ensuring cost-effectiveness. The examination of disease and symptom severity plays a pivotal role in addressing fundamental questions in plant stress biology (Fox et al., 2006; Martinelli et al., 2015). In recent decades, extensive research efforts have been dedicated to uncovering the intricate dynamics between the plant immune response and pathogens. The wealth of data generated from these investigations stems from significant advancements in genomics and proteomics. Traditionally, scientists utilized large-scale genetic screening and genomic approaches to pinpoint genes and proteins of interest. However, with the emergence of machine learning algorithms, a set of analytical methods automating the model-building process and iteratively learning from data to glean insights without explicit programming, more robust and efficient tools have become available. These tools not only facilitate the identification of genes and proteins involved in plant-pathogen interactions but also enable the classification of plant diseases based on images of infected leaves (Harakannanavar et al, 2022). The advent of precision agriculture introduces innovative technologies to enhance farm output. Leveraging these advanced techniques holds the potential for substantial economic growth in agriculture. Precision agriculture encompasses various applications, including pest detection, weed identification, crop yield prediction, and plant disease detection. Farmers commonly employ pesticides to control pests, prevent diseases, and increase crop yield. Diseases in crops pose significant challenges, leading to reduced production and economic losses for farmers and agricultural industries. Hence, the identification of diseases and their severity has become a critical necessity (Shruthi et al., 2019).

Machine learning, the scientific discipline focused on programming computers to learn from data, has found extensive applications in both classical and in vitro-based plant breeding studies. Its role extends to interpreting the flow of information about plants, spanning from DNA sequences to observed phenotypes. Machine learning methods can be classified in three ways: firstly, into supervised and unsupervised models; secondly, into linear and nonlinear algorithms; and thirdly, into shallow and deep learning models (Niazian and Niedbała, 2020). The integration of diverse machine learning techniques has ushered in a new era in plant disease detection, significantly elevating the accuracy and efficiency of identification processes. Methodologies such as the CNN classifier, SVM classifier, FUZZY classifier, Deep Learning, and ANN classifier collectively contribute to the field's advancement. These techniques, each playing a crucial role, provide a comprehensive toolkit for nuanced and precise disease identification. The CNN classifier excels in image-based tasks, capturing intricate patterns for accurate classification. SVM ensures robust categorization through optimal decision boundaries, while the Fuzzy classifier adds adaptability to handle uncertainty. Deep Learning techniques, including CNNs, revolutionize the field by automatically learning complex data representations. Inspired by the human brain, the ANN classifier deciphers intricate relationships for subtle pattern identification. Together, these machine learning techniques empower plant disease detection systems, addressing diverse challenges and enhancing agricultural practices (Shruthi et al., 2019). Recent years have witnessed the evolution of computational systems, notably Graphical Processing Units (GPUs) embedded processors. This evolution has propelled the exponential growth of Machine Learning-related Artificial Intelligence applications, giving rise to innovative methodologies and models, which collectively form the burgeoning field of Deep Learning (Ferentinos, 2018). A comprehensive review synthesizes various plant disease detection techniques, as conducted by Sandhu and Kaur in 2019.This review critically analyzed studies from multiple authors, covering diverse methodologies and outcomes, providing invaluable insights for researchers in the field of plant disease identification. The study explored a range of techniques, including k-means clustering, genetic algorithms, and color difference-based algorithms. Notable findings highlighted the effectiveness of k-means clustering in handling large datasets and the application of genetic algorithms for clustering unlabeled points. The review placed significant emphasis on the importance of feature extraction methods, such as color histograms, morphology, and texture-oriented features.Each study contributed novel approaches to plant disease detection, showcasing achievements in accuracy, system efficiency, and diagnostic capabilities. The proposed algorithms demonstrated promising results, with overall system accuracies ranging from 82% to an impressive 96%. The utilization of advanced techniques like SVM and BPNN in classification yielded high accuracies, addressing the trade-off between training time and precision.

Alagumariappan et al., (2020) presented an advanced system for plant disease identification. The research centers on extracting informative features, utilizing Hu moments and Haralick texture to describe leaf outlines and quantify leaf texture, respectively.The Hu moments methodology, producing seven invariant numbers to various image transformations, and Haralick texture, derived from a gray level co-occurrence matrix (GLCM), distinguish healthy and diseased leaves based on texture variations. The study adopts SVM with linear and polynomial kernels for binary classification, achieving accuracy by selecting an efficient hyperplane. The research explores three machine learning algorithms, including SVM with linear kernel (LSVM), SVM with polynomial kernel (PSVM), and Extreme Learning Machine (ELM). Notably, the accuracy of the extreme learning machine surpasses the other two classifiers. The study emphasizes the performance of SVM and ELM in detecting three different plant diseases. Miller et al. (2019) proposed a comprehensive methodology for identifying plant diseases through a series of steps including preprocessing, feature extraction, and classification. Image preprocessing involves resizing to a standardized format. Feature extraction incorporates Histogram of Oriented Gradients (HOG) descriptors along with Hu moments and Haralick texture to capture shape and texture attributes. The algorithm employs a robust random forests classifier known for its flexibility and high accuracy compared to other machine learning techniques. This classifier proves effective even with a relatively small dataset. The study emphasizes the significance of distinguishing healthy and diseased leaves based on texture variations. In the study conducted by Xin Yang and Tingwei Guo (2017), machine learning methods take the forefront in identifying plant diseases, with a primary focus on disease detection and classification. Notably, these methods were frequently employed to detect Huanglongbing (HLB) in citrus trees. In the study by Niazian and Niedbała (2020), the application of different machine learning algorithms is explored for plant breeding studies. The authors emphasize the importance of comparing algorithm performance to identify the most suitable one for specific datasets. In the context of tea plants (Camellia sinensis L.), partial least squares discriminative analysis (PLS-DA) and least squares-support vector machines (LS-SVM) were compared for classifying different nitrogen nutrition statuses under field conditions. LS-SVM outperformed PLS-DA in correct classification. The study provides a detailed comparison of pros and cons for different machine learning and deep learning methods, including ANNs, CNNs, SVMs, and RF. Each algorithm is evaluated based on its advantages and disadvantages in classical and in vitro-based plant breeding studies. The study highlights the limitations of classical statistical methods in assessing plant biological features, emphasizing the need for considering various influential factors involved in plant responses to environmental stress.

Collectively, these entire studies underscore the diversity and limitations of methodologies and techniques in the field of plant disease detection. From feature extraction using advanced algorithms to the application of machine learning and deep learning models, these studies contribute significantly to the evolving landscape of intelligent plant disease identification systems while providing the foundation for future research and improvements.

METHODOLOGY

This study employs a structured approach to investigate and analyze plant disease detection techniques. The methodology unfolds in three key stages: literature review, empirical studies, and a comparative analysis of diverse machine learning algorithms. The aim is to extract valuable insights into the effectiveness of different methodologies, contributing significantly to the advancement of diagnostic capabilities in the field of plant pathology. Data on plant disease detection from Kaggle typically include image datasets of various plant species affected by different diseases. These datasets often consist of thousands of labeled images, where each image is associated with a specific plant disease or a healthy state. The images are captured under different conditions, such as varying lighting, angles, and backgrounds, to simulate real-world scenarios. The labels provided with the images specify the type of plant, the type of disease (if any), and sometimes additional metadata such as the severity of the disease or the stage of infection. These datasets are meticulously curated to ensure accuracy and consistency in labeling. The primary source of data for this project is the Plant Village dataset (PlantVillage), obtained from Kaggle. Plant Village provides a diverse collection of images specifically focused on Bell Pepper, Tomato and Potatoes leaves, encompassing various stages of plant health such as Bell Pepper Bacterial Spot, Bell Pepper Healthy, Potato Early Blight, Potato Healthy, Potato Late Blight, Potato Target Spot, Tomato Mosaic Virus, Tomato Yellow Leaf Curl Virus, Tomato Bacterial Spot, Tomato Early Blight, Tomato Healthy, Tomato Late Blight, Tomato Leaf Mold, Tomato Septoria Leaf Spot, Tomato spider Mites and two Spotted Spider Mite. The diagram (figure 1.0) below represents a flowchart of the model system components and processes involved in building the system.

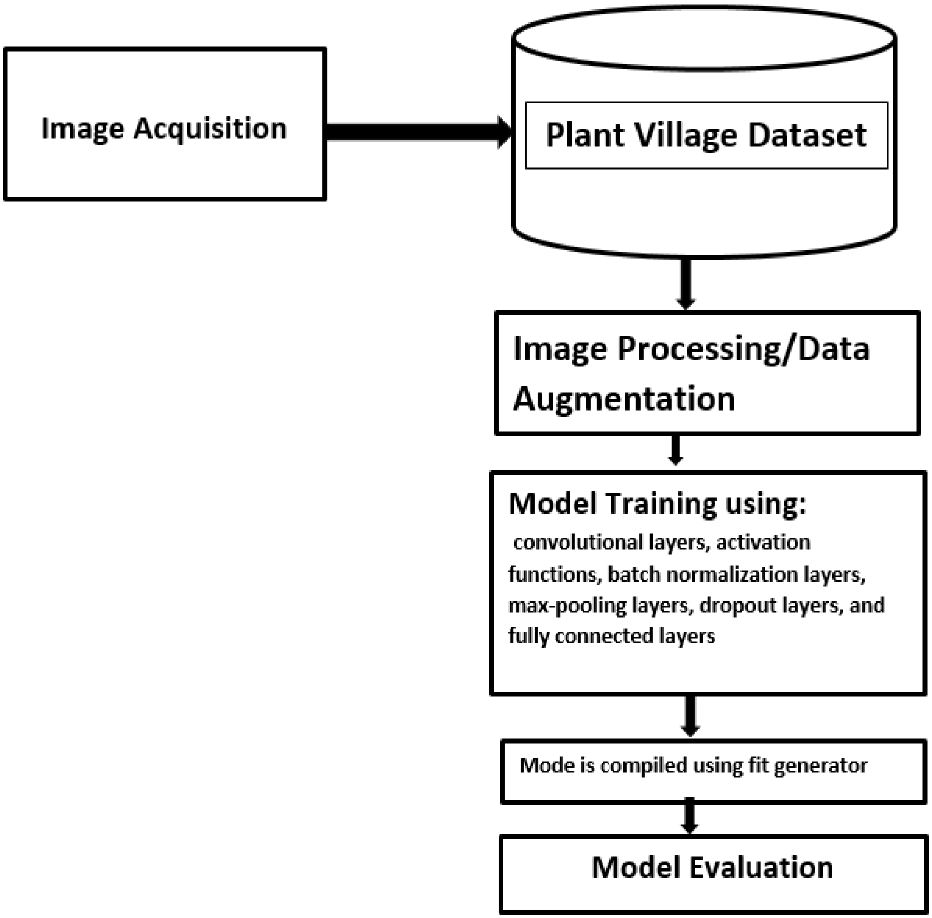

Figure 1.0: The System Model Architecture

System Algorithm

The algorithm to design and implement a plant disease identification using combined machine learning and deep learning involves the following steps:

Step 1. Start

Step 2. Download Plant Village dataset from kaggle.

Step 3. Upload dataset to google drive and load it in google Colab.

Step 4. Design your classification model.

Step 5. Export the model and import it into visual studio code.

Step 6. Set up a python flask web page.

Step 7. Connect the imported model with the flask server.

Step 8. Link the flask server with the client side for user accessibility.

Step 9. Evaluate the system performance and readability to assess the system accuracy and prediction state.

Step 10. End.

Interface Design

The graphical user interface (GUI) serves as a pivotal component in facilitating user interaction with the plant disease detection system. It provides users with a user-friendly platform to access and utilize the system's functionalities effectively. In the interface design phase, Python Flask is utilized to build the web application's frontend. Flask, a lightweight and flexible web framework, enables rapid development of web applications with minimal overhead.

RESULT AND DISCUSSION

The System working Procedure

The user visits the web browser with the provided link, click a button and choose a plant leaf image from the local machine, uploads the image and click on the predict button which will trigger the model from the server to accurately classify the uploaded image using SVM, RFA and CNN algorithm and return the predicted result to the client user which is the plant health state. The developed model based on the three classification techniques of CNN, Random Forest and SVR gave 82.6%, 74.4% and 26.8% accuracies respectively.

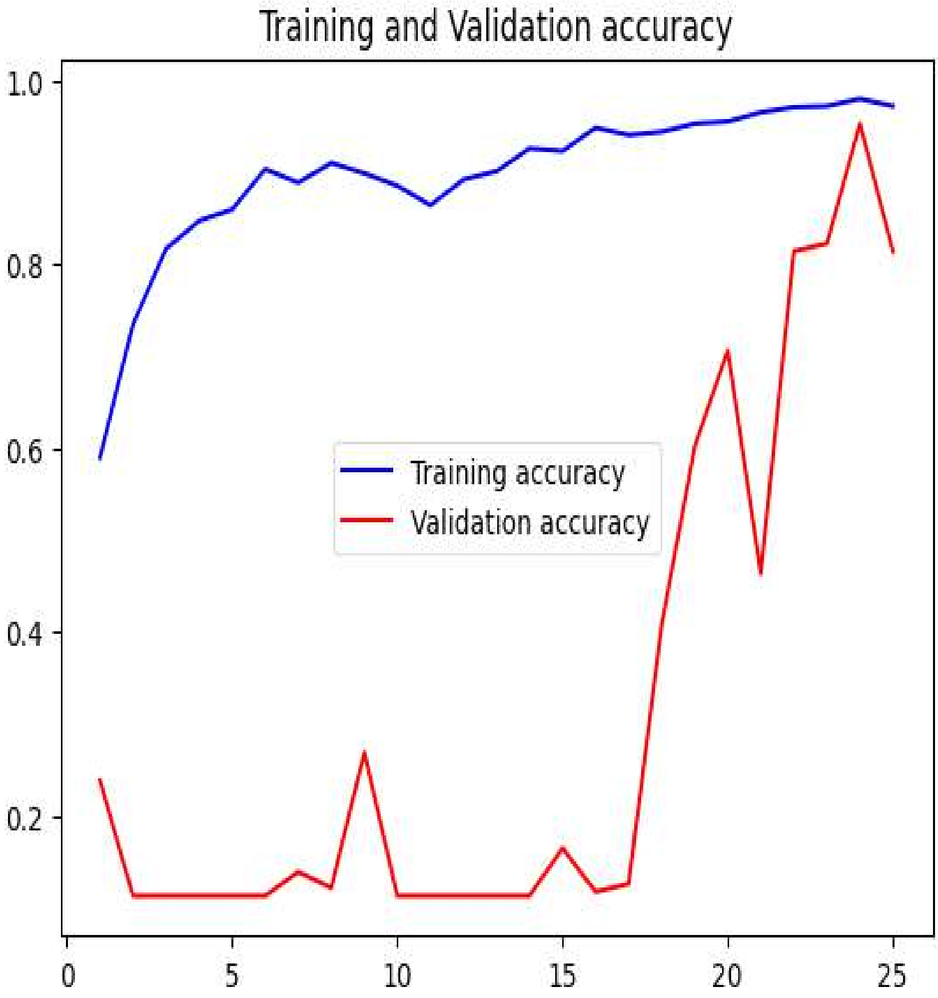

Figure 2.0: Training and Validation Accuracy Graph

Figure 2.0 above depicts the Training and Validation Accuracy Graph generated during the training of the machine learning model. This graph illustrates the performance of the model over twenty-five epochs, with the x-axis representing the number of training iterations (epochs) and the y-axis representing the accuracy which is approximately 84%. We also have Precision: 0.86, Recall: 0.84, and F1 Score: 0.84. The blue line graph represents the training accuracy, which measures the model's performance on the training data during each epoch. As training progresses, the training accuracy typically increases, indicating that the model is learning and improving its ability to classify examples in the training dataset. The red line represents the validation accuracy, which evaluates the model's performance on a separate validation dataset that it has not seen during training. Initially, the validation accuracy lags behind the training accuracy, but it gradually increases as the model learns to generalize to new data. As we can see from our graph above the validation accuracy starts from a low value and begins to increase between epochs 8 and 9, decreases again and begins to increase around epoch 17. This upward trend suggests that the model is starting to generalize well to unseen data, as evidenced by the improved performance on the validation dataset.

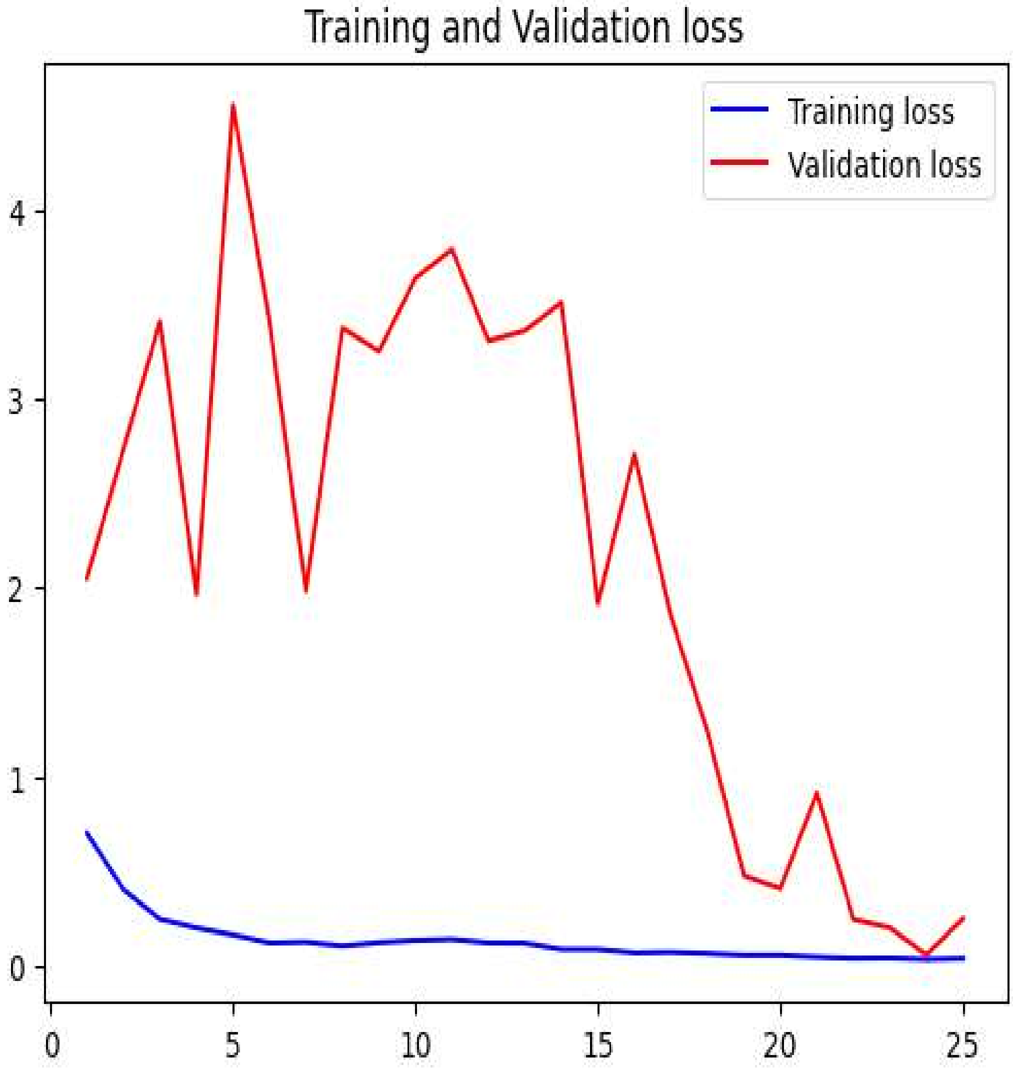

Figure 3.0: Training and Validation Loss Graph

Figure 3.0 displays the Training and Validation Loss Graph generated during the training the machine learning model. This graph illustrates how the loss, or error of the model changes over the course of training epochs. The x-axis represents the number of training epochs which in our case it's 25, while the y-axis represents the loss value. The loss value indicates how well the model's predictions match the actual targets during training. Lower loss values indicate better agreement between predicted and actual values. In the graph, the blue line represents the training loss, which measures the error on the training dataset at each epoch. As training progresses, the goal is to minimize the training loss, indicating that the model is effectively learning the patterns in the training data. The red line represents the validation loss, which measures the error on a separate validation dataset that the model has not seen during training. The validation loss helps assess how well the model generalizes to unseen data. A decreasing validation loss indicates that the model is learning to generalize well and is not overfitting to the training data. Initially both training and validation loss are somehow high as the model learns to make predictions. As training continues, the training loss typically decreases, while the validation loss also decreases.

CONCLUSION

This project focuses on the development of a machine learning-based plant disease identification system to enhance agricultural practices. By leveraging advanced techniques such as Convolutional Neural Networks (CNNs), Random Forest, and Support Vector Regression (SVR), the system accurately identifies plant diseases based on leaf symptoms. The study utilizes the Plant Village dataset for model training and evaluation, incorporating data augmentation techniques for enhanced performance. Through the implementation of a user-friendly interface, farmers can easily upload images for disease prediction, empowering them with timely and precise diagnostic tools. Based on the three models developed, CNN gave the best accuracy and prediction. By harnessing the power of advanced algorithms and the availability of comprehensive datasets, the developed system showcases promising results in disease diagnosis. The integration of user-friendly interfaces enhances accessibility, facilitating widespread adoption among farmers and agronomists. Moving forward, continued research and development in this field hold the potential to further optimize agricultural practices, mitigate economic losses, and ensure food security in the face of evolving environmental challenges

REFERENCES

Alagumariappan, P., Dewan, N. J., Muthukrishnan, G. N., Raju, B. K. B., Bilal, R. A. A., & Sankaran, V. (2020). Intelligent plant disease identification system using Machine Learning. *Engineering Proceedings*, *2*(1), 49.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture, 145, 311-318.

Fox, R. T. V., & Narra, H. P. (2006). Plant disease diagnosis. In *The epidemiology of plant diseases* (pp. 1-42). Dordrecht: Springer Netherlands.

Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., & Pramodhini, R. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings*, *3*(1), 305-310.

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P. & Dandekar, A. M. (2015). Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*, *35*, 1-25.

Miller, S. A., Beed, F. D., & Harmon, C. L. (2009). Plant disease diagnostic capabilities and networks. *Annual review of phytopathology*, *47*, 15-38.

Niazian, M., & Niedbała, G. (2020). Machine learning for plant breeding and biotechnology. *Agriculture*, *10*(10), 436.

Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Shashank, N., & Vinod, P. V. (2018, April). Plant disease detection using machine learning. In *2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C)* (pp. 41-45). IEEE.

Sandhu, G. K., & Kaur, R. (2019, April). Plant disease detection techniques: a review. In *2019 international conference on automation, computational and technology management (ICACTM)* (pp. 34-38). IEEE.

Shruthi, U., Nagaveni, V., & Raghavendra, B. K. (2019, March). A review on machine learning classification techniques for plant disease detection. In 2019 5th International conference on advanced computing & communication systems (ICACCS) (pp. 281-284). IEEE.