A Hybrid Deep Learning and Machine Learning Model to Multicrop Disease Detection and Fertilizer Recommendation on Leaf Images and Soil Nutrients

Prof. Manishkumar Dubey
Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai, India
Dby.manish@gmail.com

Vaishnavi Deshmukh
Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai, India
vaishnavideshmukh636@gmail.com

Deepak Chauhan
Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai, India
deepakchauhan887972@gmail.com

Abstract—**The evolving climatic condition and the existing agricultural techniques such as grey soils have negative effects on incorporation of sustainable agricultural systems in the world. Timely and appropriate diagnosis of plant disease results in the reduction of crop wastage and maximum production. The comparative research design appeared in this work because we used artificial intelligence to identify crop diseases and the recommended type of fertilizers with the help of leaf images and the content of soil nutrients. We collected the crop leafs and surface analysis nitrogen, phosphorus and potassium in soil and used deep and machine learning. process algorithms with the data. Deep learning techniques used on. match diseases and classify fertilizers convolutional. neural networks (CNNs) and randomly forested.The performance indicators that we compared the models based on include. as recovery, accurateness, recall and F-measure. Findings indicated that CNNs detected diseases more successfully than traditional approaches and Random Forest made excellent suggestions of fertilizers. An online tool of identifying real-time crop diseases based on a Flask-based application was created with the focus on demonstratingthe significance of AI-based decision support DSS (Systems) in enhancing crop productivity and efficient resources distribution.**

I. Introduction

One of the major risks worldwide is crop diseases, which diminishes the yields, influences the food supply, and income of farmers. Research indicates that fungus and bacteria will increase as a result of climatic changes, change in temperature and humidity[9].In other nations such as India where farming is a major source of income, farmers continue to stick to traditional farming technologies, thus making it hard to diagnose in time and low productivity[17].

Innovations in AI in the agriculture field have increased dramatically like crop health monitoring and automated disease detectors. According to a study such as that by Mohanty et al[7],machine-based deep learning models were very accurate in diseases classification. On the same note, Ferantinos[5]demonstrated that deep learning is more successful than with the standard machine learning to identify the non-uniform crops. Random Forests and Support Vector Machine can also offer some useful results in detection of an illness and in prediction of yield[10].

Integration of image data and soil data are more informative and can be used to make better decisions. Research shows the necessity of the combination of the two because the nutrients in soil (N, P, K, etc.) have a direct impact on the plant health and its yield. Survival and adaptation to the changes in climate require disease resistance and adaptative capability to support the survival and growth of plants [15][16].

An example of ensemble machine learning (e.g. Randdom Forest, majority voting) has been shown to be effective in agriculture, frequently outperforming individual models. These overlapping algorithms such as ANN, SVM and others are used to enhance better prediction which is more reliable[12][13]. The next generation development is to address web based as well as mobile based in order to transform complex models into tools that are easy to use. These systems enable farmers to diagnose disease, administer fertilizer choices and treatment suggestions in interactive dashboards in real-time in the absence of any technical skills they may have had[18][19].

A. Motivation

There are several main reasons for developing a Crop Disease Detection and Fertilizer Recommendation System:

- Farmers are losing crops from late detection of fungal or bacterial diseases that spread quickly in the right environment. [6].
- Farmers rely heavily on manual methods of inspection and traditional methods of control; therefore, they may misidentify disease and apply incorrect control measures.[3].
- Many rural areas have little availability of agricultural expertise and extension services to assist farmers in obtaining timely and accurate information. [5].
- Many farmers are not aware of their soil nutrient levels so they improperly apply fertilizers or use too much/too little fertilizer, reducing crop yield, and degrading their soils. [9].
- There are currently no integrated decision-support systems that provide both plant disease detection and fertilizer recommendations in one place. [8].
- Farmers have been slow to adopt artificial intelligence (AI) tools in day-to-day farming because of the complex-

ity of the tool, lack of access to tools, and low levels of digital literacy among farmers. [7].

B. Objectives

The Crop Disease Detection and Fertilizer Recommendation System was proposed with various main goals:

1) Designing a crop health-analysis system that relies on data-driven decisions for farmers by providing accurate recommendations based on deep learning technology combined with machine learning technology.
2) Creating an automatic detection system built with analysis of leaf photos through CNN to determine the disease from the image of the leaf. It can also automatically classify the image to save time and give high accuracy versus a manual system for identifying disease.
3) Creating a Random Forest (RF) model to develop a fertilizer recommendation system based on the amount of N, P, K in the soil. This will help maintain a balance in soil nutrient content and provide healthy crops throughout all stages of growth.
4) Unifying (all disease detection and fertilizer recommendation) to form a complete yield-management system. This unification provides the farmer with full assistance for managing their crop: visually identifying the crop's disease and allowing the farmer to determine which fertilizer to use.

II. Related Work

Researchers have attempted to implementing different techniques in detecting plant diseases, managing crops, minimising productivity losses related to such diseases, and increasing the effectiveness of precision agriculture techniques.

A. Plant Disease Identification via Deep Learning

The experiments conducted by Pawar and Patil [1] on CNNs proved that they are a great way to automatically detect plant diseases under the conditions in India. CNNs have been shown to perform much more accurately than other traditional image processing algorithms. Kumar and Ghosh [2] also built a smartphone app able to identify diseases based on images captured in real-time, demonstrating that AI can be practically used by farmers while they are working in their fields.

Researcher Sladojevic et al. [4] introduced a neural network model to divide leaf images as having different diseases, and they found that their model performed very well across many different diseases affecting different plants. Ferentinos [5] did further testing of models to diagnose plant disorders, found that CNN-based models performed much better than classical machine learning approaches when using images as input.

Singh et al. [6] showed improved accuracy of deep CNNs for detecting tomato leaf diseases in different environments. Mohanty et al. [7] undertook a large-scale experiment using the PlantVillage dataset and found that deep learning models work well when classifying diseases for each of the crop species tested. Brahimi et al. [8] conducted experiments on classification of tomato diseases using deep learning approaches and obtained very good results with respect to also being able to see which parts of a leaf were affected by a disease. The research conducted by Kamilaris and Prenafeta-Boldu´ [10] was a detailed analysis of the various deep learning applications that have been implemented in agriculture, and showed that convolutional neural network (CNN) models outperformed traditional techniques for crop disease identification.

B. Use of Traditional Machine Learning for Crop Evaluation

Many researchers have used traditional machine learning methods before the introduction of deep learning to identify the occurrence of diseases on plants. One of the examples provided by the Tata Institute of Fundamental Research [3] was using classical machine-learning methods such as Support Vector Machines and Decision Trees for predicting the occurrence of crop diseases.

Traditional classifiers like K-Nearest Neighbors, Random Forest, and Support Vector Machines have been used extensively using handcrafted features from images; however, they achieved moderate-to-good accuracy (less than 80%) in classifying plant diseases [4], [6]. Some studies have found Random Forest models to be helpful for various decisions made within agriculture, including providing fertilizer recommendations, and evaluating the nutrients contained in soil because they have been previously verified to be robust and interpretable decision-making tools [5], [6].

C. Agricultural Decision Support Systems

Today's research has made a major move to include AI models in just-in-time, real-time, decision support systems for use by farmers. The Food and Agricultural Organization (FAO) agricultural report [9] noted that early detection of disease and the use of AI-generated recommendations to improve crop yields and eliminate financial losses from disease were critical in supporting crop yield production.

There are several systems that integrate image-based detection of disease together with soil nutrient analysis to provide full crop management capabilities. Such systems provide farmers with not only the capability to diagnose and treat disease, but also offer fertilizer recommendations via the web or mobile media access through AI tools to help their crops be more productive [2], [7], [9].

The results of this work demonstrate the importance of integrating deep learning for disease detection with traditional ML models for soil nutrient management; however, the vast majority of systems available today focus either on disease detection or fertilizer recommendations as separate tools. Therefore, this work proposes a new model that combines these two important functions into one unified application.

III. Problem Statement

Agricultural productivity is hampered due to the challenges of plant pests and diseases as well as improper use of fertilizers for small and mid-sized producers. Crop health problems many times go undetected until large affected areas become evident and crop damage has already occurred. Conventional ways to diagnose plant pests and diseases are typically time-consuming, subjective, and untrustworthy as they rely on the manual inspection of crops by farmers and the expertise of agricultural experts. There are several key challenges this project addresses:

- **Late Disease Detection:** Crop diseases are usually only detected after damage has become apparent on a large percentage of leaves and/or stems. Farmers normally rely on their own experience or reports from agricultural technicians, leading to delays that reduce the effectiveness of control measures and increase production losses.
- **Reliance on Manual Diagnosis:** The traditional way that diseases are diagnosed by farmers is using visual inspection by agriculture experts or by sending plant samples to a laboratory. Many farmers in rural and remote locations cannot easily access such services, leading to misdiagnosis and ineffective pesticide application.
- **Inappropriate Fertilizer Use:** Because farmers often apply fertilizer according to generic recommendations rather than soil testing, crops may be under- or over-fertilized for nitrogen, phosphorus, and potassium, resulting in lower yields, degraded soils, and wasted input expenses.
- **Lack of Integrated Management Systems:** Current solutions available to farmers address either disease diagnosis or fertilizer recommendations separately. Farmers are required to use multiple tools, making effective crop management decisions difficult without a unified system.
- **Limited Access to Data-Driven Tools:** Most crop production decisions made by farmers are based on experience rather than scientific analysis. Intelligent systems that can process real-time crop images and soil data are not readily available for timely decision-making.
- **Difficulty Understanding AI Solutions:**Difficulty Understanding AI Solutions: Although a number of machine learning systems achieve very high rates of accuracy, due to their complexity, they often cannot easily be interpreted by farmers or utilized as intended because of the absence of effective interfaces and intuitive recommendations.
- **Limited Use in Real-World Applications:**Many of the systems developed for plant disease detection remain at either prototype or research levels with no meaningful application or testing in the field and therefore lack true impact as a result within the agricultural sector.
- **No Support for Preventative Actions:**If a farmer cannot quickly identify a problem, all that he/she can do is perform a corrective action after a crop has been damaged, which increases input costs and decreases total production. Early identification of a problem and proper fertilizer recommendations are essential to preventing crop losses.

IV. Proposed Plant Disease Detection and Fertilizer Recommendation System

A. Overall System Design

The overall system architecture consists of four layers:

1) **Data Collection** – Collects leaf images and soil nutrient information for raw input data to the system.
2) **Data Pre-processing and Image Processing** – Deals with preparing the data for model training (correcting quality issues of the collected image data by resizing, normalizing, augmenting and cleaning up of the soil nutrient data).
3) **Various Models** – Applies CNN to classify images of diseased leaves and applies Random Forest to recommend fertilizer.
4) **UI and Application Layer** – Provides access to the system via an interactive web application powered by the Flask framework which allows the user to upload pictures of their leaves, input NPK values, and receive live predictions and recommendations for treatment.

B. Proposed System Modules

The proposed system is composed of five main modules:

1) **Soil/Image Data Pre-processing** – Cleans image files, fills in missing soil data, normalizes inputs and applies augmentation techniques to all inputs to allow for improvement in the quality of data.
2) **Leaf Disease Classification using CNN** – Uses CNNs to classify leaves into separate categories of Healthy, Early Blight, Late Blight, Leafspot, Rust and Mildew.
3) **Fertilizer Recommendation using Random Forest** – Uses Random Forest to classify soils based upon NPK values and determine specific fertilizers to be used for various crops.
4) **Model Performance Evaluation and Analysis Module** – Outlines performance metrics and models based on their accuracy, precision, recall, F-measures, and Confusion Matrix.
5) **Treatment Suggestion and Recommendation Module** – Provides actionable and corrective treatment Suggestions along with the report on Disease Detection.
6) **Flask-based Interactive Web Application Module** – This is a user interface that allows the user to upload photos and input soil data to see Predictions and recommendations in Real-time.

V. Working

The system for detecting plant diseases and recommending fertilizers accepts two different kinds of inputs: photos of plant leaves to diagnose disease and numeric values of soil N, P, K content to determine related fertilizer. Once both datasets are collected, leaf images are preprocessed to remove corrupted or poor-quality samples. Each leaf image with standard dimension and normalized for ensure consistent pixel distributions. Data augmentation techniques such as rotating images, flipping, and brightness adjustment are implemented for generate additional difference in the training dataset. For the fertilizer module, numeric soil data is cleaned and scaled so that nitrogen, phosphorus, and potassium contribute equally to model training.

In the second phase, a convolutional neural network (CNN) is trained to classify plant leaves into disease categories: healthy, early blight, late blight, leaf spot, and rust. Simultaneously, a Random Forest model is trained on NPK values to recommend the appropriate fertilizer. Both models are evaluated using accuracy, precision, recall, F-measure and confusion matrix analysis.

The trained models are then integrated into an interactive web application built on Flask. Users can upload leaf images to receive disease identification, and enter soil nutrient values to receive fertilizer recommendations. The application processes each input in real-time and provides results along with appropriate treatment options through a user-friendly interface.

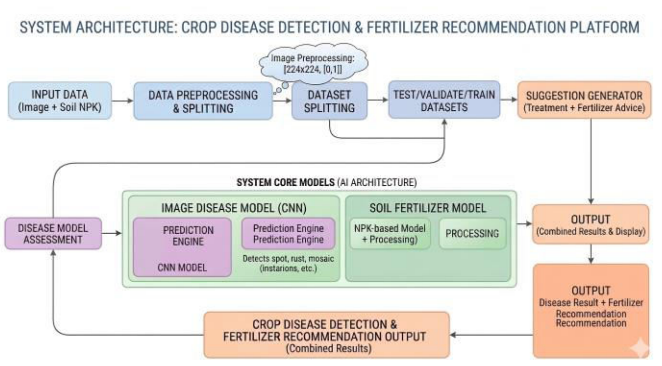

Fig. 1: System Architecture Diagram

A. System Architecture Diagram

The system architecture diagram captures the complete operational flow of the System. It illustrates how input data in the form of leaf images and soil NPK values passes through preprocessing and dataset splitting, then into the core AI model layer comprising the CNN disease model and the soil fertilizer model. The final layer delivers combined results including disease identification, fertilizer suggestions, and treatment recommendations to the user.

VI. Dataset Description

A. Image Dataset

This is an image dataset made up of leaf images from different crops such as potatoes and tomatoes. Each image is representative of the health condition of the specific plant, and the dataset is divided into 6 classes: Healthy, Early Blight, Late Blight, Leaf Spot, Rust, and Mildew. This dataset is for train and test the disease detection model.

B. Soil Nutrient Dataset

This dataset contains numerical attributes related to soil, specifically the amounts of NPK. These attributes serve as input for the fertilizer recommendation model, enabling analysis of soil nutrient content and generation of appropriate fertilizer recommendations for specific crops.

C. Crop Information Dataset

This dataset contains labels for different types of crops associated with their corresponding NPK values and diseased plant images. This information allows the system to generate fertilizer recommendations for specific crops and provide disease treatment suggestions.

D. Dataset Split

Image and soil nutrient data has been divided into a ratio of 70% to 30% for both the image data and the soil nutrient data, ensuring proper evaluation of both the CNN disease detection model and the Random Forest fertilizer recommendation model.

TABLE I: Dataset Statistics

Dataset Type	Samples	Classes
Crop Disease Dataset	2,500+ leaf images	6 disease categories

VII. Data Preprocessing

A. Cleaning of Image Data

- Deleting damaged or otherwise unusable leaf images
- Creating a consistent file naming convention based on disease classification

B. Image Processing

- All leaf images resized to 256 × 256 pixels
- Pixel values normalized to the 0 to 1 range

C. Data Augmentation

- Rotation, flipping, zooming, and brightness adjustment applied to increase dataset variety and improve CNN generalizability

D. Soil Data Preparation

- Missing or inconsistent NPK values handled and outlier records removed

E. Feature Engineering

- Nitrogen, Phosphorus, and Potassium values transformed using min-max normalization
- Crop type labels encoded to determine the proper fertilizer for each crop type

VIII. Machine Learning Models

A. Disease Classification Model

- Convolutional Neural Network (CNN)

B. Fertilizer Recommendation Model

- Random Forest

C. Baseline Models

- Decision Tree
- Logistic Regression

IX. Training Strategy

- Cross–Entropy Loss - CNN Disease Classification
- Split Data - Training and Testing, Validation
- Performance Metrics: Accuracy, Precision, Recall and F-measure
- Confusion Matrix Analysis of Both Disease Detection and Fertilizer Prediction Models

X. End-to-End Workflow Diagram

The workflow shows how the entire crop disease and fertilizer recommendation systems operate from start to finish. It describes how a user uploads an image of a leaf and enters the nutrient data into the system using a web interface. Pre-processing steps include resizing, normalizing and augmenting the leaf images for input into the CNN model. The Random Forest model requires scaling and cleansing to prepare for receiving its inputs (nutrient content). After each model generates its predictions, the two sets of outputs are combined and provided to the user as information on the leaf's potential disease, what treatment should be applied, and how much fertilizer needs to be applied based on soil content.

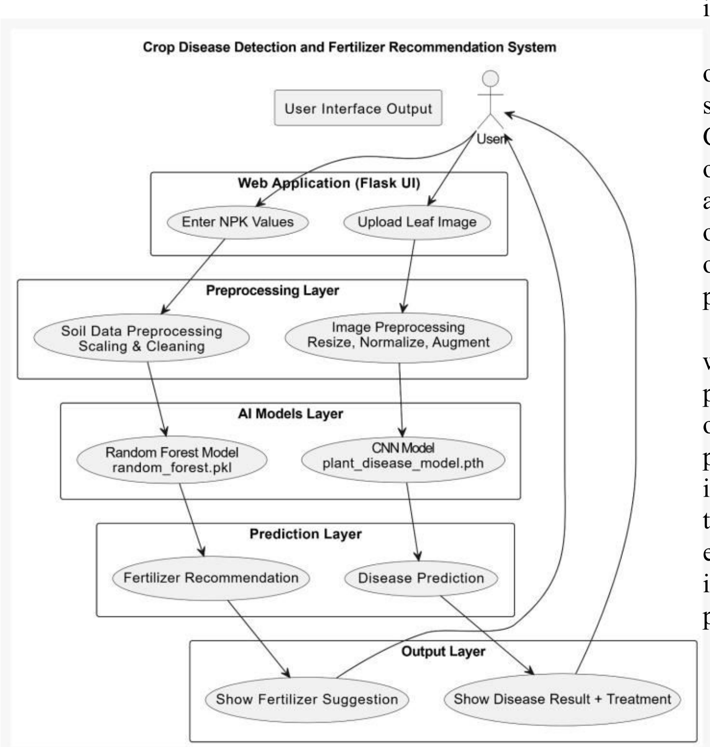

Fig. 2: End-to-End Workflow

This workflow diagram illustrates the overall approach used in the study from data collection through AI model predictions and ultimately presenting the combined model outputs back to the user via the interactive web-based application.

XI. Evaluation Metrics

The efficiency of system is analyzed using the following metrics:

- Accuracy
- Precision
- Recall
- F-measure
- Confusion Matrix

Accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

XII. Discussion

Experimental results show that the plant disease detection and fertilizer recommendation system is effective in making. agricultural decisions, using image-based deep learning and machine learning in processing soil nutrient data. Comparative results show that the CNN-based disease detection model accurately classifies leaf diseases, while the Random Forest algorithm reliably recommends fertilizers matched to soil nutrient requirements. Considering both crop images and soil conditions together produces more accurate and practical agricultural decision-making than considering either source of information alone.

Data Preparation and expansion were key for reliability of the disease detection system. standardization, adjusting size and expansion with rotation and flipping allowed the CNN to learn from a different set of disease examples by overfitting. Preprocessing the NPK values with normalization and appropriate encoding improved the stability and reliability of the fertilizer recommendation model. The combination of imagebased and numerical features enhanced the overall predictive. capability of integrated system.

The modular design of the application architecture, along with the implementation of the web interface utilizing Flask, produces a system capable of supporting a diverse range of applications quickly and easily, while providing a single point of entry that allows users to obtain disease diagnosis information, fertilizer application or treatment recommendations, and how to treat affected crops based on previous experience/knowledge, therefore providing accurate and timely information to farmers, agricultural majors, and extension personnel related to improving crop health and productivity.

XIII. Limitations

- Disease detection can be adversely affected by the accuracy. affected by the low-quality, or dark images in the input dataset.
- The fertilizer recommendation model is the one that is based on the sole recommendation alone NPK values, and therefore the other valuable soil properties like the pH and moisture content are left out, which also influence plant growth.
- There is no field testing that has been conducted on a large scale. of statistics or in various areas and types of crops which restricts capability of ascertaining generalizability of the system.

XIV. Conclusion and Future Work

This initiative has seen to it that artificial intelligence systems can be effectively used in agriculture for detection of Plant diseases as well are recommended to prescribe the right fertilizers. Developed system uses a CNN to categorise leaf diseases and Random Forest model to give recommendations on fertilizer. according to the NPK levels of soils. Observations based on experiment prove that the CNN precisely determines diseases and that the Random. Forest has been reliable in prescribing the fertilizers that can be used in the measured soil conditions. In this way, the worth of this strategy is shown hybridization of machine and computer vision. to help in making agricultural decisions using data.

The future study will increase the amount of images used in the study. take field pictures of different weather and light to improve model generalizability. Additional soil characteristics including pH, moisture, organic carbon and micronutrients will be added to the model of fertilizer recommendations to give more detailed recommendations. Opportunities also exist to create the estimation of disease severity with the help of computer vision. and to incorporate weather information to promptly notify about disease outbreaks. Possible improvements are the mobile implementation, the multilingual nature and the comprehensive in-field experimental tests to. enhance usability and real-world adoption by farmers.

Appendix

Algorithmic Overview of the Proposed System

An algorithmic perspective of the process taken by the Crop Disease Detection And Fertilizer Recommendation System will be discussed in this appendix. Validation, cleaning and standardising of both leaf image and soil NPK values will occur as an initial step in acquiring an input data set. The images of the leaves will be resised and normalised, followed by the classification of the images using convolutional neural networks (CNN) to identify whether they are diseased, and the numerical values of NPK nutrients will require scaling and be ready to be used to recommend a fertiliser.

The processed Input is inputted into a trained model with CNN (name = plant_disease_model.pth) utilized in predicting the disease, and the Random Forest (name = random_forest.pkl) utilized in predicting fertilizer recommendations. The model(s) output does not only consist of the predicted characteristics of the plant disease but also the recommended fertilizer considering the N, P and K (NPK).

Predicted output will be validated before coming up with any verified performance metrics and delivered by way of a simple system interface. Web-Based Solution will enable farmers to detect the diseases of their crops in real-time, as well as offering them fertilizer advice and treatment recommendations to help them enhance agricultural decision making.

References

[1] S. Pawar and R. Patil, (2019) Automated Detection of Plant Diseases using CNN for Indian Agriculture will be published in IJIRCCE.

[2] P. Kumar and A. Ghosh, (2020) Smartphone-Based Real-Time Detection of Plant Diseases, Technical Report, Indian Institute of Technology Kharagpur.

[3] Tata Institute of Fundamental Research, (2021) Machine Learning in Crop Disease Prediction, Technical Seminar Series, Mumbai.

[4] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, (2016) Deep Neural Networks-Based Recognition of Plant Diseases by Leaf Image Classification was published in Computational Intelligence and Neuroscience.

[5] K. P. Ferentinos, (2018) Deep Learning Models for Plant Disease Detection and Diagnosis are published in Computers and Electronics in Agriculture.

[6] R. Singh, V. Sharma, and M. Pandey, (2020) Detection of Tomato Leaf Diseases Using Deep Convolutional Neural Networks will be included in the proceedings of the International Conference on Computing, Communication and Automation.

[7] A. Mohanty, D. Hughes, and M. Salathe´, (2016) Using Deep Learning for Image Based Plant Disease Detection were published in Frontiers in Plant Science.

[8] S. Brahimi, K. Boukhalfa, and A. Moussaoui,(2017) Deep Learning for Tomato Diseases: Classification and Symptoms Visualization was published in Applied Artificial Intelligence.

[9] The Food and Agriculture Organization's "Early disease detection in crops using artificial intelligence," was published as an FAO Agricultural Research Report in Rome by FAO in 2022.

[10] Kamilaris and Prenafeta-Boldu published their comparative survey of "Deep learning in agriculture," in the journal Computers and Electronics in Agriculture, volume 147 (pp. 70-90, 2018).

[11] Barbedo's work entitled "Impact of the dataset size and dataset variety on the effectiveness of deep learning and transfer learning for plant disease classification," was published in Computers and Electronics in Agriculture, volume 153 (pp. 46-53, 2018).

[12] Too, Yujian, Njuki, and Yingchun worked collaboratively to provide a comparison of "Fine-tuning deep learning models for plant disease identification," in Computers and Electronics in Agriculture, volume 161 (pp. 272-279, 2019).

[13] Cruz, Ampatzidis, Pierro, and Materazzi investigated the application of "Deep learning-based crop disease classification using field images," in the journal Biosystems Engineering, volume 192 (pp. 80-94, 2020).

[14] Zhang, Wu, and Chen applied "Real-time crop disease recognition using convolutional neural networks on mobile devices," in IEEE Access, volume 8 (pp. 16166-16175, 2020).

[15] R. P. Singh and S. P. Singh, produced "Machine learning approach for soil nutrient analysis and fertilizer recommendation," featured in International Journal of Agricultural Science and Research, volume ten, number three (pp. 45-52, 2020).

[16] S. K. Behera, R. K. Jha, and P. Singh, presented "Soil nutrient based fertilizer recommendation – random forest algorithm," to Journal of Agricultural Informatics, volume twelve, number two (pp. 12-21, 2021).

[17] The 2020 publication of the Indian Council of Agricultural Research titled "Soil Health Management and Fertilizer Recommendations".

[18] The 2021 governmental publication of the Ministry of Agriculture and Farmers Welfare titled "Guidelines and Recommendations of the Soil Health Card Scheme".

[19] FAO, "Digital agriculture and artificial intelligence for crop health monitoring," FAO Technical Report, 2021.

[20] The 2006 Springer publication by C.M. Bishop titled "Pattern Recognition and Machine Learning".

[21] The 2009 Springer publication by T. Hastie, R. Tibshirani and J. Friedman titled "The Elements of Statistical Learning".

[22] The 2013 Springer publication by G. James, D. Witten, T. Hastie and R. Tibshirani titled "An Introduction to Statistical Learning".

[23] The 2019 O'Reilly Media publication by A. Geron titled "Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow".

[24] The 2012 MIT Press publication by K.P. Murphy titled "Machine Learning: A Probabilistic Perspective".