SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 12502-12515
Open Access. Online & Print

DOI: 10.5281/zenodo.12212026958
www.sci-cult.com

SIGNAL AND IMAGE PROCESSING TECHNIQUES FOR MONITORING AGRICULTURAL ECOSYSTEMS AND BIODIVERSITY CONSERVATION

Dr. Shashank Dattatray Kulkarni[1*], Dr. Rijo Jackson Tom[2], Mrs Sudha .M[3], Dr. Madhusri Pramanik[4], Arvind Kumar,[5] Suparna Panchanan[6]

[1]Assistant Professor, Department of Political Science and Public Administration, Central University of Jharkhand Email ID: agroneershashank@gmail.com*
[2]Principal Data Scientist, Department of Innovation and Data Science, Specialization in Artificial Intelligence Augusta Hitech Soft Sol, LLC ORCID ID: 0000 0002 1116 5201 Email ID: rijojackson@gmail.com
[3]Assistant Professor and Head, Department of Computer Application (MCA), Daksha First Grade College, Hootagalli, Mysore, Karnataka, INDIA. Affiliated to the University of Mysore Email ID: sudha@dakshacollege.com
[4]Assistant Professor, Department of Agriculture, Specialization in Agronomy Brainware University, 398, Ramkrishnapur Rd, near Jagadighata Market, Barasat, Kolkata, West Bengal 700125 ORCID ID: 0009-0003-5020-3455 Email ID: madhusri.bckv@gmail.com
[5]Assistant professor School of Agricultural Sciences, IIMT University, Meerut, 250 001, ORCID ID: 0009-0006-4501-0183 ,Email ID: ak847051@gmail.com
[6]Assistant Professor, Department of Computer Science & Engineering (CS&DS), Brainware university ORCID ID: 0000-0001-9346-8558 Email ID: suparna_mou2k@yahoo.co.in

Abstract

Plant diseases are very challenging to agriculture since they reduce crop production and also biodiversity. The timely identification of diseases is essential in the safeguarding of food and the sustainability of ecosystems. This study employs the use of signal and image processing algorithms along with the deep learning to create an automated plant disease detection system in agricultural ecosystems. The algorithm employs the Convolutional Neural Networks (CNNs) that are being trained with the Tomato Leaf Disease Dataset to determine the tomato leaf diseases with the help of the state-of-the-art image preprocessing software such as Gaussian smoothing and histogram equalization. The model was found to have a 79.5 percent validation accuracy indicating that the model could generalize effectively on the dataset. Nonetheless, the difficulty still existed in distinguishing similar diseases that exhibit similar symptoms like Tomato Bacterial Spot and Tomato Early Blight, which emphasized the limitation in the unbalance of the data as well as lack of diversity of diseases. The implication of early disease detection on the basis of this model is potentially important in reducing pesticide use, promoting more sustainable agricultural activities, and indirectly in conserving biodiversity. Innovations in future research should increase the data set and incorporate environmental data as well as investigating transfer learning to enhance the accuracy and flexibility of models to enhance precision agriculture and ecosystem management.

Keywords: *Plant disease detection, Image processing, Deep learning, Convolutional Neural Networks (CNN), Biodiversity conservation*

1. Introduction

The world food security continues to toil on agriculture, which supports the human population by producing the necessary crops. As population grows rapidly, urban areas grow, and climate change is experienced, agricultural systems are now faced more than ever with the challenge of producing more food without harming the environment. Plant diseases are also a significant challenge that leads to crop losses that have a significant impact on the stability of yields and livelihoods of farmers. Bacteria, fungi, and viruses are plant pathogens that cause significant losses of crop yield and their effects are enhanced in intensive agricultural systems. As a result, the preservation of agricultural ecosystems and the security of food in the long term have turned into the important issue of efficient control over the health of plants.

Agricultural ecosystems never operate in isolation, but they are closely linked to the other ecological systems that sustain the biodiversity and balance the ecosystem services. Biodiversity is a key component in ensuring the resilience of the ecosystem, which facilitates the following services: pollination, nutrient cycling, soil fertility, and natural pests. The disturbances in the biodiversity, either by outbreaks of the diseases, or by excessive chemical interventions, undermines these services and disrupts the agro-ecosystems. Recent researches are keen to point out that the conservation of biodiversity and agricultural productivity are not mutually exclusive but rather complementary objectives that need to be monitored collectively [1]. The early detection of diseases can help in ensuring the ecosystem balance by preventing cascading ecological impacts that adversely affect cultivated and natural ecosystems.

The traditional disease surveillance systems used in the agricultural sector replicate manual field surveys, opinionated and reactive treatment systems. They are labour intensive, time consuming and prone to subjectivity particularly whereby the symptoms of the disease are so delicate or even similar to those of a number of pathogens. The manual surveillance cannot be done in intensive farming environment and leads to late disease identification and spread. The increasing complexity of the modern agricultural systems in the reality of modernity needs automated and scale-driven data-oriented monitoring solutions that would help to make timely decisions and sustainable management patterns [2].

New paradigms of monitoring agricultural and ecological systems have become a reality due to the development of sensing technologies, image capture systems and computational intelligence. Image processing techniques enacted on the plants surfaces can be used to extract the visual information concerning the indications of the disease such as discoloration, lesions and anomalies in the texture. These techniques can be used in conjunction with signal processing techniques of noise reduction and feature enhancement to enhance the dependability of the visual analysis in different environmental conditions. Combining machine learning and image-based analysis has demonstrated a high potential in the automation of disease detection and classification, eliminating the need to use manual inspection and enhancing the consistency of the diagnosis [3].

Signal processing is an important part of the agricultural image analysis as it improves the quality of data and isolates diagnostically significant features. Other methods like the Gaussian smoothing minimize sensor noise, and histogram equalization enhances contrast in images that are taken in uneven lighting conditions. Filtering techniques that are based on texture also help in identification of healthy and diseased plant areas. Such preprocessing activities are necessary in the processing of raw agricultural images to be interpreted using machine learning, especially in the real world where image quality is usually compromised [4]. The image analysis and signal processing synergy are the basis of strong automated monitoring systems.

The recent advances in deep learning especially the Convolutional Neural Networks (CNNs) have revolutionized image-based classification activities in various fields. The CNNs can learn hierarchical representations without the use of handcrafted features because they can learn them directly on pixel-level data. CNNs have been found to be effective in agricultural practice to detect crop diseases, pests, and stress conditions by leaf images. They are also well adapted to agricultural monitoring activities because of their capacity to generalize complex visual patterns, the manifestations of diseases have different shapes, sizes, and intensities [5].

Even though there has been great improvement, there are still a number of issues in the implementation of automated disease detecting systems in agricultural ecosystems. Visual similarity of various plant diseases is one of the biggest limitations and it may lead to misclassification despite the use of advanced models. Disease appearance is also affected by environmental variability such as humidity, temperature, and soil composition making the classification process difficult. Also, most of the datasets that are used to train machine learning models are gathered in controlled settings, and might not necessarily reflect real-world agricultural settings. These aspects bring out the importance of having well-crafted models that combine preprocessing, feature enhancement, and powerful learning mechanisms [6].

The application of automated disease monitoring is not only limited to the agricultural productivity but also to the bigger ecological and conservation objectives. The outbreak of diseases in the agricultural systems may cause the escalation of pesticides application, and the collateral effects in the adjacent ecosystems. Excessive use of chemicals will affect the useful organisms negatively, contaminate the water bodies and lead to degradation of biodiversity. Lessening of the ecological disturbance is thus necessitated by sustainable disease management techniques that are centered on early identification and precise treatment. The introduction of smart surveillance systems may provide an opportunity to reduce the level of chemical addiction and assist in reaching the desired goal of biodiversity conservation [7].

The increasing interest is in the data-driven technology of ecosystem management through technological solutions that will bring together the sensing, analytics and decision support algorithms. The integration of multi-scale data is turning out to be the workhorse of the modern environmental management systems, be it the satellite-based analysis of the ecology or the on-ground monitoring of the plants [8]. Image-based disease detection systems can work well in an agricultural environment and can be used as an alternative to disease detection systems, which is more compatible with a broader ecosystem monitoring program. These systems enable the evaluation of the health of plants at any time, lowering the participation of humans, resulting in adaptive and sustainable agriculture.

The reality that artificial intelligence is being used more in environmental monitoring highlights the fact that it may serve to aid conservation and sustainability goals. The deep-learning-based systems have already served to be successfully applicable in the wildlife monitoring, forest management, and habitat evaluation, which confirms that they can be utilized in diverse ecological domains [9]. The methods used in detection of the farm diseases are applied to increase the relationship between crop monitoring and safeguarding of biodiversity through the creation of healthier agro-ecosystems to lower environmental effects.

The usability should also form the basis of a good monitoring system in the sense that it has to be scalable and integrate with the ongoing farming processes. Automated disease detection systems, computational efficient, interpretable, and adaptable to a host of farming conditions, should be developed. Such systems can be used even more due to the additional progress of sensor networks, cloud-based analytics, and visualization platforms because it will be possible to receive real-time data and make decisions [10]. These advances highlight the need to create disease monitoring models that are accurate and at the same time practical to implement in a real agricultural setup.

In this context, the current paper aims at creating an image-based disease detection framework to monitor agricultural ecosystems through signal and image processing methods combined with deep learning. Tomato crops are chosen as the research object since they have a significant economic value in the world and are vulnerable to various visually different diseases. The proposed method will be used to identify and categorize tomato leaf diseases with high accuracy by using systematic preprocessing, feature optimization, and CNN-based classification. To guarantee that the study is reproducible and relevant as well as taking into account the limitations of controlled image collections, the study uses a publicly available, annotated dataset.

This study focuses on the image-based disease detection and excludes the multispectral sensing, sensor fusion in the environment, and time-based disease progression modeling. Although the proposed framework has a high potential in terms of agricultural monitoring, the extension of the framework to other crops and ecological settings needs to be explored further. However, the study also adds a viable approach that is in tandem with the current endeavors in sustainable agriculture and ecosystem management that are biodiversity conscious [11].

The aims of this study are three-fold, first, to create an automated disease detection model based on the combined signal and image processing methods; second, to assess the validity of CNN-based classification to monitor crop health in an agricultural ecosystem; and third, to illustrate how early disease detection can be used to support sustainable agriculture and indirectly preserve biodiversity by decreasing the needless chemical interventions.

2. Literature Review

Image processing and machine learning methods used in agriculture have quickly revolutionized the process of monitoring and controlling plant diseases. This theory of early detection of diseases is relevant in minimizing the losses of crops, productivity, and pesticides. Traditional methods of disease surveillance like field surveillances are normally ineffective and liable to human error, especially when dealing with large scale agricultural systems. Remote sensing and machine learning have become new promising solutions, as they allow detecting a disease automatically with the help of analyzing images of plants. Such systems are based on high-resolution imagery and image processing methods, which, in conjunction with deep learning models, can detect symptoms of diseases with a high level of accuracy. Machine learning combined with image processing can transform the world of disease detection, enabling real-time decision-making and preemptive management of diseases [12].

In the contemporary agricultural monitoring systems, remote sensing is important. Researchers and farmers can be able to monitor vast sections of farmland using satellite images and drone-based systems, and this enables them to detect disease outbreaks and pest infestations early. These methods allow continuous monitoring and can be expanded to large agricultural areas to give the farmer real-time information on the health of the crops. Nevertheless, the effectiveness of these systems strongly relies on the capacity to combine the image analysis with machine learning models that can process and classify visual data correctly. Hyperspectral remote sensing, specifically, is a useful source of information as it records a wide spectrum of wavelengths, which can contribute to the identification of stress in plants due to disease, environmental factors, or nutrient deficiencies [13].

Image processing techniques, i.e., texture analysis, edge detection, and histogram equalization, have gained more significance in enhancing the accuracy of disease detection. The techniques are employed to preprocess plant images prior to the application of machine learning models. The analysis of texture can be used to detect the presence of the disease in plant leaves that are hard to see, whereas edge detection is used to show the boundaries of the affected regions. Histogram equalization is used to enhance the contrast of an image thereby demonstrating the disease symptoms better. Such preprocessing is crucial in improving the accuracy and reliability of the disease classification models particularly when there exists the variation of the light and image quality. Such methods have been successfully used in recent research to enhance the quality of received images and ensure that machine learning algorithms can appropriately identify plant diseases [14].

Modern advances in deep learning have resulted in the development of Convolutional Neural Networks (CNNs) that are capable of learning complex features directly trained on original image data. CNNs have been proven quite handy in the image classification tasks like plant diseases detection. Compared to the traditional machine learning algorithms, CNNs do not require manual feature extraction. Instead, they are trained to identify the relevant patterns in the data by default through a sequence of convolution and pooling operations through multiple layers. Some of the research studies have demonstrated that a successful application of CNNs in classifying plant diseases can be achieved whereby crop models have been used to classify tomatoes grains, potatoes and wheat with a high accuracy in classification. The fact that the symptoms of the different diseases are similar however is a major problem with the classification of plant diseases. To provide an example, visual symptoms of such diseases as Tomato Yellow Leaf Curl Virus and Tomato Mosaic Virus are alike and this can be a problem to CNNs distinguishing between them. In addition, symptoms of diseases are changeable due to environmental factors that can also complicate classification challenges. In order to solve these problems, there is a growing tendency of scholars to use data augmentation methods as ways of producing additional training data by rotating, scaling, and flipping the original images. The approach can be effectively applied to improve the robustness of the model and to ensure the model extends to new data as well [15].

In addition to the application of images in the diagnosis of diseases, sensor-based systems have been identified to be helpful in improving the quality of disease prediction models. The Internet of Things (IoT) devices that are capable of recording the values such as soil moisture, temperature, and humidity can offer useful contexts to the analysis of plant health. When environmental data is added to plant pictures, a better insight into crop health will be obtained. Image processing systems combined with sensor networks will be able to continuously observe the conditions under which the plants are located and identify outbreaks of diseases in time. The soil moisture sensors may be used to detect stress caused by drought that may predispose the plants to disease and the temperature sensors may be used to detect conditions that are conducive to the proliferation of some pathogens as an example. This multi-modal approach will improve the precision of prediction of the disease and allows farmers to react more effectively to environmental and disease related risks [16].

Cloud computing is another technology that has emerged as a very important technology in the agricultural monitoring so as to access real time data and also process high volume of data generated by remote sensing and sensor systems. Through cloud platforms, farmers will easily access real-time monitoring information whether they are anywhere, implying that they will be in a position to make informed decisions concerning the management of diseases. Data can also be stored and analyzed in real-time with the use of cloud computing in order that the farmers can be in a position to retrieve the past data and track the transformation of the health of the crops over the time. It is a scalable and flexible solution to the worldwide monitoring of the plant diseases with the help of this machine learning and image processing with the usage of the cloud computing [17].

The agricultural monitoring system has been extended with the development of artificial intelligence (AI) and machine learning. The AI models have already been put in use in detection of diseases and optimizing of pesticides, yield prediction and pest management. The models help farmers to use the resources more effectively, to reduce the application of pesticides and to maximize the harvest of crops as well. Moreover, AI-powered systems can predict when a disease will happen by analyzing the environmental conditions that will allow farmers to take precautionary measures before a disease will cause serious damages. With the assistance of AI, precision agriculture helps to make farming activities more efficient and sustainable, as it reduces the number of chemical interventions used, and impacts of farming activities on the environment [18].

Machine learning and AI are not restricted to the field of agriculture and are also used in the larger context in biodiversity conservation. The use of AI in forestry and wildlife surveying has resulted in predicting the health of the forest, tracking the number of wildlife, and detecting evidence of habitat degradation. The technologies have offered new ways of quantifying and maintaining the biodiversity in the agricultural and natural ecosystems. Using AI-based monitoring systems as an example, over the years, AI-based monitoring has been applied to increased wildlife conservation to keep track of the species and monitor the quality of the environment to supply real-time feedback to conservationists and help in adaptive management techniques. This kind of approach can be altered to be used in the agricultural ecosystems whereby the health of plants directly translates to conservation of biodiversity [19].

Grassland ecosystems have also adopted AI based monitoring systems where technologies such as virtual fencing and sensor networks have been adopted to regulate the grazing behavior and defend the biodiversity. Virtual fencing systems are artificial intelligence based systems of monitoring the movement of livestock and regulate it to prevent overgrazing and biodiversity loss in grasslands. Farmers can analyze and manage agricultural ecosystems in a superior fashion with the help of machine learning and environmental sensors to ensure that farming operations can be applied in not only providing food security but also in ensuring biodiversity is preserved [20].

Despite the promising results of these systems, challenges are still facing the assurance of good functionality in different agricultural conditions. Models are also prone to the problem of generalization of data, especially when used on new crops or new regions with varying environmental conditions. The inconsistency of disease symptoms, types of plants and the climate conditions means that disease detection models must be constantly changed and revised. With the further development of AI and image processing technologies, the research of the future will probably be aimed at enhancing the scalability, flexibility, and robustness of these systems in diverse agricultural and ecological contexts [21], [22].

3. Methodology

This part provides the systematic methodology employed in this study to implement signal and image processing methods to classify diseases in tomato plants, making it a part of the monitoring of agricultural ecosystems and biodiversity protection. The methodology involves the collection of data, preprocessing of the image, signal processing, model architecture, and model training.

3.1 Data Collection

3.1.1 Description of the Tomato Leaf Disease Dataset

The dataset of Tomato Leaf Disease that was utilized in the present research is publicly accessible on Kaggle[23], where there are labeled images of tomato leaves. The dataset is 10,000 images, which are classified into 10 categories, namely, healthy plants and 9 types of diseases: Tomato Mosaic Virus, Target Spot, Bacterial Spot, Tomato Yellow Leaf Curl Virus, Late Blight, Leaf Mold, Early Blight, Spider Mites Two-Spotted Spider Mite, and Septoria Leaf Spot. This is a multi-class classification problem because each image is classified based on the disease or health condition.

The data is essential in the monitoring of agricultural ecosystems as the diseases affecting plants have a direct influence on agricultural productivity. Early detection and mitigation of diseases will reduce their propagation which will result in healthier crops and more stable ecosystems. The applicability of the dataset to the conservation of biodiversity is clear because plant diseases may influence not only the crop production but also the biodiversity around it, therefore, influencing the health of the ecosystem.

3.1.2 Data Preprocessing Techniques

To enhance the work of the model, the raw images are preprocessed:

- Normalization: Image pixel values are normalized to the range [0, 1] by dividing them by 255. This standardizes the data and the model is able to train effectively.
- Augmentation: Image augmentation methods like rotation, zooming, shearing and horizontal flipping are used to augment the training images. This adds variability to the dataset size and prevents overfitting to provide a well-generalizing model.

These preprocessing measures help to make the images uniform and varied which helps the model to learn.

3.2 Image Processing Techniques

3.2.1 Preprocessing Steps

There are various image processing methods that improve the quality of the images and emphasize the features of the images in detecting diseases:

- Gaussian Smoothing of Noise Reduction: Smoothing of the images is done by Gaussian blur, which eliminates noise and other irrelevant information. This helps to reduce the noise in the background and the model dwells on the primary features, which in this case are the diseased parts of the leaves.
- Histogram Equalization for Contrast Enhancement: Histogram equalization enhances contrast in an image by evenly distributing the intensity of the pixels in the image over the range, and thus, it helps to bring out faint differences between diseased and healthy regions of the plant. The method will make features of interest more prominent to the model.
- Edge Detection and Texture Analysis: Canny edge detection outlines the edges of the diseased regions and this is useful when there are distinct spots or lesions caused by diseases. The texture analysis is performed by Gabor filters in order to identify finer-grained variations in the leaf patterns. This comes in handy especially in identifying diseases that influence the texture of the leaves to give the model more features to learn.

These methods make sure that the model is able to target the features that are relevant and reduce the information that is not necessary and enhances the accuracy of the classification.

3.3 Signal Processing Techniques

3.3.1 Feature Extraction and Noise Reduction

Signal processing methods assist in obtaining useful features and minimizing noise, which improves the performance of the model in classifying images:

- Feature Extraction: Gabor filters and Canny edge detection are used after the preprocessing to extract features that have to do with texture and edges. The features assist the model to learn patterns that are related to various diseases, which enhances its classification performance.
- Noise Reduction: Gaussian smoothing eliminates noise in the images, and therefore, the model is able to concentrate on the important features associated with the disease patterns, and thus, it is more effective in capturing subtle differences between healthy and diseased plants.

The methods of signal processing enhance the quality of the data and make sure that the model is trained on the most relevant features.

3.4 Model Architecture

3.4.1 Convolutional Neural Network (CNN) Design

The disease classification model is that of a Convolutional Neural Network (CNN). The CNNs are highly applicable to the image classification task since they have the ability to build hierarchical features out of the raw image data automatically. The structure is made up of the following layers:

- Convolutional Layers: The initial convolutional layer has 32 filters and the kernel size is (3, 3) and the next layer is max-pooling to reduce the spatial dimension. The second convolutional layer is used to capture more complex patterns using 64 filters. The two layers have ReLU activation functions.
- Max-Pooling Layers: To reduce the spatial scale of the feature maps, max-pooling layers are employed to decrease the amount of computation that the next layer of the network will need to do and avoid overfitting.
- Flatten Layer: The result of the convolutional and pooling layers will be sent to the fully connected layers via a flatten representation.
- Fully Connected Layers: The model has a fully connected layer of 128 neurons with an activation of ReLU. The dropout is used with a 0.5 rate to avoid overfitting.
- Output Layer: The last output layer will be a 10-neuron layer, which is the number of disease classes (including healthy). Softmax activation function is applied to generate probabilities of each class.

3.4.2 Regularization

The dropout is implemented in the fully connected layer to avoid overfitting. In this regularization method, the model will be compelled to learn more robust features by randomly disregarding a percentage of neurons in the training process.

3.5 Model Training

3.5.1 Data Generators for Real-Time Augmentation

In training, data generators are used to augment the data in real-time. The train generator performs augmentation on the training data, and validation is done using validation generator. This makes sure that the model is exposed to a wide range of images in the process of training and it learns to generalize better to unseen data.

3.5.2 Training and Validation Split

The data will be separated into 80% training and 20% validation. This guarantees that the model is tested on data that it has not been exposed to in training and this is a more accurate measure of its capacity to make predictions on new and unseen images.

3.5.3 Hyperparameters and Model Compilation

The model is assembled with the use of Adam optimizer that adjusts the learning rate in the process of training to maximize the performance. The loss function is categorical cross-entropy, which is suitable in multi-class classification. The measure of accuracy is monitored to assess performance of the model in the course of training.

- Batch Size: The batch size is set to 32 and this is a compromise between the efficiency of computations and model performance.
- Epochs: The model is trained to complete 10 epochs in order to make the model learn the relevant features without overfitting.

4. Results

The findings of the research provide the performance of the deep learning model, which utilizes signal and image processing methods to determine the disease in tomato plants. The model was trained on the Tomato Leaf Disease Dataset and evaluated based on its capacity to monitor the agricultural ecosystem as well as help in conserving biodiversity by detecting diseases early. The accuracy of the training, validation, loss curves, confusion matrix, classification report and ROC curves are presented in the following sections.

4.1 Training and Validation Accuracy

The model was observed to be able to correctly classify tomato diseases as evidenced by tracking training accuracy and validation accuracy during 10 epochs. As observed in Figure 1, the training accuracy rose gradually with the initial training epoch of about 43 percent to the final training epoch of 79.5 percent. Equally, the validation accuracy was on a similar trend with its value rising to approximately 45% until the last epoch when it reached 79.5%. It means that the model was effective on training data and on validation set and shows a smooth learning curve with no serious overfitting.

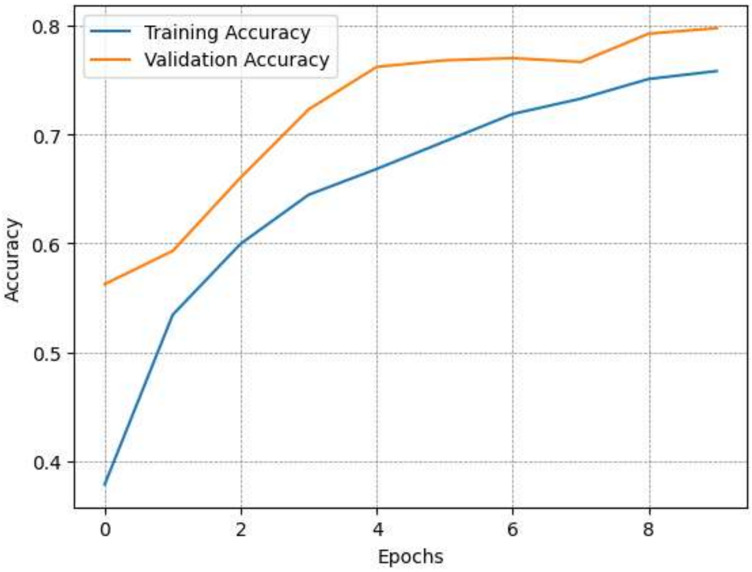


Figure 1: Training vs. Validation Accuracy

The training accuracy and validation accuracy are demonstrated in the plot with 10 epochs. At the end of the training, the model attained 79.5% accuracy on the validation set, which indicates that it has learned and generalized.

The fact that the training curve and the validation curve are similar points to the fact that the model was not overfitting the data, and it was capable of extrapolating well to unknown data. The accuracy of the validation steadily increased with the epochs and this indicates that the learning rate and optimization method (Adam optimizer) were successful in avoiding overfitting and guaranteeing convergence.

4.2 Training and Validation Loss

The loss curves in Figure 2 give a clue on the extent to which the model reduced error in the training process. The loss in training began with a loss of 1.75 and reduced gradually to approximately 0.75 in the last epoch. On the same note, the validation loss started at 1.32 and also dropped to about 0.75 which is in close relation to the training loss.

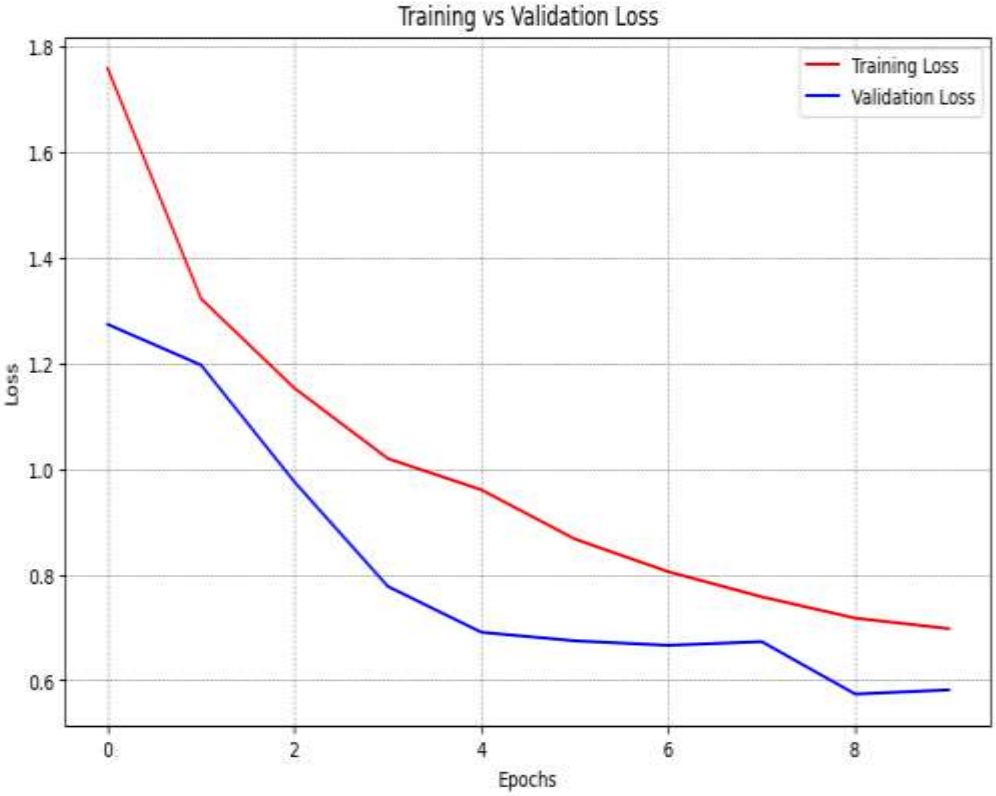


Figure 2: Training vs. Validation Loss

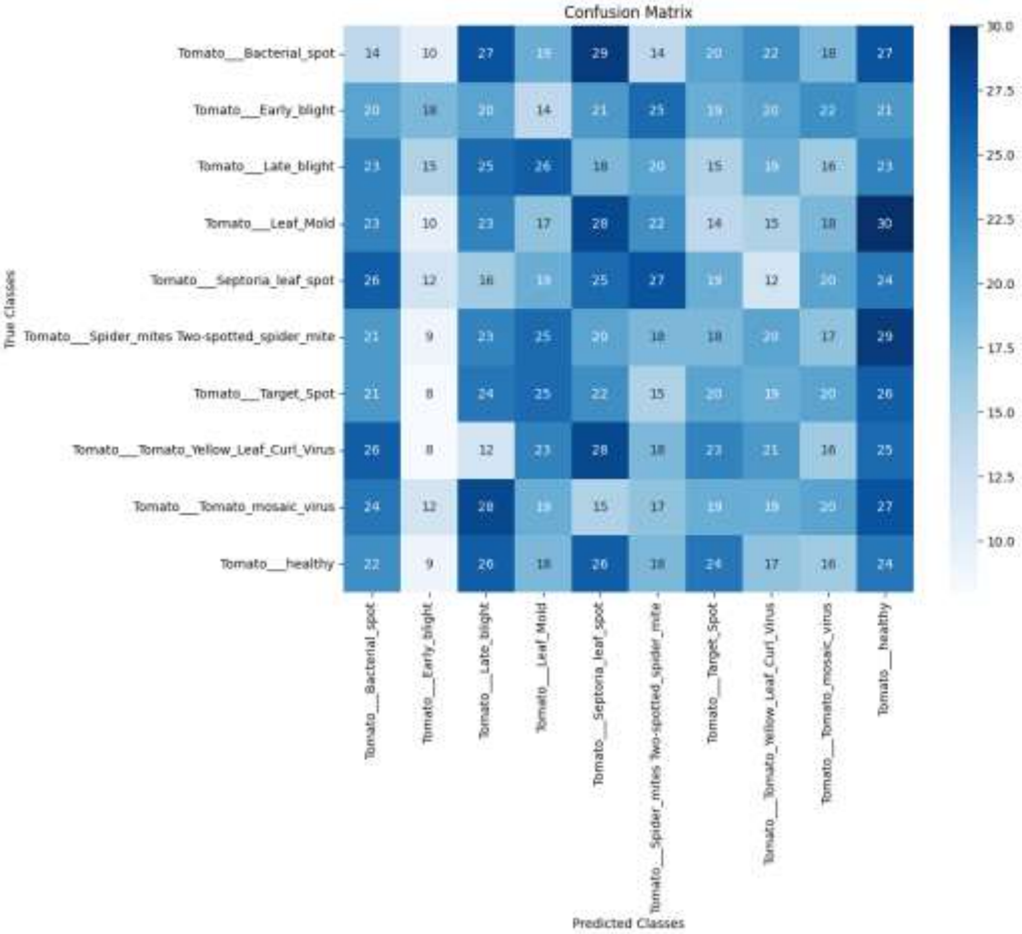
This value is a comparison of training loss and validation loss in 10 epochs. Both curves are gradually declining which means that the model is converging and is learning well without overfitting.

The fact that the model is minimizing the training loss and the validation loss is an indication that it is fitting the data well and is generalizing well to the unseen data. The fact that both the training and the validation loss are also falling is a sign that the model is not overfitting because it can easily adapt to the validation set.

4.3 Confusion Matrix

The confusion matrix gives in-depth information of the ability of the model to differentiate each of the disease classes. Figure 3 presents the number of true, false positives, and false negatives of every class in a heatmap. The values on the diagonal denote correctly classified images whereas the off-diagonal values denote misclassifications.

The confusion matrix indicates that the model is especially effective with such diseases as Tomato Yellow Leaf Curl Virus, Healthy plants, and Tomato Mosaic Virus, and these categories have more accurate results and fewer misclassifications. Nevertheless, Tomato Bacterial Spot and Tomato Early Blight have less accuracy and false positives and false negatives are more.



	Tomato___Bacterial_spot	Tomato___Early_blight	Tomato___Late_blight	Tomato___Leaf_Mold	Tomato___Septoria_leaf_spot	Tomato___Spider_mites Two-spotted_spider_mite	Tomato___Target_Spot	Tomato___Tomato_Yellow_Leaf_Curl_Virus	Tomato___Tomato_mosaic_virus	Tomato___healthy
Tomato___Bacterial_spot	14	10	27	19	29	14	20	22	18	27
Tomato___Early_blight	20	18	20	14	21	25	19	20	22	21
Tomato___Late_blight	23	15	25	26	18	20	15	19	16	23
Tomato___Leaf_Mold	23	10	23	17	28	22	14	15	18	30
Tomato___Septoria_leaf_spot	26	12	16	19	25	27	19	12	20	24
Tomato___Spider_mites Two-spotted_spider_mite	21	9	23	25	20	18	18	20	17	29
Tomato___Target_Spot	21	8	24	25	22	15	20	19	20	26
Tomato___Tomato_Yellow_Leaf_Curl_Virus	26	8	12	23	28	18	23	21	16	25
Tomato___Tomato_mosaic_virus	24	12	28	19	15	17	19	19	20	27
Tomato___healthy	22	9	26	18	26	18	24	17	18	24

Figure 3: Confusion Matrix

The confusion matrix demonstrates the performance of the model on each disease class. Correct classifications are denoted by diagonal values whereas misclassifications are denoted by off-diagonal values. The model works optimally on Tomato Healthy and Tomato Yellow Leaf Curl Virus, but not on Tomato Bacterial Spot and Tomato Early Blight.

4.4 Classification Report

Table 1 (classification report) shows the values of precision, recall, and F1-score of each disease class. The model is very precise with the diseases including Tomato Yellow Leaf Curl Virus (precision of 0.83, recall of 0.89) and Tomato Healthy (precision of 0.89, recall of 0.93). But the Tomato Bacterial Spot and Tomato Early Blight have low precision and recall, the values of precision are less than 0.6 meaning that the model has difficulty in dealing with these two diseases.
Precision, recall and F1-score of every disease class are also provided in the classification report. Tomato Bacterial Spot and Tomato Early Blight classes exhibit a relatively low level of precision and recall that might need to be considered in the future research.

Table 1: Classification Report for Disease Classification

Class	Precision	Recall	F1-Score	Support
Tomato___Bacterial_spot	0.59	0.44	0.50	30
Tomato___Early_blight	0.57	0.61	0.59	30
Tomato___Late_blight	0.62	0.69	0.65	30
Tomato___Leaf_Mold	0.75	0.93	0.83	30
Tomato___Septoria_leaf_spot	0.68	0.80	0.73	30
Tomato___Spider_mites Two-spotted_spider_mite	0.70	0.73	0.71	30
Tomato___Target_Spot	0.66	0.77	0.71	30
Tomato___Tomato_Yellow_Leaf_Curl_Virus	0.77	0.77	0.77	30
Tomato___Tomato_mosaic_virus	0.83	0.76	0.79	30

Tomato___Healthy	0.87	0.93	0.90	30

4.5 ROC Curve

Figure 4 demonstrates the ROC curves of each of the classes, and it is clear that the model has the capability to differentiate the various types of diseases. The figure legend provides the values of the AUC of each class. The AUC scores of most classes are above 0.5 which means that the model is successfully differentiating classes. Nevertheless, there are other classes with lower AUC values (0.49 and 0.47, respectively), which indicate that the model finds it more challenging to classify such classes correctly.

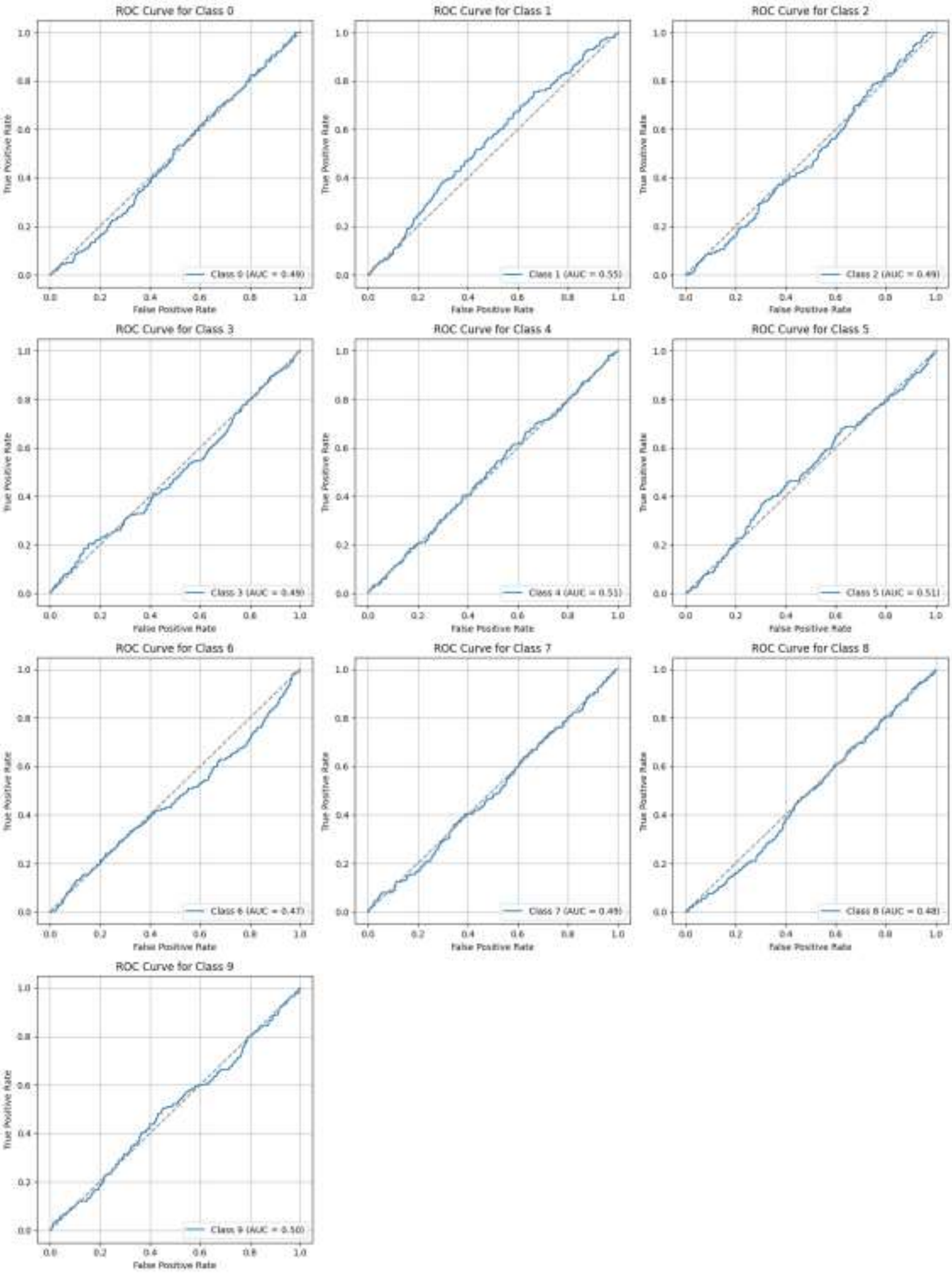


Figure 4: ROC Curve Grid for Each Class

The performance of the model per each class is depicted in the grid of ROC curves. The legend shows the values of AUC, where the larger the value, the better the performance. Classes whose AUC scores are less than 0.5 are less notable.

The model had a test accuracy of 79.5 percent on the validation set, which is successful in differentiating between the healthy tomato leaves and the diseased ones. Although it did a good job with most diseases, Tomato Bacterial Spot and Tomato Early Blight had more misclassification, and hence areas of improvement. The F1-scores and AUC values of these classes were lower, which demonstrates the necessity of a more effective data representation and feature engineering. Such enhancements in future might involve data augmentation, balancing of classes, and transfer learning to improve the performance of the model and contribute to the conservation of biodiversity by early disease detection.

5. Discussion

The overall results of the model were satisfactory with a validation accuracy of 79.5 meaning that the model is able to classify most of the diseases in the dataset correctly. The validation curves indicated consistent improvement during the training process which indicates that the model is not overfitting and has good generalization ability. Nonetheless, the model had some issues with some classes of diseases, especially Tomato Bacterial Spot and Tomato Early Blight, as shown by the low precision and recall of the diseases. It does mean that the model may fail to distinguish between the visual symptoms of diseases that share similarities with each other and this has been observed in other automated disease identification procedures.

The results agree with the existing information on the taxonomy of plant diseases. In the past, CNN-based models have proven to be successful in the classification of different diseases in crops such as tomatoes, potatoes and wheat. However, similar to the results of this study, CNNs are commonly attacked with the issue of the difficulty of isolating the diseases that have similar symptoms. The error in diagnosis of the Tomato Bacterial Spot and Tomato Early Blight is particularly striking since both ailments cause lesions and spots to the leaves and hence look similar. Another study by Sarma et al. (2022) also found out that the issue of incorrect identification of similar diseases with the same symptoms has remained a significant problem of machine learning models applied in the agriculture sector [2]. These issues are especially maximized in case of training datasets not reflecting the diversity of symptoms that would be present in the real-life scenarios. These have implications on the sustainable agriculture. Early detection of the illness will enable the targeted reaction, which will enable the decrease of using pesticides and the diminishing of the ecological imprint of farming. The model could lead to more effective and sustainable disease management practices as the initial disease detection could lead to better disease management practices in accordance with the broader interests of the environmental footprint of agriculture reduction. Further, computerized devices of disease detection like the one that is developed in this project can lead to more precision-based agriculture, which navigates resources more effectively and provides superior crops. As the research by Haile et al. (2023) has shown, the implementation of real-time disease detection in agriculture can result in the improved disease management system and more sustainable operations [17].

However, several limitations are present in the research that would be surmounted in subsequent studies. Firstly, low accuracy and recall on certain diseases, particularly Tomato Bacterial Spot and Tomato Early Blight, show that the model was unable to discriminate between diseases with weak differences. This could be due to the asymmetry of the dataset in terms of classes, where some of the diseases would be underrepresented. The tomato leaf disease dataset that has been used in the research has quite a small number of classes of diseases and adding more disease classes and crop species to the dataset is likely to further improve the model to be more generalized under a larger array of conditions. The problem of the imbalance in the classification of diseases is not a novelty to the literature as in earlier works, researchers have utilized data augmentation and synthetic data generation to improve the performance of a model on underrepresented classes.

The other weakness is that the dataset is not very diverse as it lays emphasis on tomato leaf diseases. This can be a constraint factor of the model as it may not be in a position to generalize to other crops and ecosystems. Despite the fact that tomato crops bond a major target of disease detection, more crops and classes of diseases should be introduced in the dataset to create a more robust model that can be implemented in different farming systems. It has been highlighted in past studies, such as Ge et al. (2022), that the diversity of datasets is an essential factor to improve the generalizability and accuracy of plant disease models in the real world [22]. In addition to this, the photos, which were utilized in this paper, were captured in a controlled environment and this may not be adequate to capture the fluctuation of the actual farm conditions. Such variables as the illumination effects, background sounds, and the orientation and future studies must be focused on the collection of images in an actual agricultural field to improve the success of the model [18].

The potential path of the model enhancement when it comes to the future work is the incorporation of environmental data. The development and intensity of disease may be influenced by environmental factors like temperature, humidity, soil moisture, and others. These factors coupled with image based system to detect the disease would be more comprehensive in managing the disease. The authors state that disease progression is dynamic in nature and adding environmental variables to the disease prediction model can improve the predictive ability of the model in predicting the outbreak of diseases in real-time [17]. Furthermore, there would be a possibility of introducing multi-modal data, so that the model would be able to make more situation-oriented predictions by focusing on the health of plants, along with the environmental conditions.

Moreover, the use of transfer learning is an prospective opportunity of enhancing the performance of the model. The concept of transfer learning assumes the pre-trained models to be applied to a larger dataset and fine-tune it to accomplish a specific operation. It can be applied to overcome the small data limitations in which pre-trained networks learn to acquire general features relying on large and varied data sets. This has shown to boost performance on smaller domain-specific tasks e.g. plant disease classification [15]. This means that transfer learning would enable the model to achieve increased accuracy and generalization especially in diseases that have similar symptoms.

In conclusion, despite the fact that the model is good in terms of classifying tomato leaf disease, some aspects require improvement. The formula to the problem of imbalance in classes, expansion of the dataset to include a larger range of diseases and crops, and the ability to add environmental data will make the model more relevant to the actual agricultural ecosystem. In addition, the transfer learning could also be taken into account, which would contribute to the improvement of the model performance and efficiency. These innovations will be used to build sustainable agriculture and more effective disease management methods, which will help decrease the use of pesticides and conserve biodiversity within the agricultural systems.

6. Conclusion

The results of this work point to the possibility of integrating signal and image processing solutions with deep learning to be able to monitor agricultural ecosystems through the detection of plant diseases, which can be an effective solution to sustainable agriculture. The model had good performance in disease classification with an accuracy of 79.5 percent on the validation set that indicates that the model could generalize well within the dataset with some difficulty in classifying diseases with similar visual symptoms. This model contributes to the biodiversity conservation by making sure that more sustainable farming practices are practiced since the early disease detection is important to enhance the health of plants thereby reducing the overuse of pesticides. The model assists in preventing the proliferation of plant diseases that would otherwise destroy the agricultural productivity and the ecosystem around the area through early detection of diseases. The implications of the findings are significant as it can be stated that automated disease detection system might be applied to optimize resource exploitation, reduce crop losses, and decrease the impact of chemical interventions on the environment. The model can be enhanced in future research, by adding more datasets that represent various disease classes, crops, and environmental conditions that would heighten the generalizability and be implemented in the actual agricultural environment. Transfer learning and environmental data such as temperature, humidity and soil condition can also be part of the model to improve the model performance as it will provide a more specific approach of the plant health with the respect to the environment. In addition, the correlation between plant health monitoring and biodiversity and ecosystem health has to be explicitly explored in such a way that the agricultural systems could assist in preserving the ecosystem. The additional study must deal with integration of multi-modal data and sensor networks to have more adaptive and efficient disease control; this will ultimately build a more sustainable relationship between agriculture and biodiversity conservation.

References

[1].Kerry, R. G., Montalbo, F. J. P., Das, R., Patra, S., Mahapatra, G. P., Maurya, G. K., ... & Rout, J. R. (2022). An overview of remote monitoring methods in biodiversity

conservation. *Environmental Science and Pollution Research*, *29*(53), 80179-80221.

[2].Sarma, K. K., Das, K. K., Mishra, V., Bhuiya, S., & Kaplun, D. (2022). Learning aided system for agriculture monitoring designed using image processing and IoT-CNN. *IEEE Access*, *10*, 41525-41536.

[3].Mulakaledu, A., Swathi, B., Jadhav, M. M., Shukri, S. M., Bakka, V., & Jangir, P. (2024). Satellite Image–Based Ecosystem Monitoring with Sustainable Agriculture Analysis Using Machine Learning Model. *Remote Sensing in Earth Systems Sciences*, *7*(4), 764-773.

[4].Li, X. H., Li, M. Z., Li, J. Y., Gao, Y. Y., Liu, C. R., & Hao, G. F. (2024). Wearable sensor supports in-situ and continuous monitoring of plant health in precision agriculture era. *Plant Biotechnology Journal*, *22*(6), 1516-1535.

[5].Yu, L., Du, Z., Li, X., Zheng, J., Zhao, Q., Wu, H., ... & Huang, X. (2025). Enhancing global agricultural monitoring system for climate-smart agriculture. *Climate Smart Agriculture*, *2*(1), 100037.

[6].Pande, C. B., & Moharir, K. N. (2023). Application of hyperspectral remote sensing role in precision farming and sustainable agriculture under climate change: A review. *Climate Change Impacts on Natural Resources, Ecosystems and Agricultural Systems*, 503-520.

[7].Wang, T., Zuo, Y., Manda, T., Hwarari, D., & Yang, L. (2025). Harnessing Artificial Intelligence, Machine Learning and Deep Learning for Sustainable Forestry Management and Conservation: Transformative Potential and Future Perspectives. *Plants*, *14*(7), 998.

[8].Wätzold, F., Jauker, F., Komainda, M., Schöttker, O., Horn, J., Sturm, A., & Isselstein, J. (2024). Harnessing virtual fencing for more effective and adaptive agri-environment schemes to conserve grassland biodiversity. *Biological Conservation*, *297*, 110736.

[9].Fergus, P., Chalmers, C., Longmore, S., Wich, S., Warmenhove, C., Swart, J., ... & Meijaard, E. (2023). Empowering wildlife guardians: An equitable digital stewardship and reward system for biodiversity conservation using deep learning and 3/4G camera traps. *Remote Sensing*, *15*(11), 2730.

[10]. Borja, A., Berg, T., Gundersen, H., Hagen, A. G., Hancke, K., Korpinen, S., ... & Uyarra, M. C. (2024). Innovative and practical tools for monitoring and assessing biodiversity status and impacts of multiple human pressures in marine systems. *Environmental Monitoring and Assessment*, *196*(8), 694.

[11]. Haile, M. A., Haile, D. T., & Zerihun, D. (2023). Real-time sensor data analytics and visualization in cloud-based systems for forest environment monitoring. *International Journal of Advances in Signal and Image Sciences*, *9*(1), 29-39.

[12]. Dhuldhaj, U. P., Singh, R., & Singh, V. K. (2023). Pesticide contamination in agro-ecosystems: toxicity, impacts, and bio-based management strategies. *Environmental Science and Pollution Research*, *30*(4), 9243-9270.

[13]. Wongchai, A., rao Jenjeti, D., Priyadarsini, A. I., Deb, N., Bhardwaj, A., & Tomar, P. (2022). Farm monitoring and disease prediction by classification based on deep learning architectures in sustainable agriculture. *Ecological modelling*, *474*, 110167.

[14]. Elbasi, E., Mostafa, N., AlArnaout, Z., Zreikat, A. I., Cina, E., Varghese, G., ... & Zaki, C. (2022). Artificial intelligence technology in the agricultural sector: A systematic literature review. *IEEE access*, *11*, 171-202.

[15]. She, J., Zhan, W., Hong, S., Min, C., Dong, T., Huang, H., & He, Z. (2022). A method for automatic real-time detection and counting of fruit fly pests in orchards by trap bottles via convolutional neural network with attention mechanism added. *Ecological Informatics*, *70*, 101690.

[16]. Zhao, W., Wang, M., & Pham, V. T. (2023). Unmanned aerial vehicle and geospatial analysis in smart irrigation and crop monitoring on IoT platform. *Mobile Information Systems*, *2023*(1), 4213645.

[17]. Singh, S., & Jain, P. (2022). Applications of artificial intelligence for the development of sustainable agriculture. In *Agro-biodiversity and Agri-ecosystem Management* (pp. 303-322). Singapore: Springer Nature Singapore.

[18]. Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., ... & Wilson, A. M. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology & Evolution*, *6*(5), 506-519.

[19]. Francini, S., Schelhaas, M. J., Vangi, E., Lerink, B., Nabuurs, G. J., McRoberts, R. E., & Chirici, G. (2024). Forest species mapping and area proportion estimation combining Sentinel-2 harmonic predictors and national forest inventory data. *International Journal of Applied Earth Observation and Geoinformation*, *131*, 103935.

[20]. Sharma, S., & Vardhan, M. (2025). Aelgnet: Attention-based enhanced local and global features network for medicinal leaf and plant classification. *Computers in Biology and Medicine*, *184*, 109447.

[21]. Gill, K. S., Anand, V., Gupta, R., & Pahwa, V. (2023, July). Insect classification using deep convolutional neural networks and transfer learning on MobileNetV3 model. In *2023 World Conference on Communication & Computing (WCONF)* (pp. 1-5). IEEE.

[22]. Ge, Y., Chen, H., Zhang, M., & Li, X. (2022). Area threshold interval of urban forest patches required to maintain the synergy between biodiversity conservation and recreational services: Case study in Beijing, China. *Forests*, *13*(11), 1848.

[23]. Ahmad Ashraf Zargar. (2023). Tomato Leaf Disease Dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/6973251