## An extensive survey of Machine Learning inputs for the identification & forecasting of crop diseases & pests

**Sailaja Madhu1\*, Ravisankar vadali2**

1Gandhi Institute of Technology and Management

Hyderabad, Rudraram, Telangana 502329

1sailajamadhu214@gmail.com

2Gandhi Institute of Technology and Management

Hyderabad, Rudraram, Telangana 502329

2ravisankarphd213@gmail.com

\*Correspondence:sailajamadhu214@gmail.com

**Abstract**

In light of the recent population growth for upcoming 2050, the productivity of all crops worldwide is predicted to double. A major barrier to reaching this production output is pests and illnesses. Therefore, it is crucial to create effective procedures. The need to develop efficient methods for automatically identifying, classifying, and predicting illnesses and pests in agricultural crops. In order to carry out this automation, methods involving from the data being worked on machine learning will be used to extract information and relationships. Classification, identification for forecasting of diseases, with a focus on tomato crops in particular are the main topics of the literature review on machine learning (ML) approaches employed in the agricultural industry. The research aims to advance precision agriculture and smart farming by enticing farmers to use less chemicals and pesticides in maintaining and enhancing crop quality and output.

**Keywords:** Classification, Machine Learning, Tomato disease classification, Precision Agriculture, Smart Farming

# 1. INTRODUCTION

The requirement for crop products has increased in recent decades due to the rapid population growth, which has led to a major increase in agricultural [Fróna, Szenderák, et al., (2019)]. In order to satisfy the increasing population's demand for food, biofuels, and animal products by 2050, crop yield production will need to triple its current output. Major yields for crops must increase by 2.4% yearly to meet this aim as this could be barely doing so at the moment [Ray, Mueller, et al., (2013)]. However, if this requirement is met, the ecosystem will suffer from the emissions of greenhouse gases are raising while biodiversity is disappearing. It is crucial to make the best use of resources like water and soil to enable high- yielding crops because traditional agricultural production is neither economically viable nor environmentally sustainable.

Furthermore pest insects and illnesses are a constant threat to agricultural production. Plant diseases and insect attacks are thought to cause between 20% and 40% of the world's annual agricultural production to be lost, causing the equivalent of $220 billion and seventy billion dollars in losses to the world economy, correspondingly. These losses range in size and frequently result from transnational plant pests and diseases. For instance, compared to other geographic regions, between 1950 and 2000, agricultural diseases and pests were more prevalent in North America [3].

Climate change's increase in average world temperature has an effect on the growth and damage caused by pests. Insects' metabolic rates increase when the temperature rises, causing larvae to eat larger amounts of food and cause additional harm. Numerous factors influence insect growth rates, including temperature. For every degree the earth's surface heats on average, it is estimated that insect pest losses to agriculture will increase by 10% to 25% [Deutsch, Tewksbury, et al., (2018)].Pests have traditionally been managed by farmers using pesticides and chemical treatments. The increased usage of pesticides to preserve crops has a negative impact on human health and damages soil and groundwater more [5].

Expert observation using only one's eyes is how plant diseases are typically found and identified. This is labor-intensive and impractical as a method of overseeing vast farms. Therefore, automated methods for crop monitoring and forecasting are necessary to overcome the shortcomings of manual detection [Kartikeyan, Shrivastava, (2021)]. A system that is capable of carrying out these duties can be crucial in preventing the overuse of pesticides and chemicals, minimizing the harm to the environment and the production costs related to their use [Kamilaris, Kartakoullis, et al., (2017)].

Big data analysis techniques are becoming more widely available, which could lead to more research and development in smart farming. It intends to aid in disease identification, event forecasting, and soil and water management in addition to encouraging more sustainable, better yield crops. By gathering information from weather stations, remote sensors, historical data, and publicly accessible data sets, big data is entering the agriculture sector [Rajpurkar, Irvin, et al., (2017)].Numerous applications of ML have been successful, including the detection of diseases from medical images [Wang, Lee, et al., (2015)], the categorization of images on massive data sets [Soni, Dharmacharya, et a., (2021)], advancements in academic disciplines like physics [Mohanty, Hughes, et al., (2016)] and autonomous vehicles [Carleo, Cirac, et al., (2019)].

The usage of Long Short Term Memory (LSTM) networking may be employed to identify relationships between bug incidence and weather-related information. While deep learning and object segmentation methods can be used to find insects on leaves. For example, it is possible to categories diseases from images for many different kinds of plants with a range of diseases using renowned Convolutional Neural Network (CNN) architectures [Xiao, Li, et al., (2019)]. Predicting impending insect attacks. Farmers now have access to commercial smart farming equipment and services that heavily rely on machine learning. Here are a few instances. The android-based agricultural assistance programme named Plantix was developed by Progressive Environmental and Agricultural Technologies (PETA) in Germany. It assists in identifying plant diseases using visual analysis as well as deep learning methods.

Agrio and Crop Diagnosis are two other instances of related applications [15, 16]. A startup firm called Gamaya is situated in Switzerland and provides a heavy range active agricultural options received on the strategy of photos from remote inspection devices to unnamed aerial vehicle (drones) connected to IoT devices [17].A literature review included in this study intends to guide the creation gives Machine Learning solutions for agricultural holders for accessing to systems that support data-driven decision making. Farmers can be helped in this way to preserve and improve crop quality and yield while also minimizing the necessity of using pesticides and the risks involved. This helps ensure that there will always be enough food to feed everyone on the earth while causing the least amount of environmental harm.

# 2. LITERATURE REVIEW

The three fundamental processes of Machine Learning applications are depicted in Fig-1 gathering information, extracting features, and ML models for classification make up the process of pre-processing data (i.e., data preparation). The next sections illustrate and describe the various techniques used in these three stages.



## Figure 1: Machine learning process

Data acquisition is the procedure used to collect information from various sources and systems [Choudhuri, Mangrulkar, (2021)].For usage with ML approaches, previous research has collected data from a variety of sources. By taking photos of plants in greenhouses, the authors produce their own images based on the findings of Gutierrez et al. [Gutierrez, Ansuategi, et al., (2019)] and Raza et al. [Raza, Prince, et al., (2015)].

However, the construction of efficient ML-based models can be hampered by the fact that many people acquire image data manually, which typically leads to limited image data sets. The literature also recommends acquiring meteorological data for example, by Rustia and Lin [Rustia, Lin, (2017)], who used sensors in greenhouses. Weather stations in rural locations, which frequently maintain data for an extended amount of time, may additionally provide meteorological information. Search engines alone can be used to obtain images. A significant number of images can be obtained using this method, but the accuracy of the data must be verified by subject-matter specialists, and data cleaning is usually employed to remove images that don't correspond to the requirements.[Mokhtar, Ali, et al., (2015); Skawsang, Nagai, et al., (2019)]

Utilizing satellite and drone-taken remote sensing photos has many benefits for large agricultural regions are the ability to access picture data. Crop growth can be assessed using the multi- temporal and hyper-spectral images data that make up satellite-derived remote sensing data. This task can be completed by keeping an eye on the development of vegetation indicators. the capacity to calculate several vegetative indices, such as those provided in [23] [Abdulridha, Ampatzidis, et al., (2020)] [25].A major benefit of spectrum image over conventional full spectral photography is that it produces images that are resistant to variations in the sun's brightness

It is possible to complete this work in observing the inputs of evolution in vegetation indexes. The ability to compute many types of vegetation indexes, which will tell in [Evangelides, Nobajas, (2020); Albetis, Duthoit, et al., (2017); Chandel, Khot, et a., (2021)] and resilient changes in formation of natural energy of Sun illumination is a major benefit of spectral imagery over visible light spectrum photography [Wang, Huang, et al. (2007)].It is possible to recreate a bigger representation of the entire field using images from drones, but there are additional requirements, including specifying the device's path, organizing the drone's place with the camera for taking pictures, and rectifying distortions in geometry on each image acquired. Therefore, it is possible to say that data is composed of a variety of modalities and variables.[30].It may be able to comprehend their interaction and how it relates to a researched outcome using ML-based and data analysis tools. Which illness is hurting crops, is a common topic in the context of farming areas. What bug is harming my property?

# 3. FACTORS AFFECTING CROP INFECTIONS AND PESTS

Accurate agricultural disease and pest detection and identification, as well as the ability to predict when they will manifest, are crucial. Utilizing both the long-term study of weather station data and the current meteorological data collected by unmanned observation aircraft, models that can forecast the presence of diseases have been constructed. The Comprehensive Infectious Approach [Kim, Lee, et al., (2018)] put forward in was utilized in to evaluate the system's predictive capabilities [Yin, Kropff, et al., (1995)]. It was demonstrated that accurate plant disease prediction systems may be built if combined structures like both of these are formed and the different input data sets necessary towards interdependence studies gathered..

Knowing which factors will affect the event being projected is essential when making predictions about future events. This was accomplished in the study by Henderson et al. [Henderson, Williams, et al., (2007)] by determining the weather variables that affect the forecast.. On the other hand Lasso et al. [Lasso, Corrales, et al., (2020)] identified the temporal frame associated with every climatic factor and crop component that is necessary for the growth of the bean leaf rust virus in coffee plantations [Small, Joseph, et al., (2015)].

Using meteorological data volatile organic molecules, The research of Gafferi et al. is concentrated on the extremely initial stage of detection of diseases in tomato crops. In response to physical, biotic, and infectious stress, plants create a wide range of volatile organic chemicals [Ghaffari, Zhang, et al., (2010)]. The illnesses being researched in were powdery mildew and spider mites. To comprehend how location and temperature affect crops, an agricultural model created by Diepeveen et al. in

[Diepeveen, Armstrong, et a., (2008)] might be employed. Additionally, it was discovered that factors like soil, humidity, rainfall, and wetness have an impact on crop output [Holopainen, (2004)].

Weather and environmental factors have a significant impact on the development of plant diseases and pests [Patil, Saiyyad, (2019)]. A favorable environment for the growth of fungal illnesses is humidity. Because of excessive leaf moisture brought on by improper watering techniques or the weather, tomatoes are more prone to diseases like leaf mould and bacterial spots. [Rosenzweig, Iglesius, et al., (2001)].Plants reflects the remaining solar light while absorbing some of it. The radiation absorbed and reflected varies according on the plant's health. This distinction between healthy and ill plants can be utilized to determine the extent of the damage [41]. Figure 2 Temperature illustrates the idea.



## Figure 2: Photon immerse and effectiveness of health of a plant

Since they are ectothermic, insects must obtain their heat from the environment because they are unable to control their internal temperature. Insect population growth and metabolic rates are influenced by temperature [Dake, Chengwei, (September 2006)]. Thus, the number of days with a temperature that is conducive to development has a substantial impact on how long an insect's life cycle lasts. It is possible to distinguish between two temperature thresholds: A greater boundary whereby development in insects slows or stops, and a lower limit where none grows.. This criterion varies depending on distinct bug species.

The relationship among the total number of nuisance catches and temperatures is stronger as compared to other variables in the case of ML-based applications that explore the impact of weather on the development of pest insects. [Henderson, Williams, et al., (2007); Lasso, Corrales, et al., (2020)].The pace at which a plant transpires and, subsequently, its temperature, are both impacted by some illnesses. The temperature of plant leaves can therefore be used to detect diseases.. When thermal and visible light images are combined, ML models can identify diseases more accurately. The advantages are more helpful for early diagnosis when the plant has not yet shown signs that can be seen with the naked eye.

# 3.1 Leaf Reflectance

Photosynthesis-active radiation, which is between 400 and 700 nm in wavelength and roughly corresponds to the visible light area, is absorbed by plants. The substantial increased leaf reflectivity on a magnitude scale that happens for wavelengths longer than 700 nm (red) in the Near Infra-Red (NIR) region is known as the red edge. [Lasso, Corrales, et al., (2020].Due to the fluctuating levels of chlorophyll and the harmed leaf tissue in ailing plants, their leaf spectral reflectance differs from that of a healthy plant. Plants that are ill end up absorbing more NIR light and less visible light. Based on this understanding, disease detection can be carried out utilizing data on leaf reflectance

Remote sensing can be used to retrieve different vegetation indices [23]. The Normalized Difference Vegetation indicator (NDVI) (Figure 2) is a widely used indicator for determining an area's level of vegetation using data on leaf reflectance. NDVI can be calculated from customized cameras or from satellite data It was discovered that NDVI and temperature work better together than they do separately to predict the appearance of pests. To effectively assess the severity of an illness, ML models can also use NDVI as input data [45].

# 3.2 Agriculture Data-Sets

Pictures of agricultural diseases or pests are frequently added to data sets used in agriculture so that they can be categorized. The openly accessible data sets include Plant Village, PlantDoc, IP102, Flavia, and MalayaKew Leaf, to name just a few. Here is a quick summary of each of these: Plant Village is a well-liked data set for categorizing plant diseases [Hughes, Salathé, (2015)]. It includes 18,160 photos of tomato leaves specifically showing those with bacterial spots, early and late blights, leaf mould, target spots, tomato yellow leaf curl virus, spider mites, two-spotted spider mites, septoria leaf spots, and spider mites. There are moreover pictures of wholesome vegetation. Two samples of these data-set images are shown in Figure 3. IP102 [Wu, Zhan, et al., (2019)]: A data set for classifying pests that includes more than 75,000 photos divided into 102 categories. Annotations for bounding boxes are also included in a portion of the image set (19,000 photos).

Due of the wide diversity of insects, their associated developmental stages (egg, larva, pupa, and adult), and visual backgrounds, this data collection is particularly challenging to deal with. Additionally, the data set is very unbalanced. Two samples of photographs from this data collection are shown in Figure 4.

Images from tomato illnesses picked up in the fields are shown in PlantDoc [Singh, Jain, et al., (2020)]. In addition to healthy tomatoes, diseases such as tobacco bacterial spot, tomato early blight, tomato late blight, mould, tomato mosaic virus, tomato septoria leaf spot, tomato yellow virus, and healthy tomatoes have also been considered.Images of lone plant leaves against a white background, without stems, can be found in Flavia [Wu, Bao, et al., (2007)] . As 33 plant species are included in this data set.

The fact that Tomato Powdery Mildew Disease (TPMD) is tied to meteorological data makes it a unique form of data set. It provides statistics on leaf wetness, temperature, wind speed, global radiation, and humidity as indicators of the susceptibility of plants to the powdery mildew disease. [Bakeer, Abdel-Latef, et al., (2013)]



* 1. Mosaic virus illness affects tomato leaves. (b) The pathogen late blight has harmed a tomato leaf.

## Figure 3: Plant Village data set scenarios involving leaves from tomatoes



**with illness**

(a) Rolling of Rice Leaves (Cnaphalocrocis exigua) (b) Small arachinds (Penthaleidae)

**Figure 4**: Images of insects obtained from the IP102 data collection

# 3.3 Data Collected in the Field vs. in the Laboratory

The calibre and kind of input (an image or another type) has an impact on how well ML models work. Photographs captured in the wild versus photographs captured in a controlled laboratory environment can produce entirely distinct procedures and/or outcomes. Those captured in the wild have a far harder time being classified as illnesses and pests than those captured in a controlled environment. There is frequently a solitary plant on a white artificial background in photographs taken in a controlled laboratory setting [Brahimi, Boukhalfa, et al., (2017)].

One such instance is provided by the Plant Village data-set [Hughes, Salathé, (2015)]. On these data sets, excellent performance is achievable. These data sets must be produced; however it takes time and money to do so.Due to the existence of several leaves in a single shot, additional plant parts, different ground textures, different backdrops, different shadow and lighting circumstances, etc.., field photographs are far more complex than images taken in a lab. The research claims that when ML models are trained using laboratory photos and tested in the field, the results are subpar, rendering them unusable for the task. On the other hand, testing on laboratory images and training on field photographs yield reasonable results [Ferentinos, (2018)].

Table 1 displays the findings of some experiences which looked at the conditions of procurement impacted for efficiency predictions categorization models. The letters "L" for lab images, "F" for environment pictures, and "L + F" among different types of photographs are written at each table cell.

**Table 1: Assessment of performance of data from the field and the lab**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pre-trained Weights** | **Training** | **Testing** | **Performance** | **Referred** |
| Imagenet |  | L | 15 % | [68] |
| Image net | F | L | 65 % |  |
| Imagenet + Plant | F | L | 95 % | [63] |
| Image net | L | L | 25 % |  |
| Image net + Plant | F | L | 70 % | [13] |

# 3.4 Pre-Processing of Data

The majority of ML-based applications frequently pre-process prior for supplying it to the model. In order to eliminate noise, improve contrast, extract regions of interest, Image pre- processing sometimes involves utilizing computer vision techniques to extract picture properties and other data. In general, picture pre-processing procedures produce superior model results. The following sub-sections discuss the methods of data pre-processing that are used the most frequently.

# 3.5 Segmentation of Images

Segmenting an image involves classifying pixels into areas of interest. These points of intersection involved in diseased spots on the leaf of the plant, which can be used to determine the complexity of the problem can be counter with number spots that are affected, as well as they can be used to clear background clutter, which enables the regions of interest to be highlighted for further examination. Figure 5 displays a backdrop removal example.



(a) Fragaria Vesca (O) (b) Fragaria vesca (segmented)

## Figure 5. Plant Village information -set.

Blob identification is a method used in image processing to identify pixel regions with similar characteristics. These areas' characteristics, such color and brightness, are very different from those of their surrounds. For example, this method can be used to find and count insects in photos [Albanese, d’Acunto, et al., (2019)]. Popular unsupervised machine learning (ML) algorithm for image segmentation is k-means clustering. Clusters of pixels are formed that have comparable pale and contrast values. For example, this method is useful for identifying damaged leaf areas [Sannakki, Rajpurohit, et al., (2011)]. When using the fuzzy c-means soft clustering technique, A pixel could belong to multiple groups. This method has been used by SekulskaNalewajko and Goclawski [Sekulska-Nalewajko, Goclawski, (2011)] and Zhou et al. [Zhou, Zang, et al., (2013)] for plant.





(a) Botryosphaeria obtusa (o) (b) Botryosphaeria obtusa (g)

**Figure 6:** Conversion of Picture to Grey Scale in plant information set

# 3.6 Feature Extraction:

It is a common step in the preparation of images for shallow imaging.. The Histogram of common image feature extraction algorithms Scale Invariant Feature (SIF), Speeded Up Robust Features (SURF), and oriented Gradient (HoG) (SIFT) transform. Distinct quality Different traits which might or might not appear to be more or less distinct are produced by extractors. appropriate for the current problem.

HoG emphasizes the architecture and By identifying edges on photos that are oriented in various orientations, one may determine the geometry of the image objects. The pattern of distribution of gradients in various directions provides the basis for the development of features. The histogram mostly illnesses have symptoms that influence the color of the leaves, can be utilized to tell healthy plants apart from sick ones .As the Two examples of Computer inputs algorithms for feature extraction in conversion of pictures to grayscale through Haralick and edge approach [Haralick, Shanmugam, et al., (1973)]

A matrix called a "Grey Level Co-occurrence Matrix" (GLCM) tracks how often adjacent grey levels appear together in a given image used to compute the features of the Haralick texture. Every set of The GLCM counts the number of pairs of gray-level pixels in the image. Because a healthy leaf's texture is flatter and a diseased leaf's surface is more uneven, the textures of the two are different. Another method for extracting picture texture features that is resistant to variations in lighting conditions is called Local Binary Pattern (LBP) [Ojala, Pietikainen, et al., (1994)]. Tan et al. in extracted information regarding illnesses from tomato leaves using the LBP approach.[ Tan, Lu, et al., (2021)]

It is possible to provide new data and improve model performance; multispectral image data sets are used. For instance, in [Rosenzweig, Iglesius, et al., (2001)], the fields were originally captured using NIR images, and the authors developed new images from this data using principal component analysis, Band ratios, spectral discrepancies (in the range of NIR and green bands as well as among green and blue bands).The authors evaluate the optimum data type for the models in terms of performance.

# 3.7 Image Cropping and Resizing

Images are cropped and resized to make them smaller for input, to speed up processing, or to meet hardware specifications. It can also be utilised to generate extra a lot of low resolution data, for example, is needed to train the models. photos can be obtained from a small number of high quality images [Rosenzweig, Iglesius, et al., (2001)].

# 3.8 Tabular Data Pre-Processing

The literature reviewed for this paper was regularly stumbled upon tabular data made up of weather records. There are two techniques to integrate Data records with various dates and places, including cross-year data records and cross-location data records and cross-location, when models are tested for the same year in several places. For all ML algorithms examined, cross-year models had better average coefficients of determination. Calling/standardizing data and processing missing values are frequent pre-processing techniques Others, The normalization of feature values can enhance training and minimize the effects of vanishing gradients in systems like neural networks [Amara, Bouaziz, et al., (2017)]. The vast majority of algorithms require that the data be complete and without any missing values.

When there are many records, down sampling is a helpful method of processing the data. The reflectance of leaves was measured in with a 1 nm space between 760 and 2500 nm. Following the compression of the 1740 wavelength data into 174, with the help of the stepwise process, 10 wavelengths were selected. Regression analysis results showed that the link between these wavelengths and leaf severity had a r 2 coefficient of determination of 0.94.Less than those 10 wavelengths would result in inferior performance, according to experiments.

# 3.9 Deep Learning Preprocessing

Since the ability of deep learning to produce features on its own is one of its most important and advantageous characteristics, deep learning pre-processing does not concentrate on feature extraction. Pre-processing is therefore primarily concerned with expanding the number photographs used as input are scaled to match the input parameters of the model and enhanced with data to create a new set of images. Robust learning focus of pre-processing is not extraction of features but have both been compared in several research. Manually chosen features were unable When it came to classifying insects in the field, we were unable to handle the noise of real-world photos or jot down all the important details relating to bug infestations. Additionally, manually chosen traits failed to distinguish minute changes between various insect species that have a similar look [Kamilaris, Prenafeta-Boldú, (2018)].

.In the study by Brahimi et al. in [Brahimi, Boukhalfa, et al., (2017)], models utilizing feature extractors are not far behind deep learning models in terms of accuracy, with values over 94%. However, deep learning models obtain better accuracy, with values above 98%.

The original study found that deep learning versions fared better color photographs when compared to the utilization of images that had been transformed to grey scale or background segmentation [Xiao, Li, et al., (2019)]. These results are supported by [82], which contrasts the performance of color and grayscale images. This affirms the notion that deep learning does not necessitate considerable image pre-processing.

The technique of artificially extending and broadening the training data set is known as data augmentation. By adding variability to the data and enabling better domain generalization, this technique enhances for effectiveness of the models [83]. Many approaches transformations are utilized, including movement, trimming, expansion, and flipping.

Data cleaning is the process of evaluating the data's quality and either modifying or erasing it. It is typically used in studies that automatically gather photographs for their data sets from search engines, excluding images that do not match the labels intended for them or do not meet the minimal resolution standards .Typically, To fit the input parameters of the models, image scaling is carried out[68].Studies that investigated how well Results of the models using various image sizes as inputs led to the following findings that larger images resulted in higher accuracy but also longer training times and greater hardware requirements [71].

The pre-processing methods used on the deep learning classification models examined for this research are shown in Table 2. The employed data preprocessing method is displayed in the 'type' column, and more information about it is provided in the 'info' column.

## Table 2. Data processing of Techniques in Pre-processing of Deep Learning

|  |  |  |
| --- | --- | --- |
| **Type** | **Info** | **Study** |
| Grayscale | - | [Xiao, Li, et al., (2019)] |
| Frame work Segmentation | disguise |
| Rescale | 256 \* 256 |
| Input data Cleaning | Perspective, rotation, and affine | [Duarte-Carvajalino, Alzate, et al., (2018)] |
| Data Expansion |
| Resize |
| Data Expansion | Scale, flip, Gaussian noise, rotation, crop | [15] |
| Rescale | 300 \* 300 , 600 × 1024 |
| Grayscale | - | [82] |
| Rescale | 60 \* 60 |
| Data Cleaning | - | [Kamilaris, Prenafeta-Boldú, (2018)] |
| Resize | 224 \* 224 |

It is clear from the table that image scaling was used in all of the analyzed studies that used deep learning-based techniques. It's also important to note that 25% of the depicted tasks used data augmentation, and the same is true for data cleaning and image color conversion to grayscale.

## 3.10 Models for Machine Learning

Machine learning (ML) models enable investigators to gain knowledge about data and existing links among a variety of Variables that affect how frequently crops are attacked by pests and diseases.. Among other goals, models can be used for regression and classification. After the data has been processed and the features have been extracted. In classification, a label is given to a fresh data sample based on the relationships discovered during training. Regression uses input variables to estimate a continuous output value.

## 3.11 SVM, Support Vector

The SVM [84] model divides two classes into a hyper-plane; it can also be modified and used for multi-class issues. SVM selects the ideal hyper-plane to divide the data by maximizing the separation, or margin, among each class's nearest points of data (support vectors), measured in terms of distance from the hyper plane. The so-called kernel trick method can be used to improve SVM's performance on non-linear information. The SVM kernel converts a low-dimensional input space into a higher-dimensional, linearly separable space.

SVM can therefore be quite successful in high-dimensional spaces. Regression issues can also be solved with SVM [Wang, Huang, et al., (2007); Rosenzweig, Iglesius, et al., (2001)]. As illustrated by Bhatia et al. in [Bhatia, Chug, et al., (2020)], who used Support Vector Machine with the logistic regression technique to predict the powdery mildew infection in tomato plants, Support Vector Machine will also be used in a hybrid form..

Table 3 displays a compound/ combination of studies in agriculture utilizing SVM as the ML model. It is possible to observe the type of SVM employed, as well as its kernel and outcome. The most often utilized kernels using methods for regression and classification based on SVM employed in agricultural applications appear to be linear, polynomial, and RBF kernels.

## Table 3: Performance of Support Vector Machine

|  |  |  |  |
| --- | --- | --- | --- |
| **Regression / Categorization** | **Model** | **Outcome** | **Study** |
| Classification | Radial Basis Function | 90.5% | [Ferentinos, (2018)] |
| Multi-Layer Perception | 91.0% |
| Quadratic |  |
| Polynomial |  | [Rosenzweig, Iglesius, et a., (2001)] , [Wang, F.M., Huang, et al., (2007)] |
| Regression | Linear | 90.0 % |
| Regression | Linear | r2 =.45 |
| Classification |  |  | [25] |
| Classification | Quadratic | 90.1 % |
| Linear |  | [Kim, Lee, et al., (2018)] |
| Multi Layer Perception |  |
| Polynomial |  |
| Radial basis function | 92.0 % |

**3.12 Random Forest:**

Random Forest (RF) is a popular composite of non-parametric (decision Tree) that have been structured & trained using different types of Structured, Unstructured and semi-structured data of the training. Additionally, rather than taking into account all of the features, RF evaluates a random collection of variables when determining how a node should split a variable.

Each tree casts a vote, and the class with the most support is returned during categorization. The computation is quick since each tree is trained using a small selection of characteristics and data. They are resistant to noise and outliers since they have a large number of trees and that each one is diverse. Table 4 displays a few research that used Random Forest (RF).

**Table 4:** Random Forest Inputs and Performance

|  |  |  |  |
| --- | --- | --- | --- |
| **Classification** | **No of Tress** | **Performance** | **Study** |
| Regression | 100 | r 2 = .75 | [Rosenzweig, Iglesius, et a., (2001)] |
| Classification |  | 70.0 % |  |
| Regression | 200 | r 2 = .75 | [Yin, Kropff, et al., (1995)] |
| Classification |  | 94.0 % |  |

## 3.13 Artificial Neural Networks:

The biological brains serve as inspiration for artificial neural networks (ANN). The input, hidden, and output layers of an ANN are composed of neurons and each layer may contain a number of hidden layers and neurons. A representation of an ANN is shown in Figure 7 .where it may build high-order features by learning complicated connections from the hierarchical fusion of multiple inputs. ANNs with many layers are thought to be used in deep learning.

Optimization had been repetitive approach in minimizing blunder that in most cases is based on the Gradient Descent algorithm. Optimization is how learning happens. The optimization procedure often employs groups of records known as batches rather than computing the incline of the total data set.



## Figure 7: Artificial Neural Network

The weights are then changed in accordance with how much each layer contributed to the error as the error is carried back through the network one layer at a time. Back-propagation is the name of this update procedure. A training period is finished once all the inputs are recorded in the data set have been constrained. It may take several epochs to train the network to the desired state.

A type of deep learning network called a CNN is frequently used for image categorization jobs. The inclusion of so-called convolutional layers in this sort of network permits a hierarchical extraction of features, with more specific and complicated information being extracted in deeper levels while basic elements like edges are extracted in the initial layers.Another kind of deep learning network is recurrent neural networks (RNN), which are frequently used with time series data. RNNs can capture temporal correlations and automatically extract features from data. The changes in incline utilized for updating the weights in these networks have the potential for becoming unsteady, either rising or declining too quickly (exploding gradient or vanishingincline). Different RNNs are created by structuring the recurrent layers in a variety of different ways. Hochreiter and Schmidhuber proposed the LSTM cell in [Hochreiter, Schmidhuber, (1997)]. Here, the conventional recurrent cell's memory capacity was enhanced to address unwelcome reliance on the long-term.

## 4. NETWORK ARCHITECTURES THAT USERS DEFINE

Studies with custom neural network architectures defined by the authors are presented in this subsection.Duarte-Carvajalino et al. constructed two distinct neural network models and compared the results [Rosenzweig, Iglesius, et a., (2001)]. The initial model was a Multilayer Perception (MLP) with two hidden layers, each with half the number of nodes as the preceding one.. The authors employed batch normalization, a dropout probability of 0.2 in all layers, the Adamax optimizer, a rate of learning of 0.01, and ReLU as an activating mechanism.. The same MLP hyper parameters were used to train the second model, a CNN. Two convolutional layers with 20 filter kernels of size 3 each are followed by a size 2 2 max pooling layer.

In time series data, i.e., winter and fall data, were processed using an LSTM network. Two completely connected layers with five hidden units each made up the LSTM network. In comparison to RF, SVM, and K-Nearest Neighbors (KNN), the results showed that the LSTM network had the greatest performance with 92% accuracy. The interpretability was checked using the Apriori technique [Agrawal, Srikant, (1994)].

A model put up by Patil and Kumar [Patil, Kumar, (2021)] sought to establish a causal relationship between meteorological factors and the formation of four different rice disease kinds. In this study, the scientists employed an ANN to carry out disease detection, identification, and prediction in rice fields. The meteorological data set included information from 1989 to 2019.

In the time series of data for the LSTM network with winter and autumn data. As the fully connected lay As the LSTM network contains 2 fully connected layers .Winter and fall time series data were processed using an LSTM network [Gutierrez, Ansuategi, et al., (2019)]. Two completely interaction of every single layer consisted with five concealed units each made up LSTM network. In comparison to RF, SVM, and K-Nearest Neighbors (KNN), the results demonstrate that, with a precision of 90%, the network using LSTM performed at its best. For interpretability, the Apriori method was employed. A model put up by Patil and Kumar in sought to establish a causal relationship between meteorological factors and the formation of four different rice disease kinds. In this study, the scientists employed an ANN to carry out disease detection, identification, and prediction in rice fields. The meteorological data set included information from 1989 to 2019.

## 4.1 Architectures of Convolutional Neural Networks

With the development of numerous model designs over the previous ten years, image classification has produced excellent results. The "Large Scale Visual Recognition Challenge" (ILSVRC), It was the environment in which most of these deep learning models were proposed. Popular designs like ResNet, VGG, and AlexNet, which have been successfully used to classify images in a variety of application domains, are among these models.

Table 5 highlights several studies that used already present CNN designs, outlining the construction techniques they used and the results they produced. Table 5 shows that numerous CNN architectures created over the past ten years have been effectively applied and have a lot of potential for use in agriculture applications.

## 4.2 Transfer Learning:

To solve the issue at hand, TF applies information that has already been acquired for a related task or subject. Typically, For the study's specific data set, methods which were previously developed for image categorization on large data sets are used and adjusted.. A method that's regularly employed to modify a trained network for a different classification objective is to swap out as dense layers. Following that, at the Current stage of the networks it is maintained the layer weights of the trained model do not change when reused on a subsequent downstream mission as they remain Frozen. So only the recently added layers are trainable. The refinement of this strategy is also frequently employed. In addition to Fine-tuning enables training of additional base model layers, often the network's deeper convolutional layers, after the newly added layers have been trained.

## Table 5: Study and Comparison of methods of CNN Model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sno** | **Data Set****Retrained** | **Method** | **Model** | **Study** | **Comparison** |
| 1 | Imagenet and/or Plant Village | All layers trainable | Alex Net, GoogLeNet | [Xiao, Li, et al., (2019)] | 2% acc. |
| 2 | Imagenet | original layers learn at a low rate (0.1), while the top layer learns ata high rate (10) | CaffeNet | [Duarte-Carvajalino, Alzate, et al., (2018)] | 0.50% acc. |
| 3 | Imagenet | learning at a modest rate for the base layers (0.1), but at a high ratefor the top layer (10) | CaffeNet | [Duarte-Carvajalino, Alzate, et al., (2018)] | 0.50% acc. |
| 4 | Coco | Fine tune | RCNNs that run more quickly (ResNet101, Inception V2, and InceptionResNet V2) | [Yang, Guo, et al., (2022)] | No comparison |
| 5 | Imagenet | Fine tune | Alex Net, GoogLeNet | [Yang, Guo, et al., (2022)] | 2% |
| 6 | Imagenet | Fine tune |  |  |  |
| 7 | Imagenet | Fine tune | Alex Net, GoogleNet, VGGNet,ResNet | [Kamilaris, Prenafeta-Boldú, 2018] | Best model of 14 % (ResNet) |

**4.3 Data Gathering**

ML models' behavior in terms of successful outcomes and robustness will depend on the amount and quality of data, therefore on how well the ML model performs; a lot will depend on how the data is collected. When leaves are photographed under laboratory settings, featuring a solitary leaf on a background of neutral hues, disease classification has produced encouraging results [Xiao, Li, et al., (2019)]. Typically, regulated lighting and white backdrops are used to gather laboratory picture data. However, because of the variety of leaves and plants, the different shades of light and shade, the distinctive ground textures, and the background items, photographs taken in cultivation fields are significantly more difficult to categories.

Since weather conditions have a significant influence in the development of illnesses and pests, ML-based algorithms have been used in several research to analyze meteorological data sets. Data records are filled with meteorological data from sensors, including temperature, humidity, and rainfall, in order to forecast these kinds of changes. You can use these data records to determine whether there is a connection among the information and the increase of insects or diseases that affect plants.

Deep learning methods can be used to examine historical records of climatic observations and insect occurring, as shown in. To manually quantify pest occurrence, insects can be counted or computer vision techniques can automatically find insects in traps or on plants

When determining the cause of plant temperature-related disorders, thermal photography mixed with colored photos performs more accurately than utilizing color features alone. This is particularly helpful for spotting specific illnesses early on when the plant has not yet shown any outward symptoms. When choosing what kinds of data should be collected, Understanding the elements that will influence how the plantation is being worked on is essential.

# 4.4 Pre-processing of Data

Different data pre-processing methods are employed depending on the ML-based strategy. Feature extraction for picture Deep learning can be used to process data automatically or manually using computer vision techniques.

Typically, the use of manual feature extraction methods necessitates the use of pre-processing techniques like noise reduction and contrast amplification. Which feature extractors are better suited for the current issue must be decided by the researchers. Data augmentation, which involves expanding the training data set to increase model generalization, is a common topic of pre-processing for deep learning. When analyzing the original photos, deep learning performs better than using images that have been transformed to grayscale or have had their backgrounds removed

This is a helpful discovery because, for photographs captured in the outdoors with a complex and variable background, background removal can be a challenging and time-consuming task [68]. In research that evaluated both methods using the same input data, the deep learning models outperformed the models based on human feature selection when performance was compared grayscale or had the backdrop removed were two examples.

# 4.5 Models for Machine Learning

The majority of the investigations discussed in this paper used ML models that were based on Support vector machine (SVM), Random Forest (RF) or Deep learning (DL). These are having positive outcomes, demonstrating the promise of applying ML approaches to the disease and pest classification, detection, and forecasting. Due to its usage of the kernel method, Support vector machine have reliable in places with high dimensions. Due to the enormous number of branches trained on various data subsets, RF is able to prevent over fitting. Using the inputs as a starting point, deep learning can produce and extract hierarchical features; it typically produces the best classification results. When leveraging Deep learning performs better than other ML models, especially in the fields of image classification, when compared to already existing CNN models like Inception and ResNet [Wu, Zhan, et al., (2019), Singh, Jain, et al., (2020)].

Despite the fact that models based on Deep Learning are receiving accuracy scores, Support Vector Machine and RT are also the big inputs with approximate 91 % for disease strategy classification from the practical expose and here SVM gives the performance of detection of vegetable tomato affliction towards outputs of 90 % [Duarte-Carvajalino, Alzate, et al., (2018)].RNNs outperform other models like RF and SVM in their ability to create correlations under forecasting information and weather data and pest occurrence .Models trained with less data may benefit from using TF rather than starting from scratch in situations when data is difficult to come most pre-train their models using sizable image classification data sets like Image Net or COCO.

Pre-training algorithms for disease classification on field-collected photos are made more accurate by including the Plant Village dataset with Image Net [Wu, Zhan, et al., (2019)]. In order to apply TF, it is frequently necessary to train both the newly created classifier and a portion of the highest layering of the previously learned model together. An alternative would be to use few-shot learning techniques to solve lack-of-data issues, as mentioned in [Yang, Guo, et al., (2022)].

# 5. CONCLUSION:

This review summarized recent research on machine learning methods detects prediction, and categorization of diseases and pests. Input information sets of Data related to the environment, illness, and pests ought to keep records for a very long period. The intervals of series Machine Learning models like Recurrent Neural Network, which are based on a series of meteorological measurements, can be used to precisely anticipating presence of diseases and pests. NDVI measures are also beneficial because they reveal more details regarding the development of the crop.

Using more effective deep learning techniques built around models created by CNN and computer vision, it is possible to detect and classify pests and diseases.

When compared to earlier methods for classifying images that relied over "manual" feature analysis. Some of the models from Deep learning will give a call about different sorts of information, which will be challenging to get. This issue might be solved using transferring knowledge or few-shot learning approaches.. As Evaluation of field-taken photographs taken under genuine circumstances, still needs more investigation, even if deep learning-based systems perform well for photographs gathered under controlled conditions.

As the consideration of the this article to make put on some of the Machine learning implementations on different ways of information and here Significant research on infectious disease and pest prediction using a variety of data modalities is still lacking in the literature. This was done to make way for potential future beneficial projects.

## Acknowledgement

## Not applicable to this paper

## Conflict of Interest

All authors declare no conflict of interest

## Data Availability

## Not applicable to this paper

## Ethics Approval

Not applicable to this paper

## REFERENCES:

Abdulridha J, Ampatzidis Y, Kakarla SC, Roberts P (2020) Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. Precis. Agric. 21:955–978.

Agrawal R, Srikant R (September 1994) Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago, Chile, 12–15. 1215:487–499

Albanese A, d’Acunto D, Brunelli D (2019) Pest detection for precision agriculture based on iot machine learning. In International Conference on Applications in Electronics Pervading Industry, Environment and Society; Springer: Berlin/Heidelberg, Germany 65–72.

Albetis J, Duthoit S, Guttler F, Jacquin A, Goulard M, Poilvé H, Féret JB, Dedieu G (2017) Detection of Flavescence dorée grapevine disease using unmanned aerial vehicle (UAV) multispectral imagery. Remote Sens. 9:308.

Amara J, Bouaziz B, Algergawy A (2017) A deep learning-based approach for banana leaf diseases classification. In Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband; Gesellschaft für Informatik e.V.: Bonn, Germany.

Bakeer A, Abdel-Latef M, Afifi M, Barakat M (2013) Validation of tomato powdery mildew forecasting model using meteorological data in Egypt. Int. J. Agric. Sci. 5: 372

Bhatia A, Chug A, Singh AP 9(February 2020) Hybrid SVM-LR classifier for powdery mildew disease prediction in tomato plant. In Proceedings of the 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 27–28. 218–223

Brahimi M, Boukhalfa K, Moussaoui A (2017) Deep learning for tomato diseases: Classification and symptoms visualization. Appl. Artif. Intell. 31:299–315.

Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborová L (2019) Machine learning and the physical sciences. Rev. Mod. Phys. 91:045002

Chandel AK, Khot LR, Sallato B (2021) Apple powdery mildew infestation detection and mapping using high-resolution visible and multispectral aerial imaging technique. Sci. Hortic. 287:110228.

Choudhuri KBR, Mangrulkar RS (2021) Data Acquisition and Preparation for Artificial Intelligence and Machine Learning Applications. In Design of Intelligent Applications Using Machine Learning and Deep Learning Techniques; Chapman and Hall/CRC: Boca Raton, FL, USA 1–11.

CropDiagnosis. Available online: https:/[/www](http://www.cropdiagnosis.com/portal/crops/en/home).[cropdiagnosis.com/portal/crops/en/home](http://www.cropdiagnosis.com/portal/crops/en/home) (accessed on 11 January 2022)

Dake W, Chengwei M (September 2006) The support vector machine (SVM) based near-infrared spectrum recognition of leaves infected by the leafminers. In Proceedings of the First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC’06), Beijing, China, 30 August–1. 3:448–451.

Deutsch CA, Tewksbury JJ, Tigchelaar M, Battisti DS, Merrill SC, Huey RB, Naylor RL (2018) Increase in crop losses to insect pests in a warming climate. Science 361:916–919

Diepeveen D, Armstrong L, Vagh Y (August 2008) Identifying key crop performance traits using data mining. In Proceedings of the IAALD-AFITA-WCCA Congress 2008 (World Conference on Agricultural Information and IT), Tokyo, Japan 25–27.

Duarte-Carvajalino JM, Alzate DF, Ramirez AA, Santa-Sepulveda JD, Fajardo- Rojas AE, Soto-Suárez M (2018) Evaluating late blight severity in potato crops using unmanned aerial vehicles and machine learning algorithms. Remote Sens. 10:1513

Earth Observing System Vegetation Indices to Drive Digital Agri Solutions. Available online: https://eos.com/blog/vegetationindices/ (accessed on 11 January 2022).

Evangelides C, Nobajas A (2020) Red-Edge Normalised Difference Vegetation Index (NDVI705) from Sentinel-2 imagery to assess post-fire regeneration. Remote Sens. Appl. Soc. Environ. 17:100283. [CrossRef]

Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145:311–318

Food and Agriculture Organization of the United Nations FAOSTAT Pesticides Use. Available online: https:/[/www](http://www.fao.org/).[fao.org/](http://www.fao.org/) faostat/en/#data/RP/visualize

Fróna D, Szenderák J, Harangi-Rákos M (2019) The challenge of feeding the world. Sustainability 11:5816.

Gamaya. Available online: https:/[/www](http://www.gamaya.com/).[gamaya.com/](http://www.gamaya.com/) (accessed on 11 January 2022).

Ghaffari R, Zhang F, Iliescu D, Hines E, Leeson M, Napier R, Clarkson J (July 2010) Early detection of diseases in tomato crops: An electronic nose and intelligent systems approach. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23. 1–6.

Gutierrez A, Ansuategi A, Susperregi L, Tubío C, Ranki´c I, Lenža L (2019) A benchmarking of learning strategies for pest detection and identification on tomato plants for autonomous scouting robots using internal databases. J. Sens. 5219471

Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621

Henderson D, Williams CJ, Miller JS (2007) Forecasting late blight in potato crops of southern Idaho using logistic regression analysis. Plant Dis. 91:951–956.

Henderson D, Williams CJ, Miller JS (2007) Forecasting late blight in potato crops of southern Idaho using logistic regression analysis. Plant Dis. 91:951–956.

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput. 9:1735–1780

Holopainen JK (2004) Multiple functions of inducible plant volatiles. Trends Plant Sci. 9:529–533

Hughes D, Salathé M (2015) An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv. arXiv:1511.08060.

Kamilaris A, Prenafeta-Boldú FX (2018) Deep learning in agriculture: A survey. Comput. Electron. Agric. 147:70–90.

Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX (2017) A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143:23–37

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

Kim S, Lee M, Shin C (2018) IoT-based strawberry disease prediction system for smart farming. Sensors 18:4051. [CrossRef]

Lasso E, Corrales DC, Avelino J, de Melo Virginio Filho E, Corrales JC (2020) Discovering weather periods and crop properties favorable for coffee rust incidence from feature selection approaches. Comput. Electron. Agric. 176:105640.

Lasso E, Corrales DC, Avelino J, de Melo Virginio Filho E, Corrales JC (2020) Discovering weather periods and crop properties favorable for coffee rust incidence from feature selection approaches. Comput. Electron. Agric. 176:105640.

Mohanty SP, Hughes DP, Salathé M (2016) Using deep learning for image-based plant disease detection. Front. Plant Sci. 7:1419

Mokhtar U, Ali MA, Hassanien AE, Hefny H (2015) Identifying two of tomatoes leaf viruses using support vector machine. In Information Systems Design and Intelligent Applications; Springer: Berlin/Heidelberg, Germany 771–782.

NDVI and EVI.https://earthobservatory.nasa.gov/features/MeasuringVegetation/ measuring\_vegetation\_2.php , Measuring \_vegetation

Ojala T, Pietikainen M, Harwood D (1994) Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1. 582–585.

Patil NN, Saiyyad MAM (2019) Machine learning technique for crop recommendation in agriculture sector. Int. J. Eng. Adv. Technol. 9:1359–1363

Patil RR, Kumar S (2021) Predicting rice diseases across diverse agro-meteorological conditions using an artificial intelligence approach. PeerJ Comput. Sci. 7:e687.

PlantVillage Tomato | Diseases and Pests, Description, Uses, Propagation. Available online: https://plantvillage.psu.edu/ topics/tomato/infos (accessed on 11 January 2022)

Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, et al. (2017) Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv. arXiv:1711.05225.

Ramesh S, Hebbar R, Niveditha M, Pooja R, Shashank N, Vinod P (April 2018) Plant disease detection using machine learning. In Proceedings of the 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India, 25–28. 41–45.

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. PLoS ONE 8:e66428.

Raza SEA, Prince G, Clarkson JP, Rajpoot NM (2015) Automatic detection of diseased tomato plants using thermal and stereo visible light images. PLoS ONE 10:e0123262.

Rosenzweig C, Iglesius A, Yang XB, Epstein PR, Chivian E (2001) Climate Change and Extreme Weather Events-Implications for Food Production, Plant Diseases, and Pests; NASA Publications: Nebraska, NU, USA.

Rustia DJA, Lin TT (2017) An IoT-based wireless imaging and sensor node system for remote greenhouse pest monitoring. Chem. Eng. Trans. 58:601–606

Saillog. Available online: https:/[/www](http://www.saillog.co/).[saillog.co/](http://www.saillog.co/) (accessed on 11 January 2022).

Sannakki SS, Rajpurohit VS, Nargund V, Kumar A, Yallur PS (2011) Leaf disease grading by machine vision and fuzzy logic. Int. J. 2:1709–1716.

Sekulska-Nalewajko J, Goclawski J (May 2011) A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means clustering. In Proceedings of the Perspective Technologies and Methods in MEMS Design, Polyana, Ukraine, 1–14. 172–175.

Significant Remote Sensing Vegetation Indices a Review of Developments and Applications. Available online: https:/[/www](http://www/). hindawi.com/journals/JS/2017/1353691/ (accessed on 11 January 2022).

Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N (January 2020) PlantDoc: A dataset for visual plant disease detection. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, Hyderabad, India, 5–7. 249–253.

Skawsang S, Nagai MK, Tripathi N, Soni P (2019) Predicting rice pest population occurrence with satellite-derived crop phenology, ground meteorological observation, and machine learning: A case study for the Central Plain of Thailand. Appl. Sci. 9: 4846.

Small IM, Joseph L, Fry WE (2015) Development and implementation of the BlightPro decision support system for potato and tomato late blight management. Comput. Electron. Agric. 115:57–65.

Soni A, Dharmacharya D, Pal A, Srivastava VK, Shaw RN, Ghosh A (2021) Design of a machine learning-based self-driving car. In Machine Learning for Robotics Applications; Springer: Berlin/Heidelberg, Germany 139–151

Tan L, Lu J, Jiang H (2021) Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods. AgriEngineering 3:542–558

The Future of Food and Agriculture: Trends and Challenges; FAO: Rome, Italy, 2017.

Wang FM, Huang JF, Tang YL, Wang XZ (2007) New vegetation index and its application in estimating leaf area index of rice. Rice Sci. 14:195–203.

Wang L, Lee CY, Tu Z, Lazebnik S (2015) Training deeper convolutional networks with deep supervision. arXiv. arXiv:1505.02496

Wu SG, Bao FS, Xu EY, Wang YX, Chang YF, Xiang QL (December 2007) A leaf recognition algorithm for plant classification using probabilistic neural network. In Proceedings of the 2007 IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 15–18. 11–16.

Wu X, Zhan C, Lai YK, Cheng MM, Yang J (June 2019) Ip102: A large-scale benchmark dataset for insect pest recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20. 8787–8796.

Xiao Q, Li W, Kai Y, Chen P, Zhang J, Wang B (2019) Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. BMC Bioinform. 20:688

Yang J, Guo X, Li Y, Marinello F, Ercisli S, Zhang Z (2022) A survey of few-shot learning in smart agriculture: Developments, applications, and challenges. Plant Methods 18:28.

Yin X, Kropff MJ, McLaren G, Visperas RM (1995) A nonlinear model for crop development as a function of temperature. Agric. For. Meteorol. 77:1–16

Zhou Z, Zang Y, Li Y, Zhang Y, Wang P, Luo X (2013) Rice plant-hopper infestation detection and classification algorithms based on fractal dimension values and fuzzy C- means. Math. Comput. Model. 58:701–709