



E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating (2026): 5.20
www.agronomyjournals.com
2026; 9(5): 371-379
Received: 12-03-2026
Accepted: 16-04-2026

Onkar R Hirwe
Department of Agronomy, School of
Agriculture, Lovely Professional
University, Phagwara, Punjab, India

Ekamjot Kaur
Department of Soil Science, School of
Agriculture, Lovely Professional
University, Phagwara, Punjab, India

Nitish Ahlawat
Department of Soil Science, School of
Agriculture, Lovely Professional
University, Phagwara, Punjab, India

Manzoor A Khanday
Department of Statistics, Lovely
Professional University, Phagwara,
Punjab, India

Surya Dev Singh
Department of Agronomy, School of
Agriculture, Lovely Professional
University, Phagwara, Punjab, India

Aniket Singh
Department of Agronomy, School of
Agriculture, Lovely Professional
University, Phagwara, Punjab, India

Kiran More
Department of Agronomy, School of
Agriculture, Lovely Professional
University, Phagwara, Punjab, India

Corresponding Author:
Kiran More
Department of Agronomy, School of
Agriculture, Lovely Professional
University, Phagwara, Punjab, India

International Journal of Research in Agronomy

Crop monitoring and disease detection: Using statistical analysis of sensor data to detect crop diseases and pests

Onkar R Hirwe, Ekamjot Kaur, Nitish Ahlawat, Manzoor A Khanday, Surya Dev Singh, Aniket Singh and Kiran More

DOI: <https://www.doi.org/10.33545/2618060X.2026.v9.i5e.5533>

Abstract

This review paper provides the state of art in using statistical analysis of sensors for accurate and early identification of diseases and pest in crops. Discussed the multiplied applications of hyperspectral, multispectral, thermal and LiDAR approaches in precisions agriculture and highlighting their roles in distinguishing small changes that are Potential stress or infestation. The essence of the review centers on the intricately applied methodological analysis of the sensor data through the application of machine learning algorithms, multivariate analysis of variance for the reduction of the voluminous sensor data into intelligent and manageable patterns, and time series modeling. Discussed the relative advantages and disadvantages of diverse analytical methods, pay attention to the importance of the proper choice of an approach depending on the data type and the nature of a disease/pest. Moreover, some of the main issues in the preprocessing of text, feature engineering, and model evaluation to consider the importance of stable analytical workflows are discussed. Last but not the least, outline possible future directions of development in this swiftly growing field, with a focus on the primary future directions that consist in increasing the use of artificial intelligence and big data analytics for the purposes of more accurate predictions and in development of user-friendly tools for farmers.

Keywords: Precision, crop, disease, pest, detection, statistical

1. Introduction

Crop diseases and pests essentially continue to be some of the major concerns to food safety worldwide since they cause low yields, great financial losses and quality down-grades of produce. Currently, diseases and pests are said to be lowering crop productivity by almost 40 percent globally (Khan *et al.*, 2023) ^[16]. Such losses do not only harm the livelihoods of millions of farmers, they also contribute to the already well-known fact that food insecurity is a problem in the world today depending on the environmental conditions of the region where the agricultural practices are being carried out. Market diseases and pests control methods include the common conventional strategies of disease and pest control, which involve pesticide use, manual infestation monitoring, and others. However, it has several critical drawbacks that arise due to the application of the conventional methods. For instance, when the farmer applies a chemical pesticide the pest species may mutate and after sometime they become immune to the chemical causing the pesticide to be less effective (Kumar *et al.*, 2025) ^[17]. In addition, such chemicals are environmentally damaging in that they contribute to soil degradation, water pollution, and disruption of life forms beneficial to other species of insects and other natural predators. There are, however, two major disadvantages –these are manual scouting and field inspections which entails a lot of time, is very expensive, and does not help to identify early stages of infestation or diseases. Such changes may include the appearance and development of sensor-based technologies as well as improvements in statistical analysis of crop protection. These new emerging techniques include hyperspectral, thermal imaging, and fluorescence imaging which are used to monitor the crop health in real-time mode (Zhang *et al.*, 2024) ^[46]. For instance, it has become possible to use hyperspectral imaging to identify diseases such as those affecting crops like wheat and soybean, including the early stage that clinical symptoms have not yet manifested. In the same way, thermal sensors can also measure small variations in

the temperatures of the canopy, which may indicate mild water stress or pest infestation, so irrigation and pest control measures can be applied appropriately.

The combination of these sensor technologies with the so-called data analytics and new technologies underpinning machine learning took precision agriculture to the next level. Using machine learning approach, the large datasets obtained from sensors capture able to predict disease outbreak and pest's invasion with a lot of precision (Mittal *et al.*, 2024) ^[20]. For example, in analyzing image data from drones, deep learning has accuracy rates of more than 95% for various plant diseases. The use of climate data, soil information, and history of pest appearances increases the capabilities of statistical models and allows farmers to act not based on reaction on some events but on the prevention of them. They also enhance the specific nature of the intervention and, by the same token, the efficiency in the use of inputs to reduce the negative impact on the environment (Kumar *et al.*, 2024) ^[18]. Furthermore, real-time data collection through IoT (Internet of Things) devices facilitates continuous monitoring and rapid decision-making. In various settings such as green houses, the networks enable murmurs for managing disease and improving yield by changing irrigation, ventilation, and fertility provision depending on what the plant has to say. The combination of sensors, connectivity and big data is redefining crop conditions and management and this is made

possible due to ability of identifying stress signal in the early stage and application of targeted treatment that is cost effective and has least impact on the environment. The other benefit of the new tools is that they are flexible and easy to scale from professional to amateur use. Drones are ideal for large scale farming ventures that oversee large acreage while an effective device that smallscale farmers could use include smart phones with sensors to monitor fields (Paul *et al.*, 2022) ^[27]. This has had the effect of reducing the distance between research and practice in the advancement of agriculture to help develop a stronger, more food secure globe. The increasing effects of diseases and pest on crops demanding increase yield, and which is beyond the ability of traditional technologies to manage, have pushed the development of sensor and data technology in modern agriculture. The following strategies, based on real-time detection, the use of analytics, and machine learning, have a great potential to revolutionize crop protection. Apart from preventing yield losses, these technologies also constitute sustainable and environmentally friendly farming methods since the problems are detected at an early stage and intervenel (Raj *et al.*, 2021) ^[32]. Further studies, innovations in technology as well as application of multiple discipline approaches will be necessary for enhancing the effectiveness of utilisation of crop protection through data-intensive approaches (Figure 1).

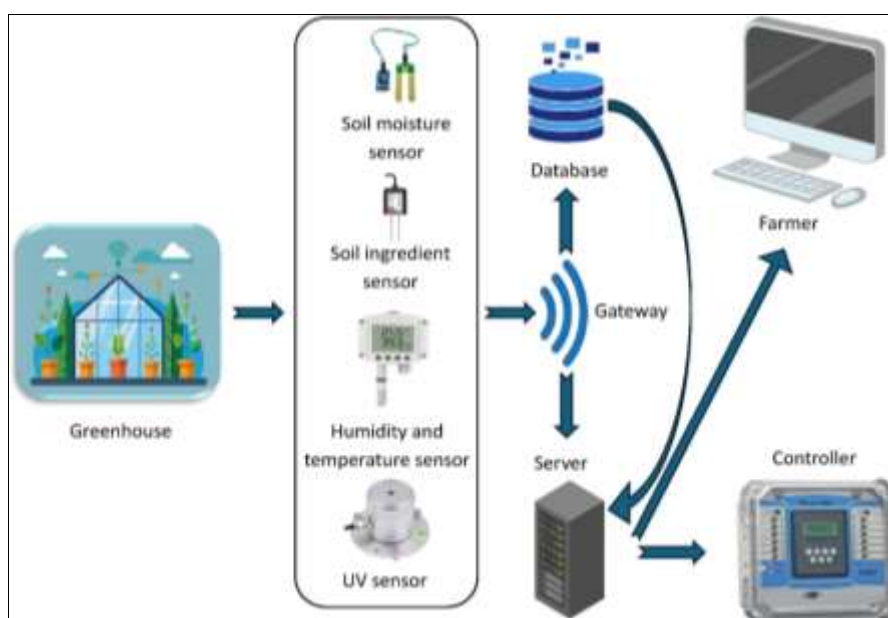


Fig 1: A representation of internet of things applications for monitoring agricultural conditions within a greenhouse

2. Sensor Technologies in Precision Agriculture: A Multifaceted Approach

Technological improvements in sensors are starting to greatly improve the capacity to measure crop and environmental parameters at the micro level and in real time. This participatory process brings together various sensors which help sense the gradual changes in the plant's physiology, efficient granting of resources and improvement of the yield estimate. I therefore infer that hyperspectral imaging, multispectral and thermal imaging and LiDAR-based 3D sensing sensors are the key to this revolution. These sensor technologies hyperspectral imaging, multispectral and thermal imaging, LiDAR based 3D sensing each have strengths for identifying plant stress and growth in order to make valid from field studies as supported by recent studies, multispectral and thermal imaging and LiDAR based 3D sensing are all have powerful and attractive methods

to monitor plant health and stress in precision agriculture (Omia *et al.*, 2023) ^[24]. The application of the advanced practices allows detecting the stress signs at an early stage, determine the yield with high precision, and use the resources more effectively. As the technology in sensors unfolds alongside with the data analytics as well as machine learning, the employment of the new technology in the agricultural sector will enable the agricultural industry to embrace efficient, sustainable, and precise or enhanced crop management thus feeding the world's population despite the emergence of environmental factors such as climate change.

2.1 Hyperspectral Imaging: Unveiling Spectral Signatures of Stress

Hyperspectral imaging acquires many bands of narrow bandwidth and provides total reflectivity of tissues of plants.

This technique helps the researcher to have a direct profile of disease spectral signature as well as pest and other abiotic stresses (Dhaliwal *et al.*, 2024) ^[6]. This can be done since hyperspectral imaging is also capable of detecting somewhat biochemical changes that are present before the manifestation of typical symptoms.

2.1.1 Spectral Fingerprinting

Others are hyperspectral sensors which are able to penetrate more chlorophyll content, carotenoids, and water contents (Wieme *et al.*, 2022) ^[41]. For instance, depending of the shift in the position of the red edge, which is the steep transition from red to near infra red reflectance, it may point to pathogen attack or lack of nutrients.

2.1.2 Data Integration with Machine Learning

Hyperspectral images are employed for the definite classification of diseases through the application of more refined techniques such as CNN. A group for instance successfully utilized hyperspectral aerial cameras mounted on UAVs that provided over 90% success rate in detecting the early signs of fungal attacks on a soybean field. This helps in early determination of the right preventive measures that accrue to lowered overall use of chemicals and losses of crops (Noshiri *et al.*, 2023) ^[23].

2.1.3 Real-Time Monitoring

Conventional hyperspectral imaging is steady and offers an assessment of crop vigour, facilitating decision-making on irrigation, fertilization and pest management in crops. It is most advantageous in large scale plants for scouting by vehicles as manually this technique cannot be employed.

2.2 Multispectral and Thermal Imaging

Assessing Plant Health and Stress Indicators

Multispectral imaging uses a limited number of narrow bands of wavelength under TCB which include blue, green, red and near infrared to estimate some vegetation indices for instance the NDVI. All these indices are useful in measuring vigor, biomass, and chlorophyll content of plants. Whereas, thermal imaging detects the temperature of the canopy which reflects the status of water and stress of the plant (Issa *et al.*, 2020) ^[12].

Vegetation index and other coefficients based on multispectral information are calculated for NDVI and similar values now associated with photochemical processes and the health condition of plants. It helps in early indication of stress arising from lack of nutrients, diseases pests or lack of water among the plants. For instance, a study done on maize field, depicted that the reduced values of NDVI can be used to estimate that the occurrence of drought stress is about to begin before the wilting signs appears. Heat sensing cameras aims to record thermal gradients on plant top surfaces. Higher temperatures in certain regions are the signs of lowered transpiration and water deficit (Smigaj *et al.*, 2024) ^[37]. For instance, thermal imaging has been applied in vineyards where it was used to define areas of water shortage so that irrigation can be affected in the right areas. In research conducted on vineyards in California, it was established that thermal mapping assisted in the management of water hence leading to the reduction in charge of irrigation and conservation of water. On the one hand, multispectral imaging gives the measurements of the biochemical and structural state of the crop, while on the other hand, thermal imaging gives the actual physiological picture of the crop at the same time. It improves the general crop practice since farmers can make decisions on its

irrigation, nutrient as well as pest control.

2.3 LiDAR and 3D Sensing

Analyzing Canopy Structure and Growth Patterns

Detailed characterization of plant canopies that is done by LiDAR (Light Detection and Ranging) which makes use of lasers to scan and create 3D images of the plants. It offers detailed information of the upper strata especially on height, volume, and LAI all of which are important for biomass, growth and yield rates. LiDAR maps out the structure of the underneath topography, hence enables the determination of the tree or crop canopy dimensions. These measurements are useful in determining the production rate of crops as well as the wellbeing of the plants (Paturkar *et al.*, 2021) ^[26]. For instance, LiDAR mapping in apple orchards has helped growers pattern canopy density and relate the density to the production levels of the fruit; this has made growers to give the best treatment in the pruning and irrigation of such trees. When done frequently, the LiDAR scans enable the farmer to monitor the growth of crop canopies from time to time. Sophisticated longitudinal data used in farming aids in identifying a young plant or crop that has a different growth progress from normally growing plants of the same type as this may indicate problems such as the need for nutrients, pest or diseases attack. Temporal monitoring is, therefore, very useful when it comes to management as well as early interferences. When integrated with hyperspectral and thermal data, LiDAR gives an overall idea of the health status of plant (Ecke *et al.*, 2022) ^[7]. Whereas LiDAR gives an amount of information about the structure and the stored mass, hyperspectral and thermal data describe biochemical and physiological features. This system enables precise management decision to be made which in turn increase crop yield and productivity (Figure 2).

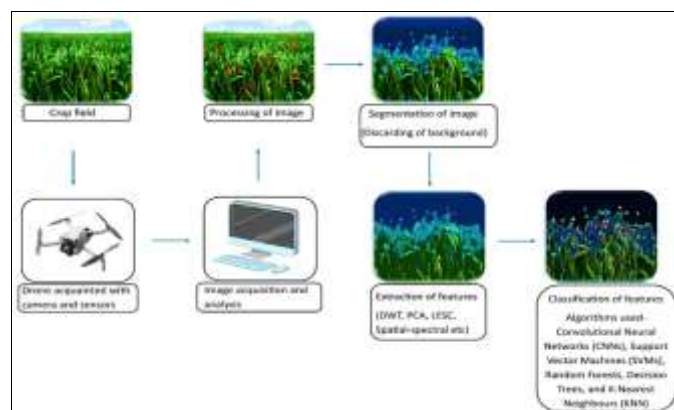


Fig 2: The plant disease detection and classification model include image acquisition, preprocessing, segmentation, feature extraction, and classification. Drones have sensors and cameras to photograph plant diseases. Background noise is removed by processing and segmenting pictures. Traditional machine learning models like K-nearest neighbour (KNN) and support vector machine (SVM) are used to retrieve and identify features, as well as deep learning models like CNN. Principal component analysis (PCA), local embedding based on spatial coherence algorithm (LESC), and discrete wavelet transform (DWT)

3. Statistical Modeling for Disease and Pest Detection

Extracting Meaningful Patterns

Statistical modeling is making a revolution in plant pathology and integrated pest management by finding out significant patterns from large data generated from sophisticated sensors. Since machine learning algorithms, multivariate analysis and temporal analysis can be used to identify and anticipate diseases

and pest attacks researchers can accurately and promptly intervene. Application of statistical modeling in precision agriculture provides a great approach to early detection of diseases and pests through using multi spectral sensors (Shin *et al.*, 2022) [35]. In spectral analysis, machine learning is highly effective at categorizing plant conditions according to their spectra, whereas with multivariate analysis techniques, the new biochemical changes can be found in the plant under stress conditions. These approaches are complemented by time-series analysis, that helps to predict further development of the disease and pests. Altogether, the methods inform farmers on the most effective ways of protecting their crops, minimizing on the use of resources and the appropriate practices that can be adopted to embrace sustainable farming in the face of rising environmental issues (Srivastav *et al.*, 2021) [38].

3.1 Machine Learning Algorithms

Classifying Disease and Pest Infestations

Contemporary sensor devices starting from hyperspectral cameras up to thermal imaging systems create large amounts of data pertaining the physiological conditions of plants. The data collected in such instances is then processed by machine learning algorithms that look for signs thereof, in relation to disease or pests. These methods, such as decision tree, support vector machines (SVM), convolutional neural networks (CNN), are trained in labeled dataset to classify the plant condition with high efficiency (Lu *et al.*, 2021) [19], one case study was conducted in soybean crops where hyperspectral images captured by UAV were analyzed using CNN. The model was able to further distinguish between healthy and fungus-infected plants basing its analysis on minor variations in the spectral reflectance in the red-edge region. The model brought out classification rates that were over 92% which made it possible for early assessment to be done before signs manifest themselves. This work lays the foundation on the ability of machine learning techniques to analyze spectral data, and come up with result that can be useful in the management of diseases (Sahu *et al.*, 2022) [33].

3.2 Multivariate Analysis

Identifying Key Spectral Features

Multivariate analysis methods are useful in analyzing the huge data arising from multispectral and hyperspectral imagery. Techniques like Principal Component Analysis (PCA), Partial Least Squares Regression (PLSR), and Canonical Correlation Analysis (CCA) are used to find out lodges that are stressed and those infested (Ban *et al.*, 2025) [3]. These tools enable the reduction of sensor data dimensionality whereby only the important variance useful in diagnosing plant health problems is retained. Among those, one can highlight the case of PCA used for hyperspectral data to determine which bands reminded more on the early-stage leaf rust infection in the case of wheat (Whetton, 2016) [40]. Thus, it was named that some particular wavelengths at the near-infrared range were highly associate with the disease progression.

3.3 Time-Series Analysis

Tracking Disease Progression and Pest Population Dynamics

Time-series data analysis is very important as crops may go through changes some of which are seasonal and therefore must be tracked periodically. From such information as might be obtained from sensors taken at discreet time intervals, one is able to simulate diseases and variations in pests. Analyzes of future stress events are done with the help of models such as

ARIMA, exponential smoothing, or even more complex machine learning approaches like recurrent neural networks help to evaluate the effectiveness of intervention (Mojtahedi *et al.*, 2025) [21]. One striking example was the assessment of thermal profiles of vineyards for determining the indications of water stress or pest/ disease attack by monitoring the changes in canopy temperature. The data trend analysis suggested the outlines of temperature changes that occurred before the manifestation of visible symptoms of the disease. The researchers used forecasting with the use of ARIMA for temperature changes since this is expected to increase with the infestation of pests. This was the situation because the functions provided the vineyard managers with a predictive capability that helped determine the conditions for irrigation as well as applying pest control measures that reduced yield loss (Wu *et al.*, 2023) [42]. The use of time-series analysis in the management of crop in as therefore useful in giving an early signal to producers so that management intervention can be made to ensure survival of the crop (Figure 3).

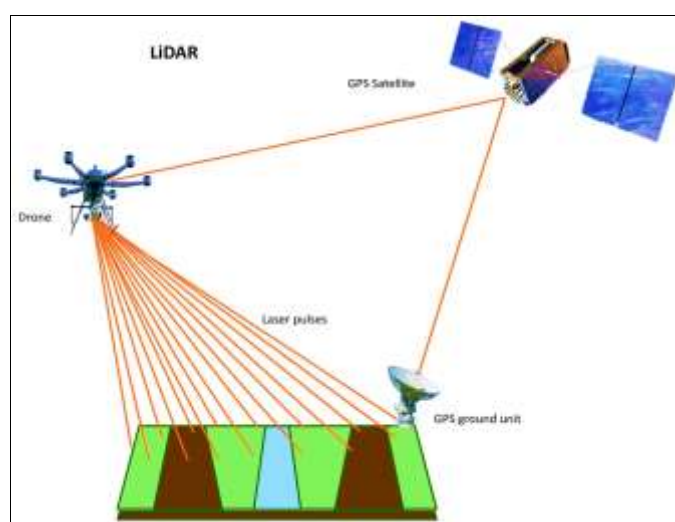


Fig 3: Light detection and ranging (LiDAR) survey data acquisition

4. Data Preprocessing and Feature Engineering: Preparing Data for Robust Analysis

Data analysis, especially a robust one, does not begin when the algorithm is applied; it begins with cleaning and processing of data. In most contemporary environments, real-life data is available preprocessed to contain much noise and variability and to include redundant data that may interfere with the analysis process. These two steps play a very important role in the determination of reliability and integrity of the data, increase in speed and accuracy of the model and attainable beneficial results from the output of the sensors (Alahmad *et al.*, 2023) [1]. Vigilant data preprocessing and feature engineering are critical for the extraction of valuable insights of the data collected from the sensors. Thus, by analysing the presented data reducing noise, using expert's knowledge in data selection, and focusing on data quality, analysts can create accurate models which will predict the proper outcomes to be successfully used in the decision-making process. Incorporation of these practices goes a long way in improving the preciseness and speed of models and, at the same time, paves way for proper and more fulfilling utilization in different fields in the future (Júnior *et al.*, 2024) [14]. These are the principles on which the next generation of innovative and efficient data-centric strategies will stand on, as technology develops further.

4.1 Addressing Noise and Inconsistencies in Sensor Data

Data from sensorial data, Internet of things devices, environment monitoring or from industry generally contains various types of noise such as random errors, outliers and missing values. It is therefore important to clean the data to eliminate or deal with these irregularities. Processes like filtering (moving average, Kalman filters) and others such as detecting and eliminating outliers as well as imputing missing observations (by mean, median or model) are used for eliminating noise and missing values (Girimurugan *et al.*, 2023) ^[10]. In this way, the collected data is made more credible which is useful for further analysis. Thus, filtering has an advantage of not only increasing the accuracy of the predictions made, but also the interpretability of the data. In the process of data acquisition, it is also important to achieve similarity across different channels or sources of data. Since the sensor data can be acquired at different frequencies, scales and in different formats, this needs to be preprocessed in a way that can be standardized. It may include rescaling the data to have the same range, standardizing the time stamp and converting units. Standardization makes it easier for the features to be made comparable hence proper model training together with eradicating differences that may be brought about by data format differences. Suppose there is a system of environmental monitors continuously monitoring the rate of temperature, humidity level, and particle density in an urban environment. Unprocessed series sampled from the sensors can be irregular due to noise and different calibration at certain time instances. Through the combination of such kinds of filters and normalization procedures, the data is cleaned enhancing the accuracy of analyzing trends in the concentration of determinants in the air quality in urban areas (Chen *et al.*, 2022) ^[5]. The cleaned and standardized data is therefore a sound corpus that forms the basis on which to base the models that dictate the public health and planned urbanizations.

4.2 Optimizing Feature Selection for Improved Model Accuracy and Efficiency

Feature selection may be defined as one of the subsequent important steps in data preprocessing whereby the most suitable features for training a model or a machine learning algorithm to be used are selected to be used for the task. Big data can also mean that the data set size is high-dimensional which results to overfitting, increase in computational time and reduction in the model's accuracy. Many procedures like Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), and the use of Lasso, Ridge, among other techniques enable feature dimensionality reduction because only the most descriptive and predictive components are retained (Tanimu *et al.*, 2022) ^[39]. These techniques help in setting only the relevant features in the model so that the overall model becomes less complex and more generalized. Even the selection of features should inculcate the domain knowledge as this plays a great role in determining the results from a model. It is possible for the experts to identify those features that are likely to have an effect on the outcomes considering what has been tested by other researchers and the theoretical frameworks governing such systems. It not only improves the step of identifying important features but also helps to build new features that represent combinations of characteristics. For example, in the analysis of sensor data, combining the two features that are temporal and spatial dimensions can help to improve the model's capability in predicting more accurately on a time series (Xu *et al.*, 2021) ^[43]. Let us supposed the source of information that is to be gathered through sensors for estimating water need is crop productivity.

This indicates that rather than feeding FSCS with raw data from moisture, temperature and humidity and sensors, a set of engineered features is obtained by deriving indices for water stress and evapotranspiration rates. Finally, other elaborate dimensionality reduction algorithms identify the most important indices and ultimately construct an effective and precise model (Singh *et al.*, 2022) ^[36]. Not only does this type of feature selection lessen training time but also boosts up the performance of the specified models by giving stress to essential features.

4.3 The Importance of Data Quality for Reliable and Meaningful Results

Quality data is the lifeblood of any analytical process focus on data quality must be a high priority in an analytical project. It is achieved when the data is accurate, whole some, consistent and updated or timely. A high level of validation and cross check should be used so as ensure that the data emerging in the process is accurate measure rather than systematic or random errors (Yates *et al.*, 2023) ^[45]. Some common approaches include using statistical activities on the data, auditing the data as well as performing quality control checks with a view of rectifying errors before they go a round in the analysis process. High quality data is a foundation of business insight that drives success of the predictive model outputs since they must be credible to be usable. This is a potential disadvantage for any business firm since it makes it hard to come up with accurate models, arrives at the wrong decisions and ends up in operation costs. On the other hand, accurate clean data which has been properly arranged would result into models that give precise estimates and recommendations. This is because when data quality is given special importance then the results of a model and the information that is extracted from the model are easily understandable by the decision makers and also it creates confidence among the stakeholders who depend on the model for making their decisions (Gil *et al.*, 2021) ^[9]. This is especially true for sectors where the decisions made have consequences over resource usage and food production as is the case with the agriculture sector. Suppose that smart city required the installation of a large number of sensors that will be interconnected to track energy usage and environmental factors. When data is preprocessed and contains limited noise data and variations, it enables city planners to design precise models of the future state perfectly. These models be used to predict energy requirements and effects it will have on the environment hence allowing for timely optimization of the resources (Rahman *et al.*, 2021) ^[30]. The general increase in quality of data collected promotes organisational efficiency and satisfy the public thus improves the overall decision-making process.

5. Model Validation and Performance Assessment: Ensuring Reliability and Accuracy

In the field of machine learning, accuracy and reliability of the models hold the prime place. It is also important that these models behave well when operated on examples different from those used in the training process. This opens the calls for such validation and performance assessment procedures to be very stringent.

5.1 Methods for Evaluating Model Performance and Generalizability

5.1.1 Training-Validation-Test Split

The first and simplest split of the data set is accomplished through division of the set into three different types of set: the training set, the validation set, and the test set. The training set is

used to build the model the validation set helps in tuning the hyperparameters and to also avoid overfitting of the model while the test set gives the accurate assessment of the final model (Pfob *et al.*, 2022) ^[28]. This is done to ascertain that the model is able to make a proper prediction on any new data and that the high level of performance on the training data that does not translate to poor performance on the testing data known as overfitting.

5.1.2 Resampling Methods

Other forms of developing train and test sets include bootstrapping and permutation tests as the other available ways for performance assessment. Cross Validation is attained by taking samples from the dataset and using them in iterations to form training and testing sets in which variability of the model can be estimated. Using the permutation tests one can evaluate whether the performance of the model is significantly better than expected by chance: the distribution of performances is generated by permuting data. Assume that a financial institution wants to create a model that provides evaluation of the ability of loan default. This training-validation-test split allows the institution to train the model on the historical data, improve the hyperparameters using the validation set to make accurate predictions for refined accuracy, and test the generalizability of the model with the help of the test set (Kaur *et al.*, 2020) ^[15]. This enhances the validity of the credit scoring model for the fact that it accurately estimates new loan applicants' fail to pay thus helpful in on credit decisions.

5.2 Cross-Validation Techniques and Independent Testing Datasets

5.2.1 K-Fold Cross-Validation

K-Fold Cross Validation is known as a reliable method of layering the available data set into 'K' groups or folds. The entire model is trained for 'k-1' fold and the next fold is used for testing the model. The process is repeated 'k' times with each fold forming the testing set in certain iteration. They are then averaged out to give an overall performance indicator of the assessed machine (Zhang and Liu, 2023) ^[47]. It also aims to make full use of all the data such that issues like over fitting are not experienced with the model.

5.2.2 Leave-One-Out Cross-Validation (LOOCV)

The other type of cross-validation is called LOOCV or leave one out cross validation where each point of data constitutes a test set and the rest of the data as the training set. This is good for a few reasons especially with limited data because it helps to make excellent use of data that one gets for every iteration (Pang *et al.*, 2023) ^[25]. However, the method may consume a lot of time when dealing with large datasets and more iterations are required.

5.2.3 Independent Testing Datasets

Independent testing sets are normally used to test the model on a set of data not used in the development of the model or its validation. This comes in handy in measuring the model's performance in practice situations rather than in academic scenarios that specifically fit the training set data. It is even more vital in areas where the model is going to make decisions that will affect people's lives in one way or the other, such as in diagnosis of diseases, self-driving vehicles, etc. Suppose healthcare analytics is establishing a model to analyse readmission rates of patient. Thus, when the company uses k-Fold Cross-Validation, it assesses the performance of the model

on more than one subset the data of the patient, therefore increases, its reliability (Allgaier and Pryss, 2024) ^[2]. Then, to evaluate the performance of the proposed model when applied to a different hospital system different from the one used to develop the model, the latter is applied on an independent dataset. This broad validation approach allows the model to be accurate in making Centre for epidemiological skills predictions than making extrapolations when identifying patient populations.

5.3 Metrics for Assessing Model Accuracy, Precision, and Recall

5.3.1 Accuracy

Accuracy is one of the simplest evaluation metrics it measures how any of the tests issued by the model was correct out of all tests it made. Though quite sensible, accuracy gives a biased representation of the dataset especially when the test data set has more of one class over the other. For example, in the set with 95% of negative cases and 5% of positive cases, the model that predicts all the cases as negative gains 95% of accuracy but misses all positive cases (Rainio *et al.*, 2024) ^[31].

5.3.2 Precision and Recall

Positive Predictive Value or Precision in simple terms is the ability of the model to make least wrong predictions among all the positive ones. It illustrates ability of the model not to confuse with other types of signals other than the targeted one, that is, low false positive rate. Recall, also known as Sensitivity, shows the extent to which the deployed model is successful in detecting all the actual positive findings, without been diluted by actual negative findings. Another benefit of the F1 Score is in balancing between precision and recall because a high value of F1 can be obtained when both are balanced; especially in a situation where one class has more samples than the other (Gupta *et al.*, 2023) ^[11].

5.3.3 Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC)

The ROC curve is commonly used and is a graphical representation of the true positive rate against the false positive rate as they vary depending on the threshold levels. This is because the AUC sums the probability of correct classification of all the samples in the ROC space and is a good measure of the model's power to differentiate between classes. The closer AUC value is to 1, it shows that the particular model performs well in classification task and on the other hand, if AUC value is near to 0.5 then it means that model has no discriminative power. In the case of a cybersecurity firm that has come up with, for instance, a model that identifies cases of fraudulent transactions (Movahedi *et al.*, 2023) ^[22]. Due to the scale of fraud transactions against legitimate ones the firm is keen in the level of accuracy that is provided to avoid false honeypots which may inconvenience the clients.

6. Challenges and Future Directions: Addressing Limitations and Expanding Capabilities

6.1 Addressing the Complexities of Diverse Environmental Conditions and Crop Types

Agricultural systems are environmental systems found in a very broad class of other environments these include, climatic conditions, types of soil and geographical features. Also, the availability of a large number of crop species all of which have different growth and required resources characteristics also makes the management of data and models more complicated

(Javaid *et al.*, 2023) ^[13]. Since they were trained on specific regions, the models do not transfer as they are due to regional climatic and soils. This variability, then, has to be addressed in models that are to generalize to different environments. Due to the physiological differences in crops, the models developed for one particular crop cannot be used for the other and thus crop specific models or highly mobile models need to be developed (Yang *et al.*, 2020) ^[44]. To achieve these objectives, environmental data that include environmental variability called enviromics data are integrated into models' testing. In this way, the patterns are more suitable for the environment since the influence is controlled by environmental factors. This makes a way to train models for one crop or region and apply them to another which takes much less time compared to training a model from scratch.

6.2 Developing Robust and User-Friendly Analytical Tools for Widespread Adoption

The usefulness of the analytical tools is determined by the degree to which they can be applied by the end-users, which in this case includes farmers and other agricultural personnel. Most of the contemporary models are computationally intensive and present certain technicalities that make it difficult for end-users to embrace them. However, lack of adequate access and sometimes nonexistent cellular data connectivity in developing countries, shortage of adequate hardware and software may greatly slow down the utilization of analytical tools (Bayih *et al.*, 2022) ^[4]. While designing interfaces as well as converting complicated figures into understandable formats, specialized persons can make them intuitive to other persons too. Mobile technology and cloud computing, the two technologies can help in making analytic available to anybody who can avail a normal internet connection without requiring much of computational resources in his/her local machine (Gaire *et al.*, 2020) ^[8].

6.3 The Need for Large, High-Quality Datasets for Model Training and Validation

Analytical models can be only as good as the data that is used in the creation of the model and the more precise and up to date the information is the better the functional model will be. It is therefore becoming more challenging to find clear-cut and exhaustive datasets of different crops and weather conditions especially in developing world (Pongratz *et al.*, 2021) ^[29]. Randomness and errors in the data similar to how operations are done will result in low quality data that will give a low quality model output. It is also important that more research organizations and governments as well as semi-public and private institutions support the accumulation of various and huge databases. It is useful to maintain consistency when collecting and validating the data so that the datasets could be effectively used in training the models. Mitigating these challenges by implementing special approaches will therefore go along way in enhancing the advancement and implementation of analytical instrument in support of farming systems (Sharma *et al.*, 2022) ^[34].

7. Conclusion

A Vision for the Future of Crop Protection

The next technical revolutions have improved the way crops diseases and pests are identified and controlled through the use of sensor technologies applied in precision agriculture. The next advancements in hyperspectral, multispectral, the thermal, and LiDAR depiction have made early detection and even monitoring possible to make perfect decisions, decreasing losses

and delivery of chemicals accordingly. These sensor systems have changed the scouting techniques by complementing a new dimension into the plant health by identifying stressed areas before visibly depicting the signs. However, more challenges persist here and there as highlighted below. Limitations to the current advancement include; Inability to gather information from a number of heterogeneous sensor systems, the variability in sequence of information between the various platforms, high cost of the advanced equipment. Furthermore, another aspect also complicates that the conditions the plants live in are variable and the responses of the plants themselves are feeble thus adding more variables that need further studies and more refined models. Closing these gaps is crucial in ensuring that researchers' advancements in technology are effectively taken to the farmers of all scales. It must be mentioned that the sphere of artificial intelligence (AI) and big data analysis is capable of offering numerous chances for transformation. That is why, through the usage of machine learning models and statistical patterns, scholars are in a position to analyze enormous chunk of data and forecast disease incidence and pest infection. Machine learning technologies can be applied in addressing complex land use decisions by receiving data from satellite imagery, field sensors or any data and provide a risk analysis and solution of the problem encountered. This is not only effective in terms of resource management but also contributes positively to sustainable agricultural practices as the application of chemicals such as pesticides and other materials with adverse impacts to the ecosystem would be minimized as much as possible. The future trend of crop protection is already moving into the phase of prevention instead of the current treatment or after effect method. As far as sensors and AI integration becomes smarter, the ultimate goal is to achieve a network of interconnected systems that include sensors, AI networks, IoT connections that would help in detecting stress signals to launch immediate and local countermeasures. They would allow one to go from fixed time-based treatments to precise interventions on crop health while maintaining biological balance. All in all, it is possible to note that the usage of sensors for disease and pest detection has advanced substantially, but the existing problems have to be addressed to provide for even greater possibilities of sensor-based technologies. The active AI and big data promotion in the seed industry will allow to achieve the proactive and environmentally friendly crop protection strategies as a result trends for the contemporary progressing world to guarantee the sustainability of the food production accordingly.

References

1. Alahmad T, Neményi M, Nyéki A. Applying IoT sensors and big data to improve precision crop production: a review. *Agronomy*. 2023;13(10):2603. doi:10.3390/agronomy13102603.
2. Allgaier J, Pryss R. Cross-validation visualized: a narrative guide to advanced methods. *Machine Learning and Knowledge Extraction*. 2024;6(2):1378-1388. doi:10.3390/make6020065.
3. Ban S, Tian M, Hu D, Xu M, Yuan T, Zheng X, *et al.* Evaluation and early detection of downy mildew of lettuce using hyperspectral imagery. *Agriculture*. 2025;15(5):444. doi:10.3390/agriculture15050444.
4. Bayih AZ, Morales J, Assabie Y, De By RA. Utilization of internet of things and wireless sensor networks for sustainable smallholder agriculture. *Sensors*. 2022;22(9):3273. doi:10.3390/s22093273.
5. Chen L, Wang J, Wang H, Jin T. Urban air quality

- assessment by fusing spatial and temporal data from multiple study sources using refined estimation methods. *ISPRS International Journal of Geo-Information*. 2022;11(6):330. doi:10.3390/ijgi11060330.
6. Dhaliwal SS, Sharma V, Shivay YS, Gupta RK, Verma V, Kaur M, *et al.* Assessment and detection of biotic and abiotic stresses in field crops through remote and proximal sensing techniques—evidence from earlier findings. *Arabian Journal of Geosciences*. 2024;17(6):188. doi:10.1007/s12517-024-11993-6.
 7. Ecke S, Dempewolf J, Frey J, Schwaller A, Endres EJ, Klemmt HJ, *et al.* UAV-based forest health monitoring: A systematic review. *Remote Sensing*. 2022;14(13):3205. doi:10.3390/rs14133205.
 8. Gaire R, Sriharsha C, Puthal D, Wijaya H, Kim J, Keshari P, *et al.* Internet of Things (IoT) and cloud computing enabled disaster management. In: *Handbook of Integration of Cloud Computing, Cyber Physical Systems and Internet of Things*. Cham: Springer; 2020. p. 273-298. doi:10.1007/978-3-030-43795-4_12.
 9. Gil Y, Garijo D, Khider D, Knoblock CA, Ratnakar V, Osorio M, *et al.* Artificial intelligence for modeling complex systems: taming the complexity of expert models to improve decision making. *ACM Transactions on Interactive Intelligent Systems*. 2021;11(2):1-49. doi:10.1145/3453172.
 10. Girimurugan R, Selvaraju P, Jeevanandam P, Vadivukarassi M, Subhashini S, Selvam N, *et al.* Application of deep learning to the prediction of solar irradiance through missing data. *International Journal of Photoenergy*. 2023;2023(1):4717110. doi:10.1155/2023/4717110.
 11. Gupta P, Varshney A, Khan MR, Ahmed R, Shuaib M, Alam S. Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques. *Procedia Computer Science*. 2023;218:2575-2584. doi:10.1016/j.procs.2023.01.231.
 12. Issa S, Dahy B, Ksiksi T, Saleous N. A review of terrestrial carbon assessment methods using geo-spatial technologies with emphasis on arid lands. *Remote Sensing*. 2020;12(12):2008. doi:10.3390/rs12122008.
 13. Javaid M, Haleem A, Khan IH, Suman R. Understanding the potential applications of artificial intelligence in agriculture sector. *Advanced Agrochem*. 2023;2(1):15-30. doi:10.1016/j.aac.2022.10.001.
 14. Júnior MRB, de Almeida Moreira BR, dos Santos Carreira V, de Brito Filho AL, Trentin C, de Souza FLP, *et al.* Precision agriculture in the United States: A comprehensive meta-review inspiring further research, innovation, and adoption. *Computers and Electronics in Agriculture*. 2024;221:108993. doi:10.1016/j.compag.2024.108993.
 15. Kaur S, Aggarwal H, Rani R. Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease. *Machine Vision and Applications*. 2020;31:1-15. doi:10.1007/s00138-020-01078-1.
 16. Khan BA, Nadeem MA, Nawaz H, Amin MM, Abbasi GH, Nadeem M, *et al.* Pesticides: impacts on agriculture productivity, environment, and management strategies. In: *Emerging Contaminants and Plants: Interactions, Adaptations and Remediation Technologies*. Cham: Springer International Publishing; 2023. p. 109-134.
 17. Kumar P, Choudhury D, Dey SR, Raj A. Nanofertilizers/nanopesticides for national economy. In: Kumar P, Dubey RC, editors. *Nanofertilizers for Sustainable Agriculture*. Cham: Springer; 2025. doi:10.1007/978-3-031-78649-5_17.
 18. Kumar P, Raj A, Kumar VA. Approach to reduce agricultural waste via sustainable agricultural practices. In: Srivastav AL, Bhardwaj AK, Kumar M, editors. *Valorization of Biomass Wastes for Environmental Sustainability*. Cham: Springer; 2024. doi:10.1007/978-3-031-52485-1_2.
 19. Lu J, Tan L, Jiang H. Review on convolutional neural network (CNN) applied to plant leaf disease classification. *Agriculture*. 2021;11(8):707. doi:10.3390/agriculture11080707.
 20. Mittal M, Gupta V, Aamash M, Upadhyay T. Machine learning for pest detection and infestation prediction: A comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2024;14(5):e1551. doi:10.1002/widm.1551.
 21. Mojtahedi FF, Yousefpour N, Chow SH, Cassidy M. Deep learning for time series forecasting: Review and applications in geotechnics and geosciences. *Archives of Computational Methods in Engineering*. 2025:1-31. doi:10.1007/s11831-025-10244-5.
 22. Movahedi F, Padman R, Antaki JF. Limitations of receiver operating characteristic curve on imbalanced data: assist device mortality risk scores. *The Journal of Thoracic and Cardiovascular Surgery*. 2023;165(4):1433-1442. doi:10.1016/j.jtcvs.2021.07.041.
 23. Noshiri N, Beck MA, Bidinosti CP, Henry CJ. A comprehensive review of 3D convolutional neural network-based classification techniques of diseased and defective crops using non-UAV-based hyperspectral images. *Smart Agricultural Technology*. 2023;5:100316. doi:10.1016/j.atech.2023.100316.
 24. Omia E, Bae H, Park E, Kim MS, Baek I, Kabenge I, *et al.* Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sensing*. 2023;15(2):354. doi:10.3390/rs15020354.
 25. Pang Y, Wang Y, Lai X, Zhang S, Liang P, Song X. Enhanced Kriging leave-one-out cross-validation in improving model estimation and optimization. *Computer Methods in Applied Mechanics and Engineering*. 2023;414:116194. doi:10.1016/j.cma.2023.116194.
 26. Paturkar A, Sen Gupta G, Bailey D. Making use of 3D models for plant physiognomic analysis: A review. *Remote Sensing*. 2021;13(11):2232. doi:10.3390/rs13112232.
 27. Paul K, Chatterjee SS, Pai P, Varshney A, Juikar S, Prasad V, *et al.* Viable smart sensors and their application in data driven agriculture. *Computers and Electronics in Agriculture*. 2022;198:107096. doi:10.1016/j.compag.2022.107096.
 28. Pfob A, Lu SC, Sidey-Gibbons C. Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. *BMC Medical Research Methodology*. 2022;22(1):282. doi:10.1186/s12874-022-01758-8.
 29. Pongratz J, Schwingshackl C, Bultan S, Obermeier W, Havermann F, Guo S. Land use effects on climate: current state, recent progress, and emerging topics. *Current Climate Change Reports*. 2021:1-22. doi:10.1007/s40641-021-00178-y.
 30. Rahman MM, Shakeri M, Tiong SK, Khatun F, Amin N, Pasupuleti J, *et al.* Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. *Sustainability*. 2021;13(4):2393.

- doi:10.3390/su13042393.
31. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*. 2024;14(1):6086. doi:10.1038/s41598-024-56706-x.
 32. Raj A, Kumar A, Vaghasiya TP. Multidisciplinary approaches to clean up heavy metal contaminated soil: a review. *Bulletin of Environmental Pharmacology and Life Sciences*. 2021;10(6):53-65.
 33. Sahu M, Gupta R, Ambasta RK, Kumar P. Artificial intelligence and machine learning in precision medicine: A paradigm shift in big data analysis. *Progress in Molecular Biology and Translational Science*. 2022;190(1):57-100. doi:10.1016/bs.pmbts.2022.03.002.
 34. Sharma V, Tripathi AK, Mittal H. Technological revolutions in smart farming: Current trends, challenges & future directions. *Computers and Electronics in Agriculture*. 2022;201:107217. doi:10.1016/j.compag.2022.107217.
 35. Shin J, Mahmud MS, Rehman TU, Ravichandran P, Heung B, Chang YK. Trends and prospect of machine vision technology for stresses and diseases detection in precision agriculture. *AgriEngineering*. 2022;5(1):20-39. doi:10.3390/agriengineering5010003.
 36. Singh KN, Devi SD, Devi HM, Mahanta AK. A novel approach for dimension reduction using word embedding: An enhanced text classification approach. *International Journal of Information Management Data Insights*. 2022;2(1):100061. doi:10.1016/j.jjime.2022.100061.
 37. Smigaj M, Agarwal A, Bartholomeus H, Decuyper M, Elsherif A, de Jonge A, *et al.* Thermal infrared remote sensing of stress responses in forest environments: a review of developments, challenges, and opportunities. *Current Forestry Reports*. 2024;10(1):56-76. doi:10.1007/s40725-023-00207-z.
 38. Srivastav AL, Dhyani R, Ranjan M, Madhav S, Sillanpää M. Climate-resilient strategies for sustainable management of water resources and agriculture. *Environmental Science and Pollution Research*. 2021;28(31):41576-41595. doi:10.1007/s11356-021-14332-4.
 39. Tanimu JJ, Hamada M, Hassan M, Kakudi H, Abiodun JO. A machine learning method for classification of cervical cancer. *Electronics*. 2022;11(3):463. doi:10.3390/electronics11030463.
 40. Whetton RL. Multi-sensor and data fusion approach for determining yield limiting factors and for in-situ measurement of yellow rust and fusarium head blight in cereals. Cranfield: Cranfield University; 2016.
 41. Wieme J, Mollazade K, Malounas I, Zude-Sasse M, Zhao M, Gowen A, *et al.* Application of hyperspectral imaging systems and artificial intelligence for quality assessment of fruit, vegetables and mushrooms: A review. *Biosystems Engineering*. 2022;222:156-176. doi:10.1016/j.biosystemseng.2022.07.013.
 42. Wu B, Zhang M, Zeng H, Tian F, Potgieter AB, Qin X, *et al.* Challenges and opportunities in remote sensing-based crop monitoring: A review. *National Science Review*. 2023;10(4):nwac290. doi:10.1093/nsr/nwac290.
 43. Xu L, Chen N, Chen Z, Zhang C, Yu H. Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. *Earth-Science Reviews*. 2021;222:103828. doi:10.1016/j.earscirev.2021.103828.
 44. Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, *et al.* Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular Plant*. 2020;13(2):187-214. doi:10.1016/j.molp.2020.01.008.
 45. Yates LA, Aandahl Z, Richards SA, Brook BW. Cross validation for model selection: a review with examples from ecology. *Ecological Monographs*. 2023;93(1):e1557. doi:10.1002/ecm.1557.
 46. Zhang K, Yan F, Liu P. The application of hyperspectral imaging for wheat biotic and abiotic stress analysis: A review. *Computers and Electronics in Agriculture*. 2024;221:109008. doi:10.1016/j.compag.2024.109008.
 47. Zhang X, Liu CA. Model averaging prediction by K-fold cross-validation. *Journal of Econometrics*. 2023;235(1):280-301. doi:10.1016/j.jeconom.2022.04.007.