



RESEARCH ARTICLE

AGRIGURU: A smart artificial intelligence solution for crop recommendation and plant disease detection

Nandhini Vineeth*, Rajeshwari Madli, Sheetal V A, Bharath M, Subramanya L, Anitha U & Poorvika S K

Department of Computer Science and Engineering, B. M. S. College of Engineering, Bengaluru 560 019, Karnataka, India

*Correspondence email - nandhiniv.cse@bmsce.ac.in

Received: 16 January 2026; Accepted: 06 June 2026; Available online: Version 1.0: 07 August 2026

Cite this article: Nandhini V, Rajeshwari M, Sheetal VA, Bharath M, Subramanya L, Anitha U, Poorvika SK. AGRIGURU: A smart artificial intelligence solution for crop recommendation and plant disease detection. Plant Science Today. 2026;13(sp5):01–12. <https://doi.org/10.14719/pst.13701>

Abstract

Agriculture plays a crucial role in influencing the economy of any country, as it is the major source of food, raw materials and employment to the majority of the population. In India, agriculture contributes a major portion to the country's gross domestic product (GDP) and hence it is crucial to enhance the agricultural yield using the latest tools and technologies. Agricultural yield depends on factors such as plant diseases, fertilisers used, improper crop choices and inaccurate yield prediction techniques. Information Technology (IT) is growing rapidly and is utilised in every walk of human life. Recent advances in artificial intelligence (AI) and machine learning (ML) have enabled data-driven solutions to address these challenges in agriculture. Hence, this work targets to design and develop a system that identifies the infected plants, detects the type of disease and performs crop recommendation. The crop recommendation system uses the support vector classifier (SVC) as the core algorithm. The convolutional neural networks (CNNs) that are trained and tested with the Plant Village dataset, have given higher performance in identifying the disease types of plants, compared to K-nearest neighbours (KNN), Logistic Regression and Decision Trees. This recommendation system determines the nitrogen, phosphorus and potassium (NPK) contents, the pH value and the moisture percentage levels of the given soil and recommends appropriate crops for farmers. Soils from different regions have been considered for experimentation. The decision tree algorithm has been utilised for crop yield prediction by processing climatic and historical agricultural data. Remarkable precision has been achieved with the operations. The proposed system resolves the challenges by utilising an integrated software platform that delivers timely and relevant information to farmers, thereby improving agricultural productivity and performance. The integrated software platform provides farmers with timely and actionable insights, supporting informed decision-making and improved agricultural productivity.

Keywords: convolutional neural network; crop recommendation; decision tree algorithm; disease detection; machine learning; yield prediction

Introduction

As agriculture acts as a vital economic structure of India, the country stands second highest in the production. India experiences varying climatic conditions, possesses diverse soil types and as a result, farmers need to handle complex agricultural ecosystems. Current prevalent methods for disease detection, which involve visual examination and laboratory testing, have been slow-paced, prone to misinterpretation and labour-intensive. This has resulted in incorrect pesticide choices, the trigger to pathogen resilience and damage to crops (1). The manual assessment done using mobile images has been proven unreliable. Convolution neural networks (CNN), when used, help in quicker processing, as well as provide accurate results, in turn helping the farmers to diagnose the disease before there is no damage. The productivity of crops depends clearly on soil health. When the soil is tested appropriately, crops that match the soil conditions can be chosen, resulting in nutrient-rich crops. The significant soil parameters include nitrogen, phosphorus and potassium (NPK) values, the pH value and moisture measurements. The machine learning (ML) model, support vector

classifier (SVC), gives appropriate recommendations when used on such data. This method helps in better planning and increasing yields. Decision tree-based algorithms analyse crop, soil and environmental data to predict yields. The results from this algorithm ensure accuracy and support farmers in making decisions by giving better estimates.

If an integrated system is developed with the features of crop recommendation, yield prediction and disease detection, it can aid farmers in enhancing sustainability and their yield can increase without hindrances.

Some of the existing systems address crop recommendation and some others work on disease detection. An integrated system addressing all these with the usage of real-time internet of things (IoT), ML models, that is easily accessible to farmers, is lacking. Some existing systems are limited by crop variety, scalability and ease of use, where there is a gap observed in providing a user-friendly platform that could support holistic decision-making. This work, titled AgriGuru, aims to bridge this gap with an all-in-one solution personalised to farmers' needs.

The previous decade has seen substantial advancements in ML and deep learning (DL) approaches in forecasting crop yields and detecting plant diseases. The identification of patterns within the available agricultural data by these AI and DL models has brought in improvements in farm productivity and management of crops with sustainability. Support vector machines (SVM), K-nearest neighbours (KNN) and artificial neural networks (ANN) have been identified as leading ML models for the agricultural challenges discussed, based on the performance of these models in the literature survey. Previous research papers demonstrate how SVM, neural networks, decision tree, Naïve Bayes and random forest models successfully identify crop pests, while random forest delivers the best predictive results among these approaches. The field of ML for agricultural disease prediction shows continuing growth potential because it provides important solutions for the industry (2, 3).

Plant health anomalies are identified through the visual symptoms with the help of images of the various parts of a plant, like leaves, stems and root systems. These can be classified on the specific types of disease through the mathematical approaches with pattern matching and statistical functionalities.

An end-to-end ML and DL technique was proposed for plant disease detection and crop selection. The study demonstrates how SVM, random forest and Naïve Bayes computation techniques perform when predicting crop yields using structured data. However, it faces issues when processing unstructured data. Disease detection and soil-based crop recommendation systems benefit significantly from applications of DL methods, where CNNs and RNNs stand out as the preferred analytical approaches. Hybrid systems composed of ML and DL models result in enhanced system performance. The goal of this study involves speeding up localised sustainable agricultural modelling systems, which enhance both forecasting precision and agricultural protocol effectiveness (4).

The work, proposed by early researchers, makes use of methods to predict crop yields, by integrating farm data and the data from experts. Although significant environmental factors such as rainfall, temperature and soil conditions have been used with neural networks, the accuracy of prediction has been limited and the reasons spotted are the temporal gaps in observing the data. Hence, the authors have suggested new data collection timelines to increase the model performance (5).

The authors employ ML algorithms on satellite imagery data to study agricultural output patterns, while making predictions using harvest prognosis. This research implements K-means clustering together with Mask R-CNN to track vegetation, but K-means demonstrates superior accuracy rates for change detection. Researchers employ ML models such as random forest and linear regression to make crop yield predictions and random forest demonstrates superior precision levels. The combination of time series models, including autoregressive integrating moving average (ARIMA) and long short-term memory (LSTM), produces improved modelling accuracy. Vegetation mapping through Python demonstrates superior outcomes in crop monitoring when compared to Matlab-based approaches (6).

A Sine Cosine Algorithm Weighted Regularized Extreme Learning Machine (SCA-WRELM) was proposed, which brings together ELM with regularisation methods and weighting functions alongside SCA optimisation for enhanced rice crop yield prediction.

The framework exists to support yield forecasting accuracy in rice-producing South Asian regions because precise agricultural plans are required. This research uses SCA optimisation on Weighted Regularized ELM (WRELM) parameters to achieve improved prediction accuracy, which will deliver a dependable system for managing rice yields in dominant regions (7).

The authors proposed an application of machine learning algorithms for plant disease prediction, which is examined alongside the increasing popularity of AI in agricultural research. This research explores supervised, unsupervised and reinforcement learning model types, which are proven to be suitable for agricultural challenges (8).

Early researchers demonstrated how the research in the field of crop disease detection has progressed from using early models, like the adoption of CNNs, superseded previous agricultural applications of ANN and k-NN, because these earlier models had fitting challenges and processing problems. Early disease detection models achieved strong test results yet performed poorly with big datasets and generalisations. Research demonstrates that CNNs, specifically ResNet, established themselves as powerful solutions because of their distinctive approach to data processing, higher accuracy and better handling of large, noisy data. Models like VGG16 and EfficientNetModels were used for specific crop detection, but their effectiveness was restricted by limited input data and a restricted focus scope. Acceptable crop disease detection results come from the combination of CNNs, most prominently ResNet, for generalised crop disease detection (9).

The authors developed multiple crop disease identification models with upgraded classification technologies to enhance the accuracy results. The models Unified Matrix-Based CNN (UMB CNN) and leaf wetness sensor (LWS) demonstrate effectiveness for wheat and rice crops simultaneously. However, they face limitations in their ability to scale across different agricultural systems. EfficientNet V2 performs disease detection tasks with both better accuracy and reduced computational complexity across various diseases. The Dilated Multiple Scaled AlexNet, Ensemble Nonlinear SVM, augmented CNN and related models require large training datasets yet employ image segmentation techniques to boost their performance capability. The failure to implement feature selection technology prevents real-time deployment (10).

The work presents the RoBERTa-BiLSTM hybrid model that will help in overcoming long dependencies, lexical diversities and dataset imbalances while analysing sentiment. The model makes use of RoBERTa for word embedding generation and BiLSTM for contextual semantics extraction from text. Experimental results show superior performance to the baseline models, which obtained an accuracy of 92.36 % on the IMDB dataset and showed strong results on the Twitter datasets. Future work includes diverse datasets and fine-tuning hyperparameters to obtain enhanced performance (11).

Previous researchers have applied DL techniques for wheat disease verification. This work uses hyper neural network-based learning (HNNL) methodology and CNN for improving disease detection in wheat crops. These have shown better diagnosis of the diseases with these techniques. The HNNL has helped in detection by extracting the vital component features required for the same with the deep learning algorithms that are pretrained models. The fundamental features are extracted during the initial training

processes. Standard CNN systems have proven to be excellent in the classification of disease detection. The study has used 1500 images, which include 11 disease categories and one control sample. Classification accuracy of 96.46 % has been seen when these techniques are applied. The dataset also has repetition of images seen with varying perspectives, whole-angle observations, which have helped improve the reliability of disease detection. Compared to other methods, HNNL has shown significant improvement. However, 1500 images may not be sufficient to represent all variations of wheat diseases (12).

The authors proposed research that connects IoT instrumentation to ML frameworks to develop such systems. The system evaluates soil quality by examining nitrogen (N), phosphorus (P) and potassium (K) content, as well as pH values. It also considered environmental factors such as humidity measurements and temperature records. The processing methods utilised here, enhance decision trees and random forests within the system. Models analyse records in combination with current information to uncover relationships between variables. Live operational measurements emerge from environmental sensors, which feed directly into the system. The system uses current farm measurements to generate instructions tailored to individual conditions. Precision farming decisions improve because of soil analysis optimisation. Machine learning models such as decision trees and random forests may not generalise well to different geographical regions without retraining. Also, large-scale real-time data processing can increase computational complexity (13).

Early researchers have shown that in agriculture, soil nutrient management can be improved, as well as crop recommendations can be done better with the integration of IoT and ML. Traditional methods of suggestions given on manual soil analysis and deciding on crops have been proven inefficient, as they are labour-intensive as well as rely on subjective assessments. The usage of IoT devices that include sensors for moisture, pH and nutrients provides real-time data on soil conditions. The ML algorithms, like random forest, logistic regression and neural networks, process this data for recommending crops. These have enabled timely decisions, enhanced resource utilisation, improved crop yields and promoted sustainable agriculture (14).

The authors have proposed an advanced system to monitor the crops while identifying any health issues in real-time. Crop diseases present an ongoing threat to agricultural production. The worldwide food supply system remains endangered because of unsuccessful crop production. At present, crops fail to perform as expected because diseases affect both production rates and farm revenue. Traditional disease diagnosis through inspections demonstrates both prolonged duration and substantial dependency on human judgment skills that produce errors. Environmental changes detected through IoT technology sensors enable measurement of environmental conditions. The system measures environmental factors through temperature measurements and soil moisture monitoring and humidity checks. The system uses sensors to monitor soil moisture together with temperature measurements to identify diseases at their earliest stages. The analysis conducted by ML algorithms searches through large datasets to detect distinctive patterns and symptoms that human analysts miss. The random forest classifier method achieves higher prediction success rates by finding problems early. Through electronic systems, farmers gain access to essential data points that

allow them to enhance their farming operations. Correct pesticide applications and water management practices under controlled farming settings allow farmers to observe crop diseases in developing stages (15).

Early researchers review the various approaches of ML that suit prediction systems using classification. They suggest exploring new crop diseases as soon as they are seen and sharing the same with farmers or others involved in agricultural work. The next step has been to explore more on symptoms, the type of disease, spreading methods and other risky crops. Further, authors have explored how different ML algorithms can be applied to these. The authors also reported, as an outcome of their review, that a combination of ML algorithms has shown better results in comparison to the ones used alone (16).

Previous researchers analyse soil-water-air inputs to select the best crops, determine soil health status and fertiliser needs. They offer fertilisation advice based on soil testing results. They make use of previous agricultural data and precipitation records for decision-making. The system successfully identifies the crops to be grown for an upcoming season with an accuracy of 85.55 %. The model analyses soil nutrients to benefit sustainable agriculture and produce better harvests and higher yields. This technology prevents crop losses and reduces wastage of valuable resources to improve farming operations, make every day processes run better. Through ML-based models, farmers find the best crops for their needs to support higher income and resource conservation. The paper refers to soil, weather and water as key factors to achieve effective agricultural planning (17).

The effect of soil nutrients like N, P, K, pH, temperature, rainfall, etc., on agricultural production were discussed. The work also recommends the use of modern methods based on information technology in place of traditional farming techniques. They have shown the usage of the KNN algorithm for the prediction of crops and fertilisers, addressing soil and climatic conditions. Accurate results are ensured using data pre-processing, feature extraction and classification steps. This approach has been claimed to empower farmers in enhancing their yield, reducing costs and promoting sustainable agriculture (18).

Early researchers have applied Naive Bayes classifiers for predicting and recommending the crops suitable for the soil, weather conditions and rainfall. Data sets seen in farmers.gov, APMC, IMD, etc., are pre-processed and categorised for training and testing. Rainfall prediction has been done with the SVM model and the feature selection is done with the correlation matrices. A crop prediction accuracy of 97 % has been achieved by the model. The main objective of this work is to aid farmers with a better profit through maximising crop yield by suggesting the most appropriate crops for their soil and climatic conditions prevailing. The authors have listed an Android app and a website for broader accessibility in their future work (19).

The integration of IoT, ML and weather APIs to enhance crop prediction and aid farmers in better crop selection were explored. A prototype has been developed which uses IoT sensors and APIs collect physical (N, P, K, pH) and virtual (temperature, humidity, precipitation, wind) data, communicating via the message queuing telemetry transport (MQTT) protocol. The random forest algorithm, achieving 99 % accuracy, was applied to an 80–20 training-testing dataset. Predictions and

recommendations were generated using cloud-based tools like Jupyter and Anvil. Field tests validated the model's effectiveness for both swampy and solid soils. Future improvements include extending weather forecasts to crop life cycles, adding more crops from local data and allowing user-contributed crop data (20).

The different methodologies used for accelerating crop yield are discussed by the authors. Various AI techniques, especially ML algorithms and big data analytics used for precision agriculture are also discussed. The usage of various crop recommender systems using KNN, ensemble-based models, neural networks, etc. have been demonstrated (21).

The authors have come up with precision farming method using IoT and a new method named Crop Yield Prediction Algorithm (CYPA). To forecast annual crop yields, the integration of data of various kinds, like chemical, climate, meteorological and agricultural yields, with CYPA was used by policymakers. The authors wanted to aid farmers and decision makers with a decision support tool to analyse meteorological data specific to regions, predicting agricultural yields. The optimisation of CYPA was implemented by including active learning implementation, which reduced the requirements of labelled data for training. Fresh samples were used for labelling in this active learning, which in turn supported CYPA with field environments like pest outbreaks or unpredictable weather conditions (22).

The authors discuss their observation that the agricultural productivity and crop quality are significantly reduced by plant diseases. Identification of leaf diseases, when done early and accurately large-scale crop losses can be prevented. Traditional disease detection methods were not always accessible, as they were time-consuming, demanded expert inspection, which was very expensive. Hence, the automatic leaf disease detection using image processing and ML methods is recommended to overcome these challenges. These techniques analyse leaf images and identify the patterns of diseases accurately. The authors aim to extract the hybrid features from leaf images and apply the supervised ML algorithms like SVM, random forest, KNN and ANN to classify the diseases. Datasets or images shot by a camera of both healthy and diseased leaves are considered, which have different lighting conditions and varying backgrounds. The leaf images that are labelled are used by the classifier for training and to predict the classes of diseases of new images (23).

As an outcome of the literature survey the summarised information is that the set objectives of crop recommendation, crop disease detection and yield prediction and in comparison, to the standalone models, integrated frameworks exhibit a superior performance. The work done earlier focus on any the above three in isolation which is a limitation of their practical applicability in the real-world decision making. The decision of the hybridisation done in Agriguru instead of the usage of single ML or DL algorithm, results in robust predictions. The multimodal approach of using plant images, soil parameters and climatic conditions benefits the system as heterogenous data sources are used. The limitations observed are the dataset greatly influences the performance as they are specific to certain regions, soil, type of crops etc. The usage of DL models increases accuracy at the cost of computational complexity making deployment in real-time a challenge. Infrastructure limitations and power consumption may become a challenge in areas in which agriculture is practised. The black-box nature of DL

algorithms i.e., the opaqueness observed in the inner working of the deep neural network justifying their outputs, reduces trust in the end users who may prefer a decision-making system which is transparent.

This work implements agricultural-based research that integrates ML algorithms, DL techniques and IoT capabilities. These IoT sensors such as soil moisture, temperature, humidity, pH and NPK sensors, help in sensing the current crop parameters and use this data for disease detection of different plants. The ML algorithms, like random forest and SVM, are fed with required inputs like NPK levels, moisture levels and pH readings and combined with temperature and humidity data, which are processed and used to assess the soil. The CNN model is trained with leaf images and is used to detect diseases accurately. The system gives precise yield predictions with the usage of the decision tree algorithm, which makes use of previous crop yields, environment metrics and indicators of soil health.

Materials and Methods

Methodology

The crop recommendation system starts with the collection of sensor data related to soil moisture, temperature, humidity, pH levels and NPK sensors (Fig. 1). Soils from different regions - Raichur, Tumkur and Bangalore have been used for experimentation to check for real-time data. This data is to be cleaned, normalised and formatted, which are the various tasks in the pre-processing stage of a ML model. The crop recommendation system then analyses the data from the pre-processed stage and generates the output. These recommendations are then displayed to the user through the application.

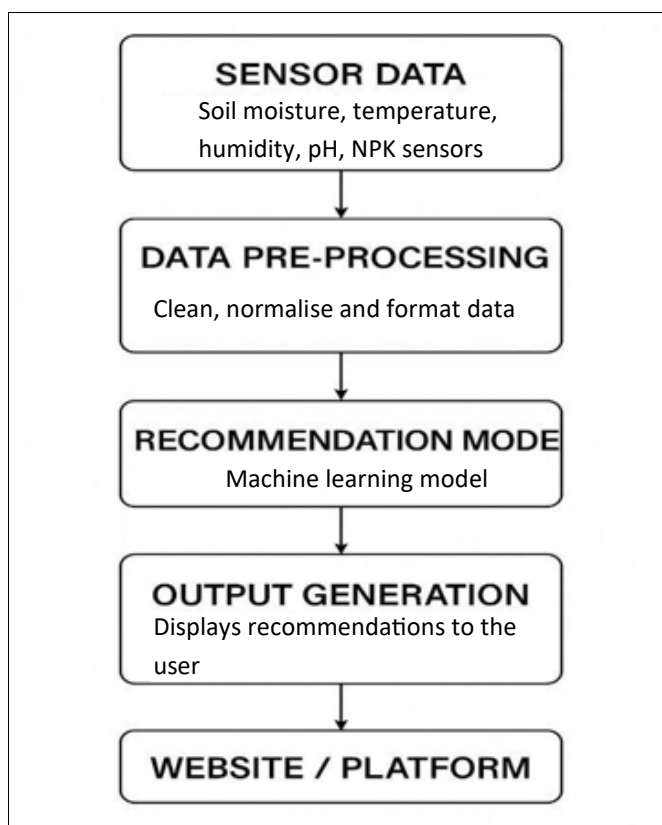


Fig. 1. Crop recommendation.

Accurate predictions and reliable decisions are possible with the integration of historical yield statistics, weather patterns and plant images. A single smart framework includes these three agricultural capabilities that support sustainable farming operations. This aids farmers in planning their tasks as transparent actions, with clear information provided to them through the refined operations in the application.

The plant disease detection system here involves various steps. As seen in Fig. 2, it starts with the uploading of the image of the plant. This image is then resized, normalised and enhanced as part of pre-processing. This pre-processed image is then sent for analysis by a DL model that has been designed for disease detection. The results are displayed to the user through the application.

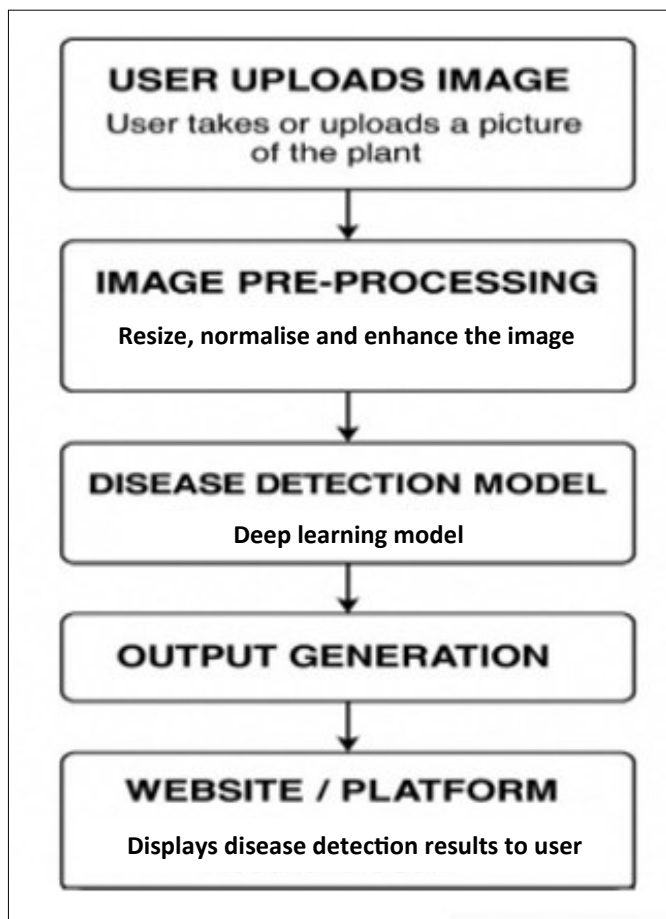


Fig. 2. Plant disease detection.

The yield prediction system starts by acquiring essential input variables like year, country, crop type, pesticide usage, average rainfall and temperature (Fig. 3). These are critical parameters that influence crop production. The system processes the gathered data through a decision tree algorithm to analyse patterns based on historical facts and environmental elements. It then generates estimated crop harvest amounts to support farmers in their planning and decision-making processes.

The selection of ML models has been based on data characteristics, predictive performance, computational efficiency and practical deployment constraints in agriculture. The SVC has been used because it performs well on structured data, can model complex relationships and gives accurate results. For disease detection CNN has been chosen as this can automatically learn important features while processing images. For yield prediction a decision tree model has been chosen as this works with less computing power even with complicated models.

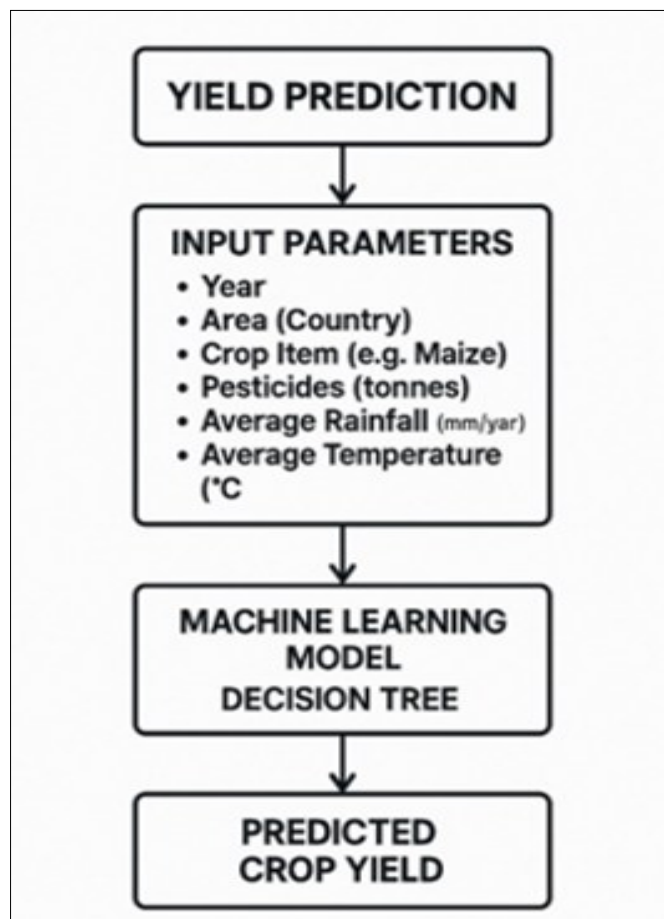


Fig. 3. Crop yield prediction.

The hyperparameters in SVC are set and the justification for the same are also briefed. The kernel is set as radial basis function (RBF) as this supports in capturing nonlinear relationships between soil nutrients (N, P, K), temperature etc. The balance between minimising classification errors and maximising the decision boundary margin is done by the regularisation parameter C which is set as 10 balances margin maximisation and misclassification. It is observed that lower values underfit, higher values overfit. To avoid overly complex decision boundaries and ensure moderate influence of data points, the gamma value is set to 0.01.

The hyperparameters in CNN based disease detection model has been set with the following hyperparameter details. The number of hidden layers used is 2 and filter size used is 3×3 with rectified linear unit activation function, adaptive moment estimation optimiser and adaptive learning. The model has been trained with 30 epochs.

As the depth is an important parameter in decision trees, maximum depth has been set to 10 as this avoids trees that are overly deep avoiding memorising of noise. To ensure meaningful splits, the minimum samples split is set as 5 and the minimum samples leaf has been set as 2 to improve generalisation. The criterion has been set as squared error as it works well for continuous yield prediction.

Real-time environmental and soil data consisting of NPK values, pH levels and temperature information are gathered through IoT sensors during the initial phase. The first stage involves data cleaning, which removes noisy and inconsistent information from the input before proceeding to the next step of data preparation. The data preparation step applies normalisation combined with augmentation to achieve stable and performant

data conditions for subsequent analysis. A testing set, along with a training set, gets established from prepared data, which builds strong ML models. Such models require five essential input parameters, which include nutrient levels together with humidity levels and soil moisture, temperature data and images regarding diseases.

The next step of the system operates a three-part functionality model.

- Soil and climatic factors serve as input for the crop recommendation to determine the optimal crop selection.
- The disease detection system implements CNNs to analyse plant leaf images so it can detect plant infections at their earliest stage.
- The yield prediction module that uses decision trees and makes estimates of the yield, which is dependent on the environmental factors as well as the cultivation parameters (Fig. 4).

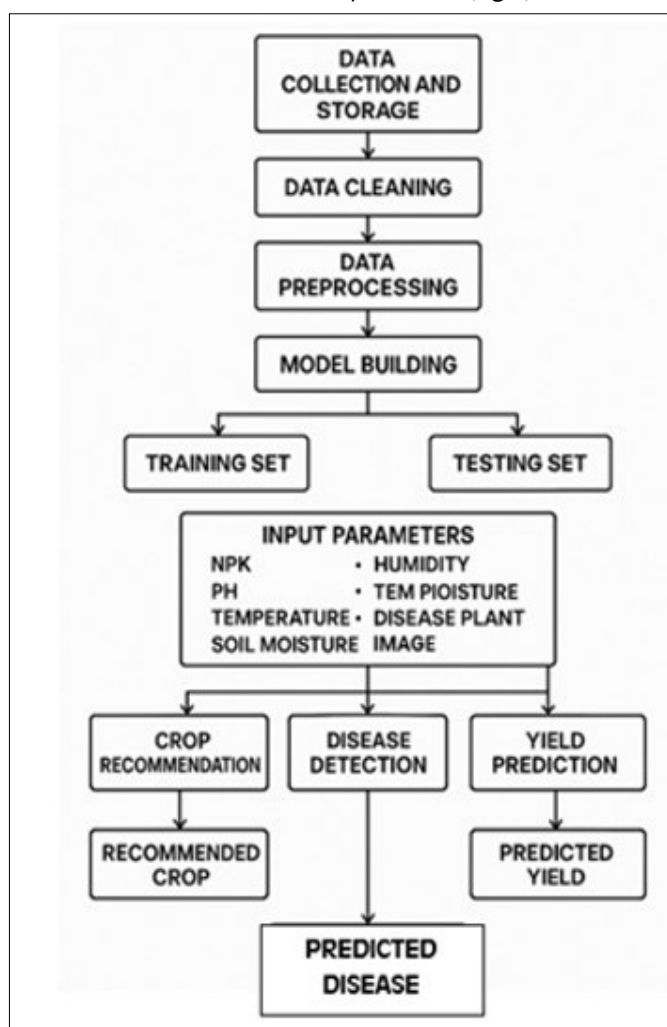


Fig. 4. Integrated crop recommendation and plant disease detection system.

The models have been trained and tested using the Kaggle datasets crop recommendation for recommendation of crops and the PlantVillage dataset for disease detection. In addition to the above, the field testing has also been done using soils from three different regions in Karnataka-Raichur, Tumkur and Bangalore. Sensor data have been collected in real time and the parameters such as N, P, K, pH, moisture etc. are used for testing the model. For disease detection, the model has been tested with real time images of leaves collected from local botanical garden.

As seen in this Fig. 4, a single framework is used to suggest precise agricultural decisions on the proper selection of the crop, as well as give timely warnings on diseases. The Flask platform is used to deploy such a system, which offers scalable features and provides ease of use, scalable and real-time functionality that targets ecological sustainability and better production.

Implementation

Input modules (sensors used)

1. Soil moisture sensor:

Measures soil moisture levels that help in determining the need for irrigation. This also aids in optimal usage of water influencing better crop health.

2. Ultrasonic sensor

Detects obstacles using ultrasonic waves. This facilitates safe navigation for automated systems such as robotic irrigation.

3. NPK sensor

This sensor helps in measuring the N, P and K levels in soil. This supports in determining soil fertility and in turn, in recommending suitable fertilisers.

4. pH sensor

Measures the acidity or alkalinity of the soil. Essential for determining the suitability of soil for different crops.

5. Temperature and humidity sensor (DHT11/DHT22). This monitors atmospheric conditions, including temperature and humidity. This helps in climate-based crop recommendations.

Control units

1. Arduino

This processes data from all sensors. Controls motors and makes real-time decisions based on inputs.

2. ESP32

This manages the web interface for remote monitoring and control, as well as integrates GPS for tracking the location of the system. This also controls servos for various mechanisms like irrigation and spraying.

Output Modules

1. LED display / LCD screen

This shows real-time sensor readings (NPK values, pH, moisture, temperature, humidity). This also displays crop recommendations based on collected data.

2. Cloud data storage (AWS IoT / firebase)

This storage sends sensor data to the cloud for remote monitoring and analysis, as well as stores historical data to track soil health trends over time.

3. Automated irrigation system

This has a feature that allows turning on/off water pumps based on soil moisture levels. This also ensures efficient water usage.

4. Mobile/Web interface

Users can remotely monitor soil conditions and crop recommendations. This provides alerts and notifications for irrigation and fertilisation needs.

Here, Fig. 5 provides sensor details, which are crucial in data

collection for the ML models. The sensor data serves as the input, influencing the overall model performance and accuracy. The IoT model is shown in Fig. 6. The effectiveness of each algorithm depends on how well it interprets and processes this sensor data, contributing to the variations in accuracy are observed in the Table 1 below.

An IoT is designed for smart agriculture and consists of an NPK sensor, pH sensor, soil moisture sensor, temperature and humidity sensor and LED display.

The sensors are used to provide data on the real-time soil and environment, which can be processed and displayed on the screen. In view of better crop management, this data is transferred to the AWS cloud for storage and further analysis, resulting in efficient monitoring and better decision-making.

Real time challenges such as computational cost, internet dependency, latency, or performance in low-connectivity rural environments can be evaluated with field deployment which is included as a part of future work.

The datasets used are the ones stated below

1. PlantVillage dataset

<https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset?resource=download>

2. Crop recommendation dataset

<https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>

Results and Discussion

Crop recommendation system

The dataset used for crop recommendation considers data preprocessing, exploratory data analysis (EDA) and appropriate ML models for classifying and recommending crops. A comprehensive analysis was conducted to evaluate model accuracy and performance.



Fig. 6. Internet of things (IoT) model.

Dataset overview

- The dataset consists of 2200 entries with 8 columns:
N, P, and K (macronutrients in soil)
Temperature, humidity, pH, rainfall (Environmental conditions)
Label (crop type)
- No missing values are found and the dataset appears clean.
- The training testing split has been 80 % and 20 % respectively.

Model performance and accuracy

The crop recommendation system was evaluated using various ML models, achieving the following accuracies.

Overall, ensemble models (Naïve Bayes, decision tree and bagging) performed exceptionally well.



Fig. 5. Internet of things (IoT) sensors used for collecting soil and environmental parameters.

Table 1 presents the accuracy comparisons of various ML models. Among them, random forest achieved the highest accuracy at 99.09 %, followed closely by bagging (98.86 %), decision tree (98.63 %) and Naïve Bayes (98.54 %). Models like gradient boosting (98.18 %) and support vector machine (96.81 %) also demonstrated strong performance. However, AdaBoost exhibited significantly poor accuracy at 14.09 %, indicating its inefficacy for this task. AdaBoost has a much lower accuracy rate (14.09 %) than other models as it is more sensitive to noisy and misclassified samples. However, AdaBoost sequentially emphasises incorrectly classified instances, which increases the effect of noise and outliers. This is different from ensemble methods like random forest, which reduce variance through parallel learning. These factors together make it perform worse.

The model's performance was also tested for values collected in real-time from soils of three different regions.

Table 1. Accuracy comparison of different models

Model	Accuracy (%)
Logistic regression	96.36
Naive Bayes	98.54
Support vector machine	96.81
K-nearest neighbours	95.90
Decision tree	98.63
Random forest	99.09
Bagging	98.86
AdaBoost	14.09 (poor performance)
Gradient boosting	98.18
Extra trees	92.95

Plant disease prediction using convolutional neural networks

Dataset overview

The work utilises the PlantVillage dataset from Kaggle, which contains a vast collection of labelled images depicting healthy and diseased plant leaves. The training testing split has been 80 % and 20 % respectively.

Model architecture

Convolutional neural network

The work implements a CNN model designed to extract features from leaf images and classify them into respective disease categories. The CNN based disease detection model has the following architectural details. The number of hidden layers used is 2 and filter size used is 3×3 with rectified linear unit activation function, adaptive moment estimation optimiser and adaptive learning. The model has been trained with 30 epochs.

Model performance and accuracy

While the specific accuracy metrics of the model are not detailed in the provided information, similar studies have demonstrated high accuracy in plant disease detection using CNNs. For instance:

- Early prediction of plant diseases using CNN and generative adversarial networks (GANs): This study achieved a classification accuracy of 97 % using a CNN model and improved to 98 % by incorporating GANs for data augmentation.
- The model's performance was also tested using real images of leaves during disease detection.

- Deep learning-based model for plant lesion segmentation: This research reported an average accuracy of 92 % in lesion segmentation and up to 99.04 % in lesion subtype recognition for certain plants.

These studies indicate that CNN-based models can achieve high accuracy in plant disease detection tasks.

Data augmentation

As mentioned above, for plant disease prediction, CNN has given a classification accuracy of 97 %, without applying data augmentation. When data augmentation was incorporated using GANs, the accuracy improved to 98 %.

Model optimisation

Incorporating advanced techniques such as squeeze-and-excitation blocks and residual skip connections can further enhance the model's ability to detect and classify plant diseases accurately. The integration of CNNs in plant disease prediction offers a promising approach to accurately identify and classify various plant diseases. By leveraging comprehensive datasets and advanced DL techniques, such models can significantly aid in early disease detection, thereby contributing to improved crop management and yield.

Crop yield prediction using a decision tree

Dataset overview

The yield prediction model operates on structured agricultural data, which contains historical environmental factors such as year, country, crop types, like coffee, cotton, jute, maize, mango etc., together with pesticide usage and metrics of average annual rainfall (in mm) and average temperature (°C). The data points provide critical knowledge for creating models that study different factors affecting crop yields.

Model architecture

Decision tree algorithm

As the decision tree can analyse the numerical data effectively as well as the categorical data, it has been considered as the base model for this work. As this method considers the feature thresholds that are identified for branching when the dataset is segmented, it is easier to visualise and interpret the dataset.

Input parameters

Year: The year of cultivation.

Area: The geographic region or country. Average rainfall: Annual rainfall (mm/year).

Average temperature: Mean temperature (°C).

Model output

The model presents its analysis through predicted crop yield, establishing agricultural productivity predictions based on historical data and climatic patterns. This prediction assists farmers and agricultural planners in resource allocation and risk assessment.

Model performance and accuracy

Decision tree models demonstrate reliable performance for yield forecasting, according to current industry understanding. While exact metrics for this specific application are not provided, previous field research works dealing with similar data have reported

successful results consistent with the performance of this work. The proposed work applies machine learning techniques for agricultural yield prediction demonstrating 85–90 % prediction accuracy. Decision trees matched random forests in terms of yield prediction during the study, while using fewer computational resources than other regression models. The screenshots of the user interface designed to help farmers know the crop recommendation system and disease detection system (Fig. 7–14).

Conclusion

Modern technologies such as precision agriculture and smart farming systems help advance agricultural production and contribute to India's economic development. The proposed system integrates plant disease detection, crop recommendation and crop yield prediction using convolutional neural networks, support vector classifier and decision tree algorithms. The deployment of IoT sensors enables real-time data acquisition on nitrogen, phosphorous and potassium (NPK) values, soil moisture, temperature and humidity, which enhances the precision of all the forecasted outcomes.



Fig. 7. Home page.



Fig. 8. Register and login.

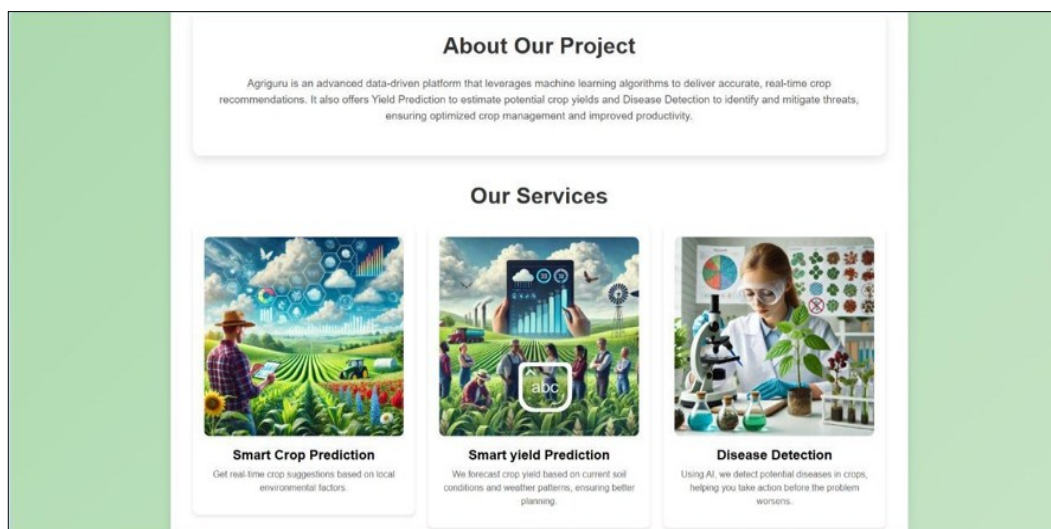


Fig. 9. Our services.

The screenshot shows the 'Crop Recommendation System' interface. At the top is a green navigation bar with links: Home, About, Crop Recommendation, Yield Prediction, and Disease Detection. On the right are 'Login' and 'Register' buttons. The main content area has a light green background with a white card. The card features the slogan 'Grow Smarter, Not Harder 🌱 – Your Crop Recommendation Awaits!' and the title 'Crop Recommendation System' with plant icons. Below this is a dark blue 'Input Details' form with fields for Nitrogen, Phosphorus, Potassium, Temperature (°C), Humidity (%), pH, and Rainfall (mm). A green 'Get Recommendation' button is at the bottom of the form.

Fig. 10. Crop recommendation system.

The screenshot shows the 'Plant Disease Detection' interface. It has the same green navigation bar as Fig. 10. The main content area features a white card with the title 'Plant Disease Detection' and the instruction 'Upload an image of the plant leaf to detect any potential diseases. Supported formats: PNG, JPG, JPEG.' Below this is a file upload section with a 'Choose File' button and a 'No file chosen' text. A green 'Predict Disease' button is positioned below the upload area. At the bottom of the page, a green footer contains the text '© 2025 Crop Recommendation. All rights reserved.'

Fig. 11. Plant disease detection.

The screenshot shows the 'Prediction Result' page. It features the same green navigation bar. The main content area has a white card with the title 'Prediction Result'. In the center is an image of a potato leaf showing early blight symptoms. Below the image, the text reads 'The Model Predicts: Potato__Early_blight'. A green button with a circular arrow icon and the text 'Try Another Image' is located below the prediction text. A green footer at the bottom contains the text '© 2025 AgriDerm. All rights reserved.'

Fig. 12. Plant disease detection results.

Home About Crop Recommendation Yield Prediction Disease Detection Login Register

Unlock The Power Of Precision 🌱 – Predict Your Crop's Yield Like Never Before!

Crop Yield Prediction Per Country 🌍

Input All Features Here

Year	Average Rainfall (mm/year)
2013	1485
Pesticides (tonnes)	Average Temperature (°C)
121	16
Area	Item
Albania	Maize

Predict Yield

Fig. 13. Crop yield prediction.

Home About Crop Recommendation Yield Prediction Disease Detection Login Register

Unlock The Power Of Precision 🌱 – Predict Your Crop's Yield Like Never Before!

Crop Yield Prediction Per Country 🌍

Input All Features Here

Year	Average Rainfall (mm/year)
2013	
Pesticides (tonnes)	Average Temperature (°C)
Area	Item

Predict Yield

Predicted Yield:
6712.0

Fig. 14. Crop yield prediction-Results.

With the help of this system, farmers can detect diseases well in advance, while receiving specific crop recommendations and accurate yield forecasts to guide better choices and resource utilization. The easy-to-use interfaces on web and mobile platforms help different groups of farmers to quickly adapt to this system. The future work may include a system that amalgamates features that combine disease prediction capabilities with wholesale market price forecasts for enhanced sustainability of agriculture. Also, the future work orients towards removal of dataset bias, a greater number of real-time leaf images for training and testing of disease detection model and real-time field deployment to evaluate the challenges such as cost and usability.

Acknowledgements

The authors extend their sincere gratitude to B.M.S. College of Engineering for providing an encouraging research environment and institutional support.

Authors' contributions

NV conceived the problem statement, conducted the survey, contributed to the crop recommendation system design and assisted in budgeting and hardware procurement. RM conducted the survey and contributed to the disease detection system design. SVA conducted the survey and assisted in selecting hardware components and datasets. BM participated in the study and implemented the crop recommendation system. SL participated in the study and implemented the disease detection system. AU participated in the study and implemented the crop recommendation system. PSK participated in the study and implemented the disease detection system. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

References

1. Srivastava D. AI-powered innovations in plant pathogen detection: transforming agriculture through technology. *Int J Plant Pathol Microbiol.* 2025;5(2):60–6. <https://doi.org/10.22271/27893065.2025.v5.i2a.145>
2. Kim YH, Yoo SJ, Gu YH, Lim JH, Han D, Baik SW. Crop pests prediction method using regression and machine learning technology: survey. *IERI Procedia.* 2014;6:52–6. <https://doi.org/10.1016/j.ieri.2014.03.009>
3. Kranth GPR, Lalitha MH, Basava L, Mathur A. Plant disease prediction using machine learning algorithms. *Int J Comput Appl.* 2018;182(25):1–7. <https://doi.org/10.5120/ijca2018918049>
4. Jeevaganesh R, Haris D, Priya B. A machine learning-based approach for crop yield prediction and fertilizer recommendation. In: *Proceedings of the IEEE 6th International Conference on Trends in Electronics and Informatics (ICOEI)*; 2022; India. IEEE; 2022. p. 1330–4. <https://doi.org/10.1109/ICOEI53556.2022.9777230>
5. Patki V, Wazurkar P. Application of data sampling for crop yield prediction using stochastic gradient descent neural networks. In: *Proceedings of the 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*; 2021; India. IEEE; 2021. p. 685–9. <https://doi.org/10.1109/CSNT51715.2021.9509682>
6. Shahrin F, Zahin L, Rahman R, Hossain AJ, Kaf AH, Azad AA. Agricultural analysis and crop yield prediction of Habiganj using multispectral bands of satellite imagery with machine learning. In: *Proceedings of the 11th International Conference on Electrical and Computer Engineering (ICECE)*; 2020; Bangladesh. IEEE; 2021. p. 21–4. <https://doi.org/10.1109/ICECE51571.2020.9393066>
7. Thirumal S, Latha R. Automated rice crop yield prediction using sine cosine algorithm with weighted regularized extreme learning machine. In: *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*; 2023; India. IEEE; 2023. p. 35–40. <https://doi.org/10.1109/ICICCS56967.2023.10142403>
8. Choudhary V, Thaku A. Comparative analysis of machine learning techniques for disease prediction in crops. In: *Proceedings of the 11th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*; 2022; India. IEEE; 2022. p. 190–5. <https://doi.org/10.1109/CSNT54456.2022.9787661>
9. Mehetre S, Mangle R, Karvanje P, Sarode RY. Crop disease diagnosis using convolutional neural network. In: *Proceedings of the 3rd International Conference on Intelligent Technologies (CONIT)*; 2023; India. IEEE; 2025. p. 1–7. <https://doi.org/10.1109/CONIT59222.2023.10205929>
10. Lohi SA, Bhatt C. Design of a crop disease detection model using multi-parametric bio-inspired feature representation and ensemble classification. In: *Proceedings of the 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*; 2023; India. IEEE; 2023. p. 1–6. <https://doi.org/10.1109/ICITIIT57246.2023.10068649>
11. Dhande A, Malik R. Empirical study of crop-disease detection and crop-yield analysis systems: a statistical view. In: *Proceedings of the International Conference on Emerging Smart Computing and Informatics (ESCI)*; 2022; India. IEEE; 2022. p. 1–4. <https://doi.org/10.1109/ESCI53509.2022.9758284>
12. Nithyasundari B, et al. HNNL: experimental analysis of wheat crop disease detection using hyper neural network-based learning methodology. In: *Proceedings of the International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*; 2024; India. IEEE; 2024. p. 1–7. <https://doi.org/10.1109/ACCAI61061.2024.10602195>
13. Kancharagunta KB, et al. Crop recommendation system using machine learning and IoT: a survey. In: *Proceedings of the International Conference on Innovative Computing and Communication.* Singapore: Springer Nature; 2024. p. 63–86. https://doi.org/10.1007/978-981-97-4149-6_5
14. Islam MR, Oliullah K, Kabir MM, Alom M, Mridha MF. Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. *J Agric Food Res.* 2023;14:100880. <https://doi.org/10.1016/j.jafr.2023.100880>
15. Ekanayaka EMHI, Kumara PPNV. Machine learning approach for new crop disease predict and alert system: a review. In: *Proceedings of the International Conference on Emerging in Computational Intelligence (ICETCI)*; 2022; India. IEEE; 2022. p. 45–9. <https://doi.org/10.1109/ICETCI55171.2022.9921364>
16. Deshmukh T, Rajawat AS, Potgantarwar A. Machine learning technique for crop selection and prediction of crop cultivation. In: *Proceedings of the International Conference on Advanced Computing Technologies and Applications (ICACTA)*; 2023; India. IEEE; 2024. p. 1–7. <https://doi.org/10.1109/ICACTA58201.2023.10392348>
17. Kumar R, Shukla N. Plant disease detection and crop recommendation using CNN and machine learning. In: *Proceedings of the International Mobile and Embedded Technology Conference (MECON)*; 2022; India. IEEE; 2022. p. 168–72. <https://doi.org/10.1109/MECON53876.2022.9752173>
18. Vaishnnave MP, Kamaitachi SB, Kumar CR, Premkumar S. Prediction of crop based on various predictive algorithmic models. In: *Proceedings of the 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*; 2023; India. IEEE; 2024. p. 725–8. <https://doi.org/10.1109/ICAC3N60023.2023.10541486>
19. Chana AM, Batchakui B, Nges BB. Real-time crop prediction based on soil fertility and weather forecast using IoT and a machine learning algorithm. *Agric Sci.* 2023;14(5):645–64. <https://doi.org/10.4236/as.2023.145044>
20. Senapati BR, Sanskar, Trishna A, Swain RR. Recommendations of crop yield and fertilizers using machine learning algorithm. *J Inf Optim Sci.* 2022;43(5):1029–37. <https://doi.org/10.1080/02522667.2022.2094541>
21. Katarya R, Raturi A, Mehndiratta A, Thapper A. Impact of machine learning techniques in precision agriculture. In: *Proceedings of the 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*; 2020; India. IEEE; 2020. p. 1–6. <https://doi.org/10.1109/ICETCE48199.2020.9091741>
22. Talaat FM. Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. *Neural Comput Appl.* 2023;35(23):17281–92. <https://doi.org/10.1007/s00521-023-08619-5>
23. Kumar SR, Sowrirajan S. Automatic leaf disease detection and classification using hybrid features and supervised classifier. *Int J Adv Res Electr Electron Instrum Eng.* 2016;5(6):4556–63. <https://doi.org/10.15662/IJAREEIE.2016.0506008>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.