IoT-driven automated water irrigation Along with Crop Prediction and Disease Detection Using ML model for smart agriculture

Arijit Ghosh, Rishav Raj, Mohammad Adil Imam, Snehasis Paul, Debarka Das, Gaurav Kumar, Shivam Raj, Avinav Rai, Tanisha Goswami, Soumik Kumar Kundu, Samit Karmakar, Gobina Sen, Amitava Ghosh, Gautam Ghosh, Malay Gangopadhyay
Department of Electronics and Communication Engineering
Institute of Engineering & Management, Kolkata, India,
Arijit.Ghosh2022@iem.edu.in, Rishav.Raj2022@iem.edu.in , MOHAMMAD.ADILIMAM2022@iem.edu.in,
Snehasis.Paul2022@iem.edu.in, Shivam.Raj2022@iem.edu.in, soumik.kundu@iem.edu.in,

***Abstract*— These days, agriculture is dealing with a lot of big Climate change, waste of resources, and too many a lot of work by hand. To fix these issues, smart Farming is very important right now. We have done this project. built a whole system for smart farming using the Internet of Things, Deep Learning and Machine Learning to make the making. We are using an ESP8266 to run our system. microcontroller. We used different sensors for the soil and to see how the field is doing in the environment. It can begin the automatically waters based on some set limits. There There is a web server on the ESP8266 that lets us see real- You can get real-time data and control the water even when there is no cloud. help. We also have a web platform called Agrosense to You can look at live and old data from anywhere. For better crops We made a Machine Learning model for management. The The model looks at the pH, nutrients, temperature, and rain in the soil. We tried a lot of different models and found that the Random Forest The algorithm gives the best results. We also put in a system that uses Deep Learning to find different plant diseases (CNN with ResNet-50). By combining IoT automation and Our system is very reliable and can do a lot of things because it uses smart prediction models.**

Keywords*— *Smart Agriculture, Internet of Things (IoT), ESP8266, Automatic Irrigation, Crop Prediction, Crop Disease Detection, Machine Learning, Deep Learning.

I. Introduction

Agriculture is very important for food security and economy, especially in countries that are still developing. But old Farming methods rely excessively on manual labor and farmer's skill. Because of this, water is sometimes wasted. People pick the wrong crops and don't find diseases quickly. Their problems are made worse by climate change and less rain. worse. We really need smart and automatic farming now. system to make production better. IoT has made it possible to You can keep an eye on the farm in real time with cheap sensors and devices. time. We can measure how wet the soil is with IoT. all the time, the temperature and humidity. This helps get things going. Irrigation happens on its own with very little help from people. The The ESP8266 microcontroller is perfect for this because it uses It uses little power and has a web server and Wi-Fi built in. Also, Machine Learning is also very useful for things other than IoT. farming. These models can figure out how things are related in complicated ways. between the soil, the weather, and the growth of crops. This helps to suggest which crop would be best for the soil. Well-liked Some examples of algorithms are Random Forest, SVM, Decision Tree, and KNN is used because it works better for crops. guessing. Crop diseases are another big problem for farmers. Most of the time, Experts look at the plants to find diseases, but it is a very A process that takes a long time and a lot of work. Now, with Deep Learning We can easily find diseases with image analysis. For example, ResNet-50 and other CNN models use "transfer learning" to figure out what diseases a plant has by looking at pictures of its leaves very accurate. We put forward a smart agriculture system in this paper. It includes IoT sensing for automatic irrigation and machine learning. for predicting crops and Deep Learning for finding diseases. The ESP8266 has a local web server and a Agrosense is a platform that shows data. We even put in an an interactive chatbot to give farmers simple advice. This The model is meant to help with crop yield and long-lasting precision farming.

II. LITERATURE REVIEW

Kamilaris et al. examined the applications of big data analytics in precision agriculture, which looked at how big amounts of data from IoT devices, remote sensing, and advanced Tools for precision farming can make farming better. making decisions. [1]. S. Li, J. Zhang et al. in their review paper compared different Machine Learning and Deep Learning Models utilized for the predicting crop growth and yield by comparing the strengths and the pros and cons of different models. It brings out the more and more people are using ensemble learning techniques, Support vector machines and recurrent neural networks, such as LSTM for understanding complicated relationships in farming information. They also talked about how accurate these are. Techniques can be enhanced by integrating various Natural factors like climate, the environment, and soil variables, and also pointing out problems with data generalization of quality and models. [2]. V. VijayLaxmi et al. looked at how things have changed recently with the utilization of diverse machine learning models and the comparative examination of methodologies including regression models, Support vector machines, random forests, and neural networks networks for predicting how much crops will yield when utilized on various environmental and agronomic datasets. They also stressed how decisions based on data over goes beyond standard statistical methods [3]. I. Nagaraju et al. assessed different Machine Learning algorithms for predicting crop yields in the using common agricultural datasets in the field of agriculture. The results showed that the Random Forest method generally got better at predicting than to the other algorithms like Decision Trees and Neural Networks. Support Vector Machines and Trees. [4] T. Gupta et al. proposed a machine learning-based methodology. to forecast the appropriateness of crops and their performance utilizing the datasets that are available, which include the soil traits and different weather conditions. They used classification and regression algorithms to set up the link between environmental factors and the best crop Selection, which showed how data-driven models can help with planning crop decisions instead of traditional crop ways to plan. [5] R. R. Sharma et al. talked about the pros and cons of the machine learning models, including decision trees and k- Support vector machines, nearest neighbors, and ensemble models, concentrating on how these models can support and aid The farmers make decisions based on data about the weather and the soil. and the state of the environment. Their analysis brought out the significance of precise data preprocessing and feature selection to improve the accuracy of the recommendation of the models for machine learning. [6] X. Hu et al. gave a summary of how generative Frameworks can help with adding more data, images, and improvement and finding strange things. It elucidated the theoretical underpinnings of diffusion processes and examined their ability to solve problems with small datasets that are common in tasks involving machine learning in agriculture. The writers also compared diffusion models to other generative models approaches and emphasized their efficacy in generating superior- high-quality synthetic agricultural data that can be used to train strong Models of AI. [7] A. K. Singh et al. looked into another important use. Utilization of machine learning in intelligent agriculture encompasses crop finding and predicting diseases. Systematic reviews People who study precision agriculture have said that image- identifying diseases using deep learning models, especially CNNs, has reached a high level of classification precision across diverse crop species [8]. Hasan et al. conducted an extensive survey on machine learning methods to find and predict crop diseases and pests' diseases and pests, emphasizing the application of sensor-based data and visual characteristics for early disease detection [9]. Metagar et al. examined an extensive array of machine learning models for predicting plant diseases and stressed that hybrid methods that mix old-school ML methods with deep learning models, you can make things stronger and generalization performance [10].

III. **System Architecture and Working Principle**

The proposed smart farming platform is being made as an an IoT system that works together to make the whole monitoring the environment, making decisions automatically, and management of irrigation and the ability to control it from a distance the field conditions into one embedded setup. At the microcontroller ESP8266 is at the heart of the system. NodeMcu is the main controller and the module for communication. The microcontroller gathers the It gets data from different sensors and processes the sensor values in in real time, it controls the irrigation pump and also gives users a way to keep an eye on things.

The system architecture was split into four main parts. sensing, control, actuation, and user are the functional layers. interaction. The sensing layer has a sensor for soil moisture. a sensor, a temperature and humidity sensor, a rain sensor, and a sensor for light. These sensors keep an eye on the soil and the weather in the field. The sensors have all the information they need. sent to the ESP8266 microcontroller, which the controller is always comparing and analyzing the sensor values using set threshold values and programmed logic for making decisions.

The control layer was in charge of deciding when needs to be watered. If the soil is too wet and there is no rain when it goes below a certain level When it finds something, the ESP8266 turns on the relay module to start the water pump. Once the soil has enough moisture the system turns off the pump on its own when it gets to a certain point. The automated decision-making process helps stop over-watering and using water wisely. The actuation layer does the physical work of the system for watering. The relay-controlled part is the main part of it. water pump, which sends water to the crops whenever when irrigation is needed. A buzzer is also included with this. in the system to make sounds that let you know when something is wrong situations like very dry soil or a system that could break down problems.

The system had both local and remote interfaces for watching and talking to users. Readings of soil in real time the IP address of the temperature, humidity, and moisture host, which let the farmers quickly look at all the were able to see the conditions and the system status right in the field. displayed on a 16×2 LCD screen that was attached to the device. The ESP8266 also had a simple a web-based dashboard where it could see all the conditions of the field, which could be reached with a WiFi-enabled smartphone, tablet, or computer. The system was linked to a special project website. called AgroSense to make it even easier to use. Here, the sensor data was shown in a way that was easy for users to understand and follow. way. The website also has an inbuilt chatbot where users could ask questions often and get answers. The website let people see both live readings and and historical data, which makes it easier to keep an eye on farm conditions and see how they change over time. The farmers might have been able to AgroSense lets you keep an eye on the system from afar and stay informed about the conditions in their field from almost anywhere.

The suggested system was made up of several different parts. sensors, automated irrigation control, and both local and the ESP8266 and AgroSense let you keep an eye on things from far away. platform gave us a quick and useful way to fix modern farm management while cutting down on the need for supervision by hand all the time.

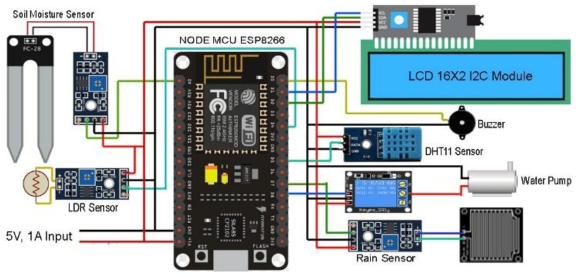

Fig 1: Hardware Circuit Diagram of AgroSense System

IV. Hardware Implementation

The parts for the smart farming system that was suggested were not too expensive. and easy to find, so it would be cheap and able to grow. with the needs of real farms. The ESP8266 NodeMCU is the main controller and the way things talk to each other. the part that makes everything work. It gets information from the sensors that are connected to it, processes the information, and then controls the irrigation system works based on what it sees in the field. A sensor for soil moisture is put in the ground to keep a watch the soil to see how it's doing. This sensor looks at the system uses the moisture level in the soil to decide. if the plants need water or not. It's important to look into the amount of water in the soil so that you only water it when It's too dry.

The DHT11 sensor can tell you how hot and wet it is in the area around it. These things help us learn about the weather has an effect on how crops grow. The system Also, it has a rain sensor module that stops watering the field. when it rains. This sensor can tell when it is raining and stops watering when there is enough water from the weather. A relay module links the controller to a water pump that gives the plants water. The ESP8266 checks the soil and sees that the level of moisture has gone below what it needs to be, it turns on the switch. After that, the pump turns on the relay. that sends water to the plants. The controller shuts off the It will turn on by itself when the right amount of water is back.

When something strange happens, like a buzzer goes off, very dry soil or other system alerts to help people learn a little more about the system. You can see it on a 16×2 LCD screen. Check out what's going on in your area. This display shows the temperature, humidity, and moisture levels in the soil, as well as how the irrigation pump is doing right now. This makes It's simple for users to check the system in the field. The system It also has a water level indicator that shows how full the reservoir is. If the water level is very low or the reservoir is If the battery is dead, an LED light will turn on to let you know.

V. Crop Prediction Model

A. Problem Formulation

Crop prediction means to recommend the best suitable crop according to soil traits and current weather conditions for farming. Traditionally, crop selection depends on farmer experience, but with changing weather and different soil types, this method is not effective every time. Due to this, machine learning models are used, which handle complex and non-linear relationships in agricultural factors in the best way.

.

This task is defined as a multi-class classification problem where the input vector is

$$\mathbf{X} = [N, P, K, pH, T, H, R]$$

This represents soil nutrients and environmental factors. Output **Y** represents the best crop category.

B. Dataset and Feature Space

The crop prediction model is trained using standard crop recommendation datasets, which are widely adopted in ML research. In this dataset, approximately **2,200 samples** and **22 crops** are included

Input Features

The feature space includes:

- **Nitrogen (N)**: Governs vegetative growth and chlorophyll synthesis
- **Phosphorus (P)**: Influences root development and flowering.
- **Potassium (K)**: Controls water regulation and stress tolerance
- **Soil pH**: Determines nutrient bioavailability
- **Temperature (°C)**: Affects enzymatic activity and crop growth cycles
- **Relative Humidity (%)**: Impacts transpiration and disease susceptibility
- **Rainfall (mm)**: Primary water source for rain-fed agriculture

Unlike generic datasets, we deliberately retain potassium and rainfall values. These values represent the agronomic conditions of grapes and rice, not just noise..

C. Data Preprocessing Strategy

To make the model reliable, we apply preprocessing steps:

- Remove inconsistent and duplicate records
- Apply Z-score normalization in distance-based models (SVM, kNN)
- Label encoding on crop categories
- Use an 80:20 train-test split along with implementing 5-fold cross-validation

We train tree-based models like Random Forest on both scaled and unscaled types of data.

D. Model Architectures Evaluated

To identify the best method for predicting crop outcomes several supervised learning were used.

Random Forest (RF)

Random Forest is an ensemble learning method that combines multiple decision trees. These trees are trained on randomized data subsets and features. Due to this design, the risk of overfitting is significantly reduced and the model is able to learn non-linear decision boundaries present in agricultural systems.

Support Vector Machine (SVM)

Radial basis function kernel with SVM was used to model non-linear class separation. This is effective, but with an increase in dataset size, its computational cost significantly increases.

Decision Tree (DT)

Decision Trees provide Decision Trees provide transparent rule-based decisions but, without pruning, suffer from a high variance problem.

k-Nearest Neighbors (kNN)

kNN classifies the crops on the basis of proximity to historical samples. While simple, it scales poorly during inference.

E. Comparative Performance Analysis

Table 1. The comparison of crop prediction models performance

Model Architecture	Accuracy	Strengths	Limitations
Decision Tree	90–98%	Fast training	Overfitting
kNN	97–98%	Simple	Slow inference
SVM	98.6%	Non-linear	High cost
Random Forest	**99.5%**	**Stable/Robust**	**Model size**

F. Discussion

The experimental results indicated that the Random Forest model outperformed other machine learning models. learning models for figuring out which crops would work best. Forest of Random took the predictions from a lot of decision trees and put them together them to make the last guess more correct and less likely to be wrong. This method lets it deal with difficult agricultural data and still get the right answers, even when There are a lot of different things happening in the dataset. Another good thing about Random Forest is that it can find important things that matter, like temperature, the amount of rain, the type of soil, and the amount of humidity. The model can, for example, say how rain affects different crops because it is important for so they can grow. This is very helpful when you want to offer plants that could thrive in similar climates but need different levels of Nitrogen, Phosphorus, and Potassium in the ground. Random was able to see all of these things at once. Forest can give better and more well-rounded advice. Random Forest is a good choice for real-world crop recommendation systems because it can handle a variety of datasets, find important agricultural features, and create predictions that are right. Farmers can use this to pick the best crops that will do well in their land and the weather.

VI. Crop Disease Prediction Model

A. Problem Definition

If crop diseases are not identified timely, then crop losses may happen. Identifying crop diseases manually is a very time-consuming process and can be subjective. This emphasizes the necessity of computer vision-based automated disease detection. The purpose is to classify images of plant leaves as either healthy or diseased.

.

This problem is formulated as a multi-class image classification task, where every image is associated with the corresponding image sample.

This problem is formulated as a multi-class image classification task, where each input image is linked to a specific disease label.

B. Dataset Description

Here, the presented disease prediction model uses the PlantVillage dataset. This dataset includes more than 54,000 leaf images from 38 disease classes and covers 14 crop species.

The dataset includes diseases such as:

- Early Blight
- Late Blight
- Leaf Mold
- Bacterial Spot
- Mosaic Virus

C. Image Preprocessing and Augmentation

To improve generalization, the following steps were applied:

- Image resizing to uniform resolution
- Pixel normalization
- Data augmentation (rotation, flipping, brightness variation)

Augmentation helps the model manage real-world differences, like changes in lighting and complex backgrounds.

D. Deep Learning Architecture

Convolutional Neural Networks (CNNs)

CNNs work well for detecting leaf diseases because they maintain spatial relationships. Convolutional layers pick up low-level features like edges and textures. Meanwhile, deeper layers learn patterns that are specific to diseases.

ResNet-50 with Transfer Learning

ResNet-50 was chosen as the main architecture because its residual connections allow for stable training of deep networks. Transfer learning was used by starting the network with ImageNet weights and adjusting the final layers for crop disease images.

E. Model Training Configuration

- Optimizer: Adam
- Loss Function: Categorical Cross-Entropy
- Epochs: 25
- Batch Size: 32

Training convergence was reached without overfitting because of data augmentation and transfer learning.
Performance Evaluation

F. Performance Evaluation

Table 2. Performance comparison of disease detection models

Model	**Accuracy (%)**	**Key Advantage**	**Limitation**
SVM (handcrafted features)	82–85	Low compute	Limited accuracy
Basic CNN	90–92	End-to-end learning	Needs more data
ResNet-50 (Transfer Learning)	**94–99**	High accuracy, robust	Higher compute
MobileNet	96–97	Lightweight, mobile-ready	Slight accuracy drop

G. Discussion

The results made it very clear that deep learning methods were much better at finding crop diseases than old-fashioned methods. methods for machine learning. Deep learning models can learn complicated patterns from pictures of things on their own plants, which helps them find diseases a lot better. For instance, ResNet-50 did really well because it was deep. The way the network is set up lets it pick up on very small details of Plants that are sick, like spots, color changes, and changes in the feel of the leaves. This research additionally employs transfer learning. which is another important way to do things. This is when a model that has already been trained on a lot of data is changed so That it only looks for diseases in crops. This method worked really well, even when there weren't There aren't many farming pictures to use, which is a common problem. with actual farming data. Lightweight models like MobileNet was also made to use less processing power. and memory. This is why they work well on phones and other phones and tablets. Farmers can use these to quickly and easily check on how healthy their crops are while they are working in the field.

VI. CONCLUSION

Machine learning (ML) is making modern farming a lot better. by letting people make decisions based on what happened in the past and right now. These models help farmers do more and use better use of their resources and farm in a way that is good for the world. They are often better at handling changes. in farming more than old-fashioned and statistical ways. Random forests, decision trees, and other machine learning methods Support vector machines and forests look at things like the weather, soil conditions, and moisture to help you choose the best guess how much they will make and guess how much they will grow. Advanced Deep learning models help it recognize things even better. over time and space, but they need a lot of data and power to work. ML is also very helpful for finding and predicting illnesses in plants. Convolutional neural networks and other types of image- Finding plant diseases is easy with methods that are based on. On the On the other hand, models based on sensors can predict outbreaks. before any signs of sickness show up. Hybrid methods that Combine deep learning with classical machine learning to make them more useful and stronger in the real world. When ML and IoT mix with smart farming systems, one can keep an eye on things in real time and make predictions and change how things work. They use methods that make farming more efficient, lower risks, and ML-driven solutions are good for the environment. very good at fixing problems that come up in modern farming.

ACKNOWLEDGEMENT

The authors sincerely thank Dr. Satyajit Chakrabarti, the Director, and Malay Gangopadhyay, our department head at the Institute of Engineering and Management in Kolkata, for providing the project grant through the IEM-UEM grant-in-aid scheme.

REFERENCES

[1] M. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," Computers and Electronics in Agriculture, vol. 143, pp. 23–37, 2017.

[2] S. Li, J. Zhang, and H. Wang, "Crop yield prediction using machine learning and deep learning techniques: A systematic review," Artificial Intelligence in Agriculture, vol. 9, pp. 1–15, 2025.

[3] V. VijayLaxmi and G. Singla, "Advancements in agricultural forecasting: A review of machine learning based crop yield prediction," Research Review Journal of Agriculture and Allied Sciences (RRJoAST), vol. 4, no. 1, pp. 45–52, 2025.

[4] I. Nagaraju, R. Kumar, and P. S. Reddy, "Machine learning-based crop yield prediction: A comparative study," Journal of Advanced Zoology, vol. 46, no. 2, pp. 123–132, 2025.

[5] T. Gupta, A. Verma, and S. Patel, "Crop prediction using machine learning," IRE Journals, vol. 6, no. 9, pp. 112–118, 2023.

[6] R. Sharma, P. Jain, and M. Kulkarni, "Review of machine learning techniques for crop recommendation systems," International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 10, no. 5, pp. 2150–2156, 2022.

[7] X. Hu, Y. Chen, and L. Wang, "A comprehensive review of diffusion models in smart agriculture," arXiv preprint arXiv:2501.08432, 2025.

[8] A. K. Singh, S. Roy, and D. Mukherjee, "Precision agriculture in the age of artificial intelligence: A systematic review of machine learning for crop disease detection," Agricultural Technology, vol. 3, no. 2, pp. 98–112, 2025.

[9] M. Hasan, A. Rahman, and N. Islam, "Machine learning for detection and prediction of crop diseases and pests: A survey," Agriculture, vol. 12, no. 9, pp. 1350–1365, 2022.

[10] S. Metagar and G. Walikar, "Machine learning models for plant disease prediction and detection: A review," Agricultural Science Digest, vol. 44, no. 1, pp. 1–8, 2024