

Real-Time Crop Prediction Based on Soil Fertility and Weather Forecast Using IoT and a Machine Learning Algorithm

Anne Marie Chana, Bernabé Batchakui, Boris Bam Nges

Department of Computer Engineering, National Advanced School of Engineering, Yaoundé, Cameroon
Email: anne.chana@univ-yaounde1.cm, bernabe.batchakui@univ-yaounde1.cm, bam.nges@enspy-uy1.cm

How to cite this paper: Chana, A.M., Batchakui, B. and Nges, B.B. (2023) Real-Time Crop Prediction Based on Soil Fertility and Weather Forecast Using IoT and a Machine Learning Algorithm. *Agricultural Sciences*, 14, 645-664.

<https://doi.org/10.4236/as.2023.145044>

Received: April 5, 2023

Accepted: May 22, 2023

Published: May 25, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).
<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

The aim of this article is to assist farmers in making better crop selection decisions based on soil fertility and weather forecast through the use of IoT and AI (smart farming). To accomplish this, a prototype was developed capable of predicting the best suitable crop for a specific plot of land based on soil fertility and making recommendations based on weather forecast. Random Forest machine learning algorithm was used and trained with Jupyter in the Anaconda framework to achieve an accuracy of about 99%. Based on this process, IoT with the Message Queuing Telemetry Transport (MQTT) protocol, a machine learning algorithm, based on Random Forest, and weather forecast API for crop prediction and recommendations were used. The prototype accepts nitrogen, phosphorus, potassium, humidity, temperature and pH as input parameters from the IoT sensors, as well as the weather API for data forecasting. The approach was tested in a suburban area of Yaounde (Cameroon). Taking into account future meteorological parameters (rainfall, wind and temperature) in this project produced better recommendations and therefore better crop selection. All necessary results can be accessed from anywhere and at any time using the IoT system via a web browser.

Keywords

Smart Farming, Crop Selection, Recommendation of Crops, IoT, Machine Learning, Weather Forecast

1. Introduction

Agriculture, the backbone of the economy to most developing countries, should normally contribute to its economic growth. But ignorantly, farmers grow crops

by simply taking a piece of land, clearing it, plowing it and planting crops without formal knowledge of the nutrient content of the soil. Some even move further to apply fertilizer with the hope of maximizing their yields. As a result, some farmers may end up in excessive fertilizer use, leading to lower yields than expected.

Some common problems that prevent the development of the economy of agriculture include but are not limited to: lack of adequate storage facilities, limited market for marketing of agricultural products, climatic conditions, lack of basic infrastructure, pest and diseases, lack of adequate research, shortage of capital, etc. [1].

As a general rule, the soil must have the right ratio of soil nutrients, and weather conditions must be favorable for crops to grow well and produce good yields [2]. Some of these parameters may vary depending on the type of soil and crop under consideration. Most of the time, fertilizers are used because the soil does not contain enough nutrients to grow the desired crops. This occurs in part because farmers are unable to identify the best crops that can naturally grow on a particular soil, hence the use of fertilizer to promote growth. Poor crop selection without consideration of the soil leads to low yields, resulting in excessive fertilizer use that degrades the environment. Based on soil nutrients and weather conditions, a farmer will be able to know which crops perform best in a particular soil [3].

Over three cropping systems, the increase in crop output is exponentially associated with the growth in an Integrated Fertility Quality Index (IFQI). Production variability between years is demonstrated to be adversely, linearly associated to integrated fertility quality index ($p < 0.001$) across treatments and systems [4].

The use of advanced technologies in the agricultural sector has a positive impact as it significantly contributes to crop production. Smart agriculture based on IoT and AI technologies, allow growers and farmers to monitor and optimize the resources involved [5] [6] [7].

Muangprathub and al [8] propose an optimal watering system for agricultural crops based on a wireless sensor network to analyze data and predict soil temperature, moisture and water content. Thorat and al [9], in addition to monitoring environmental parameters also detect leaf diseases. IoT, when used alone, might not permit the accurate recommendation of best crops because of the lack of weather forecast parameters. Machine learning which is a method of data analysis that automates analytical model building using weather data, can be used to predict or forecast future weather conditions [10]. Also, Machine learning algorithms due to its sense of intelligence, can generally be used to do recommendations based on past data and the analysis of current data. Patil [11] use semi-supervised learning with the Q-learning algorithm to detect soil fertility parameters in a real-time environment and identify chemical properties such as PH values, moisture, temperature, soil humidity. Classification and prediction of soil parameters such as phosphorus, potassium, pH and others help to determine

the appropriate amount of nutrients to be added. The report by Suchithra *et al.* [12] uses the fast learning technique known as Extreme Learning Machine (ELM) with different activation functions to improve soil health and environmental quality. In the report by Girish [13] and Nischitha [14], the authors use the predictive analysis carried out by using various machine learning algorithms like linear regression, Support Vector Machine (SVM), K-Nearest Neighbour (K-NN) method and Decision Tree algorithm to allow farmers to choose the particular crop that is suitable for specific rainfall and market crop price. The report by Nischitha [14] depicts a suggested system for predicting the best crop for a specific plot of land based on soil composition and weather parameters such as temperature, humidity, soil PH, and rainfall using machine learning predictive algorithms like Support Vector Machine (SVM) and Decision Tree.

The authors Patil [11], Suchithra [12] and Nischitha [14] propose IoT architectures with machine learning algorithms but the prediction accuracy was low. Soil with weather parameters were tested, and results showed that SVM with 89% had the highest efficiency in rainfall prediction compared to Linear regression with 86%, Decision Tree with 49% and K-NN with 52%. An important aspect here is that soil data is collected using sensors which allows for real time prediction.

Motivated by the limits of the above review, we ask ourselves the question, how is AI helpful in the real time recommendation of crops for a particular type of soil based on soil fertility and weather forecast? More precisely, how can IoT and machine learning be used to test soil fertility and help in choosing the appropriate crop with weather forecast?

The goal of this work is therefore to help farmers make better crop selection decisions using IoT, machine learning, and weather forecast. The prototype created accepts input parameters for the IoT sensors such as nitrogen, phosphorus, potassium, moisture, and pH, as well as the weather API for data forecasting. The machine learning algorithm used is Random Forest as it gave the best accuracy compared to the other algorithms for crop prediction. This algorithm was trained with jupyter in anaconda on data-set, put together by [15]. In this work, the authors used the Random Forest algorithm in order to get better prediction accuracy while considering similar input conditions. A key particularity of this work is to factor in the aspect of weather forecast and give real time recommendation to some actions that can be taken by the farmer seven or more days before an actual unwanted climatic condition occurs. This has initially been ignored by most authors though being a significant aspect in the kind of crop that can be predicted. In this article, the authors use the predicted crop based on fertility and monitor them closely with weather forecast data.

The Section 2 presents the methodological approach of data collection, the training principle of the model and the implementation of the prototype that has been developed. Section 3 presents the results regarding data collection, weather forecasting and crop recommendations. Section 4 is devoted to the interpreta-

tion and discussion of the results. Section 5 concludes the work.

2. Materials and Methods

2.1. Data Collection

2.1.1. Collection of Nitrogen, Phosphorus, Potassium and pH

Here historical data that will be used for the training and testing of our model is needed. Real time data used to validate and predict the suitable crop based on the trained model is also needed. Historical data of N, P, K, pH, Temperature and Humidity corresponding to specific crops was gotten from an already existing data-set [15]. This data-set had 2200 records for four crop types. In order to measure the contents of any soil in real time, sensors were used to measure the amount of N, P, K, pH, T and H. The specifications of the soil sensor can be seen in **Table 1**.

2.1.2. Collections of Weather Data

The authors used the Short-range forecast (SRF) since a period of 7 days and above was needed to get the forecast values of Precipitation (P), Temperature (T), Humidity (H) and Wind (W) which is powered by convolution machine learning solutions. In order to get the actual values of P, H, T and W, daily API calls to the open weather forecast website was done using geographical coordinates of the area under forecast like latitude, longitude, count of future days (7+ in this case), and also the API key of the connection. The response from this API request comes in JSON format with the values of Temperature, Precipitation,

Table 1. Soil sensor specifications.

Item Type	Soil Sensor
Maximum Power Consumption	12 - 24 V DC
Operating Temperature	0°C to 55°C
Nitrogen Phosphorus Potassium	Nitrogen phosphorus potassium 1 - 1999 mg/kg (mg/L) range; resolution rate: 1 mg
Temperature Measurement Range	-40°C - 80°C
Temperature Accuracy	±0.5°C
Temperature Resolution	0.1°C
Moisture Measurement Range	0% - 100%
Moisture Accuracy	±3% within the range of 0% - 53%; ±5% within the range of 53% - 100%
Moisture Resolution	0.10%
pH Measurement Range	3 - 9 pH
pH Measurement Accuracy	±0.3 pH
Brand	Sorandy

Wind Speed, Humidity, and many others which were then extracted and interpreted.

2.2. Machine Learning

The Random Forest algorithm was used to examine crop growth in relation to the present soil parameters: pH, N, P, K, T and H [16] and [17]. The dataset here represents a collection of our data-set made up of N, P, K, T, H, pH and the recommended crop. The trees represent the different models for the Random Forest which are grouped into classes. Every class brings out the best outcome based on the decision tree samples. When this is done, a majority voting is now conducted to choose the best class. The Random Forest algorithm generates Decision Trees from distinct data samples, predicts the data from each subset, and then provides a better answer for the system via voting. The data is trained using the bagging approach which improves the accuracy of the outcome. The Random Forest training algorithm applies the common technique of bootstrap aggregation, or bagging, to tree learners. Given a set of training data $X = x_1, \dots, x_n$, where X represents the values of N, P, K, H and T respectively in the set; with responses $Y = y_1, \dots, y_n$, where Y represents to the results of the set X bagging repeatedly (B times) selecting a random sample with replacement of the training set and fits trees to these samples.

For $b = 1, \dots, B$: we sample with replacement, n training examples from X and Y ; call these X_b and Y_b . We then train a classification or regression tree f_b on X_b, Y_b . After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x' given by:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (1)$$

In the case of classification trees, the majority vote is used because it reduces the variance of the model without increasing the bias. This bootstrapping approach improves model performance. It indicates that, while a single tree's predictions are highly sensitive to noise in its training set, the average of several trees is not, as long as the trees are not coupled. Training several trees on a single training set would result in heavily correlated trees (or the same tree many times if the training procedure is deterministic); bootstrap sampling is a method of decorrelating the trees by exposing them to different training sets. Random-forest does both row sampling and column sampling with decision tree as a base (Figure 1).

In this case, the "Instance" represent a collection of our data-set made up of N, P, K, H, T, pH and the crop type. The algorithm now selects random rows of data and groups similar variables in the form of decision trees.

The decision tree models are then voted to get the class with the highest similar values of N, P, K, H, T, pH and crop type which then forms our tree model to be used. This tree model is then used for testing and predicting the result with

actual real time values. The number of trees formed depends on the training capacity and size of the data-set.

The chosen tree with trained data is saved in the database as seen in the table represented by **Figure 2**. The trees represent the different models for the Random Forest which are grouped into classes. Every class brings out the best outcome based on the decision tree samples. When this is done, a majority voting is now conducted to choose the best class. As you increase the number of rows for N, P, K, H, T, pH and crops, the variance will decrease. When you decrease the number of rows for N, P, K, H, T, pH and crops, variance increases. The table represented in **Figure 3** shows an extract of 20 rows from the 2200 rows of the data-set which is used in our training model. It contains sample values of N, P, K, temperature, pH and humidity of the predicted crops.

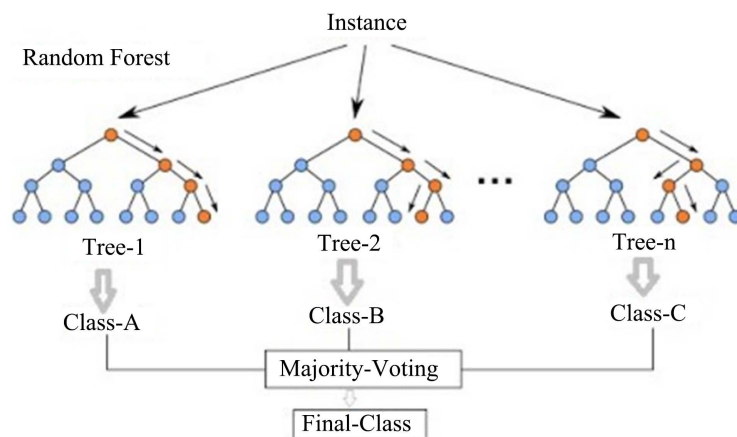


Figure 1. Random forest architecture [15].

1	N	P	K	temperature	humidity	pH	crop
2		13	72	21	24.32116642	21.0278674	5.821194486 kidneybeans
3		73	58	21	19.97215954	57.68272924	6.596060648 maize
4		17	57	21	19.88394011	20.31564139	5.789214289 kidneybeans
5		75	56	18	19.39851734	62.35750641	5.696205468 maize
6		40	73	20	21.59343016	20.31871249	5.811314232 kidneybeans
7		3	77	25	24.84906168	22.89464642	5.608165195 kidneybeans
8		88	38	15	25.08239719	65.92195844	6.455116637 maize
9		87	60	23	20.27317074	63.91281869	6.439071996 maize
10		37	74	15	24.92360104	18.22590825	5.582178402 kidneybeans
11		78	58	15	25.00933355	67.816568	6.528631266 maize
12		64	35	23	23.02038334	61.89472002	5.680361038 maize
13		31	79	25	23.18864385	22.3104551	5.902033406 kidneybeans
14		37	65	16	22.8352024	18.97267518	5.683548308 kidneybeans
15		63	58	22	18.25405352	55.28220433	6.204747653 maize
16		62	52	16	22.27526694	58.84015925	6.967057762 maize
17		27	72	23	19.92889503	21.79992115	5.961934481 kidneybeans
18		17	77	23	24.51324787	20.81527638	5.670062975 kidneybeans
19		76	39	24	24.2547451	55.64709899	6.995843776 maize
20		84	57	25	22.53510514	67.99257471	6.489040367 maize
21		67	51	24	23.50297882	61.32026065	5.584171461 maize

Figure 2. Sample data-set for crop prediction.

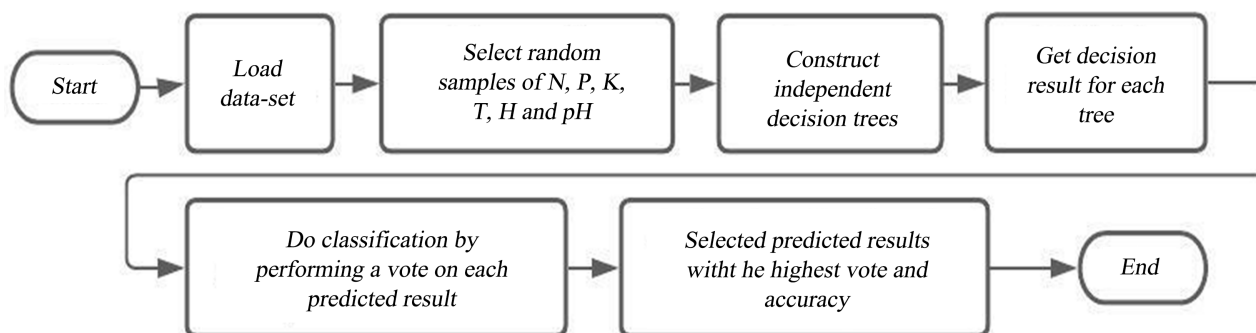


Figure 3. Flow chart for crop prediction.

The table in **Figure 3** represents just about 20 rows of the actual data-set for informational purposes. It is always important to share the data-set into the training data-set and the validation data-set. The training model will use about 80% of the data-set for training and about 20% for testing or validation of the model. **Figure 3** describes the Random Forest algorithm's application in this use case as in Equation (1).

The crop with the best accuracy is selected and visualised as seen in the architecture in **Figure 4**.

2.3. IOT Architecture

The proposed IoT architecture in **Figure 4** has four parts: 1) the input unit which is made up of the sensors (pH and NPK) and the weather forecast API (T, H and W). 2) The control unit which is made up of the microcontroller for data acquisition and wireless internet access control to the cloud. 3) The cloud unit which is made up of the database, Software control and processing of acquired data with Machine learning algorithms. 4) The display unit that will be represented as a web page on all smart devices. All units communicate over the internet.

Figure 4 shows the proposed architecture described above.

Figure 4 is made up of different sections. From the Figure, the data input is made up of sensors used to collect the values of N, P, K within a scale of 11,999 mg/kg, and pH within a scale of 3 - 9, Temperature in degree Celsius and Humidity in % gotten from the weather API. Next, is the control unit which is made up of the micro-controller and a Global System for Mobile (GSM) communication unit for internet connectivity. This unit receives data from the sensors and APIs and then sends it to the server unit using the MQTT protocol. Thirdly, the cloud server here helps in housing the database, software and Processor. All data is stored at the database in the cloud. The software section of the cloud also represents where the Random Forest algorithm is written as code and compiled by the processor. The weather API which streams data from its online repository to act as complementary inputs was used. The display is used to visualise the crop prediction and the recommendations.

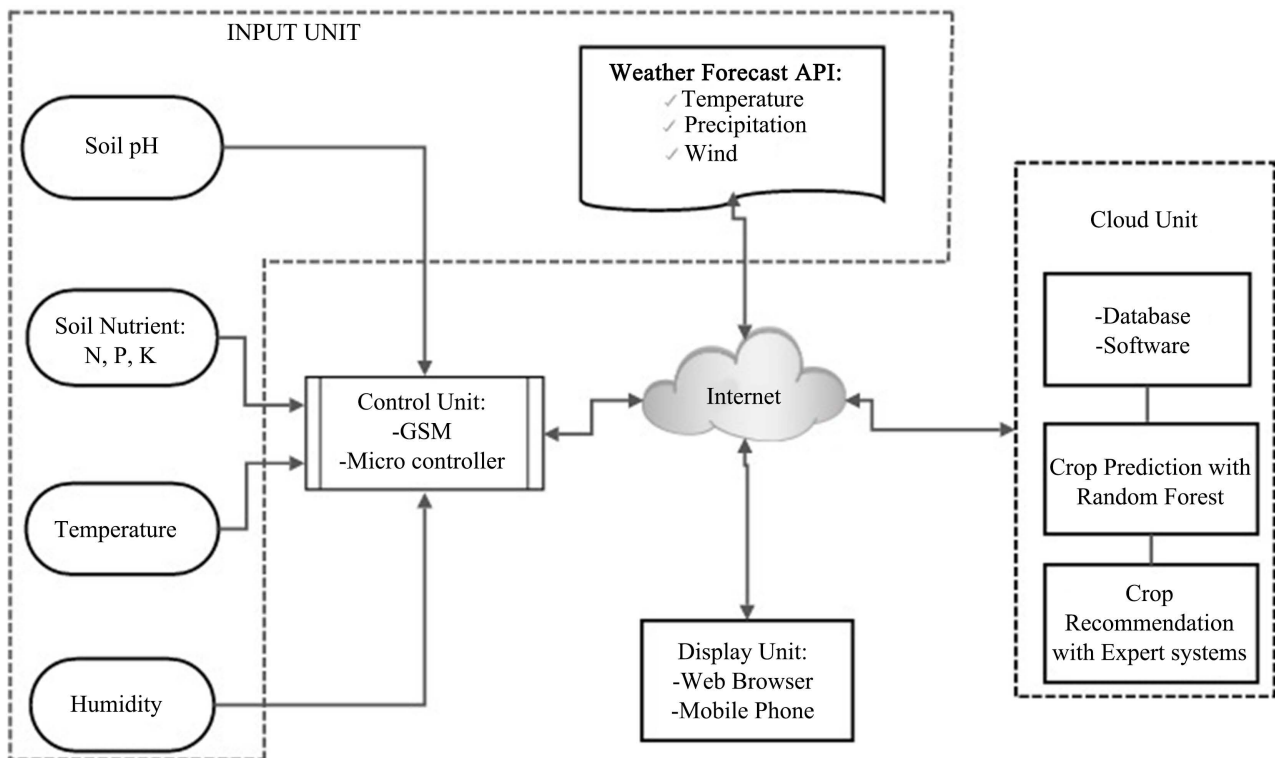


Figure 4. Crop prediction architecture based on soil fertility and weather forecast using IoT.

2.4. Crop Recommendations

The crops recommendation process takes into account future weather conditions. It is important to note that the soil can be good but if weather conditions are not taken into consideration, the planting of the predicted crop might not be helpful, so it will depend on the farmer to know if it is necessary to change the area, choose a running up crop or better still take precaution to protect the crop from the upcoming adverse weather conditions. So there are two process, first the prediction of the crop using the trained data-set and second the recommendation using weather forecast and the expert system. It is also assumed that the farmer is well aware of the planting seasons, so this system does not detect planting or cultivation seasons but tries to check harmful weather conditions to crops. **Table 2** and **Table 3** from Prabhu [18], Robert [19], Grant [20] and Togari [21], show the soil fertility conditions and normal weather conditions needed for effective crop growth in a few use cases. The information is then used to develop the expert system for the recommendations.

For the predictive model, it is assumed that all other best practices of agriculture as far cleaning the farm, having good seed and following local guidelines have been taken into consideration. **Table 2** shows the normal soil nutrients needed for the effective growth of Maize, Tomatoes, Rice and Kidney Beans and **Table 3** shows the normal weather parameters needed for the effective growth for the corresponding crops. The forecast is gotten daily over a period of seven (7) days all at once. Given that a crop is predicted using the Random Forest in

Section 2.2, the expert or knowledge system is then used as described in **Figure 5** for recommendations.

Information has been gathered from some experts, articles, publications and compiled into a knowledge database. This knowledge database is a set of possible outcomes for a large number of different scenarios of weather on each of the crops. The rules engine helps to select the appropriate knowledge from the knowledge base and present it to the farmer who is a non-expert user upon the farmer's request.

Table 2. Soil fertility conditions.

Normal soil nutrients and parameters						
<i>Crop Duration (Days)</i>	<i>N</i>	<i>P</i>	<i>K</i>	<i>Humidity</i>	<i>Temperature</i>	<i>Soil pH</i>
Maize 60 - 100	79	35	20	55% - 60%	21°C - 37°C	5.8
Tomatoes 60 - 100	120	80	80	65% - 90%	12°C - 29°C	6.2 - 6.8
Rice 120 - 140	100	45	40	60% - 80%	21°C - 35°C	5.6
Kidney Beans 60 - 100	85	39	20	60% - 80%	18°C - 26°C	5.5 - 6.2

Table 3. Normal weather information.

Normal weather parameter			
<i>Crop Duration (Days)</i>	<i>Wind</i>	<i>Temperature</i>	<i>Precipitation</i>
Maize 60 - 100	0 m/s - 16 m/s	15°C - 37°C	5 mm - 10 mm
Tomatoes 60 - 100	0 m/s - 3 m/s	12°C - 29°C	6 mm - 7 mm
Rice 120 - 140	0 m/s - 26 m/s	21°C - 37°C	6 mm - 10 mm
Kidney Beans 60 - 100	0 m/s - 26 m/s	18°C - 26°C	6 mm - 15 mm

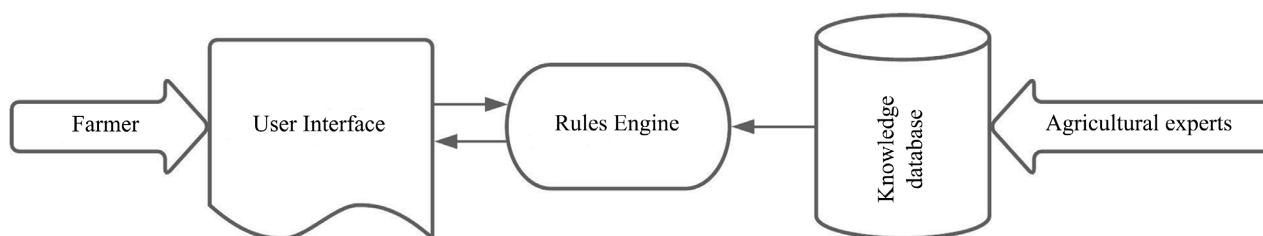


Figure 5. Recommendation expert system.

2.5. Rules Engine of the Expert System

The rules engine helps to select the appropriate knowledge from the knowledge base and present to the farmer who is a non-expert user upon the farmer's request. The farmer and user interface is represented as a display. Looking at the knowledge database found in **Table 3**, the following rules can be written.

Let x_1, x_2, \dots, x_n represent the different crops. Let w_1, w_2, \dots, w_n represent the set of weather forecast for x_1, x_2, \dots, x_n respectively with $w_1 = \{W_1, T_1, P_1\}$, $w_2 = \{W_2, T_2, P_2\}, \dots, w_n = \{W_n, T_n, P_n\}$ where $W = \text{Wind}$, $T = \text{Temperature}$ and $P = \text{Precipitation}$.

Also, let the normal range for W , T and P from the expert knowledge base be

$$\begin{aligned} & \{[W_1 \text{ min}, W_1 \text{ max}], [T_1 \text{ min}, T_1 \text{ max}], [P_1 \text{ min}, P_1 \text{ max}]\}, \\ & \{[W_2 \text{ min}, W_2 \text{ max}], [T_2 \text{ min}, T_2 \text{ max}], [P_2 \text{ min}, P_2 \text{ max}]\}, \dots, \\ & \{[W_n \text{ min}, W_n \text{ max}], [T_n \text{ min}, T_n \text{ max}], [P_n \text{ min}, P_n \text{ max}]\} \end{aligned}$$

for x_1, x_2, \dots, x_n respectively.

From the above definition, the following can be deduced:

- For any crop x_p , after successful prediction, to be recommended, the forecast weather parameter w_i should belong to the normal set for each forecast day *i.e.* $W_i \text{ min} \in [W_i \text{ min}, W_i \text{ max}]$, $T_i \text{ min} \in [T_i \text{ min}, T_i \text{ max}]$ and $P_i \text{ min} \in [P_i \text{ min}, P_i \text{ max}]$
- If any of the conditions in the item above is not satisfied but which can be remedied, the engine will have to decide based on available remedies. Some of these remedies include but not limited to irrigation, sun shading, plant covering and wind blocking.

These rules are a generalization for the recommendation of crops in the system.

2.6. Implementation

An electronic prototype has been implemented in order to get real time results for the prediction and recommendations of suitable crop as seen in **Figure 6**.

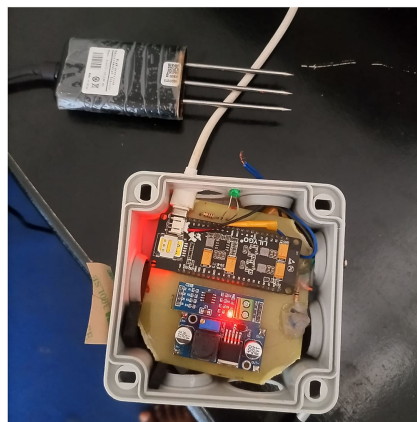


Figure 6. Physical prototype.

The prototype includes a microcontroller module, NPK and pH sensors, Battery Management System (BMS), GSM, RS485 and power. After testing the breadboard connections, a mini Printed Circuit Board (PCB) was produced to connect the parts without using overhead wires and assembled into a tiny box. The database is MySQL; the machine learning section is programmed in python with Jupiter in Anaconda and the visualization front end in anvil.

3. Experimentation and Results

The test and results are structured in two phases: Training model and Field measurements.

3.1. Training Model

During the programming of the Random Forest prediction algorithm, a series of tests was done on the jupyter notebook to see how accurate and strong the training model could be before proceeding. Over-fitting and under-fitting were tested by doing a cross validation on the data-set. The accuracy was about 99% on training data-set and about 98% on the test data-set. The result showed no over-fitting or under-fitting because of the low bias and low variance. Below are some of the major test results obtained during the training and validation of the model. All graphs were generated using jupyter notebook-Density graphs:

Density graphs are plotted as function of density vs the variable under consideration. The graphs basically show the concentration of the variables (N, P, K, Temp, Hum and pH) over the entire data-set. The figures below show the captured density graphs. **Figure 7** shows the density concentration of Nitrogen and Phosphorous, Potassium, pH, Temperature, and Humidity within the maximum and minimum values of their respective ranges.

A series of tests were done on jupyter notebook during the Random Forest prediction algorithm programming to see how accurate and strong the training model and data-set could be the overall training accuracy was 99%. An important statistical results amongst others was the soil nutrient ratio comparison ans seen in **Figure 8**.

3.2. Field Measurements

After the data was trained, code compiled and loaded to the micro-controller and with the anvil server running, tests were then manually made on the field to confirm the prototype. For a more practical result, two different locations in Nomayos were considered. The procedure outlined in the methodology on how to use the sensors was followed. **Figure 9** shows some pictures taken in the field during the tests.

For a more accurate measurement, the entire field was covered by taking the average value of the values of the N, P and K of the soil at five different positions at the study site. To be sure the user understands what they are doing, a button and a visual indicator on the prototype was used. When the N, P, and K probes

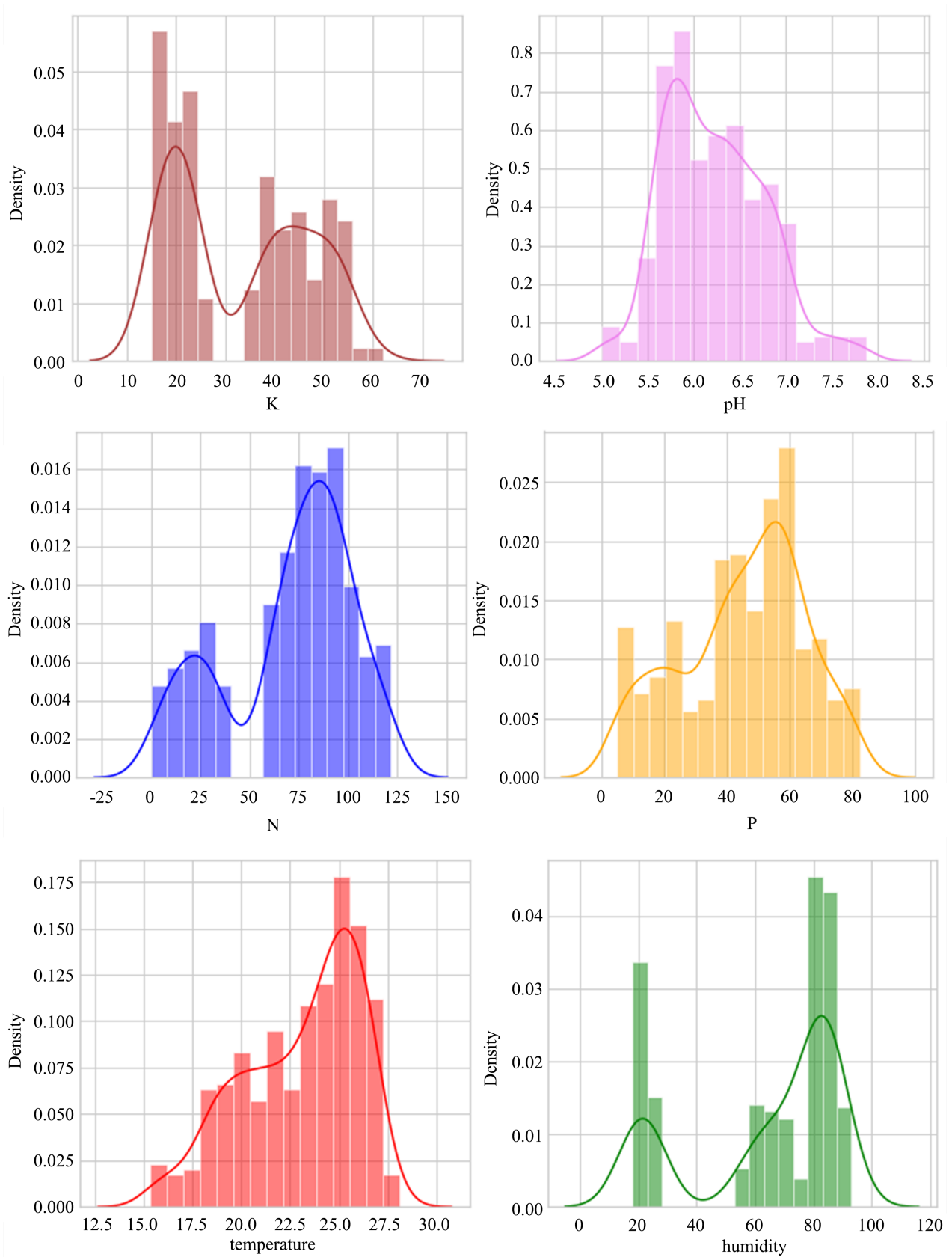


Figure 7. Density concentration of K, pH, N, P, T and H.

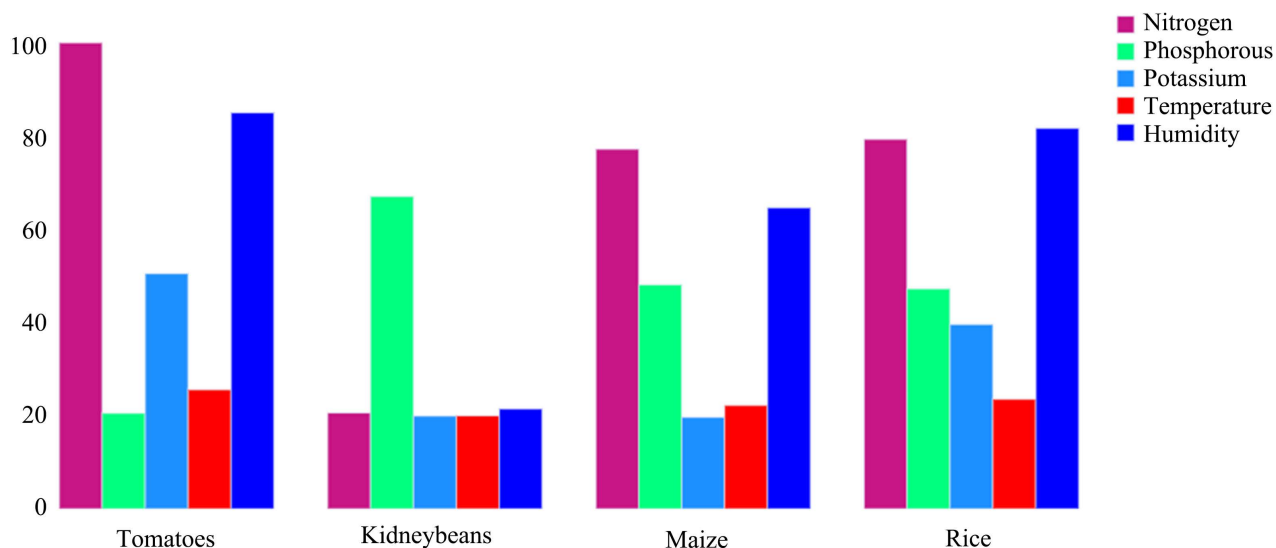


Figure 8. N, P, K, T, and H ratio comparison between crops.



Figure 9. Nomayos field test.

were inserted into the ground, the button was pressed once in order to get the value of the N, P, K, temperature, humidity and pH. The blinking light on the device actually indicates that the values are being collected and stored in the local memory of the micro-controller. This process is repeated for four other positions on the field and the 5th time, the values are then averaged and sent to the database on the server indicated by a steady light, which goes back off once the data transmission is complete. **Table 4** shows the results of the predicted crops collected at the 2 locations.

Table 4 shows the values obtained during the measurements taken in the field at Nomayos with the corresponding predicted crops.

Table 5 shows the weather forecast of temperature, wind and precipitation for seven days. These values are used in comparison with the standard values in order to make helpful recommendations to the farmer. A sample of the recommendations gotten for each of the forecast days for Rice and Beans both tested at Nomayos swamps and solid soil is seen in **Table 6**.

Table 4. Soil fertility prediction parameters for crops.

		LOCATION	
		Nomayos swamps	Nomayos solid ground
Soil parameters	N	90	80
	P	40	43
	K	36	16
	Temperature	32.4	32.6
	Humidity	29	29
	pH	6	6
Predicted cop		Rice	Kidney Beans

Table 5. Weather forecast for recommendations.

Predicted crop	Rice							Kidney beans						
Location	Nomayos swamps							Nomayos solid ground						
Day	1	2	3	4	5	6	7	1	2	3	4	5	6	6
Temp (°C)	31.4	32.05	32.51	33.08	30.59	31.23	32.75	31.34	32.05	32.51	33.08	30.59	31.23	32.75
Wind (Km/h)	1.95	2.47	2.52	1.81	2.03	2.15	2.57	1.95	2.47	2.52	1.81	2.03	2.15	2.57
Precipitation (mm/h)	na	na	na	3.05	1.64	2.05	3.43	na	na	na	3.05	1.64	2.05	3.43

Table 6. Recommendations based on weather forecast.

Predicted Crop	Date	temperature	Wind	Precipitation
Rice	05-01-2022 12:00:00	Within required values	Within the required range	Little or no rain, consider irrigation
	06-01-2022 12:00:00	Within required values	Within the required range	Little or no rain, consider irrigation
	07-01-2022 12:00:00	Within required values	Within the required range	Little or no rain, consider irrigation
	08-01-2022 12:00:00	Within required values	Within the required range	Little or no rain, consider irrigation
	09-01-2022 12:00:00	Within required values	Within the required range	Little or no rain, consider irrigation
	10-01-2022 12:00:00	Within required values	Within the required range	Little or no rain, consider irrigation
	11-01-2022 12:00:00	Within required values	Within the required range	Little or no rain, consider irrigation

Continued

	05-01-2022 12:00:00	Above normal, consider another option	Within the required range	Little or no rain, consider irrigation
	06-01-2022 12:00:00	Above normal, consider another option	Within the required range	Little or no rain, consider irrigation
	07-01-2022 12:00:00	Above normal, consider another option	Within the required range	Little or no rain, consider irrigation
Kidney Beans	08-01-2022 12:00:00	Above normal, consider another option	Within the required range	Little or no rain, consider irrigation
	09-01-2022 12:00:00	Above normal, consider another option	Within the required range	Little or no rain, consider irrigation
	10-01-2022 12:00:00	Above normal, consider another option	Within the required range	Little or no rain, consider irrigation
	11-01-2022 12:00:00	Above normal, consider another option	Within the required range	Little or no rain, consider irrigation

4. Discussion

In accordance with this research question and objectives, the results, as expressed in Section 3, demonstrate that the Internet of Things (IoT) is useful in the prediction of the best crops that can be grown on cultivated or uncultivated soils without the use of fertilizers directly. Furthermore, it has been demonstrated that crop evolution recommendations made for the predicted crop can be used to assist in the selection of the most suitable crop. Because of this, weather forecasts are critical in the selection of the most suitable crops for planting at the appropriate time of the year.

Figure 7 and **Figure 8** show some major statistical results which helps in the pre-validation of the Random Forest algorithm used for predictions. Each of these results is further discussed in the following sections.

4.1. Density Graphs

The graphs in 7 show the density distribution of the respective soil parameters. They give firsthand information to the useful range of the data-set for each parameter. The x-axis represents the parameter under control. In this case, N, P, K, pH, Temperature and Humidity are used. The y axis represents the probability density of the variable. It actually shows the areas of concentration of the parameter. For example, if the first two diagrams of **Figure 7** are considered, it will be noticed that for Potassium, the data is found within two intervals, from about 15 to 28 and from 35 to 62. This definitely means that any data during prediction that lies in the empty interval which is from 28 to 35 does not directly fit in the data-set, but can be approximated by the training algorithm during crop prediction. These graphs are important because it makes us see where we can easily expect our values. Each of the graphs in 7 are differentiated by colors just for clarity but all represent density graphs for N, P, K, Temp, Humidity and pH.

4.2. Soil nutrient Ratio Comparison

Figure 8 shows a distinct view of the average required of the prediction parameters of N, P, K, T, H and pH. The x-axis shows the crops under prediction and the y-axis shows the amount or quantity of the needed soil parameter. By looking at the Figure, one easily sees that the quantities or ratios needed for one crop varies with respect to the other individual parameters. For example, if just maize and rice is considered, their prediction would be the same in case only the Nitrogen content of the soil was considered. However, as Phosphorous, Potassium, temp, humidity and pH is increased, a clear and distinct change becomes noticeable. From this graph we can easily conclude that the soil parameters needed for the growth of some crops can at times be similar and if we are to limit to them, the results would most of the time be erroneous. In our case, we chose Rice, Maize, Kidney Beans and Tomatoes as use cases for our tests. The prediction of crops was made using the Random Forest algorithm with a training accuracy of about 99%, which is better than the Q-learning algorithm presented in Section 1. This accuracy can be seen as an improvement to crop prediction. Other machine learning algorithms were also evaluated like the SVM, K-Nearest Neighbors and Decision tree whose accuracy were significantly lower: 89%, 86% and 49% respectively.

4.3. Field Results

The Section 3.2 shows the results gotten during the testing of the prototype. Considering the research question, which was based on how to use IoT in the prediction of best crop and in the recommendations of best practices using machine learning algorithms, field results can be analysed. In Section 1, an extreme learning machine learning algorithm was also used which by default is a feed forward neural network, supposedly better than the supervised Random Forest used. The algorithm presents an accuracy of about 90% which is smaller than our obtained accuracy of about 99%. This gave a conclusion that even the best methods are not always suitable for some problems. In this same section, a similar tool using IoT technology and rechargeable batteries was developed but it was too cumbersome to use or to even easily get good results because of the way it was manipulated (dragged on animals or trolleys). Our IoT solution proved to be far smaller and portable which permits the measurement of any soil at any time. The results presented in **Table 4**, show the values of N, P, K, Temperature, Humidity and pH obtained from Nomayos for both for a swampy and solid ground. Upon insertion in any soil, the user presses a button on the device to collect the data of that location. The data of five such locations on the same field or area under consideration is collected. The data sent to the server is an average of the five different positions collected. This helps reduce any impurities that might occur during the prediction phase. It further predicts that Rice can easily grow in swampy areas while beans can easily grow in solid grounds of that area. It equally shows that the major difference is between the values of Nitrogen and

Potassium (10 and 20 mg/kg) respectively. These results were not only validated by the data-set but also from the fact that inhabitants of this area grow these crops and they do well with little or no fertilizer. Crop prediction has been the major achievement in most articles cited in the literature without consideration of future weather parameters. In this study, it was found out that, a soil can be truly fertile for a particular crop and also truly predicted but not really suitable for cultivation because of future weather disturbance, which can be remedied by a recommendation. **Table 5** shows weather forecasts for Temperature, Wind speed and precipitation 7 days into the future. These values are used to examine how the predicted crop can be affected by weather forecast. It was actually noticed that, though Kidney Beans is suitable for the solid grounds at Nomayos, it might not actually do well because the future temperature at the time of measurement was higher than the normally required temperature for beans growth.

Based on the period and season in which the results were gotten, we noticed a calm nature of wind and little or no precipitation. Using our expert system and collected data from experts, we were able to make the recommendations found in **Table 6**. The table shows the predicted crops and the forecast dates from when the prediction was made. The recommendations are actually made with respect to temperature, wind speed, and precipitation which are the major factors sometimes identified to have a negative impact on crops as far as weather conditions are concerned. These results can be refreshed on a daily basis to continuously see updated recommendations 7 days at any point in time without necessarily going back to the field.

Nevertheless, we were able to note some of the following limitations:

- The scope of this work can only permit a forecast of up to 7 days which at times might be short to take a drastic decision in the event of a drastic weather change.
- Finding some very critical parts for the IoT prototype was not very easy making development time actually longer than anticipated.

It was not very easy to get many local agricultural experts. This actually increased the difficulty of finding relevant data, especially during predictions.

5. Conclusions

The purpose of this paper was to examine the usefulness of weather forecast on crop prediction using IoT, machine learning and weather APIs in order for farmers to be able to make better crop selection decisions.

Several existing methods and architectures were evaluated before developing the architecture presented in this study. This allowed us to visualize the results of the predictions and recommendations from any location and at any time. In order to accomplish this, an IoT device prototype was created which communicates with the cloud via the MQTT protocol over the internet. This device uses interconnected sensors to collect the data it requires, both physically (for N, P, K, and pH collection) and virtually (using APIs) (for T, H, P and W).

This was followed by the application of the Random Forest algorithm to our data set, which proved to be accurate (with a 99% for the trained algorithm). Approximately 80% and 20% of the data set were used to train and test the data-set respectively. The expert or knowledge systems were used for recommendations. The implementation of these algorithms was done on the cloud jupyter in Anaconda and anvil (drag and drop we app builder). Prediction of the best crops and recommendations were made. A key aspect of this work was the addition of a weather forecast for prediction, which was actually seen in the results section to sometimes have an impact on the predicted crop.

At the end of the work, real tests were carried out on the field for both swampy and solid ground and the results actually met the objectives initially defined in the introduction.

Considering the limitations of our work we plan to increase the weather forecast period to the time representing the complete life cycle of the predicted crop. Also, increasing the number of crops represented in the data set coming from local communities and giving users the ability to add their own crops based on their geographic location would be a great improvement.

Acknowledgements

We would like to thank Dr. Bernadette Dia Kamgnia, Dr. Emmanuel Kamgnia and Dr. Bertrand Fesuh who made themselves available to review this document. Special thanks go to Agronomist Nkiambouh Lengouh Takuh who contributed to the experimental phase of this project.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Neumark, S.D. (1959) Some Economic Development Problems of African Agriculture. *Journal of Farm Economics*, **41**, 43-50. <https://doi.org/10.2307/1235197>
- [2] Food and Agriculture Organization (2022) Soils for Nutrition: State of the Art.
- [3] Pittenger, D. (2021) Crop Nutrient Requirement, University of California. https://ucanr.edu/sites/Nutrient_Management_Solutions/stateofscience/Meet_Crop_Nutrient_Requirements/
- [4] Shang, Q.Y., Ling, N., Feng, X.M. Yang, X.X., Wu, P.P., Zou, J.W., Shen, Q.R. and Guo, S.W. (2014) Soil Fertility and Its Significance to Crop Productivity and Sustainability in Typical Agroecosystem: A Summary of Long-Term Fertilizer Experiments in China. *Plant and Soil*, **381**, 13-23. <https://doi.org/10.1007/s11104-014-2089-6>
- [5] Zhao, J., Zhang, J., Feng, Y. and Guo, J.X. (2010) The Study and Application of the IOT Technology in Agriculture. 2010 *3rd International Conference on Computer Science and Information Technology*, Chengdu, 9-11 July 2010, 462-465.
- [6] Madakam, S., Ramaswamy, R. and Tripathi, S. (2015) Internet of Things (IoT): A

- Literature Review. *Journal of Computer and Communications*, **3**, 164-173.
<https://doi.org/10.4236/jcc.2015.35021>
- [7] Verdouw, C., Sundmaeker, H., Tekinerdogan, B., Conzon, D. and Montanaro, T. (2019) Architecture Framework of IoT-Based Food and Farm Systems: A Multiple Case Study. *Computers and Electronics in Agriculture*, **165**, Article ID: 104939.
<https://doi.org/10.1016/j.compag.2019.104939>
- [8] Muangprathub, J., Boonnam, N., Kajornkasirata, S., Lekbangpong, N., Wanichsombat, A. and Nillaor, P. (2019) IoT and Agriculture Data Analysis for Smart Farm. *Computers and Electronics in Agriculture*, **156**, 467-474.
<https://doi.org/10.1016/j.compag.2018.12.011>
- [9] Thorat, A., Kumari, S. and Valakunde, N.D. (2017) An IoT Based Smart Solution for Leaf Disease Detection. 2017 *International Conference on Big Data, IoT and Data Science (BIG DATA)*, Pune, 20-22 December 2017, 193-198.
<https://doi.org/10.1109/BIGDATA.2017.8336597>
- [10] Katarya, R., Raturi, A., Mehndiratta, A. and Thapper, A. (2020) Impact of Machine Learning Techniques in Precision Agriculture. 2020 *3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, Jaipur, 7-8 February 2020, 1-6
<https://doi.org/10.1109/BIGDATA.2017.8336597>
- [11] Patil, B., Maske, R., Nakhate, R., Nikam, R. and Javheri, P.S. (2020) Soil Fertility Detection and Plant Prediction Using IoT and Machine Learning Approach. *International Journal of Innovative Research in Science, Engineering and Technology*, **9**, 3329-3334.
- [12] Suchithra, M.S. and Pai, M.L. (2020) Improving the Prediction Accuracy of Soil Nutrient Classification by Optimizing Extreme Learning Machine Parameters. *Information Processing in Agriculture*, **7**, 72-82.
<https://doi.org/10.1016/j.inpa.2019.05.003>
- [13] Girish, L., Gangadhar, S., Bharath, T.R., Balaji, K.S. and Abhishek, K.T. (2018) Crop Yield and Rainfall Prediction in Tumakuru District Using Machine Learning. *National Conference on Technology for Rural Development*, 61-65.
- [14] Nischitha, K. and Dhanush, V. (2020) Crop Prediction Using Machine Learning Approaches. *International Journal of Engineering Research and Technology*, **9**, 23-26.
- [15] Ingle, A. (2021) Crop Recommendation Dataset, Maximize Agricultural Yield by Recommending Appropriate Crops.
<https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>
- [16] Anakha, V., Aparna, S., Jinsu, M., Rima, M. and Vinu, W. (2021) Crop Yield Prediction Using Machine Learning Algorithms. *International Journal of Engineering Research and Technology*, **9**, 117-123.
- [17] Singh, R. (2021) Mathematics behind Random Forest and XGBoost.
<https://medium.com/analytics-vidhya/mathematics-behind-random-forest-and-xgboost-ea8596657275/>
- [18] Prabhu, S., Revandekar, P., Shirdhankar, S. and Paygude, S. (2020) Soil Analysis and Crop Prediction. *International Journal of Scientific Research in Science and Technology*, **7**, 117-123. <https://doi.org/10.32628/IJSRST207433>
- [19] Precheur, R., Greig, J.K. and Armbrust, D.V. (2021) The Effects of Wind and Wind-plus Sand on Tomato Plants. *Journal of the American Society for Horticultural Science*, **103**, 351-355. <https://doi.org/10.21273/JASHS.103.3.351>
- [20] Grant, A. (2021) Learn How to Grow Kidney Beans.

<https://www.gardeningknowhow.com/edible/vegetables/beans/how-to-grow-kidney-beans.htm>

- [21] Togari, Y. (1940) Damages to Rice Crops Caused by a Wind Storm, with Special Reference to a Typhoon That Visited Kansai District, Japan. Part I. Local Differences in the Damage Caused and a Discussion on the Cause of Damage from Both the Meteorological and Geographical Points of View. *Japanese Journal of Crop Science*, **12**, 250-264. <https://doi.org/10.1626/jcs.12.250>