Ph.D Proposal Seminar

on

An Intelligent IoT-Driven Maize Crop Farming Support System

By

TOSIN COMFORT OLAYINKA (Mrs.)

CSC/19/3569

Supervisor: Prof. Adetunmbi A.O

Co-Supervisors: Prof. Obe O.O

Dr. Ibam E.O

Department of Computer Science

Federal University of Technology, Akure, Ondo State

February, 2023

OUTLINE

- Introduction
- Internet of Things (IoT)
- Machine Learning
- Maize
- Objectives of Study
- Motivation
- Literature Review
- Methodology
- Contribution to Knowledge
- References

INTRODUCTION

- Rapid development in Information and Communication Technology (ICT) over the last few decades has led to new frontiers in computing and computational science world.
- Recent trends include Artificial Intelligence, Machine Learning, Reinforcement Learning, Deep learning, Big Data and Internet of Things (IoT) among others.

INTRODUCTION

- Since the coining of the word IoT in 1999 by Kevin Ashton, IoT has attracted enormous attention from researchers across the globe because of its diverse applications in several areas of human endeavors.
- The Internet of Things (IoT) or Internet of Everything is seen as the third technology wave (Yacob, 2020). The Internet of Things (IoT) is a unique concept which merges “Internet” and “Things”.
- IoT can be viewed as a worldwide network that allows numerous things, including humans and things, to communicate with one another.

INTERNET OF THINGS (IoT)

- IoT can be defined as the collection of many devices or things that communicate, detect, and interact with their interior and exterior states using embedded technology (Lee & Lee, 2015).
- IoT enables unintelligent objects to become intelligent by exploring the power of the internet.
- IoT with machine learning and artificial intelligence have led to the development of smart cars, smart farms, smart cities, smart homes and smart environments among others (Prathibha, 2017; Farooq et al., 2019; George et al., 2019).
- IoT assist farmers in agricultural processes in order to improve the quality and quantity of agricultural products.

INTERNET OF THINGS (IoT)

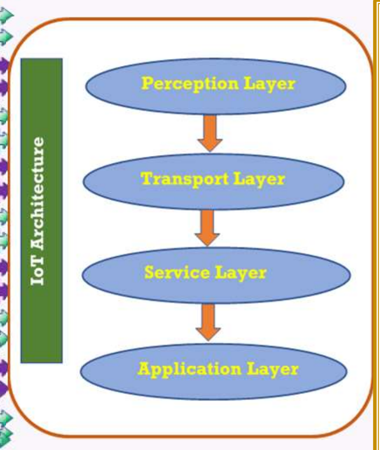

Fig. 1: IoT Architecture

Perception layer of IoT includes sensor motes that serve as the building blocks of sensor technology e.g. cameras, RFID (Di, 2018).

Transport layer offers network interconnectivity and transport capabilities to communicate objects effectively.

Service layer encompasses the storage of data, visualization of data, and processing of data resources.

Application layer is capable of providing intelligence for definite control tasks using virtual objects .

MACHINE LEARNING

- Machine learning (ML) is one of the most stimulating recent technologies in Artificial Intelligence. ML is used to design algorithms based on data trends and historical relationships between data.
- The primary aim of machine learning is to allow computers to learn automatically without human intervention or assistance and adjust actions accordingly.
- ML have been used in several areas like; network intrusion detection, disease prediction, universal learning system, currency exchange forecasting, cyber-crime monitoring, smart deforestation monitoring and control, e.t.c. (Adetunmbi et al. 2021, Adewale et al., 2022, Obe et al. 2022, Alese et al. 2022, Ibam & Olowokere, 2021 and Ojokoh et al. 2022)

MACHINE LEARNING

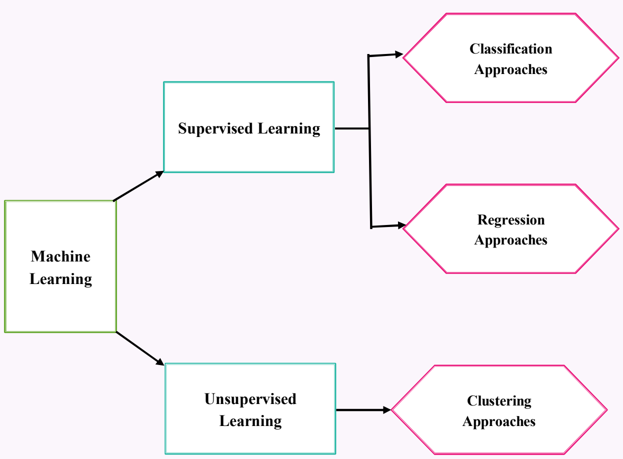

Fig. 2: Types of Machine Learning

For the purpose of this research, supervised machine learning will be used, under the classification approach.

The classification algorithms for the maize yield prediction are Artificial Neural Network, (ANN), Naïve Bayes (NB) and K-nearest Neighbor (KNN), Decision Tree (DT), Apriori and CNN-RNN algorithms.

MAIZE

- Maize or corn (Zea mays L.) is a plant belonging to the family of grasses called Poaceae. Maize is a cereal crop cultivated globally.
- Maize serves as a staple human food, feed for livestock and raw material for many industrial products.
- Maize is the world's highest supplier of calorie (19.5%) compared to rice (16.5%) and wheat (15.0%).
- Maize crop is mostly cultivated during raining season because of the need of water for the plant growth.
- With IoT there is possibility of automated irrigation such that maize crop can be planted all through the year, to increase food supply in our country.

MAIZE

- The world population has been projected to reach 9.8 billion from the current 7.7 billion by the year 2050 and 11.2 billion in 2100.
- The population of sub-Saharan Africa is predicted to increase by 99% by the year 2050 (UN, 2019).
- With this rapidly growing global population especially in the sub–Sahara Africa where Nigeria's population is dominant, there is an urgent need for a commensurate increase in food supply to meet the food demand of the ever-increasing population.

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/ Future work
Shakoor et al. (2017). Agricultural Production Output Prediction Using Supervised Machine Learning Techniques.	Provide a learning agent that can aid in taking decisions.	Decision Tree and K-Nearest Neighbour Regression (KNNR) algorithms.	Discover the patterns in the dataset which contains average temperature and rainfall value.	Unable to predict close values due to inconsistencies in the prediction.
Jahan (2018). Applying Naive Bayes Classification Technique for Classification of Improved Agricultural Land soils	To predict soil types by their characteristics and fertility, via classification technique .	ZeroR, Stacking and Naïve Bayes algorithms.	Naive bayes classifier generates the good performance based on its accuracy for this soil data set compared to zeroR and stacking.	Other researchers can add other soil properties.

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Balducci et al. (2018). Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement.	Aims to show how to manage heterogeneous information and data coming from real datasets that collect physical, biological, and sensory values.	Decision trees, K-nearest neighbors, neural networks, and polynomial predictive models were used.	The decision tree model works very well, compared to others.	Smart systems that provide real-time suggestions and make long-term forecasts based on user choices and preferences must be studied and tested.
Priya et al. (2018). Predicting yield of the crop using machine learning Algorithm.	Focused on predicting the yield of the crop based on the existing data by using Random Forest algorithm..	Random forest algorithm.	Random Forest algorithm achieves a largest number of crop yield models.	Determine the efficient algorithm based on their accuracy metrics that will helps to choose an efficient algorithm for crop yield prediction.

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Vincent et al. (2019). Sensors Driven AI-Based Agriculture Recommendation Model for Assessing Land Suitability.	Develop a model to automate assessment of present and future conditions for the better crop yield.	Neural networks and Multi-Layer Perceptron (MLP) algorithms.	The system was able to make four (4) decision classes, namely more suitable, suitable, moderately suitable, and unsuitable.	Multiclass classification can be done.
Bhanumathi et al. (2019). Crop Yield Prediction and Efficient use of Fertilizers.	To predict the most efficient model and predict the output of the crop.	Random Forest algorithm and Backpropagation algorithms for prediction for different crops in various regions.	Lesser error rate for backpropagation compared to random forest.	To develop web application that can help farmer to understand the crop yield.

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Gopal and Bhargavi (2019). A novel approach for efficient crop yield prediction.	To develop an hybrid MLR-ANN model	Hybrid MLR-ANN algorithm.	Hybrid MLR-ANN model gives better prediction accuracy compared to conventional MLR and ANN model and other models.	Other related datasets can be added to agricultural and weather data on the hybrid MLR-ANN model.
Chaganti et al. (2019). Prediction Based Smart Farming.	Focused mainly on prediction analysis and image processing techniques that could be utilized in the field of Precision Farming.	Satellite Imagery based Machine Learning, Image Processing, and the Internet of things .	Predictive investigation service was utilized to recommend the main three most appropriate yields dependent on the nourishment levels of the Soil, temperature, and so on.	A mechanical based robotized vehicles can be developed for future improvement .

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Dewi and Chen (2020). Decision Making Based on IoT Data Collection for Precision Agriculture.	Proposes IoT for local information data collection for precision agriculture.	Decision Tree Learning-ID3 and K Nearest Neighbors Regression algorithms.	Data collected are water level, temperature, humidity, soil moisture, and light from various crop (Potato, Onion, Sugar Beet, Tomato, Hay, Citrus, Grape, and Sugarcane) .	Monitoring farm both on website and mobile phones should be developed.
Shahhosseini et al. (2020). Forecasting Corn Yield with Machine Learning Ensembles	To design a machine learning based framework to forecast corn yield using weather, soil, plant population, and planting date data.	Linear regression, LASSO regression, Extreme Gradient Boosting (XGBoost), LightGBM, and random forest.	Ensemble models perform better compared to linear and LASSO regression.	Addition of other features such as forecasted weather data and N-fertilization inputs by county can improve the model performance.

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Zhang et al. (2020). Combining Optical, Fluorescence, Thermal Satellite, and Environmental Data to Predict County-Level Maize Yield in China Using Machine Learning Approaches.	Integrated optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield.	Regression-based method (LASSO), two machine learning (ML) methods (RF and XGBoost), and deep learning (DL) network (LSTM)..	ML and DL methods evidently outperformed traditional regression models.	Need to integrate multi-spectral satellite data and environmental variables for predicting crop yield.
Ramos et al. (2020). A random forest ranking approach to predict maize yield using only multispectral UAV-imagery.	Proposed a ranking-based approach to potentialize the Random Forest (RF) method for maize yield prediction.	Random forest algorithm.	RF algorithm performed better in all of the configuration scenarios.	Adopted in future research to evaluate different types of crop yield

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/ Future work
Saleh et al. (2020). Determination of Corn Quality using the Decision Tree of C 4.5 Algorithm.	The research aimed to test several methods/algorithms of data mining (Naive Bayes and Decision Tree (C4.5)).	Naïve Bayes and Decision Tree (C4.5) algorithms	Decision Tree method (C 4.5) in the process of determining the quality standard of corn is more effective than the naïve Bayes method.	Other algorithms can be used to compare with decision tree.
Mwaura and Kendulywo (2020). County level maize yield estimation using artificial neural network.	*To use satellite data and ANN's model to predict maize yields.*	*ANNs of multilayer perceptron, feed forward back propagation to predict maize yield.*	*ANNs results are better compared with ordinary Regression.*	*To integrate physical and management factors for maize yield prediction..*

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/ Future work
Bouras et al. (2021) Cereal yield forecasting with satellite drought-based indices, weather data, and regional climate indices using machine learning in Morocco.	*To investigate the potential of machine learning for developing dynamic decision support systems for cereal production*	*combined satellite-based drought indices, weather and climate data. Multiple Linear Regression, MLR), and non-linear (Support Vector Machine, (SVM), Random Forest (RF), and eXtreme Gradient Boost (XGBoost) algorithms.).*	*Combining data from multiple sources outperformed models based on one dataset only. The XGBoost method exhibited the best metrics.* .	*Other related factors, such as planting date, soil properties, local climate conditions and physical properties can be considered.*
Rafi et al. (2021). IoT and Machine Learning based Smart Agri -Farming System.	Proposed a smart farm irrigation system using IoT and Machine Learning technique algorithms.	IoT and Machine learning algorithms (such as SVR, and SVM with Radial basis function kernel were used.	Inexpensive approach towards automating the agriculture industry	To implement a mobile system.

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Glória et al. (2021). Sustainable Irrigation System for Farming Supported by Machine Learning and Real-Time Sensor Data.	Developed an automatic irrigation control system for agricultural fields.	Decision Trees, Random Forest, Neural Networks, and Support Vectors Machines.	A mobile application was developed for real-time data collection and analyses the data.	The mobile application can be enhanced.
Suruliandi et al. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques.	*To find the best feature selection technique, with a classification method, to predict the most suitable crop for cultivation, based on factors such as soil and environment..*	*Wrapper feature selection techniques used are RFE, Boruta, and Sequential Forward Feature Selection (SFFS). K-Nearest Neighbour, Naïve Bayes, Decision Tree, SVM, Random Forest and Bagging*	*Boruta, Sequential Forward Feature Selection (SFFS), and RFE feature selection techniques with the bagging classifier performing best.*	*Other parameters can be added to the soil and environmental data for robust classification.*

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Chen et al. (2021). Prediction of Maize Yield at the City Level in China Using Multi-Source Data	*Combined climate, satellite data, and meteorological indices to forecast maize yield using four machine learning methods.*	*Cubist, random forest (RF), extreme gradient boosting (Xgboost), and support vector machine (SVM) algorithms.*	*The climate data or satellite data inputs from all growth stages were essential for maize yield prediction. Cubist algorithm has best performance.*	*Research could be done for other crop types and other regions.*
Moraye et al (2021). Crop Yield Prediction Using Random Forest Algorithm for Major Cities in Maharashtra State.	Focused on predicting crop yield using a different machine learning algorithm and to build a user-friendly web application to predict the crop yield based on the factor of climate change.	Random forest algorithm.	Random Forest algorithm gave high accuracy and best prediction. Smart Farm application was developed for crop prediction.	Other factors like quality of soil, pest, chemicals were not used.

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Sarijaloo et al. (2021). Yield performance estimation of corn hybrids using machine learning algorithms.	Aimed to use different machine learning algorithms to predict the yield of corn hybrids	Decision tree, gradient boosting machine, random forest, adaptive boosting, XGBoost and neural network algorithms.	XGBoost is more accurate than other models .	Use novel clustering approaches that might be more informative for yield prediction.
Nyéki et al. (2021). Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods.	*Focused to predict maize yields using AI that uses spatio-temporal training data.*	*CP-ANN, XY-F, SKN, XGBoost and SVM algorithms.*	*XGBoost was the best prediction method.*	*Physical parameters should be added to the soil, meteorological, satellite data for robust prediction.*

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Gosai et al (2021). Crop Recommendation System using Machine Learning.	Aims to recommend the most suitable crop based on input parameters like Nitrogen (N), Phosphorous (P), Potassium (K), PH value of soil, Humidity, Temperature, and Rainfall.	Decision Tree, Naïve Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and XGBoost algorithms.	Implemented an intelligent crop recommendation system, which can be easily used by farmers.	To improve dataset with larger number of attributes and build website with mobile app for easy to use. .
Guo, et al. (2021). Integrated phenology and climate in rice yields prediction using machine learning methods.	*Integrated site-based phenology, climate, preseason, geography and yields data to predict rice yield using machine learning techniques.*	*Multiple linear regression (MLR), backpropagation neural network model (BP), support vector machine (SVM) and random forest (RF).*	*The ML methods performed better than the MLR method, and the applied SVM and RF were better than the BP method.*	*Assembled data of all phenology, climate, preseason and GEOI should be used for similar crops.*

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Oikonomidis et al. (2022). Hybrid Deep Learning-based Models for Crop Yield Prediction.	To investigate the power of XGBoost and hybrid CNN-DNN models to build crop yield prediction models.	XGBoost algorithm, Convolutional Neural Networks, Deep Neural Networks, CNN-XGBoost, CNN-Recurrent Neural Networks (RNN), and CNN-Long Short Term Memory (LSTM). Weather, soil and farm management parameters were used.	The hybrid CNN-DNN model outperforms other models.	Datasets other than that soybean crop should be used for modelling.
Attia et al. (2022). Coupling Process-Based Models and Machine Learning Algorithms for Predicting Yield and Evapotranspiration of Maize in Arid Environments.	Combining crop models (CMs) with different ML algorithms to create a hybrid approach with robust predictions.	Multiple linear regression (MLR), backpropagation neural network model (BP), support vector machine (SVM) and random forest (RF) algorithms.	The Ensemble model (super learner) and XGBoost outperform other models in predicting Grain Yield and Evapotranspiration for maize.	Expanding the current approach to include crop models–machine learning–deep learning as a hybrid under future climate change

LITERATURE REVIEW

Authors/Title	Objective(s)	Methodology	Contributions	Limitations/Future work
Adesanya and Yinka-Banjo (2022). Classification of Nitrogen Deficiency for Maize Plants Using deep learning algorithms on Low-End Android Smartphones.	To develop a mobile app for low-end android phones that use a machine learning model to detect nitrogen insufficiency.	Single Shot Detector (SSD) Mobilenet model.	The Single Shot Detector (SSD) Mobilenet model generated eight-one (81) percent accuracy.	Get more robust datasets. .
Haque et al. (2022). Deep learning-based approach for identification of diseases of maize crop.	Proposed a deep convolutional neural network-based approach to automatically identify digital images of diseases.	A deep convolutional neural network-based algorithm.	Ensemble model (super learner) and XGBoost outperform other models in predicting Grain Yield and Evapotranspiration for maize.	The proposed model will be integrated with a mobile application for real-time disease identification tool.

MOTIVATION OF STUDY

- Having reviewed recent literatures both on the internet of things and machine learning related to maize crop yield prediction. The following are the summaries of the limitations in the study reported in (Mwaura and Kendulywo, 2020, Bouras et al., 2021, Suruliandi et al., 2021, Chen et al., 2021, Nyeki et al., 2021, Guo et al., 2021, Attia et al., 2022, Adesanya and Yinka-Banjo, 2022 and Haque et al., 2022).

i. The datasets were secondary datasets collected from agricultural data bank which consists of soil data, weather data, environmental data and some with satellite data.

ii. ANN model were compared with ordinary and multiple linear regression, which is not robust enough.

iii. Hybrid deep learning was not considered.

This research is therefore motivated by the need to address some these issues for robust maize crop yield prediction. There is need to integrate physical properties of maize such as maize plant height, stem diameter, stem girth, leaf length, leaf width, leaf area and number of leaves, which can be collected from the farm land primarily, together with weather data, environmental data, soil properties collected in real-time.

OBJECTIVES OF STUDY

1) To collect real-time data over a period of time, by deploying IoT sensor-based system in a maize crop farm.
2) To produce robust maize yield dataset based on soil, atmospheric and physical parameters.
3) To perform data preprocessing and relevant data selection.
4) To design machine learning models.
5) To implement the machine-learning models developed in number 4 to predict maize yield.
6) Evaluate the model using standard performance metrics.

METHODOLOGY

1) **Data Elicitation**
- IoT sensor-based and Openwaether API dataset
- Physical dataset
- Maize yield dataset

2) **Data Pre-processing**
- Data Cleaning and Merging
- Encoding Categorical Variables
  - One hot encoder:
  - Label Encoder
- Data Normalization

3) **Feature Selection** - Boruta algorithm

4) **Classification Models – ANN, NB, KNN, DT, Apriori & CNN-RNN**

5) **Model Evaluation**
- Accuracy, Recall, Precision. F1-Score and ROC curve.

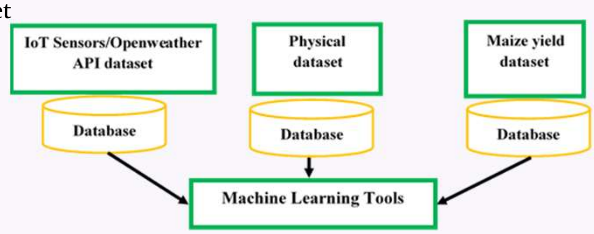

Fig. 3: Dataset Source

METHODOLOGY

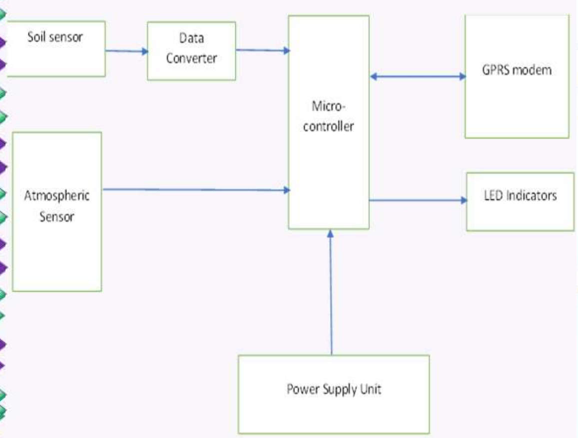

Fig.. 4: Block Diagram of IoT Sensor-based microcontroller system

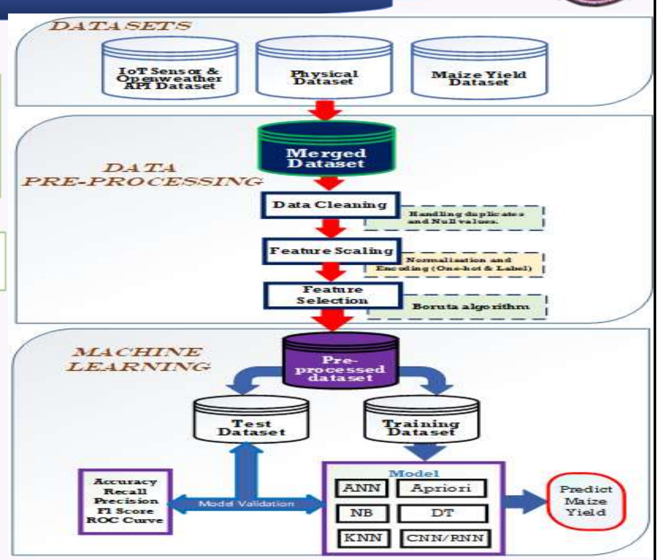

Fig. 5: Proposed System Architecture

METHODOLOGY

- The dataset used in this work was collected primarily from a physical maize plantation for a particular season between the month of August 2022 and November, 2022.
- The dataset consists of three (3) tables namely: IoT sensor and openweather API dataset, the physical dataset measured from the field and the maize yield dataset measured during the harvest period.
- The total number of records 63,231, were stored in a CSV file.
- The number of records in the training set is 50,585 records and the testing set is 12,646 records.

METHODOLOGY

Dataset Features

•The training data set contains the features on Table 1. Some of them are nominal, integer, float, binary and time stamp.

•Studies shows need of data discretization and normalization.

Need of features selection because large number of features decrease accuracy of the machine learning techniques used.

Table 1: Features of the Physical dataset

#	Column	Non-Null	Data_type	Description
0	Day_No.	2025	int64	Planting day (1) to harvest day (75)
1	Replication	2025	int64	Replication (1 or 2 or 3)
2	Plot_Code	2025	int64	Plot code (1 –9)
3	VarField_PltNo_Rep	2025	object	Unique code for Variety, field type, plot number and replication
4	Plant height (cm)	2025	float64	Plant height is the average plant height of the 10 plants tagged in each plot. Measured daily for the 27 plots using meter rule.
5	Stem_Diameter1(mm)	2025	float64	Stem_Diameter1 is the average of the first stem diameter measurement of the 10 plants tagged in each plot. Measured daily for the 27 plots using digital vernier caliper.
6	Stem_Diameter2 (mm)	2025	float64	Stem_Diameter2 is the average of the second stem diameter measurement of the 10 plants tagged in each plot. Measured daily for the 27 plots using digital vernier caliper.
7	Avg_Stem diameter (mm)	2025	float64	Avg_Stem Diameter is the average of the Stem_Diameter1 and Stem_Diameter2
8	Stem girth (mm)	2025	float64	Calculated Stem girth from the average stem diameter (Avg_Stem diameter*(22/7).
9	No. of leaves per plant	2025	int64	Average number of leaves per plant counted daily for 10 tagged plants per plot
10	leaf length (cm)	2025	float64	Average leaf length measured using meter rule for 10 tagged plants per plot
11	Leaf Width (cm)	2025	float64	Average leaf width measured using meter rule for 10 tagged plants per plot
12	Leaf area (cm^2)	2025	object	Calculated Leaf area from average leaf length and leaf width {(Leaf length*leaf width) * 0.75}
13	Tasseling	2025	int64	Number of plants that is tasseling from day 50 upwards.
14	Silking	2025	int64	Number of plants that is silking from day 50 upwards.

METHODOLOGY

Table 2: Features of the IoT sensor and openweather API dataset

#	Column	Non-Null	Data_type	Description
0	New_Id	66123	int64	New Identification Number
1	id	66123	int64	Identification Number
2	Day_No.	66123	int64	Planting day (1) to harvest day (75)
3	longi_op	66117	float64	Longitude of the location
4	lat_op	66117	float64	Latitude of the location
5	conditn_op	66117	object	Weather condition *(Rain or Clouds or Clear)*
6	description_op	66117	object	Weather description *(broken clouds or Clear sky or Few clouds or Light rain or Moderate rain or Overcast clouds or scattered clouds*
7	temperature_op	66117	float64	Temperature (ºC)
8	min_temp_op	66117	float64	Minimum Temperature (ºC)
9	max_temp_op	66117	float64	Maximum Temperature (ºC)
10	pressure_op	66117	float64	Atmospheric Pressure (hPa) *(on the sea level, if there is no sea_level or grnd_level data)*
11	humidity_op	66117	float64	Atmospheric Humidity (%)
12	sea_level_op	66117	float64	Atmospheric pressure on the sea level (hPa)
13	ground_level_op	66117	float64	Atmospheric pressure on the ground level (hPA)
14	visibility_op	66117	float64	Visibility (m)
15	wind_speed_op	66117	float64	Wind speed (m/s)
16	wind_deg_op	66117	float64	Wind degree (degrees)
17	wind_gust_op	66117	float64	Wind gust (m/s)
18	rain_op	66123	float64	Rain (mm)
19	cloud_op	66117	float64	Cloud (%)
20	country_op	66117	object	Country code (NG – for Nigeria)
21	sun_rise_op	66123	object	Sun rise time
22	sun_set_op	66123	object	Sun set time
23	temp	66123	float64	Soil temperature (ºC)
24	humidity	66123	int64	Soil humidity (%)
25	ph	66123	float64	Soil pH
26	nitro	66123	int64	Soil Nitrogen
27	phosphorus	66123	int64	Soil Phosphorus
28	potassium	66123	int64	Soil Potassium
29	at_temp	66123	float64	Sensor measured Atmospheric Temperature (ºC)
30	at_humidity	66123	int64	Sensor measured Atmospheric Humidity (%)
31	battery	66123	float64	IoT Sensor system Battery voltage (v)
32	location	66123	object	
33	longitude	66123	object	Plot longitude
34	latitude	66123	object	Plot Latitude
35	Replication	66123	int64	Replication (1 or 2 or 3)
36	Plot_Code	66123	int64	Plot code
37	VarField_PltNo_Rep	66123	object	Unique code for Variety, field type, plot number and replication
38	ddata_time	66123	object	Date and time for data logging
39	timestamp	66123	object	Date and time in epoch time.

CLASSIFICATION MODELS

ARTIFICIAL NEURAL NETWORKS (ANN) MODEL

- The neural network is also called Artificial Neural Network (ANN)
- ANN is derived from the biological concept of neurons. A neuron is a cell-like structure in a brain.
- ANN is a form of a mathematical model which uses a learning set of rules or algorithms inspired by the brain for the storage of information.

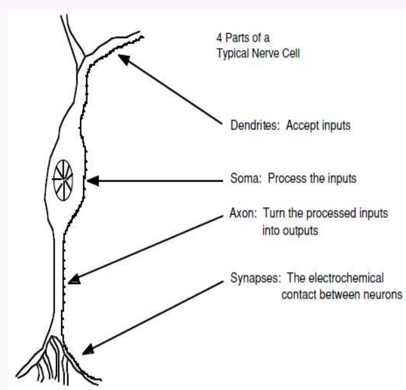

Fig. 6: Structure of a Simple Neuron (Anderson, 1992)

CLASSIFICATION MODELS

ARTIFICIAL NEURAL NETWORKS (ANN) MODEL

- ANN is an extensive class of flexible non-linear regression and discriminant models, data reduction models, and non-linear dynamical systems.
- ANN has been successfully applied in many areas of human endeavour like disease prognosis, economic forecasting, predictive modeling, smart agriculture, network intrusion detection, autonomous vehicle, material modeling, climatic control among others (Cook, 2020; Olayinka et al., 2020, Adetunmbi et al. 2021, Ibam et al., 2021 & Obe et al. 2022)

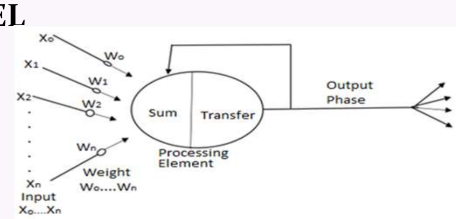

Fig. 7: Fundamental Structure of an Artificial Neuron

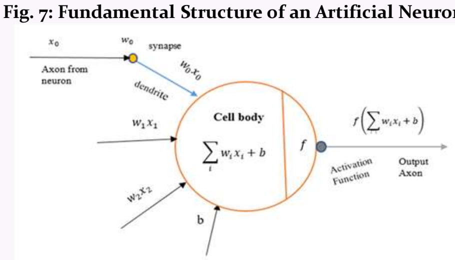

Fig. 8: Mathematical Model for ANN

CLASSIFICATION MODELS

ARTIFICIAL NEURAL NETWORKS (ANN) MODEL

- The architecture of ANN consists of: Input layer, Hidden layer and Output layer.
- There must be a connection from the nodes in the input layer with the nodes in the hidden layer and from each hidden layer node with the nodes of the output layer.
- Inputs are inserted into the input layer, and each node provides an output value via an activation function. The outputs of the input layer are used as inputs to the next hidden layer.
- ANN is capable of solving more complex and sophisticated tasks based on the number of hidden layers they possess (Olayinka et al.2022).

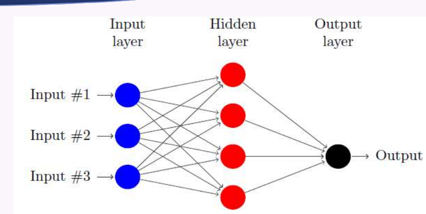

Fig. 9: ANN Architecture

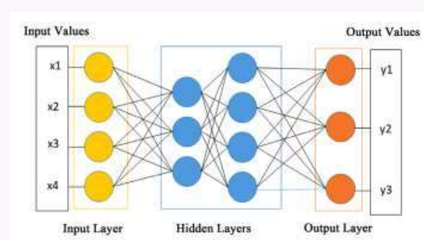

Fig. 10: Artificial Neural Network

CLASSIFICATION MODELS

Feed-Forward Propagation

- The feed-forward propagation calculates the predicted output ŷ. It takes the input elements (x_i), then, multiplied by the weights (w_i) such that:

$$\hat{y} = w_i * x_i = (w_1 * x_1) + (w_2 * x_2) + (w_3 * x_3) \quad (1)$$

- The weighted summation result passed through a sigmoid formula in order to calculate the neuron's output. The Sigmoid function is the step function used and is represented as:

$$\frac{1}{1+e^{-y}} \quad (2)$$

- The sigmoid function is used to normalize the result in between 0 and 1. Figure 11 depicts the graph of sigmoid function as well as other popular activation functions (Tanh and ReLU).

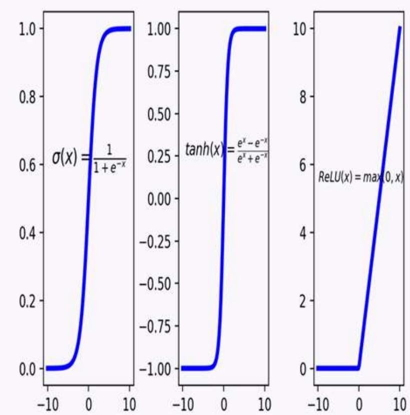

Fig.11: Activation Functions

CLASSIFICATION MODELS

BACKWARD PROPAGATION

- Backward propagation also referred to as back-propagation usually bring up-to-date the weights and biases.
- Back-propagation calculates the error, that is, the difference between the actual output and the expected output. Depending on the error, adjust the weights by multiplying the error with the input and again with the gradient of the Sigmoid curve:

(Weight += Error Input Output (1-Output) *(3)*

Output (1-Output) is derivative of sigmoid curve)

- Back propagation can be used only when the network is learning a training set. The sequential chart in figure 8, illustrates the process of training neural network.

CLASSIFICATION MODELS

Feed-Forward Propagation Algorithm

A forward propagation algorithm is given by:

$$u_j = f(Net_j) = \frac{1}{1+e^{-(\sum_i w_{j,i}u_i+\theta_j)}} \quad (4)$$

Such that;

$$Net_j = \sum_i w_{j,i} \cdot u_i + \theta_j \quad (5)$$

$$u_j = f(Net_j) \quad (6)$$

Backward Propagation Algorithm

1) Correction calculation of the weights connected to the output

i. $\Delta out_i = (t_j - u_j)f'(u_j)$ (7)

ii. $\Delta w_{j,i} = r.\Delta out_j.u_j$ (8)

2) Correction calculation of the weights not connected to the output

i. $\Delta hidden_i = f'(u_i).\sum_j \Delta out_j w_{j,i}$ (9)

ii. $\Delta w_{i.k} = r.\Delta hidden_i.u_k$ (10)

3) Fulfillment of the corrections on the weights

For weight ($w_{j,i(n)}$ *and* $w_{i,k(n)}$) and change in weight ($\Delta w_{j,i}$ *and* $w_{i,k}$)

i. $w_{j,i(n+1)} = w_{j,i(n)} + \Delta w_{j,i}$ (11)

ii. $w_{i,k(n+1)} = w_{i,k(n)} + \Delta w_{i,k}$ (12)

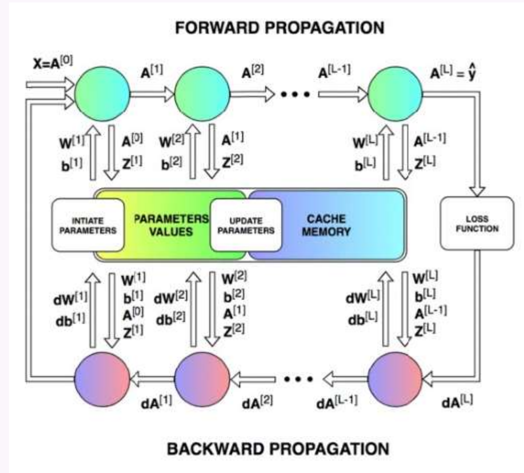

Fig. 12: ANN Blueprint

CLASSIFICATION MODELS

NAIVE BAYES (NB)

NB is a robust supervised learning method which have the ability to perform relatively well in complex classification situations. Gaussian naïve bayes (GNB) is a variant of Naïve bayes (NB).

Gaussian naïve bayes (GNB) incorporates distribution modeling into Naïve Bayes to compute the likelihood of an instance by estimating the mean (μ) and variance (σ) of the features given its class label. The mathematical representation of GNB is captured in equation (13).

$$P(\vec{x}\,|c_k) = \frac{1}{\sqrt{2\pi\sigma_{c_k}^2}} \exp^{\left(-\frac{(x_i-\mu)^2}{2\sigma_{c_k}^2}\right)} \quad (13)$$

Where μ represents the mean and (σ^2) represents the variance as described in equation (14) and (15) respectively.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (14)$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \quad (15)$$

Maximum A Posteriori (MAP) Decision rule represented in equation (16) decides which class to be assigned by simply picking the class with the largest probability given the features.

$$P(\vec{x}|c_k)P(c_k) > P(\vec{x}|c_i)P(c_i) for\ 1 \le i \le m, k \ne i \quad (16)$$

CLASSIFICATION MODELS

K-NEAREST NEIGHBOUR (k-NN)

- KNN predicts the class of unlabeled test object x_t by finding a group of k training instances that are closest to x_t, then assign the most frequent class labels of k instances as the class label for x_t .
- To determine the nearest neighbor list, the *Euclidean distance function is* presented in (17), to find the proximity of all instances in the training set $D = \{(x_1, y_1), (x_2, y_2),..., (x_n, y_n)\}$ with x_t, where is an instance and y_i is attached class label.

$$E_{Dis} = \sqrt{\sum_{i=1}^{m}(x_{qi} - x_{pi})^2} \quad (17)$$

- where m denotes the number of features, i is the sequence, q is the feature value of the training instance feature x_i, and p is the feature value of the test instance feature x_i. Thereafter, the number of k training instances with the least distance values are selected as the nearest neighbors, then the majority voting function is applied to find the class label that is most frequent in these nearest neighbors.

CLASSIFICATION MODELS

DECISION TREE (DT)

- Decision Tree is the most powerful and popular tool for classification and prediction.
- A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.
- Decision trees are able to generate understandable rules.
- Decision trees perform classification without requiring much computation.

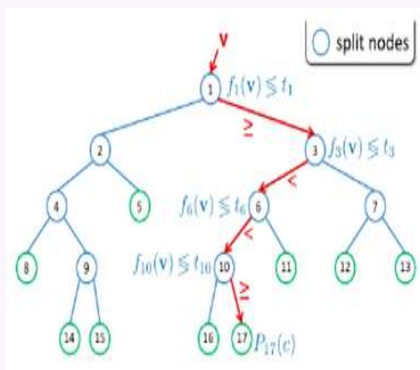

Fig.13: Binary Decision Tree

CLASSIFICATION MODELS

APRIORI ALGORITHM

- Apriori algorithm is given by R.
- Apriori algorithm is used for finding frequent item sets in a dataset for boolean association rule.
- Apriori algorithm uses prior knowledge of frequent item set properties, which apply an iterative approach or level-wise search where k-frequent itemsets are used to find k+1 itemsets.
- Apriori algorithm is used to improve the efficiency of level-wise generation of frequent item sets and reducing the search space.
- All non-empty subset of frequent itemset must be frequent. The key concept of Apriori algorithm is its anti-monotonicity of support measure. Apriori assumes that:

All subsets of a frequent itemset must be frequent(Apriori property).
If an itemset is infrequent, all its supersets will be infrequent.

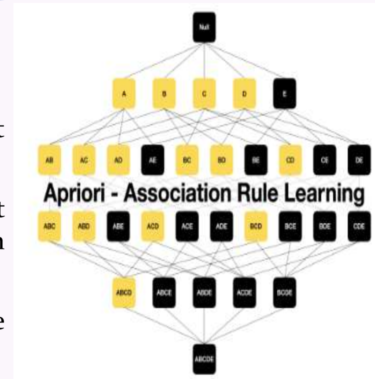

Fig.14: Apriori algorithm for Association Rule learning

CLASSIFICATION MODELS

HYBRID CONVOLUTIONAL NEURAL NETWORKS & RECURRENT NEURAL NETWORK (CNN-RNN) MODEL

- An Hybrid CNN and RNN architecture takes the positive features of a RNN are used to improve the CNN.
- The key module of this RCNN are the recurrent convolution layers (RCL), which introduce recurrent connection into a convolution layer. With these connections the network can evolve over time though the input is static and each unit is influenced by its neighboring units.
- This property integrates the context information of an image, which is important for object detection.

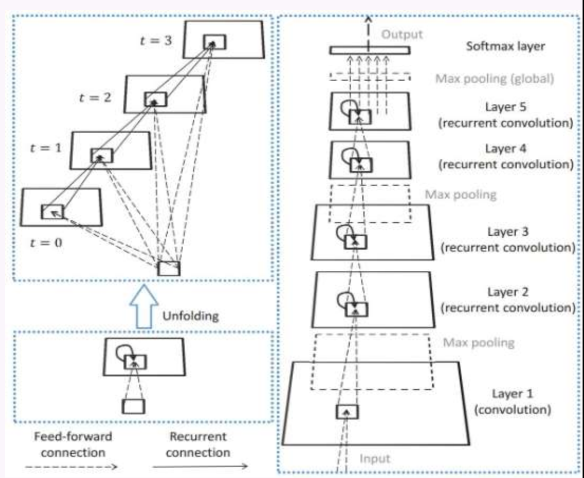

Fig.14: Hybrid CNN-RNN Architecture

Classification Performance Evaluation

$$\text{Sensitivity} = \frac{\text{Number of 'True Positives'}}{\text{Number of 'True Positives'} + \text{Number of 'False Negatives'}}$$

$$\text{Specificity} = \frac{\text{Number of 'True Negatives'}}{\text{Number of 'True Negatives'} + \text{Number of 'False Positives'}}$$

$$\text{PPV} = \frac{\text{Number of 'True Positives'}}{\text{Number of 'True Positives'} + \text{Number of 'False Positives'}}$$

$$\text{NPV} = \frac{\text{Number of 'True Negatives'}}{\text{Number of 'True Negatives'} + \text{Number of 'False Negatives'}}$$

- *False positive rate (FP rate) = FP / (FP + TN) = 1 - specificity;*
- *False negative rate (FN rate) = FN / (TP + FN) = 1 - sensitivity;*
- *Likelihood ratio positive (LR+) = sensitivity/(1 - specificity);*
- *Likelihood ratio negative (LR-) = (1 - sensitivity)/specificity.*

CONTRIBUTION TO KNOWLEDGE

- This research produced a novel dataset for maize crop yield based on soil, atmospheric, weather and physical parameters for three (3) standard maize varieties over three (3) distinct soil treatments.
- This research work is expected to develop an intelligent IoT-driven climate-independent maize crop farming support system.

REFERENCES

- Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine learning applications on agricultural datasets for smart farm enhancement. Machines, 6(3). https://doi.org/10.3390/machines6030038
- Bhanumathi, S., Vineeth, M., & Rohit, N. (2019). Crop yield prediction and efficient use of fertilizers. Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019. https://doi.org/10.1109/ICCSP.2019.8698087
- Chen, X., Feng, L., Yao, R., Wu, X., Sun, J., & Gong, W. (2021). Prediction of maize yield at the city level in China using multi-source data. Remote Sensing, 13(1). https://doi.org/10.3390/rs13010146
- Dewi, C., & Chen, R. C. (2020). Decision making based on IoT data collection for precision agriculture. In Studies in Computational Intelligence (Vol. 830). https://doi.org/10.1007/978-3-030-14132-5_3.
- Di, M. B. (2018). Internet of things reference architectures, security and interoperability: A survey. Elsevier–IoT. Vol 1-2 ,99–112.
- Farooq M.S., Shamyla R., Adnan A., Kamran A. and Muhammad A. N. (2019). Special section on new technologies for smart farming 4.0: research challenges and opportunities: a survey on the role of IoT in agriculture for the implementation of Smart Farming. IEEE Access Vol. 7, 156237. DOI: 10.1109/ACCESS.2019.2949703.
- George S., Ioana M., Cristina B., Elena B. and Marius D. (2019). Efficient IoT system for Precision Agriculture. 15th International Conference on Engineering of Modern Electric Systems (EMES). 978-1-7281-0773-8/19 European Union. Pp 173-176
- Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., Robin Bryant, C., & Senthilnath, J. (2021). Integrated phenology and climate in rice yields prediction using machine learning methods. Ecological Indicators, 120. https://doi.org/10.1016/j.ecolind.2020.106935.
- Jahan, R. (2018). Applying Naive Bayes Classification Technique for Classification of Improved Agricultural Land soils. International Journal for Research in Applied Science and Engineering Technology, 6(5). https://doi.org/10.22214/ijraset.2018.5030

REFERENCES

- Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. Business Horizons, 58(4). https://doi.org/10.1016/j.bushor.2015.03.008
- Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio da Silva Junior, C., Fernando Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., Marcato Junior, J., Eduardo Teodoro, P., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. Computers and Electronics in Agriculture, 178. https://doi.org/10.1016/j.compag.2020.105791.
- Prathibha, S. R., Hongal, A., & Jyothi, M. P. (2017). IOT Based Monitoring System in Smart Agriculture. Proceedings - 2017 International Conference on Recent Advances in Electronics and Communication Technology, ICRAECT 2017. https://doi.org/10.1109/ICRAECT.2017.52
- Priya, Muthaiah, & Balamurugan. (2018). Predicting yield of the crop using machine learning algorithms. International Journal of Engineering Sciences and Reseach Technology, 7(4).
- Shahhosseini, M., Martinez-Feria, R. A., Hu, G., & Archontoulis, S. v. (2019). Maize yield and nitrate loss prediction with machine learning algorithms. Environmental Research Letters. https://doi.org/10.1088/1748-9326/ab5268
- Shakoor, M. T., Rahman, K., Rayta, S. N., & Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques. 2017 1st International Conference on Next Generation Computing Applications, NextComp 2017. https://doi.org/10.1109/NEXTCOMP.2017.8016196
- Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. Mathematical and Computer Modelling of Dynamical Systems, 27(1). https://doi.org/10.1080/13873954.2021.1882505
- UN. (2019). World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100. UN Department of Economic and Social Affairs.
- Yacob, A. (2020). A Review of Internet of Things (IoT): Implementations and Challenges. International Journal of Advanced Trends in Computer Science and Engineering, 9(1.3). https://doi.org/10.30534/ijatcse/2020/5891.32020.

DATASETS

id	Day_No.	longi_op	lat_op	conditn_op	description_op	temperature_op	min_temp_op	max_temp_op	pressure_op	humidity_op	sea_level_op	ground_level_op	visibility_op	wind_speed_op	wind_deg_op	wind_gust_op	rain_op	cloud_op	country_op	sun_rise_op	sun_set_op	temp	humidity	ph	nitro	phosphorus
5149	1	5.5851	6.2808	Clouds	overcast clouds	23.29	23.29	23.29	1014	87	1014	1009	10000	2.07	241	7.18	0	85	NG	6:32:16	18:50:16	30.2	47	1.79	100	68
5148	1	5.5851	6.2808	Clouds	overcast clouds	23.29	23.29	23.29	1014	87	1014	1009	10000	2.07	241	7.18	0	85	NG	6:32:16	18:50:16	30.2	47	1.79	100	68
5151	1	5.5851	6.2808	Clouds	overcast clouds	23.29	23.29	23.29	1014	87	1014	1009	10000	2.07	241	7.18	0	85	NG	6:32:16	18:50:16	30.14	47	1.79	100	68
5150	1	5.5851	6.2808	Clouds	overcast clouds	23.29	23.29	23.29	1014	87	1014	1009	10000	2.07	241	7.18	0	85	NG	6:32:16	18:50:16	30.17	47	1.79	100	68
25310	1	5.5851	6.2808	Rain	moderate rain	23.49	23.49	23.49	1012	97	1012	1006	10000	1.61	233	5.79	1.22	100	NG	6:27:11	18:34:31	29.18	47	5.49	89	68
84155	1	5.5851	6.2808	Clouds	scattered clouds	23.29	23.29	23.29	1009	98	1009	1003	10000	0.76	242	0.85	0	48	NG	6:24:05	18:18:19	25.29	48	9.96	100	70
84436	1	5.5851	6.2808	Rain	light rain	33.48	33.48	33.48	1009	50	1009	1004	10000	0.86	166	3.26	0.26	26	NG	6:24:05	18:18:19	35.48	50	9.96	100	72
84924	1	5.5851	6.2808	Clouds	scattered clouds	22.95	22.95	22.95	1010	97	1010	1005	10000	0.59	308	0.59	0	39	NG	6:24:13	18:18:09	24.95	52	5.84	100	73
55318	1	5.5851	6.2808	Rain	light rain	25.42	25.42	25.42	1010	91	1010	1004	10000	2.03	237	5.74	0.89	99	NG	6:24:01	18:24:33	27.42	52	9.96	100	75
55315	1	5.5851	6.2808	Rain	light rain	25.42	25.42	25.42	1010	91	1010	1004	10000	2.03	237	5.74	0.89	99	NG	6:24:01	18:24:33	27.42	52	9.96	100	75
20301	1	5.5851	6.2808	Clouds	overcast clouds	23.31	23.31	23.31	1014	96	1014	1008	10000	1.39	234	6.01	0	100	NG	6:28:34	18:38:18	31.39	52	2.38	100	75
54640	1	5.5851	6.2808	Rain	light rain	30.63	30.63	30.63	1011	65	1011	1005	10000	2.19	208	1.57	0.5	78	NG	6:24:01	18:24:33	32.63	54	9.96	100	77
54552	1	5.5851	6.2808	Rain	light rain	31.67	31.67	31.67	1011	59	1011	1005	10000	1.64	208	1.25	0.37	73	NG	6:24:01	18:24:33	33.67	53	9.96	100	77
54550	1	5.5851	6.2808	Rain	light rain	30.63	30.63	30.63	1011	65	1011	1005	10000	2.19	208	1.57	0.5	78	NG	6:24:01	18:24:33	32.63	53	9.96	100	77
78528	1	5.5851	6.2808	Clouds	scattered clouds	23.43	23.43	23.43	1013	97	1013	1008	10000	0.33	1	0.32	0	46	NG	6:23:32	18:20:12	25.43	54	9.96	100	77
55181	1	5.5851	6.2808	Rain	moderate rain	27.63	27.63	27.63	1009	84	1009	1004	10000	1.86	234	4.14	1.25	100	NG	6:24:01	18:24:33	29.63	54	9.96	100	77
54644	1	5.5851	6.2808	Rain	light rain	30.63	30.63	30.63	1011	65	1011	1005	10000	2.19	208	1.57	0.5	78	NG	6:24:01	18:24:33	32.63	54	9.96	100	77
83775	1	5.5851	6.2808	Clouds	broken clouds	31.75	31.75	31.75	1007	53	1007	1001	10000	0.85	167	1.58	0	64	NG	6:23:58	18:18:30	33.75	52	9.96	100	76
83776	1	5.5851	6.2808	Clouds	broken clouds	31.75	31.75	31.75	1007	53	1007	1001	10000	0.85	167	1.58	0	64	NG	6:23:58	18:18:30	33.75	52	9.96	100	76
20303	1	5.5851	6.2808	Clouds	overcast clouds	23.31	23.31	23.31	1014	96	1014	1008	10000	1.39	234	6.01	0	100	NG	6:28:34	18:38:18	31.39	52	2.38	100	75
20305	1	5.5851	6.2808	Clouds	overcast clouds	23.31	23.31	23.31	1014	96	1014	1008	10000	1.39	234	6.01	0	100	NG	6:28:34	18:38:18	31.39	52	2.38	100	75
20123	1	5.5851	6.2808	Rain	light rain	28.27	28.27	28.27	1009	77	1009	1004	10000	2.56	239	3.28	1	76	NG	6:28:34	18:38:18	31.88	52	2.39	100	75
55314	1	5.5851	6.2808	Rain	light rain	25.42	25.42	25.42	1010	91	1010	1004	10000	2.03	237	5.74	0.89	99	NG	6:24:01	18:24:33	27.42	52	9.96	100	75
55317	1	5.5851	6.2808	Rain	light rain	25.42	25.42	25.42	1010	91	1010	1004	10000	2.03	237	5.74	0.89	99	NG	6:24:01	18:24:33	27.42	52	9.96	100	75

PHYSICAL MAIZE FARM PICTURES

PHYSICAL MAIZE FARM PICTURES

Thank you for Listening