



Maago – A Smart System for Mango Plantation Management

Project ID : 2023-309

Introduction

- Mango cultivation is a developing sector in the world.
- Sri Lanka has good environmental conditions for mango cultivation.
- There is a problem with the lack of systems in Sri Lanka to increase the productivity of mango plantations.
- This leads to a significant loss of mango yield and quality.
- It is essential to create smart mango plantation management systems that make use of technological advancements and data analysis to address these issues.



Research Problem



This research aims to identify key challenges in managing mango plantations in Sri Lanka



Related to pest and disease control, irrigation methods, and post-harvest management.



The effectiveness of current management practices will be evaluated.



The study aims to offer solutions to improve mango plantation management for sustainable production.

Research Objective

- To implement an integrated solution to increase the production of the mango plantation and help the management to monitor the plants.

Commercialization

- We plan to sell the system to the Mango Farmers
- Leased or rented to farmers who cannot afford to purchase
- Participate in the government projects
- Offer maintenance contracts to farmers who have purchased the system.
- Collaborate with Non-Governmental Organizations (NGOs) to provide the system to small scale farmers.

IT20466008

Withanaarachchi S. P.



Specialization: Computer Systems and Network Engineering

Smart Watering System Enabled by IoT and Machine Learning

Introduction

Sri Lanka has a long and rich history of mango cultivation. The objective is to create a smart watering system based on IoT and Machine Learning to boost mango production and reduce water usage.



Research Question

- How to identify the water needs of the plantation provide water through a smart system using IoT and machine Learning technologies?



Objective

The main objective of this component is to implement a smart water management system that provides water to the plantation at the correct time and in the correct amount.

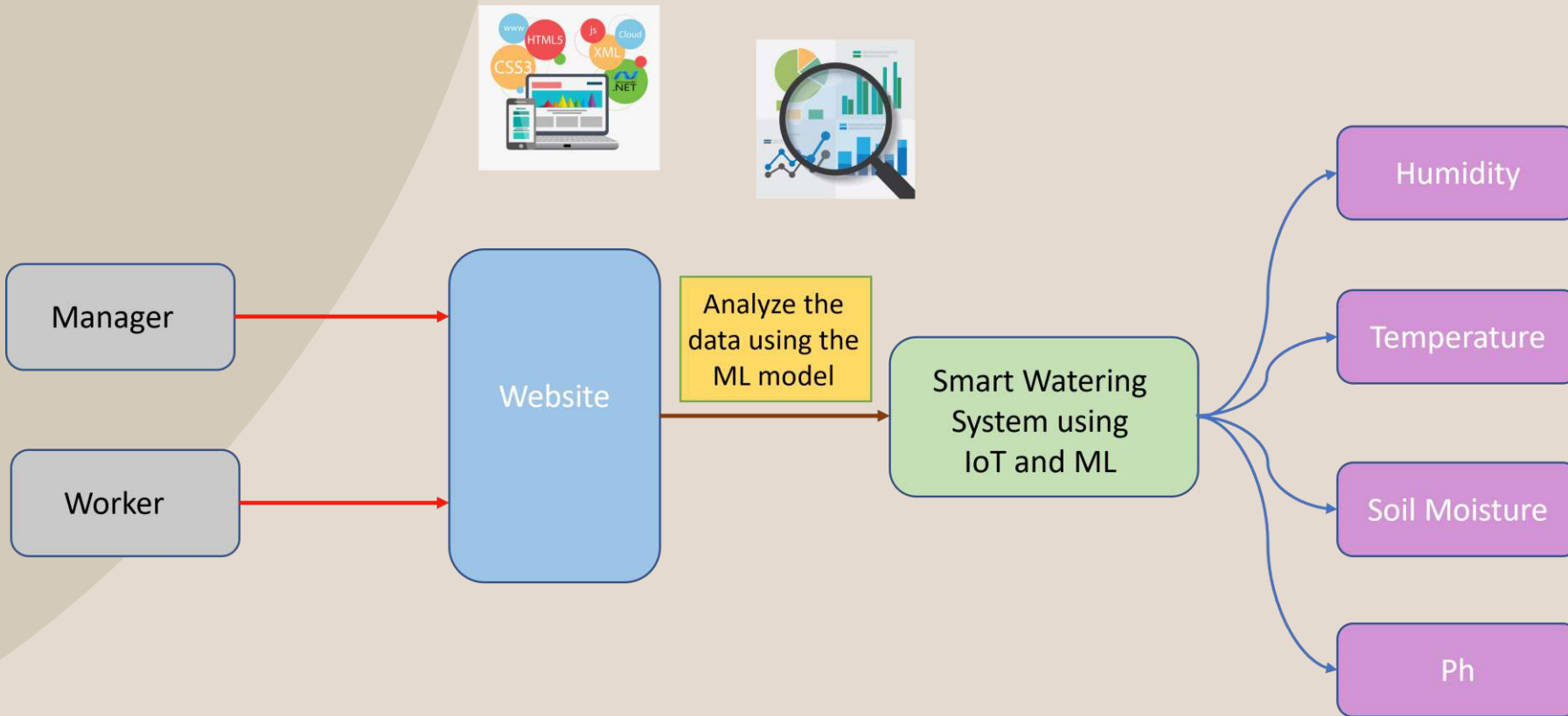


Methodology

- An IoT device is developed to measure the soil conditions. The device will be equipped with temperature, humidity, and soil moisture sensors.
- A machine learning model is developed to predict the water needed for the plantation according to the sensor readings. The IoT device and machine learning model will be integrated to create the proposed smart watering system.
- A mobile application is used to update the real-time data to the users.



System Diagram



Technologies

- Arduino
- Python
- C
- Google Colab
- Algorithms - Linear Regression Algorithm



Tools

- ESP32 module
- DHT-11 sensor
- Soil moisture sensor
- PH sensor



Completion of the IoT device

```
sketch_may23a | Arduino 1.8.13
File Edit Sketch Tools Help

sketch_may23a

///// Humidity & Temperature Monitoring sensor readings /////
#include <DHT.h>
#define DHT_SENSOR_PIN 21 // ESP32 pin GPIO21 connected to DHT11 sensor
#define DHT_SENSOR_TYPE DHT11

////////////////////////////////////// Moisture PID Initialisation ////////////////////////////////////////
const int moisturePin = A0; // moisture sensor pin
float moisturePercentage; //moisture reading

DHT dht_sensor(DHT_SENSOR_PIN, DHT_SENSOR_TYPE);

// Define a struct to return values of Humidity and Temperature ///
struct HumidityTemperatureData {
    bool success;
    float humidity;
    float temperatureC;
    float temperatureF;
};

void setup() {
    Serial.begin(9600);
    dht_sensor.begin(); // initialize the DHT sensor
}

void loop() {
    HumidityTemperatureData data = readHumidityTemperature();
    if (data.success) {
        Serial.println("Humidity: ");
        Serial.println(data.humidity);
        Serial.println(" ");
        Serial.println(" ");
    }
}
```



```
sketch_may23a | Arduino 1.8.13
File Edit Sketch Tools Help

sketch_may23a

// wait a 2 seconds between readings
delay(2000);
}

float ReadMoisture() {
  Serial.println("Start Moisture Reading");
  moisturePercentage = 100 - ( (analogRead(moisturePin) / 1023.00) * 100 );
  Serial.print("Soil Moisture is: ");
  Serial.print(moisturePercentage);
  Serial.println("%");
  return moisturePercentage;
}

HumidityTemperatureData readHumidityTemperature() {
  HumidityTemperatureData data;
  data.humidity = dht_sensor.readHumidity(); // read humidity
  data.temperatureC = dht_sensor.readTemperature(); // read temperature in Celsius
  data.temperatureF = dht_sensor.readTemperature(F); // read temperature in Fahrenheit

  // check whether the reading is successful or not
  bool isReadingSuccessful = (!isnan(data.temperatureC) || !isnan(data.temperatureF) || !isnan(data.humidity));
  data.success = isReadingSuccessful;

  if (!isReadingSuccessful) {
    data.humidity = 0.0;
    data.temperatureC = 0.0;
    data.temperatureF = 0.0;
  }

  return data;
}

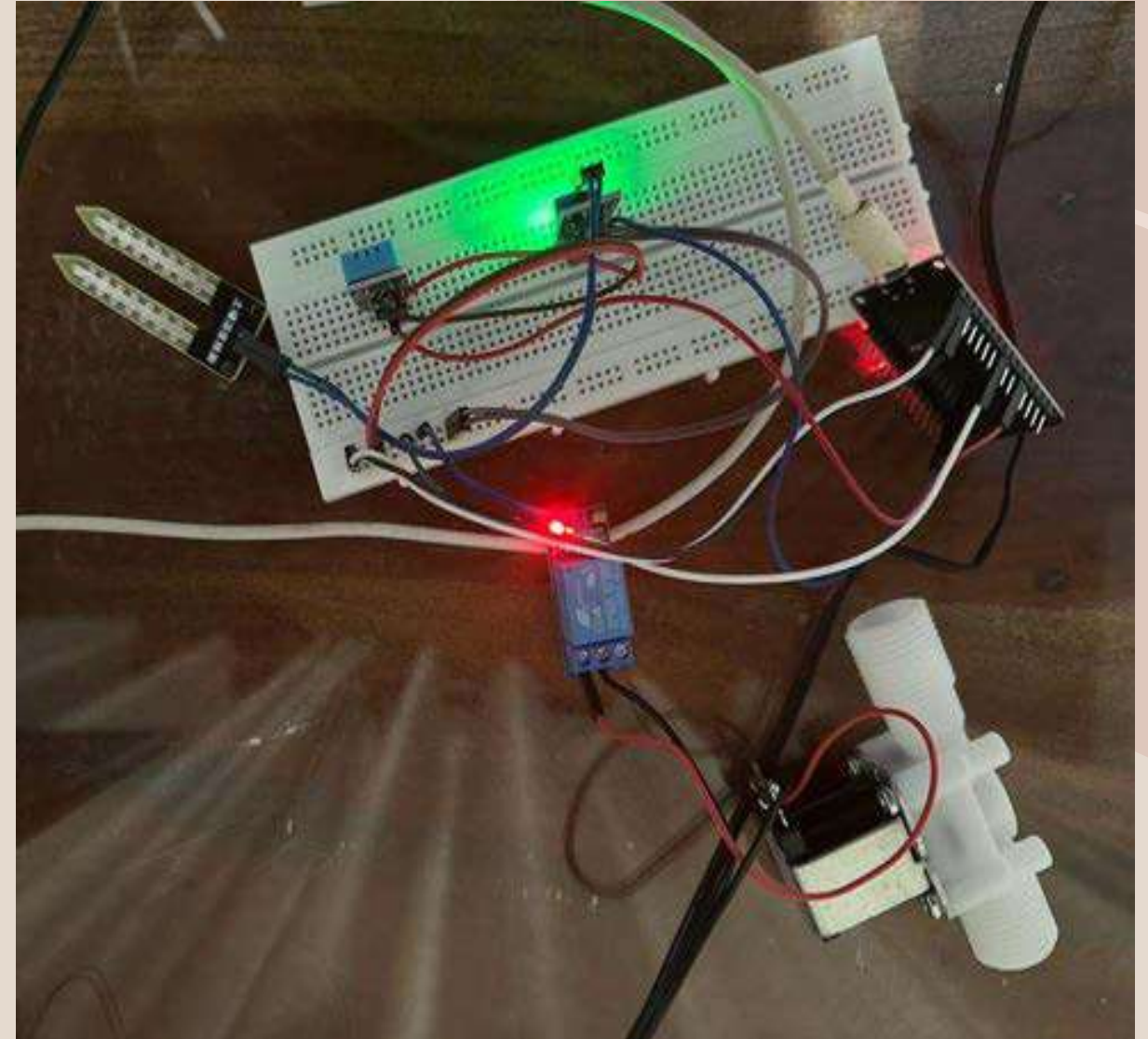
Done compiling

Sketch uses 268425 bytes (20%) of program storage space. Maximum is 1310720 bytes.
Global variables use 22552 bytes (6%) of dynamic memory, leaving 305128 bytes for local variables. Maximum is 327680 bytes.

83 ESP32 Dev Module, Disabled, Disabled, Default 4MB with spiffs (1.2MB APP), 5MB SPIFFS, 240MHz (WiFiBT), QIO, 80MHz, 4MB (32MB), 921600, Core 1, Core 1, None, Disabled on COM4
```

COM4

```
Humidity: 63.00% | Temperature: 33.90°C ~ 93.02°F  
Start Moisture Reading  
Soil Moisture is = -300.29%  
Humidity: 63.00% | Temperature: 33.90°C ~ 93.02°F  
Start Moisture Reading  
Soil Moisture is = -300.29%  
Humidity: 63.00% | Temperature: 33.90°C ~ 93.02°F  
Start Moisture Reading  
Soil Moisture is = -300.29%  
Humidity: 63.00% | Temperature: 33.90°C ~ 93.02°F  
Start Moisture Reading  
Soil Moisture is = -300.29%  
Humidity: 63.00% | Temperature: 33.90°C ~ 93.02°F  
Start Moisture Reading  
Soil Moisture is = -300.29%  
Humidity: 63.00% | Temperature: 33.90°C ~ 93.02°F  
Start Moisture Reading  
Soil Moisture is = -300.29%  
Humidity: 63.00% | Temperature: 33.90°C ~ 93.02°F  
Start Moisture Reading  
Soil Moisture is = -300.29%
```



Completion of the ML model

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3 from sklearn.linear_model import LinearRegression
4 from sklearn.metrics import mean_squared_error, r2_score

5 # Assuming you have already created the dataframe 'df' with the provided data

6 # Split the data into features (X) and target variable (y)
7 X = df.drop('Amount of water Needed (ml)', axis=1)
8 y = df['Amount of water Needed (ml)']

9 # Split the data into training and testing sets
10 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

11 # Initialize and train the linear regression model
12 reg_model = LinearRegression()
13 reg_model.fit(X_train, y_train)

14 # Make predictions on the test set
15 y_pred = reg_model.predict(X_test)

16 # Evaluate the model
17 mse = mean_squared_error(y_test, y_pred)
18 r2 = r2_score(y_test, y_pred)

19 print("Mean Squared Error:", mse)
20 print("R-squared score:", r2)
```

```
21 Mean Squared Error: 5.290136427517269
22 R-squared score: 0.7462031199185647
```

```
23 [ ] import pandas as pd
24     import numpy as np
```

```
1 # Get the total number of rows in the dataframe

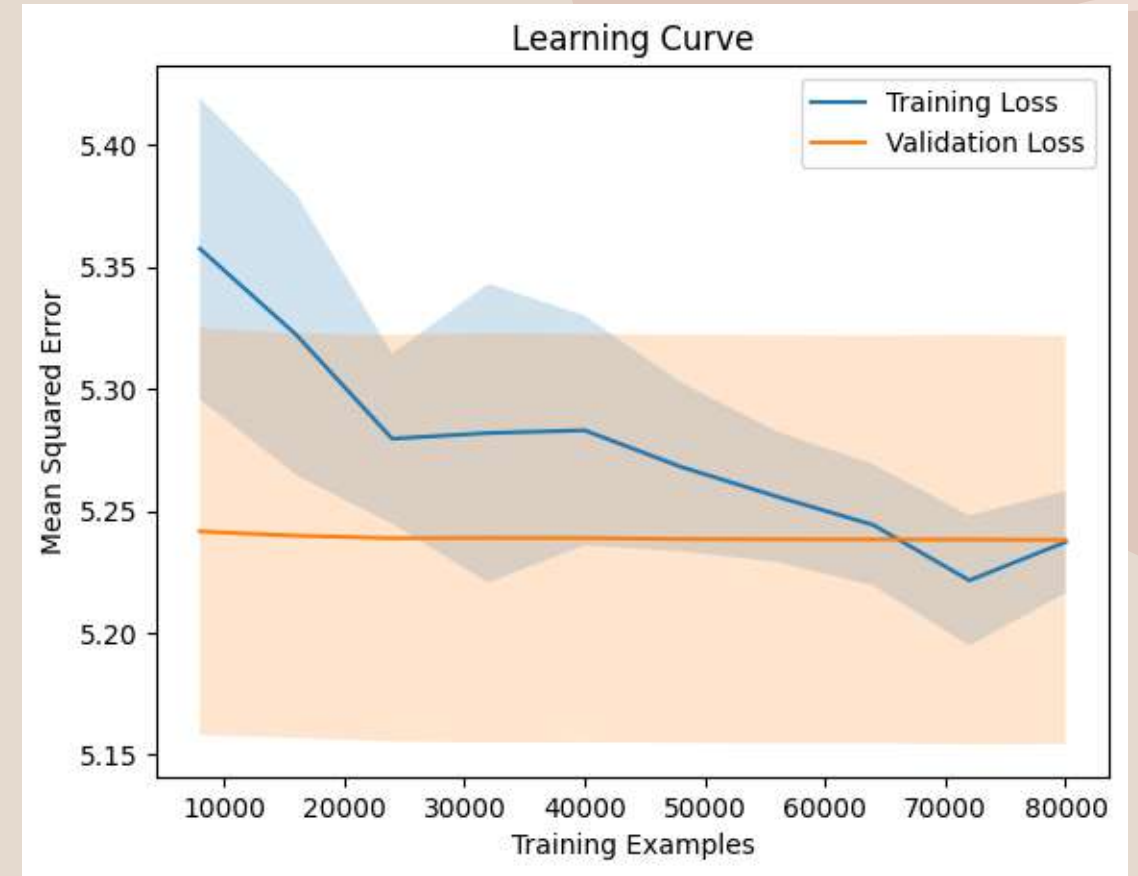
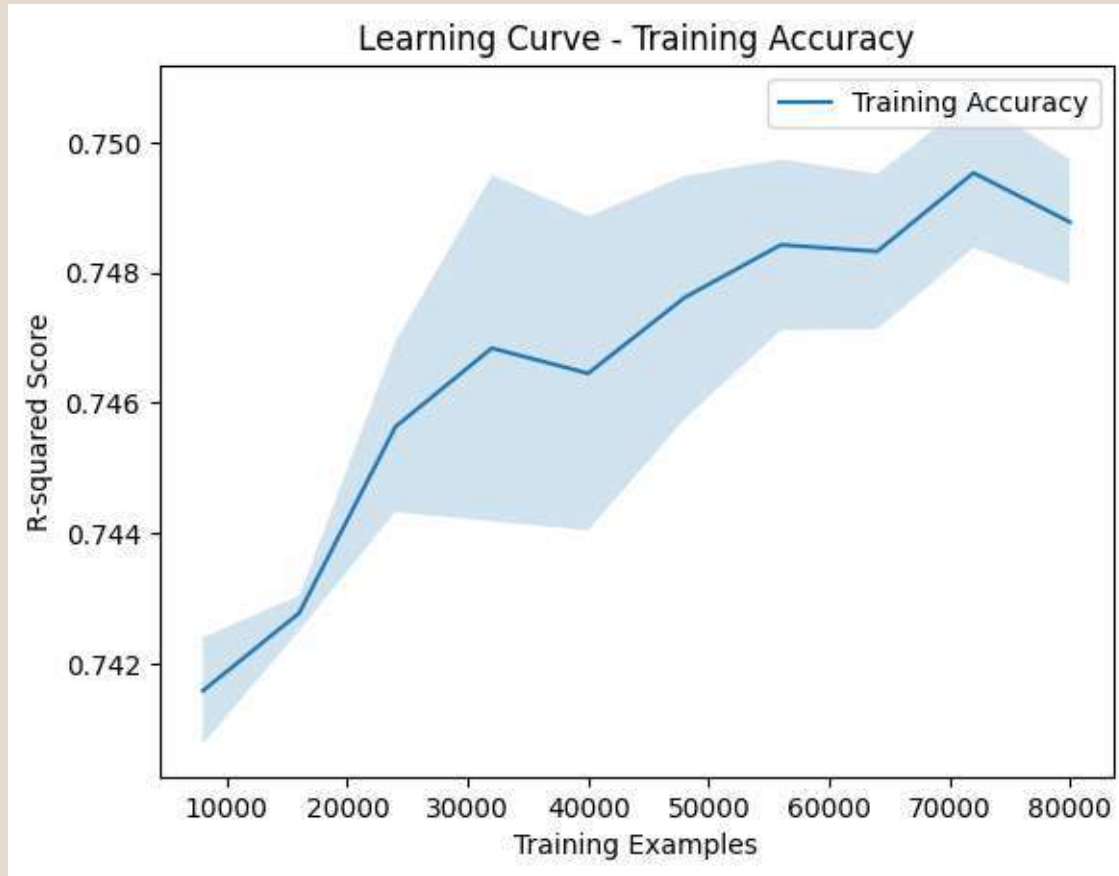
2 df = pd.read_csv('/content/drive/MyDrive/Kaggle/Kaggle_project-2023052320702412-801/Kaggle_project/watering_data.csv')
3 total_rows = df.shape[0]

4 # Select 20% of the indices randomly
5 selected_indices = np.random.choice(total_rows, size=int(total_rows * 0.2), replace=False)

6 # Multiply the yield values in the selected indices by 0.20
7 df.loc[selected_indices, 'Amount of water Needed (ml)'] *= 0.20
8 df
```

	pH Level	Humidity Level	Moisture Level	Temperature	Amount of water Needed (ml)
0	6.213000	34.106645	0.861426	30.204245	21.589984
1	8.972373	62.073150	0.600127	36.886552	30.885282
2	6.824781	60.884988	0.077513	28.794518	21.014545
3	7.662937	54.746297	0.094007	25.583974	24.644907
4	3.040124	27.272004	0.850259	34.522049	26.019462
...
9995	8.670932	68.547809	0.507732	22.270037	17.182713
9996	11.988047	86.436557	0.207808	25.064946	23.886887
9997	7.097638	62.802012	0.788198	28.656678	20.774722
9998	6.544100	68.551031	0.732225	31.024891	24.302440
9999	3.614915	33.915484	0.024769	33.580151	24.321158

Completion of the ML model



Progress at the moment

- Implementation of the soil moisture sensor and temperature and humidity sensor
- Train the ML model to identify the water needed for the plantation

What's to be done

- Update the database with real-time data.
- Finish coding the back end of all pages.
- Testing the finished system using a live environment.
- Introduce the system to the clients and research how they interact.

IT20280260

Aksham M.Z.M



Specialization: Information Technology

Identify diseases using a scanning system

Introduction

- To implement a system to identify the disease on the mango leaves using the scanner
- Identifying disease on mango leaves is to facilitate early detection and control of plant diseases to increase the mango production

Research Question

- How to identify the mango disease and increase the production through a scanning system?



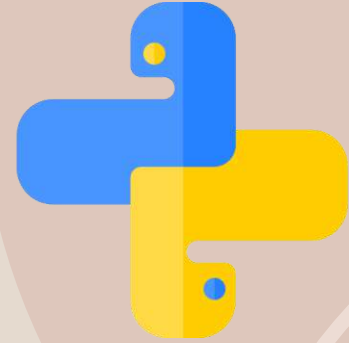
Methodology

- Machine learning-based image analysis:

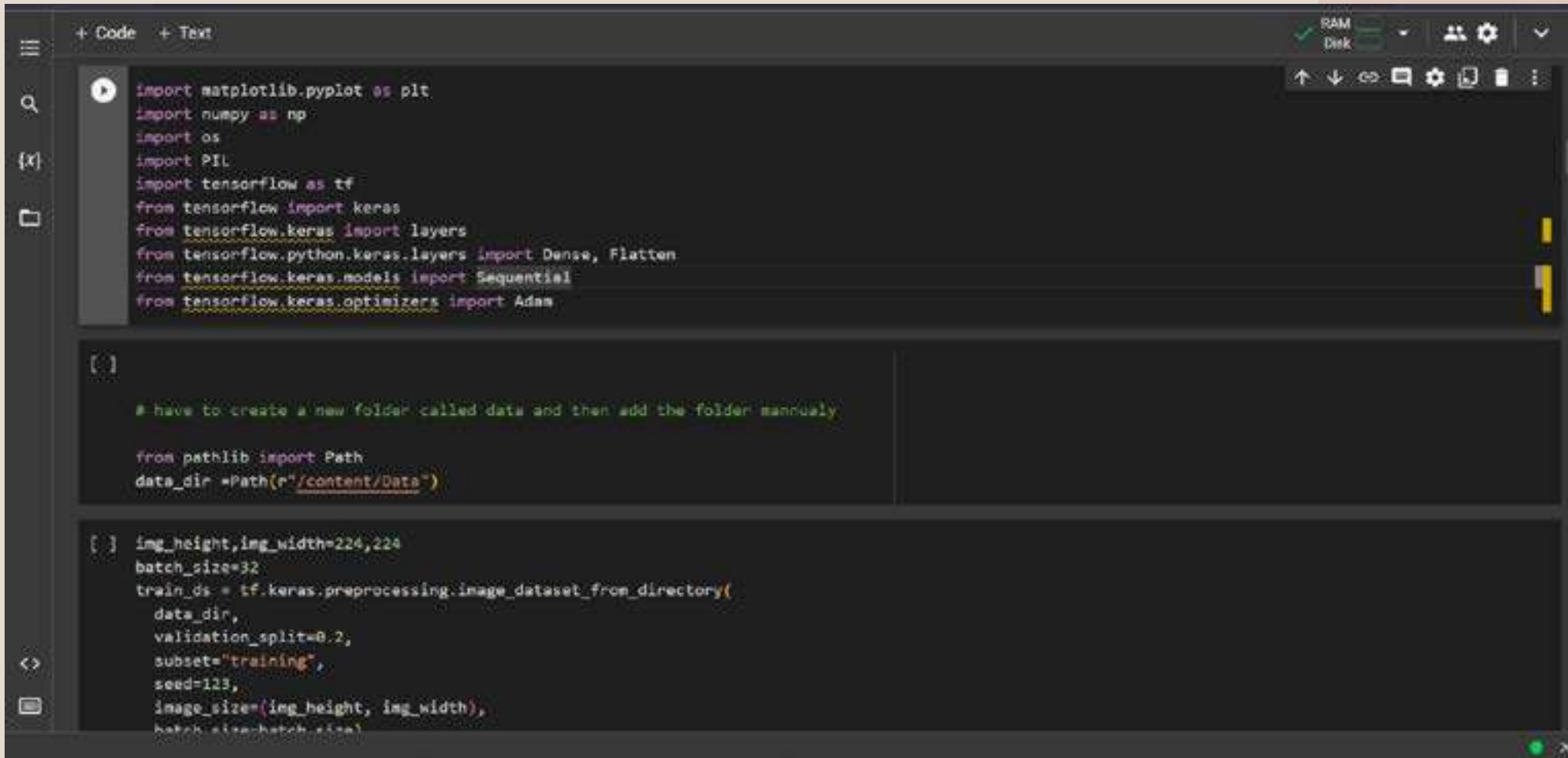
Method for spotting diseases of mango tree leaves is to analyze photographs of the leaves and look for patterns that correspond to various diseases. In order to construct a predictive model for disease identification, the system may need to be trained a large dataset of images of both healthy and diseased mango leaves.

Tools and Technologies

- Python
- Google Colab
- Pandas
- NumPy



Train the model



```
+ Code + Text
import matplotlib.pyplot as plt
import numpy as np
import os
import PIL
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from tensorflow.python.keras.layers import Dense, Flatten
from tensorflow.keras.models import Sequential
from tensorflow.keras.optimizers import Adam

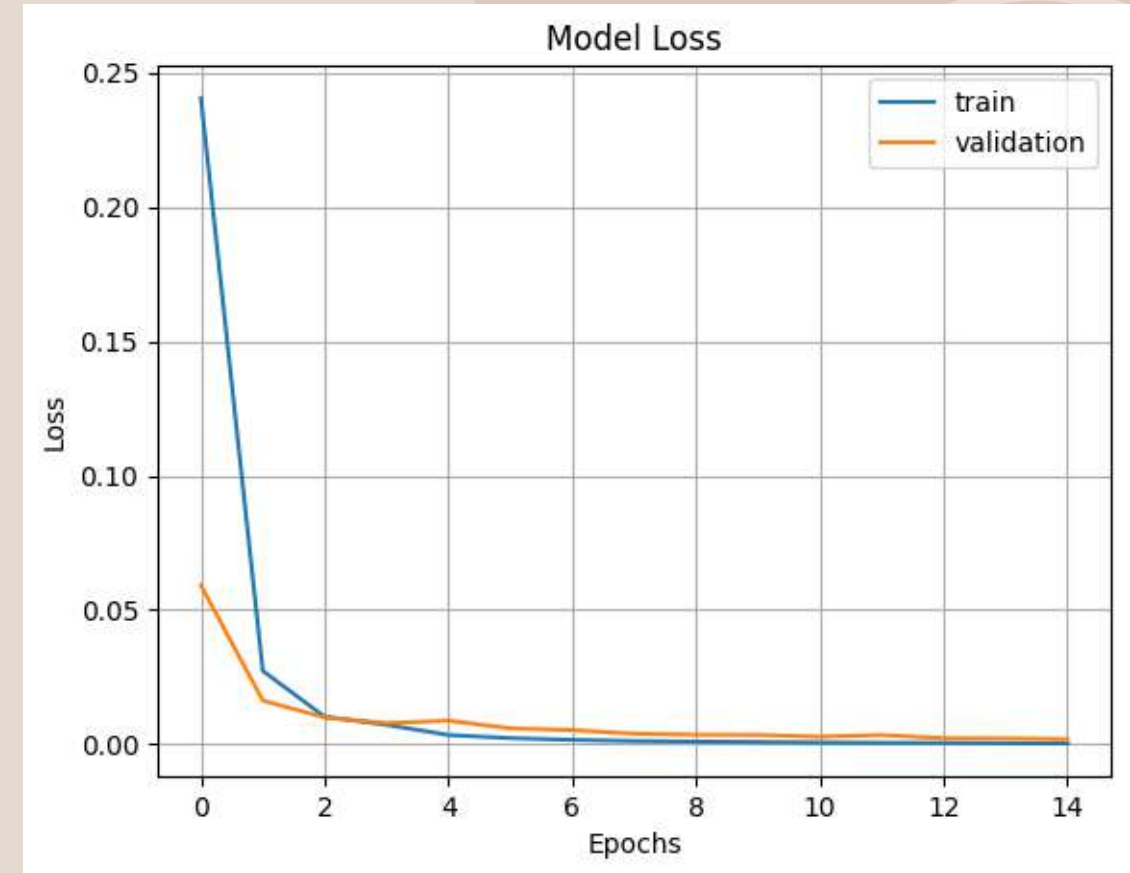
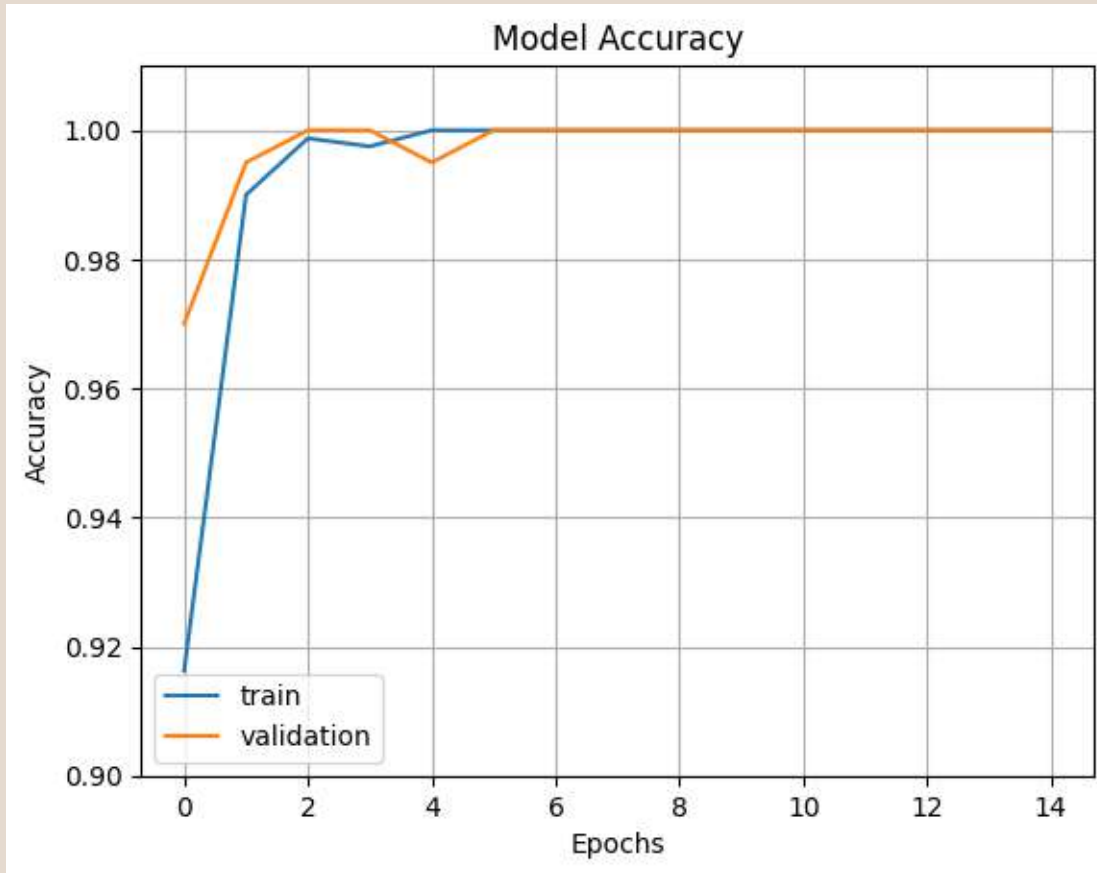
[ ]

# have to create a new folder called data and then add the folder manually

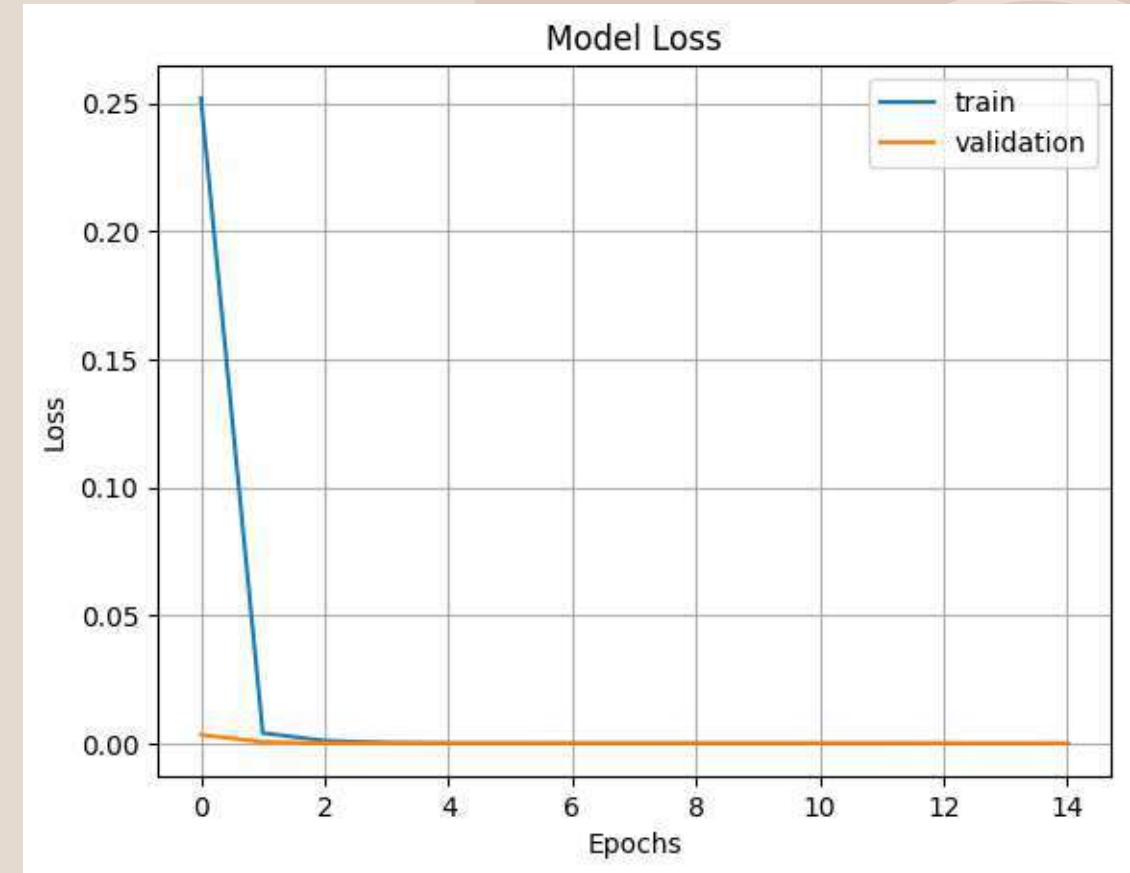
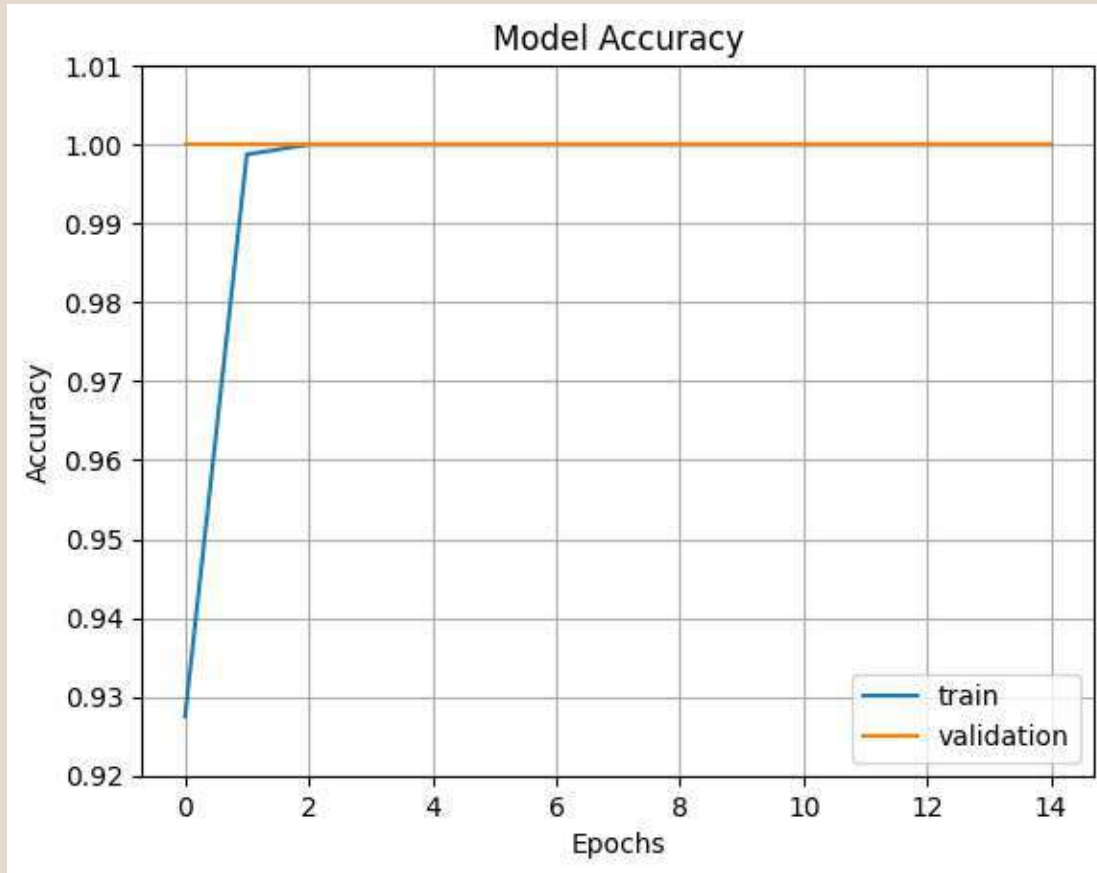
from pathlib import Path
data_dir = Path(r"/content/Data")

[ ] img_height, img_width=224,224
batch_size=32
train_ds = tf.keras.preprocessing.image_dataset_from_directory(
    data_dir,
    validation_split=0.2,
    subset="training",
    seed=123,
    image_size=(img_height, img_width),
    batch_size=batch_size)
```

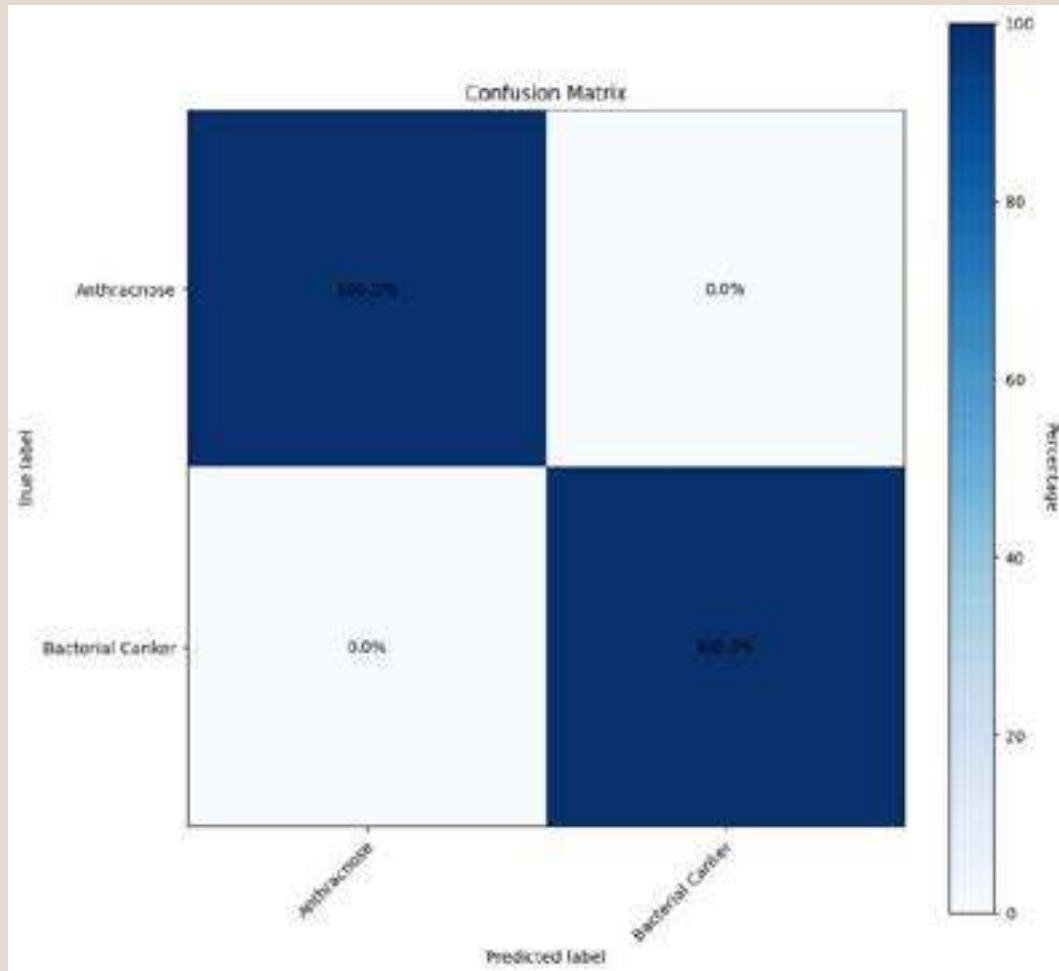

Completion of the model



Completion of the model



Confusion Matrix



	precision	recall	f1-score	support
Anthracnose	1.00	1.00	1.00	25
Bacterial Canker	1.00	1.00	1.00	25
accuracy			1.00	50
macro avg	1.00	1.00	1.00	50
weighted avg	1.00	1.00	1.00	50

Completed Implementations



- Collect Dataset.
- Train the Resnet50 ML model using 80% of the training data
- Adjusting and training the model to achieve high accuracy.
- Validate the model using the 20% of test data
- Identify the mango disease.

Future Implementations



- Add 3 more disease to train model and add to system.
- Suggest the prevention solution for the disease.
- Design and Develop the Mobile App.

IT20276928

Jayamanne B.D.N

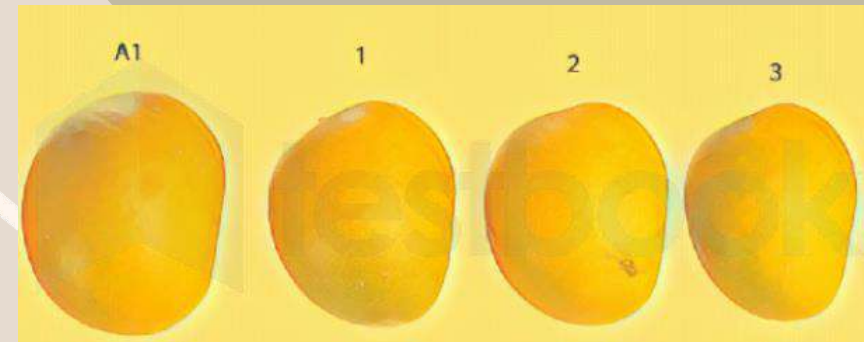


Specialization: Computer Systems and Network Engineering

IoT Based Mango Quality Grading System

Introduction

- Traditional mango grading methods are manual and subjective, with human experiences.
- In grading systems, mangoes are classified into different grades, such as “A, B, C” classes.
- The proposed system will aim to implement an IoT device and machine learning to develop an automated mango grading system.





RESEARCH QUESTION

- ✓ **How mango farmers increase their cost of mango production ?**
- ✓ **How to increase accuracy of mango grading ?**
- ✓ **How farmers reduce time consuming and workforce?**

Sub Objective



Implement of a machine learning model to identify the grade of mangos from the images.

An IoT device captures mango images and weighs them using a camera and weight sensor.

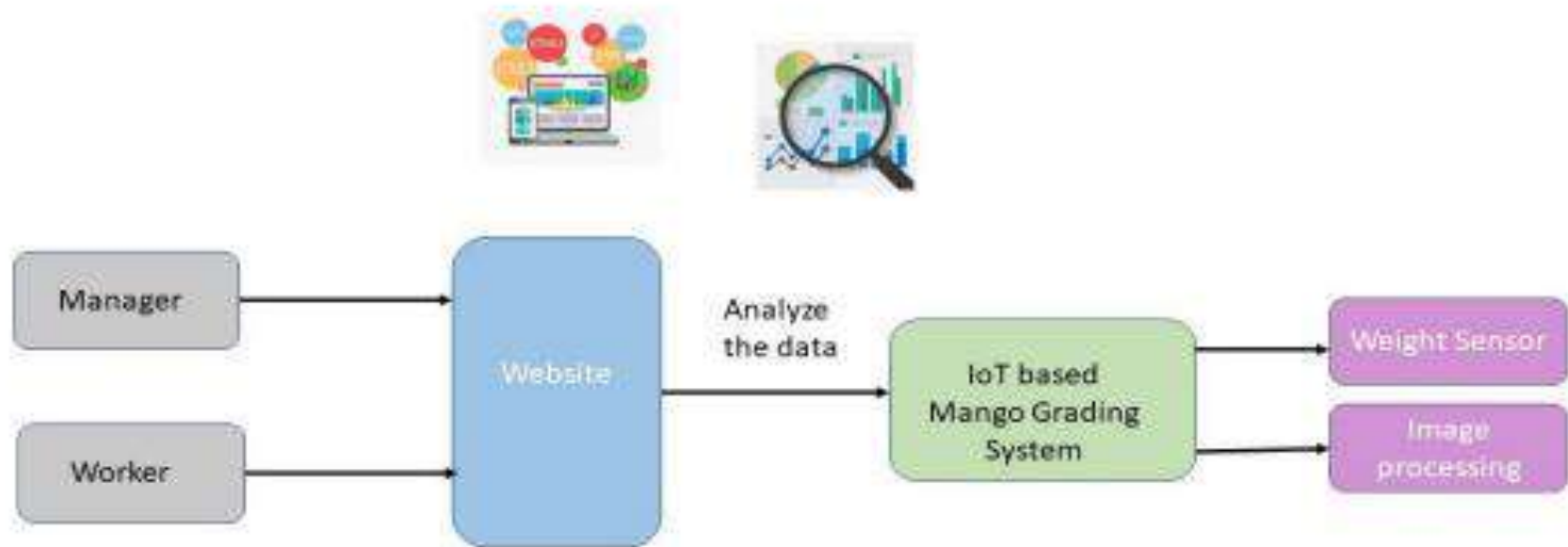
Methodology

A machine learning model is developed to classify mangoes based on their maturity stage and quality grade. The model is trained using the collected dataset of mango images.

An IoT device will be developed to capture images of mangoes and weigh them. The device will be equipped with a camera and a weight sensor, and it will be connected to a central server using wireless communication protocols such as Wi-Fi or Bluetooth.

The IoT device and machine learning model will be integrated to create the proposed mango grading solution. The IoT device will capture images and weight the mango, which will be analyzed by the machine learning model to determine its grade.

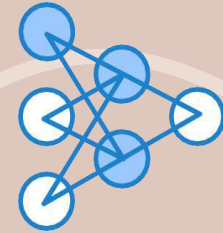
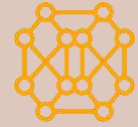
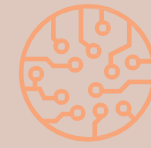
System Diagram



Technologies and Tools

- Python
- Google Colab
- CNN –Transfer Learning

- Python
- Raspberry Pi
- Load Sensor



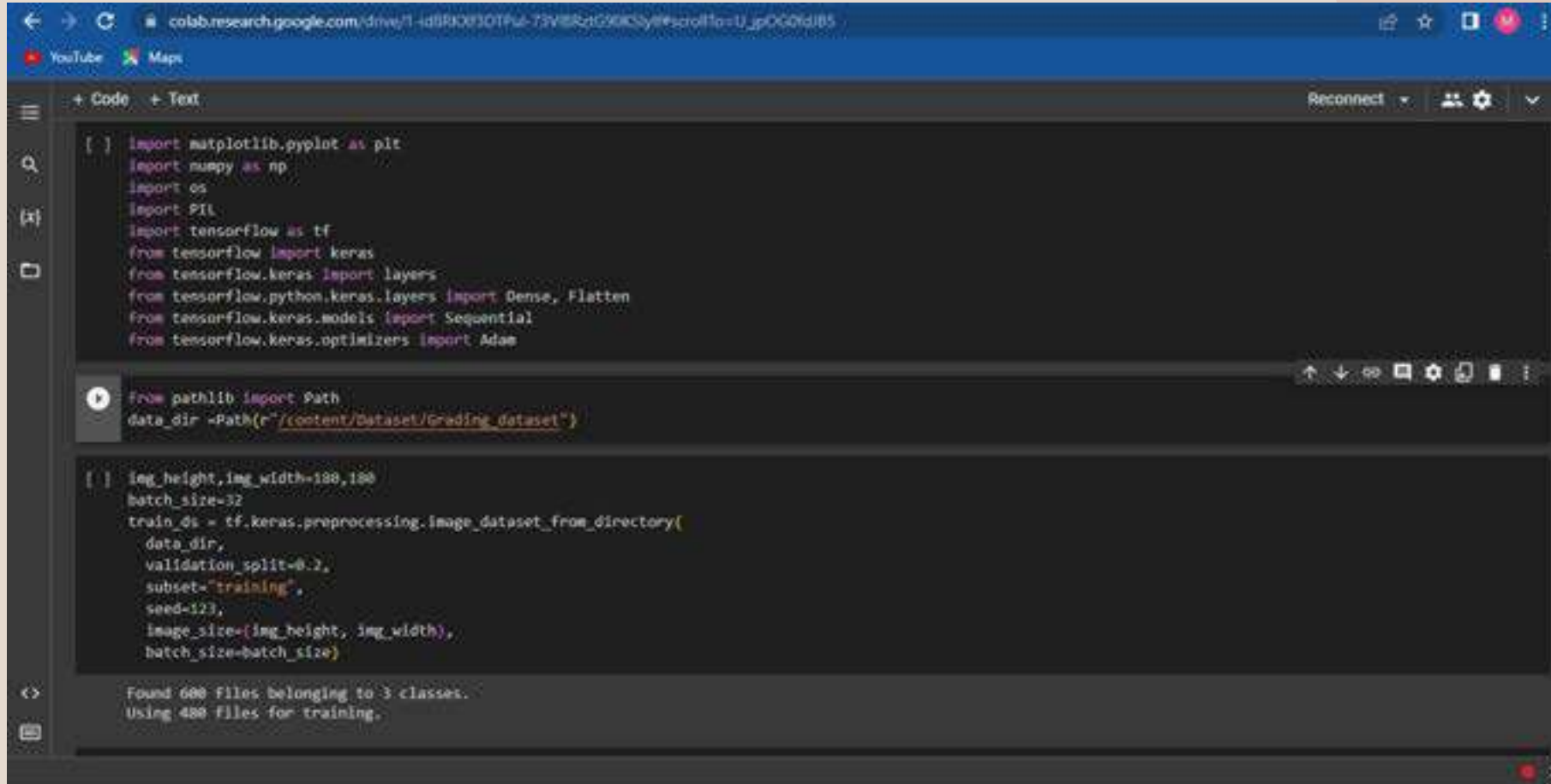
Collected Dataset

```
! unzip /content/mango-varieties-classification.zip

Archive: /content/mango-varieties-classification.zip
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_102834.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_102839.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_102859.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_102913.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_102920.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_102934.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_102950.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103000.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103004.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103018.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103027.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103032.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103040.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103046.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103050.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103118.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103128.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103131.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103135.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103138.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103141.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103144.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103155.jpg
  inflating: Dataset/Classification_dataset/Anwar Ratool/IMG_20210630_103158.jpg
```

Completion

Training the model

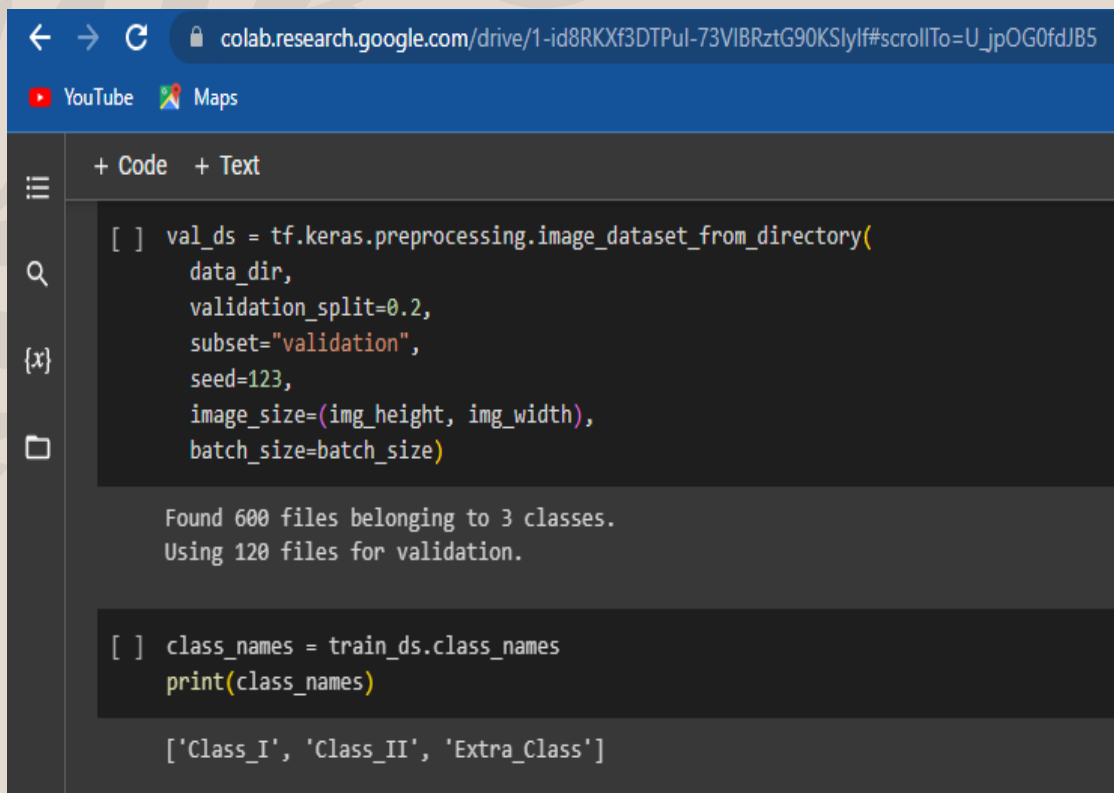


```
[ ] import matplotlib.pyplot as plt
import numpy as np
import os
import PIL
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from tensorflow.python.keras.layers import Dense, Flatten
from tensorflow.keras.models import Sequential
from tensorflow.keras.optimizers import Adam

from pathlib import Path
data_dir = Path(r"/content/Dataset/Grading_dataset")

[ ] img_height, img_width=180,180
batch_size=32
train_ds = tf.keras.preprocessing.image_dataset_from_directory(
    data_dir,
    validation_split=0.2,
    subset="training",
    seed=123,
    image_size=(img_height, img_width),
    batch_size=batch_size)

Found 600 files belonging to 3 classes.
Using 480 files for training.
```



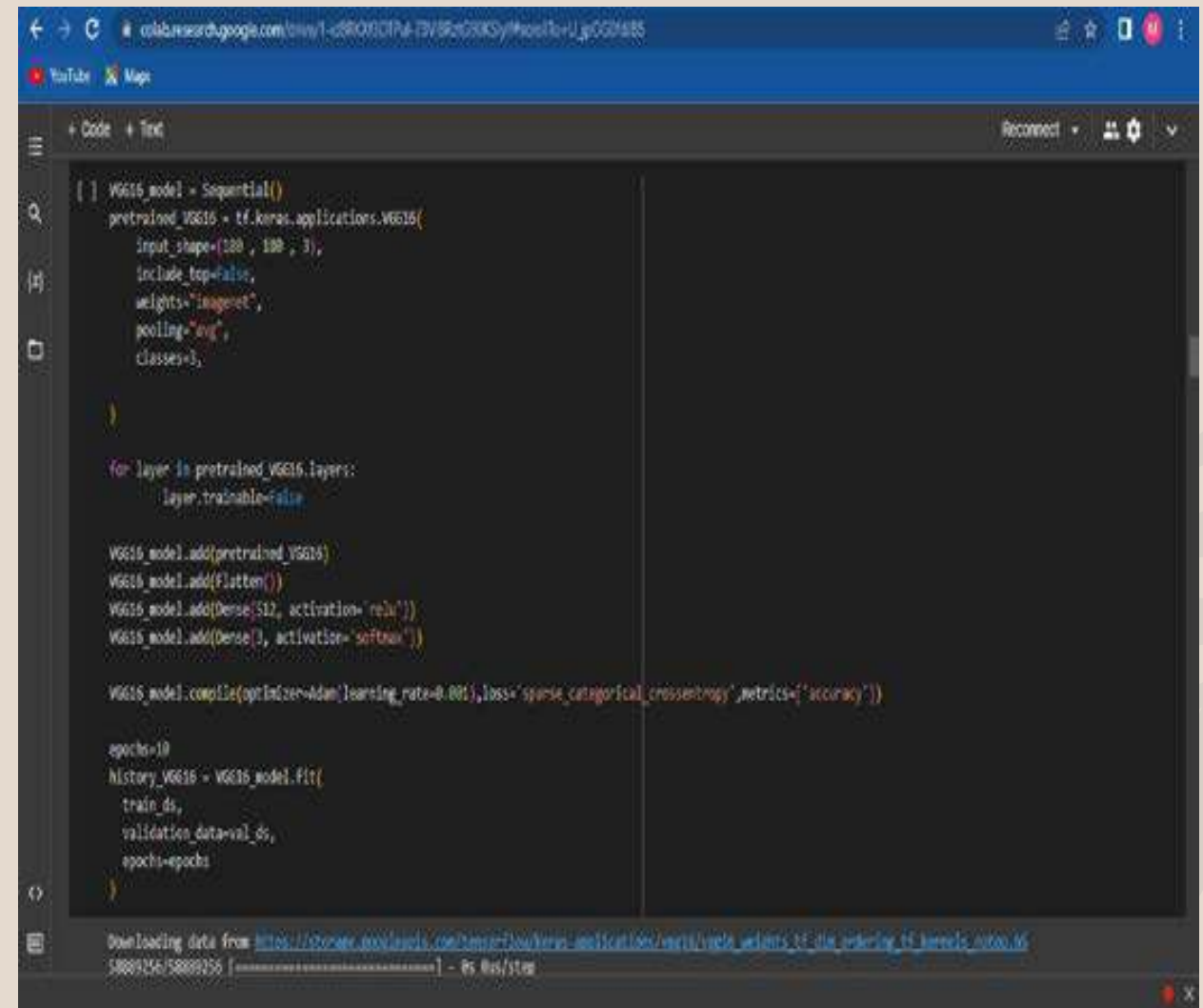
The screenshot shows a Google Colab interface with a code editor and an output area. The code defines a validation dataset from a directory. The output shows that 600 files were found belonging to 3 classes, with 120 files used for validation. The class names are printed as 'Class_I', 'Class_II', and 'Extra_Class'.

```
[ ] val_ds = tf.keras.preprocessing.image_dataset_from_directory(
    data_dir,
    validation_split=0.2,
    subset="validation",
    seed=123,
    image_size=(img_height, img_width),
    batch_size=batch_size)

Found 600 files belonging to 3 classes.
Using 120 files for validation.

[ ] class_names = train_ds.class_names
print(class_names)

['Class_I', 'Class_II', 'Extra_Class']
```



The screenshot shows a Google Colab interface with a code editor. The code constructs a VGG16 model using a pre-trained VGG16 base, adds a Flatten layer, and two Dense layers. The model is compiled with Adam optimizer and sparse categorical crossentropy loss. It is then trained for 10 epochs. The output at the bottom shows the progress of downloading data from the TensorFlow Hub.

```
[ ] VGG16_model = Sequential()
pretrained_VGG16 = tf.keras.applications.VGG16(
    input_shape=(128, 128, 3),
    include_top=False,
    weights='imagenet',
    pooling='avg',
    classes=3,
)

for layer in pretrained_VGG16.layers:
    layer.trainable=False

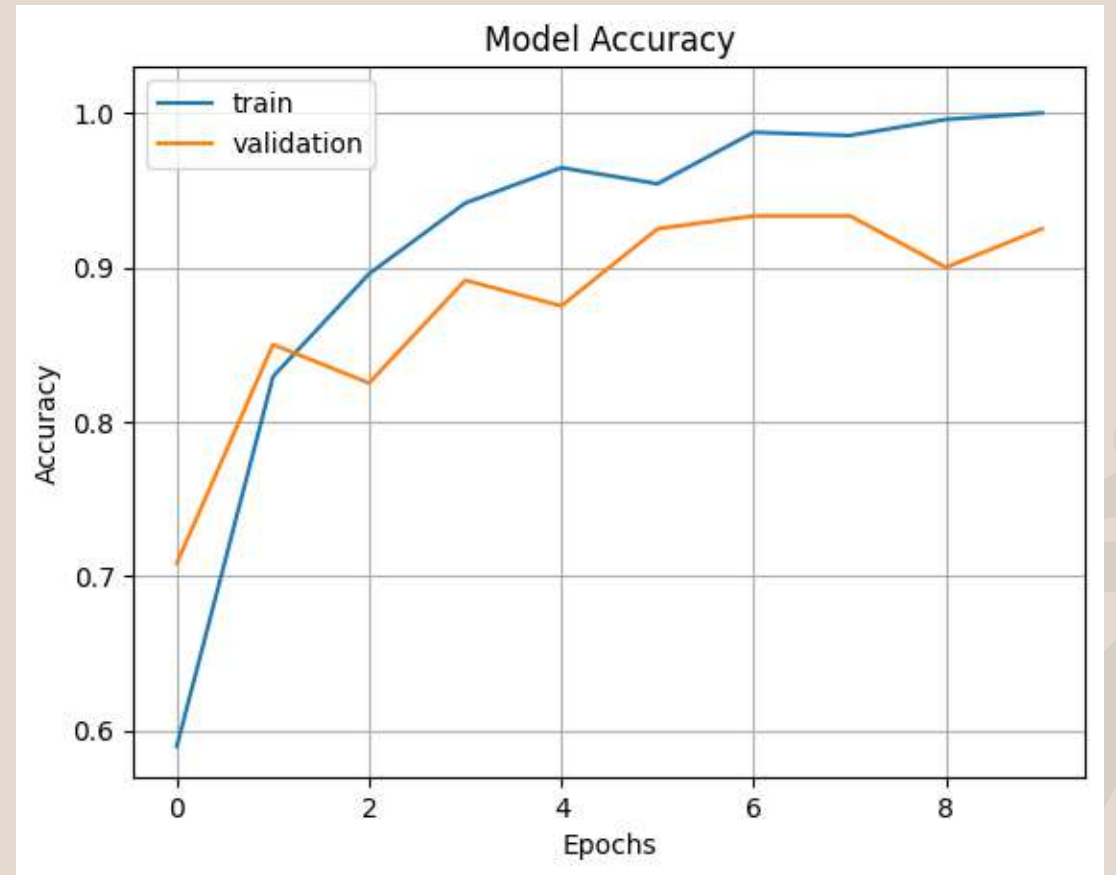
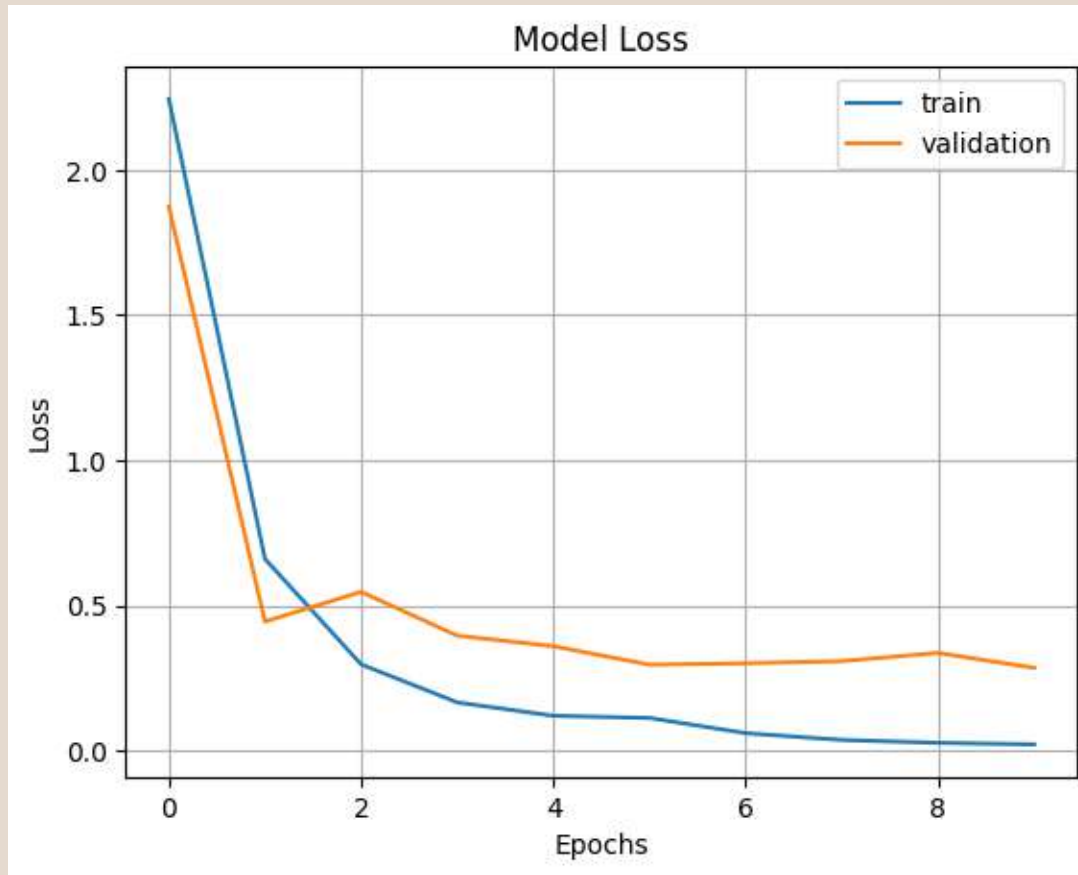
VGG16_model.add(pretrained_VGG16)
VGG16_model.add(Flatten())
VGG16_model.add(Dense(512, activation='relu'))
VGG16_model.add(Dense(3, activation='softmax'))

VGG16_model.compile(optimizer=Adam(learning_rate=0.001), loss='sparse_categorical_crossentropy', metrics=['accuracy'])

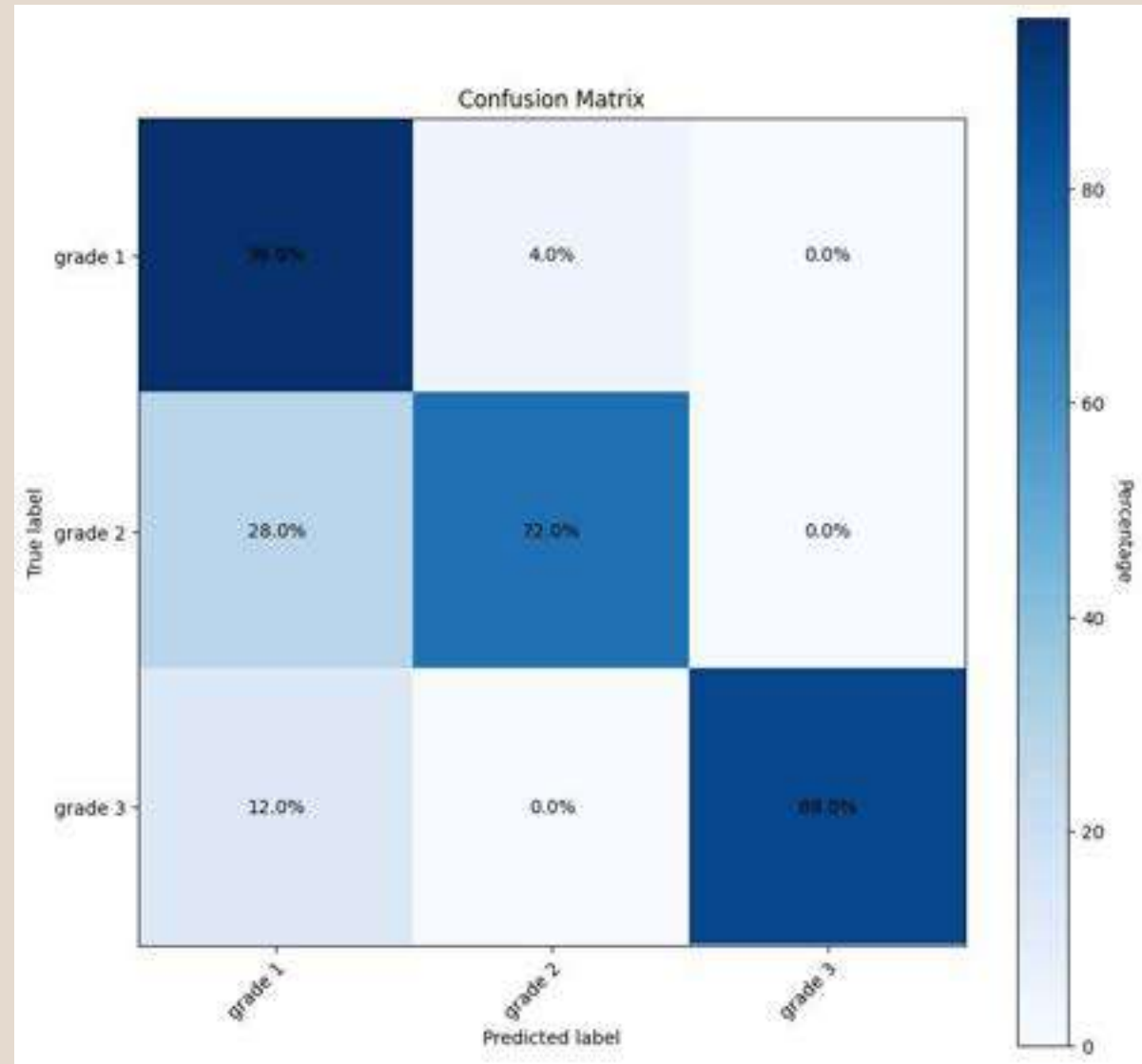
epochs=10
history_VGG16 = VGG16_model.fit(
    train_ds,
    validation_data=val_ds,
    epochs=epochs)

Downloading data from https://tfhub.dev/tensorflow/tf_keras_applications/vgg16_weights_tf_defaults_notop/1
58881256/58881256 [#####] - 85.0us/step
```

Accuracy And Loss



Confusion Matrix



Classification Report

```
[17] from sklearn.metrics import classification_report

target_names = ["grade 1", "grade 2", "grade 3"]

classification_rep = classification_report(y_true, y_pred, target_names=target_names)
print(classification_rep)
```

	precision	recall	f1-score	support
grade 1	0.71	0.96	0.81	25
grade 2	0.95	0.72	0.82	25
grade 3	1.00	0.88	0.94	25
accuracy			0.85	75
macro avg	0.88	0.85	0.86	75
weighted avg	0.88	0.85	0.86	75

Completed Implementations



- **Collect Dataset.**
- **Train the VGG16 ML model using 80% of the training data**
- **Validate the model using the 20% of test data**
- **Trained the ML model to identify the grade of mangos from the images.**

Future Implementations



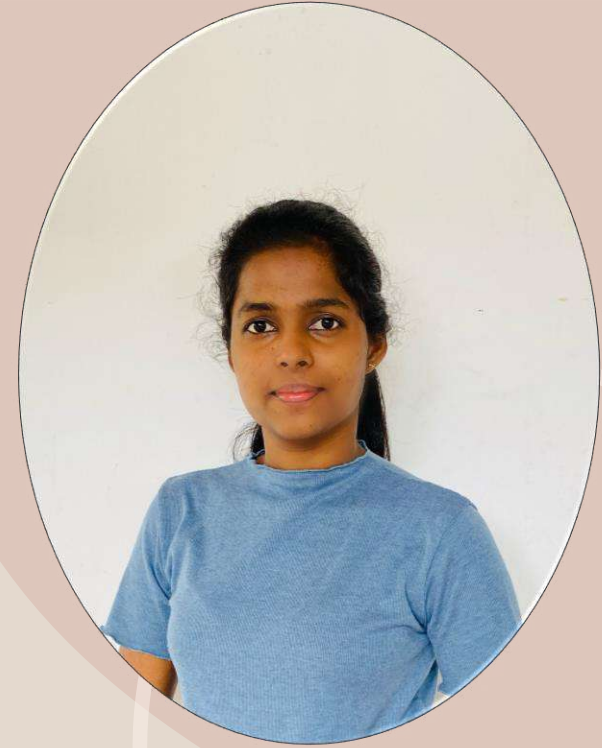
- **An IoT device will be developed to capture images of mangoes and weigh them.**
- **Implementing machine learning Resnet50 model to increase the accuracy of the system by adding weight and pre trained "mango grade" model**
- **Testing the finished system .**
- **Introduce the system to the clients and research how they interact.**

References

- [1] p. P. L. Yi, "Influences of Different Storage Conditions on Postharvest Quality of Mango," Influences of Different Storage Conditions on Postharvest Quality of Mango , 2019-09-26.
- [2] S. Krug, "AgriEnvironment," 02 May 2023. [Online]. Available: <https://www.mdpi.com/2624-7402/5/2/50>.
- [3] T. R. Razak, "Towards Capturing Mango Grading From Human Experts - A Comprehensive User Study," Towards Capturing Mango Grading From Human Experts - A Comprehensive User Study, 14 february 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9703830>. [Accessed 10 march 2023].
- [4] L. Pauly, "IEEE," 15 July 2015. [Online]. Available: <https://ieeexplore.ieee.org/document/7154891>.
- [5] A. K. R. K. A. M. R. K. Virender Singh, "Adoption of post-harvest management practices," 02 10 2020. [Online]. Available: https://www.researchgate.net/profile/Rajesh-Kumar314/publication/343627242_Adoption_of_postharvest_management_practicesby_Mango_growers_of_Haryana/links/5f34c659a6fdcccc43c5ac9e/Adoption-of-post-harvest-management-practicesby-Mango-growers-of-Haryana.pdf. [Accessed 5 May 2023].
- [6] [Online]. Available: <https://ja-si.com/gps-tracking-technology/>. [7] C. N. d. Ricerche, "Marking Standing Trees with RFID Tags," Consiglio Nazionale delle Ricerche, 29 January 2020. [8] "Development of higher yield and high-quality mango production system based on Internet of Things," 02 April 2019.

IT20103354 Niroshani A.

Specialization: Information Technology



Smart Mango Yield Prediction System

Introduction

- Mango is a highly valued fruit crop globally.
- The ability to predict mango yield in advance helps farmers to make informed decisions regarding resource allocation, harvest planning, and marketing strategies.
- In this research, we aim to develop a Regression model for predicting the yield of mango based on some important key factors.

Research Question

How can the yield of mango be predicted based on factors such as soil PH, soil moisture, temperature, humidity, rainfall, light exposure, life span, disease, and fertilizer usage?



Objective

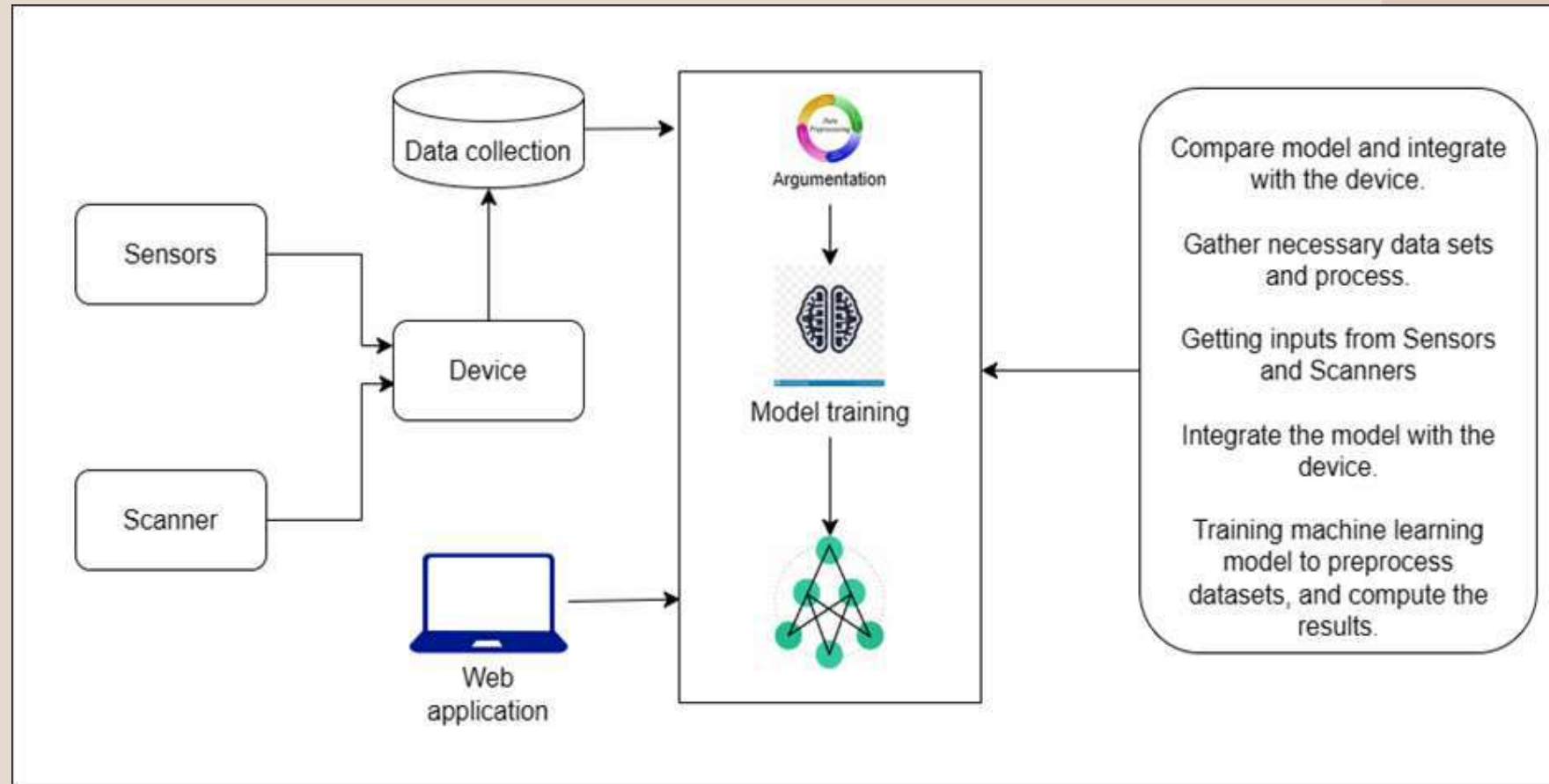
To develop a Linear Regression predictive model that accurately estimates mango yield based on the factors of soil pH, soil moisture, temperature, humidity, rainfall, light exposure, lifespan, disease, and fertilizer, thereby providing valuable insights for farmers to optimize mango cultivation and maximize crop productivity.

Methodology

- Collect data on various factors that affect mango yields such as soil PH, soil moisture, temperature, humidity, rainfall, light exposure, life span, disease, and fertilizer. And pre-processed collected data.
- Using linear regression machine learning algorithm creating machine learning models for Yield prediction with web application.
- The website will be used to update the real data to the user.



System Diagram



Tools and Technologies

Python

Google colab

Pandas

NumPy

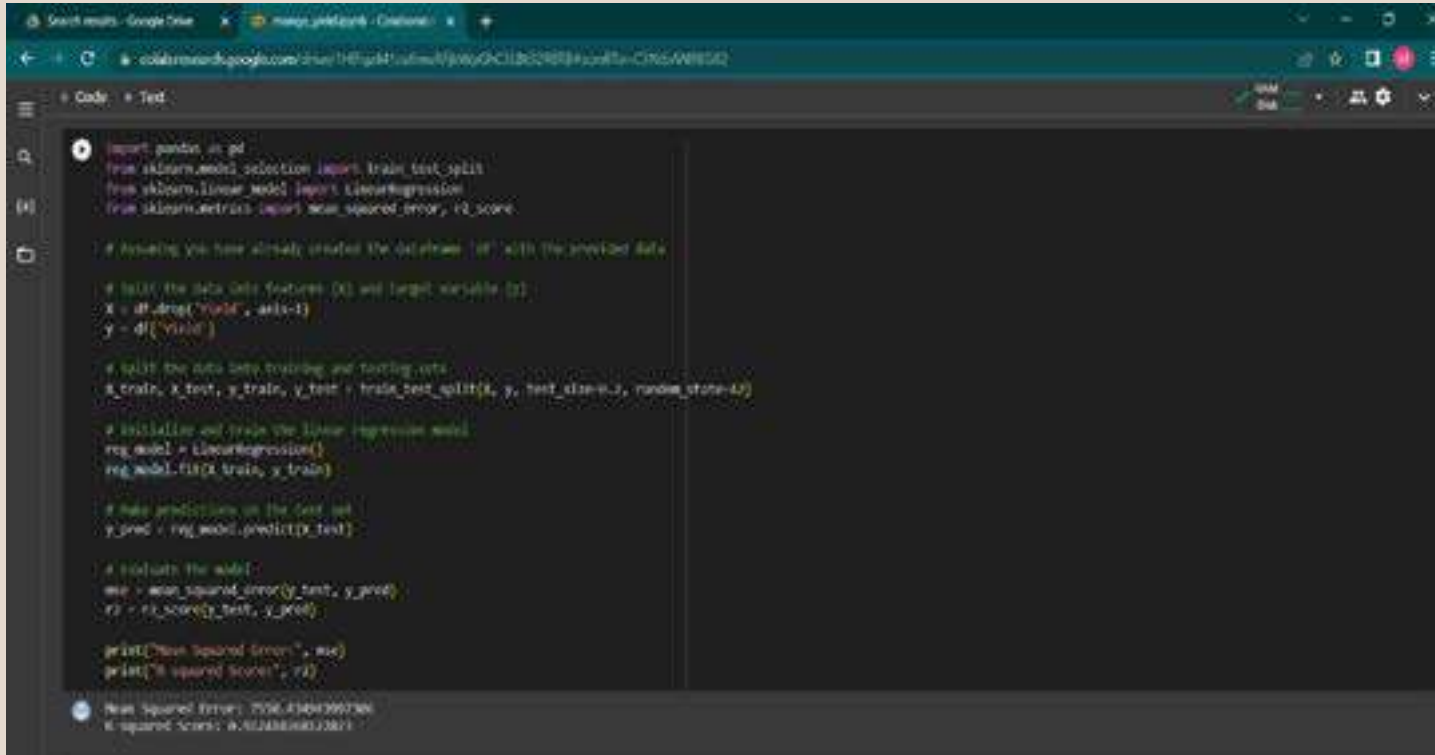


Dataset

A	B	C	D	E	F	G	H
Soil pH	Soil Moisture	Temperature	Humidity	Rainfall	Light Exposure	Life Span	Yield
5.592530449	0.4728880193	37.89890025	49.6772381	1615.167403	9.661759319	76.46022885	456.5401036
4.722204137	0.2805666112	39.29586185	30.31741978	477.5824121	2.978145635	15.21388495	177.6788351
5.761179642	0.09031500466	8.09894735	18.30529534	2308.356148	10.67439193	95.87837855	311.1653353
6.124665484	0.5108153096	15.42829299	80.39357692	2180.486764	18.09088132	62.56039669	936.6031243
7.43321794	0.6723658048	14.09918752	55.453696	1760.748061	7.972517802	95.674312	649.1567049
7.623064009	0.9103698222	16.14327768	85.84430575	164.6020406	7.217473809	73.46381903	894.9636337
5.479866675	0.3001829174	12.99905088	10.9145057	2224.788691	13.01585617	28.18743003	225.6819991
7.256988884	0.3890290181	11.02858617	38.16837812	1448.59849	8.432238809	87.13779701	474.9725568
5.421231874	0.05568393094	21.39826799	39.04128871	1959.500361	14.14087635	89.42770838	458.2993329
6.119803179	0.3771212659	30.17791638	30.99488975	245.86488	4.512982075	14.19323201	244.9998651
7.722803312	0.5286243783	7.731203222	43.45903234	1147.125455	11.06305942	53.5900184	559.0509233
7.638132346	0.5560486355	26.78913422	85.42210969	438.1980238	9.16069635	10.08463613	653.4152939
7.683748135	0.8988055854	17.4212079	36.98815192	959.736544	0.8993979323	53.73565776	392.8483072
6.055587327	0.360905887	17.4973084	23.36042732	2116.289734	15.84379229	60.9985506	335.8197158
7.433524626	0.5567551105	21.44282787	76.47709083	1922.348571	14.21867664	42.23757654	852.8916494
6.403297832	0.6234785439	19.79588143	61.05845258	807.0318554	5.287948644	23.51240441	833.0464441
6.199430288	0.3424612988	25.11189041	82.58851104	1177.978365	7.258874819	45.6960911	626.1284024
6.469795073	0.6838679104	18.48510694	38.65855024	1610.385712	14.3238067	33.61555429	477.739417
7.501090044	0.4011548044	17.80888858	18.82722757	1278.540171	8.891846046	33.38080304	238.9876333
7.101573373	0.5483300725	16.74036677	55.59217672	636.1400944	11.28774114	89.47674375	611.8053479
6.582336033	0.01142694861	30.11745498	42.65656026	670.365844	14.9236133	71.41460641	415.3007672
7.6994176	0.4714657831	10.83680166	79.42912891	2072.348325	12.45300168	44.10550931	935.1747855
8.28234654	0.9269928489	17.70369453	68.64614204	2132.486663	7.947347517	92.72425821	783.3004821
6.013083626	0.113663944	34.95816749	35.87499468	2141.878446	9.937777871	90.50459317	338.3323827
6.828172228	0.4978209407	14.78510421	43.04370102	399.3078733	6.163372469	23.87883026	458.0063279
6.135141348	0.5827830283	22.96930819	45.31336539	1074.430576	7.328254971	92.12324328	467.2765695
4.906093531	0.1178580029	39.38071961	12.37966896	658.0211342	1.623838713	73.44010024	-5.474208056
5.089430245	0.1248967242	40.12512647	25.49319027	990.4401709	12.62948138	65.77233079	188.0843468
6.753534958	0.3881913825	10.15943874	43.54870795	1606.087043	11.19301225	23.54681693	550.7772973
7.734238216	0.5559698016	36.70536399	46.58628833	1070.240656	7.224007232	33.19487529	408.1367724

- Collected a dataset with various factors that affect mango yield. (Agriculture department and jewelex agro plantation)

Train the ML Model



```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Assuming you have already loaded the dataset 'df' with the provided data

# Split the data into features (X) and target variable (y)
X = df.drop("House", axis=1)
y = df["House"]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Initialize and train the linear regression model
reg_model = LinearRegression()
reg_model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = reg_model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error:", mse)
print("R-squared Score:", r2)
```

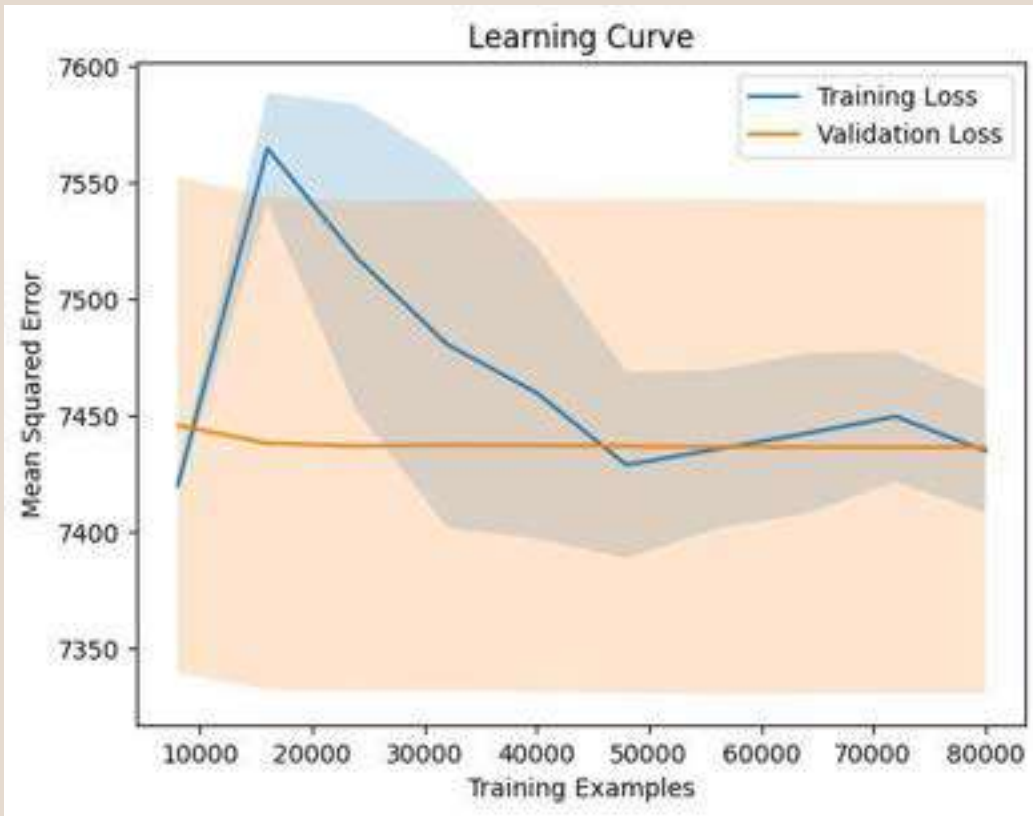
Mean Squared Error: 7506.43493997366
R-squared Score: 0.924886832871

- The Model trained using Google Colab online tool
- Separate dataset is used to validate the model

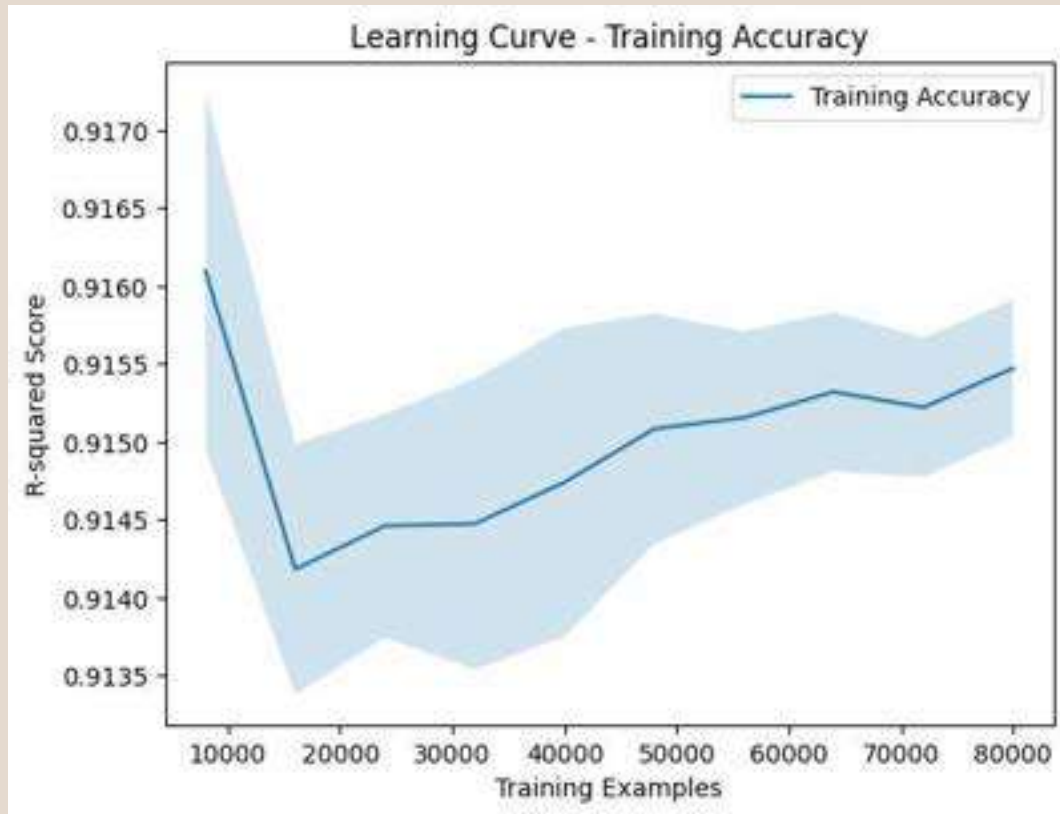
Learning Curve

Loss: 7550.434

Validation Loss: 7448.531



Learning Curve-Training Accuracy



Accuracy: 0.912418368122873

Test and load ML Model

Predicted Yield: 259.86 t/ha
(metric tons per hectare)

The screenshot shows a Google Colab notebook with the following code:

```

+ Code + Test
+ Test
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

report pandas is pd
report numpy is np
report joblib

+ load the trained ml model
model = joblib.load('/content/drive/MyDrive/Wagyu/Wagyu_project_2020/ml/20200404/ml/Wagyu_project/regression_model_yield.joblib')
model

+ define the synthetic data row
synthetic_data = pd.DataFrame({
    'sell price': (4.5),
    'sell window': (4.4),
    'temperature': (25.0),
    'humidity': (20.0),
    'rainfall': (200.0),
    'light exposure': (4.0),
    'life span': (4.0)
})

+ make predictions on the synthetic data row
predictions = model.predict(synthetic_data)

+ print the predicted yield
predictions
print('Predicted Yield: ', predictions[0])
  
```

The notebook interface shows the code is executed, with a status bar at the bottom indicating "28 completed at 10:04 PM". The Windows taskbar at the bottom shows the time as 11:51 AM on 5/24/2020.

Functional Requirements

- The user should be able to generate accurate and reliable predictions.
- The system should be designed to handle a large volume of data and provide timely predictions.
- The system should have an intuitive and user-friendly interface that allows users to easily manage.

Completed Implementations



- Collect Dataset.
- Train the Linear regression ML model using 80% of the training data
- Adjusting and training the model to achieve high accuracy.
- Validate the model using the 20% of test data
- Predict the mango yield.

Future Implementations



- Add disease and fertilizer parameters as independent variables.
- Implement the real time data input feature.
- Design and Develop the Frontend.

The background features a light gray base with large, soft-edged organic shapes in muted red and olive green. A thin white line outlines a shape on the right. In the top left, there is a faint, stylized illustration of a leafy branch.

Thank you