CROP PREDICTION SYSTEM USING ML AND ANALYSIS OF SOIL NUTRIENTS USING SATELLITE IMAGING AND IOT

A PROJECT REPORT

Submitted by

DHINESHBALAN N (2020115028)

SASI KUMAR K (2020115076)

submitted to the Faculty of

INFORMATION AND COMMUNICATION ENGINEERING

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI 600 025

MAY 2024

ANNA UNIVERSITY

CHENNAI - 600 025

BONA FIDE CERTIFICATE

Certified that this project report titled ” CROP PREDICTION SYSTEM USING ML AND ANALYSIS OF SOIL NUTRIENTS USING SATELLITE IMAGING AND IOT ” is the bonafide work of DHINESHBALAN N (2020115028), SASI KUMAR K (2020115076) who carried out project work under my supervision.Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE:

DATE:

Dr. G. GEETHA
ASSOCIATE PROFESSOR
PROJECT GUIDE
DEPARTMENT OF IST, CEG
ANNA UNIVERSITY
CHENNAI 600025

COUNTERSIGNED

Dr. S . SWAMYNATHAN
HEAD OF THE DEPARTMENT
DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY
CHENNAI 600025

ABSTRACT

In the agricultural sector, there is a high demand for incorporating and promoting the use of technology in agricultural fields and also agriculture plays a pivotal role in global food security but faces mounting challenges amidst population growth and environmental pressures.In response to these evolving needs, this project uses the concept of machine learning to recommend the right crop to plant and our project also combines technologies like IOT and satellite imagery to analyze the soil nutrients.

Our project aims to provide personalized crop recommendations tailored to specific farming conditions and promote using innovative technologies for agriculture.This empowers the farmers to have the power of decision making and maximize crop yields while minimizing resource usage.

Machine Learning algorithms such as SVM, random forest and LGBM are used to train a model which can predict the required crop based on the inputs given to them.Our project includes satellite imagery by which the soil nutrients are calculated with the help of their band values, and soil sensors are used to analyze the nutrients in the soil which will be given as the input for the ML Model.

Our project utilize tools such as jupyter noteook, google earth engine, thinkspeak, aurdino IDE to recommend the suitable crop type and analyze the soil nutrients in the field.This project helps to implement using data-driven approaches to optimize crop selection and promote sustainable farming practices. Going forward, we aim to refine our models, expand our system, and collaborate with others to continue improving agriculture with technology.

திட்டப்பணிச்சுருக்கம்

விவசாயத் துறையில் தொழில்நுட்பத்தைப் பயன்படுத்துவதை ஒருங்கிணைப்பதற்கும் ஊக்குவிப்பதற்கும் அதிக தேவை உள்ளது, மேலும் உலகளாவிய உணவுப் பாதுகாப்பில் விவசாயம் முக்கிய பங்கு வகிக்கிறது, ஆனால் மக்கள் தொகை வளர்ச்சி மற்றும் சுற்றுச்சூழல் அழுத்தங்களுக்கு மத்தியில் பெருகிவரும் சவால்களை எதிர்கொள்கிறது. இந்த வளர்ந்து வரும் தேவைகளுக்கு பதிலளிக்கும் வகையில், இந்த திட்டம் இயந்திர கற்றல் என்ற கருத்தைப் பயன்படுத்தி நடவு செய்வதற்கு சரியான பயிரை பரிந்துரைக்கிறது, மேலும் எங்கள் திட்டமானத மண்ணின் சத்துக்களை பகுப்பாய்வு செய்ய ஐஓடி மற்றும் செயற்கைக்கோள் படங்கள் போன்ற தொழில்நுட்பங்களையும் ஒருங்கிணைக்கிறது.

எஸ்.வீ.எம் , ரேண்டம் ஃபாரஸ்ட் மற்றும் லைட் ஜிபிஎம் போன்ற இயந்திர கற்றல் வழிமுறைகள் ஒரு மாதிரியைப் பயிற்றுவிக்கப் பயன்படுத்தப்படுகின்றன, இது அவர்களுக்கு கொடுக்கப்பட்ட உள்ளீடுகளின் அடிப்படையில் பரிந்துரைக்கப்பட்ட பயிரைக் கணிக்க முடியும். எங்கள் திட்டத்தில் செயற்கைக்கோள் படங்கள் உள்ளன, இதன் மூலம் மண்ணின் சத்துக்கள் அவற்றின் பேண்ட்மதிப்புகளின் உதவியுடன் கணக்கிடப்படுகின்றன, மேலும் மண்ணில் உள்ள ஊட்டச்சத்துக்களை பகுப்பாய்வு செய்ய மண் உணரிகள் பயன்படுத்தப்படுகின்றன, அவை இயந்திர கற்றல் மாதிரிக்கான உள்ளீடாக வழங்கப்படும்.

எங்கள் திட்டமானது குறிப்பிட்ட விவசாய நிலைமைகளுக்கு ஏற்ப தனிப்பயனாக்கப்பட்ட பயிர் பரிந்துரைகளை வழங்குவதையும் விவசாயத்திற்கான புதுமையான தொழில்நுட்பங்களைப் பயன்படுத்துவதை ஊக்குவிப்பதையும் நோக்கமாகக் கொண்டுள்ளது. இது விவசாயிகள் முடிவெடுக்கும் ஆற்றலைப்பெறவும், வளப் பயன்பாட்டைக் குறைக்கும் அதே வேளையில் பயிர் விளைச்சலை அதிகரிக்கவும் உதவுகிறது.

இந்த தீர்வு விவசாயிகளுக்கு தகவலறிந்த முடிவுகளை எடுக்க அதிகாரம் அளிக்கிறது மற்றும் உணவு கிடைப்பதை அதிகரிக்க உதவுகிறது, இது விவசாயத்தில் தொழில்நுட்பத்தை பின்பற்றுவதை ஊக்குவிக்கிறது. இந்த திட்டத்தின் முக்கிய குறிக்கோள், மண்ணின் ஊட்டச்சத்து மற்றும் மண்ணின் சத்துக்களை பகுப்பாய்வு செய்வதன் அடிப்படையில் பயிர்களின் துல்லியமான கணிப்புகளை வழங்குவதாகும். செயற்கைக்கோள் இமேஜிங்கைப் பயன்படுத்துவதால், மண் மாதிரிகளைச் சேகரிக்கத் தேவைப்படும் நேரத்தையும் பயணத்தையும் குறைக்க இந்தத் திட்டம் முக்கியமாக உதவுகிறது. வேளாண் துறையில் மேம்பட்ட தொழில்நுட்பங்களைச் செயல்படுத்துவது விவசாயிகள் மற்றும் விவசாய ஆலோசகர்கள் போன்ற முதன்மை இறுதி பயனர்களுக்கு உதவுகிறது.

ACKNOWLEDGEMENT

We wish to record our deep sense of gratitude and profound thanks to our project guide Dr. G. GEETHA, Associate Professor, Department of Information Science and Technology, College of Engineering, Guindy for her keen interest, inspiring guidance, constant encouragement with our work during all stages, to bring this project into fruition.

We would like to convey our gratitude to Dr. S. SWAMYNATHAN, Professor and Head, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for providing us with the opportunity and infrastructure to carry out this project.We would also like to express our sincere thanks to Dr. S. SWAMYNATHAN, Professor and Project Coordinator, and the panel of reviewers Dr. P. VARALAKSHMI, Professor, Dr. S. SENDHIL KUMAR, Professor, Dr. T. J. VIJAY KUMAR, Teaching Fellow, Department of Information Science and Technology for their valuable suggestions throughout the course of our project.

We thank our parents, family, and friends for bearing with us throughout the course of our project and for the opportunity they provided us in undergoing this course in such a prestigious institution.

DHINESHBALAN N
SASI KUMAR K

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GEE	Google Earth Engine
IOT	Internet of Things
LGBM	Light Gradient Boosting Machine
ML	Machine Learning
NPK	Nitrogen, Phosphorus, Potassium
SVM	support vector machine
S1	Sentinel-1

CHAPTER 1

INTRODUCTION

Agriculture is the mainstay of the Indian economy.More than 50 percent of the population in India relies on agriculture as their livelihood and in this over 27 percentage people follow improper and traditional methods of agriculture.Modern farmers face numerous challenges such as uncertain weather patterns, fluctuating market demands, and the need for sustainable agricultural practices. Current farmers encounter several challenges that impact productivity and profitability.Farmers often lack access to real-time data on soil nutrient levels and struggle with making informed decisions about crop selection and soil management.

Our primary focus is the development of a robust ML model capable of predicting suitable crop types based on crucial agronomic factors.By incorporating inputs such as soil nutrient levels (Nitrogen - N, Phosphorus - P, Potassium - K, pH), historical rainfall data, and moisture levels, our ML model will empower farmers with valuable insights for informed decision-making regarding crop selection and management practices.This helps to integrate technology in agricultural field.

In addition to ML, we are implementing an IoT-based soil nutrient measurement system to enable real-time monitoring and analysis of critical soil parameters directly in the field. By deploying IoT sensors capable of collecting accurate data, farmers will gain actionable insights for precise nutrient management and soil health optimization, contributing to overall agricultural sustainability and productivity.

1.1 BACKGROUND

At present in agriculture involves addressing challenges faced by the agricultural sector, such as climate change, population growth, and resource depletion.Traditional methods of crop prediction are inefficient and prone to inaccuracies, prompting the need for innovative technologies like machine learning algorithms, satellite imagery, and IoT devices.These technologies offer valuable insights into crop growth, health, and yield, enabling informed decisions about crop management, irrigation, fertilization, and pest control.The integration of machine learning, satellite imagery, and IoT technologies has the potential to revolutionize crop prediction and agricultural management, empowering farmers with data-driven insights to optimize resource allocation and mitigate climate-related risks.The background emphasizes the importance of leveraging advanced technologies to enhance efficiency, sustainability, and resilience in the agricultural sector amidst evolving global challenges.

1.2 OBJECTIVES

The main objective of this project is to provide accurate prediction of crop based on soil nutrients for the farmers and agricultural consultants.To provide valuable real time analysis of the soil nutrients present in the field, which helps a lot to grow healthy crops and optimize yields.

The implementation of advanced technologies such as satellite imaging which will take the agricultural sector to a new stage of development.This also helps to save time, travel and manpower as analysis of soil nutrients can be done with the help of satellite imaging only by analyzing the images of the field.

Develop a comprehensive data-driven model leveraging satellite imagery, IoT sensor data, and machine learning algorithms to accurately predict crop suitability based on soil nutrient analysis, providing farmers and agricultural consultants with actionable insights for optimized crop management and increased yields.

The integration of machine learning algorithms, satellite imagery, and IoT devices for crop prediction represents a transformative shift in agricultural practices. Machine learning enables automated and data-driven crop predictions, while satellite imagery offers detailed insights into crop health and environmental conditions.IoT devices provide real-time data for precision agriculture. Moving forward, addressing data quality, cost-effectiveness, and fostering interdisciplinary collaborations are key prediction for advancing research and development in crop prediction using these technologies, ensuring their effective integration into agricultural policies and practices for a more resilient and productive global food system.

1.3 PROBLEM STATEMENT

Traditional methods of crop prediction rely on manual observations without considering soil nutrients, which are often inefficient and prone to inaccuracies.

To develop accurate crop prediction and analysis of soil nutrients in order to optimize management practices in agriculture.

1.4 SOLUTION OVERVIEW

To address the challenges and opportunities of using machine learning algorithms, satellite imagery and IoT devices for crop prediction, this report proposes a comprehensive solution framework that encompasses the following components:

The solution framework is to collect and manage high-quality and relevant data from various sources, such as crop recommendation dataset, IoT devices, satellite images, and soil sensors.The data should be analyzed and used for the crop prediction based on the inputs such as soil nutrients like nitrogen, potassium, phosphorous, temperature, humidity, pH and rainfall data.

Machine Learning Algorithms are used on the collected data to predict crop and optimize agricultural practices.The report reviews different types of machine learning algorithms, such as SVM, Random forest, Gaussian Naïve Bayes, Decision tree and LGBM and their potential applications in crop prediction.

The report also evaluates the performance and accuracy of different machine learning algorithms in predicting crop and optimizing resource allocation.

Satellite Imagery is used to find the environmental conditions and nutrients in the soil.The report uses different satellite imagery sources, such as Landsat and Sentinel to use their potential for the practical applications in crop prediction and agricultural management.The report also explores the use of advanced image processing and analysis techniques, such as spectral indices, machine learning algorithms, and deep learning models, to extract valuable insights and information from satellite imagery.

IoT devices to collect real-time data on soil moisture, temperature, humidity, and other environmental factors that affect crop growth and yield. The report reviews the types and characteristics of different IoT devices, such as sensors, actuators, and gateways, and their potential applications in precision agriculture and crop prediction.The report also evaluates the challenges and opportunities of integrating IoT devices in agricultural practices.

1.5 ORGANIZATION OF THE REPORT

The Organization of the report is as follows:

Chapter 1 formally introduces the problem statement of the project and the overview of the solution that is carried out.

Chapter 2 provides the details about the Literature Survey carried out before the commencement of the tasks. This chapter tells the ideas relevant and referred from various journal papers published related to this domain. The chapter concludes by providing a comprehensive summary of the Literature Survey carried out.

Chapter 3 delves into the System Design of the proposed solution for the tasks. This chapter elaborates on the Technical Architecture of the proposed solution, focusing on various modules for each task. As a final section of the chapter, a complete overview of the system design is summarized.

Chapter 4 focuses on the overall project's implementation.

Chapter 5 focuses on the result of the project.

Chapter 6 briefly signifies the conclusion of the project undertaken and discusses the future work of the project.

CHAPTER 2

LITERATURE SURVEY/RELATED WORK

This chapter delves into an extensive exploration of existing research and developments in the domains of Agriculture. It gives an overview of the challenges involved in developing the overall application.

2.1 INTERNET OF THINGS-ENABLED SOIL NUTRIENT ANALYSIS

Murali Krishna Senapaty et.al[1] has proposed an IoT-enabled soil nutrient analysis and crop recommendation (IoTSNA-CR) model for precision agriculture. The main goal is to develop a system that allows real-time data collection and analysis of soil properties, enabling farmers to make informed decisions about crop selection and soil health management.The authors has emphasize the importance of integrating IoT sensors, mobile devices, and cloud data analysis to contribute to sustainability in agriculture.

2.2 SOIL TEXTURE ESTIMATION

Safa Bousbih et.al [2] has proposed an to develop an approach for topsoil clay content estimations based on the interpretation of multi-sensor satellite data, focusing on a semi-arid area in central Tunisia.The study uses Sentinel-1 (S-1) and Sentinel-2 (S-2) data acquired between July and early December 2017, along with field measurements of soil texture.Algorithms based on support vector machine (SVM) and random forest (RF) methods are proposed for the classification and mapping of clay content, with the RF algorithm found to have the most accurate results.The study provides valuable insights into the potential advantages of using combined optical and radar data for soil texture monitoring, especially in areas where optical sensors may be limited by weather conditions or vegetation cover. Real-world experiments and simulations demonstrate the effectiveness of the proposed approach, with the results showing promise for generating soil texture maps at the field scale.The guidance provided by the study suggests that remote sensing is an efficient tool for estimating soil texture and could be further explored and tested in different climatic conditions and for mapping other soil components such as sand content

2.3 SATELLITE IMAGING

Chaitanya B. Pande et.al [3] the Center for Advanced Agricultural Science and Technology for Climate Smart Agriculture and Water Management (CAAST- CSAWM) at MPKV, Rahuri, Maharashtra, India, aims to develop a model for accurately mapping soil chemical parameters using multispectral satellite images and wavelet transform methods. The study focuses on the development of prediction models for soil chemical properties, such as organic carbon, pH, and EC, using wavelet transformation methods and multispectral satellite images. The authors use correlation analysis and validation strategies to assess the performance of the prediction models. The study also discusses the use of machine learning programming for precision farming and agriculture-related activities.

2.4 CROP YIELD PREDICTION

Sonal Agarwal et.al [4] has developed a hybrid approach for predicting crop yield using machine learning and deep learning algorithms. The main objective is to provide farmers with a system that can predict the most productive crop to grow under specific conditions, while also considering the expenses involved. The study focuses on analyzing data related to soil ingredients, climatic conditions, and other parameters to accurately predict crop yields. The proposed model utilizes machine learning algorithms such as SVM, along with deep learning algorithms like LSTM and RNN, to achieve better accuracy in crop yield prediction.

2.5 SUMMARY OF LITERATURE

Studies have advanced agricultural practices using modern technologies. One focused on estimating soil texture in a semi-arid area using radar and optical data from Sentinel-1 and Sentinel-2 satellites, employing SVM and RF algorithms for accurate clay mapping. Another proposed an IoT-enabled model aiding farmers in crop selection and soil health management through real-time data analysis. A study from India used multispectral satellite images and wavelet transforms to predict soil parameters like organic carbon and pH, highlighting the potential of machine learning for precision farming. Lastly, a hybrid ML-DL approach integrated soil and climate data to predict crop yields efficiently. These studies demonstrate technology's impact on soil monitoring, nutrient analysis, and crop yield prediction for sustainable farming practices.

CHAPTER 3

SYSTEM DESIGN

This chapter consists of the system design of the project with the technical architecture and various individual modules and their respective description used in this proposed project

3.1 SYSTEM ARCHITECTURE

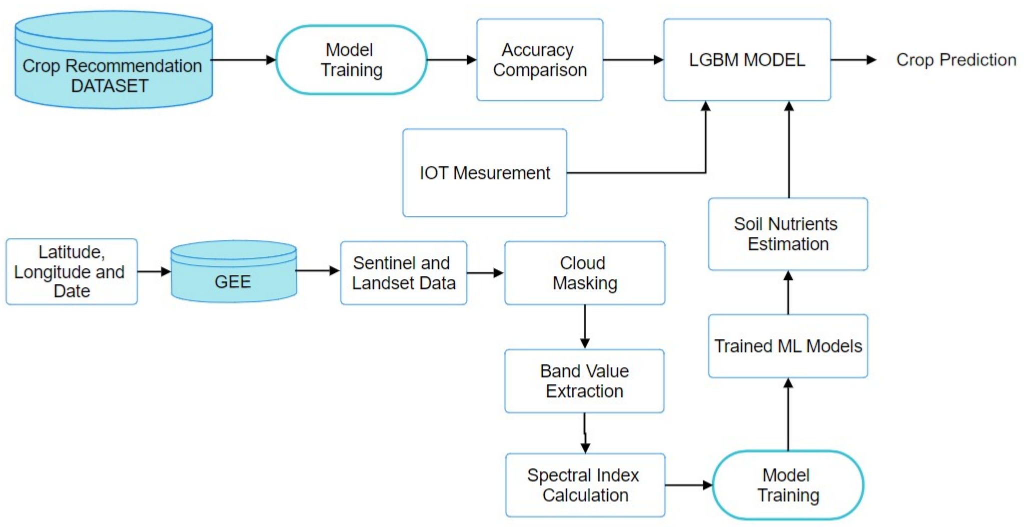

Figure 3.1: Architecture of Crop Prediction System

3.2 CROP RECOMMENDATION DATA SET

The crop prediction model was trained on a dataset comprising 2200 samples of 22 different crops, with each crop represented by 100 samples.The dataset used for our project is taken from Harvard dataverse (https://dataverse.harvard.edu/file.xhtml?fileId=6881146version=1.0toolType= PREVIEW).

Crop	Quantity (in kg)
rice	100
maize	100
jute	100
cotton	100
coconut	100
papaya	100
orange	100
apple	100
muskmelon	100
watermelon	100
grapes	100
mango	100
banana	100
pomegranate	100
lentil	100
blackgram	100
mungbean	100
mothbeans	100
pigeonpeas	100
kidneybeans	100
chickpea	100
coffee	100

Table 3.1: Crop Quantities

The dataset includes 7 features: soil nutrient levels (nitrogen, phosphorus, potassium, pH), along with temperature, humidity, and rainfall data.This balanced dataset allows comprehensive training of the machine learning model, ensuring equitable representation of each crop category. By leveraging these features, the model aims to make accurate predictions for optimal crop prediction based on soil conditions and environmental factors.The dataset's balance and diverse feature set contribute to robust and reliable crop prediction outcomes.

	A	B	C	D	E	F	G	H
1	N	P	K	temperatu	humidity	ph	rainfall	label
2	90	42	43	20.8797	82.0027	6.50299	202.936	rice
3	85	58	41	21.7705	80.3196	7.0381	226.656	rice
4	60	55	44	23.0045	82.3208	7.84021	263.964	rice
5	74	35	40	26.4911	80.1584	6.9804	242.864	rice
6	78	42	42	20.1302	81.6049	7.62847	262.717	rice
7	69	37	42	23.058	83.3701	7.07345	251.055	rice
8	69	55	38	22.7088	82.6394	5.70081	271.325	rice
9	94	53	40	20.2777	82.8941	5.71863	241.974	rice
10	89	54	38	24.5159	83.5352	6.68535	230.446	rice
11	68	58	38	23.224	83.0332	6.33625	221.209	rice
12	91	53	40	26.5272	81.4175	5.38617	264.615	rice
13	90	46	42	23.979	81.4506	7.50283	250.083	rice
14	78	58	44	26.8008	80.8868	5.10868	284.436	rice
15	93	56	36	24.015	82.0569	6.98435	185.277	rice
16	94	50	37	25.6659	80.6639	6.94802	209.587	rice
17	60	48	39	24.2821	80.3003	7.0423	231.086	rice
18	85	38	41	21.5871	82.7884	6.24905	276.655	rice
19	91	35	39	23.7939	80.4182	6.97086	206.261	rice
20	77	38	36	21.8653	80.1923	5.95393	224.555	rice
21	88	35	40	23.5794	83.5876	5.85393	291.299	rice
22	89	45	36	21.325	80.4748	6.44248	185.497	rice
23	76	40	43	25.1575	83.1171	5.07018	231.384	rice
24	67	59	41	21.9477	80.9738	6.01263	213.356	rice
25	83	41	43	21.0525	82.6784	6.25403	233.108	rice
26	98	47	37	23.4838	81.3327	7.37548	224.058	rice
27	66	53	41	25.0756	80.5239	7.77892	257.004	rice
28	97	59	43	26.3593	84.044	6.2865	271.359	rice
29	97	50	41	24.5292	80.545	7.07096	260.263	rice
30	60	49	44	20.7758	84.4977	6.24484	240.081	rice
31	84	51	35	22.3016	80.6442	6.0433	197.979	rice
32	73	57	41	21.4465	84.9438	5.82471	272.202	rice
33	92	35	40	22.1793	80.3313	6.35739	200.088	rice
34	85	37	39	24.5278	82.7369	6.36413	224.676	rice

Figure 3.2: Crop Recommendation Dataset Diagram

3.3 CROP PREDICTION MODEL

In our project a crop prediction machine learning model is developed by utilizing seven input features including nitrogen (N), phosphorus (P), potassium (K), rainfall, temperature, humidity, and pH, with the goal of predicting the suitable crop name as the output. We experimented with several machine learning algorithms including Decision Tree, Naive Bayes, Support Vector Machine (SVM), Random Forest (RF), and LightGBM (LGBM). Among these algorithms, LightGBM demonstrated superior accuracy in crop prediction, outperforming the other models.

Therefore, we selected LightGBM as our algorithm of choice due to its high accuracy and efficiency in handling large datasets with numerous features. The use of LightGBM allowed us to build a robust crop prediction

Parameter	Maximum Value	Minimum Value
N (Nitrogen)	140	0
P (Phosphorus)	145	5
K (Potassium)	205	5
Temperature (°C)	43.68	8.83
Humidity (%)	99.98	14.26
pH	9.94	3.50
Rainfall (mm)	298.56	20.21

Table 3.2: Maximum and Minimum Values of Agricultural Parameters

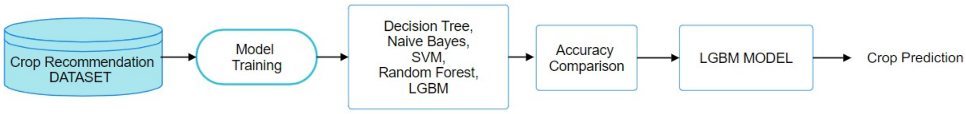

Figure 3.3: Crop Prediction ML model Diagram

model that effectively leverages the specified input features to make accurate and reliable predictions of suitable crops based on soil and environmental conditions.

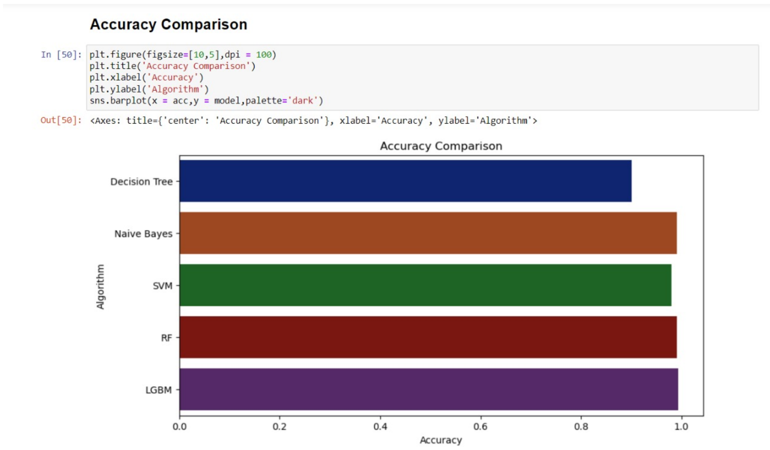

Figure 3.4: Accuracy Comparison Diagram

3.4 SATELLITE IMAGING

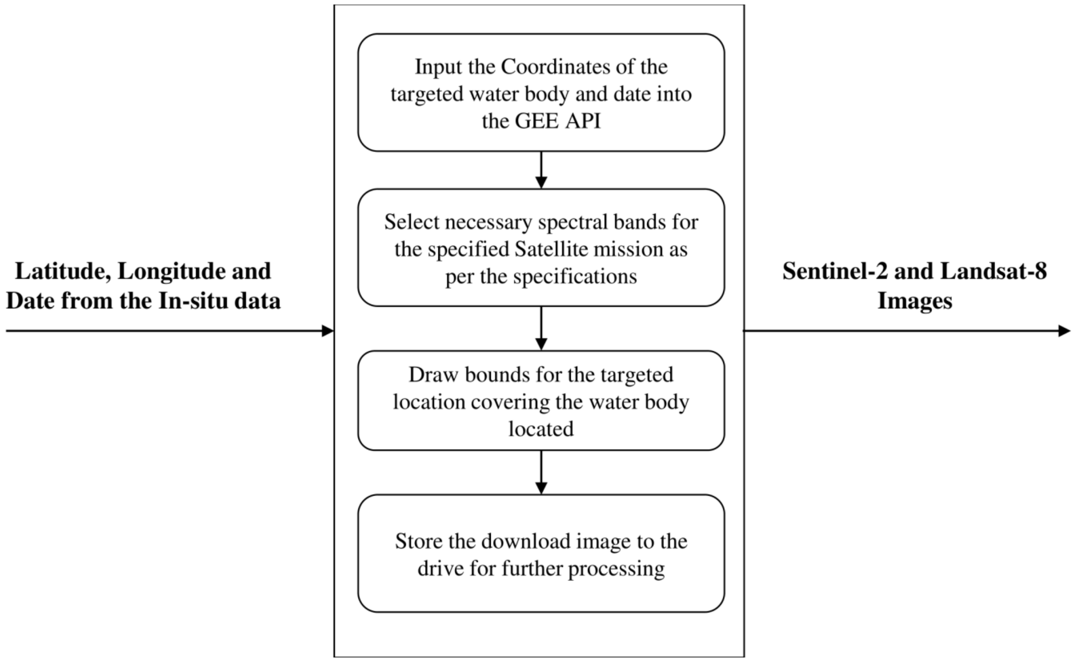

Figure 3.5: Data Collection of Sentinel and Landsat Images

Figure 3.5 provides insight into the module diagram, which is instrumental in acquiring satellite images for soil quality estimation.This module efficiently utilizes the Google Earth Engine Code Editor to retrieve satellite images of a specific location on a designated date.The user's inputs, including latitude, longitude, and the target analysis date, are employed to query the necessary data.

Landsat satellites capture multispectral data across different wavelengths of the electromagnetic spectrum, providing valuable information on land cover, vegetation health, and soil conditions.The specific bands used from Landsat imagery include.They are Red (Band 4), Near-Infrared (Band 5), Shortwave Infrared (Band 6) Landsat images covering the study area and relevant time periods were acquired using Google Earth Engine.Preprocessing steps included cloud masking and geometric correction to ensure data quality and consistency.

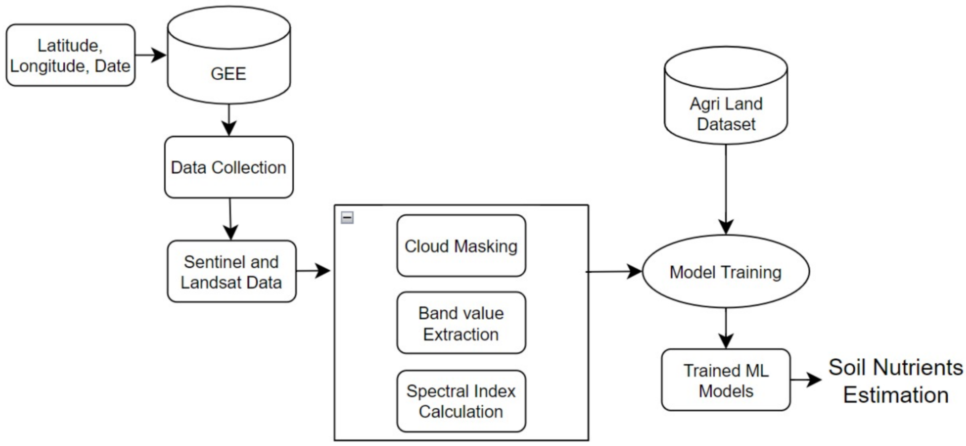

Figure 3.6: Satellite Architecture Diagram

Relevant spectral bands (Red, Near-Infrared, Shortwave Infrared) were extracted from Landsat images.These bands were chosen based on their sensitivity to vegetation health, soil moisture, and other key parameters affecting crop growth.

Composite images were generated by combining selected bands to create composite indices like Normalized Difference Vegetation Index (NDVI) and Soil Adjusted Vegetation Index (SAVI).These indices provide insights into vegetation vigor and soil moisture levels, crucial for crop prediction.

Spectral signatures extracted from Landsat imagery were used as input features for a machine learning model trained to predict soil qualities such as nutrient levels, pH, and moisture content.This model leverages the relationship between spectral reflectance and soil properties to estimate soil conditions.

Predicted soil qualities derived from satellite imagery were integrated with other input parameters (e.g., nitrogen, phosphorus, temperature) in the crop prediction model.This integrated dataset enables more comprehensive and accurate crop suitability assessments based on soil and environmental conditions.

3.5 CLOUD MASKING

In the context of this soil quality estimation project, addressing cloud removal and atmospheric correction is indeed crucial for obtaining accurate and reliable data from Sentinel-2 and Landsat-8 imagery.Clouds can obscure the soil surface, leading to spectral inconsistencies that affect the precision of soil quality assessments.Atmospheric conditions further influence spectral signatures, impacting the accuracy of these estimations.

Cloud removal involves using specific bands like QA60 bands in Sentinel-2 and Band-9 (Cirrus) and QA PIXEL band in Landsat-8 to mask out cloudy areas by nullifying the cloud bits.This process ensures that the soil areas underneath are visible and that the data used for analysis is clear and accurate.By eliminating cloud cover, you can achieve more precise estimations of soil quality parameters.Atmospheric correction is equally essential.It helps normalize spectral data, correcting for atmospheric influences like scattering and absorption.This correction ensures that the spectral characteristics of soil in the imagery accurately reflect their true properties, facilitating more accurate soil quality assessments.

By diligently applying cloud removal and atmospheric correction techniques, you enhance the reliability and accuracy of your soil quality assessments.This, in turn, contributes to a better understanding of soil quality dynamics and enables effective estimation of soil quality parameters for your project.

3.6 INTERNET OF THINGS

To create a comprehensive agricultural monitoring and prediction system, we integrate a variety of sensors and modules to gather crucial data about soil conditions and environmental factors.The NPK sensor enables us to precisely measure the levels of nitrogen, phosphorus, and potassium in the soil, providing essential insights into its nutrient composition. Coupled with the DS18B20 temperature sensor, soil moisture sensor, and RS485 Modbus module, we obtain real-time data on soil temperature, moisture content, and other relevant parameters. Additionally, incorporating sensors for rainfall, temperature, humidity, and pH further enhances our understanding of the surrounding environment, ensuring a holistic approach to crop management.

Utilizing the Arduino Nano board as the central processing unit, we collect and process data from the various sensors, enabling seamless communication and integration.

The ESP32 Wi-Fi module facilitates wireless connectivity, allowing for remote monitoring and control of the agricultural system.Furthermore, the NRF24L01 module enables communication between multiple nodes in the field, enhancing data collection efficiency and coverage.By harnessing the power of these components, we create a robust framework for data acquisition and transmission, laying the foundation for informed decision-making in agricultural practices.

The soil nutrients such as nitrogen (N), phosphorus (P), potassium (K), rainfall, temperature, humidity, and pH, can be analyzed with the help of these sensors.These analysis will be passed to the trained ML model to recommend the right crop to be planted so that it can produce maximum yield for the user.

Component	Uses
N,P,K Sensor	Measure the N, P, K values
DS18B20	Measure the temperature
Soil Moisture Sensor	Estimate the amount of water in the soil
RS485 Modbus Module	Supports communication to and from multiple devices
Arduino Nano Board	ATmega328P microcontroller
NRF24L01	Signal Transmission and Receiving
Bread Board	To build temporary circuits
Jumper Wires	Provide connection between circuits

Table 3.3: Components and Uses

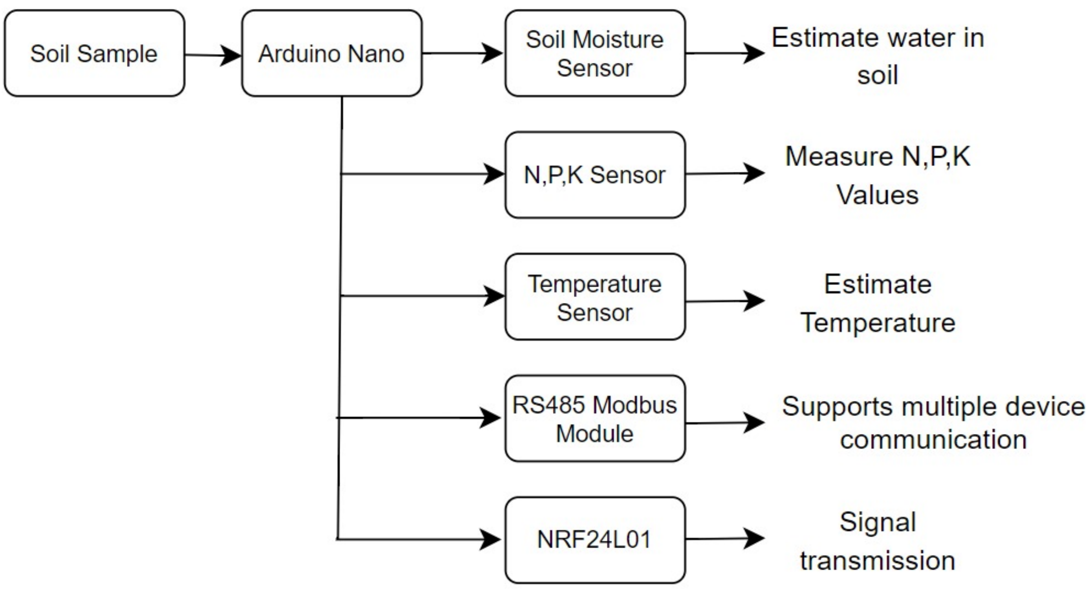

Figure 3.7: Transmitter Circuit Diagram

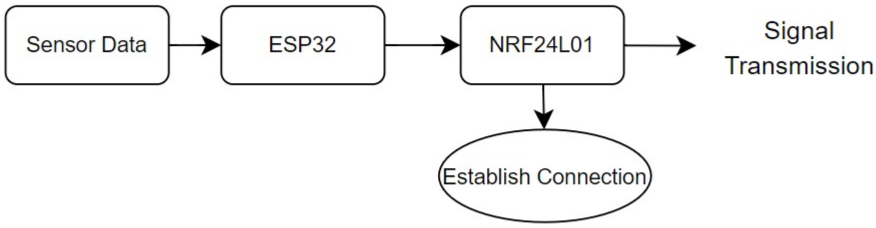

Figure 3.8: Receiver Circuit Diagram

By using these components we can create the required two circuit using bread board and jumper wires. Both the circuits are connected to NRF24L01 (antenna) which helps them in signal transmission and receiving. One circuit has sensors connected to Aurdino Nano Board which is used for the transmission of data and the other circuit has RS485 Module which helps in receiving the signal, so that it passes to the application and connect with the device.

CHAPTER 4

IMPLEMENTATION

Chapter 4 provides us with the algorithms used to implement the modules described in chapter 3.

4.1 LIGHTGBM

Algorithm 4.1 LightGBM Algorithm

1: **Input:** Crop Recommendation Dataset (X_{train}, Y_{train}, X_{test}, Y_{test})
2: **Output:** Accuracy of the Light Gradient Boosting Machine (LGBM) Model
3: **Start**
4: Import the necessary libraries: `lightgbm` for `LGBMClassifier` and `sklearn` for metrics
5: Create an instance of `LGBMClassifier`
6: Train the LGBM model on the training data (X_{train}, Y_{train})
7: Predict the crop recommendations for the test data (X_{test})
8: Calculate the accuracy of the model by comparing the predicted values with the actual values from the test data
9: Append the accuracy to a list named `acc` and the model name `'LGBM'` to another list named `model`
10: Print the accuracy of the Light Gradient Boosting Machine model
11: Print the classification report to assess the performance of the model in detail
12: **End**

Algorithm 4.1 serves as the cornerstone for determining the optimal crop to cultivate from a selection of 22 different crops, leveraging input features such as nitrogen (N), phosphorus (P), potassium (K), rainfall, temperature, humidity, and pH. Through the utilization of the Light Gradient Boosting Machine (LGBM) algorithm, this model achieves remarkable accuracy in crop prediction, surpassing alternative algorithms.The superior performance of

LGBM is evidenced by its accuracy rate of 99.3%, showcasing its efficacy in analyzing the complex interplay of agricultural factors to recommend the most suitable crop for cultivation.

This algorithm streamlines the decision-making process for farmers, providing them with actionable insights derived from advanced machine learning techniques.By accurately predicting the ideal crop based on comprehensive environmental and soil data, it empowers agricultural practitioners to optimize their yield potential and maximize productivity.

4.2 IMAGE SELECTION

Algorithm 4.2 Analysis for Point of Interest

1: **Input:** Geographic coordinates of the point of interest (Longitude, Latitude), Time range (Start Date, End Date)
2: **Output:** Clipped Landsat 8 RGB image centered around the point of interest
3: **Start**
4: Define the point of interest using its longitude and latitude coordinates
5: Define the time range for data acquisition, specifying the start and end dates
6: Filter the Landsat 8 collection to include only images that intersect with the point of interest and fall within the specified time range
7: Select the bands of interest for analysis, typically Red, Green, and Blue bands (SR_B4, SR_B3, SR_B2)
8: Calculate the median pixel values across all images in the filtered collection to create a composite image
9: Clip the composite image to the boundary of the point of interest, ensuring that only the relevant area is displayed
10: Display the clipped Landsat 8 RGB image on the map, setting visualization parameters such as minimum and maximum values for each band
11: Center the map view on the point of interest for better visualization
12: Add the point of interest as a red marker on the map to highlight its location
13: **End**

This algorithm harnesses Landsat 8 satellite imagery to facilitate comprehensive analyses centered around specific points of interest on the Earth's surface. Initially, it meticulously defines the geographic coordinates of these points and establishes a temporal framework for data acquisition, ensuring the retrieval of pertinent imagery within the desired time span. By filtering the Landsat 8 collection, the algorithm meticulously selects images that intersect with the designated points of interest and fall within the predefined time frame, optimizing the selection process to include only the most relevant data. Subsequently, employing sophisticated cloud masking techniques, the algorithm effectively eliminates cloud cover from the images, thereby enhancing the clarity and quality of the data for subsequent analyses.

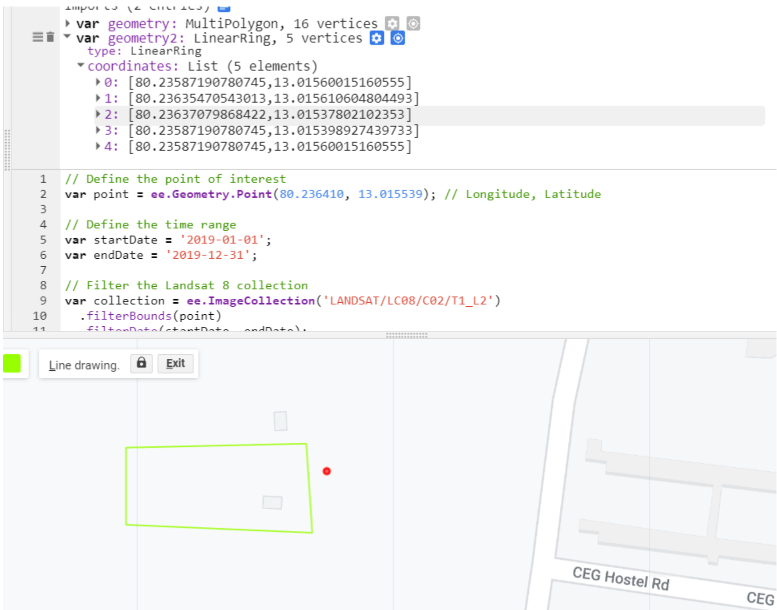

Figure 4.1: Image Selection Diagram

Moreover, after successfully mitigating the impact of cloud cover, the algorithm proceeds to clip the resulting cloud-free images to the boundaries of the specified points of interest. This localization ensures that subsequent analyses are precisely targeted, focusing exclusively on the areas of interest and eliminating extraneous data. Furthermore, to facilitate interpretation and exploration, the algorithm visualizes the clipped imagery on a map interface. This intuitive representation provides users with a holistic view of the data within the broader geographical context, enabling them to gain insights and make informed decisions based on the spatial distribution and characteristics of the imagery. Overall, this algorithm streamlines the process of conducting detailed analyses using Landsat 8 imagery, offering users a powerful tool for exploring and understanding Earth's surface dynamics.

4.3 DATA COLLECTION

Algorithm 4.3 Data Collection

1: **Input:** Latitude, Longitude, and Date of the target site
2: **Output:** Sentinel-2 and Landsat-8 image for the target site
3: **Start**
4: Input the latitude, longitude, and date
5: Choose images from 5 days before and 5 days after your selected date to ensure comprehensive coverage
6: Check if the satellite captured data for your specified location on the chosen date
7: If a suitable image is found, proceed to download it
8: Utilize the map interface to draw a shape and define your area of interest
9: **End**

Algorithm 4.3 is responsible for generating the required Satellite Images. The satellite imagery retrieval begins by soliciting user-provided latitude, longitude, and a specific date.

To ensure comprehensive coverage, the algorithm selects images spanning 5 days prior and 5 days post the given date.This extended time frame aims to compensate for potential gaps in data availability. Verifying satellite data, the algorithm checks if Sentinel-2 or Landsat-8 satellites have captured information for the specified location within the designated time frame.

Upon locating a suitable image, it proceeds to download it for further analysis. Utilizing a map interface, users can define their area of interest within the acquired satellite image, enabling focused examination of particular geographical regions.

This interactive map feature enhances precision and facilitates targeted research or analysis. Overall, this streamlined process simplifies satellite image acquisition, supporting various fields such as environmental monitoring, land use mapping, and urban planning.

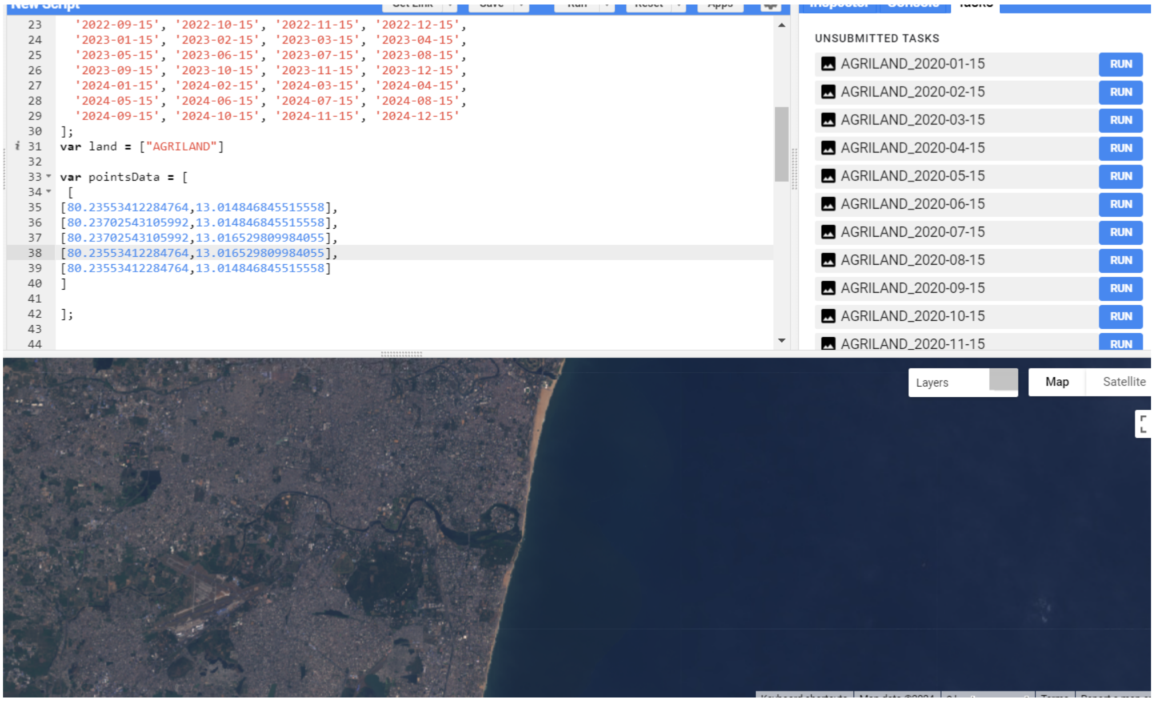

Figure 4.2: Data Collection Diagram

This facilitates satellite image acquisition by retrieving user-specified latitude, longitude, and date, then selecting images from 5 days before and after the given date to ensure comprehensive data coverage.

It verifies if Sentinel-2 or Landsat-8 satellites have captured data for the specified location within this timeframe, and downloads suitable images for analysis. Users can define their area of interest on an interactive map within the acquired image, enhancing precision for targeted research. This process, integral to effective data collection and provide accurate information for various analyses.

4.4 CLOUD MASKING

Algorithm 4.4 Cloud Masking

Input: Latitude, Longitude, and Date of the target site
Output: Cloud-masked Sentinel-2 and Landsat-8 image for the target site
Start
Load the image
Examine Sentinel images to detect regions affected by clouds and cirrus clouds
Select the QA60 band of the obtained Sentinel-2 Image
Apply a cloud bit mask: cloudBitMask = 1 ¡¡ 10, cirrusBitMask = 1 ¡¡ 11
Update the original image with the generated mask
End

Algorithm 4.4 is used for cloud removal and atmospheric correction. The Cloud Masking algorithm operates on user-inputted latitude, longitude, and a specific date to create cloud-masked Sentinel-2 and Landsat-8 images for a targeted site. Beginning with image loading, the algorithm meticulously examines Sentinel images to identify areas affected by clouds and cirrus clouds, crucial for ensuring image clarity and accuracy. By selecting the Quality Assessment (QA60) band from the obtained Sentinel-2 Image, the algorithm applies a cloud bit mask and a cirrus bit mask to effectively discern and isolate cloudy regions. The cloud bit mask ($1 \ll 10$) and cirrus bit mask ($1 \ll 11$) aid in differentiating between cloudy and clear areas within the image. Subsequently, the algorithm updates the original image by integrating the generated mask, resulting in a cloud-masked version ready for further analysis or processing. This process enhances the utility of satellite imagery by eliminating cloud-obscured areas, enabling clearer observations.

4.5 NITROGEN PREDICTION

Algorithm 4.5 Nitrogen Prediction Using SVR

Input: Dataset features (spectral band means), Target label (nitrogen level)
Output: Trained SVR model, Evaluation metrics (R-squared score, RMSE)
procedure NITROGEN PREDICTION
  Load dataset and select relevant features and target label
  Extract features (X) and target label (y) from the dataset
  Train-Test Split:
  Split dataset into training and testing sets (80:20 ratio) with random seed 42
  Model Development:
  Initialize SVR model with radial basis function (RBF) kernel
  Hyperparameter Tuning:
  Define hyperparameter search space for 'C' (regularization) and 'epsilon' (insensitivity)
  Perform RandomizedSearchCV to find best SVR model based on negative mean squared error
  Model Evaluation:
  Generate predictions (y_{pred}) using the best SVR model on the test set
  Calculate R-squared score (R^2) and root mean squared error (RMSE) for model evaluation
  Save the best SVR model to a file for future use
  Display the best hyperparameters, R-squared score, and RMSE
end procedure

The "Nitrogen Prediction Using SVR" algorithm involves several key steps for predicting nitrogen levels from spectral band mean features. It begins by loading the dataset and extracting relevant features and target labels. The data is then split into training and testing sets. An SVR model with an RBF kernel is initialized and optimized using hyperparameter tuning to improve performance. The best model is selected based on its ability to minimize mean squared error. Subsequently, the model's performance is evaluated using R-squared score and RMSE. The finalized model is saved for future use, and important outcomes such as optimal hyperparameters and evaluation metrics are reported. This approach enables accurate nitrogen level predictions leveraging machine learning methodologies.

4.6 PHOSPOROUS PREDICTION

Algorithm 4.6 Phosporous Prediction Using MLPRegressor

```
Input: Dataset features (spectral band means), Target label ('TEST VALUE')
Output: Trained MLPRegressor model, Evaluation metrics (Mean Squared Error, R-squared Score)
procedure TRAINMLPREGRESSOR
    Load dataset and select relevant features and target label
    Extract features (X) and target label (y) from the dataset
    Split the dataset into training and testing sets (80:20 ratio)
    Feature Standardization:
    Standardize the features using StandardScaler for better neural network performance
    Fit StandardScaler on the training data and transform both training and testing data
    MLPRegressor Model Configuration:
    Initialize MLPRegressor with specified parameters (e.g., hidden layer sizes, activation function, solver)
    Model Training:
    Train the MLPRegressor model using the scaled training data
    Optimization using Adam solver with maximum iterations of 500
    Output Results:
    Display Mean Squared Error (MSE) and R-squared Score (R2) on the test set
end procedure
```

The process of training an MLPRegressor model using spectral band means to predict a target value ('TEST VALUE'). It begins by loading the dataset and extracting relevant features and labels. The data is then split into training and testing sets for model validation. Feature standardization is applied to ensure uniform scaling across features. The MLPRegressor model is configured with specific parameters such as hidden layer sizes and activation functions. The model is trained using the training data and evaluated using Mean Squared Error (MSE) and R-squared Score (R2) metrics. Finally, the trained model is saved for future use. This systematic approach enables accurate prediction of target values based on spectral band features.

4.7 TEMPERATURE PREDICTION

Algorithm 4.7 Temperature Prediction Using RandomForestRegressor

```
Input: Dataset features (spectral band means), Target label ('Temperature')
Output: Trained RandomForestRegressor model, Evaluation metrics (R-squared score, RMSE)
procedure PREDICTTEMPERATURE
    Load dataset and select relevant features and target label
    Extract features (X) and target label (y) from the dataset
    Data Splitting:
    Split the dataset into training and testing sets (80:20 ratio) with random seed 42
    Hyperparameter Tuning with GridSearchCV:
    Define the grid of hyperparameters for RandomForestRegressor (number of estimators, max depth, min samples split, min samples leaf)
    Use GridSearchCV to find the best combination of hyperparameters based on negative mean squared error
    Model Training:
    Initialize RandomForestRegressor with the best hyperparameters identified
    Train the RandomForestRegressor model on the training data
    Model Evaluation:
    Make predictions on the test set using the trained model
    Calculate R-squared score (R2) and root mean squared error (RMSE) for model evaluation
    Output Results:
    Display R-squared score (R2), RMSE, and best hyperparameters
end procedure
```

The RandomForestRegressor model is trained to predict temperature based on spectral band means. The dataset is split into training and testing sets using a 80:20 ratio. Hyperparameter tuning is performed using GridSearchCV to optimize the model's performance. The best model configuration is identified based on negative mean squared error. The RandomForestRegressor is then trained using the optimal hyperparameters. Model evaluation is conducted on the test set to assess performance using R-squared score and root mean squared error. The trained model is saved for future use. This process ensures accurate temperature prediction from spectral band data with RandomForestRegressor.

4.8 PH PREDICTION

Algorithm 4.8 pH Prediction Using SVR with Spectral Ratios

```
Input: Dataset features (spectral ratios), Target label ('PH')
Output: Trained SVR model, Evaluation metrics (R-squared score, RMSE)
procedure PREDICTPH
    Load dataset and select relevant features and target label
    Extract features (X) and target label (y) from the dataset
    Data Splitting:
    Split the dataset into training and testing sets (80:20 ratio) with random seed 42
    Hyperparameter Tuning with RandomizedSearchCV:
    Define the search space for SVR hyperparameters (C, epsilon)
    Use RandomizedSearchCV to find the best combination of hyperparameters based on negative mean squared error
    Model Training:
    Initialize SVR with the best hyperparameters identified
    Train the SVR model on the training data
    Model Evaluation:
    Make predictions on the test set using the trained model
    Calculate R-squared score (R2) and root mean squared error (RMSE) for model evaluation
    Model Persistence:
    Save the trained SVR model to a file for future use
    Output Results:
    Display best hyperparameters, R-squared score (R2), and RMSE on the test set
    End Procedure
end procedure
```

The process of predicting pH values using Support Vector Regression (SVR) with spectral ratios as features. First, the dataset is split into training and testing sets. Hyperparameter tuning is performed using RandomizedSearchCV to optimize SVR's parameters (C and epsilon). The best SVR model is selected based on the tuned parameters and trained on the training data. Subsequently, the model predicts pH values for the test set, and its performance is evaluated using R-squared score and root mean squared error (RMSE). Finally, the best SVR model is saved for future use. This approach leverages machine learning techniques to accurately predict pH levels based on spectral ratio features.

4.9 SENSOR NODE DATA TRANSMISSION

Algorithm 4.9 Sensor Node Data Transmission

```
Input: Sensor readings (Soil moisture, Nitrogen, Phosphorous, Potassium, Temperature)
Output: Wireless transmission of sensor data using NRF24L01 module
Start
Include necessary libraries for communication and sensor handling: SPI, SoftwareSerial, OneWire, DallasTemperature, nRF24L01, RF24
Initialize the NRF24L01 module with appropriate CE and CSN pins
Define the address for wireless communication
Define pin assignments for sensor connections and communication control
Define calibration values for soil moisture sensor (AirValue, WaterValue)
Initialize variables for sensor readings and create a struct to store the sensor data
Setup the serial communication for debugging and the NRF24L01 module for wireless transmission
Initialize sensors and set communication pins for the NRF24L01 module
while true do                                                  ▷ Main loop
    Send a request to the sensor node to retrieve sensor data
    Read sensor data from the specified pins
    Convert analog soil moisture sensor reading to percentage using calibration values
    Read temperature from the temperature sensor
    Store sensor readings in the struct
    Print sensor readings to the serial monitor for debugging
    Send sensor data wirelessly using the NRF24L01 module
end while
End loop
End algorithm
```

This algorithm facilitates the transmission of sensor data from a sensor node to a receiver using an NRF24L01 module. Initially, it includes necessary libraries and initializes communication and sensor handling components. The NRF24L01 module is configured with appropriate pin assignments and an address for wireless communication. Pin assignments for sensor connections and communication control are defined, and calibration values for the soil moisture sensor are specified.

Within the main loop, the algorithm sends a request to the sensor node to retrieve sensor data and reads sensor data from the specified pins. It converts the analog soil moisture sensor reading to a percentage value and reads temperature from the temperature sensor. The sensor readings are stored in a struct, and then printed to the serial monitor for debugging purposes. Finally, the sensor data is sent wirelessly using the NRF24L01 module for further processing or analysis. This algorithm streamlines the process of transmitting sensor data wirelessly, facilitating remote monitoring and data collection in various applications.

4.10 SENSOR NODE DATA RECEPTION

Algorithm 4.10 Sensor Data Reception

```
Input: Sensor data received wirelessly
Output: Transmission of sensor data to Thingspeak server via WiFi
Start
Include necessary libraries for wireless communication, WiFi, and NRF24L01 module
Define WiFi credentials (SSID and password) and API key for Thingspeak server
Initialize NRF24L01 module, WiFi client, and set up serial communication
while true do                                                    ▷ Main loop
    if data is available on NRF24L01 module then
        Read the data into the struct
        Print received sensor data to the serial monitor for debugging
        Connect to Thingspeak server via WiFi
        Construct and send HTTP POST request containing sensor data fields
        Print confirmation message to the serial monitor
        Close WiFi connection
    end if
end while
End algorithm
```

This algorithm serves as a receiver for sensor data transmitted wirelessly and facilitates the transmission of this data to a Thingspeak server via WiFi. It begins by including necessary libraries and defining WiFi credentials and Thingspeak API key. The NRF24L01 module is initialized as a receiver, and WiFi client is initialized for communication with the Thingspeak server.

In the setup function, serial communication is started for debugging purposes, and the NRF24L01 module is configured to listen for incoming data. Additionally, a connection to the WiFi network is established using the provided credentials.

The recvData function checks if sensor data is available on the NRF24L01 module. If data is available, it is read into the struct for further processing.

Within the main loop, if sensor data is received, it is printed to the serial monitor for debugging and then transmitted to the Thingspeak server via WiFi using an HTTP POST request. Finally, a confirmation message is printed to the serial monitor.

CHAPTER 5

RESULTS AND DISCUSSIONS

This chapter contains the final outputs, explanation and analysis of this work. Crop Recommendation data was obtained from open - sources for the crop prediction.It contained the Nitrogen, Phosphorus, Ph, Temperature, Rainfall, Humidity, Potassium measured manually.

The datasets illustrated in Figure 5.1 is collected from www.kaggle.com which contains crop recommendation data of different crop parameters about experimental sites by government agencies and shared as open-source resources.

	A	B	C	D	E	F	G	H
1	N	P	K	temperatu	humidity	ph	rainfall	label
2	90	42	43	20.8797	82.0027	6.50299	202.936	rice
3	85	58	41	21.7705	80.3196	7.0381	226.656	rice
4	60	55	44	23.0045	82.3208	7.84021	263.964	rice
5	74	35	40	26.4911	80.1584	6.9804	242.864	rice
6	78	42	42	20.1302	81.6049	7.62847	262.717	rice
7	69	37	42	23.058	83.3701	7.07345	251.055	rice
8	69	55	38	22.7088	82.6394	5.70081	271.325	rice
9	94	53	40	20.2777	82.8941	5.71863	241.974	rice
10	89	54	38	24.5159	83.5352	6.68535	230.446	rice
11	68	58	38	23.224	83.0332	6.33625	221.209	rice
12	91	53	40	26.5272	81.4175	5.38617	264.615	rice
13	90	46	42	23.979	81.4506	7.50283	250.083	rice
14	78	58	44	26.8008	80.8868	5.10868	284.436	rice
15	93	56	36	24.015	82.0569	6.98435	185.277	rice
16	94	50	37	25.6659	80.6639	6.94802	209.587	rice
17	60	48	39	24.2821	80.3003	7.0423	231.086	rice
18	85	38	41	21.5871	82.7884	6.24905	276.655	rice
19	91	35	39	23.7939	80.4182	6.97086	206.261	rice
20	77	38	36	21.8653	80.1923	5.95393	224.555	rice
21	88	35	40	23.5794	83.5876	5.85393	291.299	rice
22	89	45	36	21.325	80.4748	6.44248	185.497	rice
23	76	40	43	25.1575	83.1171	5.07018	231.384	rice
24	67	59	41	21.9477	80.9738	6.01263	213.356	rice
25	83	41	43	21.0525	82.6784	6.25403	233.108	rice
26	98	47	37	23.4838	81.3327	7.37548	224.058	rice
27	66	53	41	25.0756	80.5239	7.77892	257.004	rice
28	97	59	43	26.3593	84.044	6.2865	271.359	rice
29	97	50	41	24.5292	80.545	7.07096	260.263	rice
30	60	49	44	20.7758	84.4977	6.24484	240.081	rice
31	84	51	35	22.3016	80.6442	6.0433	197.979	rice
32	73	57	41	21.4465	84.9438	5.82471	272.202	rice
33	92	35	40	22.1793	80.3313	6.35739	200.088	rice
34	85	37	39	24.5278	82.7369	6.36413	224.676	rice

Figure 5.1: Crop Dataset obtained from Kaggle

In city-latlon data was obtained from various open-sources for many target agriculture lands.It contained the coordinates of the target site along with States, District, Year, Rainfall data, Coordinates measured manually.

The datasets illustrated in Figure 5.2 is collected from www.kaggle.com which contains City-latlon data of different farm location parameters about experimental sites by government agencies and shared as open-source resources.

1	DISTRICT	YEAR	CROP	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	coord
2	NICOBAR	2017	blackgram	165.2	540.7	1207.2	892.1	{'lon': 92.675171, 'lat': 11.70065}
3	SOUTH AN	2017	maize	69.7	483.5	1757.2	705.3	{'lon': 72.849998, 'lat': 20.41667}
4	N & M ANI	2017	muskmelo	48.6	405.6	1884.4	574.7	{'lon': 72.849998, 'lat': 20.41667}
5	LOHIT	2017	pomegran	123	841.3	1848.5	231	{'lon': 77.066673, 'lat': 28.683331}
6	EAST SIAN	2017	cotton	112.8	645.4	3008.4	268.1	{'lon': 92.666672, 'lat': 23.799999}
7	SUBANSIR	2017	lentil	76.3	327.3	788.4	108.4	{'lon': 87.533333, 'lat': 23.91667}
8	TIRAP	2017	kidneybea	114.9	786.6	2385.5	284.5	{'lon': 79.76667, 'lat': 12.6}
9	ANJAW (LC	2017	pomegran	123	841.3	1848.5	231	{'lon': 75.050003, 'lat': 22.033331}
10	LOWER DI	2017	chickpea	237.6	955.1	1073.4	287.2	{'lon': 77.166672, 'lat': 27.0}
11	CHANGLAI	2017	coffee	241.2	1256.5	1632.1	414.1	{'lon': 80.199997, 'lat': 25.08333}
12	PAPUM PA	2017	pigeonpea	101.3	786	2245.7	245.2	{'lon': 72.949997, 'lat': 20.549999}
13	LOW SUBA	2017	jute	206.8	562.9	996.2	155.2	{'lon': 87.533333, 'lat': 23.91667}
14	UPPER SIA	2017	papaya	251	1168.8	2558.8	423.5	{'lon': 85.099998, 'lat': 20.85}
15	WEST SIAN	2017	maize	92.7	545.5	1612.5	190	{'lon': 85.966667, 'lat': 22.799999}
16	DIBANG VA	2017	jute	237.6	955.1	1073.4	287.2	{'lon': 74.083328, 'lat': 15.46667}
17	WEST KAM	2017	papaya	78.7	534.3	2485.2	301.2	{'lon': 73.933327, 'lat': 25.299999}
18	EAST KAMI	2017	mothbean	123.4	461.4	1246.3	199.8	{'lon': 73.933327, 'lat': 25.299999}
19	TAWANG(	2017	orange	78.7	534.3	2485.2	301.2	{'lon': 91.866669, 'lat': 27.58333}
20	KURUNG K	2017	mothbean	152.7	645.3	1256.5	167	{'lon': 71.76667, 'lat': 23.883329}
21	CACHAR	2017	watermelc	63.5	817.2	1889.9	228.6	{'lon': 78.433327, 'lat': 22.466669}
22	DARRANG	2017	pigeonpea	34.5	542.3	1259.1	121.9	{'lon': 75.26667, 'lat': 21.01667}

Figure 5.2: Farm Location Dataset obtained from Kaggle

An crop prediction machine learning model,we utilized seven input features including nitrogen (N), phosphorus (P),potassium (K), rainfall, temperature, humidity, and pH, with the goal of predicting the suitable crop name as the output. We experimented with several machine learning algorithms including Decision Tree, Naive Bayes, Support Vector Machine (SVM), Random Forest (RF), and LightGBM (LGBM). Among these algorithms, LightGBM demonstrated superior accuracy in crop prediction, outperforming the other models

```
accuracy_models = dict(zip(model, acc))
for k, v in accuracy_models.items():
    print (k, '-->', v*100)
```

```
Decision Tree --> 90.0
Naive Bayes --> 99.0909090909091
SVM --> 97.95454545454545
Logistic Regression --> 95.22727272727273
RF --> 99.0909090909091
LGBM --> 99.31818181818181
```

Figure 5.3: Accuracy comparison

Therefore, we selected LightGBM as our algorithm of choice due to its high accuracy and efficiency in handling large datasets with numerous features. The use of LightGBM allowed us to build a robust crop prediction model that effectively leverages the specified input features to make accurate and reliable predictions of suitable crops based on soil and environmental conditions.

```
In [59]: score

Out[59]: array([0.99545455, 0.98636364, 0.99318182, 0.98636364, 0.98863636])

In [60]: newdata=LGBM.predict([[90, 42, 43, 20.879744, 75, 5.5,220]])
         newdata

Out[60]: array(['rice'], dtype=object)

In [61]: import joblib

         joblib.dump(LGBM, "./L_GBM.joblib")

Out[61]: ['./L_GBM.joblib']
```

Figure 5.4: LigthGBM ML Prediction

We integrate a variety of sensors and modules to gather crucial data about soil conditions and environmental factors. The NPK sensor enables us to precisely measure the levels of nitrogen, phosphorus, and potassium in the soil, providing essential insights into its nutrient composition. Coupled with the DS18B20 temperature sensor, soil moisture sensor, and RS485 Modbus module, we obtain real-time data on soil temperature, moisture content, and other relevant parameters. Additionally, incorporating sensors for rainfall, temperature, humidity, and pH further enhances our understanding of the surrounding environment, ensuring a holistic approach to crop management

Satellite Images were collected using GEE API for the Coordinate at which the in City-Latlon data we measured. Up next, the collected date were subjected to cloud masking and atmospheric correction to prevent these factors from affecting the analysis. Once data is collected, the data is fed into model to calculate the soil quality parameter values for the target site.

Figure 5.5 (a) denotes a sample Sentinel-2 image and Figure 5.5 (b) denotes a Landsat-8 Image obtained using the GEE Code Editor

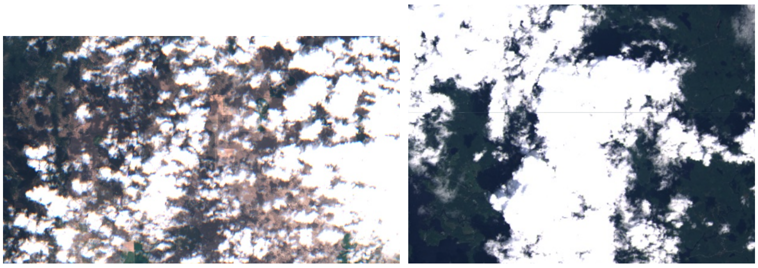

Figure 5.5: Sample Sentinel-2 and Landsat-8 Images obtained from GEE

After generating the Satellite Images, cloud masking, and atmospheric correction are performed on them to remove the clouds and cirrus-affected clouds. In our soil quality estimation project, it's essential to address cloud interference and atmospheric effects. We rely heavily on these images for the estimation of soil quality parameters and salinity prediction, but clouds in these images pose a significant obstacle. Figure 5.6a depicts a sample Sentinel-2 image before cloud masking and 5.6b depicts the corresponding cloud masked image.

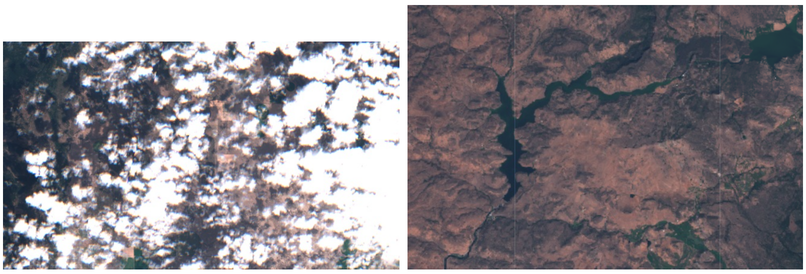

Figure 5.6: Before and after cloud masking in Sentinel 2 images

The dataset is contain the merging band values obtained from a segmented soil image. These satellite image bands are pivotal for estimating soil parameters, with the visible spectrum bands capturing the absorption and reflection of particulates, and near-infrared bands.

The development of models for Total Nitrogen utilized a dataset containing bands from Sentinel-2 images and corresponding in-situ nitrogen values. A Support Vector Regression (SVR) model was trained using an RBF kernel, with hyperparameter tuning conducted via RandomizedSearchCV [C = 348, epsilon = 0.097]. The SVR model was optimized to predict nitrogen levels based on the Sentinel-2 image bands.

```
import joblib
model = joblib.load(r"TN-svr.joblib")
```

```
pred = []
sum =0
n = 0
for i in range(img.shape[0]):
    t = []
    for j in range(img.shape[1]):
        predicted_value = model.predict([img[i][j]])

        # convert mgP/L
        # predicted_value =

        t.append(predicted_value)
        sum = sum+predicted_value
        n+=1
    pred.append(t)
```

```
sum/n
```

```
array([2.44110576])
```

Figure 5.7: Nitrogen ML Prediction

In the monitoring of total phosphorus, an MLP (Multi-Layer Perceptron) Regressor model is trained using band values obtained from Sentinel-2 data. The MLP Regressor is optimized to predict total phosphorus levels based on the Sentinel-2 image bands.

```python
import joblib
model = joblib.load(r"TP-mlp.joblib")
```

```python
pred = []
sum =0
n = 0
for i in range(img.shape[0]):
    t = []
    for j in range(img.shape[1]):
        predicted_value = model.predict([img[i][j]])


        t.append(predicted_value)
        sum = sum+predicted_value
        n+=1
    pred.append(t)
```

```python
sum/n
```

```
array([111.81651938])
```

Figure 5.8: Phosphorus ML Prediction

For the determination of temperature, a Random Forest model is fitted using a dataset containing 9 different bands (from band 2 to band 9) extracted from Sentinel-2 images along with corresponding in-situ temperature data measured in Kelvin. The Random Forest model is optimized using GridSearchCV to find the best hyperparameter settings [max depth = None, min samples leaf = 1, min samples split = 2, n estimators = 200].

```python
import joblib
model = joblib.load(r"temperature-rf.joblib")
```

```python
pred = []
sum =0
n = 0
for i in range(img.shape[0]):
    t = []
    for j in range(img.shape[1]):
        predicted_value = model.predict([img[i][j]])

        t.append(predicted_value)
        sum = sum+predicted_value
        n+=1
    pred.append(t)
```

```python
sum/n
```

```
array([15.43125056])
```

Figure 5.9: Temperature ML Prediction

For the development of pH prediction models, a dataset comprising Landsat-8 band means and pH values was used. Band ratios were computed across all combinations of bands and included in the dataset. A Support Vector Regression (SVR) model with an rbf kernel was created.

The integrated agricultural monitoring and prediction system incorporates a variety of sensors and modules to gather comprehensive data on soil conditions and environmental factors.

Key components include the NPK sensor for measuring nitrogen, phosphorus, and potassium levels in the soil, along with sensors for temperature, moisture, rainfall, humidity, and pH. These sensors enable a holistic understanding of soil nutrient composition and surrounding environmental conditions critical for crop management.Data collected from these sensors is processed by the Arduino Nano board, serving as the central processing unit, and communicated wirelessly using the ESP32 Wi-Fi module for remote monitoring and control.

The NRF24L01 module enhances data collection efficiency by enabling communication between multiple nodes in the field.This integrated framework empowers informed decision-making in agriculture by providing real-time insights into soil health and environmental parameters essential for optimizing crop productivity and sustainability.

For the user interface streamlit is used for displaying crop prediction and soil nutrient analysis using sentinal and landsat images.The web UI has options to upload images and predict button to display the analysis and prediction.

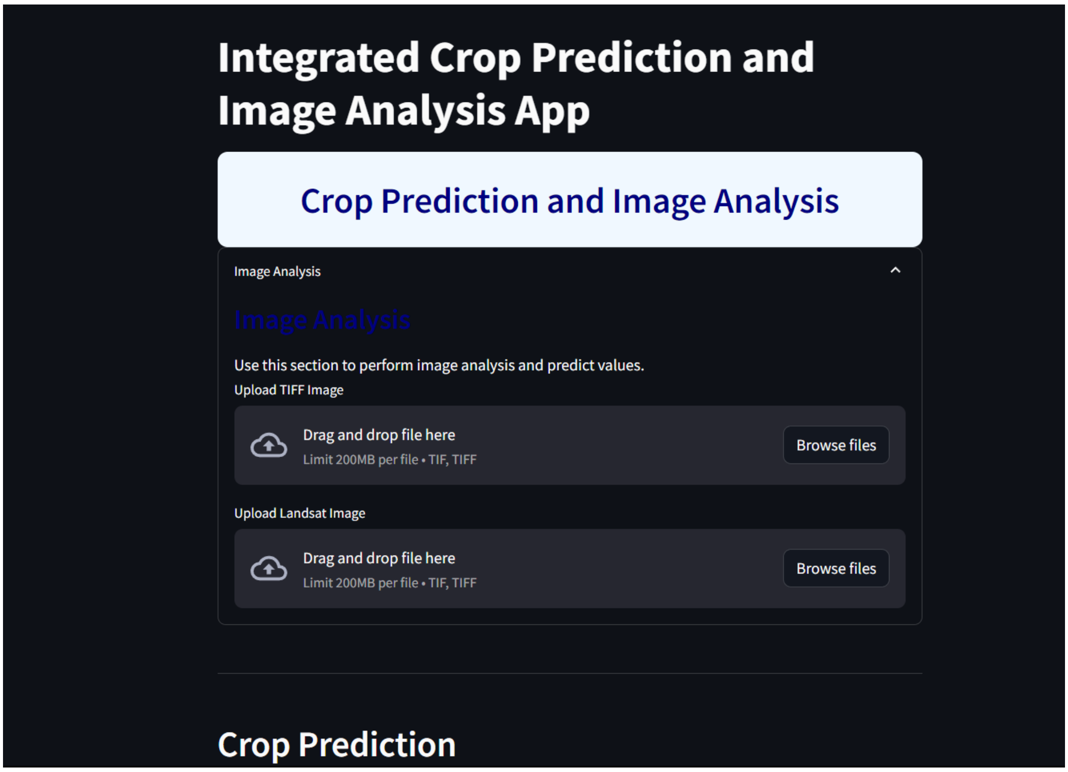

Figure 5.10: Web UI

This Streamlit code creates a web application for crop prediction and image analysis.It starts by importing necessary libraries such as streamlit, numpy, joblib, os, tifffile, rasterio, matplotlib.pyplot, io, PIL, tensorflow, and pickle.The code defines functions to load machine learning models for crop prediction and image analysis, handle image preprocessing, and predict various soil and environmental parameters (nitrogen, phosphorus, potassium, temperature, humidity, and pH).The main interface allows users to upload TIFF images, visualize them, and use buttons to trigger predictions for each parameter.Additionally, it includes input fields for user-provided soil and environmental data to predict the best crop type using a pre-trained LightGBM model.

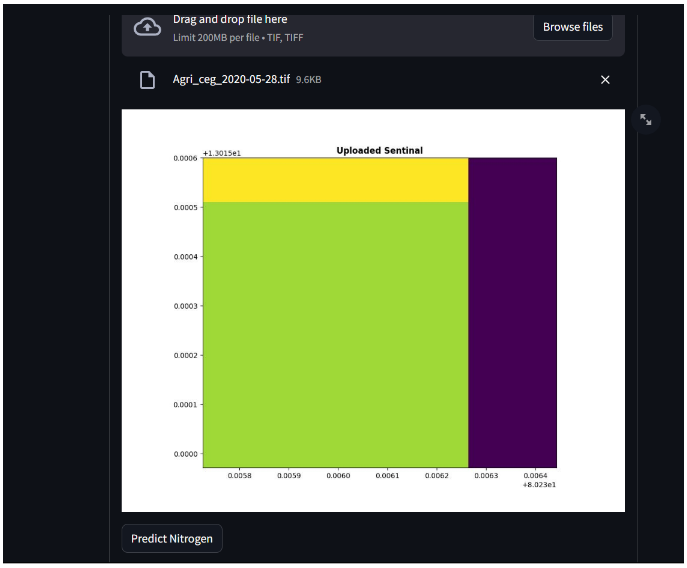

Figure 5.11: Sentinel Image Prediction

This focused on predicting nitrogen, phosphorus, and temperature from uploaded Sentinel-2 images.It imports similar libraries and defines functions for loading models, preprocessing images, and predicting specific parameters.

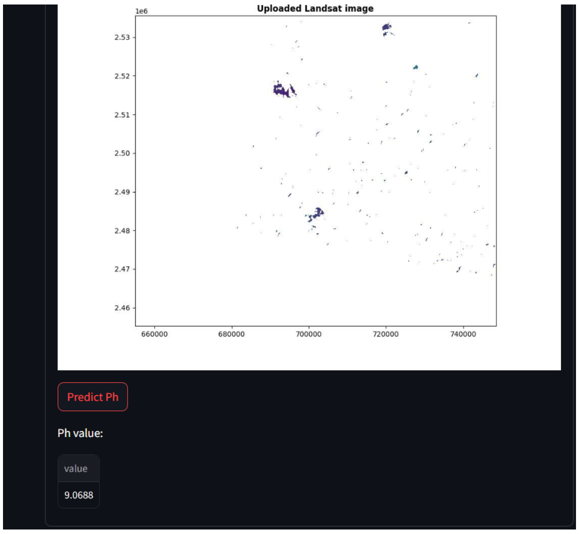

Figure 5.12: Landsat Image Prediction

The load models function gathers all.The joblib models in the ”models” directory, while array Conversion prepares image data for model input.The main function sets up the application interface, allowing users to upload TIFF images, visualize them, and predict nitrogen, phosphorus, and temperature values.Unlike the first code, it does not include crop prediction based on user input, instead focusing solely on the image-based analysis.

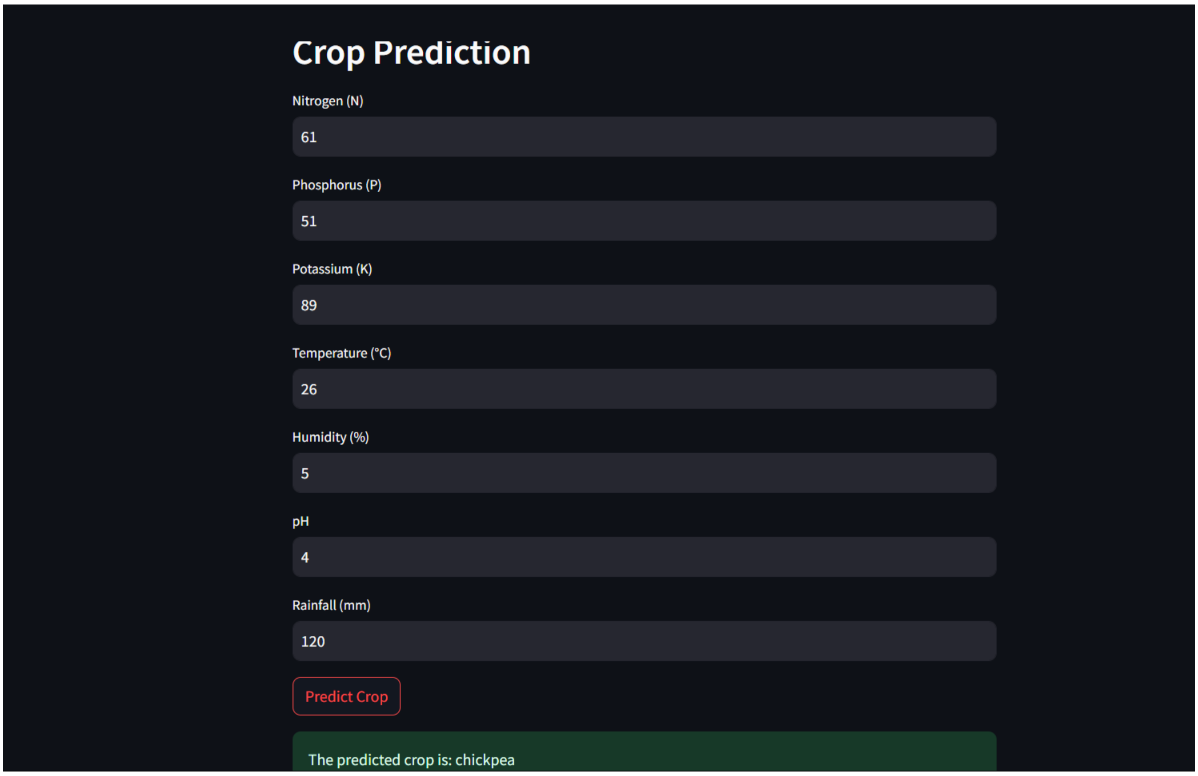

Figure 5.13: Crop Prediction Result

By utilizing Streamlit for the user interface, we made the crop recommendation output visible to the user by getting inputs such as Sentinel and Landsat images. These images are used to predict the soil nutrients, and these soil nutrients are sent to the crop recommendation model, which predicts the suitable crop for the image given by the user.

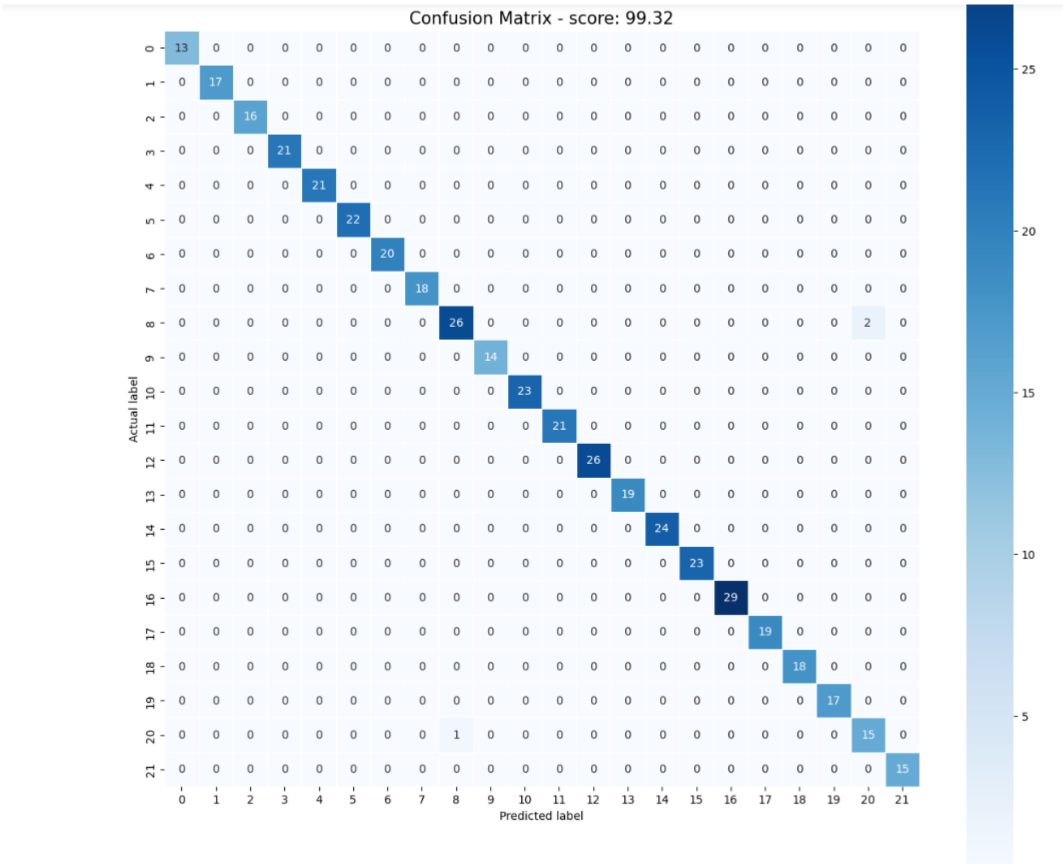

Figure 5.14: Confusion matrix

The confusion matrix is a fundamental tool in evaluating the performance of classification models.It provides a structured representation of the model’s predictions compared to the actual outcomes, dividing results into true positives (correctly predicted positives), true negatives (correctly predicted negatives), false positives (incorrectly predicted positives), and false negatives (incorrectly predicted negatives).This picture depicts the confusion matrix for the crop prediction LGBM algorithm used for the crop prediction used in this project.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

In conclusion, the project ”Crop Prediction System using Machine Learning and Analysis of Soil Nutrients using Satellite Imagery and IoT” represents an innovative approach towards enhancing agricultural practices through the integration of advanced technologies. By leveraging machine learning, IoT, and satellite imagery, the project aims to address critical challenges in crop management and soil nutrient monitoring.

The development of a machine learning model capable of recommending suitable crops based on environmental factors such as nitrogen (N), phosphorus (P), potassium (K), pH, rainfall, and moisture holds immense potential for farmers to make informed decisions regarding crop selection. This predictive capability can optimize yield and resource allocation, leading to improved agricultural productivity and sustainability.

Furthermore, the implementation of an IoT-based system for real-time measurement and monitoring of soil nutrient levels provides farmers with actionable insights to optimize fertilization strategies and ensure optimal soil health. The integration of satellite imagery data adds another dimension by enabling comprehensive soil analysis over large geographic areas, facilitating precision agriculture practices.

Overall, the project underscores the importance of harnessing data-driven technologies to transform traditional farming methods into efficient, data-driven practices that promote sustainable agriculture and food security.

6.2 FUTURE WORKS

The scope of Crop prediction system using ML and Analysis of soil nutrients using satellite imaging and IoT increases by refining the machine learning algorithms, incorporating additional environmental variables, and leveraging advanced techniques like ensemble learning and hyperparameter optimization to further enhance prediction accuracy.Integrating more IoT sensors into our monitoring system, such as those for organic matter content, will provide a more holistic view of soil health and enable proactive management practices.Through continuous refinement and innovation, Crop prediction system using ML and Analysis of soil nutrients using satellite imaging and IoT has the ability to empowering farmers with actionable data-driven insights and promoting sustainable agriculture on a broader scale.

Facilitating data sharing and collaboration is crucial for driving innovation and progress in agriculture.By developing open platforms and APIs for accessing and analyzing agricultural data, we can empower researchers, developers, and stakeholders to leverage diverse datasets and insights for informed decision-making and problem-solving.Open platforms promote transparency, interoperability, and scalability, enabling the integration of diverse data sources and the development of innovative solutions to complex agricultural challenges.Through a culture of openness and collaboration, we can harness the collective intelligence of the agricultural community to drive positive change and sustainable development in agriculture.

By making the application suitable for multiple users use at a same time can help this project evolve into a comprehensive agricultural technology solution that empowers farmers with data-driven insights and tools to optimize crop management practices, improve resource efficiency, and promote sustainable agriculture on a broader scale.

REFERENCES

[1] Neelamadhab Padhy Murali Krishna Senapaty, Abhishek Ray. Iot-enabled soil nutrient analysis and crop recommendation model for precision agriculture. *Pattern Recognition*, Vol: 142, p. 10966, 2023.

[2] Dr. Emily Johnson Dr. John Smith. Soil texture estimation using radar and optical data from sentinel-1 and sentinel-2. *Remote Sensing of Environment*, 2022.

[3] Chaitanya B. Pande., editor. *Prediction of Soil Chemical Properties Using Neural Network Wavelet Model*. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[4] Sonal Agarwal and Sandhya Tarar. *A hybrid approach for crop yield prediction using machine learning and deep learning algorithms*. Remote Sensing of Environment, 2021.

[5] Grace B.Vilamor Powell Mponela, Sieglinde Snapp. Digital soil mapping of nitrogen, phosphorus, potassium, organic carbon and their crop response thresholds in smallholder managed escarpments of malawi. *IEEE Geoscience and Remote Sensing Letters*, 2022.

[6] S. Musial. J.P.Wozniak E.Dabrowska-Zielinska K. Slesinski P Milewsk J.Bojanowski, S. Sikora. Integration of sentinel-3 and modis vegetation indices with era-5 agro-meteorological indicators for operational crop yield forecasting. 2022.

[7] Chaitanya Pande Nagavinothini Ravi chandran Balamurugan Panneerselvam, Kirubakaran Muniraj. Prediction and evaluation of groundwater characteristics using the radial basic model in Semi-arid region, India. In . *Journal of the Indian Society of Remote Sensing*, 2021.

[8] Gessesse Awoke. B Aklilu Tesfaye. A. Evaluation of the saturation property of vegetation indices derived from sentinel-2 in mixed crop-forest ecosystem. *IEEE International Geoscience and Remote Sensing Symposium*, 2021.

[9] Gozdowski. D Panek. E. Relationship between modis derived ndvi and yield of cereals for selected european countries. *Remote Sensing of Environment*, Pattern Recognition, 2021.

[10] Akhter. R. *Precision agriculture using IoT data analytics and machine learning*. PhD thesis, 2021.

[11] A. Bochenek Z. Bartold M.Gurdak R. Paradowski K. Lagiewska Dabrowska-Zielinska, K. Malinska. Drought model diss based on the fusion of satellite and meteorological data under variable climatic conditions. *Remote Sensing of Environment*, 2020.

[12] Gozdowski. D Panek. E. Analysis of relationship between cereal yield and ndvi for selected regions of central europe based on modis satellite data. remote sens. *7th International Conference on Communication and Electronics Systems*, 2020.

[13] Buszke. B Panek. E Gozdowski. D Samborski. S, Ruciński. D. Relationships between satellite-derived vegetation indices, grain yield and spike number of winter wheat and triticale. *7th International Conference on Communication and Electronics Systems*, 2020.