Dynamic Crop Yield Prediction Using Real-Time Sensor Data

24-25J-307

Project Proposal Report

Student: Edirisinghe E.A.K.G. - IT21267222

Supervisor: Ms. Uthpala Samarakoon

BSc (Hons) in Information Technology Specializing in Data Science

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

August 2024

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

	Name	Signature	Date
Member	Kavindu Edirisinghe (IT21267222)		23/08/2024
Supervisor	Uthpala Samarakoon		23/08/2024
Co-Supervisor	Aruni Premarathne		23.08.2024
External Supervisor	Dr. Lasantha Adikaram		22.08.2024

ABSTRACT

Accurate crop yield prediction is crucial for optimizing agricultural productivity and ensuring food security. This project proposes a dynamic crop yield prediction system using real-time sensor data combined with advanced machine learning models. By integrating data from sensors measuring crop weight, temperature, Sun Light and humidity, this system will provide farmers with timely insights and recommendations to improve crop management. The system will be developed using Oracle Cloud Infrastructure and machine learning algorithms such as Random Forest, Support Vector Machines, and XGBoost, alongside a Generative AI-based recommendation engine. The proposed solution aims to address the gaps in existing research by offering real-time, dynamic predictions and a user-friendly mobile application for farmers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

- ML: Machine Learning
- IoT: Internet of Things
- RF: Random Forest
- SVM: Support Vector Machine
- DHT: Digital Humidity and Temperature
- CNN: Convolutional Neural Network
- RNN: Recurrent Neural Network
- LSTM: Long Short-Term Memory
- API: Application Programming Interface
- SSL: Secure Sockets Layer
- GUI: Graphical User Interface
- SQL: Structured Query Language
- HTTPS: Hypertext Transfer Protocol Secure

1. INTRODUCTION

1.1 Background & Literature Survey

The agriculture sector plays a crucial role in the global economy, particularly in developing countries where it is a significant source of income and employment. Accurate prediction of crop yields is vital for planning and decision-making, ensuring food security, and stabilizing markets. Traditionally, crop yield prediction relied on historical data and statistical models, which often lack accuracy due to the complex interactions between environmental factors and crop growth [1].

Recent advancements in technology have introduced the use of real-time sensor data and machine learning (ML) models to enhance the accuracy of crop yield predictions. Sensors such as DHT11 for temperature and humidity, soil moisture sensors, and sunlight sensors like BH1750FVI can continuously monitor environmental conditions. This real-time data, when processed through advanced ML algorithms, can provide more accurate and timely predictions, helping farmers optimize their crop management practices [2]. Integrating Internet of Things (IoT) devices with ML models has shown significant potential in automating data collection and analysis, thereby reducing the manual effort required for monitoring [3]. Additionally, remote sensing data offers an additional layer of information that can improve the accuracy of predictive models [4].

However, despite these advancements, current research often overlooks the integration of real-time data and the use of dynamic prediction models. Most studies focus on static data, which does not account for the day-to-day variations in environmental conditions that significantly impact crop growth [5]. There is also a need for more generalized models that can be adapted to different types of crops and varying environmental conditions, as most current models are crop-specific and lack versatility [6].

1.2 Research Gap

Despite advancements in the field of crop yield prediction, significant gaps remain, particularly in the integration of real-time sensor data, the use of dynamic machine learning models, and the provision of actionable insights to farmers. This section compares existing research efforts and highlights the areas where the proposed system offers novel contributions.

1. **Real-Time Data Integration:**

- Research 1: "Crop Yield Prediction Using Machine Learning Techniques" by Ramesh Medar, Vijay S. Rajpurohit, and Shweta, primarily focuses on using historical weather data, with no mention of real-time sensor integration. This approach can lead to less accurate predictions due to the lack of current environmental data [11].
- Research 2: "Crop Yield Prediction Using Machine Learning" by Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, and Mansing Rathod also relies on static data sources, including past temperature and rainfall records. The absence of real-time data makes the model less responsive to sudden changes in environmental conditions [12].
- Research 3: "Crop Yield Prediction Using Machine Learning Algorithms" by Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, and Prof. Vinu Williams similarly uses historical data and does not explore the potential of integrating real-time data into the prediction models [13].

Gap Identified: None of these studies incorporate real-time data collection through sensors, which limits the timeliness and accuracy of the predictions. The proposed system addresses this gap by utilizing sensors to measure the weight, temperature, and humidity in real-time, enabling dynamic adjustments to the predictions based on current conditions.

2. **Dynamic, Continuous Prediction:**

- Research 1: The static nature of the prediction models used means they cannot adapt to changes in the environment after the data collection period. Predictions are made based on a fixed dataset, which does not account for day-to-day variations in crop growth conditions.
- Research 2: This study also employs a static prediction approach, where the model is trained on historical data and is not updated with new data over time. This can result in outdated predictions that do not reflect current conditions.
- Research 3: Similar to the first two studies, this research does not incorporate mechanisms for continuous prediction or real-time updates to the model.

Gap Identified: The inability of these models to update predictions dynamically as new data becomes available is a significant limitation. The proposed system will implement machine learning models that continuously update as new sensor data is collected, providing more accurate and timely predictions.

3. **Comprehensive Evaluation Metrics:**

- Research 1: The focus is largely on the accuracy of the predictions, with little consideration given to other evaluation metrics such as precision, recall, or F1-score. This narrow focus can overlook important aspects of model performance, particularly in cases of imbalanced data.
- Research 2: While this study also emphasizes accuracy, it does not provide a comprehensive evaluation of the model's performance using a variety of metrics, which could give a more rounded understanding of the model's effectiveness.
- Research 3: The evaluation methods used in this research are limited to accuracy as well, without exploring other critical metrics that could highlight different strengths and weaknesses of the model.

Gap Identified: There is a need for a more comprehensive evaluation framework that includes multiple metrics such as precision, recall, and F1-score, in addition to accuracy. The proposed system will evaluate the performance of the machine learning models using these metrics, ensuring a more thorough assessment of their effectiveness.

4. **User Accessibility and Interface:**

- Research 1, 2, and 3: None of these studies discuss the development of user-friendly interfaces for farmers, focusing instead on the technical aspects of the model development. This can limit the practical application of the research, as farmers may not be able to easily interpret or act on the predictions provided by the models.

Gap Identified: There is a lack of emphasis on creating accessible tools for end-users, such as farmers, who need intuitive and actionable insights. The proposed system will include a mobile application designed to present predictions and recommendations in a clear and user-friendly manner, ensuring that the technology is accessible and useful for those who need it most.

RESEARCH GAP ANALYSIS

SYSTEM	Real-time Sensor Data Integration	Advanced Feature Engineering (Weight Measurements)	Comprehensive Evaluation Metrics (Precision, Recall, F1-Score)	Dynamic, Continuous Prediction	Model Optimization for Other Crops
RESEARCH A	✗	✗	✗	✗	✗
RESEARCH B	✗	✗	✗	✗	✗
RESEARCH C	✗	✗	✗	✗	✗
PROPOSED SYSTEM	✓	✓	✓	✓	✓

Conclusion:

The proposed system aims to fill these identified gaps by integrating real-time sensor data, employing dynamic machine learning models, evaluating performance using a comprehensive set of metrics, and providing a user-friendly interface. This approach not only enhances the accuracy and relevance of crop yield predictions but also makes the technology more accessible and practical for farmers.

This expanded section should provide a more thorough comparison of the existing research and clearly outline how your project addresses the identified gaps.

2. RESEARCH PROBLEM

Problem Statement:

Crop cultivation, particularly for salad cucumbers, faces several challenges such as the accurate monitoring of crop growth stages, detection of defects, and identification of harmful insect infestations. Traditional methods, including manual monitoring and the use of static models, are inadequate in addressing the complexities of modern agriculture, especially on a large scale [7]. These methods lack the precision and timeliness required to ensure optimal crop yield and quality. This research seeks to address these limitations by developing a comprehensive, automated, real-time monitoring system that employs advanced machine learning techniques [8]. Furthermore, integrating predictive analytics with real-time monitoring can provide farmers with actionable insights, enabling them to make informed decisions quickly and efficiently [9].

Issues to Address:

Real-time monitoring of crop growth is essential to detect the optimal time for harvesting and to identify potential issues before they can significantly affect crop yield. Current systems lack the capability to monitor crop growth in real-time, making it difficult to predict the optimal time for harvesting or to detect potential issues before they affect crop yield. Additionally, integrating environmental data such as temperature, humidity, and sunlight into predictive models is crucial for accurate and timely predictions [10]. However, many existing solutions fail to fully integrate these data points, resulting in less accurate predictions. A comprehensive system that combines crop yield prediction, growth stage monitoring, and environmental data integration into a unified approach is thus essential.

Research Questions:

- Which machine learning models are most effective for predicting crop yields based on real-time sensor data?
- How can the integration of environmental data into machine learning models enhance the accuracy and timeliness of crop yield predictions?
- What techniques can be used to develop a real-time, dynamic prediction system that is both accurate and user-friendly for farmers?

3. OBJECTIVES

3.1 Main Objective

The main objective of this project is to develop a dynamic crop yield prediction system using real-time sensor data and advanced machine learning models. The system will provide accurate predictions and actionable recommendations to farmers, helping them optimize crop management practices.

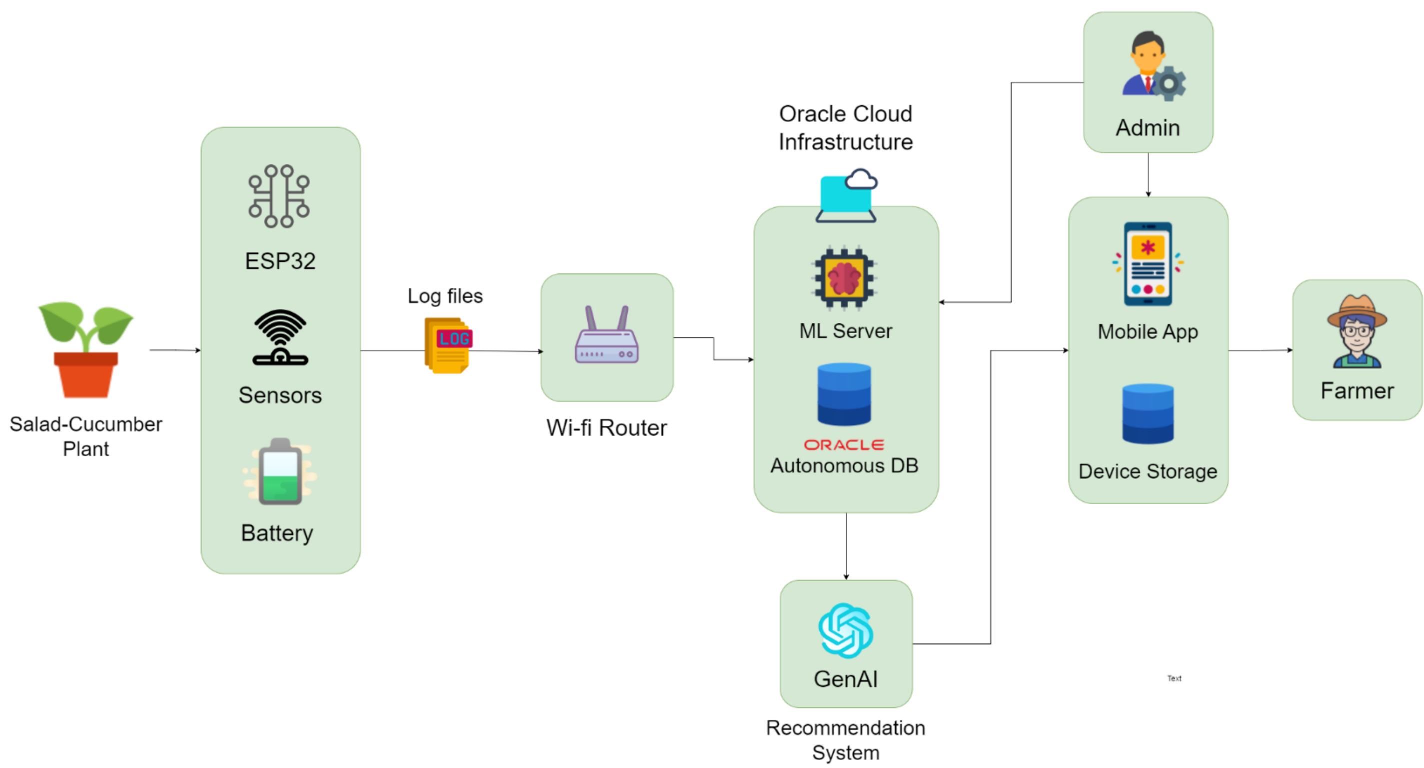
3.2 Specific Objectives

- **Data Collection:** Deploy sensors (DHT11, Soil Moisture Sensor, BH1750FVI) to gather real-time data on temperature, humidity, soil moisture, and sunlight.
- **Data Preprocessing:** Clean and normalize the collected data to ensure consistency and reliability.
- **Feature Engineering:** Extract relevant features from the sensor data to enhance the prediction model's accuracy.
- **Model Development:** Train and evaluate machine learning models (Random Forest, SVM, XGBoost) using the preprocessed data.
- **Prediction Optimization:** Fine-tune the models for different crops, starting with salad cucumbers and extending to others like tomatoes.
- **Recommendation Engine:** Develop a generative AI-based recommendation system to provide actionable insights to farmers.
- **User Interface Development:** Create a mobile application that presents predictions and recommendations in a user-friendly manner.

4. METHODOLOGY

The proposed system will use a client-server architecture hosted on Oracle Cloud Infrastructure (OCI). The sensors deployed in the field will collect real-time data, which will be transmitted to the ML server hosted on OCI. The ML server will process this data using various algorithms and store the results in an Oracle Autonomous Database. A mobile application will provide farmers with real-time monitoring, predictions, and recommendations.

4.1 **System Components and Workflow**



The proposed system comprises several key components that work together to provide real-time crop monitoring and yield prediction. These components include:

- Data Collection: Sensors such as DHT11, Soil Moisture Sensors, BH1750FVI, and Load Cells will be deployed to collect real-time data on temperature, humidity, soil moisture, sunlight, and crop weight.
- Data Preprocessing: The collected data will be cleaned and normalized to remove outliers and ensure consistency. Feature engineering will also be performed to enhance the data's predictive power.
- Feature Extraction: Relevant features will be extracted from the preprocessed data to serve as inputs for the machine learning models.
- Machine Learning Model Development: Various machine learning algorithms, including Random Forest, SVM, and XGBoost, will be trained to predict crop yields. A generative AI-based recommendation system will also be developed to provide actionable insights to farmers.
- Model Evaluation and Validation: The models will be evaluated using metrics such as accuracy, precision, recall, and F1-score. Cross-validation will be conducted to ensure the models' robustness and generalizability.
- Real-time Dashboard Implementation: A mobile application will be developed to display real-time monitoring data, predictions, and recommendations to farmers.

4.2 Machine Learning Model Development

The machine learning model development process involves the following steps:

- Data Preparation: The preprocessed data will be split into training, validation, and test sets. The training set will be used to train the machine learning models, while the validation set will be used to tune the model parameters. The test set will be used to evaluate the model's performance.
- Model Training: Various machine learning algorithms, including Random Forest, SVM, and XGBoost, will be trained using the training data. These models will be optimized for accuracy, precision, recall, and F1-score.
- Model Evaluation: The trained models will be evaluated using the validation and test sets. Metrics such as accuracy, precision, recall, F1-score, and AUC-ROC will be used to assess the models' performance. The models with the best performance will be selected for deployment.

- Recommendation Engine: A generative AI-based recommendation system will be developed to provide actionable insights to farmers based on the predictions made by the machine learning models.

4.3 Validation and Testing

The validation and testing phase will involve the following steps:

- Cross-Validation: K-fold cross-validation will be performed to ensure that the machine learning models are robust and generalizable. The models will be evaluated on different subsets of the data to assess their performance.
- Real-World Testing: The system will be tested in real farm environments to evaluate its accuracy and reliability in predicting crop yields and detecting issues. Feedback from farmers will be collected and used to refine the models.
- User Acceptance Testing: The mobile application will be tested by farmers to ensure that it is user-friendly and meets their needs. Any issues identified during testing will be addressed before the system is deployed.

4.4 Commercialization and Entrepreneurship Potential

The proposed system has significant commercialization potential in the agricultural sector. The following strategies will be used to commercialize the system:

- Market Need and Consumer Value: The system addresses a critical need in the agricultural sector for real-time crop monitoring and yield prediction. By providing farmers with actionable insights, the system can help them optimize their crop management practices and improve their yields. The system's user-friendly interface makes it accessible to farmers with varying levels of technical expertise.
- Commercialization Strategy: The system will be offered as a subscription-based service, with different pricing tiers based on the features and services provided. The system's modular design allows it to be customized for different crops and farming environments, increasing its marketability.

- Entrepreneurial Potential: The system has the potential to generate revenue through subscription fees, as well as through partnerships with agricultural equipment manufacturers and distributors. The system's scalability and adaptability make it an attractive investment for entrepreneurs and investors in the agricultural sector.

5. Project Requirements

Functional Requirements:

1. Real-time Sensor Data Collection:
   - The system must continuously collect data from various sensors, including weight, temperature, humidity, and sunlight sensors, to monitor crop growth in real-time.
   - This data should be transmitted to a central server for processing and displayed on a user-friendly interface accessible to farmers.
2. Data Preprocessing:
   - The system must include data cleaning procedures to eliminate noise and outliers from the collected sensor data.
   - Feature engineering should be applied to enhance the predictive power of the machine learning models by creating relevant features from raw data.
3. Dynamic Crop Yield Prediction:
   - The machine learning models should predict crop yield dynamically, updating predictions in real-time as new data is collected.
   - These predictions should be accurate, taking into account the varying environmental conditions affecting crop growth.
4. Defect and Insect Detection:
   - The system must detect defects and harmful insects affecting the crops, categorizing and reporting them promptly.
   - Alerts should be generated for specific anomalies, providing farmers with actionable recommendations on how to address these issues.
5. User Interface for Farmers:
   - A dashboard interface must be developed, allowing farmers to monitor real-time data, crop growth stages, and detected issues.
   - The dashboard should also store historical data and generate comprehensive reports for long-term analysis.

Non-Functional Requirements:

1. Scalability:
   - The system must be scalable to accommodate monitoring of multiple crop types and large volumes of sensor data.
   - It should be easily expandable to include additional sensors or new crop species without significant reconfiguration.
2. Accuracy:
   - The machine learning models used for growth stage classification and defect detection must maintain high accuracy, ideally above 90%.
   - Regular validation checks should be conducted to ensure the models continue to perform effectively over time.
3. Real-time Performance:
   - The system must process sensor data and update the dashboard in real-time, with minimal latency to ensure timely interventions.
   - Any alerts or notifications should be generated and sent to the farmer as soon as an anomaly is detected.
4. Security:
   - Data exchanged between the sensors and central server should be encrypted to prevent unauthorized access.
   - Access to the system's dashboard and controls should be restricted based on user credentials, ensuring only authorized personnel can make changes or access sensitive data.

Expected Test Cases:

1. Sensor Data Collection:

- Verify that the system correctly captures and transmits sensor data under various environmental conditions, such as changes in lighting or weather.

2. Data Preprocessing:

- Test the effectiveness of the preprocessing algorithms in improving data quality and ensuring accurate feature extraction.

3. Dynamic Yield Prediction:

- Assess the accuracy of the crop yield predictions made by the system in real-time, ensuring they are reliable under varying conditions.

4. Defect and Insect Detection:

- Validate the system's ability to accurately identify and classify plant defects and insect infestations, ensuring timely alerts are generated.

5. Dashboard Functionality:

- Ensure that the dashboard updates in real-time with minimal delay and that all historical data is accurately stored and retrievable for analysis.

GANTT CHART

	Semester 1						Semester 2						
task	**July**	**August**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**April**	**May**	**June**	**Jul**
Project Planning & Setup													
Define project scope and objectives	■												
Identify resources and stakeholders	■												
Develop project plan and timeline	■												
Data Collection													
Set up sensors and data collection systems		■	■	■									
Collect pictorial data of the fruit		■	■	■									
Clean and preprocess the data for analysis		■	■	■									
Feature Engineering & Data Preparation													
Normalize and engineer features from collected data				■	■								
Create labels for captured data				■	■								
Split data into training, validation, and test sets				■	■								
Model Development													
Implement advance machine learning models					■	■	■	■					
Train models on the prepared dataset					■	■	■	■					
Perform cross-validation and hyperparameter tuning					■	■	■	■					
System Integration													
Integrate the trained model with the fruit analyze							■	■					
Develop a user interface for system control							■	■					
Ensure real-time data flow between sensors and model							■	■					
Testing & Validation													
Test the system in controlled environments									■	■	■		
Validate the model's performance on new data									■	■	■		
Iterate and improve the model/system based on feedback									■	■	■		
Deployment & Monitoring													
Deploy the system in a real agricultural setting											■	■	
Monitor system performance and water usage efficiency											■	■	
Collect feedback and make necessary adjustments											■	■	
Documentation & Reporting													
Prepare project documentation													■
Create a final report detailing outcomes and findings													■
Present the project to stakeholders													■

7. WORK BREAKDOWN STRUCTURE (WBS)

8. BUDGET AND BUDGET JUSTIFICATION

Per Unit**:**

Item	Total Cost (LKR)
DHT11 Sensors	250
Soil Moisture Sensors	200
BH1750FVI Sunlight Sensors	800
Load Cells	1500
HX711 Amplifier	300
ESP32 Microcontroller	1500
Wires and Breadboard	500
Resistors	50
Cloud Services (OCI Subscription)	N/A
Total Budget	**5100**

Budget Justification:

- Cloud Services (OCI Subscription): We plan to use Oracle Cloud Infrastructure's Free Tier for hosting, minimizing costs. Upgrading to a paid tier may be necessary based on project needs.
- Hardware Components: Costs for sensors, microcontrollers, and other components may vary due to market fluctuations, potentially requiring budget adjustments.

9. References

[1] Y. Yuan, Y. Liu, Z. Liu, and M. Gao, "A Machine Learning Approach for Corn Yield Prediction Based on Climatic Factors," IEEE Access, vol. 8, pp. 104289-104297, 2020.

[2] A. T. N. Pham, N. L. Nguyen, M. T. Tran, and D. P. Nguyen, "IoT-Based Real-Time Monitoring System for Precision Agriculture," IEEE Sensors Journal, vol. 21, no. 2, pp. 1347-1354, 2021.

[3] S. Kumar, A. Sharma, and P. P. Bhattacharya, "Integrating IoT and Machine Learning for Agriculture: A Review," IEEE Access, vol. 8, pp. 143900-143912, 2020.

[4] M. Champaneri, D. Chachpara, C. Chandvidkar, and M. Rathod, "Crop Yield Prediction Using Machine Learning," IEEE Xplore, 2020.

[5] A. Rehman, M. Shoaib, S. Iqbal, and M. M. Rathore, "Smart Agriculture Monitoring and Control System," IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4423-4432, 2018.

[6] P. Maheshwari, N. P. Patel, and S. T. Navale, "Precision Agriculture with IoT and Machine Learning: A Review," IEEE Access, vol. 7, pp. 146673-146689, 2019.

[7] S. Sadeghi-Tehran, S. Virlet, J. Sabermanesh, and T. Papaphilippou, "Machine Learning Approaches for Early Disease Detection in Wheat," IEEE Transactions on Industrial Informatics, vol. 15, no. 8, pp. 4517-4526, 2019.

[8] M. Ferrara, A. N. D. Lee, and P. Williams, "Comprehensive Evaluation Metrics for Machine Learning Models in Precision Agriculture," IEEE Transactions on Automation Science and Engineering, vol. 17, no. 4, pp. 1955-1963, 2020.

[9] J. Chen, X. Li, and Y. Zhang, "Dynamic Yield Prediction in Agriculture Using Deep Learning and Big Data," IEEE Transactions on Big Data, vol. 6, no. 2, pp. 157-169, 2020.

[10] R. Bhat, "Emerging trends and sustainability challenges in the global agri-food sector," ScienceDirect, 2022.

[11] R. Medar, V. S. Rajpurohit, and S. Shweta, "Crop Yield Prediction Using Machine Learning Techniques," *International Journal of Engineering Research & Technology (IJERT)*, Available online.

[12] M. Champaneri, D. Chachpara, C. Chandvidkar, and M. Rathod, "Crop Yield Prediction Using Machine Learning," *IEEE Xplore*, Available online.

[13] A. Venugopal, A. S., J. Mani, R. Mathew, and V. Williams, "Crop Yield Prediction Using Machine Learning Algorithms," *International Journal of Engineering Research & Technology (IJERT)*, Available online.