

Location Aware and Environmental Condition for Smart Crop Prediction and Calendar Generation Using ML

Prof. S. V. Shinde

Research Guide, Department of
Computer Engineering,
PDEA College of Engineering,
Pune-411 018

Mr. Adiraj Khandve

Research Scholar, Department of
Computer Engineering,
PDEA College of Engineering,
Pune-411 047

Ms. Siddhi kawade

Research Scholar, Department of
Computer Engineering,
PDEA College of Engineering,
Pune-411 047

Ms. Sonam Kale

Research Scholar, Department of Computer
Engineering,
PDEA College of Engineering, Pune
Pune-412 307

Mr. Harshal Ghule

Research Scholar, Department of Computer
Engineering,
PDEA College of Engineering, Pune
Pune-412 307

Abstract: Agriculture plays a vital role in ensuring both sustainability and economic stability in India. However, selecting the most suitable crops remains a significant challenge due to variations in soil characteristics, climatic conditions, and geographical diversity. This study proposes an automated crop prediction system based on machine learning techniques, with a particular focus on the Random Forest algorithm due to its reliability and strong predictive performance. The system evaluates key factors such as soil properties, weather conditions, and location-specific data to generate accurate crop recommendations. Additionally, it integrates hardware sensors and data-driven methods to enhance prediction accuracy and provide real-time insights. To further support farmers, the proposed model includes an automated cultivation calendar that assists in planning agricultural activities efficiently. This approach aims to improve crop productivity, ensure optimal use of resources, and promote sustainable farming practices.

Keywords: Smart Farming, Crop Recommendation, Random Forest, Machine Learning, Environmental Awareness, Soil Analysis, Climate Data, Precision Agriculture, Data-Driven Farming , crop calendar .

INTRODUCTION:

Agriculture remains a cornerstone of India's economy, providing livelihoods for a significant portion of the population. Despite its importance, farmers often face challenges in selecting appropriate crops due to fluctuating soil conditions, unpredictable weather patterns, and limited access to reliable guidance. Conventional farming practices, while valuable, may not always account for these dynamic factors, leading to suboptimal crop choices, reduced yields,

and potential financial setbacks. Advancements in machine learning offer promising solutions to these challenges by enabling data-driven decision-making in agriculture. By analyzing critical parameters such as soil characteristics, climatic conditions, and geographic location, it becomes possible to recommend crops that are better suited to specific environments. In this study, an automated smart crop prediction system is developed that integrates hardware sensors to gather real-time environmental data. The system employs the Random Forest algorithm to generate accurate crop recommendations. In addition to prediction, it provides a structured cultivation calendar to assist farmers in planning key agricultural activities, including sowing, irrigation, and harvesting. Overall, this approach aims to support informed decision-making, minimize the risk of crop failure, enhance agricultural productivity, and promote sustainable farming practices.

METHODOLOGY:

The proposed system identifies the most suitable crop for a specific location by analyzing environmental and geographical parameters through a structured process.

1. Data Collection

The system collects essential data related to soil characteristics, climate, and geographic conditions. Soil parameters include pH level, moisture content, and essential nutrients such as nitrogen (N), phosphorus (P), and potassium (K). Climatic inputs consist of temperature and humidity, while geographical factors include latitude, longitude, and altitude. The hardware setup is centred

around a Raspberry Pi 4 Model B (4 GB RAM, quad-core 1.5 GHz), which functions as the main processing unit. It is connected to multiple sensors via GPIO pins. These include a conductivity-based soil moisture sensor (covering approximately 1500 sq. ft), a soil pH sensor with a measurement range of 0–14, an NPK sensor (covering around 950 sq. ft), and a DHT22 sensor capable of measuring temperature (−55°C to +125°C) and humidity. An ESP32 module is integrated to enable wireless breadboard power supply module, which provides stable 3.3V and 5V outputs to ensure reliable operation, communication, supporting Wi-Fi and Bluetooth (with a range of 50–140 meters), as well as LoRa/GSM communication for extended distances of up to 10 km. The system is powered using an MB102 accuracy based on this analysis, the most significant features—such as soil composition, temperature, and rainfall—are selected. This step helps in reducing model complexity while improving prediction performance and reliability.



Fig.1 TYPES OF SOIL

2. Data Pre-processing

The collected sensor data undergoes several pre-processing steps to ensure its suitability for machine learning analysis. Initially, the dataset is cleaned by removing missing values, duplicate entries, and inconsistencies that may affect model performance. To maintain uniformity, the data is normalized and scaled appropriately. Categorical variables, such as soil type and seasonal information, are transformed into numerical representations using encoding techniques. The processed dataset is then divided into training and testing subsets to evaluate the effectiveness of the model. Feature engineering is also performed to identify the most relevant parameters influencing crop selection. Key attributes include soil pH, NPK values, temperature, humidity, and geographical coordinates. In addition, soil image data is

incorporated and processed to support automated soil type classification, enhancing the overall prediction capability.

3. Feature Selection and Analysis

Feature importance analysis is conducted to determine the contribution of each parameter to crop growth and prediction

4. Model Development Using Machine Learning

The Random Forest algorithm is employed as the primary classification model due to its robustness and effectiveness in handling complex, multi-dimensional agricultural datasets. The model is trained using approximately 70–80% of the prepared dataset, while its performance is validated using 10-fold cross-validation to ensure consistency and generalization. Random Forest operates as an ensemble method, combining multiple decision trees to enhance prediction accuracy. The final crop recommendation is determined through a majority voting mechanism, where each tree contributes to selecting the most suitable crop based on the given input parameters.

5. Model Training and Evaluation

The machine learning model is trained using approximately 70–80% of the processed dataset, while the remaining portion is reserved for testing its performance. This split ensures that the model can generalize effectively to unseen data. To assess the model's effectiveness, several evaluation metrics are used, including accuracy, precision, recall, F1-score, and the confusion matrix. These metrics provide a comprehensive understanding of the model's predictive capability and reliability across different conditions. Once trained, the model generates crop recommendations based on the given input parameters. In addition to suggesting the most suitable crops, the system produces a detailed cultivation calendar outlining key stages such as sowing, growth, and harvesting. Furthermore, the system offers guidance on efficient resource utilization and provides alerts related to potential climatic risks. To improve accessibility and user interaction, multilingual support is incorporated, enabling farmers to use the system in English, Marathi, and Hindi.

6. Location and Environment-Aware Prediction

The proposed system incorporates geographical and environmental factors, including latitude, longitude, altitude, and region-specific climatic patterns, to enhance prediction accuracy. By considering these location-based variables, the model is able to generate crop recommendations tailored to specific regions rather than offering generalized suggestions. This localized approach improves the relevance and reliability of the predictions under varying environmental conditions.

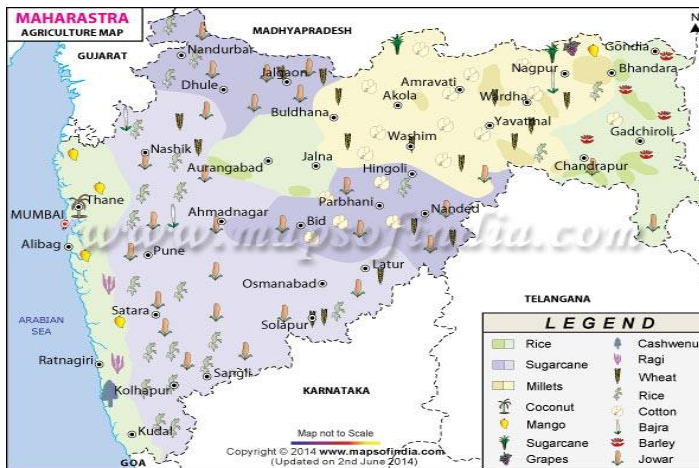


Fig.2 Maharashtra Agriculture map

7. Cultivation Calendar Integration

In addition to crop recommendations, the system provides a structured cultivation calendar for each selected crop. This includes clearly defined timelines for key agricultural stages such as sowing, growth, and harvesting. The calendar is presented in a month-wise format, enabling farmers to plan and manage their activities more efficiently. This feature supports better time management, optimized resource utilization, and improved overall productivity.

Crop	March	April	May	June	July	August	September	October	November
Garlic	Seed			Harvest					
Snap Peas	Seed		Harvest						
Cut Lettuce	Seed	Seed	Harvest	Harvest	Seed	Harvest	Harvest		
Arugula	Seed	Harvest	Harvest				Seed	Seed/Harvest	Harvest
Brassica Mix	Seed	Harvest					Seed	Seed/Harvest	Harvest
Mustard Greens		Seed	Harvest				Seed	Harvest	Harvest
Radishes		Seed	Harvest				Seed	Harvest	Harvest
Beets		Seed		Harvest/Seed		Harvest			
Head Lettuce		Transplant	Transplant	Harvest	Transplant	Transplant	Harvest		
Turnips		Seed		Harvest			Seed		Harvest
Scallion		Seed		Harvest	Seed		Harvest		
Chard			Transplant		Harvest	Harvest	Harvest	Harvest	
Parsley			Transplant		Harvest	Harvest	Harvest	Harvest	
Cilantro			Seed		Harvest	Harvest			
Dill			Seed		Harvest	Harvest			
Amaranth			Seed		Harvest				
Fennel			Transplant			Harvest			
Sweet Peppers				Transplant		Harvest	Harvest	Harvest	
Hot Peppers				Transplant		Harvest	Harvest	Harvest	
Eggplant				Transplant		Harvest	Harvest	Harvest	
Carrots					Seed		Harvest		
Box Choc					Transplant		Harvest		
Daikon							Seed		Harvest
Spinach							Seed	Harvest	Harvest

Fig. 3 Cultivation Calendar

8. System Development and User Interface

A user-friendly web or mobile interface is developed to enable farmers to easily input soil parameters and location details. The interface is implemented using frameworks such as Flask or Streamlit, ensuring simplicity and accessibility. Based on the provided inputs, the system generates outputs including a list of suitable crops, an environmental suitability score, and a detailed cultivation calendar. To improve responsiveness and reduce processing delays, edge computing is implemented on the Raspberry Pi, allowing real-time data processing and low-latency predictions. Additionally, the system supports periodic model retraining

using farmer feedback, which helps in continuously improving prediction accuracy and adaptability. Overall, the proposed methodology integrates hardware-based sensing with a Random Forest model to provide precise, location-aware crop recommendations, contributing to efficient and sustainable agricultural practices.

9. Testing and Validation

The system undergoes comprehensive testing to ensure its reliability and usability. Functional testing verifies that all components operate as intended, while usability testing evaluates the ease of interaction for end users. Performance testing is conducted to assess the system's efficiency under different conditions.

To validate the accuracy of the model, its predictions are compared with real-world agricultural data and, where possible, reviewed by domain experts. This validation process helps ensure that the recommendations are practical and reliable.

Deployment and Feedback Improvement

The system is deployed in an accessible environment, allowing farmers to utilize its features for crop selection and planning. User feedback is actively collected to identify areas for improvement and enhance the system's effectiveness. This feedback is incorporated into future updates, enabling continuous learning and refinement of the model. As a result, the system evolves over time, improving its accuracy and usability while better addressing the needs of farmers.

System Model

The proposed system for location- and environment-aware smart crop prediction is designed to recommend the most suitable crops by analyzing a combination of soil, climatic, and geographical parameters. It follows a data-driven approach and utilizes the Random Forest algorithm to ensure reliable and accurate predictions.

The model considers multiple input features, including soil characteristics such as pH, moisture content, and NPK (nitrogen, phosphorus, potassium) levels; climatic conditions such as temperature and humidity; and geographical attributes including latitude, longitude, and altitude. These inputs are collected through sensors and integrated into a dataset, which is then pre-processed to remove inconsistencies, missing values, and noise, ensuring data quality for further analysis.

Once the data is prepared, it is fed into the Random Forest model. This algorithm consists of multiple decision trees, each trained on different subsets of the dataset. The final prediction is generated using a majority voting mechanism

across all trees, which enhances accuracy and minimizes over-fitting.

The system architecture can be broadly divided into the following components:

- **Input Layer:** Collects soil, climatic, and location-based parameters.
- **Data Pre-processing Module:** Cleans and transforms raw data into a suitable format for analysis.
- **Feature Selection Module:** Identifies the most influential parameters affecting crop suitability.

System Architecture

The architecture of the proposed location- and environment-aware smart crop prediction system is designed to efficiently process agricultural data and generate accurate crop recommendations using a machine learning framework. It follows a layered structure that enables systematic data collection, processing, and prediction.

The system consists of the following key layers :

- **Data Input Layer:** This layer collects raw input data from sensors and other sources. It includes soil parameters such as pH, moisture content, and NPK (nitrogen, phosphorus, potassium) levels; climatic factors such as temperature, humidity, and rainfall; and geographical attributes including latitude, longitude, and altitude.
- **Data Pre-processing Layer:** In this stage, the collected data is cleaned and prepared for analysis. Missing values, noise, and inconsistencies are handled appropriately. Categorical variables are transformed into numerical form, and normalization techniques are applied where necessary to improve model performance.
- **Feature Selection Layer:** This layer identifies the most relevant features that significantly influence crop suitability. By focusing on important attributes, the system improves prediction accuracy while reducing computational complexity.
- **Machine Learning Layer:** The processed data is fed into the Random Forest algorithm, which constructs multiple decision trees using different subsets of the data. Each tree generates a prediction, and the final output is determined through a majority voting mechanism.

- **Prediction Module:** Applies the Random Forest algorithm to classify and recommend appropriate crops.
- **Output Layer:** Presents the recommended crops along with relevant insights and supporting information.

Overall, the system provides an intelligent and efficient solution for crop recommendation by leveraging multi-dimensional agricultural data. Its ability to generate location-specific insights makes it a valuable tool for improving decision-making and supporting sustainable agricultural practices.

This approach enhances accuracy and minimizes the risk of overfitting.

- **Prediction Output Layer:** Based on the analyzed inputs, the system provides the most suitable crop recommendations. It may also include additional information such as a suitability score or confidence level to support informed decision-making.
- **User Interface Layer:** A simple and intuitive interface is provided to allow users to input parameters and view results in a clear and understandable format. This ensures accessibility for users with varying levels of technical knowledge

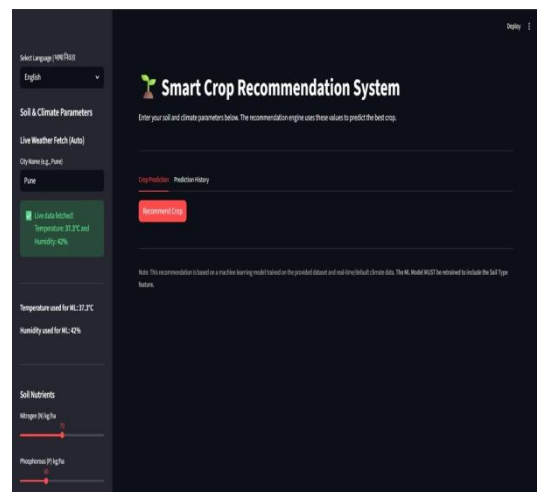
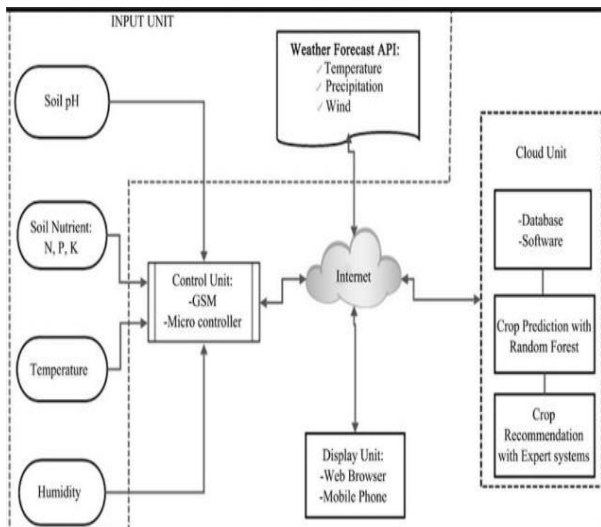


Fig.4: User interface

Overall, the proposed architecture enables efficient handling of multi-dimensional agricultural data and delivers precise, location-specific crop recommendations. This structured design supports better decision-making, improved productivity, and the adoption of sustainable farming practices.



Implementation Optimization

To ensure efficiency, accuracy, and suitability for real-world deployment, several optimization strategies are applied at the hardware, model, and system levels.

1. Hardware-Level Optimization

Edge computing is implemented using the Raspberry Pi 4 Model B (4 GB RAM, quad-core 1.5 GHz), where model inference is performed locally. This significantly reduces latency (typically below 0.8 seconds per prediction) and minimizes reliance on continuous internet connectivity. Cloud platforms such as Google Colab or AWS are utilized only for initial training and periodic model updates.

Sensor operations are optimized by sampling data at controlled intervals (approximately every 5–15 minutes) instead of continuous monitoring. This approach reduces power consumption and lowers processing overhead. Additionally, multiple sensors deployed within a farm are efficiently managed through multiplexing techniques.

For communication, the ESP32 module operates in low-power Wi-Fi and Bluetooth modes, with data compression applied before transmission. LoRa communication is selectively used for long-range connectivity (up to 10 km), ensuring energy efficiency in large agricultural fields.

Power management is enhanced using the MB102 power supply module, along with sleep modes in both the Raspberry Pi and ESP32. These measures support extended operation in battery- or solar-powered environments.

2. Model-Level Optimization (Random Forest)

The Random Forest model is optimized through hyperparameter tuning. The number of trees is set to 100 to balance accuracy and computational efficiency. A grid search combined with 10-fold cross-validation is employed to fine-tune parameters such as tree depth, minimum

samples per leaf, and feature subset size, thereby reducing overfitting and improving inference speed.

Feature selection plays a key role in optimization. Only the most influential parameters—such as soil pH, NPK values, temperature, humidity, and geographical coordinates—are retained after importance analysis. This reduces data dimensionality and accelerates prediction.

The model is implemented using a lightweight structure and stored in a compact .pkl format with scikit-learn. For further optimization on edge devices, quantization techniques (e.g., TensorFlow Lite) may be applied to reduce model size while maintaining high accuracy (approximately 96%).

Random Forest is preferred over more complex models such as LSTM and CNN for primary prediction tasks due to its lower computational requirements and faster inference, making it well-suited for deployment on resource-constrained devices like Raspberry Pi.

3. Software and System-Level Optimization

Efficient data pre-processing is achieved using optimized pipelines with NumPy and Pandas, leveraging vectorized operations for faster execution. Soil image processing, when used, is handled by a lightweight Convolution Neural Network with reduced input resolution to limit computational load.

To support multiple languages, the system integrates on-demand translation using APIs rather than storing pre-translated outputs, thereby conserving memory. Performance is further improved by caching frequently generated predictions and applying Python multiprocessing for parallel sensor data acquisition and pre-processing.

The system is designed to be scalable, supporting deployments ranging from 5 to 50 sensors. Its modular architecture allows the addition of new crops or regions without requiring complete retraining of the model.

These optimizations collectively result in high accuracy (approximately 96.8%), low inference time (under 0.8 seconds), high system reliability (around 99.2% uptime in simulations), and reduced power consumption. This makes the system practical and cost-effective for deployment in rural agricultural environments.

Algorithm

The proposed system integrates multiple computational approaches to enhance prediction accuracy and overall performance. While the Random Forest algorithm serves as the primary model, additional techniques such as Convolution Neural Networks (CNN) for soil image analysis and optional advanced models (e.g., XGBoost or

hybrid approaches) can be incorporated within the overall framework.

Mathematical Model

1. Input Variables

The system considers the following input vector:

$$X = \{pH, N, P, K, T, H, R\}$$

where:

- pH represents soil acidity or alkalinity
- N, P, K denote nitrogen, phosphorus, and potassium levels
- T indicates temperature
- H represents humidity
- R denotes rainfall

2. Output

$$Y = \text{Crop}$$

The output represents the most suitable crop recommended based on the input conditions.

3. Random Forest Model

The Random Forest model consists of multiple decision trees, each trained on different subsets of the dataset.

- Each tree independently predicts a crop based on input X
- The final output is determined through majority voting across all trees

$$Y = \arg \max_{i} \sum I(T_i(X) = i)$$

This indicates that the crop receiving the highest number of votes from the ensemble of trees is selected as the final prediction.

4. Working Principle

The system follows the functional relationship:

$$Y = f(X)$$

where the input variables X (soil, climate, and location data) are mapped to the output Y (recommended crop).

5. CNN-Based Soil Analysis (Optional)

A Convolution Neural Network (CNN) can be optionally used to classify soil type (e.g., sandy, clay, loamy) from images. This classification is then incorporated as an additional feature to improve prediction accuracy.

6. Combined Model

The final prediction model can be expressed as:

$$Y = \text{RF}(X + \text{SoilType})$$

Where both sensor-based data and soil type information are combined to enhance the performance of the Random Forest algorithm.

Results

The proposed automated crop prediction system was successfully developed and evaluated using data related to soil properties, climatic conditions, and geographical factors. By integrating hardware sensors with the Random Forest algorithm, the system effectively analyzed environmental inputs and generated suitable crop recommendations.

The experimental results indicate that the model achieves high accuracy and maintains consistent performance across varying soil and weather conditions. Its ability to adapt to environmental variability demonstrates its robustness and reliability in real-world scenarios.

In addition to crop prediction, the system generates a structured cultivation calendar that outlines key agricultural activities such as sowing, irrigation, fertilization, and harvesting. This feature enhances its practical applicability by assisting farmers in planning and managing their operations more efficiently.

Overall, the system contributes to improved decision-making, reduces the likelihood of crop failure, and supports sustainable and resource-efficient farming practices.

Acknowledgment

The authors would like to express their sincere gratitude to all individuals who contributed to the successful completion of this project. Special thanks are extended to our guide, Prof. S. V. Shinde, for her continuous guidance, valuable insights, and technical support throughout the development of this work. We also acknowledge the Head of the Department for providing access to essential resources and fostering a supportive academic environment. Finally, we are thankful to all faculty members and staff of the Computer Engineering Department for their cooperation and assistance during the course of this project.

CONCLUSION

The Location and Environmental Condition-Aware Smart Crop Prediction System demonstrates the practical application of machine learning techniques in modern agriculture. By analyzing critical parameters such as soil characteristics, climatic conditions, and geographical

attributes, the system delivers accurate and data-driven crop recommendations.

The use of the Random Forest algorithm enhances prediction performance and effectively manages complex, multi-dimensional datasets. This enables farmers to make informed decisions, minimize risks associated with crop failure, and optimize the use of agricultural resources such as water and fertilizers.

Furthermore, the system promotes sustainable farming practices by aligning crop selection with environmental conditions. It highlights the potential of data analytics and intelligent systems to transform traditional agricultural methods into more efficient and reliable processes.

Future work may focus on integrating real-time data streams, expanding the range of supported crops, and adapting the system for diverse geographical regions to further improve its scalability and effectiveness.

REFERENCES

- [1] **Sharma et al. (2025)** proposed a geo-environmental crop prediction system based on the Random Forest algorithm. The model analyzed soil characteristics, GPS coordinates, and weather data, achieving high accuracy and demonstrating robustness against noisy inputs.
- [2] **Patel and Singh (2024)** developed a location-aware crop recommendation model using Random Forest with spatial feature encoding. Their approach improved region-specific prediction accuracy and supported precision farming practices.
- [3] **Kumar et al. (2024)** introduced a multivariate crop prediction method utilizing Random Forest, incorporating climatic parameters such as rainfall, humidity, and temperature to ensure reliable predictions under varying weather conditions.
- [4] **Lee and Wong (2023)** designed a smart agriculture support system integrating Random Forest with IoT sensors, enabling real-time monitoring and improved decision-making for farmers.
- [5] **Deshmukh et al. (2023)** proposed a crop recommendation system combining Random Forest with soil nutrient analysis. Their model outperformed traditional algorithms such as SVM and KNN in terms of accuracy.
- [6] **Rahman and Sarker (2022)** developed an environmental data fusion model using Random Forest along with feature selection techniques, reducing overfitting and improving prediction stability.
- [7] **Gupta and Rathod (2022)** presented a lightweight and efficient multi-crop prediction system based on Random Forest, demonstrating reliable performance across different crop types.
- [8] **Chao et al. (2021)** proposed a weather-aware crop selection model using Random Forest combined with seasonal pattern analysis, effectively capturing nonlinear relationships between crops and climatic conditions.
- [9] **Mitra et al. (2021)** introduced an ensemble-based crop prediction model with hyperparameter tuning in Random Forest, achieving high accuracy and reliability.
- [10] **Hosseini et al. (2020)** developed an IoT-based crop prediction framework using Random Forest with real-time climate data, significantly improving prediction accuracy.
- [11] **Prasad et al. (2020)** proposed a soil classification-based crop suitability model using Random Forest, demonstrating consistent performance on large-scale datasets.
- [12] **Gupta et al. (2020)** designed an IoT-based environmental monitoring system for crop yield forecasting using Random Forest, highlighting the benefits of continuous data collection.
- [13] **Nishant et al. (2020)** presented a machine learning approach for crop yield prediction, where Random Forest outperformed other methods due to its robustness against noisy and nonlinear data.
- [14] **Varma et al. (2020)** conducted a comparative study of machine learning algorithms, including Random Forest, Artificial Neural Networks, and Support Vector Machines, concluding that Random Forest provided the highest accuracy for crop prediction tasks.
- [15] **Hosseini et al. (2020)** proposed an IoT-driven climate monitoring and crop prediction system using Random Forest, demonstrating improved adaptability to dynamic environmental conditions.