iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69641

Crop Prediction Using Sensors and Machine Learning

D. Sai kiran[1], G. Pranay Raj[2], J. Sai Krishna[3], K.R Abhinav[4], Dr. Obulesu[5]
Department of Information Technology, Vidya Jyothi Institute of Technology

Abstract: The agricultural industry is rapidly embracing emerging technologies to enhance crop forecasting and resource utilization. This project suggests an intelligent crop forecasting system based on IoT and machine learning for sustainable agriculture. Crop selection is usually done based on guesswork by the farmers, resulting in low harvest and wasted resources. Crop yield forecasting before harvesting is also a severe problem in developing countries. The system relies on Arduino-based hardware interfaced with sensors for measuring temperature, humidity, rain, soil moisture, and water levels. The values are applied to train a logistic regression model for crop recommendations. Crops taken into consideration are millets, tomato, sugarcane, strawberry, cotton, and rice. Early tests with simulated data have reflected encouraging accuracy. This location-aware, real-time solution can enable farmers to take informed decisions. It also facilitates schemes such as the Soil Health Card Scheme and PM-KISAN. Challenges involve maintaining sensor accuracy and dealing with environmental noise. Periodic calibration will be required for reliable performance. The system decreases reliance on conventional methods and enhances farm profitability. Satellite data and sophisticated ML models could be added in the future. This project provides a cost-effective, scalable step towards smart, sustainable agriculture.

Keywords: Crop Prediction,IoT(Internet of things) Machine Learning, Logistic Regression, Arduino, Environmental Sensors, Precision Agriculture, Soil Moisture, Temperature and Humidity Monitoring

I. INTRODUCTION

Agriculture has always been a vital sector, supplying food and raw materials to industries and also being a major source of livelihood for millions of people across the globe. Yet, farmers usually encounter serious challenges because of uncertain weather patterns, soil erosion, climate change, and wasteful utilization of resources like water and fertilizers. Conventional farming methods depend a lot on experience and past patterns, which do not necessarily fit today's agricultural requirements. With the speed of technological development, precision farming—a concept involving data-driven approaches—has emerged as a necessary part of modern agriculture. The Internet of Things (IoT) and Machine Learning (ML) are two technologies that are on the verge of transforming the field of farming by facilitating real-time monitoring, data gathering, and predictive analysis for improved crop management. This project, "Crop Prediction Using Sensors and Machine Learning," has the vision to offer an intelligent decision support system for farmers using real-time sensor readings fused with machine learning algorithms. The system will suggest appropriate crops depending on a range of environmental parameters like: Temperature, Humidity, Soil, Moisture, Water Level, Rainfall Through the use of IoT sensors to gather data and a Logistic Regression Machine Learning Model for prediction, this project aims to maximize crop choice, enhance productivity, and reduce wastage of resources.

II. LITERATURE SURVEY

We utilized MODIS data from NASA combined with county-level information from the United States Department of Agriculture (USDA) to parameterize empirical models estimating maize and soybean yield in the Central United States. As part of our analysis, we also examined the capacity of MODIS to monitor inter-annual variability in yields. Our findings indicate that the MODIS two-band Enhanced Vegetation Index (EVI2) offers a superior foundation for maize yield prediction compared with the popular Normalized Difference Vegetation Index (NDVI). Incorporating crop phenology information from MODIS substantially enhanced model accuracy within and among years. Contrary to expectation, applying Moderate Spatial Resolution MODIS Land Cover Type product for the identification of agricultural regions failed to impair model performance compared with the application of higher-spatial resolution crop-type maps created by the USDA. Vegetation-index-yield relations were strongest at 65–75 days greenup for corn and 80 days greenup for soybean. EVI2 was the optimal index for maize yield prediction in non-semi-arid counties ($R2 = 0.67$), but.Normalized Difference Water Index (NDWI) worked better in semi-arid counties ($R2 = 0.69$), likely due to the fact that NDWI is responsive to irrigation in semi-arid low-density agriculture.

NDVI and EVI2 were equally good at predicting soybean yield (R2 = 0.69 and 0.70, respectively). Besides, EVI2 was most capable of capturing large negative anomalies in maize yield in 2005 (R2 = 0.73). Generally, our findings indicate that the application of crop phenology and a blend of EVI2 and NDWI have considerable advantage for remote sensing-based maize and soybean yield models.

The current situation of food insecurity in the world. Achieving the 2015 international hunger goals.Approximately 795 million people are undernourished across the world, 167 million fewer than during the past decade, and 216 million fewer than in 1990–92. More dramatic reductions are occurring in developing regions, even with high population growth. Over the past few years, gains have been impeded by slower and less income-inclusive economic growth and political unrest in some developing areas, e.g., Central Africa and western Asia.2015 is the last year of monitoring of the Millennium Development Goal targets. For all developing regions as a whole, the proportion of undernourished individuals in the population has fallen from 23.3 percent in 1990–92 to 12.9 per cent. Several of these regions, including Latin America, the Asian east and southeast, the Caucasus and Central Asia, and the north and west of Africa have made rapid gains. Advances were also achieved in southern Asia, Oceania, the Caribbean and south and east of Africa, but at too modest a rate to achieve the MDG 1c goal of reducing by half the number of the chronically undernourished

72 out of 129 developing countries, or over half of the countries tracked, have achieved the MDG 1c hunger goal. The majority experienced stable political climate and economic growth, frequently paired with social protection policies aimed at vulnerable groups of people.

Agriculture is the sector that contributes significantly to making our countries economy better. Agriculture is the one which originated civilization. India is an agricultural nation and its economy mainly dependent on crop productivity. Therefore we can say that agriculture can be backbone of all business in our nation. Choosing of each crop is very crucial in the agriculture planning. The crop selection will be based on the various parameters like market price, production rate and the various government policies. Numerous changes need to be made in the agriculture sector to enhance changes in our Indian economy. We can enhance agriculture by implementing machine learning methods which are easily applied on the farming sector. Along with all developments in the machines and technologies employed in farming, helpful and precise information regarding various issues also plays an important role in it. The idea of this paper is to apply the crop selection technique so that this technique assists in resolving a lot of farmers and agriculture problems. This enhances our Indian economy by increasing the rate of crop production yield.

We all know in India that Agriculture is the backbone of our country. This paper is forecasting the yield of nearly all types of crops which are cultivated in India. This script is making new by the use of simple parameters such as State, district, season, area and the user can forecast the yield of the crop in which year he or she wishes to. The research employs sophisticated regression methods such as Kernel Ridge, Lasso and ENet algorithms to forecast yield and employs Stacking Regression for improving the algorithms to provide better prediction.

PROPOSED SYSTEM

In order to overcome the disadvantages of conventional and current crop prediction systems, we suggest an IoT-integrated, machine learning-based crop prediction system that gives real-time suggestions on the basis of real-time environmental data. The system will facilitate precision agriculture by assisting farmers in making precise decisions regarding crop selection, resource allocation,and irrigation scheduling.

Key Features of the Proposed System

1) *Sensor-Based Data Collection (IoT Integration)*

The system will employ IoT-based sensors to capture and track environmental conditions in real-time. The sensors will be installed in the fields to capture vital data that affects crop growth.

- Temperature Sensor: Captures the ambient temperature in the field, allowing crops to be cultivated under ideal thermal conditions.
- Humidity Sensor: Captures air moisture levels, which affect plant transpiration and general growth.
- Soil Moisture Sensor: Identifies the level of water present in the soil, assisting in irrigation management efficiently.
- Water Level Sensor: Tracks water levels in reservoirs and irrigation channels to avoid scarcity or excess water.
- Rainfall Sensor: Quantifies the level of rain, which assists in irrigation adjustment and forecasting drought or excess water.

2) Machine Learning-Based Crop Prediction Model

The sensor-collected data will be processed with Machine Learning (ML) algorithms to forecast the most appropriate crop based on real-time environmental conditions. A Logistic Regression model will be trained using historical agricultural information and real-time sensor data to provide precise crop suggestions.

The ML model will consider different environmental parameters and suggest one of the following crops:

- Cotton
- Strawberry
- Rice
- Sugarcane
- Tomato
- Millets

3) Real-Time Data Processing and Decision Making

The system will execute the following steps to make predictions:

- Data Collection: IoT sensors gather environmental data and transfer it to the central processing
- Data Preprocessing: The raw data gathered is cleaned, structured, and processed for machine learning analysis.
- Prediction Generation: The trained Logistic Regression model processes the data and makes crop recommendations.
- User Display & Decision Support: The last prediction is presented on a web interface, and farmers can see real-time updates and make well-informed decisions.

4) User-Friendly Web Interface (Flask-Based Web Application)

For ease of use, a Flask-based web application will be created, which will be a visual dashboard through which farmers can see live sensor readings and crop suggestions.

- Features of the Dashboard:
- Shows real-time environmental conditions from IoT sensors.
- Delineates the best crop recommendations following the analysis by the machine learning model.
- Enables farmers to monitor historical data for enhanced decision-making.
- Diagnostics for drought risk, soil dryness, or over-moisture, which can be taken on time. nThe Flask framework is efficient and light, and hence the system can be accessed on both desktops and mobiles.

5) Resource Optimization for Sustainable Farming

One of the greatest benefits of this system is its capacity to optimize agricultural resources, leading to increased efficiency and sustainability.

- Water Conservation: Through precise forecasts of soil moisture requirements, farmers can avoid excessive water usage and wastage.
- Fertilizer Management: The system guards against oversupply during fertilizer application, keeping the soilin good health at lower costs.
- Energy Efficiency: Automated irrigation scheduling lowers power consumption and reduces human intervention.
- Yield Enhancement: Selecting the most appropriate crop provides increased production output, financially rewarding farmers.

Advantages of the Suggested System

a) Real-Time Crop Forecasting for Enhanced Decision-Making

- Unlike historical models, this system examines real-time sensor data for real-time forecasts.
- Gives data-driven insights, enabling farmers to make scientifically grounded decisions instead of justdepending on experience.

b) Minimized Resource Squander

- Maximizes water utilization, avoiding unnecessary irrigation.
- Ensures optimal utilization of fertilizers, saving costs and reducing environmental degradation.

- Prevents soil depletion by choosing crops appropriate to existing soil conditions.

c) Increased Productivity and Yield
- Forecasting the correct crop at the correct time guarantees greater yields and improved farm revenues.
- Prevents crop failure risk due to inappropriate crop selection.
- Avoids unnecessary spending on unsuitable crops, making farming profitable.

d) Scalability and Adaptability
- The system can be used for small, medium, and large-scale farms, making it widely applicable.
- Farmers can expand their operations by utilizing real-time insights to manage multiple crop cycles efficiently.
- The model can be continuously improved with new data, making it more adaptable to different climatic conditions.

e) Cost-Effective and Farmer-Friendly
- Low-cost IoT sensors make the system affordable even for small-scale farmers.
- Flask-based web application has a user-friendly interface, lessening the need for technical know-how.
- Saves on labor costs as it automates monitoring and irrigation decisions.

System Implementation Flow of the System
Step 1: Data Gathering (IoT Sensors)
• Sensors are placed in the field to capture real-time environmental data.
• The Arduino microcontroller processes the raw sensor readings and sends them to a central database.
Step 2: Data Preprocessing and Storage
•Sensor readings are cleaned, formatted, and saved in CSV files to be processed later.
•Missing values and outliers are treated through data normalization and transformation methods.
Step 3: Machine Learning Model Training and Prediction
•Historical crop yield data and real-time sensor inputs are trained in a Logistic Regression model.
•The model learns patterns in soil and weather data to forecast the optimal crop.
Step 4: Web Application Development and Deployment
•A web dashboard using Flask is developed to present sensor readings and crop suggestions
•Farmers access the system using a browser-based interface, observing real-time updates.
Step 5: System Testing and Optimization
•The model is tested using actual farming conditions to confirm its accuracy.
•Performance is continuously optimized by tuning parameters and retraining the model.

III. METHODOLOGY

The system suggested in this work adopts a systematic approach to ensure a systematic and effective method of crop prediction with the help of IoT sensors and machine learning. The approach has five primary phases:

1) Step 1: Data Collection (IoT Sensors & Arduino)

This step is concerned with gathering real-time environmental data through IoT sensors placed in the field. These sensors collect vital climate and soil parameters, which are vital for making precise crop predictions.

Process:

a) Sensors Used:
- Temperature Sensor: It measures ambient temperature to decide crop suitability.
- Humidity Sensor: It monitors moisture content in the air, influencing plant transpiration.
- Soil Moisture Sensor: It monitors water content in the soil to ensure optimal irrigation.
- Water Level Sensor: It monitors the water availability for irrigation.
- Rainfall Sensor: It measures precipitation to forecast future water availability.

b) Data Transmission:
- Sensors send gathered data to an Arduino microcontroller, which interprets the readings and sends data to a central server or cloud storage.
- Data transmission can be done using communication protocols like Wi-Fi, Bluetooth, or LoRaWAN.

c) Storage:
- The data is stored temporarily in a structured database or CSV files for further processing.

2) Step 2: Data Preprocessing & Storage

Prior to inputting the gathered data into a machine learning model, it has to be processed for data preprocessing in order to eliminate errors and inconsistencies.

a) Data Cleaning & Formatting:
- Processing missing values with data normalization methods (e.g., mean substitution or interpolation).
- Deleting duplicate or erroneous sensor readings due to hardware failure.
- Specifying the format of all sensor data for consistency in dataset structure.

b) Data Storage:
- The cleaned data is stored in a CSV file, relational database, or NoSQL database.
- Each sensor reading is tagged with a timestamp to monitor changes over time.

3) Step 3: Machine Learning Model Development

The heart of the system is the machine learning model, which makes the prediction of the most appropriate crop based on live environmental data.

a) Model Selection:
- The Logistic Regression algorithm is used because it is effective in classifying crops depending on several environmental factors.
- Alternative models like Decision Trees or Random Forest can be experimented with for better accuracy.

b) Training the Model:
- The model is trained on a data set with historical crop yield data, soil parameters, and climate conditions.
- Training data is divided into 80% training and 20% testing to allow the model to generalize well to new inputs.

c) Model Optimization:

The trained model is tested with accuracy metrics like:
- Precision & Recall: To determine the quality of predictions.
- Confusion Matrix: To evaluate correct and wrong classifications.
- F1 Score: To balance between precision and recall for enhanced performance.
- Hyperparameter tuning is used to enhance accuracy and minimize overfitting.

4) Step 4: System Integration & Deployment

The system is integrated into a web-based dashboard, where farmers are able to use real-time crop recommendations based on live sensor inputs.

a) Web Application Development:
- A Flask-based web interface is constructed to provide real-time sensor readings and crop predictions.
- Farmers can login to see forecast crops, climatic conditions, and resource utilization recommendations.
- The system alerts when there is excess low soil moisture or temperature oscillations in relation to crop health.

b) Cloud Integration & API Development:

- •The data after processing and the output of the ML model are saved to a cloud server that can be accessed easily
- •APIs are generated to retrieve real-time data from the IoT sensors and send it to the web interface.

5) Step 5: Testing & Validation

The system needs to be thoroughly tested for reliability and accuracy before it is implemented in actual agricultural environments.
Testing Procedures:

- Unit Testing: Each component in isolation (e.g., sensors, database, machine learning model, web application) is tested.
- .Integration Testing: Verifies that all components of the system interact smoothly and communicate effectively.
- System Testing: The complete system is implemented in test environments (mock farms) to verify real-world functionality.
- Performance Testing: Measures how the system handles large datasets, concurrent users, and real-time processing.
- Accuracy Testing: Compares ML model predictions with actual crop growth in controlled farming conditions. After successful validation, the system is fine-tuned and deployed in real farming scenarios.

IV. RESULT

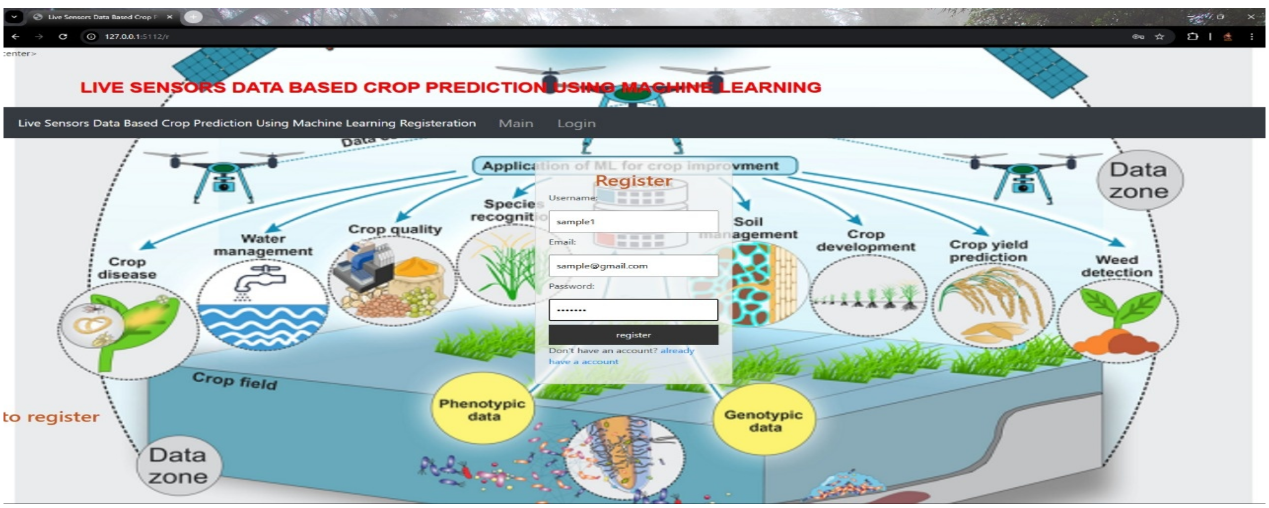


Fig. 1 Register page.

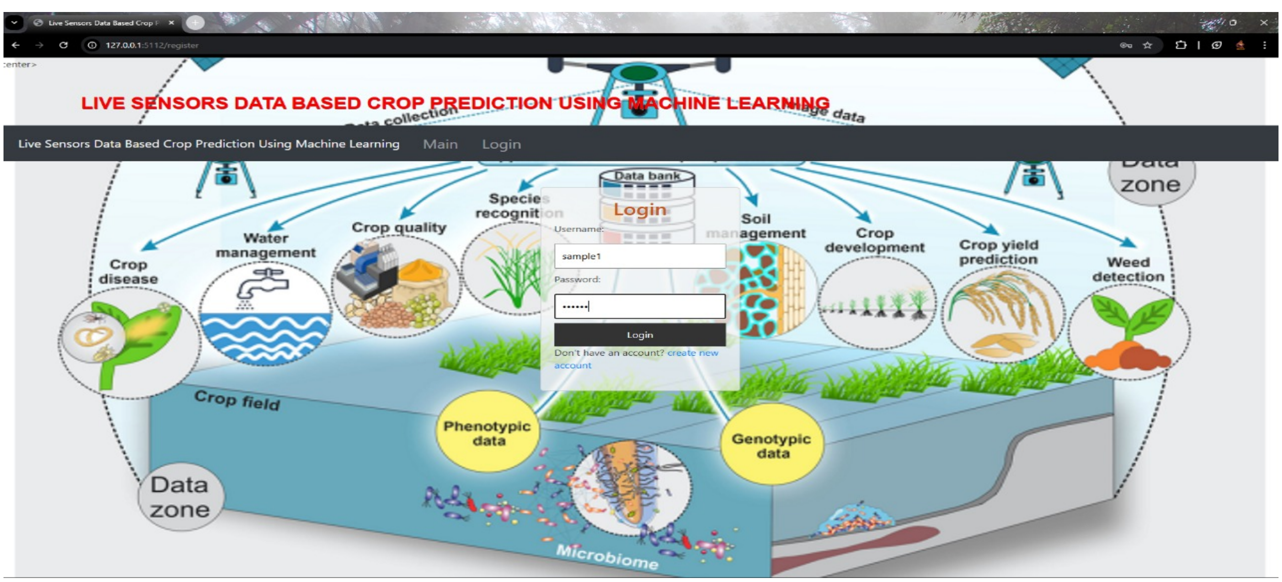


Fig. 2 Login page.

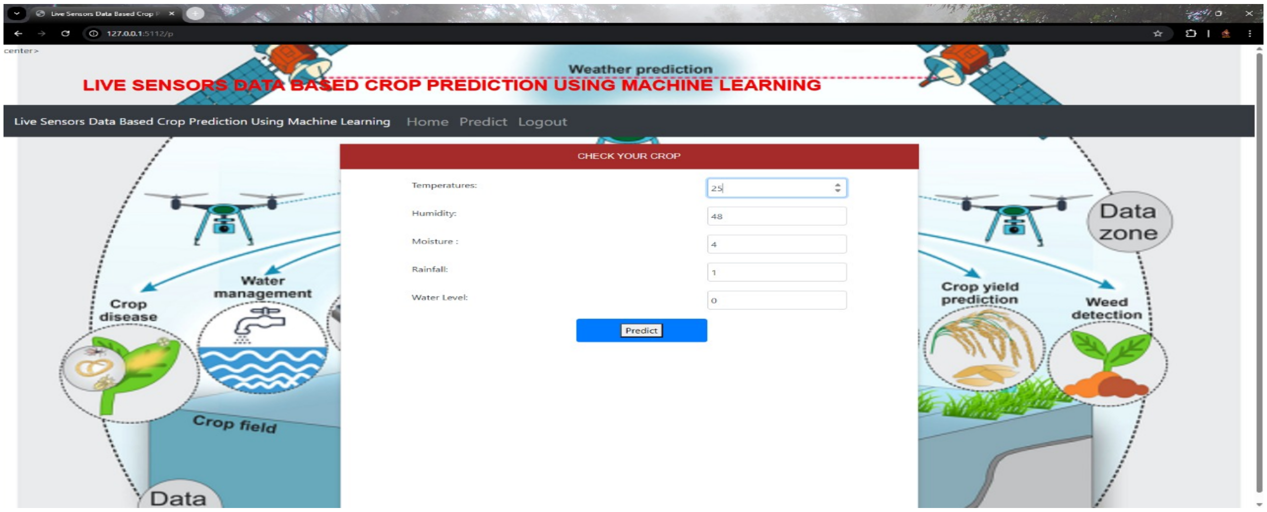


Fig. 3 prediction Analyzed page.

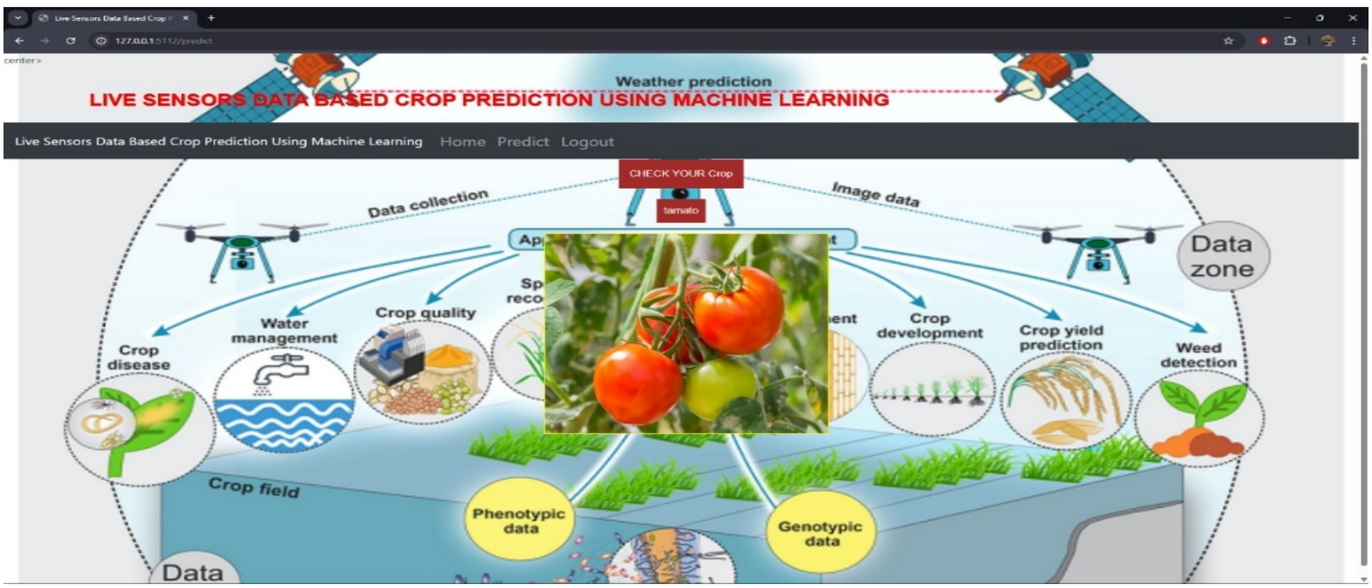


Fig. 4 Prediction page

V. CONCLUSION

In a time when climate change, resource deficiency, and demographic growth present greater threats to food security worldwide, the incorporation of Internet of Things (IoT) and Machine Learning (ML) technologies in agriculture presents an essential and game-changing move ahead. This project, "Crop Prediction Using Sensors and Machine Learning," aims to provide a solution that integrates real-time environmental sensing with predictive analysis in order to enhance crop selection choice, particularly for farmers who belong to under-resourced or rural communities. Through measurement of environmental factors like temperature, humidity, soil water content, rainfall, and water level with Arduino-sensors, and inputting these readings into highly trained ML algorithms such as Logistic Regression and Decision Tree classifiers, this system provides farmers with highly precise, real-time information about which crop is most suitable for their field under prevailing climatic and soil conditions.

This model addresses the age-old problem of farmers depending too much on conventional practices or intuition when deciding what to plant. Through the transition from intuition-driven to data-driven farming, the model greatly enhances grassroots-level decision-making. The fact that the solution can suggest crops such as cotton, tomato, sugarcane, strawberry, rice, and millets with dynamic environmental inputs makes the solution both pertinent and scalable. The performance evaluation indicates that ensemble method employing a Voting Classifier achieves higher accuracy, recall, and resilience over isolated models. Reducing the cases of misclassification and efficient imbalanced data processing, the system enables long-term sustainable farming techniques.

Moreover, the model reduces wastage of water, minimizes use of fertilizers in excess, and helps make wise land use schemes. Going beyond productivity augmentation alone, the technology provides more prosperous economic returns for farmers, improving their ability to withstand unpredictability in weather patterns. It also complements government initiatives like the Soil Health Card Scheme and PM-KISAN, bringing together innovation and policy in tandem. With minimal hardware and infrastructure investment, the model is an economical and highly effective option compared to capital-intensive precision agriculture equipment, allowing for potential deployment at large scale in different rural and semi-urban areas.

REFERENCES

[1] Bolton, D. K., and Friedl, M. A. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology 173:74–84.

[2] Boryan, C.; Yang, Z.; Mueller, R.; and Craig, M. 2011. Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program.

[3] Geocarto International 26(5):341–358. Boulton, C. A.; Shotton, H.; and Williams, H. T. 2016. Using

[4] social media to detect and locate wildfires. In Tenth International AAAI Conference on Web and Social Media.

[5] Breiman, L. 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statist. Sci. 16(3):199–231.

[6] Chu, W., and Ghahramani, Z. 2005. Preference learning with gaussian processes. In Proceedings of the 22nd international conference on Machine learning, 137–144. ACM.

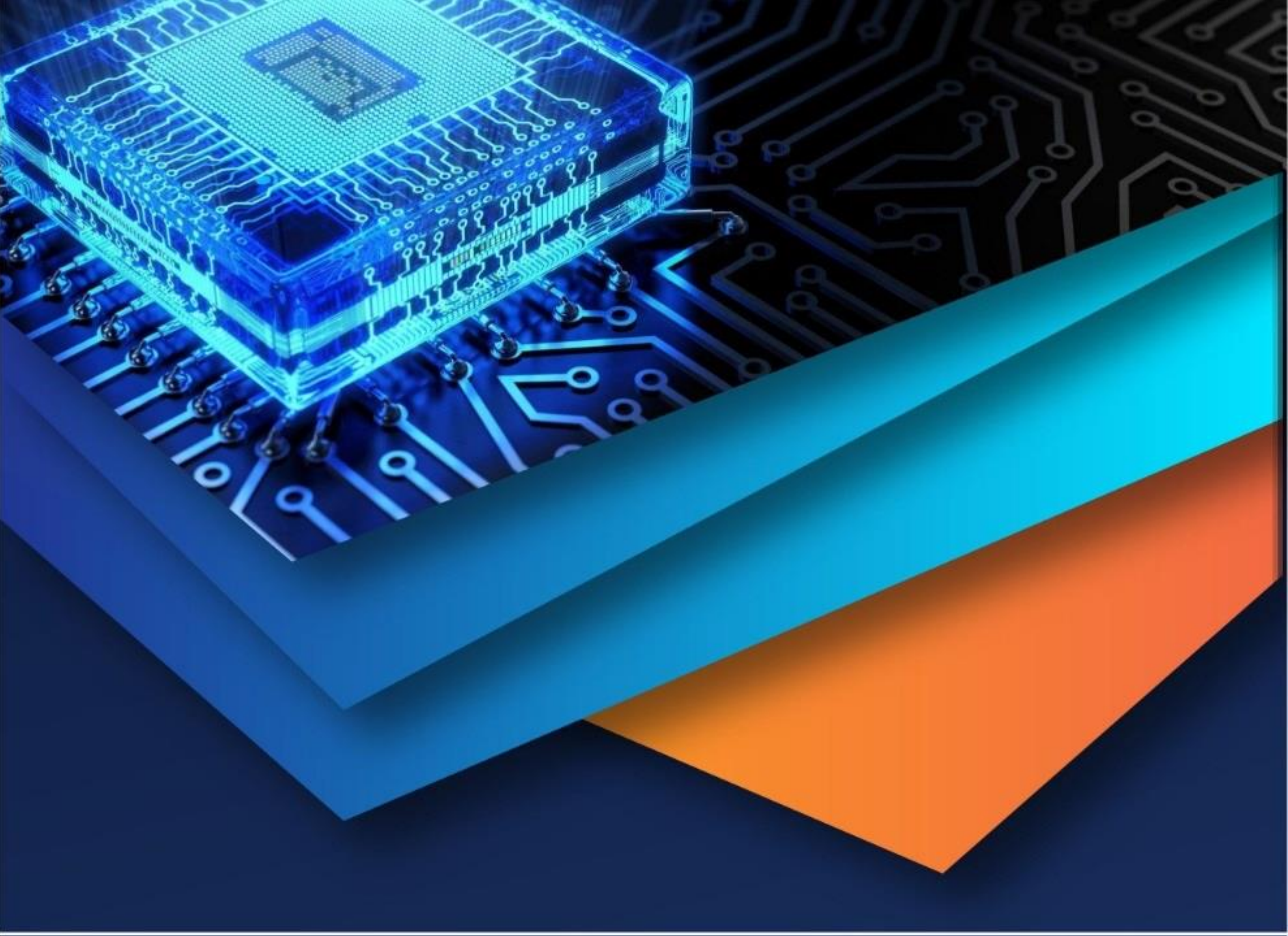
[7] Cressie, N., and Hawkins, D. M. 1980. Robust estimation of the variogram: I. Journal of the International Association for Mathematical Geology 12(2):115–125.

[8] DAAC, N. L. 2015. The modis land products. http://lpdaac.usgs.gov. Dodds, F., and Bartram, J. 2016. The Water, Food, Energy and Climate Nexus: Challenges and an Agenda for Action. Routledge.

[9] Ermon, S.; Xue, Y.; Toth, R.; Dilkina, B.; Bernstein, R.; Damoulas, T.; Clark, P.; DeGloria, S.; Mude, A.; Barrett, C.; and Gomes, C. 2015. Learning large-scale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in East Africa. In AAAI Conference on Artificial Intelligence.

[10] FAO. 2015. The state of food insecurity in the world. Meeting the 2015 international hunger targets: Taking stock of uneven progress.

[11] Fink, D.; Damoulas, T.; and Dave, J. 2013. Adaptive spatiotemporal exploratory models: Hemisphere-wide species distributions from massively crowdsourced ebird data. In AAAI.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY