Kathmandu University
Journal of Science, Engineering and Technology

A Data-Driven Decision Support System for Crop Selection Using IoT Sensors and Machine Learning

Jatin Bhusal[a], Ankit Mahato[a], Sanket Shrestha*[a], Prashanta Acharya[a], and Jeewan Khatri[a]

[a]Sunway International Business School, Birmingham City University, Kathmandu, Nepal

Abstract

Unpredictable weather patterns, soil degradation, and the unregulated use of pesticides and chemical fertilizers have severely impacted agricultural productivity in Nepal. These challenges have made crop planning increasingly difficult, leading to substantial financial losses and deteriorating mental health among farmers. In response, this study proposes a decision support system that integrates IoT-based environmental sensing with machine learning to provide real-time crop recommendations. The system collects key parameters including soil pH, NPK levels, moisture content, humidity, soil temperature, and rainfall which are fed into a supervised learning model to enable precise, data-driven crop selection. Among the models evaluated, the Support Vector Machine (SVM) classifier achieved a prediction accuracy of 96.75% with minimal overfitting, demonstrating the potential of the proposed approach to improve agricultural decision-making.

Keywords: Precision agriculture; Crop Recommendation System; Internet of Things (IoT; Machine Learning; Soil Parameter Monitoring; Agricultural Decision Support System.

1. Introduction

Even though nearly 66% of Nepalese population is engaged in agriculture [1] and has been contributing significantly in the GDP of Nepal, there are still numerous challenges our farmers are facing on a daily basis. Almost 40% of the total suicides in 2021-2023 were of farmers, while 20- 50 % crops are lost due to infestation and spoilage [2]. Main challenges lies in uncertain weather patterns and irregular seasons due to global warming. Not only that, random use of pesticides and other chemicals have mutated the soils original properties, and farmers are unaware of that. This study proposes an IoT and ensemble machine learning based architecture that can predict the best yielding crop for the farmers in real time.

The contributions of this research are:

1. A real-time crop recommendation system developed by integrating IoT-based environmental sensing with machine learning models.

2. Utilization of low-cost sensors to collect parameters such as soil pH, NPK levels, moisture, humidity, soil temperature, and rainfall.

3. A supervised learning pipeline implemented and evaluated, with the SVM classifier achieving 96.75% accuracy and low overfitting,

4. A deployment framework designed to support modular, scalable, and cost-effective usage across community, individual, and industrial levels.

5. The system validated on 20 regionally relevant crops, demonstrating its practical applicability for precision agriculture in Nepal.

*Corresponding author. Email: sanketshrestha09@gmail.com

The remainder of this paper is structured as follows: Section 2 reviews Recent Studies. Section 3 describes the System Overview. Section 4 outlines Dataset and Methodology. Section 5 gives IoT Sensor Architecture. Section 6 gives System Integration and Real-Time Prediction Pipeline. Section 7 gives Results and Discussions and finally Section 8 concludes with Conclusion and Future Work.

2. Related Work

Recent studies have explored machine learning approaches for crop recommendation and yield prediction using soil and weather data. Multiple algorithms have been evaluated, including Naive Bayes, Random Forest, K-Nearest Neighbors, Support Vector Machines, and deep learning models [3, 4, 5, 6]. These models utilize various input parameters such as soil properties (pH, N, P, K), climate data (temperature, rainfall, humidity), and other environmental factors.

Naive Bayes and Random Forest have shown particularly high accuracy, often exceeding 99% for crop recommendation [3, 4, 5]. For yield prediction, K-Nearest Neighbors and LSTM models have demonstrated strong performance[5]. These machine learning-based systems aim to assist farmers in making informed decisions about crop selection and management, potentially improving agricultural productivity and sustainability[6, 5].

Recent systematic reviews have reinforced the prevalence of Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks as dominant deep learning models for crop yield prediction, particularly when using features such as temperature, rainfall, and soil type [7] . Yadav et al. proposed a hybrid model combining traditional classifiers with ANN to enhance soil fertility analysis and crop yield prediction, demonstrating improved accuracy through deeper feature learning [8].

Similarly, Abbas et al. explored proximal sensing data and applied algorithms including Support Vector Regression (SVR), k-Nearest Neighbors (k-NN), and Linear Regression (LR), finding SVR to outperform others in predicting potato yield across multiple datasets [9].

In terms of crop recommendation, recent work by Rawat et al. evaluated diverse machine learning models—such as Decision Tree, Naive Bayes, KNN, Random Forest, and XGBoost—highlighting their capability to support data-driven agricultural decision-making [10]. Pande et al. extended this by developing a mobile-based crop recommender system that integrated GPS location and user-provided soil information, where Random Forest achieved 95% accuracy in recommending suitable crops and fertilizer schedules [11]. Collectively, these studies demonstrate the growing maturity and practical value of AI-enabled agricultural support systems, particularly in regions where climate variability and soil degradation challenge traditional farming practices.

3. System Overview

The proposed system consists of three integrated components:

1. an IoT-based environmental sensing module,
2. a machine learning model for crop recommendation,
3. a real-time data processing pipeline connecting sensor inputs to prediction outputs

Figure 1 illustrates the overall system architecture.

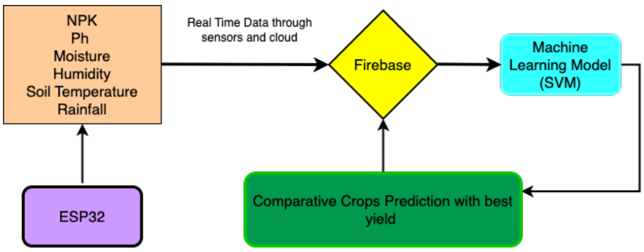

Figure 1: System architecture showing the integration of IoT sensors, Firebase, and the machine learning model for real-time crop recommendation.

4. Dataset and Methodology

4.1. Dataset Description

The dataset used in this study was sourced from an open-access repository on Figshare [?]. It contains 2,200 entries, each representing a unique set of environmental and soil conditions including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall, along with the crop best suited to those conditions. The target variable is the recommended crop type, covering 20 different classes such as rice, maize, lentil, kidney beans, mango, and watermelon. All input features are continuous variables, while the crop label is categorical. The dataset is well-balanced and offers a diverse mix of soil and climate scenarios, making it suitable for building a generalizable crop recommendation model.

4.2. Machine Learning Model

The crop recommendation model was formulated as a multi-class classification problem, where the goal was to predict the most suitable crop based on real-time environmental and soil parameters. The input features included nitrogen (N), phosphorus (P), potassium (K), temperature (°C), humidity (%), pH, and rainfall (mm), all collected via IoT sensors or cloud-based APIs.

A comparative experiment was run among three SOTA architectures : Random Forest, SVM and XG Boost.The SVM classifier was chosen due to its robustness, less overfitting, and ability to handle nonlinear relationships and multi-class targets. The dataset comprising 2,200 instances and 20 crop classes was obtained from an open-source Figshare repository [?]. The dataset was preprocessed using Min-Max normalization and label encoding, and split into training and testing sets in an 80:20 ratio.

After training, the model was exported and integrated with the real-time pipeline. Incoming sensor data from Firebase was preprocessed and passed to the trained model, which returned the top recommended crops. This prediction was then written back to Firebase for user access via the web dashboard. The lightweight nature of the model allows for near real-time inference, making it well-suited for deployment in resource-constrained environments.

5. IoT Sensor Architecture

The goal has been to setup a compact, light-weight, and robust IoT setup that can take real-time data and send them to the system for real-time prediction. Table 1 shows the sensors we have used and what value they predict:

Table 1: Sensor and Data Source Mapping

Parameter	Source/Sensor
NPK	Regional Data
pH Value	pH Sensor
Moisture	Cloud API
Humidity	DHT11
Soil Temperature	DS18b20
Rainfall	Cloud API

6. System Integration and Real-Time Prediction Pipeline

The proposed system integrated IoT-based data collection setup, cloud-based data management, and a machine learning-based crop recommendation model into a unified, real-time decision support pipeline.

6.1. Sensor Data Acquisition

To gather real-time environmental data from the field, the system used a combination of hardware sensors and external data sources. Key parameters such as soil pH, temperature, and humidity were measured using a pH sensor, the DHT11 sensor, and the DS18b20 sensor, respectively. Additional data such as NPK values and rainfall were retrieved from regional agricultural databases and third-party weather APIs like from Nepal Agriculture and Research Center (NARC). The ESP32 microcontroller acted as the central hub, reading sensor inputs and transmitting the collected data to a Firebase Realtime Database.

6.2. Cloud-Based Data Management

Firebase was used as the cloud backend to manage incoming data and keep it synchronized in real time. As new sensor readings were collected, they were automatically uploaded and stored in Firebase, ensuring that the most recent data was always available for processing. This setup allowed for seamless and low-latency communication between the sensing hardware and the machine learning model, supporting real-time decision-making and prediction.

6.3. Model Integration and Prediction Pipeline

The trained Support Vector Machine (SVM) model was integrated into a real-time pipeline designed to operate seamlessly with the IoT sensing infrastructure and Firebase cloud backend. This integration enables automated prediction and instant delivery of crop recommendations based on live environmental conditions.

A server-side script was developed to listen continuously for incoming data updates in the Firebase Realtime Database. As new data entries were pushed by the ESP32 microcontroller, including features such as pH, temperature, humidity, and other environmental parameters, the script retrieves the complete input vector from Firebase.

Upon retrieval, the input data underwent preprocessing steps including normalization and validation to ensure consistency with the format used during training. The preprocessed vector was then passed to the SVM model, which returned the predicted crop class based on the learned decision boundaries.

The prediction result was immediately written back to Firebase, making it accessible through the web-based user interface. This allows users to view updated crop recommendations in near real time. The pipeline ensures low-latency inference and maintains responsiveness even under variable data transmission rates from the field sensors.

This modular integration approach makes the system scalable and adaptable, allowing for future upgrades such as adding more features, deploying lightweight models on the edge, or integrating alert systems for abnormal environmental conditions.

6.4. User Interface and Recommendation Delivery

The predicted crop recommendations were displayed to the user via a web dashboard connected to Firebase. This enables farmers or field operators to view recommended crops in real time based on the current environmental conditions of their field. The interface has been designed to be accessible on both desktop and mobile devices.

This tightly coupled, end-to-end integration ensures a fully functional, low-cost, and scalable decision support system capable of aiding real-time agricultural decision-making.

7. Results and Discussions

The proposed crop recommendation model was evaluated using three machine learning algorithms: Random Forest, XGBoost, and Support Vector Machine (SVM). The evaluation was conducted on a test set comprising 20% of the total dataset (440 instances), and performance was measured using accuracy, precision, recall, and F1-score.

7.1. Model Performance Metrics

Table 2 presents the classification performance of each model. Random Forest achieved perfect scores on all evaluation metrics, indicating a highly accurate but potentially overfitting model. While such performance appears ideal, it may also suggest overfitting to the training data or the presence of subtle data leakage. Further validation through k-fold cross-validation or evaluation on independent datasets is recommended to assess true generalization capability. XGBoost also performed well with an accuracy of 98.92%, while SVM showed slightly lower accuracy at 96.75% but with the lowest overfitting gap.

Table 2: Performance Comparison on Test Dataset

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	1.0000	1.0000	1.0000	1.0000
XGBoost	0.9892	0.9904	0.9892	0.9893
SVM	0.9675	0.9744	0.9675	0.9636

7.2. Overfitting and Learning Efficiency

To further analyze generalization, Fig. 2 compares the overfitting gap, best validation accuracy, and learning efficiency across all three models. Although Random Forest achieved perfect scores, its overfitting gap was very large and greater than that of SVM. SVM demonstrated the best balance between validation performance and efficiency, with a learning efficiency score of 1.2094, making it the preferred model for real-time deployment.

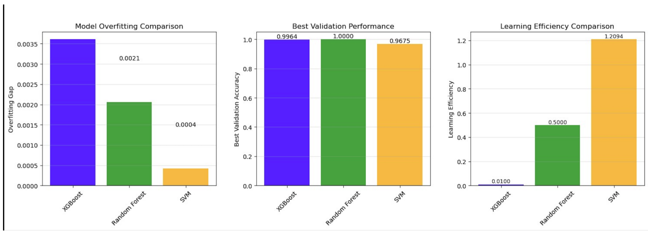

Figure 2: (a) Overfitting gap comparison, (b) best validation accuracy, (c) learning efficiency across models.

7.3. Model Selection Justification

Despite Random Forest achieving the highest performance on test data, its overfitting behavior raised concerns regarding generalization in real-world conditions. SVM, although slightly lower in accuracy, showed the lowest overfitting gap and the highest learning efficiency. Therefore, SVM was selected for deployment due to its better trade-off between accuracy, generalization, and computational simplicity.

8. Conclusion and Future Work

This research presents an end-to-end crop recommendation system that integrates low-cost IoT sensors, cloud infrastructure, and machine learning techniques to enable real-time, data-driven agricultural decision-making. The system collects environmental data such as pH, soil temperature, humidity, and rainfall using a compact IoT architecture centered around an ESP32 microcontroller and cloud-connected sensors. This data is transmitted to Firebase, preprocessed, and then passed to a trained Support Vector Machine (SVM) model for crop prediction. Among the evaluated models, SVM demonstrated the best trade-off between accuracy (96.75%), low overfitting, and computational efficiency, making it ideal for deployment in resource-constrained and real-world environments.

The results validate the feasibility of deploying AI-driven agricultural systems that support smallholder farmers in regions like Nepal, where unpredictable climate patterns, soil degradation, and limited access to expert advice pose significant threats to sustainable farming. By combining environmental sensing and machine learning, this system can serve as a reliable decision support tool to enhance productivity, reduce risks, and improve the livelihoods of farming communities.

Future Work

Future improvements to the system will focus on the following areas:

- **Data Expansion:** Collecting larger and more diverse datasets across regions and seasons to improve model robustness and generalization.
- **Feature Enrichment:** Including additional agronomic parameters such as soil organic carbon, electrical conductivity (EC), and micronutrients to enhance prediction accuracy.
- **Edge Deployment:** Deploying the trained model on low-power edge devices (e.g., Raspberry Pi, ESP32 with TinyML) to enable offline inference in remote locations.
- **Temporal Modeling:** Integrating time-series weather and crop data to support seasonal planning and early warning systems for crop failure.
- **User-Centric Enhancements:** Adding multilingual interfaces, mobile app support, and fertilizer scheduling features to improve usability and farmer engagement.
- **Field Validation:** Conducting extensive field trials to validate the system's performance under real farming conditions and refine its practical utility.

Acknowledgment

The authors thank Sunway College Kathmandu and its faculties for facilitating us with resources and guidance to make this research possible. The authors thank all previous authors and researchers for setting the foundations for this work.

References

[1] FAO. Nepal at a glance (2025). URL https://www.fao.org/nepal/our-office/nepal-at-glance/en, [Online].

[2] Poudel A. More farmers and students are killing themselves in nepal, data shows (2024). URL https://kathmandupost.com/national/2024/09/10/more-farmers-and-students-are-killing-themselves-in-nepal-data-shows, [Online].

[3] Ghilan A, El Afou Y & El Akkad N. Region-wise weather data-based crop recommendation using machine learning algorithms. In: *2024 3rd International Conference on Embedded Systems and Artificial Intelligence (ESAI)*. IEEE (2024), pp. 1–6.

[4] Changela A, Kumar Y & Koul A. Machine learning-based approaches for crop recommendations and prediction. In: *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*. IEEE (2023), pp. 370–376.

[5] Anushri T, Bhavadharani B, Suganthi P et al. Crop recommendation and yield prediction using machine learning and deep learning models. In: *2024 International Conference on Integration of Emerging Technologies for the Digital World (ICIETDW)*. IEEE (2024), pp. 1–6.

[6] Gambhir M S, Sharma M, Agarwal K, Kumar K, Kumar L & Chaudhary M, Crop recommendation system using machine learning, *International Journal for Research in Applied Science and Engineering Technology*, 11(2) (2023) 640–644.

[7] Van Klompenburg T, Kassahun A & Catal C, Crop yield prediction using machine learning: A systematic literature review, *Computers and electronics in agriculture*, 177 (2020) 105709.

[8] Yadav J, Chopra S & Vijayalakshmi M, Soil analysis and crop fertility prediction using machine learning, *Machine Learning*, 8(03).

[9] Abbas F, Afzaal H, Farooque A A & Tang S, Crop yield prediction through proximal sensing and machine learning algorithms, *Agronomy*, 10(7) (2020) 1046.

[10] Rawat P, Bajaj M, Vats S, Sharma V et al. An analysis of crop recommendation systems employing diverse machine learning methodologies. In: *2023 International Conference on Device Intelligence, Computing and Communication Technologies,(DICCT)*. IEEE (2023), pp. 619–624.

[11] Pande S M, Ramesh P K, Anmol A, Aishwarya B, Rohilla K & Shaurya K. Crop recommender system using machine learning approach. In: *2021 5th international conference on computing methodologies and communication (ICCMC)*. IEEE (2021), pp. 1066–1071.

[12] Ingle A. Crop recommendation dataset. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset (2024). [Online; accessed 10-July-2025].