Internet of Things and Artificial Intelligence Powered Crop Suitability Detection System for Sustainable Farming

A S M Ahsanul Sarkar Akib, Abu Zahid Md Jalal Uddin, Fatema Jahan Sifa, Mahadir Islam, Md. Easin Arafat, and Touhid Bhuiyan

[1] Robo Tech Valley, Dhaka, Bangladesh
ahsanulakib@gmail.com
[2] Washington University of Science and Technology, Alexandria, USA
abu.student@wust.edu
[3] Mawlana Bhashani Science and Technology University,Bangladesh
fatemajahan552@gmail.com
[4] Birshrestha Munshi Abdur Rouf Public College
mahadirislam0@gmail.com
[5] B A F Shaheen College, Dhaka, Bangladesh
easinarafat.bd.2006@gmail.com
[6] Washington University of Science and Technology, Alexandria, USA
touhid.bhuiyan@wust.edu

Abstract. This paper presents an innovative IoT-based autonomous farming system utilizing machine learning models to assist farmers in determining suitable crops based on real-time environmental data. The system integrates IoT sensors, including soil pH, NPK, temperature, and humidity sensors, to collect data from the field. The ESP-8266 NodeMCU processes this data and transmits it to a cloud database. A range of machine learning algorithms were applied to the dataset, including Logistic Regression, Gaussian Naive Bayes, Support Vector Classifier, K-Nearest Neighbors, Decision Tree Classifier, Extra Trees Classifier, Random Forest, Bagging Classifier, Gradient Boosting, and AdaBoost. The highest accuracy was achieved with the Random Forest Classifier (97.05%), followed closely by the Bagging Classifier (96.59%) and Gradient Boosting (96.36%). The AdaBoost model showed poor performance with an accuracy of 10.23%. The system's predictions are accessible to farmers via a web or mobile application, enabling them to make informed decisions about crop cultivation. This IoT and machine learning-based approach reduces human intervention, optimizes farming practices, and enhances crop yield potential. The system provides real-time crop recommendations, making farming more efficient and sustainable.Use of appropriate algorithms on the sensed data can help in recommendation of suitable crop.

Keywords: IoT · Autonomous Farming · Machine Learning · Crop Prediction · Precision Agriculture · ESP8266 NodeMCU · Crop Recommendations · Real-time Data · Algorithmic Crop Recommendation · Random

Forest · Logistic Regression · Gaussian Naive Bayes · K-Nearest Neighbors · web application · environmental sensors.

1 Introduction

Agriculture is a vital sector that underpins human sustenance and economic growth. Despite its importance, modern farming faces numerous challenges, including unpredictable weather patterns, soil nutrient deficiencies, and inefficient resource management, which often result in significant crop losses and financial hardships. In India alone, the agricultural sector incurs losses exceeding 11 billion dollars annually due to these issues. To tackle these challenges, the integration of Internet of Things (IoT) and Machine Learning (ML) offers a revolutionary solution to improve agricultural productivity and sustainability.

This paper introduces an IoT-based autonomous farming system that utilizes advanced machine learning models to optimize critical agricultural processes such as crop selection, irrigation, and fertilization. By employing a suite of IoT sensors—such as soil pH sensors, NPK sensors, soil temperature sensors, and humidity sensors—the system collects real-time data from the farming environment. This data is then analyzed using various machine learning models, including Random Forest, K-Nearest Neighbors, and Support Vector Classifiers, among others, with the Random Forest Classifier achieving the highest accuracy of 97.05%. Additionally, a web/mobile application integrates these insights, providing farmers with actionable recommendations to enhance farm management and crop yields. By combining IoT and ML technologies, the proposed system aims to minimize crop losses, optimize resource utilization, and advance autonomous farming practices.

2 RELATED WORK

The study aims to reduce crop loss in India by integrating IoT and machine learning technologies for crop selection, autonomous watering, and fertilizer recommendation [1]. The author through smart irrigation systems, India's agriculture sector uses machine learning algorithms to improve crop yield and reduce irrigation waste [2]. The paper introduces a machine learning-based crop monitoring system that utilizes environmental data to inform farmers about disease conditions, enhance detection, predict disease spread, and recommend appropriate pesticides [3]. Smart agriculture systems in India, combining IoT and ML technologies, improve productivity, predict planting locations, prevent water waste, enable strategic decision-making, save costs, and enable remote crop monitoring [4]. Precision farming using IoT and machine learning algorithms improves crop productivity by predicting suitable crops based on soil parameters, using open-source datasets for testing and training [5]. IoT-based Smart Farming and machine learning enhance agriculture by real-time monitoring and crop recommendations, ensuring farmers adapt to climate change and meet increasing food demands [6]. The paper introduces a cost-effective IoT and machine learning system in India that uses ground modules to predict crop conditions, reducing manual labor and improving agricultural efficiency [7]. The paper introduces a Smart IoT-based Agriculture system that uses an Arduino board and sensors to aid farmers in crop management, increase yield, and reduce wastage [8]. This paper explores machine learning concepts for an IoT-based Smart Agriculture system using the ThingSpeak cloud platform, analyzing current IoT state, challenges, and suggesting that integrating IoT can improve crop quality and productivity [9]. Precision agriculture, combining machine learning and IoT-enabled farm machinery, is a key solution for economic growth in agriculture. It improves soil parameter prediction, crop yield prediction, disease detection, livestock production, and sustainable productivity [10]. This study examines the use of deep reinforcement learning (DRL) in IoT-based agriculture, analyzing 51 articles using a Systematic Literature Review approach. It aims to understand DRL's development, implementation, and tool use, predicting continued development in the Industrial Revolution 4.0 [11]. The proposed IoT-based Agro Automation System uses Machine Learning Algorithms to monitor agricultural parameters, using multiple sensors and a hybrid app. It uses PTC ThingWorx cloud computing for global monitoring and future drone applications [12]. The paper proposes a smart farming system using AI to optimize plant growth and support farmers, using wireless sensor networks for data collection, cleaning, storage, and predictive processing. Future research should address AI limitations and improve training speed and accuracy [13]. The paper presents a crop monitoring system using machine learning to predict environmental diseases, improve disease detection, and provide pesticides, thereby aiding farmers in informed decision-making for profit and crop health [14]. The Indian agriculture sector is utilizing an IoT-based device to provide real-time notifications for resources and products. The system uses Python scripts to integrate sensors and electronic devices, achieving success in 84.8% of test cases [15].

3 System Architecture and Workflow

The proposed system seamlessly integrates Internet of Things (IoT) and Machine Learning (ML) technologies to address key agricultural challenges with precision and efficiency. IoT devices, including sensors and actuators, collect real-time environmental data such as soil temperature, humidity, NPK levels, and pH. This data is analyzed using advanced ML algorithms to derive actionable insights and reliable predictions. The ML component processes historical and real-time data to detect patterns, anomalies, and trends, enabling the optimization of crop selection, fertilization schedules, and disease forecasting. The system continuously improves through iterative learning, enhancing its predictive accuracy over time.

3.1 System Design Overview

The IoT-based farming system employs sensors to monitor environmental parameters such as soil temperature, humidity, NPK, and pH. Data from these

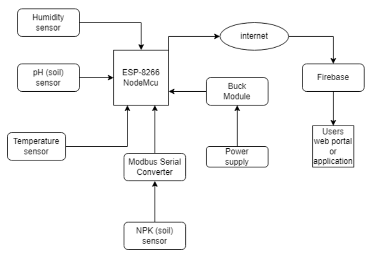

Fig. 1. Block Diagram of IoT-Based Autonomous Farming System

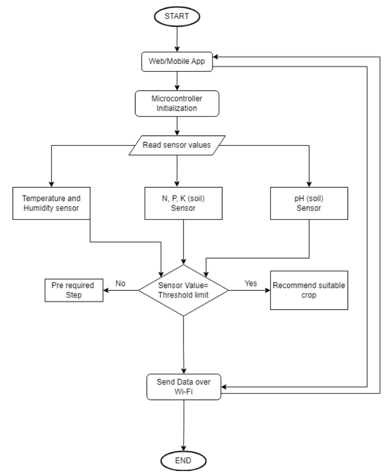

Fig. 2. Flowchart of IoT-Based Autonomous Farming System

sensors is collected by the ESP-8266 NodeMCU and transmitted to a Firebase cloud database for real-time access. The system architecture, as shown in Fig. 1, integrates machine learning models for decision-making, enabling autonomous farming practices. ML algorithms such as Random Forest Classifier, KNeighbors Classifier, and Gradient Boosting Classifier analyze the collected data to predict optimal crop choices and resource management strategies. The system also features a mobile application for user interaction. After logging into the app, users can access sensor data, manage resources like sprinklers, and receive crop recommendations based on real-time analysis. The app displays sensor readings as percentages, offering insights into soil conditions and supporting data-driven farming decisions. The workflow of the system, which processes sensor inputs to generate actionable insights, is illustrated in Fig. 2.

3.2 Hardware

Table 1 summarizes the hardware components used in this project. The ESP-8266 NodeMCU acts as the central microcontroller, supporting IoT protocols and Wi-Fi communication. Sensors ensure accurate environmental monitoring and data collection, enabling efficient system functionality.

3.3 Software

Google Colab, with TensorFlow and Keras, trained machine learning models for crop prediction using IoT sensor data. Firebase enabled real-time data storage and API-based exchange with IoT sensors via the ESP8266 NodeMCU. A web/mobile app offered farmers real-time monitoring and crop recommendations for precise farming decisions.

Table 1. Hardware Components and Their Descriptions

Component	Description
ESP-8266 NodeMCU	An IoT platform with Wi-Fi, 16 GPIO pins, 2 UART ports, and an 80–160 MHz CPU. It supports HTTP and MQTT protocols, 4 MB storage, and is programmable via Arduino IDE or Lua scripts.
Soil NPK Sensor	Measures nitrogen, phosphorus, and potassium levels. It is waterproof, durable, and suitable for precision agriculture and soil research.
Soil pH Sensor	Determines soil pH by measuring hydrogen ion concentration using pH and reference electrodes, requiring regular calibration for accuracy.
Soil Temperature Sensor	Converts electrical signals into temperature units (°C, °F, °K). Compatible with data loggers for precise temperature monitoring.
Soil Humidity Sensor	Measures soil water content using portable probes or fixed sensors, essential for agriculture and environmental monitoring.

3.4 Machine Learning

This system first checks the temperature and humidity of the chosen device before measuring the soil's NPK and pH values using respective sensors. Temperature and humidity data are collected via an IoT Agriculture App on a cloud platform. The collected data is processed through a machine learning model, where K-Means clustering is applied to group the values.

Table 2. Overview of Classifiers Used in the Project

Classifier	Description
Random Forest Classifier	An ensemble technique using decision trees and majority voting to improve accuracy, handle high-dimensional data, and prevent overfitting.
KNeighbors Classifier	Predicts based on proximity, assuming similar data points are near each other. Effective for classification but computationally intensive on large datasets.
Decision Tree Classifier	A flexible, interpretable approach for organizing decisions based on input data, aiding in predictive modeling.
Extra Tree Classifier	An ensemble method using multiple decision trees, reducing bias while increasing variance compared to Random Forest.
Support Vector Classifier (SVC)	Classifies data using decision boundaries (hyperplanes). Effective for linear and non-linear tasks, especially in high-dimensional spaces.
Gaussian Naive Bayes Classifier	Assumes features follow a normal distribution and uses Bayes' Theorem to classify based on posterior probabilities.
Logistic Regression Classifier	Predicts probabilities for binary/multiclass classification using log-loss and evaluates with metrics like accuracy, precision, recall, and AUC-ROC.
AdaBoost Classifier	Combines weak classifiers, weighting hard-to-classify cases more heavily, and aggregates predictions for improved results.
Gradient Boosting Classifier	Iteratively combines weak learners like decision trees to create a strong model, improving performance with each step.
Bagging Classifier	Uses multiple models trained on random data subsets to reduce variance, prevent overfitting, and improve accuracy.

The accuracy of ten classifiers—Random Forest, KNeighbors, Decision Tree, Extra Tree, SVC, Gaussian Naive Bayes, Logistic Regression, AdaBoost, Gradient Boosting, and Bagging—is compared. The sensor and cloud data are transmitted through the most suitable method. Finally, the system suggests the best crop for the location. The model is trained using a Kaggle dataset [16]. Table 2 outlines the classifiers used, with ensemble methods enhancing accuracy and reducing overfitting, while others like KNeighbors, SVC, and Logistic Regression focus on proximity, decision boundaries, and probability-based predictions.

4 Result and Discussion

The study used IoT sensor data to train machine learning models for autonomous farming systems. The top performers were Random Forest with 97.05% accuracy. These models can optimize crop selection and increase yields, and their integration with IoT-based data collection systems offers potential for autonomous farming. The study underscores the importance of machine learning in precision agriculture.

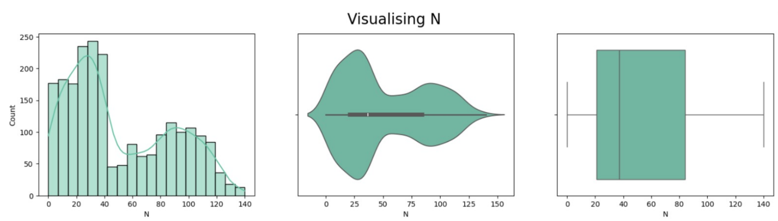

Fig. 3. The distribution of nitrogen levels (N)

Nitrogen (N) Distribution The Fig. 3 visualizes nitrogen (N) levels through three plots. The histogram with a KDE overlay shows most nitrogen values concentrated between 0 and 50, with some extending beyond 100. The violin plot highlights a left-skewed distribution, indicating a concentration of lower nitrogen levels. The boxplot reveals a median around 40, a range from approximately 10 to 90, and outliers at higher values. These plots indicate that nitrogen levels in the dataset are skewed toward lower values but display significant variance, emphasizing the need for monitoring N in crop suitability predictions.

Correlation Matrix The correlation heatmap (Fig. 4) highlights relationships between key variables such as N (nitrogen), P (phosphorus), K (potassium), temperature, humidity, and pH. A strong positive correlation (0.74) is observed between P and K, suggesting their similar influence on soil and crop suitability. Weak negative correlations are noted between N and P (-0.23) and N and K (-0.14), indicating an inverse relationship. Humidity shows a slight positive correlation with N (0.19). These insights help the model account for interactions among factors, enabling nuanced crop suitability predictions.

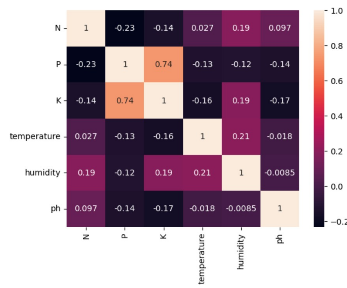

Fig. 4. The Correlation Matrix

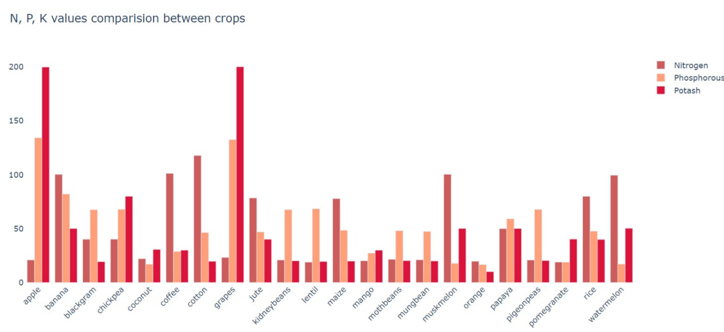

Fig. 5. The distribution of nitrogen levels (N)

Density of Nitrogen (N) The density plot Fig.5 shows a multimodal distribution for nitrogen values, peaking around 25 and 50. This indicates that there are distinct groupings of nitrogen levels within the data, potentially representing different zones or types of soil conditions within the training set.

Discussion The nitrogen data distribution, as shown in the histogram (Fig.5), violin plot (Fig.5), and density plot (Fig.5), underscores the importance of N concentration in crop health. Areas with lower N levels may require targeted supplementation to maintain productivity. The positive correlation between P and K implies these nutrients often coexist in soil, suggesting the use of balanced P and K fertilizers for effective crop growth. The variability in nitrogen levels, evident in the violin plot and density plot, highlights the need for localized soil treatment strategies to ensure sustainable agriculture. Weak correlations between N and other nutrients suggest nitrogen management should be treated independently, while managing P and K collectively may yield better results. The low correlations of temperature and humidity with nutrients, as noted in the correlation matrix (Fig. 4), indicate limited direct impacts. However, these environmental factors could indirectly affect crop health through processes like water absorption and heat stress.

4.1 Comparison

Here's a structured comparison of the reference paper number [6], based on various aspects:

Table 3. comparison of the reference paper number [6]

Aspect	**Comparison**
ML Models	10 models (Logistic, SVC, KNN, etc.) vs. 3 models (KNN, SVC, Decision Tree)
Data Parameters	Diverse: NPK, pH, temperature, humidity vs. Limited: pH, temperature, humidity, rainfall
IoT Sensor Usage	Real-time data: NPK, pH, temperature, humidity vs. Data: pH, temperature, humidity, rainfall
Prediction Objective	Best crop based on detailed soil/environmental data vs. Suitable crop based on general environmental data
User Interface	Web/mobile app for real-time recommendations vs. Website for crop and fertilizer suggestions

4.2 Suggest Recommend Crop cultivated using machine learning

Table 4. Accuracy of various classifiers

Model	Accuracy	Test acc	Train acc
LG	0.9090	0.8909	0.9005
GNB	0.9590	0.9454	0.9676
SVC	0.925	0.8454	0.9161
KNN	0.8977	0.9590	1.0
DT	0.9590	0.9636	1.0
ET	0.8090	0.9636	1.0
RF	0.9727	0.9681	1.0
BGC	0.9636	0.95092	0.92504
GBC	0.9636	0.9089	1.0
Adaboost	0.10227	0.90634	0.91002

In table 4, the accuracy of 10 distinct models has been compared.The project evaluated machine learning models' accuracy, testing, and training, with Logistic Regression demonstrating moderate accuracy, Gaussian Naive Bayes high, Support Vector Classifier 92.50%, K-Nearest Neighbors 89.77%, Decision Tree 95.90%. The pH value, along with the values for N, P, and K, are provided as inputs. The humidity and temperature values are then retrieved from the cloud for the soil specified. To determine the most suitable crop for cultivation, all these values are processed using the Random Forest machine learning model. Based on the provided values, the recommended crop for cultivation is mango.

4.3 User-Side Interface

The user-side interface of the application provides farmers with intuitive tools for crop prediction. When the farmer inputs environmental parameters such as soil pH, nitrogen (N), phosphorus (P), potassium (K), humidity, and temperature, the application processes these values using the trained machine learning model to predict the best crop to cultivate (Fig. 6).

The final product (Fig. 7) integrates the IoT sensors and components into a compact hardware setup. This device collects real-time environmental data, sends it to the cloud, and interfaces with the application to provide actionable insights for farmers.

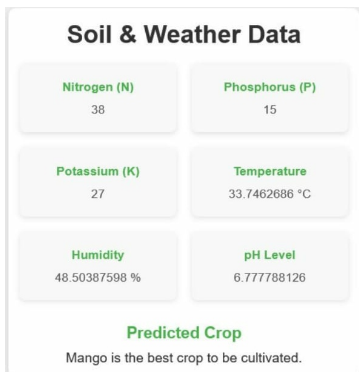

Fig. 6. Displayed Best Crop to Be Cultivated

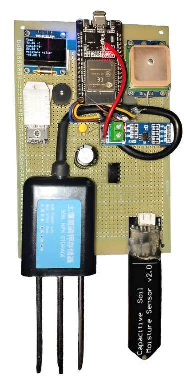

Fig. 7. The Final Product

Conclusion

The paper showcases the use of IoT-based sensors and machine learning models in autonomous farming. The sensors, including humidity, temperature, soil pH, and NPK, allow real-time monitoring of key agricultural parameters. The models, including Random Forest Classifier and Decision Tree Classifier, provide accurate predictions for crop choices. A web/mobile application offers an intuitive interface for accessing real-time sensor data and receiving recommendations. The system could be expanded with advanced sensors and machine learning techniques.

References

1. Sundaresan, S., S. Daniel Johnson, V. Mani Bharathy, P. Mohan Pavan Kumar, and M. Surendar. "Machine learning and IoT-based smart farming for enhancing the crop yield." In Journal of Physics: Conference Series, vol. 2466, no. 1, p. 012028. IOP Publishing, 2023.
2. Reddy, Kasara Sai Pratyush, Y. Mohana Roopa, Kovvada Rajeev LN, and Narra Sai Nandan. "IoT based smart agriculture using machine learning." In 2020 Second

international conference on inventive research in computing applications (ICIRCA), pp. 130-134. IEEE, 2020.
3. Chandgude, Anuja, Nikita Harpale, Diksha Jadhav, Punam Pawar, and Suhas M. Patil. "A review on machine learning algorithm used for crop monitoring system in agriculture." International Research Journal of Engineering and Technology (IRJET) 5, no. 04 (2018): 1470.
4. Patil, Rekha J., Indira Mulage, and Nishant Patil. "Smart agriculture using iot and machine learning." Journal of scientific research and technology (2023): 47-59.
5. Anguraj, K., B. Thiyaneswaran, G. Megashree, JG Preetha Shri, S. Navya, and J. Jayanthi. "Crop recommendation on analyzing soil using machine learning." Turkish Journal of Computer and Mathematics Education 12, no. 6 (2021): 1784-1791.
6. Archana, G., N. Dharmil, N. Pratiksha, and S. Atharva. "Smart Crop Prediction using IoT and Machine Learning." International Journal of Engineering Research and Technology, NTASU 2020 (2021).
7. Varghese, Reuben, and Smarita Sharma. "Affordable smart farming using IoT and machine learning." In 2018 second international conference on intelligent computing and control systems (ICICCS), pp. 645-650. IEEE, 2018.
8. Kumar, T. Raghav, Bhagavatula Aiswarya, Aashish Suresh, Drishti Jain, Natesh Balaji, and Varshini Sankaran. "Smart management of crop cultivation using IOT and machine learning." International Research Journal of Engineering and Technology (IRJET) 5, no. 11 (2018).
9. Bhanu, K. N., H. J. Jasmine, and H. S. Mahadevaswamy. "Machine learning implementation in IoT based intelligent system for agriculture." In 2020 International Conference for Emerging Technology (INCET), pp. 1-5. IEEE, 2020.
10. Sharma, Abhinav, Arpit Jain, Prateek Gupta, and Vinay Chowdary. "Machine learning applications for precision agriculture: A comprehensive review." IEEE Access 9 (2020): 4843-4873.
11. Isa, Indra Griha Tofik, Adhan Effendi, and Suhartono Suhartono. "Deep reinforcement learning in agricultural IoT-based: A review." In E3S Web of Conferences, vol. 479, p. 07004. EDP Sciences, 2024.
12. Akash Raj, N., Balaji Srinivasan, D. Deepit Abhishek, J. Sarath Jeyavanth, and A. Vinith Kannan. "Iot based agro automation system using machine learning algorithms." International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol 5 (2016).
13. Dahane, Amine, Rabaie Benameur, B. Kechar, and A. Benyamina. "An IoT based smart farming system using machine learning." In 2020 International symposium on networks, computers and communications (ISNCC), pp. 1-6. IEEE, 2020.
14. Chandgude, Anuja, Nikita Harpale, Diksha Jadhav, Punam Pawar, and Suhas M. Patil. "A review on machine learning algorithm used for crop monitoring system in agriculture." International Research Journal of Engineering and Technology (IRJET) 5, no. 04 (2018): 1470.
15. Baranwal, Tanmay, and Pushpendra Kumar Pateriya. "Development of IoT based smart security and monitoring devices for agriculture." In 2016 6th international conference-cloud system and big data engineering (confluence), pp. 597-602. IEEE, 2016.
16. Atharva Ingle. Crop Recommendation Dataset. Available online: "`https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset?select=Crop_recommendation.csv`"