PERFORMANCE ANALYSIS OF MACHINE LEARNING MODELS FOR IOT-BASED SMART AGRICULTURE IN TAMIL NADU USING MULTI CLASS METRIC

T. Parikodi
Assistant Professor, Department of Computer Science
Agurchand Manmull Jain College, Meenambakkam, Chennai

M. Jasmine
Assistant Professor, Department of Computer Science with Artificial Intelligence
Anna Adarsh College for Women, Chennai

S. Srimathi
Assistant Professor, Department of Computer Applications
St. Joseph's College (Arts & Science), Kovur, Chennai

G. Manimannan
Associate Professor, Department of Computer Applications
St. Joseph's College (Arts & Science), Kovur, Chennai
https://doi.org/10.34293/9789361639715.shanlax.ch.025

Abstract

This research introduces an integrated approach for utilizing Internet of Things (IoT) sensor data to enhance precision irrigation practices in Tamil Nadu. A real-time dataset was developed containing soil moisture, air and soil temperature, humidity, pH level, water flow, pump status, and irrigation activity readings. Following comprehensive preprocessing covering missing data treatment, feature standardization, and categorical variable encoding the dataset was initially examined using k-Means clustering ($k = 3$*) to uncover natural groupings of field conditions. Insights from this step informed the development of two supervised classification models: Random Forest and Support Vector Machine (SVM), trained and tested through stratified sampling. The Random Forest classifier reached an accuracy of 97.80%, while the SVM model achieved 96.17%, with delivering high precision, recall, and F1-scores across categories. Analysis of feature importance highlighted soil moisture, air temperature, and irrigation status as the key determinants for irrigation planning. Confusion matrix and ROC curve visualizations verified the stability and predictive reliability of both models. The proposed framework presents a scalable, data-driven tool for automated irrigation management, enabling efficient water use and increased crop yields in the region's diverse climatic settings. Recommended actions include broadening IoT sensor coverage across varied agro-ecological zones and linking these intelligent irrigation systems with government agricultural initiatives to benefit smallholder farmers.*
Keywords*: IoT Sensors, Precision Irrigation, Random Forest, Support Vector Machine, Feature Importance, Tamil Nadu Agriculture*

Introduction

The emergence of Internet of Things (IoT) technologies has introduced numerous transformations to numerous industries, such as the agricultural sector, as they could now monitor and make decisions on the basis of the data. IoT is now being used in

Tamil Nadu where agriculture is a major means of livelihood to enhance efficiency and sustainability. Connecting sensors and devices between farms, farmers are now able to monitor the key parameters of soil moisture, temperature, humidity, and nutrient levels in real time. Such an endless flow of information makes it possible to implement precision farming techniques that are economical in resource utilization, minimize wastage, and maximize yields.

IoT technology is applied to overcome the usual obstacles to Tamil Nadu agriculture, such as water shortage, sporadic weather conditions, and pest invasion. Sensor-based analytics enable farmers to use precision in their irrigation, implement fertilizer effectively, and monitor pests at their early stages. In addition to the farm, IoT-based systems assist in supply chain management and crop health monitoring, which enhance productivity and profitability. By doing so, IoT integration facilitates a transition to smarter, more resilient and sustainable agricultural practices, In the area, which will help rural economies grow.

Review of Literature

The use of IoT in the agricultural sector is also well known in terms of making the old systems more current and efficient with the use of up-to-date information about the environment and assisting in making smart decisions. Kamilaris et al. (2017) emphasized that IoT-based big data analytics are capable of processing various agricultural and environmental data to enhance resource use and product. Li et al. (2019) suggested a smart IoT system that has been used to monitor the agriculture environment, demonstrating that ongoing monitoring of both soil and weather statuses improve the timeliness and accuracy of farm activities. The article by Ahmad et al. (2020) proposes an IoT-oriented irrigation approach that aims at preserving the water quantity and keeping the crops healthy at the same time, highlighting the sustainability of such technologies. Correspondingly, Roy et al. (2021) introduced an IoT-enabled pest-detecting and control system that minimised losses and enhanced the quality of crops via automatic signals and treatments.

It has also been studied how IoT can be used together with cloud computing and machine learning. Zhang et al. (2018) have outlined a cloud-based irrigation system that was dynamic in its regulation of water supply based on real-time information, which saved on water and enhanced growth of plants. The article by Kumar et al. (2020) used the IoT data in conjunction with machine learning to create precise yield prediction models. Verma et al. (2022) and Patel and Shah (2021) investigated scalable IoT-cloud architectures, which will provide a centralized monitoring and analysis of precision agriculture. Singh et al. (2019) and Zhao and Wang (2020) also expanded this conception with the help of mobile cloud computing, providing farmers with the opportunity to access data in real-time and have control over what is happening in the field remotely.

There is study of socioeconomic aspects as well. Banerjee et al. (2018) showed that with the help of IoT, the efficiency of fertilisers applied to rice plantations was improved. Chen et al. (2021) designed greenhouse monitoring systems that provided good growth conditions, which improved the yield and quality. Das et al. (2020) and Mishra et al. (2022) affirmed the economic advantages of the Indian agricultural sector of IoT, stating that the investment is high; however, the long-term returns exceed the investment. Gupta and Kumar (2023) also noted that smart farming with the help of IoT improves profitability and contributes to sustainable development of agriculture.

Database

The information deployed in this analysis was gathered through the IoT sensor networks deployed on farms in Tamil Nadu. Such sensors constantly measure the most important environmental and soil variables that influence crop growth and yield. The parameters that are monitored are soil moisture, soil temperature, ambient temperature, humidity, pH, fertilizer application and the irrigation activity. These are intensive measurements that give information on interaction of soil health, climate and crop performance. Multi-cycle growing cycles were used to collect data and these include variation on a seasonal basis, as well as the long-term trends. Readings of sensors are performed with some delay which gives granular information that could be used in machine learning to optimize irrigation, fertilizer application, and early detection of pests. The data set was collected based on ongoing precision agriculture initiatives and publicly accessible sources of smart farming in Tamil Nadu, such as contributions of Tamil Nadu Agricultural University (TNAU), the Indian Council of Agricultural Research (ICAR) and joint efforts of government bodies, research organizations and IoT technology suppliers. Application of region specific and dependable data makes sure the analysis is correctly representative of the local agricultural situation.

Methodology

The current study will use a data-driven approach to develop smart agriculture in Tamil Nadu by exploiting sensor data acquired via IoT-supported devices installed across the areas of farmlands. These sensors will constantly record important environmental and crop related parameters, such as soil moisture, air temperature, soil temperature, humidity, pH level, status of pump operation, irrigation conditions, etc. The data gathered is the foundation of intelligent models to develop on automating the irrigation processes and helping in informed agricultural decision making. The essence and intent of the task is to transform the raw sensor data into meaningful actionable information using the supervised machine learning models namely the Random Forest and Support Vector Machine (SVM) algorithm.

In this study, the entire IoT-based agricultural workflow is presented in figure 1. It shows how sensor devices receive real-time data and send it to a central processing system and pass it through machine learning pipelines to be analyzed further. Such essential elements as irrigation pumps, water flow sensors, and soil monitoring stations are also part of the architecture, which together allow automated control over irrigation programs.

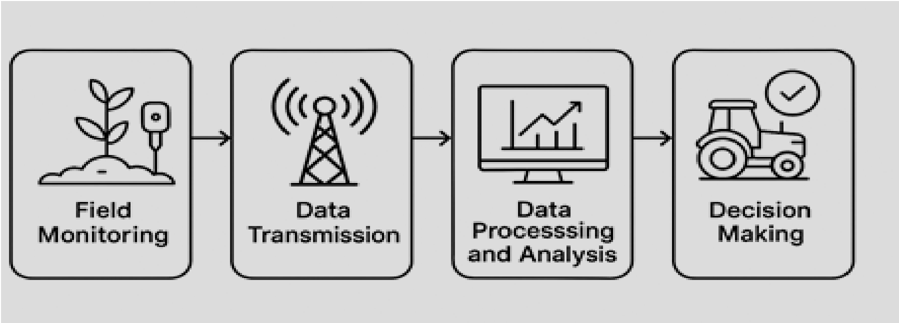


Figure 1. Workflow Diagram for IoT-Based Smart Agriculture in Tamil Nadu

The first step of the process is preprocessing of data, which is a crucial step to assure the consistency, scalability, and reliability of input features. Raw sensor data has many holes and irregularities that are overcome with proper imputation techniques. Numerical data are then normalized to make the data distributions normal as it is a necessary condition in algorithms such as SVM which are sensitive to the magnitude of the input data. Categorical variables (i.e. pump and irrigation statuses) are converted to number codes. The dataset is cleaned and stratified sampling is used to split them into the training and testing subsets to ensure the balance of classes under varying irrigation conditions.

After preprocessing, there are two classification models constructed and tested. The random forest classifier takes advantage of ensemble learning by training a set of decision trees on bootstrapped samples and combines their predictions to improve stability of prediction and minimizes overfitting. Simultaneously, an SVM model using a radial basis function (RBF) kernel is learned to detect non-linear class boundary, which maximizes the margin dividing various irrigation statuses in the feature space.

According to Figure 2, the modeling process consists of specific stages, namely the preparation of data, training a model, its validation, and prediction. Accuracy, precision, recall and F1-score are some of the metrics used to evaluate the performance of both models and they provide a comprehensive evaluation of model effectiveness. Besides, the Random Forest model gives scores of feature importance, which helps to identify such significant factors as soil moisture and pump status that influence water management efficiency significantly. Clustering plots and classification accuracy graphs are some of the visual tools used to authenticate the practicality of the models in the agricultural contexts of the real world.

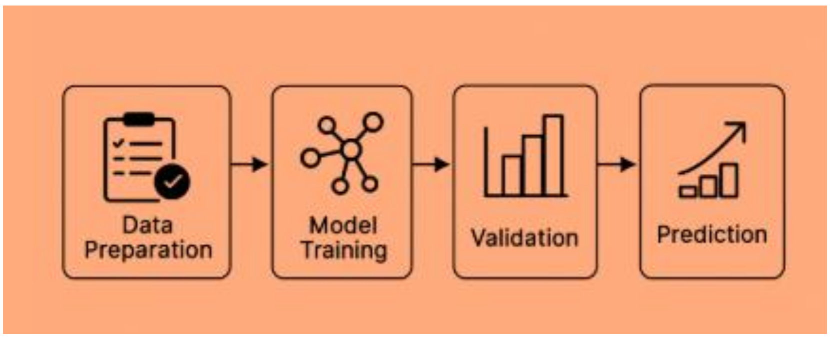


Figure 2 Workflow of SVM and Random Forest Models for IoT-Based Agriculture

Proposed Algorithm

The proposed methodology is as follows:

Step 1: Data Preprocessing

Load sensor dataset: The dataset will be loaded with sensor data collected in the smart irrigation systems spread out through Tamil Nadu.

- Complete missing data with appropriate methods of imputation.
- Convert categorical variables into numerical format through encoding
- Another: scale numerical characteristics to allow even scaling.

Step 2: Supervised Classification

- Train a Random Forest classifier using the labeled dataset.
- Develop an SVM model with an RBF kernel.
- Test the two models on test dataset with relevant performance measures.

Step 3: Model Evaluation and Visualization

- Calculate evaluation metrics: Accuracy, Precision, Recall, and F1-score.
- Generate confusion matrices and ROC curves for each model.

Example: Extract and visualize feature importance of the Random Forest pipeline.

- Compare the results of models with bar graphs and heatmaps to make performance analysis convenient.

Results and Discussion

This paper examines how IoT-based sensor data can be used to monitor and make decisions in agriculture in Tamil Nadu. The data set includes continuous real time measurements such as soil moisture, ambient and soil temperature, humidity, pH levels, water flow rates, irrigation status and pump activity. An integrated unsupervised and supervised machine learning strategy was used to identify patterns and categorize the agricultural status with an aim of enhancing efficiency and utility of irrigation and resource utilization.

The first stage implemented k-Means clustering with k = 3 which are three different agricultural conditions identified with the assistance of domain experts. To assure clustering without bias, all sensor characteristics were normalized, eliminating differences due to differences in measurement scales. The resulting clusters (Figure 1) showed a distinct bifurcation meaning that the sensor data being collected by IoT successfully reflects changes in crop water needs, irrigation patterns, and fluctuations in temperature. This attests to the ability of the IoT to capture the spatial and time dynamics of agriculture in farmlands in Tamil Nadu.

Random Forest and SVM Performance

Based on the knowledge of clustering, two supervised classification models were trained on the labeled dataset, namely Random Forest and Support Vector Machine (SVM), to make predictions of agricultural conditions. Their accuracy, precision, recall and F1-score were evaluated to achieve a complete consideration of their appropriateness in real-time implementation.

The overall accuracy of the Random Forest classifier was 97.80 percent (Table 1and Figure 3). It showed reliability in all classes and the precision and recall were over 0.97. The ensemble learning approach of the model that involves many decision trees making predictions enabled it to identify the complex non-linear patterns and tolerate a variety of sensor inputs.

Table 1. Random Forest Classification Report (Accuracy=0.97800)

Class	Precision	Recall	F1-Score	Support
1	0.98	0.99	0.99	180
2	0.98	0.97	0.98	236
3	0.97	0.97	0.97	184

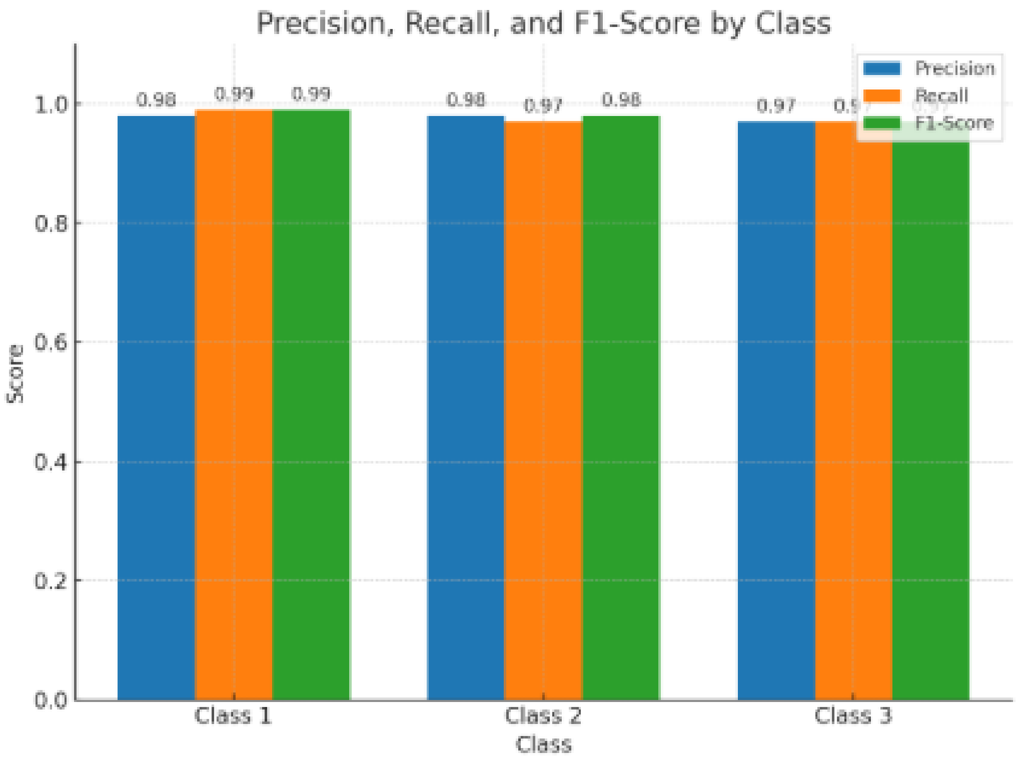


Figure 3. Random Forest Classification and their Measures

Although the SVM classifier is a bit less precise with 96.17, the predictions remain reliable and balanced (Table 2 Figure 4). The sensor data of high dimensions that the model had to deal with due to its use of RBF kernel was effectively handled by defining the best boundaries between classes.

Table 2. SVM Classification Report (Accuracy=0.9617)

Class	Precision	Recall	F1-Score	Support
1	0.96	0.98	0.97	180
2	0.98	0.95	0.96	236
3	0.94	0.96	0.95	184

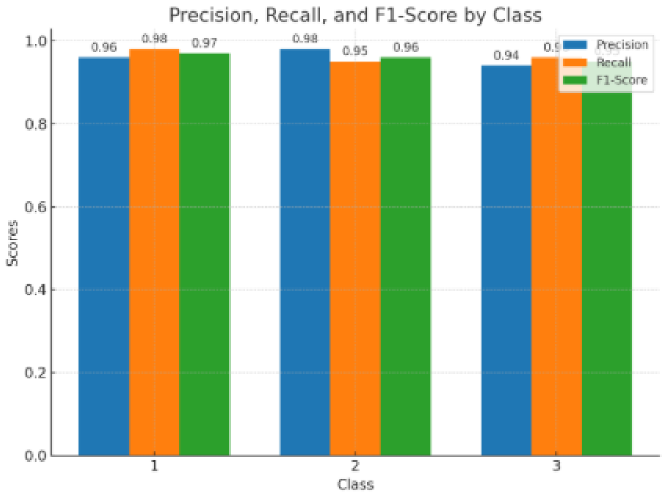


Figure 4. Random SVM and their Measures

Visual Performance Analysis

Both matrices (Figures 5 and 7) are highly dominated at the diagonal, which means that there are few cases of misclassifications in the confusion matrixes of the two models. ROC curves (Figures 6 and 8) of respective classes are close to the top-left corner, which confirms high sensitivity and specificity.

Figure 5. Random Forest Classification Matrix

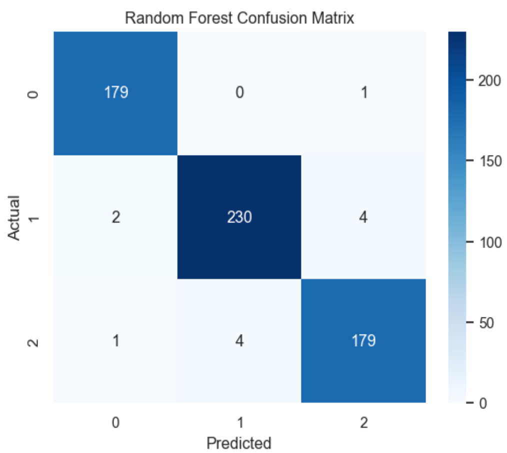


Figure 6. Random Forest ROC Curve

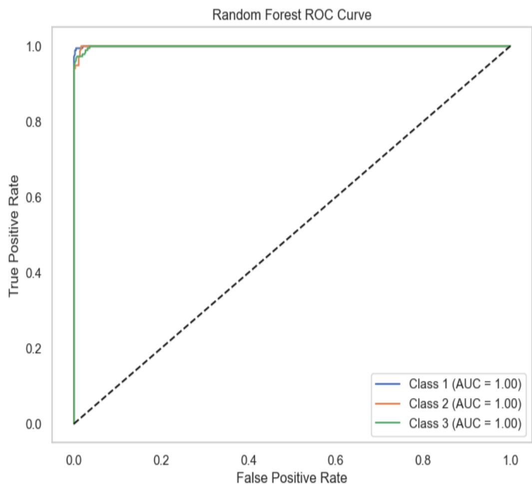


Figure 7. SVM Classification Matrix

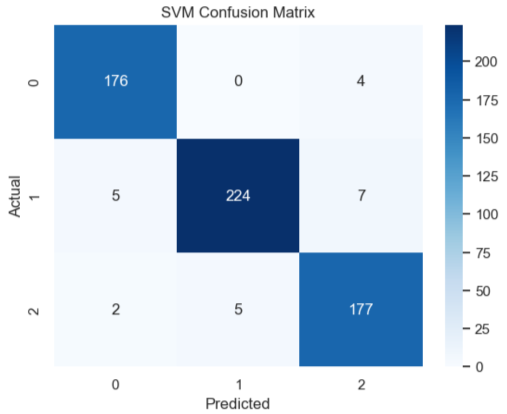
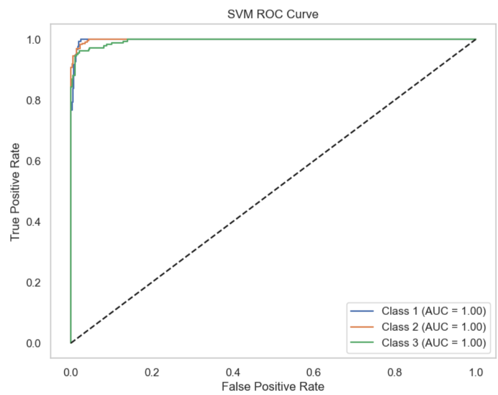


Figure 8. SVM ROC Curve



Based on the feature importance analysis of the Random Forest, it is possible to note that soil moisture, air temperature, and irrigation status have become the most significant predictors that correspond to the agricultural priorities of the region where water management is among the major issues. Such results can inform tactical sensor deployment to achieve the greatest effect.

The richness of IoT sensor data in reflecting variations in fields was supported by clustering of the natural groupings in the agricultural conditions.

Random Forest marginally overperformed SVM on all measures, but overall, both models were of high accuracy and they showed equal classification performance.

Monitoring of soil moisture has been the most decisive factor in accuracy irrigation followed by the temperature of the air and the status of the irrigation.

The integration of exploratory analysis with supervised learning in the classification proves the potential of the IoT in the development of precision farming. These models have the potential to help farmers and policymakers to minimize water wastage and optimize irrigation schedules and enhance productivity. In addition, the information on feature importance may be used to create affordable, sensor-based agricultural systems that are specific to the climate and soil nature of Tamil Nadu.

Conclusion

The paper shows how IoT-based sensor networks can be leveraged with machine learning algorithms to proceed with precision agriculture in Tamil Nadu. Both SVM and Random Forest models showed exceptional performance in real-time estimation of field conditions, the use of continuous and real-time measurements of soil moisture, temperature, pH, irrigation status, etc. to be able to classify them and inform reasonable decisions. The example of k-Means clustering gave the first idea of the natural trends in the data showing differences in soil and crop demands. The fact that the Random Forest 97.80% and the SVM 96.17% have high prediction accuracies confirms that the two methods are best fitted in terms of monitoring and management of agricultural systems on the basis of inputs provided by the IoT.

In addition to classification, this framework improves the potential to monitor key environmental factors which influence crop yield. Detection of influential factors, including soil moisture and temperatures, allow timely interventions, to allow farmers and agricultural planners to effectively address challenges. Tamil Nadu, which faces challenges of scarcity and climate uncertainty around water, presents an opportunity to leverage the IoT with machine learning to provide a scalable, data-intensive route to increasing productivity and ensuring sustainability when it comes to resource utilization in various agricultural regions.

Suggestions

1. Expanding Real-Time Applications: Introduce the use of IoT equipment on a larger scale in the Tamil Nadu farmlands to create a high-quality data set that is constantly updated. It will optimize the models and make them location-specific.
2. Connection with Government Programs: Collaborate with agricultural agencies and efforts such as the Pradhan Mantri Krishi Sinchai Yojana (PMKSY) to ensure affordable implementation of IoT-based smart irrigation and monitoring systems, especially with smallholder and marginal farmers.

References

1. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. Computers and Electronics in Agriculture, 143, 23–37.
2. Li, S., Li, Y., & Li, X. (2019). IoT based intelligent system for agricultural environment monitoring. IEEE Access, 7, 115927–115936.
3. Ahmad, F., Khan, S., & Malik, M. A. (2020). IoT based smart irrigation system for sustainable agriculture. International Journal of Computer Science and Network Security, 20(5), 234–241.
4. Roy, S., Bhattacharjee, S., & Ghosh, S. (2021). IoT-driven pest detection and control system for agriculture. Journal of Agricultural Informatics, 12(2), 35–47.
5. Zhang, Y., Xu, C., & Li, M. (2018). Smart irrigation system based on IoT and cloud computing. Journal of Cleaner Production, 196, 1724–1735.
6. Kumar, R., Sharma, S., & Verma, P. (2020). Yield prediction model using IoT and machine learning techniques. Computers and Electronics in Agriculture, 174, 105452.
7. Patel, R., & Shah, K. (2021). Cloud-based IoT platform for precision agriculture. International Journal of Computer Applications, 173(3), 10–17.
8. Verma, A., Singh, N., & Sharma, D. (2022). IoT and cloud computing in agriculture: A review. Sustainable Computing: Informatics and Systems, 33, 100610.
9. Singh, J., Kaur, R., & Singh, P. (2019). Mobile integrated IoT for agriculture monitoring. Journal of Ambient Intelligence and Humanized Computing, 10(7), 2721–2732.
10. Zhao, L., & Wang, H. (2020). IoT-enabled agriculture monitoring system using mobile cloud computing. IEEE Access, 8, 120667–120675.

11. Banerjee, S., Dasgupta, D., & Mukherjee, A. (2018). Precision nutrient management using IoT in rice cultivation. Journal of Agricultural Science, 10(3), 45–56.
12. Chen, Y., Yang, F., & Zhang, L. (2021). IoT-based greenhouse monitoring system. Computers and Electronics in Agriculture, 181, 105972.
13. Das, S., Ghosh, S., & Sen, S. (2020). Socioeconomic impact of IoT adoption in agriculture. Agricultural Systems, 178, 102729.
14. Mishra, P., Singh, V., & Kumar, A. (2022). Cost-benefit analysis of IoT adoption in Indian agriculture. Journal of Rural Development, 41(2), 153–170.
15. Gupta, R., & Kumar, S. (2023). Economic benefits of IoT-driven smart farming in India. Journal of Agricultural Economics, 74(1), 88–105.