IOT BASED SMART IRRIGATION CLASSIFICATION OF AGRICULTURAL CONDITIONS IN TAMIL NADU USING RANDOM FOREST AND SVM

K. Jeya Gowri
Head & Assistant Professor, Bachelor of Commerce-Computer Applications (Shift II), Shri Shankarlal Sundarbai Shasun Jain College for Women, T. Nagar, Chennai

R. Samuel Kirubakaran
Assistant Professor, Department of Computer Science
Vidhya Sagar Women's College, Chengalpattu
https://doi.org/10.34293/9789361639715.shanlax.ch.006

Abstract

This chapter introduces a data-driven approach to precision irrigation by leveraging Internet of Things (IoT) sensor data tailored to agricultural practices in Tamil Nadu. A curated dataset comprising real-time parameters such as soil moisture, air temperature, humidity, pH, water flow rate, pump status, and irrigation activity was systematically preprocessed using techniques like handling missing values, feature scaling, and encoding of categorical attributes. Unsupervised clustering via k-Means (k = 3) was initially applied to identify natural groupings in field conditions. Subsequently, two supervised learning models, Random Forest and Support Vector Machine (SVM) were developed and validated using stratified sampling to classify field conditions. With 96.17% accuracy, the SVM model demonstrated exceptional precision, recall, and F1-scores across all categories, while the Random Forest classifier achieved a high accuracy of 97.80%. Soil moisture, temperature, and irrigation status were shown by feature importance analysis to be the most important indicators, demonstrating their critical role in intelligent irrigation management. The reliability of the models was supported by visual evaluations utilizing confusion matrices and ROC curves. The framework presented in this research provides a scalable approach to automate irrigation choices, increasing water efficiency and crop productivity in Tamil Nadu's varied climatic circumstances. The chapter concludes with two useful recommendations: (1) extending IoT sensor networks throughout different agro-climatic zones, and (2) incorporating these predictive systems into government agricultural support programs to benefit small and marginal farmers.
Keywords: IoT in Agriculture, Smart Irrigation, Random Forest, Support Vector Machine, Feature Analysis, Tamil Nadu Farming Systems

Introduction

The rapid development of Internet of Things (IoT) technology has revolutionized a few industries, including agriculture, through the opportunity to track in real-time, and make decisions based on data. IoT has been shown to be a radically disruptive tool in Tamil Nadu, which is an agricultural-oriented state, as it enhances the productivity and sustainability of the agricultural practice. Connected sensors and equipment can erase the necessity to monitor significant variables such as soil moisture, temperature, humidity, and nutrient content in real-time using the farmers. This constant flow of information can allow farmers to use precision agriculture techniques, which will maximize the use of resources, reduce wastage, and eventually raise the yields.

The agricultural context of the state of Tamil Nadu has many issues such as water shortage, unpredictable weather patterns, and pest attacks, among others, that are being resolved with the implementation of IoT into the agricultural system. The data generated by sensors can be used by farmers to provide timely intervention (precise irrigation, specific fertilization and early control of pests). Additionally, IoT-driven systems also make it possible to monitor crop health and supply chain more efficiently, increasing productivity and profitability. Therefore, as per the general goals of smart farming and rural development, there is integration of IoT technology fostering stronger and sustainable agricultural industry in Tamil Nadu.

Review of Literature

The possibilities of IoT technologies have been widely acknowledged due to their ability to transform conventional farming through achieving real-time data collection and smart decision-making. Kaminaris et al. (2017) introduced a high-quality review that highlighted the growing importance of big data analytics in agriculture, which is enabled by IoT devices that gather various types of data associated with the environment and crops. Their study focused on how big data methods can enhance utilization of resources and increase output. On the same note, Li et al. (2019) created a smart IoT system aimed at monitoring the environment in the farm and showed that ongoing sensor information on soil and weather conditions could enhance the accuracy and time of implementing farming interventions. Ahmad et al. (2020) helped by developing a smart irrigation system that uses IoT devices to save water without impacting the health of crops, which is one of the general advantages of IoT-based agriculture. Also, Roy et al. (2021) discussed a pest detection and control mechanism based on IoT, demonstrating that timely notifications and automated feedback would minimize crop loss and increase the quality of yields.

A number of studies have integrated IoT with cloud computing and machine learning to improve agricultural management in more than just environmental monitoring. Zhang et al. (2018) developed a smart irrigation system, which is a cloud system and depends on real-time information to increase and decrease water application dynamically, which leads to significant water savings and improvement of crop growth. The study by Kumar et al. (2020) introduced a yield prediction model based on IoT with a combination of machine learning algorithms, which has accurate predictions to facilitate proactive farm management. Patel and Shah (2021) and Verma et al. (2022) have explored cloud platforms, examined scalable IoT-cloud architectures that enable remote sensing and centralized data analytics to enhance the quality of agricultural practices. Singh et al. (2019) and Zhao and Wang (2020) also expanded on it, introducing mobile cloud computing and enabling farmers to get real-time alerts and control their farms remotely based on the inputs of IoT sensors, thereby, making it more available and responsive.

Socioeconomic impact of IoT implementation in the agricultural sector has been also well researched. Banerjee et al. (2018) examined IoT in precision nutrient management in rice production and found that it yields better efficiency of utilizing fertilizers and reduces their cost. The greenhouse monitoring system, designed by Chen et al. (2021) as an IoT-based system, can sustain the necessary environment, which is likely to increase yields and improve crop quality. In addition, Das et al. (2020) evaluated the wider socio-economic value of IoT in the agricultural sector and discovered that it does not only improve agricultural output but also has positive impacts on the incomes and livelihoods of farmers. Mishra et al. (2022) performed a cost-benefit analysis of implementing IoT in Indian agriculture, which proves to be correct by demonstrating that the first investment is outweighed by long-term economic benefits. These results were supported by Gupta and Kumar (2023), who showed that smart farming with IoT will lead to a significant improvement in returns to the economy and sustainable development of agriculture in India.

Database and Parameters

The analysis uses an agricultural database containing real-time sensor measurement data on multiple farms that are connected to the IoT in Tamil Nadu. These sensors always keep a close check on some important environmental and soil aspects that directly influence the development and production of crops. Significant characteristics which are reflected in the dataset include soil moisture, soil temperature, ambient air temperature, humidity levels, soil pH, fertilizer use and irrigation status. Nevertheless, this large set of measures allows a full and changing view of the agronomic context in the database, which is why it is possible to discuss in detail the influence of climatic and soil indicators on the health of crops.

A considerable number of observations in a number of growing seasons constitute the dataset, thus ensuring the strength and variability of the data. Each of the records is a sensor data at regular intervals, which reflects the variability of the field conditions over space and time. Since this data is granular, it can be easily tracked and informed decisions can be made hence it can be used in machine learning applications that may aim at the optimization of agricultural management practice. Moreover, combining sensor data and domain knowledge can provide valuable information regarding optimization of irrigation schedules, nutrient management and pest management practices that subsequently leads to increased crop production as well as resource efficiency.

The IoT agricultural dataset data were retrieved through open-access repositories in Tamil Nadu dedicated to smart agriculture and through current precision farming projects. The collection of these data is frequently by government agricultural agencies, research agencies, and university-technology company collaborations on IoT-based solutions to agriculture. Open-access sites can provide important datasets such as the environmental sensor data of a number of Indian states, along with the Tamil Nadu Agricultural University (TNAU) and the Indian Council of Agricultural Research (ICAR). These real and area-specific datasets ensure the reliability and relevance of the analysis to the conditions of the Tamil Nadu agrarian environment.

Methodology

This study employs data-driven approach to enhance smart farming methods in Tamil Nadu by utilizing IoT-enabled devices that are installed all over the farmlands to provide sensor data. The sensors monitor a number of environmental and crop related parameters such as soil moisture, air and soil temperature, humidity, pH, pump status and irrigation status. The readings used in developing smart models can be used to automate irrigation and in making agricultural decisions. The main objective of this method is to process unedited sensor data to valuable intelligence with supervised machine learning algorithms such as Random Forest and Support Vector Machine (SVM).

The illustration in figure 1 depicts the whole IoT-driven agricultural workflow that was utilized in the study. It is a depiction of the way sensor gadgets collect data on the fly, transmit to a central processing station, and feed into machine learning pipelines to analyze. The irrigation status is automatically controlled by the use of important parts of the system including irrigation pumps, water flow sensors and soil monitoring stations.

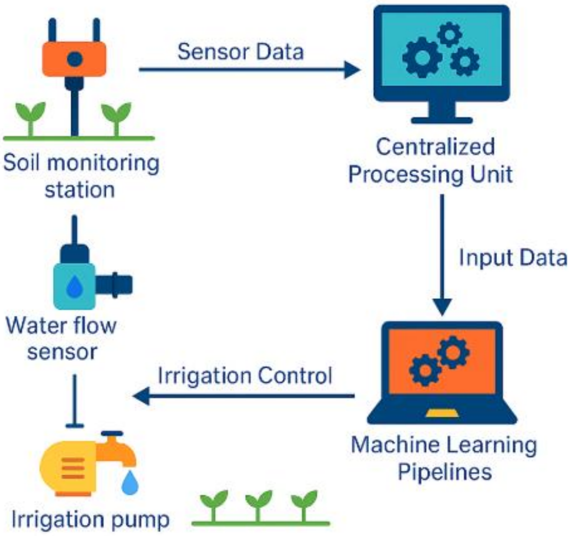


Figure 1. Workflow Diagram for IoT-Based Smart Agriculture in Tamil Nadu

The first, crucial aspect in ensuring that the input features are consistent, scalable, and free of noise and anomalies is data preprocessing. Raw sensor data can often contain noisy or missing data points, which are corrected with imputation. The standardization is then used to standardise the features to normalize the distribution of the data. Since feature magnitudes can be prone in other algorithms such as the SVM, this is a step required so as to guarantee optimal performance of the algorithm. Categorical variables such as pump and irrigation status are represented by means of numerical coding. Data is then broken into the training and testing sets after cleaning based on a stratified sampling to maintain a balance between the different irrigation status categories.

The post processing includes two classification models constructed and evaluated. The Random Forest classifier is famous in its capabilities in the ensemble learning that is trained to predict irrigation situations. It is used by combining the results of multiple decision trees which have been developed on bootstrapped data which decreases overfitting and makes the model more robust. Furthermore, an RBF kernel is used to construct an SVM model with representation of nonlinear boundaries between feature categories. This model produces hyperplanes maximizing the distinction between different irrigation status elements in the multi-dimensional feature space.

As observed in Figure 2, the process of these models consists of various stages like data preparation, training, validation and prediction. The evaluation of the models is done based on such measures as accuracy, precision, recall, and F1-score so as to present a deep analysis of their performance. Moreover, the Random Forest model generates a feature importance map, which shows the most important parameters, including soil moisture and the state of a pump, which becomes the key element of effective water management. The visual interpretation of cloning patterns and classification accuracy proves the possibility of the system application in real agriculture.

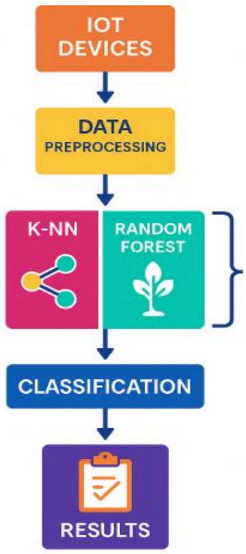


Figure 2. Workflow of SVM and Random Forest Models for IoT-Based Agriculture

Proposed Algorithm

Step 1: Data Preprocessing

The suggested methodology in terms of the division of agricultural situations with the help of smart irrigation data is a properly structured set of steps. The pre-processing phase of data begins with the importation of a data set of IoT sensors obtained at multiple sites in Tamil Nadu. Missing values are addressed using the right imputation methods to ensure that the data remains intact and complete. Encoding techniques are then used to transform the categorical features into numerical formats and the variables in the numerical form are standardized so that the features have uniform scales.

Step 2: Supervised Classification

Training two categorization models, which occurs during the supervised learning stage that follows the preprocessing stage, is the content of this learning stage. The tagged dataset is put through a Random Forest method to identify trends regarding different irrigation conditions. At the same time, a Support Vector Machine (SVM) model is constructed with a radial basis function (RBF) kernel, in order to reflect the complex relationships between the data. A test dataset is then used to compare the predictive abilities of the two models as well as using accepted performance indicators.

Step 3: Model Evaluation and Visualization

The final stage is everything about model assessment and visualization. In both models, the key metrics of the categorization accuracy are precision, recall, and F1-score, according to which the performance of the classification is measured. The predictions are explained by generating confusion matrices and ROC curves. Moreover, the importance of features as given by the Random Forest model is accessed so as to establish the most important variables and this information is presented graphically. Heatmaps and bar charts are also used to highlight the differences in model effectiveness and deliver information regarding the comparative strengths.

Results and Discussion

The study focuses on how IoT enabled devices sensor data can be used in analyzing agriculture in Tamil Nadu. The data storage will contain live data of soil moisture, temperature, humidity, pH, water flow, irrigation state and pump state. In the research, both supervised and unsupervised methods of learning are merged to classify agriculture situations and determine data trends to optimize resource utilization and facilitate smart irrigation.

The first action was to perform k-Means clustering with k = 3, which were the three distinct classes found by the means of agronomic experience. Prior to clustering, sensor data was brought to a zero mean to remove biases of different feature scales. Figure 1 illustrates the clustering output that depicts different groupings, which implies that sensor-driven IoT information captures underlying farm variations, such as field temperature variations, irrigation schedules, and crop moisture needs. This validates the promise of the Internet of Things (IoT) to capture spatial and time dynamics of the agricultural systems in Tamil Nadu.

Random Forest and SVM

In line with the unsupervised learning insights, two common supervised learning algorithms were applied (Random Forest and Support Vector Machine (SVM)) on the labeled dataset to carry out classification tasks. The suitability of the models to be used in real-time agricultural decision-making was also tested based on the multiple performance indicators, including accuracy, precision, recall, and F1-score.

The Random Forest classifier performed remarkably well as evidenced by 97.80 accuracy as shown in Table1. Having a precision and a recall of greater than 0.97, it did exceptionally well across all groups as it was able to distinguish different irrigation and crop conditions. The ensemble approach of Random Forest that involves the combination of the decisions made by multiple trees enables it to deal with the nonlinear relations and a variety of sensor values.

Table 1. Random Forest Classification Report

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

On the other hand, the SVM classifier yielded slightly low accuracy of 96.17 percent with a strong and balanced performance on all classes. Through optimization of decision boundaries between classes, the real-time capabilities of the SVM to work with a high-dimensional dataset, such as IoT sensor outputs, are highly efficient. Table 2 proves that the model is robust enough to process small variations in sensor data about different agricultural conditions as it indicates that precision and recall values remain the same.

Table 2. SVM Classification Report

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 3 to Figure 6 show the classification matrices of the two models and the corresponding ROC curves. These visualizations are useful to evaluate how the models are performing not in numerical terms only, but also in the context of decision making. As an example, strong diagonal dominance in Figure 3 (Random Forest Confusion Matrix) suggests that there are few misclassifications, whereas curves in Figure 4 (Random Forest ROC Curve) tend towards the top-left corner, suggesting high sensitivity and specificity in all classes.

Figure 3. Random Forest Classification Matrix

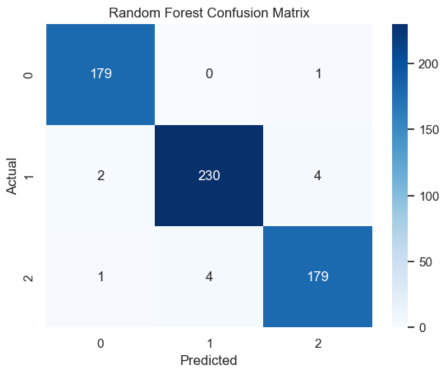
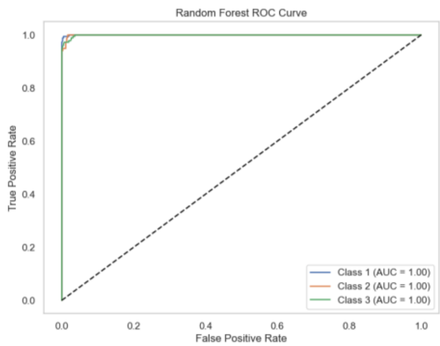


Figure 4. Random Forest ROC Curve



The figure 5 (SVM Confusion Matrix) shows high diagonal dominance indicating a low level of misclassification that does enhance interpretability. Furthermore, in Figure 6 (SVM ROC Curve), the curves are towards the upper-left hand side, revealing high sensitivity and specificity among all the classes. The water scarcity and irrigation planning are key concerns in the Tamil Nadu region, and therefore, it was identified that soil moisture, air temperature, and irrigation status are the most important contributing factors, as it is expected in agronomics. Such understandings are useful in concentrating monitoring efforts on variables with the most effect and sensor optimization.

Figure 5. SVM Classification Matrix

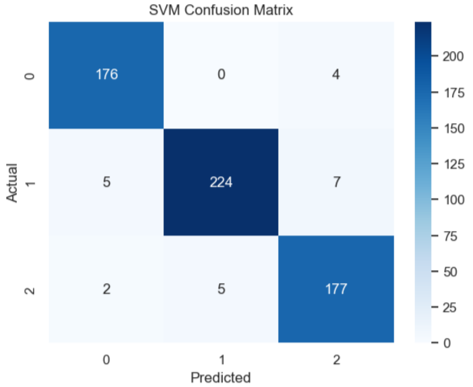
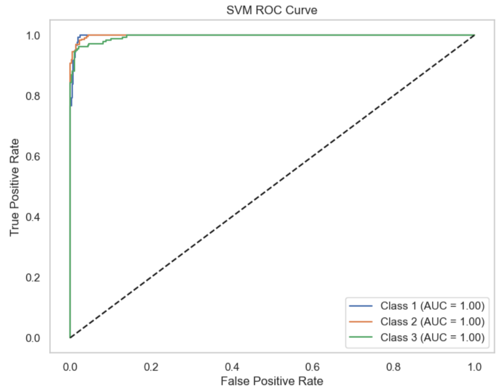


Figure 6. SVM ROC Curve



Clustering and supervised learning can be combined to get a comprehensive image of sensor-based agricultural trends in Tamil Nadu. Though the supervised models gave a good accuracy in predicting irrigation status and the crop conditions, there was a macro view of data groupings as a result of clustering. Both Random Forest and SVM have demonstrated that IoT sensor data when processed and analyzed correctly can give reliable classifications which can be utilized in automated decision-making systems.

The findings indicate that field-level monitoring is one of the ways that can be revolutionized using IoT devices in the Tamil Nadu agriculture. Farmers and legislators can use such models to optimize water use, reduce the amount of waste used and maximize production. Also, sensor-based prioritization is facilitated with the results of feature importance research, which enables low-cost and scalable precision farming solutions depending on particular soil and climatic circumstances.

Conclusion and Suggestions

Such a consideration implies the possibility of integrating IoT-driven sensor developments with machine learning computations to advance precision agribusiness in Tamil Nadu. The classification models using real-time sensor data that captured more basic elements such as soil dampness, temperature, PH values and water system, namely Arbitrary Timberland and Bolster Vector Machine (SVM) were able to predict the field conditions at a highly accurate level. The application of k-means clustering enabled the revelation of the frequently occurring designs within the data and promoted a pattern perception of soil and editing constantly. The active implementation of Irregular Woodland (97.80% accuracy) and that of SVM (96.17% accuracy) indicates their feasibility in terms of their reasonableness in classifying agrarian conditions using sensor-driven data.

This combined analysis strategy does not entrench the ability to filter field circumstances but also leaves significant experiences in the administration of moving forward trim. Being aware of the essential conditions like soil dampness and temperature, partners can develop the convenient and productive options, which will result in the increased efficiency and supportability. In the face of advancing climatic problems and water scarcity in Tamil Nadu, such a show provides a flexible, information-based configuration capable of adapting across varying agro-climatic regions.

Recommendations

1. Extended Implementation of Real-time IoT: It is advisable to increase the installation of IoT sensors within various regions of farming in Tamil Nadu. It will allow gathering data on a continuous basis and enhance the performance and regional flexibility of machine learning models.
2. Policy Integration and Farmer Support: An attempt should be undertaken to coordinate these smart agricultural solutions with government plans such as the Pradhan Mantri Krishi Sinchai Yojana (PMKSY). There is a possibility of such integration resulting in availability of advanced irrigation management tools to small and marginalized farmers to enable the delivery of equitable benefits to small farmers through affordable and technology-based support systems.

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.