International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

A Hybrid Ensemble Learning Approach for Climate-Adaptive Crop Recommendation Using Soil and Environmental Features

Dr. Anand[1], Dr. Sana Zeba[1], Nandini Ghosh[2], Zeelani Basha[2]

[1]Associate Professor, [2]Assistant Professor,[1,2]Department of Computer Science and Engineering

[1,2]Lords Institute of Engineering and Technology (UGC-Autonomous), Himayath sagar, Hyderabad, 500091, Telangana, India.

Abstract

Crop selection is an important agricultural decision that depends on soil fertility, climatic conditions, and farming practices. Selecting an unsuitable crop may result in reduced productivity, inefficient resource utilization, and economic losses for farmers. To address these challenges, this research presents a Machine Learning-based Crop Recommendation System that identifies suitable crops by analyzing soil nutrient characteristics and environmental conditions. The proposed system utilizes agricultural parameters, including Nitrogen, Phosphorus, Potassium, temperature, humidity, rainfall, sunlight hours, wind speed, region, season, irrigation type, fertilizer usage, and previous crop. A preprocessing pipeline is implemented to handle missing values, encode categorical attributes, and balance the dataset using the Synthetic Minority Over-sampling Technique (SMOTE). Four classification approaches, namely Complement Naive Bayes (CNB), Fast Interpretable Greedy Sums (FIGS) Classifier, Ridge Classifier (RC), and the proposed HistoBoost Forest Fusion (HBFF-Net), are developed and comparatively evaluated. The HBFF-Net integrates Histogram Gradient Boosting (HGB) with Random Forest (RF), where the predictions generated by HGB are incorporated as additional features alongside the original input features for final classification using RF. Model performance is evaluated using Accuracy, Precision, Recall, F1-score, Area Under the Curve (AUC), and Confusion Matrix analysis. Experimental results demonstrate that HBFF-Net achieves an accuracy of 96.47%, outperforming the other evaluated classification models. The system also supports crop prediction for unseen agricultural data, model persistence, and performance visualization. The proposed approach provides a data-driven solution for identifying suitable crops based on soil and environmental characteristics, supporting informed agricultural decisions, efficient resource utilization, improved crop productivity, and sustainable farming practices.

Keywords: Crop Recommendation, Machine Learning, Soil Nutrients, Environmental Parameters, SMOTE, Hybrid Ensemble Learning, Histogram Gradient Boosting, Random Forest.

1. Introduction

India is the leading producer of agricultural goods. The agriculture industry employs 58 percent of the total Indian population as well as contributes 17% to the country's GDP. Crops rely on a multitude of factors, including the type of soil, amount of rainfall and sunlight, irrigation, fertilizer use, insect presence, and land preparation. One of the most frequent challenges that Indian farmers must overcome is choosing crops in accordance with the terrain and the climate. Considering the fact that climatic conditions and characteristics of soil have a direct impact on crop yield, it is necessary to develop crop management practices that consider the appropriateness of the site and the soil. Weather and agriculture are closely intertwined; therefore, it is essential to adapt to changing climatic trends in a productive way. Using climate-smart agricultural practices may help increase productivity and produce quality crops. In the agriculture sector, a lot of revolutionary changes have been observed in the present scenario. Different technologies and tools, such as sensory technologies, drones, smart irrigation

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

systems, satellite image analysis, etc., are being used in farming for the enhancement of crop production. Here, the selection of a crop for the land plays an important role in farming. New farmers mainly take opinions from fellow experienced local farmers for crop selection. However, they may be misled due to human errors and their relationships with others. So, the new farmers may obtain lots of confusing information on crop selection, estimating the crop production and their profits. Precision agriculture has recently brought about significant advances in the world of agriculture, with an emphasis on irrigation systems, fertilization, crop monitoring and yield prediction. Choosing the appropriate crop in relation to location-specific soil factors and climatic conditions is also vital for enhancing production. Therefore, farmers must be equipped with instruments that allow them to choose the best crop suited for the region's unique meteorological and soil conditions.

2. Literature Survey

Liakos, K. G. et al. [1] performed an extensive review of crop recommendations and productions based on soil and water management. It has been observed that crop selection is an important factor in increasing crop yield and reducing the risk of crop loss, and the efficient use of resources like water, pesticides, and fertilizers leads to a better agricultural outcome. New farmers mainly take opinions from fellow experienced local farmers for crop selection. Doshi, Z. et al. [2] discussed identifying suitable crops based on the season, geographical location, and soil mineral properties. So, we have analysed area-wise and season-wise crop production datasets so that suitable crops can be listed for a specific season and a particular land.

Vaishnavi, S. et al. [3] analysed crop datasets based on season and productivity. Again, a data collection process was also conducted through interactions with experienced farmers based on their crop cultivation plans. The locally collected data were combined with the cloud dataset in a balanced way. Satish Babu et al. [4] elaborated on the development of the agriculture sector at a large scale in rich countries based on crop and soil parameters. In Kerala, India, to support small and marginal farmers. the soil crop databases collected from cultivation fields and crop calendars were prepared. This was achieved using different electronic devices for the reduction of agricultural expenses.

R. Balamurali et al. [5] collected soil–weather parameters, such as temperature, humidity, Potential of Hydrogen (pH value), and NPK (Nitrogen, Phosphorus, and Potassium) values using real-time observations over time. The authors used a wireless sensor network and remote server for data collection and analysis and concluded that the performance of the Medium Access Control (MAC) and routing algorithm approach is the best. Fonthal, F. et al. [6] suggested a model that contains networking and sensors to collect the environmental conditions of fields cultivated with white cabbage crops. It helped the farmer monitor the environmental needs and reduces the losses in farming. Gábor Gyarmati et al. [7] performed a brief study on precision agriculture, which can solve food problems and reduce labor costs. The authors explained leads taken in precision farming to handle the day-to-day challenges of pollution and climate change.

Palazzi, V. et al. [8] suggested collecting the leaf temperature and water needs using the sensor. It allows for better farming by identifying the season-wise crop suitability and its needs. The authors discussed Radio Frequency Identification (RFID)-based sensors and EM4325 ultra-high-frequency (UHF) chip utilization in the research. Yongsheng Wang, et al. [9] suggested different tools for improving crop production. They applied different tools like Physical Layer Signaling (PLS), Packet Switching (PS), and Connectionless Mode Service (CLS) for crop production. Dholu et al. [10] describe how the Internet of Things (IoT) connects various devices such as mobiles, tablets, and personal computers through machine-to-machine communication. They emphasized that the Internet of Things (IoT) enhanced

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

precision agriculture by optimizing the use of resources like pesticides, light, and water, thereby increasing production efficiency and reducing waste.

O. Palagin et al. [11] explained the flora test to obtain information about the state of plants, and then they discussed the two different types of data acquisition systems, which depend on portable devices that are helpful in precision agriculture for controlling and monitoring the growth of crops, the usage of water to crops, and managing the pesticides to crops so that farmer can obtain more production using precision agriculture. Jitendra Patidar et al. [12] discuss how traditional agriculture methods are currently used in India and suggest that precision agriculture can replace these methods. Their work indicated that farmers can grow various crops more effectively with precision agriculture.

Vandana B, et al. [13] explained that the agriculture field in the Indian agriculture farming industry adapts less innovative technology compared to other industries. In precision agriculture, if they adapt communication and information technologies, which can easily provide less cost-effective methods and can be helpful for smart agriculture, in this work, they prepared a model for agriculture farming that can guide the farmers performing their farming in rural and urban areas so that the production of crop yield increases rapidly, which improves their profits about the farmer's production rate. Xiaoshan Wang et al. [14] discussed the design and understanding of precision agriculture systems depending on 5S technology. Precision agriculture leads to the overall development of agriculture, like the utilization of resources and fertilizer without wastage. In proper times, they had to use them, and requirements that are needed by crops should be provided so that the production rate will increase, which can benefit the farmers compared to traditional farming. So, smart farming can replace traditional farming, which will improve the production rate of crops; ultimately, a farmer will obtain benefits using precision agriculture.

3. Proposed System

The proposed system is a machine learning-based crop recommendation model that predicts the most suitable crop using soil nutrients and environmental parameters such as temperature, humidity, and rainfall. The dataset is pre-processed by handling missing values and encoding categorical features, and class imbalance is addressed using SMOTE. Multiple classification models, including a hybrid ensemble of HGB and RF, are trained and evaluated to achieve high prediction accuracy. The best-performing model is saved and used to provide real-time, data-driven crop recommendations, ensuring improved productivity and sustainable agricultural decision-making as shown in Figure 1.

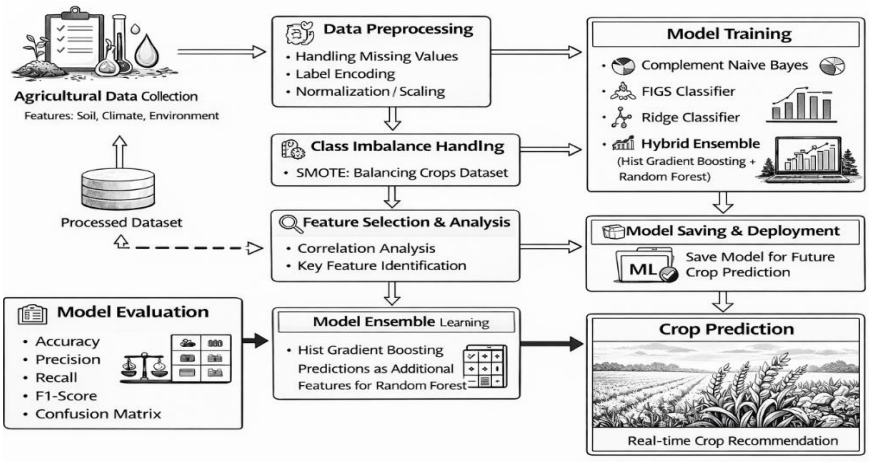

Figure 1: Proposed system architecture for crop selection.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Step 1: Data Collection

The proposed system begins with collecting agricultural data containing soil nutrients (such as Nitrogen, Phosphorus, Potassium), temperature, humidity, rainfall, and other environmental parameters. This dataset forms the foundation for training the machine learning models to understand the relationship between input features and suitable crops.

Step 2: Data Preprocessing

In this step, the collected dataset is cleaned and prepared for analysis. Missing values are handled using statistical techniques such as mean and mode imputation. Categorical features are converted into numerical form using label encoding. This ensures that the dataset becomes suitable for machine learning model training.

Step 3: Feature Selection and Correlation Analysis

The system analyzes relationships between different features using correlation heatmaps and exploratory data analysis. This helps in understanding how soil and climatic parameters influence crop selection and ensures that relevant features are used for prediction.

Step 4: Handling Class Imbalance

To ensure fair prediction across all crop categories, SMOTE (Synthetic Minority Over-sampling Technique) is applied to balance the dataset. This prevents the model from being biased toward majority crop classes and improves overall prediction reliability.

Step 5: Model Training

Multiple classification models such as CNB, FIGS Classifier, RC, and a HBFF-Net are trained using the processed dataset. Each model learns patterns between environmental inputs and recommended crops.

Step 6: Hybrid Ensemble Learning

In the hybrid approach, predictions from the HGB model are used as additional features and combined with the original dataset to train a RF classifier. This enhances prediction performance by combining boosting and bagging techniques for better generalization.

Step 7: Model Evaluation

The trained models are evaluated using performance metrics such as Accuracy, Precision, Recall, F1-score, AUC, and Confusion Matrix analysis. Comparative evaluation helps in identifying the best-performing model for crop recommendation.

Step 8: Model Saving and Deployment

The best-performing model is saved using model persistence techniques (such as joblib) to allow future predictions without retraining. This enables real-time deployment in agricultural decision support systems.

Step 9: Crop Prediction

When new soil and environmental data are provided, the trained model processes the inputs and predicts the most suitable crop category. The output is converted back to readable crop names using label decoding, ensuring user-friendly recommendations.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

4. Result Description

The experimental results demonstrate that the proposed HBFF-Net provides robust and effective crop classification across diverse crop categories. Comparative evaluation with CNB, FIGS Classifier, and Ridge Classifier shows improved predictive performance in terms of Accuracy, Precision, Recall, F1-score, and AUC. The integration of Histogram Gradient Boosting predictions as additional features for Random Forest enhances classification capability and reduces the impact of class imbalance. Confusion Matrix analysis further indicates balanced classification performance across different crop classes.

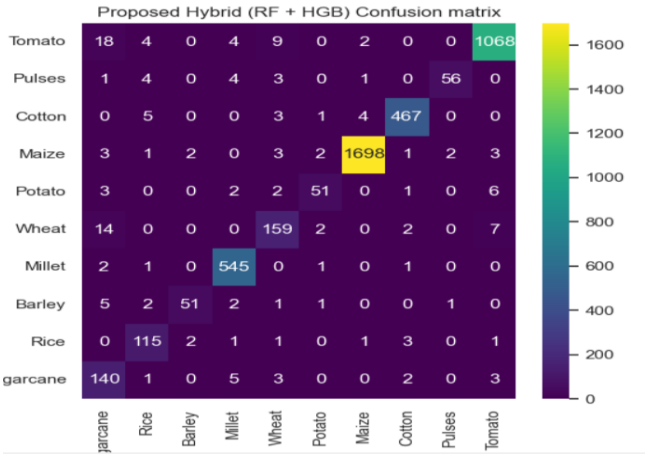

Figure 2: confusion matrix obtained using proposed hybrid model.

The Figure 2 shows confusion matrix for the Proposed Hybrid HBFF-Net Model shows strong classification performance with most values concentrated along the diagonal, indicating correct predictions for nearly all crop categories. Crops such as Maize, Tomato, Millet, Rice, and Cotton are predicted with very high accuracy, and misclassifications are minimal compared to other models. This demonstrates that the hybrid approach effectively captures complex nonlinear relationships in soil and environmental features, resulting in significantly improved and reliable crop recommendation accuracy.

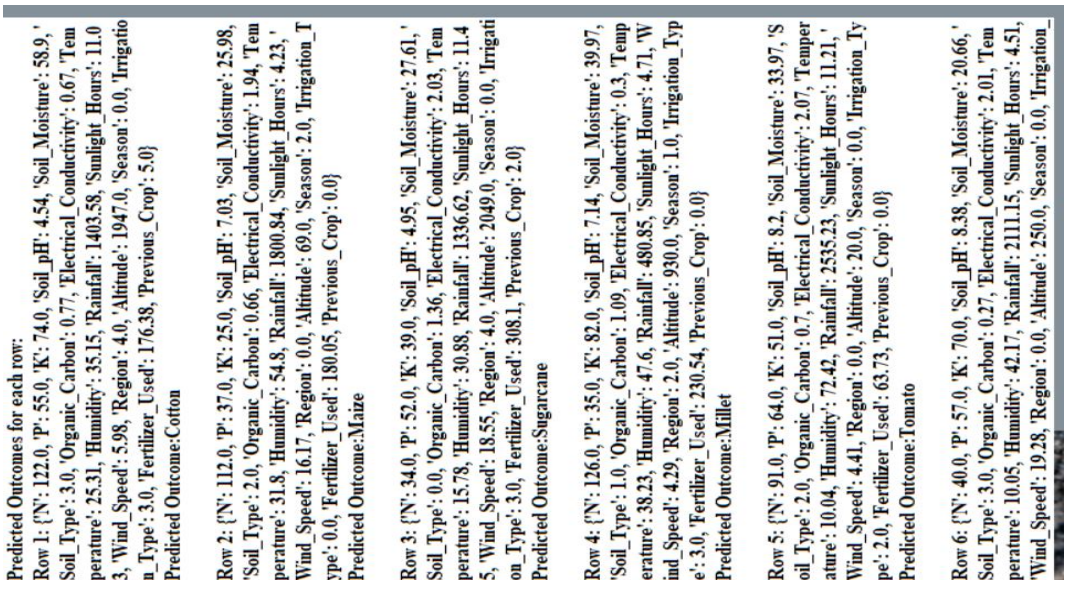

Figure 3: Prediction on Test Data using hybrid model.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The Figure 3 shows the predictions generated by the proposed Hybrid model on unseen test data. For each input record containing environmental conditions, region, season, irrigation type, fertilizer usage, and previous crop, the model outputs the predicted recommended crop. The predictions appear consistent with the input features, indicating that the hybrid model effectively learned relationships between climatic factors and suitable crops. This demonstrates the model's ability to generalize well to new data and provide reliable crop recommendations in practical scenarios.

Table 1 presents the comparative performance of four classification algorithms CNB, FIGS Classifier, RC, and the proposed Hybrid HBFF-Net model evaluated using accuracy, precision, recall, and F-score metrics. The CNB model demonstrates relatively low performance, achieving an accuracy of 41.76% and an F-score of 10.46%, indicating that the probabilistic assumptions and feature independence constraints limit its ability to capture complex relationships within the dataset. Similarly, the FIGS classifier records the weakest overall results, with an accuracy of 2.75% and an F-score of 0.54%, suggesting that although rule-based tree structures attempt to model feature interactions, they fail to generalize effectively under the given data conditions. The RC shows improved performance compared to CNB and FIGS, achieving an accuracy of 67.07%, precision of 25.56%, recall of 26.69%, and an F-score of 23.73%. This improvement indicates that linear regularization helps establish more stable decision boundaries; however, its limited capability to model nonlinear relationships restricts overall predictive performance. In contrast, the proposed Hybrid HBFF-Net model significantly outperforms all existing approaches, achieving an accuracy of 96.47%, precision of 91.61%, recall of 90.27%, and an F-score of 90.73%. The substantial improvement across all evaluation metrics demonstrates that integrating Histogram Gradient Boosting with Random Forest effectively captures complex feature interactions and enhances model generalization. The high recall indicates improved detection capability, while the strong precision confirms reliable predictions, resulting in a balanced and robust classification framework.

Table 1: Performance comparison for the CNB, RC, FIGS and Proposed Hybrid HBFF-Net Model

Algorithms Name	Accuracy	Precision	Recall	F-score
CNB	41.76%	8.04%	17.08%	10.46%
FIGS	2.75%	0.28%	10.00%	0.54%
RC	67.07%	25.56%	26.69%	23.73%
HBFF-Net	96.47%	91.61%	90.27%	90.73%

5. Conclusion

The proposed research successfully demonstrates how data-driven intelligence can improve agricultural decision-making compared to traditional experience-based practices. The system processes environmental and farming parameters such as temperature, humidity, rainfall, sunlight hours, wind speed, region, season, irrigation type, fertilizer usage, and previous crop, and transforms them into a structured format suitable for predictive modelling. Data preprocessing techniques including missing value handling, label encoding, and SMOTE balancing ensure that the dataset becomes reliable and unbiased for training. Multiple classification algorithms such as CNB, FIGS Classifier, RC, and a

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Hybrid HBFF-Net model were implemented and comparatively evaluated using performance metrics such as accuracy, precision, recall, F1-score, AUC, and confusion matrix analysis. Experimental results show that simpler models struggled to capture complex nonlinear relationships between environmental conditions and crop suitability, while the hybrid model effectively combined boosting-based pattern learning with bagging-based stability. This resulted in significantly higher prediction accuracy and more consistent performance across all crop categories. The system also supports prediction on new unseen data, demonstrating strong generalization capability and practical usability. The developed framework provides an intelligent, scalable, and reliable agricultural decision support solution that can help farmers select suitable crops, improve yield, reduce risk of crop failure, and promote sustainable resource utilization. The project highlights the importance of hybrid ensemble learning in real-world agricultural analytics and serves as a foundation for future smart farming applications and precision agriculture systems.

References

[1]. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. Sensors 2018, 18, 2674.

[2]. Doshi, Z.; Nadkarni, S.; Agrawal, R.; Shah, N. AgroConsultant: Intelligent crop recommendation system using machine learning algorithms. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–6.

[3]. Vaishnavi, S.; Shobana, M.; Sabitha, R.; Karthik, S. Agricultural crop recommendations based on productivity and season. In Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 19–20 March 2021; Volume 1, pp. 883–886.

[4]. Babu, S. A software model for precision agriculture for small and marginal farmers. In Proceedings of the 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), Trivandrum, India, 23–24 August 2013; pp. 352–355.

[5]. Balamurali, R.; Kathiravan, K. An analysis of various routing protocols for Precision Agriculture using Wireless Sensor Network. In Proceedings of the 2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR), Chennai, India, 10–12 July 2015; pp. 156–159.

[6]. Fonthal, F. Design and implementation of WSN for precision agriculture in white cabbage crops. In Proceedings of the 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Cusco, Peru, 15–18 August 2017; pp. 1–4.

[7]. Gyarmati, G.; Mizik, T. The present and future of precision agriculture. In Proceedings of the 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE), Budapest, Hungary, 2–4 June 2020; pp. 593–596.

[8]. Palazzi, V.; Gelati, F.; Vaglioni, U.; Alimenti, F.; Mezzanotte, P.; Roselli, L. Leaf-compatible autonomous RFID-based wireless temperature sensors for precision agriculture. In Proceedings of the 2019 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet), Orlando, FL, USA, 20–23 January 2019; pp. 1–4.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[9]. Wang, Y.; Liu, Y. Benefits of Precision Agriculture Application for Winter Wheat in Central China. In Proceedings of the 2018 7th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Hangzhou, China, 6–9 August 2018; pp. 1–4.

[10]. Dholu, M.; Ghodinde, K.A. Internet of things (iot) for precision agriculture application. In Proceedings of the 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11–12 May 2018; pp. 339–342.

[11]. Palagin, O.; Romanov, V.; Galelyuka, I.; Velichko, V.; Hrusha, V. Data acquisition systems of plants' state in precision agriculture. In Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, Prague, Czech Republic, 15–17 September 2011; Volume 1, pp. 16–19.

[12]. Patidar, J.; Khatri, R.; Gurjar, R.C. Precision Agriculture System Using Verilog Hardware Description Language to Design an ASIC. In Proceedings of the 2019 3rd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, India, 29–31 August 2019; pp. 1–6.

[13]. Vandana, B.; Kumar, S.S. A novel approach using big data analytics to improve the crop yield in precision agriculture. In Proceedings of the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 18–19 May 2018; pp. 824–827.

[14]. Wang, X.; Qi, Q. Design and realization of precision agriculture information system based on 5S. In Proceedings of the 2021 19th International Conference on Geoinformatics, Shanghai, China, 24–26 June 2021; pp. 1–4.