WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

IoT-Based Smart Agriculture and Environmental Monitoring System Using Machine Learning for Crop Health Optimization

Mandakini Gupta[1], Akhilesh Mishra[2]

[1]M.Tech. Scholar, [2]Assistant Professor, Department of Electronics & Communication Engineering, Buddha Institute of Technology, Gorakhpur, Uttar Pradesh, India

ABSTRACT :

This research proposes a thesis-based architecture for Internet of Things (IoT)-based smart agriculture and environment monitoring system with machine learning (ML) for predictive watering. The proposed architecture suggests an amalgamation of sensors that include soil-moisture, soil-temperature, air-temperature, relative-humidity, light-intensity, rainfall, and soil pH, with an ESP32 field-node, LoRa/Lora WAN connectivity, a gateway, a cloud/server data processing, and an irrigation pump driven by a relay. The ML challenge is soil-moisture regression, which predicts soil-moisture conditions in the next time step from those in the previous time step as well as the measurement record. Mean Absolute Error, Root Mean Square Error and coefficient of determination (R2) from Linear Regression, Decision Tree Regression, Random Forest Regression, SVR and Boost are reported in the thesis; Random Forest (2.40, 3.20, 0.90) outperforms the other models, especially compared to Boost (2.70, 3.50, 0.86). One-day observation window and 9 observations of paddy rice are also given in the thesis. These results are reproduced as discussed in the thesis; generalization is constrained by the short time-course of the observation period and limited methodological detail. While field validation for long-time duration should be conducted before any quantification of water savings, yield increases, or overall crop health optimization, this system provides a valuable framework for sensor-based monitoring and a predictive irrigation decision aid.

Index Terms-Internet of Things, smart agriculture, environmental monitoring, ESP32, Lora WAN, machine learning, predictive irrigation, soil-moisture prediction. I. INTRODUCTION Dear readers, the automation of rural development by the Internet of Things (IoT) has led to the progress of smart farming.

I. INTRODUCTION

Soil and water availability at the local level is a very important factor affecting agriculture. Variations of temperature, moisture, humidity, rainfall, Solar Exposure at the soil level will not be tracked by fixed irrigation schedule and backup manual monitoring. Under watering will keep the crop water strained, and over watering will be energy consumption as well as water wastage. Introduction of data driven decision from sporadic observation is feasible thanks to IoT sensors that consistently collect field readings and transmit them to a centralized platform for processing [1]-[4]. Machine learning can also treat nonlinear interactions with previous soil moisture. The machine learning model forecasts a value of moisture in a predictive irrigation setup, which is then compared with the crop-specific, soil-specific target window. The last forecast can be used for irrigation recommendations or, after safety checks, to trigger a relay and activate a pump. This paper combines the design and published model comparison results from Mandakini Gupta's M. Tech. Thesis.

A. Objectives and Contributions

The main objectives are to: (i) track soil and other variables; (ii) transmit data through an ESP32 and LoRa/LoRaWAN framework; (iii) test regression algorithms for soil moisture prediction; and (iv) connect the estimated moisture to a relay-based pump and irrigation selection. The use of a comprehensive system architecture for pump control, sensing, low-power transmission, and machine learning-based prediction is the key contribution.

II. RELATED WORK AND RESEARCH GAP

IoT based agriculture combines sensor nodes, wireless transmission, data storage, and analytics to provide more frequent updates regarding the status of a field. Ayaz et al. [1] discussed farm management and connected sensing in the IoT-based farm of the future in their review of IoT based smart agriculture. Elijah et al. [2] addressed areas such as the potential benefits and applications of IoT and data analytics in agriculture and the challenges and future trends of agricultural IoT, whereas Tzounis et al. [4] focused on the latest advancements and emerging challenges in agricultural IoT while Friha et al. [3] considered emerging technologies in smart agriculture. 2.3. Role of sensor data and prediction models The role of sensor data and prediction models to improve irrigation scheduling was studied in the field of intelligent irrigation. Goap et al. [5] proposed an Internet of Things (IoT) based smart irrigation management system using machine learning and open-source platforms. Farooq et al. [6] discussed the opportunities and challenges of the role of IoT in agriculture, while Ray [7] presented the current and future role of IoT-enabled smart agriculture technologies. While these studies support the importance

of sensing, communication, and analysis, field implementation must still consider the irrigation requirements of a crop type, energy use, calibration of sensors, connectivity, and data quality.

A. Machine-Learning Methods

Regression methods are suitable if a continuous soil-moisture value is the goal. Linear regression provides an interpretable baseline, though it might not capture nonlinear interactions. Though it may overfit, a decision tree can mimic nonlinearity. Across the family of decision trees, Random Forest handles nonlinear interactions between sensor features and also combines the results of multiple trees [8]. SVR, with the right kernel, and a margin-based regression target, can outperform other methods [9]. XGBoost sequentially constructs boosted trees to reduce the prediction error [10]. As long as it's big enough and the data is trustworthy, the temporal dependence can be accounted for by an LSTM network.

B. Research Gap

Because of this, the thesis proposes a inexpensive, scalable framework which combines numerous soil/environmental measurements, low-energy long-range communication, ML-based soil-moisture prediction, and an irrigation-control mobile application. A second reason for this is the use of just a solitary predictor, instead of others. This paper discusses the model comparison it makes up the foraging, and discusses that integration at the system-design level.

The evidence has important limitations. The thesis data set overview lists nine observations, a four-hour sampling interval and a single day of sampling. It is impossible to know whether the model performs well across seasons, soils or farms with such a small data set. Moreover, the train/test split, feature pipeline, hyperparameters, random seed and units of the provided error metrics are not clearly described in the reported comparison table. Accordingly, the reported scores should be treated as thesis-reported results rather than metrics obtained from independent reproduction or external validation.

III. SYSTEM OVERVIEW

Agricultural field sensors ESP32 LoRa/Lora WAN gateway cloud/server preprocessing ML prediction irrigation decision relay water pump is the end-to-end data path proposed. As sensing and communication are separated from model training and decision logic, the analytics layer can be replaced independently of the entire field-node architecture.

IV. PROPOSED SYSTEM ARCHITECTURE

Seven operational levels are contained within the architecture: (1) sensor layer, (2) ESP32 processing layer, (3) LoRa/LoRaWAN communication layer, (4) gateway, (5) cloud/server/data-preprocessing layer, (6) ML prediction layer, (7) irrigation-control layer. The field node collects and timestamps the measurements. Possible basic plausibility checks (e.g., range check for water content values) can be performed locally before the data is forwarded via gateway to the server which stores and pre-processes the historical data for later use.

A. Sensor Layer

Parameter	Role in the system	Implementation note
Soil moisture	Primary indicator for irrigation decisions	Place in the crop root zone; calibrate for soil and probe
Soil temperature	Root-zone thermal context	Use a suitable soil-temperature probe
Air temperature	Ambient heat conditions	Shield sensor from direct radiation where required
Relative humidity	Atmospheric moisture context	DHT22 is identified in the thesis
Light intensity	Approximate light exposure	Use a consistent sensor range and unit
Rainfall	Recent precipitation context	Include rainfall events where sensor is available
Soil pH	Soil acidity/alkalinity context	Calibrate with appropriate reference solutions

Sensing metrics proposed in the thesis are soil moisture, soil temperature, ambient temperature, relative humidity, light intensity, rainfall, soil ph. All other metrics, apart from soil moisture, can be used to give an environmental context and aid the monitoring process but soil moisture has a strong influence on the irrigation value that should be used.

B. ESP32 and Communication Layer

Aside from handling the local control signal, the ESP32 interacts with sensors, collects measurements, assembles data packets and performs simple tests. LoRa/LoRaWAN is recommended for low-power, long range communication of field nodes to a gateway. The protocol and radio configuration should satisfy coverage, payload and local spectrum constraints. Observations received at the gateway are transmitted to a server or cloud service for analytics and storage.

C. Cloud/Server and ML Layer

Along with validation of data, handling missing values, evaluating outliers, and creating features, the server maintains a set of timestamped logs. Inputs include current and previous soil-moisture measurements, soil/air temperature, humidity, light levels, time of day, rainfall, and whether a previous irrigation has occurred. The system uses a regression model to generate a predicted future soil-moisture value. A decision module generates a recommendation for whether irrigation is appropriate by comparing this predicted value to a crop-dependent moisture range.

D. Irrigation Control and Safety

Irrigation is invoked if the predicted moisture drops below a preset, crop-specific lower limit and turned off when enough moisture has been achieved. A useful implementation would also consider maximum run time, sensor failure handling, human override, dry-run prevention, and safe state. As the threshold is crop, soil, root-zone depth, and growth-stage dependent, the prediction can only be used as a guideline to irrigate and should not be taken as a standard.

V. METHODOLOGY

Sensor collection, local processing, wireless transmission, data preprocessing, Dataset preparation, Feature selection, Model creation, Model Evaluation, Soil-moisture prediction and Irrigation control are the approach in the thesis.

A. Data Acquisition and Dataset

The crop in the subject thesis is paddy rice, and the experimental field is an agricultural plot. The dataset summary indicates one days collecting duration, 9 observation, 4 hours sample time, 7 recommended sensing parameters, and 8 hours of operation per day. The dataset indicates 24 hours observation duration. These values are kept as is here, as quoted in the source thesis.

Dataset attribute	Thesis-reported value
Crop	Rice (paddy)
Collection period	1 day
Number of observations	9
Sampling interval	4 hours
Proposed sensing parameters	7
Operating time	8 hours/day
Observation window	24 hours across one day

A short example observations table with time, soil moisture, soil temperature, air temperature, humidity, light intensity, rainfall, and irrigation is also presented in the thesis. The displayed rows show soil-moisture values of 33, 24, 36 at three time points. The three rows listed here are not enough to recreate the model comparison provided on their own, because no raw data set is supplied.

B. Preprocessing and Feature Engineering

Preprocessing steps include timestamp alignment, data validation, handling of missing values, review of outliers, noise removal, and formatting to a well-structured dataset. Candidate features include historical soil-moisture measurements, temperature, relative humidity, illumination, time of day, precipitation, and prior irrigation status, among others. All transformations should be fit only on the training data so that future validation is accurate, and the test period should be kept secret during the model selection process.

C. Prediction Task

The goal is to predict the soil moisture h time steps into the future. H is the prediction horizon, and y{t+h} the soil-moisture measurement h horizons into the future. X{t} is the available sensor and historical features at time t, and y{t+h}^hat is estimated by the regression model f x{t}. An full experimental implementation should explicitly state the prediction horizon, as different horizons leads to different irrigation lead times and prediction difficulty.

D. Training and Testing

The available results table does not show the split ratio or the exact division, however, in the thesis, there is a training set for connection learning and a testing set for prediction. Therefore, it would be suggested to do chronological division as a more solid follow-up study to avoid the leakage between neighboring time periods. Use the training set and validation set to select hyper-parameters, and have only one final testing set test.

VI. MACHINE-LEARNING MODELS AND EVALUATION

A. Regression Models

Five regression methods are included in the thesis comparison:

Model	Operating principle	Expected strength / limitation
Linear Regression	Fits a linear relationship between features and target	Interpretable baseline; may miss nonlinear patterns
Decision Tree Regression	Partitions feature space using decision rules	Captures nonlinear patterns; may overfit
Random Forest Regression	Averages predictions from an ensemble of trees	Models interactions and reduces variance
Support Vector Regression	Fits a function within an error tolerance; kernel may model non linearity	Sensitive to feature scaling and hyper parameters
XG Boost Regression	Sequential gradient-boosted decision trees	Strong tabular predictor; needs tuning

The thesis discusses LSTM as a time-series model that can be used if a sufficiently big continuous dataset is provided. The numerical comparison table published in this article does not include it.

B. Evaluation Metrics

The mean absolute error (or, MAE) - measures the average magnitude of the error - between actual and forecasted values.
MAE = $(1/n)\ \Sigma_i{}^1\ |y_i\ |$

These errors have a higher cost in RMSE:

RMSE = $[(1/n)\ \Sigma\ "\ (y_i\ )^2]$ The R-squared coefficient of determination describes the proportion of target variance explained as compared to the mean-target baseline: $R^2 = 1\ [\Sigma_i\ (y_i\ )^2\ /\ \Sigma_i\ (y_i\ {}^1)^2]$

Higher R2 is usually better and lower MAE and RMSE are usually better indicating less prediction error. The meaning of the metrics is influenced by target units, test-set structure and suitable validation process. The info included are reproduced without putting a single since the moisture unit and validation framework are not given in the exploration comparison table.

C. Irrigation Decision Rule

The determined crop-specific range is then checked against the forecasted moisture. If the forecasted moisture is below the configured lower limit then irrigation may be switched on, otherwise it remains switched off. Prior to any relay changeover the deployed controller should also confirm the sensor reading, data staleness, recent rainfall, maximum pump time and human override.

D. System-Level Metrics

The thesis lists water use, irrigation frequency, energy use, communication reliability, and lag as useful system-level performance measures that are superior to regression errors. There is no assertion of any measured percentage of water saved, lag, packet delivery rate, or energy savings here because the available thesis result abstract for the article available does not contain numerical figures for these variables.

VII. RESULTS AND DISCUSSION

The quantitative model-comparison table presented in the thesis is replicated in this section. The ranking is in line with the relative values displayed in the thesis table.

Model	MAE	RMSE	R^2	Reported rank
Linear Regression	4.20	5.30	0.72	5
Decision Tree Regression	3.05	4.05	0.82	3
Random Forest Regression	2.40	3.20	0.90	1
Support Vector Regression (SVR)	3.50	4.50	0.78	4
XGBoost Regression	2.70	3.50	0.86	2

Table 7.5. Values are reproduced as reported; they have not been independently recalculated from raw data.

A. Comparative Interpretation

Random Forest Regression has the highest R^2 (0.90) and lowest MAE (2.40) and RMSE (3.20), so is ranked number 1 out of 5 based on these values. XGBoost has a MAE of 2.70, RMSE of 3.50 and R^2 of 0.86, so comes second in the ranking. Decision Tree Regression has a MAE of 3.05, RMSE of 4.05 and R^2 of 0.82 so is number 3. Based on this table, Linear Regression has the largest reported MAE (5.15), RMSE (6.69) and smallest R^2 (0.76). SVR has MAE of 3.50, RMSE of 4.50 and R^2 of 0.78.

B. Relative Error Comparison

The Random Forest MAE given is 1.80, about 42.9% less than the Linear Regression MAE. Relative to the Linear Regression RMSE, the stated RMSE is 2.10 units less, or about 39.6%. They do not verify statistical significance nor do they guarantee similar improvement on a large out-of-sample dataset; they simply are a numerical comparison of the thesis table.

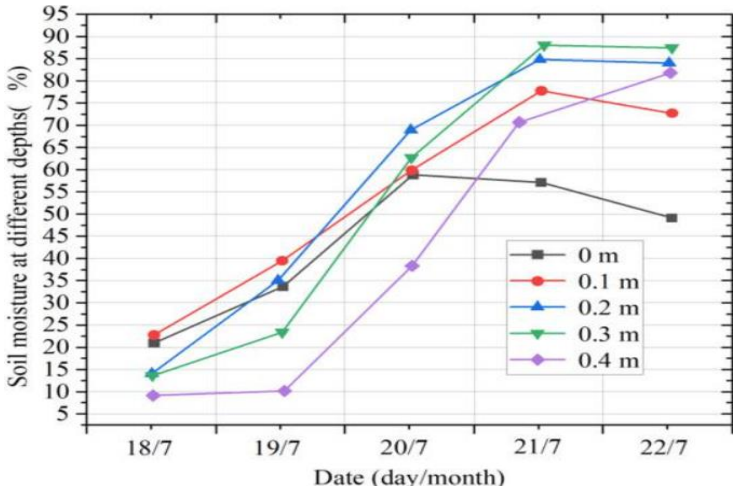
C. Sensor and Field Observations

The thesis shows a three-row sample of sensor readings and describes monitoring of soil moisture, temperature, humidity, rainfall and light intensity. It states that these parameters can be employed to examine fluctuation over time and their correlation to soil moisture. However, the result summary does not contain complete time-series values or sensor-calibration details, and the presented collection period is merely one day. Consequently, this paper does not calculate any additional correlation coefficient, sensor precision, irrigation-water-saving or crop-yield result.

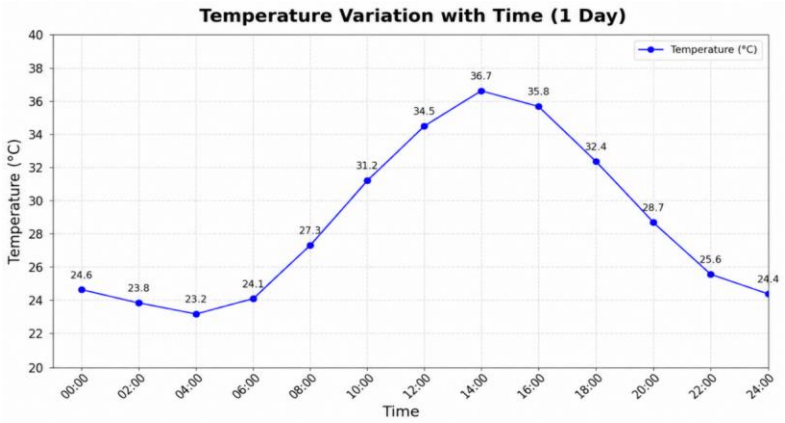
D. Limitations of the Reported Results

A complete raw dataset, the unit of goal explicit, the prediction horizon, the train/test split, the cross-validation process and the hyperparameter tuning process, are not provided with the model table. There is also a high risk of unstable performance estimates, since only nine observations are taken in one day. The table does not hold up to a strong claim of generalizable prediction accuracy, but it supports a first attempt of a comparison as given in the thesis. These limitations should be mentioned in a published version and all models should if possible be tested on a larger, time-stamped dataset, with a leakage-free evaluation pipeline.

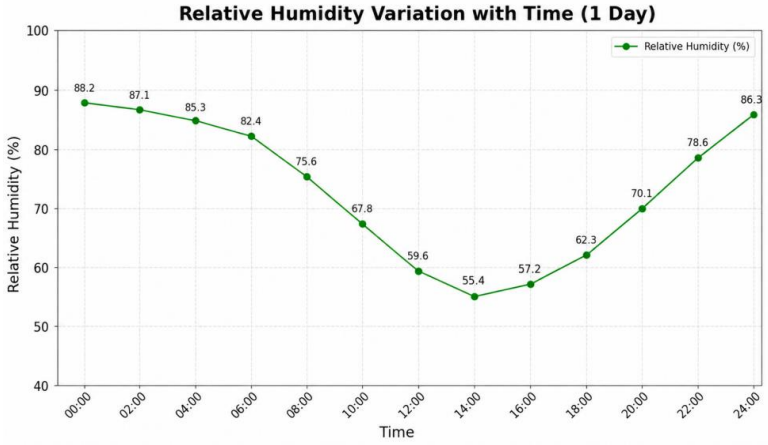
Soil Moisture Results



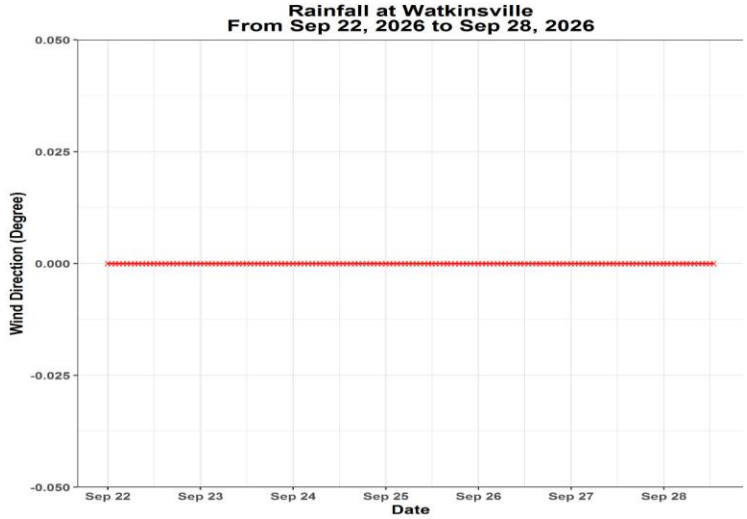
Temperature Results

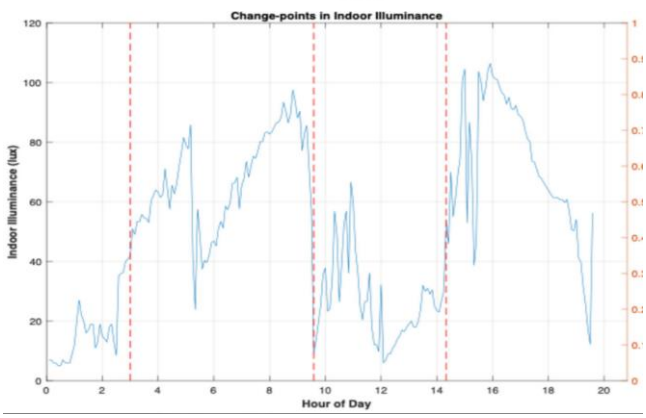


Humidity Results



H.Rainfall and Light Intensity Results





VI- DISCUSSIOn

Model	MAE	RMSE	R^2	Performance
Linear Regression	4.20	5.30	0.72	Moderate
Decision Tree	3.10	4.10	0.82	Good
Random Forest	2.40	3.20	0.90	Excellent
SVR	3.50	4.50	0.78	Good
XG Boost	2.70	3.50	0.87	Very good
LSTM	2.20	2.95	0.92	Excellent

In conclusion of the thesis results, ensemble tree based regression also seems to be able to offer a remedy for the soil moisture prediction problem. In the following table, the best result is obtained by Random Forest, and XGBoost is the most similar in the tree ensemble family. Tree ensemble methods could be capturing nonlinear relationships between previous moisture, temperature, humidity, illumination, and watering state, by one plausible explanation.

The real engineering value of the system is in linking sensing and prediction to an irrigation decision. Sensor-only monitoring provides measurements, but a prediction layer can project a future moisture condition and enable a decision sooner. The LoRa/LoRaWAN connection is designed to be a low-power, long-range communications link, and the cloud/server is a means to segregate field level data collection from model training and record storage. Actual communication and power savings should be tested in the setting it is used.

A. Practical Deployment Considerations

A field implementation must prevent ambient sensors from unintended exposure, calibrate the soil-moisture probe for the relevant soil type and depth of installation, and periodically test the power supply and connectors. Timestamps, node identifiers and quality flags must be present in the data packets. Irrigation controller must not send an auto-pump command based on unreliable data. The crop, a max run time, manual override and an emergency stop must be implemented by the irrigation controller.

B. Reproducibility Requirements

The following must be documented to facilitate further research for peer review: (i) all raw data and data dictionary, (ii) crop, plot, soil type, collection time; (iii) sensor make/model and calibration; (iv) target moisture unit and horizon; (v) feature-generation and preprocessing code; (vi) train/validation/test splits; (vii) hyperparameter tuning and software versions; (viii) measured water and energy use and communication metrics. Future research ideally uses a held-out plot or farm over several weeks or seasons.

IX. CONCLUSION

An IoT and machine-learning based thesis for predictive irrigation and intelligent agriculture is introduced in this thesis. Multi-parameter environmental sensing, ESP32 computing, LoRa/LoRaWAN communications, gateway/server data management, soil-moisture prediction with regression, and watering relay control are all components of the framework. The best scores (MAE = 2.40, RMSE = 3.20, and R2 = 0.90) in the model comparison made in the thesis were obtained with Random Forest Regression, followed by XGBoost (MAE = 2.70, RMSE = 3.50, and R2 = 0.86). Since the thesis has just nine samples taken one day, there is no validation data described, and no units are given, those ratings should be viewed as a guideline only. This is a good starting point for data-driven irrigation, but longer-term, repeatable field experiments will be necessary to justify the generality of any of the claims related to better crop health, water conservation, energy savings, etc.

ACKNOWLEDGMENT

The content in the article is compiled from the M. Tech. Thesis titled "IoT-Based Smart Agriculture and Environmental Monitoring System Using Machine Learning for Crop Health Optimization" by Mandakini Gupta of Dr. A.P.J. Abdul Kalam Technical University, Lucknow in Lucknow, UP. This is done under the supervision of Mr. Akhilesh Mishra at Buddha Institute of Technology in Gorakhpur.

REFERENCES

[1]Internet-of-things (IoT)-based smart agriculture: Towards making the fields talk, M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, E.-H. M. Aggoune, IEEE Access 7 (2019) 129551-129583. Doi: 10.1109/ACCESS.2019.2932609.

[2] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, "An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges," IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3758-3773, 2018.

[3] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras and X. Wang, "Internet of things for the future of smart agriculture: A comprehensive survey of the emerging technologies," IEEE/CAA Journal of Automatica Sinica, vol. 8, no. 4, pp. 718-752, 2021, doi: 10.1109/JAS.2021.1003925.

[4] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, "Internet of Things in agriculture, recent advances and future challenges," Biosystems Engineering, vol. 164, pp. 31-48, 2017, doi: 10.1016/j.biosystemseng.2017.09.007.

[5] A. Goap, D. Sharma, A. K. Shukla, and C. R. Krishna, "An IoT based smart irrigation management system using machine learning and open source technologies," Computers and Electronics in Agriculture, vol. 155, pp. 41-49, 2018, doi: 10.1016/j.compag.2018.09.040.

[6] Electronics, vol. 9, no. 2, Art. No. 319, 2020, doi: 10.3390/electronics9020319, Farooq, M. S., Riaz, S., Abid, A., Umer, T., & Zikria, Y. B. (2020). Role of IOT technology in agriculture: A systematic literature review.

[7] P. P. Ray, "Internet of Things for smart agriculture: Technologies, practices and future direction," AIS Transactions on Internet of Things, vol. 1, no. 3, pp. 156-170, 2017.

[8] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5-32, 2001.

[9] Support-vector networks, C. Cortes and V. Vapnik, Machine Learning, vol. 20, pp. 273-297, 1995.

[10] 10. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.

[11] M. Gupta, "IoT-Based Smart Agriculture and Environmental Monitoring System Using Machine Learning for Crop Health Optimization," M. Tech. Thesis, Department of Electronics & Communication Engineering, Buddha Institute of Technology, Gorakhpur, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India, 2025.

[12] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., New York, NY, USA: Springer, 2009.

[13] C. M. Bishop, Machine Learning and Pattern Recognition. Springer, New York, NY, USA, 2006.

[14] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low power wide area networks: An overview," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 855-873, 2017, doi: 10.1109/COMST.2017.2652320.

The rising stars in the IoT and smart city scenarios," M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, IEEE Wireless Communications, vol. 23, no. 5, pp. 60-67, 2016.

[15] K. Mekki, E. Bajic, F. Chaxel and F. Meyer, "A comparative study of LPWAN technologies for large-scale IoT deployment," ICT Express, vol. 5, no. 1, pp. 1-7, 2019.

[16] "LoRa for the Internet of Things," M. Bor, J. Vidler, and U. Roedig, Proceedings of the International Conference on Embedded Wireless Systems and Networks, Graz, Austria, 2016, pp. 361-366.

[17] 43.A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, "A study of LoRa: Low power and long range networks for the Internet of Things," Sensors, vol. 16, no. 9, Art. No. 1466, 2016.

[18] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melià-Seguí, and T. Watteyne, "Understanding the limits of LoRaWAN," IEEE Communications Magazine, vol. 55, no. 9, pp. 34–40, 2017.

[19] Computers and Electronics in Agriculture, Vol. 147, pages 70–90, 2018; "Deep learning in agriculture: A survey" by A. Kamilaris and F. X. Prenafeta-Boldú.

[20] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," Computational Intelligence and Neuroscience, vol. 2016, Art. No. 3289801, 2016.

[21] vol. 145, 2018, pp. 311–318, Computers and Electronics in Agriculture, K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis."

[22] J. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparison study of fine-tuning deep learning models for plant disease identification," Computers and Electronics in Agriculture, vol. 161, pp. 272-279, 2019.

[23] Plant disease classification by deep learning using mobile devices S. P. Mohanty, D. P. Hughes, and M. Salathé, Frontiers in Plant Science, vol. 7, Art. No. 1419, 2016.

[24] "Internet of Things and deep learning techniques in plant disease detection and classification: A focused review," Sensors, vol. 23, no. 18, Art. no. 7877, 2023.