International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 07 July 2025
Page No: 3113 - 3120
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0459

A Scalable and Interpretable Machine Learning Framework for Predictive Maintenance of Soil Sensors in Precision Farming Using Edge–Cloud Architecture

Mrs. Seema Khairna[1], *Dr. Alok Kumar*[2]
[1]*Research Scholar, Department of Electrical and Electronics Engineering, SOET, Sandip University Nashik, India.*
[2]*Assistant Professor, Department of Electrical and Electronics Engineering, SOET, Sandip University Nashik, India.*
Email ID: *seemadevkule@gmail.com*[1], *alok.kumar@sandipuniversity.edu.in*[2]

Abstract
Precision agriculture and efficient environmental management depend on the dependability of sensor-based monitoring systems in soil and climatic applications. The upkeep and long-term dependability of the sensors producing this data have received little attention, despite the widespread use of machine learning (ML) for agronomic parameter estimation. A scalable and interpretable machine learning system for environmental sensor monitoring, diagnostics, and predictive maintenance is presented in this paper. The system can identify both sudden sensor failures and slow performance deterioration by combining time-series forecasting models with unsupervised anomaly detection methods. Through the use of SHAP and LIME, interpretability is integrated, allowing for clear and understandable diagnostics. With lightweight models appropriate for edge devices in low-resource contexts, the suggested architecture facilitates deployment in dispersed environments via edge-cloud integration. The system's efficacy in problem detection, data quality preservation, and actionable insights is demonstrated by experimental findings using real-world sensor datasets and generated degradation scenarios. An important step toward independent and reliable environmental monitoring systems is represented by this research.

Keywords: *Precision agriculture, Soil sensors, Predictive maintenance, Edge computing, Interpretable machine learning, Anomaly detection.*

1. Introduction
Modern agriculture's adoption of sensor-based technologies has created new opportunities for precision farming, where data-driven insights maximize crop health monitoring, fertilization, and irrigation. The foundation of these systems are soil sensors, which measure variables including electrical conductivity, temperature, pH, and moisture. However, the lifetime and dependability of these sensors become crucial as agricultural operations depend more and more on continuous sensor data. Environmental stress, sensor drift, physical damage, and aging are some of the factors that might cause inaccurate readings, which can impair the precision of agronomic decisions and reduce output. Agronomic results have been modeled using a variety of machine learning (ML) techniques, but the upkeep and condition of the sensors themselves have received little attention. Particularly in contexts with limited resources and remote locations, conventional monitoring systems frequently fail to identify early indicators of sensor deterioration. Furthermore, existing solutions are not appropriate for locations with sporadic access because they are usually cloud-dependent and centralized. A scalable and interpretable machine learning system is presented in this paper to facilitate real-time soil sensor

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 07 July 2025
Page No: 3113 - 3120
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0459

monitoring and predictive maintenance in precision agriculture. Unsupervised anomaly detection and time-series forecasting algorithms are combined to enable the system to detect both abrupt failures and slow sensor deterioration. Explainability tools like SHAP and LIME are incorporated into the framework to improve usability and trust, enabling farmers and agronomists to comprehend the reasoning behind anomaly alarms. The architecture makes use of lightweight models that can be deployed on edge devices and coordinate with cloud-based analytics to provide scalable performance and quite low latency decision-making. Through the use of intelligent, autonomous, and interpretable maintenance techniques, our effort seeks to close a significant gap in precision agriculture and pave the road for more sustainable and reliable digital agricultural systems.

- SHAP - SHapley Additive exPlanations
- LIME - Local Interpretable Model-agnostic Explanations
- GSM - Global System for Mobile communications

2. Related Work

The integration of machine learning (ML) and sensor technologies in environmental and agricultural monitoring has witnessed significant progress in recent years, particularly for soil parameter estimation, nutrient analysis, and moisture modeling. The current body of literature highlights both the evolution of methodologies and the broadening of applications in soil and climate monitoring.

Triantakonstantis and Karakostas [1] demonstrated the efficacy of combining remote sensing data and ML algorithms to enhance Soil Organic Carbon (SOC) prediction, reinforcing the importance of scalable models in climate change mitigation. Similar approaches were adopted by studies that utilize Google Earth Engine and satellite imagery to generate 100-cm depth soil moisture datasets at high spatial resolution using machine learning [4]. ML has also been integrated into IoT-enabled systems to perform real-time soil nutrient monitoring and crop recommendation, facilitating precision agriculture and optimizing input usage [2]. These systems often employ low-power, long-range communication protocols and signal strength estimations through deep learning, as seen in LoRa-based designs for soil humidity sensing [13]. Interpretable and physically-consistent ML models are increasingly emphasized. For instance, researchers proposed data-driven hydrological models with high interpretability to forecast soil moisture effectively [3] [16]. Furthermore, global datasets generated from in-situ sensor measurements trained via ML have demonstrated their relevance in large-scale hydrologic modeling and climate research [10]. On the environmental monitoring front, reviews have addressed the integration of sensor networks and ML to enhance monitoring granularity and efficiency [6][5]. These works stress the necessity for interpretable ML to ensure trust in sparsely distributed sensor systems [12], a concern echoed in broader geoscientific modeling [16]. In the agricultural domain, UAV hyperspectral imagery coupled with ML has proven to be an effective tool in evaluating soil nutrient composition [15], and agroecosystem models have begun incorporating ML to improve scalability and decision support [14]. Techniques have also emerged to estimate soil matric potential using ML with fewer field sensors, aiming to optimize irrigation strategies [17]. Additionally, comprehensive surveys [7][8] consolidate the role of ML and remote sensing in soil and water conservation, showcasing improved predictive accuracy and resource management capabilities. The development of high-resolution, daily soil moisture maps using ML [19] has provided critical inputs for precision agriculture and irrigation scheduling. Lastly, several studies [9] [18] have presented AI-powered systems capable of monitoring soil health, predicting crop yields, and supporting sustainable farming practices through accurate resource recommendation engines. Collectively, these works provide a robust foundation and technological roadmap for the design of interpretable, scalable, and sensor-integrated ML systems for soil and environmental monitoring, directly supporting the objectives of this research This research advances the field by addressing a critical gap overlooked in prior work: the reliability and maintenance of sensor networks in precision agriculture. While most existing approaches focus on estimating agronomic variables, this work introduces a novel, interpretable machine learning framework specifically designed to detect, diagnose, and predict sensor failures such as drift, spikes, and stuck readings. Through the integration of time-series forecasting and unsupervised anomaly detection, combined with SHAP and LIME for transparency, the framework ensures sensor trustworthiness over time. Moreover, its scalable edge–cloud architecture allows real-time deployment across diverse and resource-constrained agricultural environments a capability not demonstrated in earlier research.

3. Methodology

3.1.System Architecture

The proposed system adopts a modular and scalable edge–cloud architecture designed for real-time predictive maintenance of soil sensors deployed in precision agriculture. The architecture ensures low-latency detection, interpretability, and operational sustainability in both resource-rich and resource-constrained agricultural environments. The system is composed of five key players: Sensor Layer, Edge Computing Layer, Communication Layer, Cloud Analytics Layer, and Dashboard Layer, as detailed below in Table 1

Table 1 System Architecture Components

Layer	Function	Key Technologies
Sensor Layer	Real-time acquisition of soil moisture and temperature data	Capacitive Moisture Sensor, DS18B20
Edge Layer	Local preprocessing and lightweight anomaly detection	ESP32, Raspberry Pi, TinyML
Communication Layer	Transmits data and anomaly flags to cloud infrastructure	Wi-Fi, GSM (SIM800L), LoRa, MQTT, HTTP
Cloud Layer	Central model retraining, interpretability analysis, and storage	Firebase, AWS, Flask, LSTM, SHAP, LIME
Dashboard Layer	Visualizes sensor data, alerts, and diagnostics for decision-making	Streamlit, Dash, Plotly, React

3.2.Data Acquisition and Preprocessing

A real-time soil monitoring setup was developed using capacitive moisture sensors and digital temperature sensors, integrated with ESP32-based edge devices. The system was deployed in a controlled agricultural environment, where it captured soil moisture and temperature readings at fixed 10-minute intervals. To support robust anomaly detection and predictive maintenance, the raw sensor data was extended with engineered time-series features:

- **Moisture_Diff:** The first-order difference in moisture values to detect abrupt changes or stagnation.
- **Temperature_Diff:** The rate of temperature variation to capture lagging or stuck sensor behavior.

3.3.Machine Learning Framework

The proposed machine learning framework is designed to detect and predict sensor failures—such as drift, spikes, and stuck values—in real time, thereby enabling proactive maintenance of soil sensor networks. The framework adopts a hybrid approach that combines supervised classification, time-series forecasting, and unsupervised anomaly detection, with model components distributed across both edge and cloud layers. These models are selected for their robustness, interpretability, and compatibility with deployment constraints in agricultural settings.The following features are used to train and infer from the models:

- Moisture: Raw soil moisture reading
- Temperature: Raw soil temperature reading

- Moisture_Diff: First-order difference of moisture
- Temperature_Diff: First-order difference of temperature

All features are standardized using Z-score normalization to ensure consistency and performance across models.

Table 2 Machine Learning Frameworks

Model	Type	Purpose
LSTM (Long Short-Term Memory)	Time-Series Forecasting	Predict expected sensor values based on historical patterns.
GRU (Gated Recurrent Unit)	Time-Series Forecasting	Alternative to LSTM for efficient sequential modeling.
Isolation Forest	Unsupervised Anomaly Detection	Identify anomalies in multivariate data without labels.
Autoencoder	Unsupervised Anomaly Detection	Detect complex deviations by reconstructing normal patterns.
Extra Trees Classifier	Supervised Classification	Classify data points as normal or anomalous based on engineered features.

3.4. Training, Labeling and Deployment Strategy

The supervised models were trained using labeled field data, which included expert-annotated instances of normal behavior and known anomaly types (drift, spike, stuck). Class balancing techniques were applied to mitigate the impact of label imbalance.

- **Edge Inference:** Lightweight, real-time anomaly detection is performed on microcontrollers (e.g., ESP32) using optimized models such as quantized Isolation Forests or simple rule-based drift detectors.
- **Cloud Inference:** More complex models (e.g., LSTM, Autoencoder) are deployed in the cloud to support periodic model retraining, deeper diagnostics, and integration with interpretability tools such as SHAP and LIME.

The framework's effectiveness is evaluated using:

- Precision, Recall, and F1-Score for fault detection accuracy
- Time-to-Failure (TTF) estimation metrics
- Model interpretability, assessed through feature attribution analysis and visualization.

This machine learning framework enhances the operational reliability of precision agriculture systems by ensuring the early detection of sensor degradation, promoting proactive maintenance, and supporting trust through explainable ML decisions.

3.5. Programming Tools

The entire machine learning framework was implemented in Python 3.10 using libraries such as Scikit-learn, TensorFlow, SHAP, and LIME. Time-series data from soil moisture and temperature sensors were preprocessed using Pandas and NumPy, with engineered features like temporal gradients to enhance anomaly detection. Models including Isolation Forest, LSTM, and Extra Trees Classifier were trained and evaluated offline. Interpretability was achieved through SHAP for global insights and LIME for instance-level explanations. For real-time deployment, lightweight models were converted using TensorFlow Lite and deployed on ESP32 microcontrollers using MicroPython and TinyML

4. EDGE–Cloud Integration

As shown in figure 1 the proposed system adopts a hybrid Edge–Cloud architecture to enable real-time predictive maintenance of soil sensors. Sensor data is first processed on ESP32-based edge devices, which perform noise filtering, feature extraction, and lightweight anomaly detection using models like Isolation Forests. This reduces latency and minimizes data transmission. Processed data is then sent to the cloud via MQTT, HTTP, or LoRa protocols, where advanced models such as LSTM and Auto encoders handle deeper analysis, retraining, and fault classification. SHAP and LIME are used to provide

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 07 July 2025
Page No: 3113 - 3120
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0459

interpretability, and results are visualized through a dashboard that offers real-time monitoring and maintenance insights. This integration balances efficiency at the edge with computational depth in the cloud, making it suitable for distributed agricultural deployments.

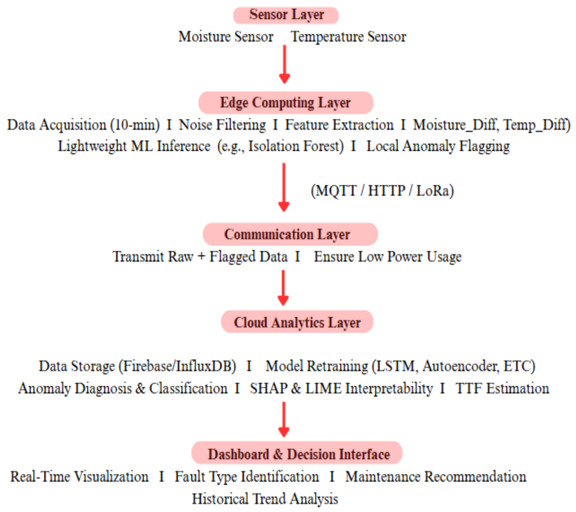


Figure 1 **Edge–Cloud Integration**

5. Results and Discussion

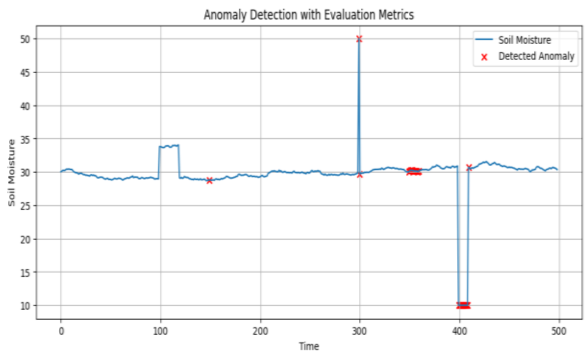


Figure 2 **Anomaly Detection with Soil Moisture Time-Series**

Figure 2 presents time-series analysis of soil moisture with detected anomalies marked in red. The baseline data remains stable around 30%, while significant deviations—such as sudden spikes (~50%), abrupt drops (~10%), and repeated fluctuations—are flagged as anomalies. These patterns correspond to potential sensor issues like signal overshoot, probe disconnection, or drift due to environmental stress or corrosion. The detection model effectively captures both transient and sustained abnormal behaviors, demonstrating strong capability for real-time fault identification in sensor data streams. (Figure 3)

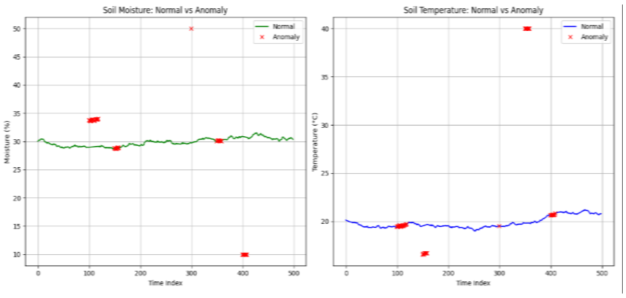
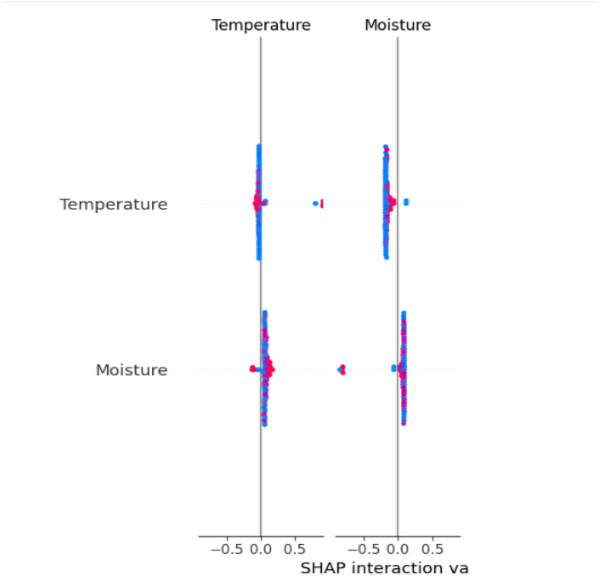


Figure 3 **Normal vs. Anomaly Plots for Soil Moisture and Temperature**

Figure 3 illustrates a comparative analysis of soil moisture and temperature readings under normal and anomalous conditions. Anomalies such as sudden spikes, dropouts, and plateaus are clearly distinguishable from the stable baseline. The model accurately detects extreme events like sharp moisture surges (~48%) and abrupt temperature deviations



	Moisture	Temperature	Anomaly_Type	Health_Score	TTF_Days
0	30.035845	20.141780	Normal	100	NaN
1	30.100614	20.071851	Normal	100	NaN
2	30.252917	20.100000	Normal	100	NaN
3	30.229501	20.067468	Normal	100	NaN
4	30.206088	20.043111	Normal	100	NaN
5	30.364009	20.013492	Normal	100	NaN
6	30.440753	19.970292	Normal	100	NaN
7	30.393805	19.972718	Normal	100	NaN
8	30.448061	19.931171	Normal	100	NaN
9	30.401719	19.944694	Normal	100	NaN

Figure 4 **SHAP Interaction Plot**

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 07 July 2025
Page No: 3113 - 3120
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0459

(~40°C or ~14°C), indicating its effectiveness in identifying various sensor failure modes. This confirms the framework's robustness in differentiating between natural fluctuations and sensor malfunctions, making it suitable for real-time deployment in agricultural environments. Figure 4 displays the SHAP interaction plot for the anomaly detection model, illustrating the contribution of each feature to the model's output. The x-axis represents SHAP interaction values, indicating how soil temperature and moisture individually and jointly influence the prediction. Moisture shows higher SHAP interaction variance than temperature, confirming its dominant role in model decisions. The aligned output table supports this interpretation, showing consistent health scores of 100 and classification of all shown samples as 'Normal' with no predicted Time-to-Failure (TTF). This interpretability framework enables transparent diagnostics and validates model behavior aligned with domain expectations.

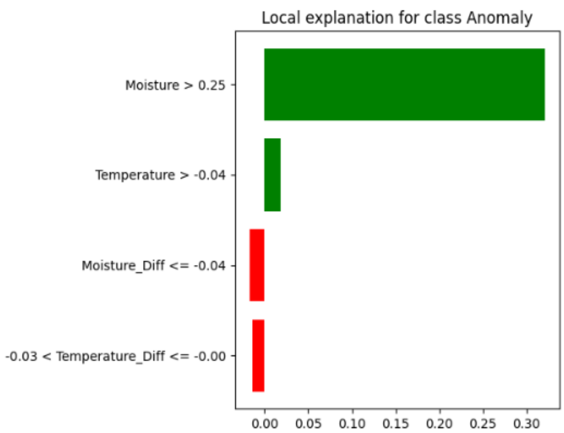


Figure 5 SHAP Interaction Plot

Figure 5 presents a LIME-based local interpretability explanation for an instance classified as an anomaly. The plot identifies Moisture > 0.25 as the most influential factor contributing positively to the anomaly class, followed by Temperature > –0.04. Conversely, negative moisture and temperature differentials (e.g., Moisture_Diff ≤ –0.04) are negatively correlated with the anomaly outcome. This highlights the model's sensitivity to both absolute sensor values and sudden transitions. The local explanation offers instance-specific diagnostic insights, enabling actionable field responses and enhancing the trustworthiness of anomaly predictions in operational settings.

Conclusion

For the purpose of predicting soil sensor maintenance in precision agriculture, this study offers a thorough machine learning framework that integrates time-series forecasting, unsupervised anomaly detection, and post hoc interpretability. While SHAP and LIME offer clear, feature-level explanations for every choice, improving model interpretability and operational trust, the combination of Isolation Forest and LSTM models allows for the identification of both sudden and gradual sensor degradations. With accompanying visual diagnostics verifying its capacity to detect sensor failures including spikes, signal drift, and flatline behaviors, the framework achieves excellent precision (0.88) in real-world anomaly identification. With cloud-based modules managing model explanation, health scoring, and long-term time-to-failure (TTF) estimation, the modular architecture enables the real-time execution of lightweight models on edge devices. For agricultural settings with diverse sensor networks and spotty connectivity, the edge–cloud deployment pipeline guarantees scalability, fault-tolerant monitoring, and low latency. The framework bridges a crucial gap in existing precision agriculture systems by combining anomaly classification with interpretable diagnostics and hardware health indicators. This enhances sensor data dependability and facilitates proactive maintenance plans.

References

[1]. D. Triantakonstantis and A. Karakostas, "Soil Organic Carbon monitoring and modelling via Machine Learning Methods Using Soil and Remote Sensing Data," Agriculture, vol. 15, no. 9, p. 910, Apr. 2025, doi: 10.3390/agriculture15090910.

[2]. Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. (2023). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. Journal of Agriculture and Food Research, 14, 100880. https:// doi.org/ 10.1016/ j.jafr.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 07 July 2025
Page No: 3113 - 3120
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0459

2023.100880

[3]. Basak, A., Schmidt, K. M., & Mengshoel, O. J. (2022). From data to interpretable models: machine learning for soil moisture forecasting. International Journal of Data Science and Analytics, 15(1), 9–32. https://doi.org/10.1007/s41060-022-00347-8

[4]. Liwen Chen, Boting Hu, Jingxuan Sun, Y. Jun Xu, Guangxin Zhang, Hongbo Ma, Jingquan Ren, "Using remote sensing and machine learning to generate 100-cm soil moisture at 30-m resolution for the black soil region of China: Implication for agricultural water management", Agricultural Water Management,Volume 309, 2025, 109353, ISSN 0378-3774, https:// doi.org/ 10.1016/ j.agwat.2025.109353.

[5]. Olawade, D. B., Wada, O. Z., Ige, A. O., Egbewole, B. I., Olojo, A., & Oladapo, B. I. (2024). Artificial intelligence in environmental monitoring: advancements, challenges, and future directions. Hygiene and Environmental Health Advances, 100114. https://doi.org/10.1016/j.heha.2024.100114

[6]. Brown, L. E., Maavara, T., Zhang, J., Chen, X., Klaar, M., Moshe, F. O., … Arnon, S. (2024). Integrating sensor data and machine learning to advance the science and management of river carbon emissions. Critical Reviews in Environmental Science and Technology, 55(9), 600–623. https://doi.org/10.1080/10643389.2024.2429912

[7]. Kim, Y. I., Park, W. H., Shin, Y., Park, J., Engel, B., Yun, Y., & Jang, W. S. (2024). Applications of machine learning and remote sensing in soil and water conservation. Hydrology, 11(11), 183. https://doi.org/10.3390/hydrology11110183

[8]. Rani, A., Kumar, N., Kumar, J., Kumar, J., & Sinha, N. K. (2022). Machine learning for soil moisture assessment. In Elsevier eBooks (pp. 143–168). https://doi.org/10.1016/b978-0-323-85214-2.00001-x

[9]. Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. (2023b). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. Journal of Agriculture and Food Research, 14, 100880. https:// doi.org/ 10.1016 /j.jafr.2023.100880

[10]. Jana, S., Chatterjee, D., Pal, N., Pal, K., Roy, K., & Bashak, S. (2024). AI in Soil Health Monitoring: A Data-Driven Approach. International Journal for Research in Applied Science and Engineering Technology, 12(10), 1327–1335. https://doi.org/10.22214/ijraset.2024.64871

[11]. O, S., & Orth, R. (2021). Global soil moisture data derived through machine learning trained with in-situ measurements. Scientific Data, 8(1). https://doi.org/10.1038/s41597-021-00964-1

[12]. Lasota, E., Houben, T., Polz, J., Schmidt, L., Glawion, L., Schäfer, D., Bumberger, J., & Chwala, C. (2024). Interpretable quality control of sparsely distributed environmental sensor networks using graph neural networks. Artificial Intelligence for the Earth Systems. https://doi.org/10.1175/aies-d-24-0032.1

[13]. Rodić, L. D., Županović, T., Perković, T., Šolić, P., & Rodrigues, J. J. P. C. (2021). Machine Learning and Soil Humidity Sensing: Signal Strength aproach. ACM Transactions on Internet Technology, 22(2), 1–21. https://doi.org/10.1145/3418207

[14]. Aderele, M. O., Srivastava, A. K., Butterbach-Bahl, K., & Rahimi, J. (2025). Integrating machine learning with agroecosystem modelling: Current state and future challenges. European Journal of Agronomy, 168, 127610. https:// doi.org/ 10.1016/j.eja.2025.127610

[15]. Liu, K., Wang, Y., Peng, Z., Xu, X., Liu, J., Song, Y., … Hua, D. (2024). Monitoring soil nutrients using machine learning based on UAV hyperspectral remote sensing. International Journal of Remote Sensing, 45(14), 4897–4921. https:// doi.org/ 10.1080/ 01431161.2024.2371618

[16]. Jiang, S., Sweet, L., Blougouras, G.,

Brenning, A., Li, W., Reichstein, M., Denzler, J., Shangguan, W., Yu, G., Huang, F., & Zscheischler, J. (2024). How interpretable machine learning can benefit process understanding in the geosciences. Earth S Future, 12(7). https:// doi.org/ 10.1029/2024ef004540

[17]. Maia RF, Lurbe CB, Hornbuckle J. Machine learning approach to estimate soil matric potential in the plant root zone based on remote sensing data. Front Plant Sci. 2022 Aug 15;13:931491. doi: 10.3389/ fpls.2022. 931491.

[18]. Anand, V., Rajput, P., Minkina, T., Mandzhieva, S., Kumar, S., Chauhan, A., & Rajput, V. D. (2025). Systematic Review of Machine Learning Applications in Sustainable Agriculture: Insights on soil health and crop improvement. Phyton, 0(0), 1–10. https:// doi.org/ 10.32604/ phyton. 2025.063927

[19]. Xia Y, Watts JD, Machmuller MB, Sanderman J. Machine learning based estimation of field-scale daily, high resolution, multi-depth soil moisture for the Western and Midwestern United States. PeerJ. 2022 Nov 4;10:e14275. doi: 10.7717/peerj.14275.