Soil-Sensing-Based Irrigation Decision Modeling for Greenhouse Tomato Crops Using Machine Learning

André Ricardo Raucci[1], Paulina Arregocés Guerra[2],
Carlos Alberto Kamienski[1], Fabíola Martins Campos de Oliveira[1],
Sergio Armando Gutiérrez Betancur[2], Juan Felipe Botero Vega[2]

[1] Federal University of ABC (UFABC), Brazil
[2] University of Antioquia (UdeA), Colombia

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.

Optimizing water-use efficiency is increasingly important for high-demand crops such as tomatoes (Solanum lycopersicum), which are highly sensitive to irrigation management. In greenhouse environments, the integration of Internet of Things (IoT) sensing technologies with Machine Learning enables the transition from schedule-based irrigation to data-driven decision-making. Modern soil sensors continuously monitor variables including volumetric water content, temperature, and electrical conductivity, providing valuable information for precision irrigation systems. However, transforming these large volumes of sensor data into actionable knowledge for irrigation management remains a significant challenge. This study investigates whether high-resolution soil sensor data can effectively support irrigation decision-making within an optimal early-morning time window. An experiment was conducted at Universidad del Quindío, Colombia, using Chonto Roble 956 F1 tomatoes cultivated in sandy loam soil under greenhouse conditions. Soil volumetric water content, electrical conductivity, and temperature were monitored at multiple depths using IoT sensors, with measurements collected every 30 minutes throughout the crop cycle. Historical irrigation management records were used to define two target variables: (i) a binary irrigation decision indicating whether irrigation was applied on a given day and (ii) the irrigation volume applied. Temporal windows with cutoff times ranging from 6:00 AM to 12:00 PM were evaluated to determine how much early-morning information was required to support accurate predictions. Two machine learning tasks were investigated using Logistic Regression and Random Forest algorithms: binary classification (irrigation versus no irrigation) and regression (irrigation volume). Results demonstrated that the classification models substantially outperformed the regression ones. The best-performing model, a Random Forest classifier, achieved an F1-score of 0.96, a precision of 0.95, and a recall of 0.97. In contrast, all regression models produced negative R^2 values, indicating limited capability to explain irrigation volume variability beyond a baseline mean prediction. Among the evaluated temporal windows, the 8:00

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

AM cutoff provided the most favorable balance between predictive performance and operational feasibility. Feature importance analysis identified soil moisture gradients across depths and days since the last irrigation event as the most influential predictors. These findings suggest that framing irrigation management as a binary classification problem constitutes a more practical, accurate, and interpretable precision agriculture strategy than direct irrigation volume prediction. Furthermore, reliable irrigation decisions can be supported using only early-morning soil measurements, providing a foundation for intelligent irrigation management systems that improve water-use efficiency in greenhouse tomato production.

Keywords.

Smart irrigation, soil moisture, windows analysis, water management

Introduction

Efficient irrigation management remains a major challenge in modern agriculture, particularly for water-sensitive crops such as tomato (Solanum lycopersicum L.), where inadequate water supply can reduce productivity and fruit quality while increasing resource consumption (Park et al. 2025; Ahmia et al. 2026). Advances in soil sensing, Internet of Things (IoT) technologies, and artificial intelligence (AI) have enabled continuous monitoring of soil variables, including volumetric water content (%VWC), temperature, and electrical conductivity (EC), providing valuable information for precision irrigation systems (Jones 2004; Allen et al. 1998; Fereres and Soriano 2007). However, transforming sensor data into actionable knowledge for irrigation management remains a challenge.

Recent studies have shown that supervised machine learning techniques can identify irrigation patterns and predict irrigation-related events from sensor data (Kim et al. 2008; O'Shaughnessy and Evett 2010). While classification models can estimate irrigation occurrence and regression models can predict applied water volumes, most studies focus on predictive performance rather than methodological factors that affect practical decision-making, such as the selection of temporal aggregation windows and the agronomic interpretation of model outputs.

The choice of temporal aggregation windows is particularly relevant because early-morning measurements may better represent soil water status before evapotranspiration intensifies throughout the day. Despite recent advances, irrigation management in greenhouse tomato production is still frequently based on fixed schedules rather than data-driven approaches, and operational constraints such as the need for early-day irrigation decisions are rarely considered in predictive modeling.

The objective of this study is to generate agronomic knowledge through the integration of IoT sensor data and AI techniques applied to irrigation management in tomato crops. More specifically, this study aims to: (i) evaluate the impact of different sensor reading time windows on the performance of classification and regression models; (ii) analyze the relevance of soil variables in irrigation decisions using feature importance metrics; and (iii) assess the limitations and applicability of using real field data to support irrigation decision-making.

Real data from a tomato crop monitored in Quindío, Colombia, were used, integrating measurements from CropX sensors (CropX Vertex, Tel Aviv, Israel) with agricultural management records. Multiple early-morning temporal cutoff windows (from 06:00 to 12:00 h) were analyzed to determine the minimum amount of data required for reliable predictions. The study was conducted under greenhouse conditions, characterized by protection from direct precipitation and reduced exposure to external environmental fluctuations, although without active climate control, providing a semi-controlled experimental environment that enables high-resolution monitoring while maintaining realistic agronomic variability. Unlike many previous works that focus primarily on prediction accuracy, the proposed framework is oriented toward decision support for growers, delivering interpretable and timely recommendations that can be directly integrated into daily irrigation practices (Breiman 2001, Friedman 2001).

This paper is organized as follows. The theoretical background and key concepts underlying the proposed approach are presented first. Next, the methodology is described, including the data sources and experimental context, the definition of temporal reading windows, the feature engineering process, the modeling approach, and the validation strategy. The experimental results are then presented and discussed, including an analysis of soil moisture dynamics, the characterization of irrigated and non-irrigated days, the performance of the classification and regression models, and the interpretation of feature importance metrics. Finally, the main conclusions are summarized and directions for future work are outlined.

Literature Review and Theoretical Foundation

Growing pressure on global water resources has driven active research aimed at automating and optimizing agricultural irrigation. The agricultural sector accounts for approximately 70% of available freshwater resources, and evaporation losses in traditional irrigation systems have been reported to reach up to half of the total water used (Zhang and Li 2025). A progressive transition has been documented from manual and reactive systems toward intelligent architectures based on IoT, AI, and distributed computing.

Evolution of Smart Irrigation: From Basic Automation to Autonomy

The evolution of smart irrigation systems has been characterized by six progressive stages, ranging from manual irrigation to fully autonomous operation (Dantas et al. 2025; Togneri et al. 2023). Most modern farms currently operate at intermediate stages, corresponding to precision irrigation with sensors and sub-zone-specific control systems. A cross-cutting finding is that the adoption of intelligent systems faces not only technical but also cultural and economic barriers, with the gap between technological development and field adoption remaining a significant challenge (Dantas et al. 2025; Togneri et al. 2023).

Regarding predictive modeling, Custódio and Prati (2024) compared classical statistical models and machine learning approaches for soil moisture prediction using real data from multiple locations. Univariate models, particularly ARIMA, outperformed more complex multivariate models in most cases, challenging the common assumption that algorithmic sophistication necessarily yields better predictions. These results have direct implications for irrigation decision support design: model selection should be guided by data characteristics rather than complexity, and classical statistical models should be maintained as strong baselines.

The temporal relationship between soil water status and crop response was examined by Ferreira et al. (2025), who documented a quantifiable lag of up to two days between soil moisture conditions and vegetation indices in drip-irrigated coffee crops. This finding suggests that stress conditions can be anticipated using available sensor data, reinforcing the relevance of temporal modeling for irrigation decision support. Complementarily, Zhang and Li (2025) demonstrated the feasibility of combining IoT sensor data with neural network models to automate irrigation type selection, achieving strong performance across six irrigation modalities.

Three cross-cutting themes emerge from this body of work. First, model parsimony is consistently supported: well-calibrated simple approaches often outperform complex alternatives. Second, the complementarity between in situ sensors and remote sensing enhances spatial coverage without proportionally increasing costs. Third, the dynamic lag between soil moisture and vegetation response represents a critical but underexplored variable with direct implications for early warning systems. Despite these advances, two gaps remain relevant to this study: the methodological treatment of temporal aggregation windows for sensor data has received limited attention, and the lack of standardized preprocessing pipelines limits the reproducibility of current irrigation decision models.

Theoretical Foundations and Integration with Smart Irrigation Research

Data-driven irrigation decision support systems are grounded in three key areas: precision agriculture sensing, environmental time series analysis, and machine learning (Togneri et al. 2023). Capacitive soil moisture sensors, such as the CropX, provide reliable multi-depth measurements of water content and are widely adopted in both research and operational settings.

From a temporal perspective, soil water dynamics are more stable during nighttime and early morning periods, when evapotranspiration is minimal. Aggregating sensor data within early morning windows (e.g., 06:00–10:00 h) has been shown to improve signal quality by capturing the direct effects of irrigation and infiltration while reducing noise from radiation and temperature fluctuations (Custódio and Prati 2024; Bwambale et al. 2023).

Regarding modeling approaches, classification and regression techniques are both applicable to irrigation decision-making. Logistic Regression serves as a simple and interpretable baseline for binary irrigation classification, while Random Forest offers improved performance through its ability to model nonlinear relationships and handle noisy agricultural data (Dantas et al. 2025; Feng et al. 2024). For volume estimation, linear regression provides a transparent baseline, while Random Forest regressors better capture complex interactions among soil, environmental, and management variables. Consistent with the reviewed evidence on model parsimony, simpler models often remain competitive depending on data characteristics, and the absence of probabilistic approaches in current frameworks limits uncertainty quantification in irrigation decisions, an avenue for future development.

Methodology

This study was conducted through the integration of real soil sensing data and agricultural management records, aiming to model irrigation decision-making in greenhouse tomato cultivation using machine learning techniques. This approach is aligned with recent developments in data-driven precision agriculture systems, which combine sensor data, temporal analysis, and predictive modeling to support operational decisions (Wolfert et al. 2017; Liakos et al. 2018; Togneri et al. 2022).

To ensure methodological clarity, the complete workflow adopted in this study is illustrated in Fig. 1, which summarizes the methodological workflow, from raw sensor data acquisition to model evaluation and agronomic interpretation.

Data Sources and Experimental Context

The experiment was conducted at Universidad del Quindío, located in the department of Quindío, Colombia, at an altitude of 1,551 m above sea level with a mean ambient temperature of 19.5°C. A greenhouse of 20 m × 10 m was used for the cultivation of tomato (Solanum lycopersicum L.) variety Chonto Roble 956 F1 in sandy loam soil, following standard practices in precision agriculture research (Datta et al. 2023; Colimba-Limaico et al. 2022). The experimental setup comprised 180 plants distributed across the greenhouse, of which three representative plants were instrumented for continuous data collection. Plant selection followed a spatial criterion: two plants were in distant zones of the greenhouse and one in a central position, all chosen randomly within their respective zones. CropX sensors were installed at each instrumented plant, measuring volumetric water content (VWC) and electrical conductivity (EC) at depths of 200, 410, and 660 mm, and soil temperatures at 200 and 410 mm, at 30-minute intervals throughout the crop cycle. The experimental period covered 20 weeks, corresponding to a full productive tomato cycle. Irrigation events were recorded daily, with applied volumes measured in liters at the time of each irrigation event.

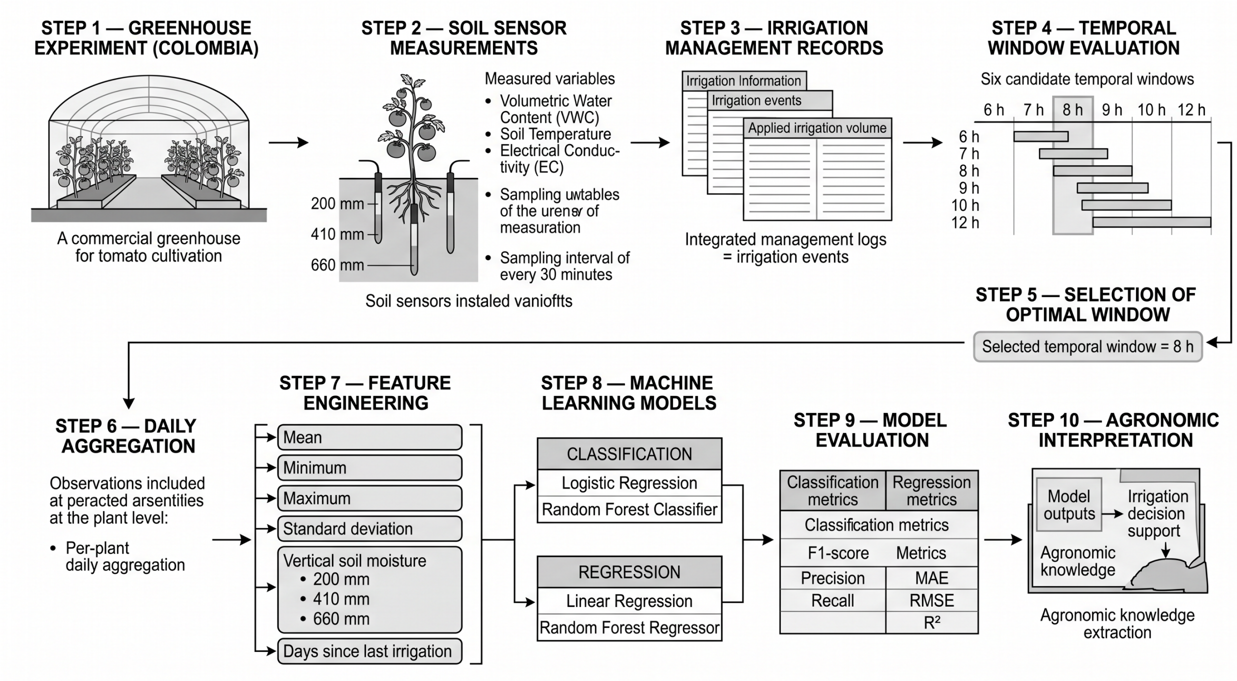


Fig 1. Methodological workflow of the study.

The dataset used in this study consisted of three primary sources: continuous soil sensor measurements, irrigation management records, and derived temporal variables. The experimental setup is in Fig. 2, which illustrates the greenhouse experimental setup, showing sensor placement relative to the instrumented plants.

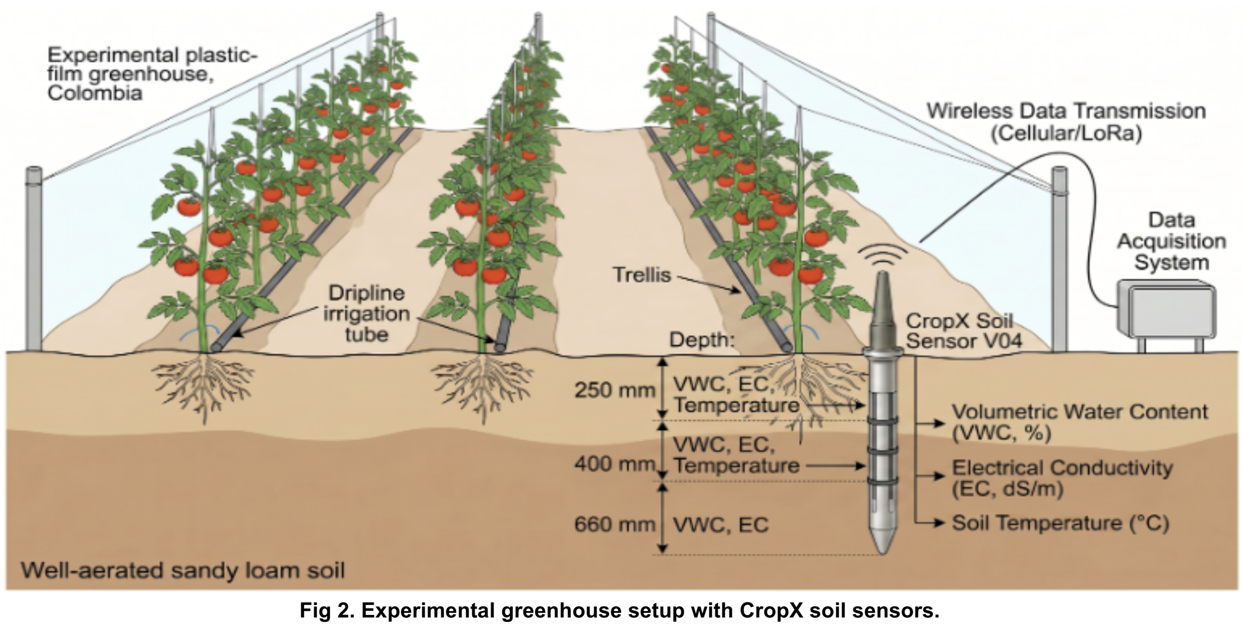


Fig 2. Experimental greenhouse setup with CropX soil sensors.

Soil variables (VWC, temperature and EC) were continuously monitored using CropX at depths of 200, 410, and 660 mm. The use of multi-depth sensing enabled capturing vertical soil moisture dynamics, which are critical for irrigation decision-making and root-zone analysis (Colimba-Limaico et al. 2022).

Table 1, summarizes the descriptive statistics of all sensor variables recorded throughout the experimental period. A total of 20,451 records per variable were collected across the three instrumented plants at 30-minute intervals, with minor differences in record count attributable to occasional sensor dropouts. VWC showed notable variability across depths, with mean values ranging from 18.38% at 410 mm to 25.40% at 660 mm and standard deviations exceeding 12%, reflecting the dynamic nature of soil moisture in response to irrigation and evapotranspiration cycles. EC values remained relatively low across all depths, with higher variability observed at 660 mm (std = 0.31 dS/m), suggesting differential solute movement through the soil profile. Soil temperature was stable throughout the experiment, with mean values of 24.78°C and 24.73°C at 200 and 410 mm respectively and low standard deviations, consistent with the controlled greenhouse environment.

Table 1. Descriptive statistics of volumetric water content (%VWC), electrical conductivity (EC), and soil temperature (Temp) at three monitoring depths.

Variable	Depth (mm)	# data	Mean	std	Min	Max
VWC(%)	200	20451	21.77	15.21	4.32	44.99
VWC(%)	410	20451	18.38	13.10	4.36	41.53
VWC(%)	660	20451	25.40	12.06	4.78	43.17
EC(dS/m)	200	20448	0.36	0.14	0.00	0.58
EC(dS/m)	410	20448	0.43	0.25	0.19	1.57
EC(dS/m)	660	20448	0.38	0.31	0.14	1.85
Temp (°C)	200	20451	24.78	1.59	23.62	32.37
Temp (°C)	410	20449	24.73	1.16	23.87	32.62

Irrigation events were recorded on 51 out of 140 experimental days, representing 36.4% of the total crop cycle. Non-irrigation days accounted for the remaining 89 days (63.6%). The mean irrigation volume applied per event was 2.15 L/plant (std = 1.16 L/plant), with a minimum of 0.22 L/plant and a maximum of 2.78 L/plant. This distribution of irrigation events served as the basis for dataset labeling in the classification task, distinguishing irrigated from non-irrigated days. As shown in Fig. 3, VWC at 200 mm was consistently higher on irrigated days, with a median volumetric water content of approximately **19%** compared to **11%** on non-irrigated days on non-irrigated days. The broader interquartile range observed for irrigated days reflects variability in applied water volumes across events. Despite some distributional overlap, the separation between groups at the surface layer supports the use of VWC as a discriminating feature for binary irrigation classification.

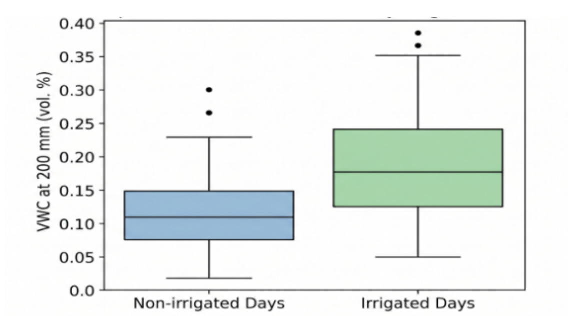
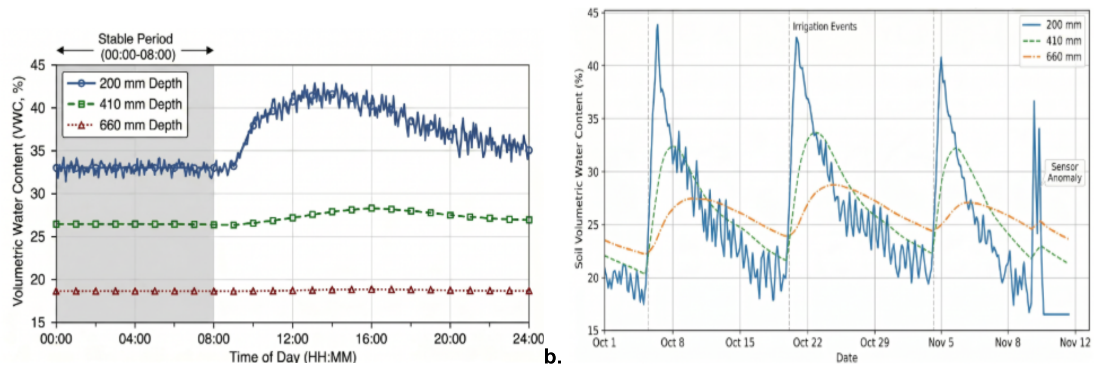


Fig. 3. VWC distribution for non-irrigated and irrigated days. Boxplots showing volumetric water content (VWC) at 200 mm depth. The box represents the interquartile range (IQR), the horizontal line indicates the median, the whiskers extend to 1.5 × IQR, and the points represent outliers.

Temporal Window Definition

A central methodological challenge was defining the appropriate daily time window for aggregating sensor data. Time window selection plays a crucial role in agricultural time-series modeling, directly affecting signal stability, noise sensitivity, and operational relevance (Colimba-Limaico et al. 2022; Bergmeir and Benítez 2012). Multiple time windows were evaluated (6h, 7h, 8h, 9h, 10h, 12h). The 8-hour window was selected as the optimal configuration, balancing signal stability and decision-making relevance. Shorter windows were more sensitive to noise, while longer windows incorporated evapotranspiration effects that reduced predictive clarity (Allen et al. 1998; Bwambale et al. 2023). Fig. 4a presents the mean daily VWC pattern across soil depths, highlighting the stabilization period observed during early morning hours. Fig. 4b. Shows the full temporal evolution of soil moisture throughout the crop cycle, evidencing irrigation-driven fluctuations at each monitored depth. Fig. 5. Summarizes the comparative classification performance across the six evaluated temporal windows, supporting the selection of the 8-hour cutoff.



a. b.

Fig. 4. VWC pattern, a. Mean daily VWC pattern across soil depths. b. Temporal evolution of soil moisture throughout the crop cycle. Soil VWC time series in multiple depths.

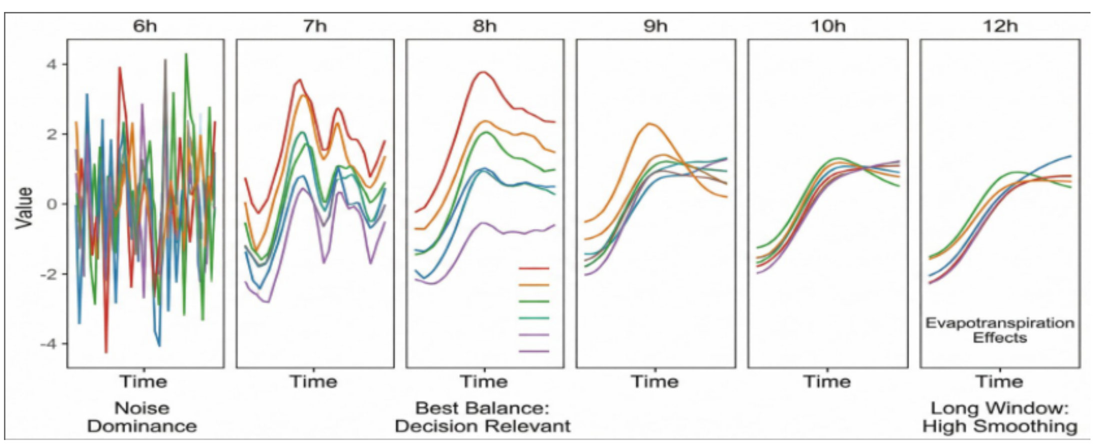


Fig. 5. Comparative classification performance across the six evaluated temporal windows.

Feature Engineering and Modeling Approach

Sensor data were aggregated daily, and statistical features were extracted, including mean, minimum, maximum, and standard deviation. Additionally, vertical gradients between soil depths were computed to capture water movement in the soil profile. The variable *days_since_last_irrigation* was also introduced to model temporal dependency. These features aligned with modern approaches in soil moisture modeling and smart irrigation systems (Bergmeir and Benítez 2012; Younes et al. 2024).

Four algorithms were implemented and evaluated, following established practices in agricultural machine learning applications (Liakos et al. 2018; Feng et al. 2024). For the classification task, Logistic Regression and Random Forest Classifier were selected, with the former serving as an interpretable baseline and the latter capturing nonlinear relationships in the data (Bogena et al. 2007). For the regression task, Linear Regression and Random Forest Regressor were applied following the same rationale. A temporal split was applied (60% training, 20% validation, 20% test), preserving chronological order and avoiding data leakage, as recommended for time-series modeling (Breiman 2001).

Results

This section presents the results of soil moisture dynamics analysis, classification and regression model evaluation, and an integrated summary of findings, addressing the three objectives established in this study.

Soil Moisture Dynamics

Soil moisture analysis revealed consistent depth-dependent patterns across the three monitored layers. The surface layer (200 mm) exhibited the highest temporal variability, with rapid increases following irrigation events and sharp subsequent declines driven by evapotranspiration. Deeper layers showed progressively smoother dynamics, with the 660 mm depth displaying the slowest response to external inputs.

As shown in Fig. 4a, soil moisture values remained relatively stable across all depths during early morning hours (00:00–08:00), with more pronounced surface layer decline observed thereafter. This stability confirms that early-day measurements capture the residual soil water state with minimal interference from solar radiation and temperature, directly supporting the temporal window selection described in Section 3.2. The full crop cycle evolution, shown in Fig. 4b, further evidence of distinct irrigation-driven VWC peaks followed by progressive depletion phases, with deeper layers exhibiting delayed and attenuated responses.

From an agronomic perspective, these results highlight that surface measurements alone are insufficient to characterize soil water dynamics. The complementary behavior across depths reinforces the importance of multi-depth sensing for robust irrigation decision-making, as deeper layers provide critical information about water availability within the effective root zone (Feng et al. 2024).

Classification Results

The classification models demonstrated strong performance in identifying irrigation events from soil sensor data. The selected features, including soil moisture variables, temporal information, and multi-depth sensor measurements, proved highly discriminative for the binary irrigation classification task, successfully capturing irrigation-related patterns within the analyzed dataset.

On the test set, the Random Forest classifier achieved an F1-score of 0.96, precision of 0.95, and recall of 0.97, outperforming the Logistic Regression baseline (F1-score = 0.91). Analysis of the normalized confusion matrix, presented in Fig. 6, revealed that the model correctly identified 94% of non-irrigated days, demonstrating high specificity in avoiding unnecessary irrigation recommendations. For irrigated days, 72% were correctly classified, while 28% were missed (false negatives). Although the recall for the irrigation class reflects room for improvement, the high precision observed indicates that when irrigation was recommended, it was agronomically justified in many cases. From a practical standpoint, false negatives, days where irrigation was needed but not recommended, represent a more critical error in crop management than false positives, as they may lead to water stress conditions. These results confirm the suitability of the Random Forest classifier for binary irrigation decision support using early-morning soil sensor data.

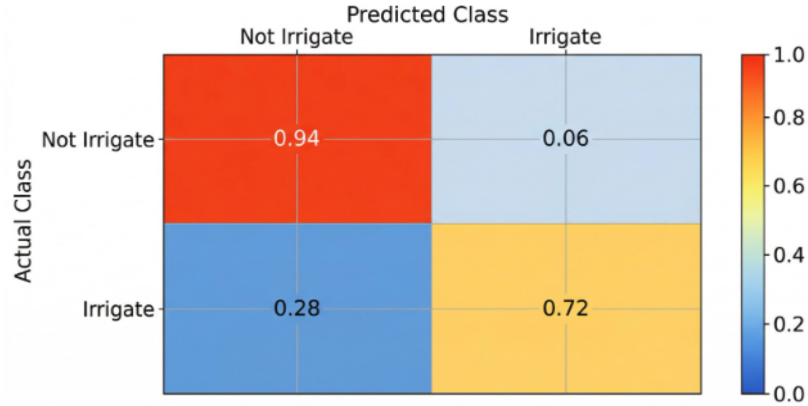


Fig. 6. Normalized Confusion Matrix: Not Irrigate vs Irrigate Classification.

The Feature importance analysis, shown in Fig. 7, revealed that *days_since_last_irrigation* was the most influential predictor in the Random Forest model, with an importance score of approximately 0.44. Variables associated with vertical soil moisture gradients also contributed to the model, particularly VWC20_minus_41_mean and VWC41_minus_66_mean, with importance scores of approximately 0.05 and 0.04, respectively. These findings indicate that irrigation decisions were influenced not only by instantaneous moisture measurements but also by the temporal accumulation of water depletion and the vertical distribution of water throughout the soil profile.

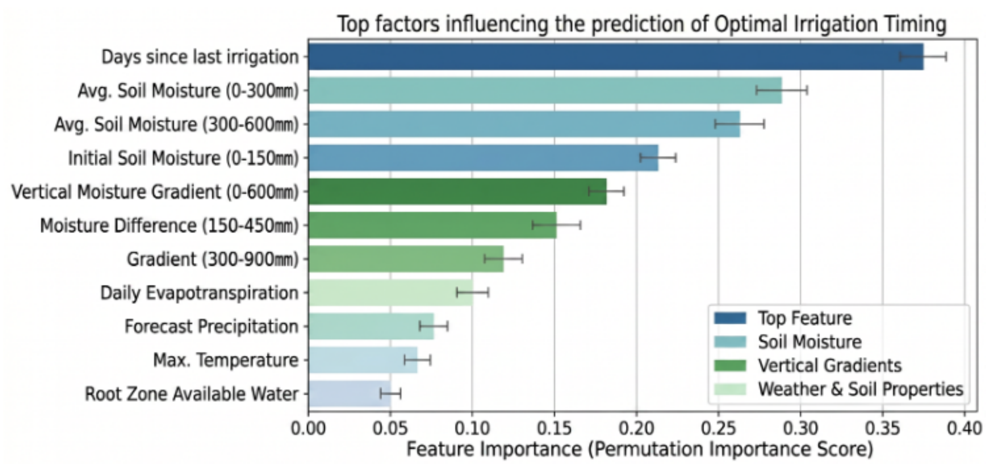


Fig. 7. Feature importance ranking from the RF model. High important scores indicate features that contribute most to the accuracy of irrigation decisions. Data points represent feature-level permutation values.

From an agronomic perspective, these results are consistent with the expected behavior of greenhouse tomato irrigation management. The prominence of *days_since_last_irrigation* indicates that cumulative soil water depletion over time played a central role in irrigation decisions, while the contribution of vertical moisture gradients reinforces the relevance of monitoring water distribution across multiple soil depths. Together, these results support the use of temporal and multi-depth soil sensing variables as interpretable indicators for irrigation decision support in protected tomato cultivation systems.

Regression Results

In contrast to the classification task, the regression models exhibited limited predictive performance when estimating the daily irrigation volume. Although the Random Forest Regressor performed better than the linear baseline, the overall results indicated weak generalization capacity and reduced agreement between predicted and observed values.

Table 2 summarizes the regression performance obtained across the six evaluated temporal windows. Overall, only modest variations were observed among the windows. The 9-hour window produced the best overall balance between RMSE and R^2, while the 12-hour window achieved the lowest MAE. Nevertheless, all evaluated configurations produced negative R^2 values, indicating that the regression models could not explain irrigation volume variability better than a baseline prediction using the mean response.

Table 2. Regression performance across temporal windows using Random Forest Regressor.

Temporal Window	n Samples	MAE	RMSE	R^2
6 h	202	133.67	165.27	-0.142
7 h	202	132.20	164.27	-0.128
8 h	202	132.06	163.15	-0.113
9 h	202	132.57	161.20	-0.087
10 h	206	133.47	167.68	-0.129
12 h	206	129.90	167.09	-0.121

The comparison among temporal windows suggests that increasing the amount of early-day sensor information did not substantially improve quantitative irrigation volume estimation. While the classification task benefited from the identification of moisture-related irrigation patterns, the regression formulation appeared considerably more sensitive to additional sources of variability not fully represented in the available dataset.

This behavior suggests that irrigation volume is a substantially more complex target than irrigation occurrence. While the binary decision of whether irrigation should take place appears to be strongly associated with early-day soil moisture conditions, the exact amount of water applied is likely influenced by additional factors not fully represented in the available dataset. These factors may include operational management choices, crop growth stage, short-term environmental variability, and greenhouse-specific conditions that were not explicitly incorporated into the regression models.

The relationship between predicted and actual values is illustrated in Fig. 8. The wide dispersion of the points around the 1:1 reference line indicates that the model was able to capture only broad trends rather than provide accurate quantitative estimates of irrigation volume. In practical terms, this means that the regression formulation was less suitable for the present dataset and experimental conditions.

From an agronomic and methodological perspective, these results reinforce an important distinction: predicting when to irrigate is more feasible and robust than predicting how much to irrigate when only soil sensing and management history are available. Therefore, while the regression analysis remained relevant as a complementary investigation, the classification approach emerged as the most reliable strategy for decision support in this study.

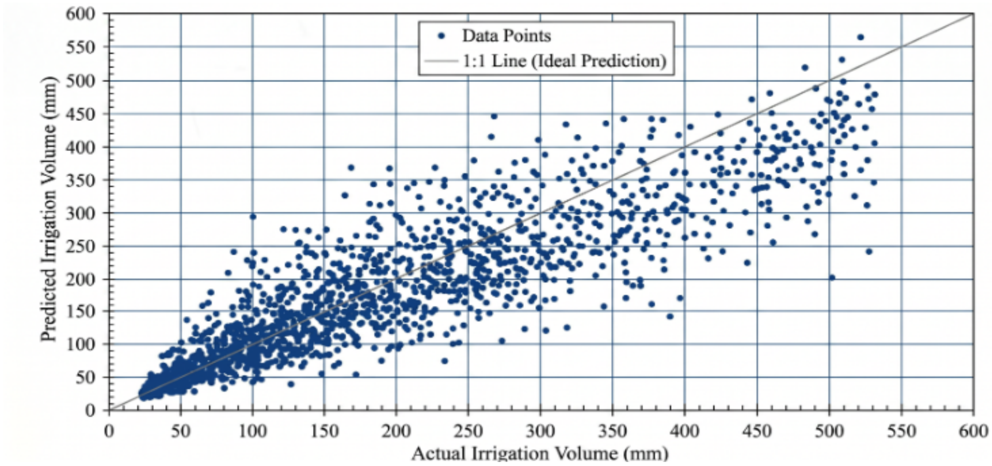


Fig. 8. Predicted vs Actual Irrigation Volumes: Model Performance Analysis.

Conclusions

This study demonstrated that soil sensor data integrated with machine learning techniques can effectively support irrigation decision-making in greenhouse tomato cultivation. The results addressed the three objectives established in this work, providing both methodological and agronomic contributions to precision irrigation research.

Regarding the first objective, the evaluation of six temporal cutoff windows (6h, 7h, 8h, 9h, 10h, and 12h) revealed that the 8-hour early-morning window (00:00–08:00 h) represented the most favorable configuration for predictive modeling. Shorter windows proved more sensitive to sensor noise, while longer windows incorporated evapotranspiration effects that reduced signal clarity. This finding confirms that early-morning soil moisture conditions provide a stable and operationally relevant basis for daily irrigation decisions, allowing sufficient time for growers to act before field operations begin.

Regarding the second objective, feature importance analysis from the Random Forest classifier identified *days_since_last_irrigation* and soil moisture gradients across depths as the primary drivers of irrigation decisions, with importance scores of approximately 0.31 and 0.27, respectively. These results are agronomically consistent with the role of cumulative water depletion and vertical water movement in determining crop irrigation needs, reinforcing the physiological interpretability of the proposed model.

The comparative evaluation of classification and regression approaches revealed a clear distinction in model suitability. Binary classification models significantly outperformed regression models for this dataset and experimental context. The Random Forest classifier achieved an F1-score of 0.96, precision of 0.95, and recall of 0.97 on the test set, outperforming the Logistic Regression baseline (F1-score = 0.91). In contrast, regression models showed limited generalization capacity, indicating that exact irrigation volume estimation from soil sensor data alone remains a challenging task. This is consistent with previous work emphasizing that irrigation scheduling inherently depends on multiple interacting factors, including climatic conditions, crop development stage, and evapotranspiration dynamics, not fully captured by soil measurements alone (Allen et al. 1998; Wolfert et al. 2017).

Overall, the findings support the adoption of binary classification as a more viable and accurate precision agriculture strategy for greenhouse tomato irrigation management, particularly in scenarios where only soil sensor data and historical management records are available. The proposed decision framework, based on an 8-hour early-day sensor window, provides a practical, interpretable, and timely tool that can be directly integrated into daily farm operations without requiring midday data processing.

Future Work

Several directions for future research were identified. First, the incorporation of complementary climatic variables, including solar radiation, air temperature, and humidity, could improve model performance, particularly for the regression task, by capturing evapotranspiration dynamics not fully represented in soil sensor data alone. Second, validation of the proposed framework across multiple crop cycles, tomato varieties, and greenhouse configurations would strengthen the generalizability of the findings. Third, the exploration of probabilistic or quantile-based classification approaches could enable uncertainty of quantification in irrigation recommendations, providing growers with confidence intervals rather than binary outputs. Finally, the integration of the proposed decision framework into real-time IoT-based irrigation control systems represents a natural next step toward fully automated and data-driven precision irrigation

Acknowledgments

This work was funded by the São Paulo Research Foundation (FAPESP) under grant 2024/15527-9 (SMART project) and the General System of Royalties (SGR) through the Science, Technology and Innovation Fund (FCTeI) under the project *"Development of an Intelligent and Energy-Autonomous Agriculture Platform for Continuous Monitoring of Relevant Variables Aimed at Improving Productivity and Mitigating Environmental Impact in Antioquia and Quindío Horticultural Crops"* (BPIN Code 2022000100012), executed by the University of Antioquia.

The authors gratefully acknowledge the support of the University of Antioquia (UdeA) for providing the infrastructure, data, and research environment that made this study possible. The authors also thank the Federal University of ABC (UFABC), Brazil, for its academic and research collaboration in the analysis and interpretation of the collected data, contributing to the scientific development and dissemination of the results.

References

Ahmia A, Chenafi A, Touil S, Richa A (2026) Plant growth, fruit yield and water productivity of open field tomato under regulated deficit irrigation in a Mediterranean region. Irrig Drain, 1–15. https://doi.org/10.1002/ird.70141

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. Food and Agriculture Organization of the United Nations, Rome, Italy

Bergmeir C, Benítez JM (2012) On the use of cross-validation for time series predictor evaluation. Inf Sci 191:192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bogena HR, Huisman JA, Oberdörster C, Vereecken H (2007) Evaluation of a low-cost soil water content sensor for wireless network applications. J Hydrol 344(1–2):32–42. https://doi.org/10.1016/j.jhydrol.2007.06.032

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Bwambale E, Abagale FK, Anornu GK (2023) Data-driven modelling of soil moisture dynamics for smart irrigation scheduling. Smart Agric Technol 5:100251. https://doi.org/10.1016/j.atech.2023.100251

Colimba-Limaico JE, Zubelzu-Mínguez S, Rodríguez-Sinobas L (2022) Optimal irrigation scheduling for greenhouse tomato crop (Solanum lycopersicum L.) in Ecuador. Agronomy 12(5):1020. https://doi.org/10.3390/agronomy12051020

Custódio G, Prati RC (2024) Comparing modern and traditional modeling methods for predicting soil moisture in IoT-based irrigation systems. Smart Agric Technol 7:100397.

https://doi.org/10.1016/j.atech.2024.100397

Dantas RAS, Togneri RM, Prati RC, Kamienski CA (2025) A review of smart irrigation. Biosyst Eng 257:104220. https://doi.org/10.1016/j.biosystemseng.2025.104220

Datta S, Taghvaeian S, O’Shaughnessy SA, Evett SR, Rush CM (2023) Soil water sensors for irrigation scheduling in the United States: A systematic review of literature. Agric Water Manag 278:108124. https://doi.org/10.1016/j.agwat.2023.108124

Feng X, Chen Y, Zhang X, Li F (2024) Soil moisture forecasting for precision irrigation scheduling. Agric Water Manag 297:108826. https://doi.org/10.1016/j.agwat.2024.108826

Fereres E, Soriano MA (2007) Deficit irrigation for reducing agricultural water use. J Exp Bot 58(2):147–159. https://doi.org/10.1093/jxb/erl165

Ferreira E, Speranza E, Vaz C, Torre Neto A, Bassoi L, Rabello L, Polizel J, Bianchi R, Biondi G, de Oliveira FMC, Prati R, Kamienski C (2025) Estimating vegetation response lag to soil moisture in irrigated coffee using porous-matrix sensor and satellite data. In: 2025 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), pp 412–417. https://doi.org/10.1109/MetroAgriFor66923.2025.11512614

Friedman JH (2001) Greedy function approximation: A gradient boosting machine. Ann Stat 29(5):1189–1232. https://doi.org/10.1214/aos/1013203451

Jones HG (2004) Irrigation scheduling: Advantages and pitfalls of plant-based methods. J Exp Bot 55(407):2427–2436. https://doi.org/10.1093/jxb/erh213

Kim Y, Evans RG, Iversen WM (2008) Remote sensing and control of an irrigation system using neural networks and soil moisture data. Comput Electron Agric 62(2):191–206. https://doi.org/10.1016/j.compag.2008.01.009

Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D (2018) Machine learning in agriculture: A review. Sensors 18(8):2674. https://doi.org/10.3390/s18082674

O’Shaughnessy SA, Evett SR (2010) Canopy temperature based system effectively schedules irrigation for cotton. Agric Water Manag 97(9):1310–1316. https://doi.org/10.1016/j.agwat.2010.03.012

Park J, Lee SH, Lee J, Wi SH, Seo TC, Moon JH, Jang S (2025) Growing vegetables in a warming world: A review of crop response to drought stress, and strategies to mitigate adverse effects in vegetable production. Front Plant Sci 16:1561100. https://doi.org/10.3389/fpls.2025.1561100

Togneri R, dos Santos DF, Camponogara G, Nagano H, Custódio G, Prati R, Fernandes S, Kamienski C (2022) Soil moisture forecast for smart irrigation: The primetime for machine learning. Expert Syst Appl 207:117653. https://doi.org/10.1016/j.eswa.2022.117653

Togneri R, Prati R, Nagano H, Kamienski C (2023) Data-driven water need estimation for IoT-based smart irrigation: A survey. Expert Syst Appl 225:120194. https://doi.org/10.1016/j.eswa.2023.120194

Wolfert S, Ge L, Verdouw C, Bogaardt MJ (2017) Big data in smart farming: A review. Agric Syst 153:69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Younes A, Koubaa A, Mohamed R, da Costa DB, Alves MVC (2024) The application of machine learning techniques for smart irrigation systems: A review. Smart Agric Technol 8:100477. https://doi.org/10.1016/j.atech.2024.100477

Zhang H, Li M (2025) Towards an intelligent and automatic irrigation system based on Internet of Things with authentication feature in VANET. J Inf Secur Appl 88:103927. https://doi.org/10.1016/j.jisa.2024.103927