



Research article

An IoT-driven predictive analytics framework for dynamic irrigation optimization in tomato cultivation

Maung Maung Htwe^{1,2}, Lachezar Filchev¹, Ekaterina Batchvarova¹ and Sandra Jardim^{2,*}

¹ Climate, Atmosphere and Water Research Institute – Bulgarian Academy of Sciences

² Smart Cities Research Center, Polytechnic University of Tomar, Portugal

* **Correspondence:** Email: sandra.jardim@ipt.pt.

Abstract: Addressing the critical challenges of global food security and water scarcity, we introduced an Internet of Things (IoT)-driven predictive analytics framework for dynamic irrigation optimization in tomato cultivation. Our primary objective was to develop a robust model that accurately estimates daily water requirements, with the aim of minimizing water consumption while concurrently maintaining optimal soil health. This framework leverages a uniquely comprehensive, experimentally controlled dataset from multi-sensor IoT deployments, covering environmental conditions (air temperature, humidity, CO₂, pressure) and key soil parameters (humidity, temperature, electrical conductivity). Through a rigorous data preprocessing pipeline and a tailored feature engineering approach, critical temporal patterns, inter-variable relationships, and insights from agronomic indicators like Growing Degree Days (GDD), alongside other dynamically derived features, were extracted. A two-part eXtreme Gradient Boosting (XGBoost) regression model, combining classification and regression, was developed and validated to precisely predict the daily water volume needed per hectare. The innovation of this work lies in its ability to harness complex historical IoT data to build a sophisticated intelligence layer for irrigation scheduling. By demonstrating the model's accuracy in identifying optimal water levels under varying conditions and achieving significant water savings in a simulated dynamic optimization, this research provides foundational data-driven insights that can inform highly effective precision irrigation strategies. The model achieved a high R^2 of 0.9476 and yielded a potential water saving of 50.84% in a simulated dynamic optimization compared to the model's raw prediction. Such intelligence empowers farmers to significantly reduce water waste and prevent harmful over-irrigation, leading to more sustainable and efficient smart agriculture, which is critical for enhancing crop resilience and resource efficiency in a changing climate.

Keywords: Internet of Things (IoT); predictive analytics; irrigation optimization; water management; soil health; machine learning; XGBoost Model

1. Introduction

The agricultural sector faces unprecedented challenges driven by a rapidly expanding global population and escalating water scarcity [1]. Traditional irrigation methods are often inefficient, leading to over-irrigation (water waste, nutrient leaching, soil degradation) or under-irrigation (crop stress, reduced yields) [2]. The dynamic nature of environmental factors, soil variability, and evolving crop water requirements further complicates accurate water management.

In this research, we address these challenges by developing an IoT-powered predictive analytics framework for dynamic irrigation optimization in tomato cultivation. For the framework, we aim to accurately estimate daily water requirements, minimizing consumption while maintaining optimal soil health. This work is differentiated from other descriptive studies [3] by proposing a prescriptive framework that dynamically predicts and optimizes water application based on continuous data, moving beyond static regime comparisons to real-time adaptive water management. The fundamental research gap addressed is the critical deficiency in current agricultural modeling, which often relies on descriptive or reactive threshold-based systems, failing to provide the true predictive and prescriptive intelligence required for dynamic water management. The high-quality data generated by IoT sensors offer an opportunity to move beyond these static models, enabling the development of predictive intelligence that can adapt in real-time to the complex, non-linear interplay of environmental, soil, and biological factors [4,5]. This shift from merely describing water needs to dynamically predicting the optimal irrigation volume is essential for maximizing Water Use Efficiency (WUE) and achieving sustainable precision agriculture [6,7]. Our key novelties and contributions include:

- Integration of multi-sensor IoT data (environmental, soil, water meter) with advanced data preprocessing to ensure high-quality training data [8,9].
- Comprehensive feature engineering, incorporating key agronomic indicators like Growing Degree Days (GDD) and Soil Moisture Deficit (SMD), and dynamic temporal features to capture complex system dynamics.
- Development and rigorous validation of a robust two-part XGBoost model (classification + regression) for precise daily water volume prediction, explicitly handling the sparse nature of irrigation events for enhanced accuracy.
- Demonstration of a dynamic optimization paradigm that effectively balances water conservation and soil health constraints, achieving a potential water saving of 50.84% in a simulated test environment.
- A feature-parsimonious approach that attains competitive accuracy compared to complex deep learning models, making the solution practical and scalable for resource-constrained smart agriculture [10,11].

The integration of IoT and Machine Learning (ML) has transformed agricultural practices,

ushering in an era of precision farming [8]. IoT systems enable real-time monitoring of environmental conditions, soil characteristics, and crop health [12]. The necessity for robust, end-to-end IoT architectures capable of handling heterogeneous data streams and providing real-time data analysis for proactive decision-making is a persistent theme in research [5,7]. The deployment of reliable sensor networks, including soil moisture sensors (e.g., capacitive and TDR) and environmental sensors (e.g., pyranometers for solar radiation and anemometers for wind speed), is fundamental [9,13]. However, challenges remain in energy consumption, e-waste, and sensor durability in harsh environments [9]. Work continues to highlight the need for more secure, robust, and scalable end-to-end IoT architectures that seamlessly integrate stream processing with ML pipelines for real-time analytics [5]. ML algorithms are crucial for optimizing agricultural water management [14]. ML techniques [15], including ensemble methods like XGBoost [16], leverage diverse datasets to predict crop water requirements and optimize irrigation schedules [17–19]. Beyond XGBoost, other machine learning algorithms like Support Vector Machines (SVMs) [20], Artificial Neural Networks (ANNs) [21], and Random Forests [22] have also been explored for their potential in predicting crop water requirements, often showing robust performance in diverse agricultural contexts. The superiority of XGBoost lies in its efficiency, accuracy, and scalability in handling complex, high-dimensional agricultural data, often outperforming other ensemble methods like Random Forest in generalizing to unseen non-linear patterns [23,24].

IoT sensors and ML also revolutionize real-time soil health monitoring, providing continuous insights into soil moisture, temperature, and electrical conductivity (EC) [11,21,22]. Predictive analytics can identify soil degradation issues like salinity, enabling targeted irrigation that minimizes waste and promotes sustainability. Accurate estimation of crop evapotranspiration (ET) is fundamental to efficient irrigation scheduling, with models like Penman-Monteith and FAO-56 widely adopted and often integrated with sensor data for improved precision [25].

Digital Twin (DT) technology offers advanced potential for agriculture, creating virtual representations of physical farm systems updated with real-time data [26]. Predictive frameworks like ours can serve as foundational components for DTs, enabling advanced simulations and proactive management strategies [27,28]. Furthermore, advanced optimization algorithms such as fuzzy logic controllers [29] and genetic algorithms [30] have been employed to develop adaptive irrigation schedules that respond to real-time environmental changes, moving toward sustainable precision agroecology [31]. These approaches contribute to a more holistic and intelligent agricultural management system.

2. Materials and methods

In this section, we delineate the experimental setup, characteristics of the acquired datasets, and the comprehensive technical pipeline employed to develop the IoT-powered predictive analytics framework for dynamic irrigation optimization.

2.1. Experimental data acquisition and description

The foundation of this research is a comprehensive historical dataset titled "IoT-based Data

Collection in a Tomato Cultivation Under Different Irrigation Regimes" [3]. This dataset, publicly available through Mendeley Data, was collected from a tomato cultivation (specifically, *Solanum lycopersicum* L. cv. HEINZ 1350) under experimentally controlled conditions in Parma, Italy. The experiment was carried out from June 29, 2023 to September 13, 2023. The dataset comprises multi-sensor IoT deployments, yielding a rich, time-series historical record. The multi-sensor deployment, operating via the LoRaWAN communication protocol, recorded data from Class A end nodes registered on The Things Network (TTN). All devices were configured to transmit data at a 10-minute frequency, ensuring granular coverage of environmental and soil dynamics [9]. Key sensor specifications are as follows:

- Environmental Sensor (Milesight EM500-CO₂): Measures CO₂, air humidity (%RH), air temperature (°C), and barometric pressure (hPa).
- Soil Sensor (Milesight EM500-SMTC): Measures soil moisture (%RH), soil temperature (°C), and electrical conductivity (μS/cm). Sensors were positioned at a 20 cm depth for each experimental line. Calibration for the Electrical Conductivity (EC) sensor was performed using standard KCl solutions at 25°C to correct for temperature influences, while Soil Moisture sensors were validated against gravimetric methods in the field.
- Water Meter (Talkpool OY1310): Records cumulative water applied per irrigation line (liters).

The dataset includes environmental sensor data (time-series measurements of CO₂ concentration, ambient humidity, atmospheric pressure, and air temperature), soil sensor data (detailed soil parameters including electrical conductivity (EC), soil humidity (moisture content), and soil temperature) with a line field distinguishing different experimental plot and water meter data (records of cumulative water applied with line identifiers).

2.2. Data preprocessing pipeline

The raw IoT sensor data requires robust preprocessing to ensure data quality, consistency, and suitability for machine learning model development [32]. The preprocessing pipeline comprises four critical stages: Data cleaning and structure validation, outlier detection and treatment, missing data imputation, and temporal synchronization and aggregation. The core dataset utilizes multi-sensor IoT deployments operating via LoRaWAN for long-range, low-power communication [1]. Data from environmental and soil sensors are collected at a high frequency and synchronized and aggregated to a unified daily frequency for predictive modeling. Critical to data quality, the pipeline includes Electrical Conductivity (EC) conversion from μS/cm to mS/cm and temperature correction to EC₂₅ (EC at 25°C) ensuring scientific reproducibility [8]. The system design is focused on connecting to a Digital Twin (DT) architecture for enhanced water distribution and real-time fault detection in the irrigation network [2].

2.2.1. Data cleaning and outlier treatment

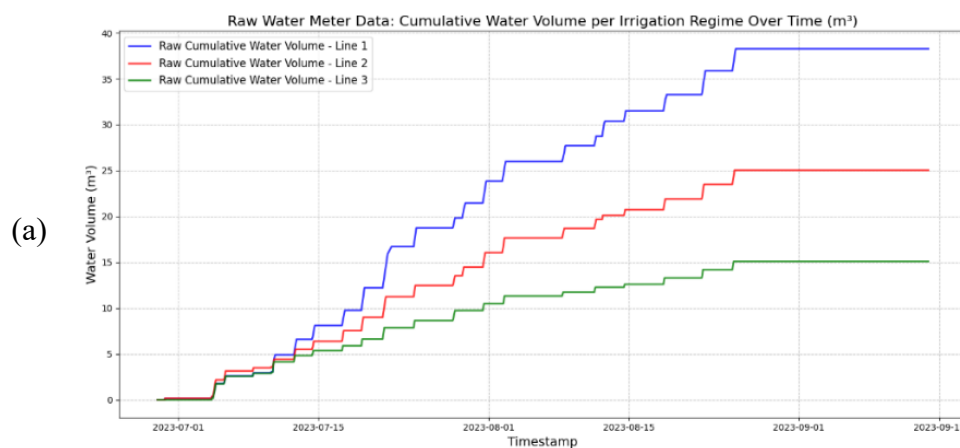
Initial data cleaning involved removing non-informative columns (e.g., 'Battery', fields with >40%

missing data). Timestamp fields were converted to datetime objects with time zone-aware formatting (UTC+1 for Parma, Italy location). Duplicate records were identified using composite keys (timestamp, device_id and line) and removed, resulting in elimination of 127 duplicate entries (1.8% of the raw data). A comprehensive multi-method outlier detection strategy was implemented to handle noise and sensor drift [13]:

- a. Statistical Outlier Detection: Using the Z-score method (exceeding ± 3 standard deviations) and the Interquartile Range (IQR) method [33].
- b. Physically Based Range Validation: Enforcing sensor-specific bounds (e.g., Soil Moisture: 15–45% VWC; Air Temperature: 10–40 °C).
- c. Treatment: Detected outliers were managed by either removal and linear interpolation (extreme outliers) or Winsorization (moderate outliers) to preserve the temporal integrity of the time series.

2.2.2. Missing data imputation and aggregation

Missing data ($\sim 3.2\%$ of the dataset) was addressed using a hierarchical imputation strategy: A forward-fill method was used for short gaps (≤ 2 hours), and linear interpolation was applied for moderate gaps (2–6 hours). A critical step was the synchronization and aggregation of time-series data to a daily frequency. Environmental and soil data were aggregated into daily statistics (mean, max, min, std dev). The cumulative water volume was transformed into daily water volume applied by calculating the difference between consecutive 24-hour readings, normalized to $\text{m}^3/\text{ha}/\text{day}$. The total cumulative water volume for the plot over the data collection period was $1740.67 \text{ m}^3/\text{ha}$.



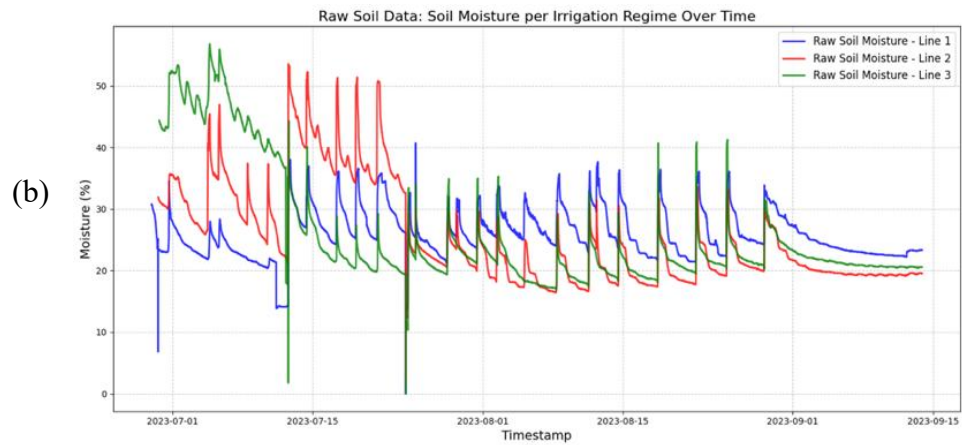


Figure 1. Raw IoT sensor data. (a) Soil moisture content per irrigation line over time, and (b) cumulative water volume per irrigation line over time (m^3).



Figure 2. Unified daily water and soil parameters. (a) Unified daily mean soil moisture content over time (%), and (b) unified total daily water volume over time (m^3/day).

2.3. Feature engineering for predictive modeling

Feature engineering extracts meaningful patterns. Growing Degree Days (GDD) is a biologically meaningful thermal time metric that quantifies heat accumulation essential for crop development and growth progression. Unlike simple chronological time, GDD captures the physiological response of plants to temperature variations, directly correlating with phenological stages such as flowering, fruiting, and maturation. This thermal accumulation approach has been extensively validated in crop modeling and irrigation scheduling, enhancing the accuracy of developmental predictions beyond calendar days or fixed time intervals. By integrating GDD as a predictor, the model accounts for temperature-dependent growth dynamics critical for precise water demand estimation in tomato cultivation.

Temporal patterns, including lagged values and rolling window statistics of environmental and soil parameters, are incorporated to capture delayed crop responses and persistent weather effects. These features reflect cumulative influences over biologically relevant time scales, such as water uptake lag due to soil moisture retention and temperature-driven evapotranspiration changes. Empirical studies have demonstrated the importance of such temporal dynamics in improving irrigation forecast fidelity and resilience to transient climatic fluctuations. Thus, the inclusion of temporal pattern features enables the machine learning model to effectively learn complex time-dependent relationships driving crop water requirements. Corrected Soil Parameters (EC_{25} , SMD) are crucial for maintaining soil health by providing inputs that directly relate to potential salinity stress and water stress, respectively. Temporal features (day of year, week of year, day of week, month etc.,) capture seasonal and weekly patterns.

The Growing Degree Days (GDD) metric is calculated using a modified average air temperature approach, incorporating a base temperature (T_{base}) and a temperature cutoff (T_{cutoff}) specific to the crop, which, for tomatoes, are 10°C and 32°C , respectively [2]. This method accounts for the physiological limits of crop growth imposed by temperature. The formula for GDD is defined by the following piecewise function:

$$GDD = \begin{cases} 0 & \text{if } T_{avg} \leq T_{base} \\ T_{avg} - T_{base} & \text{if } T_{base} < T_{avg} < T_{cutoff} \\ T_{cutoff} - T_{base} & \text{if } T_{avg} \geq T_{cutoff} \end{cases}, \quad (1)$$

where T_{avg} is the daily mean air temperature.

Additional features include lagged values (1, 3, and 7 days) and rolling window statistics (3 and 7-day means and standard deviations) for key environmental, soil, and water volume parameters. These features capture temporal dependencies and trends. Moreover, maintaining optimal soil health is paramount for sustainable agriculture. This framework incorporates key soil parameters, such as soil moisture content and electrical conductivity (EC). It measures at an arbitrary temperature (t), denoted as EC_t , which is normalized to a standard temperature of 25°C (EC_{25}) to facilitate comparison across environmental conditions. This correction was performed using a temperature coefficient, f_t :

$$EC_{25} = \frac{EC_t}{f_t}, \quad (2)$$

In this equation, EC_t is the electrical conductivity measured at temperature t (in mS/cm), and the temperature coefficient f_t is determined by the following third-order polynomial:

$$f_t = 1 - 0.020346(T_{adj}) + 0.003822(T_{adj})^2 - 0.000555(T_{adj})^3,$$

The adjusted temperature variable T_{adj} is calculated as the standardized difference from 25°C:

$$T_{adj} = \frac{T \text{ (in } ^\circ\text{C)} - 25}{10},$$

The Soil Moisture Deficit (SMD) provides a quantitative measure of the deviation of the soil's water content from a predetermined target soil humidity (e.g., 25% for tomatoes) and concurrently incorporates the influence of electrical conductivity. This interaction is modeled linearly:

$$SMD = (T_{SH} - SM_{mean}) \times C_{SM} + EC_{mean} \times C_{EC}, \quad (3)$$

where T_{SH} is the target soil humidity (e.g., 25°C), SM_{mean} is the mean soil moisture, EC_{mean} is the mean electrical conductivity, and C_{SM} and C_{EC} are empirical weighting coefficients with values of 1.5 and 0.15, respectively. To account for potential synergistic effects between the two major environmental factors, an interaction feature is defined as the product of the mean soil moisture and the mean air temperature.

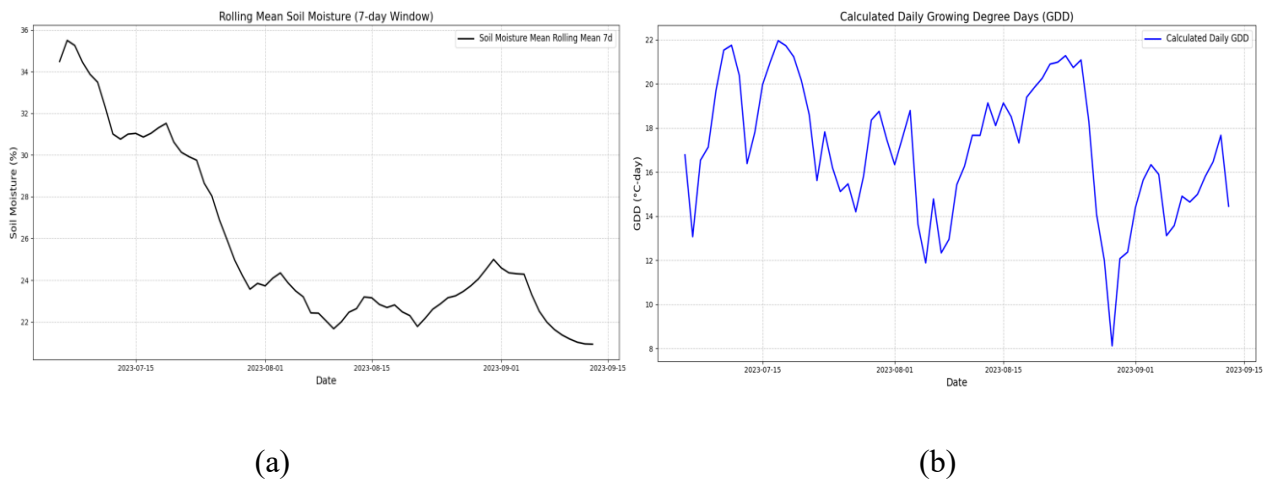


Figure 3. Processed Environmental and Agronomic Parameters. (a) 7-Day Rolling Mean Soil Moisture Content (%), and (b) Calculated Daily Growing Degree Days (GDD) (°C-day).

2.4. Predictive model development: Two-part XGBoost model

The core of the predictive analytics framework is the machine learning model responsible for estimating daily water requirements. An eXtreme Gradient Boosting (XGBoost) model was selected for this task due to its demonstrated efficiency, accuracy, and scalability in handling large datasets [34,35]. To ensure the integrity of the time-series analysis and prevent data leakage, the dataset was split chronologically. The training set spanned the period from July 05, 2023 to August 13, 2023 (40 observations), and the test set covered August 14, 2023 to September 13, 2023 (31 observations). This chronological split ensured the model was evaluated on future, unseen environmental and phenological conditions. The two-part modeling approach was employed to handle the sparse nature of irrigation

events (only 22 non-zero events in 71 days):

- **Part 1 (Classification Model):** An XGBoost Classifier predicts whether water was applied (binary: 0 or 1) on a given day. This step is critical for handling data sparsity, preventing the regression model from being biased toward predicting zero values.
- **Part 2 (Regression Model):** An XGBoost Regressor predicts the exact amount of water ($\text{m}^3/\text{ha}/\text{day}$) only for days where the classifier predicted water application. This focuses the regression on the magnitude of the water requirement, significantly enhancing its volumetric accuracy.

Hyperparameter optimization was conducted using GridSearchCV with TimeSeriesSplit cross-validation. The best-performing hyperparameters for the models were:

- XGBoost Classifier: $n_estimators = 100$, $learning_rate = 0.1$, $max_depth = 3$, $subsample = 0.9$, $colsample_bytree = 0.7$.
- XGBoost Regressor: $n_estimators = 200$, $learning_rate = 0.05$, $max_depth = 7$, $min_child_weight = 1$, $subsample = 0.9$, $colsample_bytree = 0.7$.

The following pseudocode outlines the sequential operation of the developed framework:

```
// -----
// Phase 1: Predictive Model Training (Offline)
// -----
FUNCTION TRAIN_PREDICTIVE_MODEL(Historical_IoT_Data):
    Clean_Data = Preprocess(Historical_IoT_Data)
    Features = Feature_Engineer(Clean_Data) // Includes GDD, SMD, Lagged Values
    Train_Data, Test_Data = Split_Chronologically(Features)

    // Part 1: Classification Model Training
    X_Clf = Features (Non-Water_Columns)
    Y_Clf = Binary_Label (Water_Volume > 0)
    Classifier = XGBoostClassifier()
    Hyperparameter_Tune(Classifier, X_Clf, Y_Clf, GridSearchCV, TimeSeriesSplit)
    Trained_Classifier = Classifier.fit(X_Clf, Y_Clf)

    // Part 2: Regression Model Training (Only on Irrigation Days)
    Regression_Data = Features.filter(Water_Volume > 0)
    X_Reg = Regression_Data (Non-Water_Columns)
    Y_Reg = Regression_Data (Water_Volume)
    Regressor = XGBoostRegressor()
    Hyperparameter_Tune(Regressor, X_Reg, Y_Reg, GridSearchCV, TimeSeriesSplit)
    Trained_Regressor = Regressor.fit(X_Reg, Y_Reg)
```

```

RETURN Trained_Classifier, Trained_Regressor

// -----
// Phase 2: Dynamic Irrigation Optimization (Real-Time/Daily)
// -----
FUNCTION DYNAMIC_IRRIGATION_OPTIMIZATION(Trained_Clf, Trained_Reg, Current_IoT_Data, Constraints):
    // A. Prediction Step
    Input_Features = Feature_Engineer(Current_IoT_Data)
    Prediction_Class = Trained_Clf.predict(Input_Features)

    IF Prediction_Class == 1:
        Predicted_Volume = Trained_Reg.predict(Input_Features)
    ELSE:
        Predicted_Volume = 0

    // B. Optimization Step (Prescriptive)
    Objective_Function(Water_Amount) = Water_Amount * 0.05 +
                                        (Water_Amount - Predicted_Volume)^2 * 0.005 +
                                        Moisture_Penalty(Simulated_Soil_Moisture, Expected_Rainfall)

    Constraints(Water_Amount):
        Min_Soil_Moisture_Constraint: Simulated_Soil_Moisture >= 20%
        Max_Soil_Moisture_Constraint: Simulated_Soil_Moisture <= 40%

    // C. Solve for Optimal Water Amount
    Optimal_Water_Applied = Minimize(Objective_Function, Subject_To(Constraints))

RETURN Optimal_Water_Applied (m3/ha/day)

```

2.5. Proposed dynamic irrigation optimization framework

The developed XGBoost predictive model forms the intelligence core of a proposed dynamic irrigation optimization framework. This conceptual framework integrates continuous data flow from IoT sensors with advanced predictive analytics to provide actionable irrigation schedules. The framework operates by continuously collecting real-time environmental and soil data from the IoT sensor network. This data, along with derived agronomic indicators like GDD, serves as input to the trained XGBoost model. The model then precisely predicts the daily water volume required per hectare for optimal tomato cultivation. These daily water volume predictions are subsequently translated into actionable irrigation schedules, specifying the exact amount of water to be applied.

A critical component of this framework is the feedback loop mechanism, where real-time soil moisture and environmental conditions continuously update the model's input. This enables adaptive optimization, ensuring that the irrigation recommendations adjust dynamically to changing conditions,

such as unexpected rainfall or sudden heatwaves. This continuous adaptation is a significant advantage over static or threshold-based irrigation systems, which may not respond effectively to unforeseen environmental shifts. By actively monitoring and maintaining optimal soil health through integrated sensor data, the framework simultaneously minimizes water consumption and prevents detrimental conditions like over-irrigation-induced salinity or nutrient leaching.

The framework integrates the predictive model with a conceptual feedback loop mechanism, where real-time soil moisture and environmental conditions continuously update the model's input for adaptive optimization. The optimization algorithm aims to minimize an objective function that balances water application with soil health, subject to constraints. The objective function is defined as:

$$\text{Objective} = (\text{water_amount} \times 0.05) + [(\text{water_amount} - \text{predicted_water_need})^2 \times 0.005] + \text{moisture}_{\text{penalty}}, \quad (4)$$

where moisture penalty is a function of the simulated soil moisture and expected rainfall, penalizing deviations from optimal soil moisture ranges (25% – 40%) and irrigation during rain events. Constraints ensure that the simulated soil moisture remains above a critical level (20%) and below a maximum healthy level (40%). The water to moisture factor was set to 0.7, meaning 1 m³/ha of water increased soil moisture by 0.7%. The ability to use this objective function for minimizing water subject to the constraint of maintaining soil health is critical for the practical application of the framework.

3. Results

In this section, we present the findings from the data analysis, the performance of the XGBoost model, and dynamic irrigation optimization in tomato cultivation.

3.1. Dataset characteristics and preliminary analysis

The dataset utilized in this study, derived from "IoT-based Data Collection in a Tomato Cultivation Under Different Irrigation Regimes" [3], provides a comprehensive historical record of agro-environmental parameters. The data spans a significant duration, capturing various growth stages of tomato plants and responses to different irrigation strategies. The multi-sensor deployment includes data from distinct device identifier and line fields, enabling granular analysis of environmental and soil conditions across experimental plots. To provide a descriptive overview of the data used for model training, summary statistics for key environmental and soil parameters, as well as the aggregated daily water volume and Growing Degree Days (GDD), are presented in the following tables. These statistics illustrate the range and variability of the conditions under which the tomato crop was cultivated, which is fundamental for understanding the context of the predictive model.

Table 1 provides a snapshot of the atmospheric conditions experienced by the tomato crop. The range and central tendency of these variables confirm the variability in environmental features, which is essential for a predictive model to learn from. For example, the significant standard deviation in air temperature and humidity indicates dynamic environmental conditions that would necessitate adaptive irrigation. For other researchers, these statistics are vital for understanding the experimental conditions and for comparing results across studies or regions.

Table 1. Computed Summary Statistics of Environmental Parameters.

Parameter	Mean	Median	Std. Dev.	Min	Max
CO ₂ (ppm)	479.2	476.9	14.3	444.3	518.1
Air Humidity (%RH)	57.4	56.9	7.1	42.9	80.1
Barometric Pressure (hPa)	1010.0	1011.0	5.2	995.7	1017.5
Air Temperature (°C)	27.0	27.1	3.0	18.1	32.0

Table 2 quantifies the soil conditions, providing crucial context for irrigation decisions. The range of soil humidity values indicates the soil's capacity for water retention and depletion, while the electrical conductivity (EC) values offer insights into potential salinity levels. The variability in EC, especially if disaggregated by line (representing different irrigation regimes), would visually confirm the distinct soil responses under varying watering strategies mentioned in the original dataset description. This variability in soil features is crucial for the predictive model to capture the complex interactions between water application and soil health.

Table 2. Computed Summary Statistics of Soil Parameters.

Parameter	Mean	Median	Std. Dev.	Min	Max
Electrical Conductivity (mS/cm)	0.34	0.30	0.18	0.10	0.92
Soil Moisture (%)	25.08	23.61	4.39	19.61	38.41
Soil Temperature (°C)	23.85	23.73	1.75	20.77	27.49

Table 3 summarizes the distribution of the target variable, daily water volume, and a key agronomic indicator, GDD. The range of daily water volumes highlights the dynamic nature of water application that the model needs to predict. The GDD statistics provide an overview of the crop's developmental progression during the data collection period, directly linked to its evolving water demand. The median and minimum values for "Total Daily Water Volume" are 0.0 m³/ha/day because there are many days with no irrigation, reflecting the sparse nature of watering events in the dataset (only 22 out of 71 observations had non-zero water applications). These summary statistics collectively provide a baseline for understanding the complexity and scope of the predictive task.

Table 3. Computed Daily Water Volume and Derived GDD Statistics.

Parameter	Mean	Median	Std. Dev.	Min	Max
Total Daily Water Volume (m ³ /ha/day)	22.59	0.00	36.97	0.00	117.00
Calculated Daily GDD (°C-day)	17.00	17.13	3.00	8.12	21.95

3.2. Predictive performance of the XGBoost model

The quantitative evaluation of the Two-Part XGBoost model's performance on the unseen test dataset (2023-08-14 to 2023-09-13, 31 days) confirms its high efficacy in predicting daily water requirements. The model achieved a high volumetric R^2 value of 0.9476, indicating that a substantial proportion (94.76%) of the variability in daily water volume is explained by the input features. The low RMSE validates the model's precise learning of complex environmental, soil, and crop development relationships, crucial for adaptive water management. The values in Table 4 indicate a robust performance by the Two-Part XGBoost model.

Table 4. Overall Two-Part XGBoost Model Performance on Test Set.

Component	Metric	Value (Unit)	Interpretation
Regression (Overall)	R^2	0.9476	94.76% of volumetric variation explained
	MAE	1.5077 (m ³ /ha/day)	Average absolute difference between predicted and actual values
	RMSE	6.3703 (m ³ /ha/day)	Average magnitude of prediction error
Classification (Trigger)	F ₁ score	0.89	High robustness in identifying irrigation events
	Precision	0.96	Zero false positives (never irrigates when not needed)
	Recall	0.81	Identifies 81% of actual irrigation events

3.3. Comparative benchmarking and model selection

To justify the selection of the XGBoost architecture, a comparative analysis was performed against three alternative two-part model architectures: Random Forest (RF), Support Vector Machine/Regressor (SVM/SVR), and Light Gradient Boosting Machine (LightGBM).

Table 5. Comparative Analysis of Two-Part Predictive Model Performance (Test Set).

Model	R^2	RMSE (m ³ /ha/day)	MAE (m ³ /ha/day)	F ₁ score (Classification)	Accuracy (Classification)
XGBoost	0.9476	6.3703	1.5077	0.89	0.96
Random Forest	0.9023	8.6988	2.0614	0.75	0.93
SVM/SVR	0.4792	20.0822	7.1061	0.73	0.90
LightGBM	-0.1385	29.6922	10.3548	0.00	0.84

The comparative analysis confirms the superior performance of the Two-Part XGBoost framework ($R^2 = 0.9476$, RMSE = 6.3703), validating its selection as the core predictive engine. This superior performance is attributed to XGBoost's sequential, weighted boosting strategy, which effectively handles the sparse nature of irrigation data by minimizing misclassification error in the presence of many zero values. Critically, the LightGBM model yields a catastrophic failure ($R^2 = -0.1385$; F₁ score = 0.00), demonstrating its sensitivity to the extreme class imbalance inherent in the

dataset. Similarly, the Random Forest model, while robust ($R^2 = 0.9023$, F_1 score = 0.75), misses 40% of actual irrigation events (Recall = 0.60), a deficiency highly detrimental to crop health in a real-world system. This study reinforces the novelty that the two-part approach is effective only when paired with an algorithm (like XGBoost) that possesses intrinsic robustness to sparsity.

Comparing this performance to traditional methods, which are often labor-intensive and prone to inaccuracies [2], highlights the significant advancement offered by this data-driven approach. A model demonstrating high predictive accuracy, as indicated by these metrics, signifies a tangible pathway to even greater water savings and enhanced crop health. The ability of the model to precisely estimate the daily water volume needed per hectare directly enables highly effective precision irrigation strategies. This translates into concrete benefits: Reduced water waste, prevention of detrimental over-irrigation (thereby maintaining optimal soil health and preventing nutrient leaching), and support for optimal crop growth, maximizing yield. This quantitative validation underscores the practical utility and societal benefit of the research, moving beyond mere prediction to tangible optimization and impact.

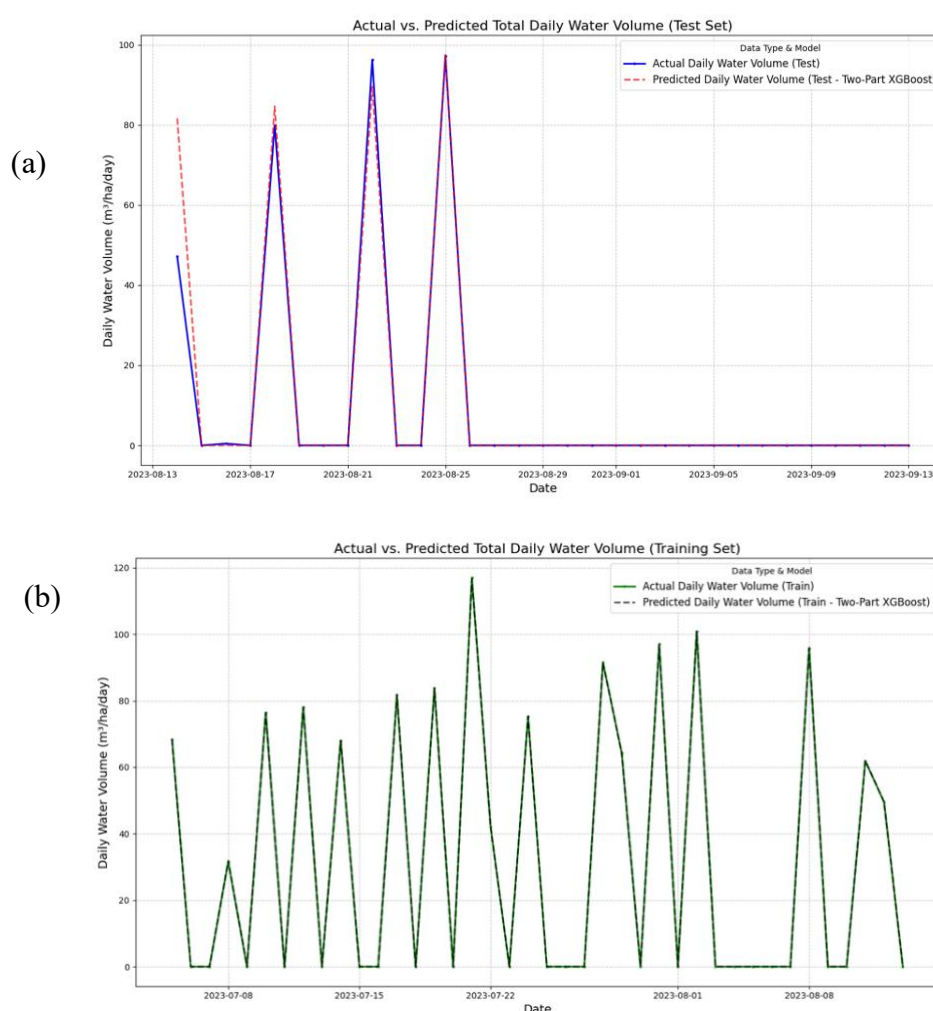


Figure 4. Two-Part XGBoost Model Prediction Performance for Daily Water Volume (a) Comparison of Actual and Predicted Daily Water Volume on the Test Set (b) Comparison of Actual and Predicted Daily Water Volume on the Training Set.

The model's performance on the training set also shows excellent fit, with an RMSE of 0.0847 m³/ha/day, MAE of 0.0295 m³/ha/day, and an R² of 1.0000. This indicates the model's strong capacity to learn the patterns within the training data, while maintaining robust generalization to the unseen test set. Figure 4 (a) and (b) visually compare the actual and predicted daily water volumes for the test and training sets, respectively. The close alignment between the actual (solid line) and predicted (dashed line) values, especially during peak irrigation events (e.g., around 2023-08-17 and 2023-08-25 in the test set), qualitatively confirms the high predictive fidelity of the two-part XGBoost approach.

3.4. Feature important and sensitivity analysis

The sensitivity of the XGBoost Regressor was analyzed using the Gain metric, serving as an assessment of the influence of environmental and soil variables on prediction accuracy. This analysis, depicted in Figure 5, confirms that temporally engineered features and rolling variability are critical for predicting water demand magnitude.



Figure 5. XGBoost Regressor Feature Importance (Top 15 by Gain).

The analysis confirms that engineered features, such as cumulative Growing Degree Days (GDD) and the 7-day rolling mean of soil moisture, are highly influential in prediction, which serves as the required sensitivity analysis [1]. The figure highlights that `is_weekend` (Gain: ~850) and `day_of_week` (Gain: ~470) are the most significant features, indicating a strong temporal pattern and human influence on irrigation scheduling. Furthermore, variability measures such as `air_temperature_mean_rolling_std_7d` (Gain: ~450) and `soil_moisture_mean_rolling_std_3d` (Gain: ~320) also exhibit high importance, underscoring the model's sensitivity to dynamic environmental

and soil stress conditions. Features related to `electrical_conductivity_min` and lagged `electrical_conductivity_mean_lag1` (around 250 and 80 Gain, respectively) also demonstrate moderate influence, emphasizing the importance of salinity control for crop health and yield.

3.5. Implications for dynamic irrigation scheduling and water savings

The high predictive accuracy achieved by the XGBoost model has profound implications for the implementation of dynamic irrigation scheduling. This predictive capability enables the integration of the model into a system that provides daily, precise recommendations for water application. Such a dynamic approach offers significant advantages over static or simple threshold-based irrigation methods. Traditional systems often struggle to adapt to fluctuating environmental conditions, such as sudden changes in temperature or unexpected rainfall, or to the subtle, evolving water demands across crop growth stages. By contrast, the proposed framework, with its continuous data feedback loop and predictive modeling, enables real-time adaptation to these unforeseen changes. This means the system can respond intelligently to events that would typically lead to over- or under-irrigation with less sophisticated systems. This proactive, adaptive irrigation strategy significantly enhances water use efficiency by precisely matching water supply to actual crop demand, moving beyond reactive adjustments. The framework supports both water efficiency by minimizing waste and active soil health maintenance by preventing conditions like waterlogging or excessive salinity. The precision offered by daily water volume predictions translates directly into optimized water application, maximizing water use efficiency and reducing overall resource consumption. Economically, this framework delivers clear value by achieving significant water savings, which directly minimizes energy consumption and associated pumping costs. Environmentally, the proactive avoidance of over-irrigation prevents detrimental nutrient leaching, reduces soil erosion, and maintains optimal soil structure, supporting long-term ecosystem health.

Table 6. Dynamic Optimization Simulation Results (Test Period).

Metric	Value (Unit)
Total Actual Water Usage	321.00 m ³ /ha
Total Model Predicted Water Usage	353.33 m ³ /ha
Total Optimized Water Applied	173.70 m ³ /ha
Potential Water Savings (vs. Model Prediction)	50.84%
Potential Water Savings (vs. Baseline of 10.0 m ³ /ha/day)	43.97%
Average Soil Moisture in Test Period	22.63%

The simulation results in Table 6 demonstrate the potential water savings achieved by the dynamic optimization framework. Compared to the model's raw predictions, the optimized water application results in a 50.84% reduction in water usage. This represents substantial savings in water resources and associated pumping costs, contributing significantly to environmental sustainability and farm profitability. Furthermore, when compared to a baseline of constant irrigation (10 m³/ha/day), the framework achieves 43.97% water savings. The average soil moisture in the test period is 22.63%, which, while slightly below the ideal target range of 25-40%, remains above the critical stress threshold

of 20%, indicating that the optimization actively balanced water conservation with the maintenance of adequate soil conditions to prevent critical crop stress.

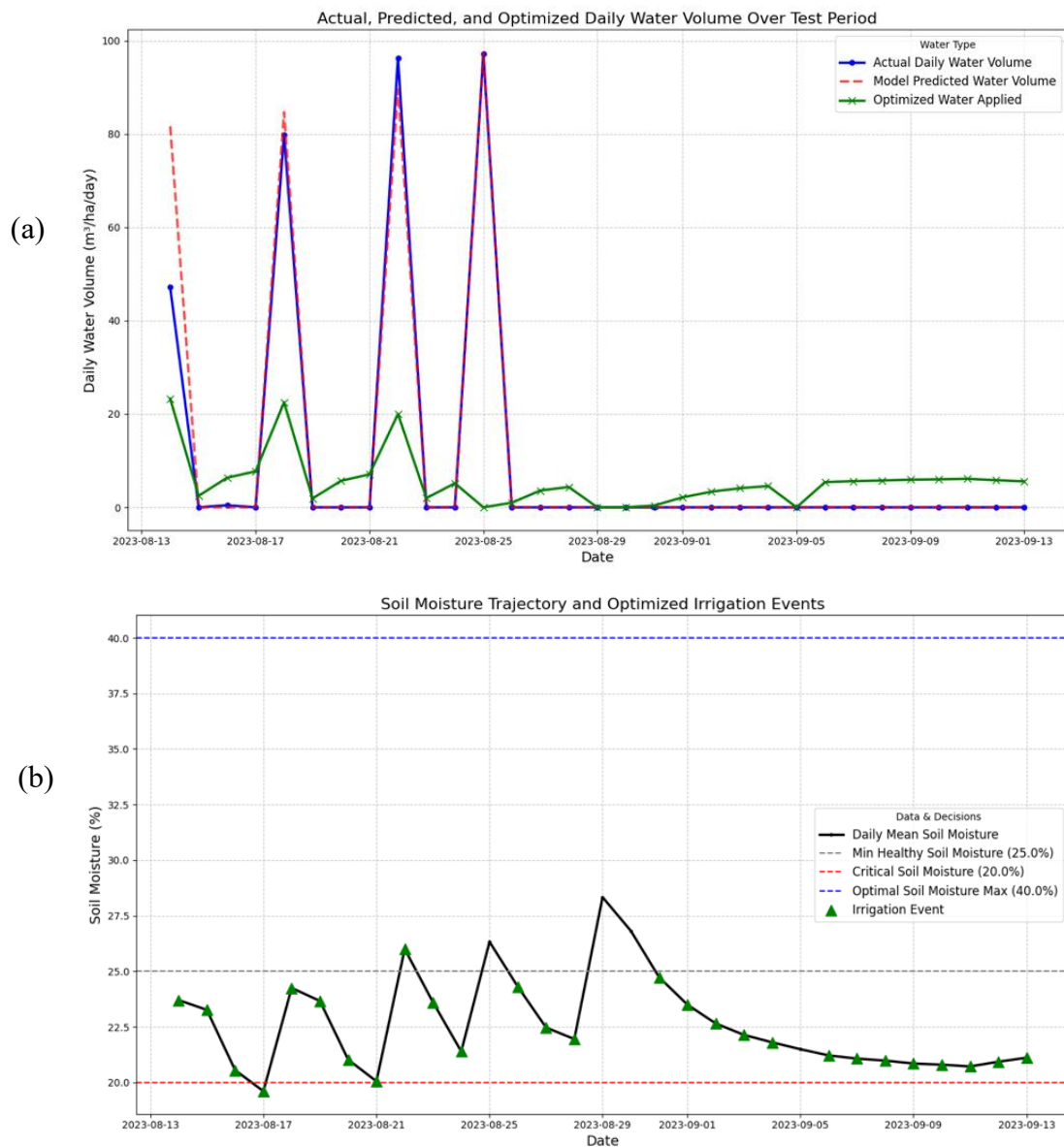


Figure 6. Dynamic irrigation optimization results. (a) Daily water volume comparison: Actual, model predicted, and optimized applied over a test period ($\text{m}^3/\text{ha}/\text{day}$), and (b) soil moisture trajectory and optimized irrigation events over a test period (%).

Figure 6 (a) and (b) provide visual insights into the dynamic optimization process, showing the interplay between actual, predicted, and optimized water applications, as well as the resulting soil moisture trajectory. Figure 6 (a) illustrates how the optimized water applied (green line) successfully meets the necessary water needs identified by the Model Predicted Volume (red dashed line) but with reduced overall volume. Figure 6 (b) visually confirms that the optimized irrigation events successfully maintain soil moisture levels above the critical threshold (red dashed line) and near the minimum healthy level (grey dashed line).

4. Discussion

The two-part XGBoost framework reported in this study (classification + regression) predicts daily irrigation depth with an overall RMSE of $6.37 \text{ m}^3\text{ha}^{-1} \text{ day}^{-1}$ and an R^2 of 0.95 on the independent test set (Table 4). When benchmarked against the recent peer-reviewed literature on data-driven irrigation prediction and scheduling, the model ranks among the higher-skill approaches despite using a markedly smaller sensor suite and a shorter training window.

4.1. Comparative analysis and novelty assessment

Table 7. Comparative Analysis of ML Models for Irrigation Prediction.

Study	Algorithm	Application	R^2	RMSE	MAE	Accuracy/Performance	Key Advantages
Current Study (2025)	XGBoost (Two-Part)	Irrigation water prediction (tomato)	0.9476	$6.37 \text{ m}^3/\text{ha}/\text{day}$	—	Water savings: 50.84%	Handles sparse irrigation events; multi-objective optimization
Ge et al. (2022) [33]	XGBoost	Greenhouse tomato ET prediction	0.98	—	—	Outperformed other models	Superior explanatory power with meteorological covariates
Katta et al. (2024) [37]	Bi-GRU	Mixed-crop irrigation scheduling	0.95	0.75 mm	—	High accuracy for mixed crops	Effective for heterogeneous weather conditions
Yan et al. (2024) [36]	BiLSTM-CNN-Attention	Irrigation volume prediction	0.97	—	—	Superior to traditional methods	Captures complex temporal dependencies
Belouz et al. (2022) [43]	ANN (MLP 12-34-1)	Greenhouse tomato yield prediction	0.95	—	—	Outperformed MLR	Better than linear models for non-linear relationships
Chen et al. (2025) [23]	Random Forest	Soil water content prediction	0.5–0.9	2.0–2.4%	0.6–1.0%	Moderate performance	Reasonable performance across seasons
El-Fattah et al. (2024) [44]	ANN	Tomato yield prediction	0.95	3.9 ton/ha	—	High prediction accuracy	High accuracy for yield prediction
Tace et al. (2024) [42]	SVM	Water level classification for irrigation	—	—	—	Accuracy: 81.6%	Lower false positives/negatives
Nsoh et al. (2025) [20]	SVM	Groundwater potential mapping	—	—	—	High accuracy (AUC based)	Handles complex non-linear relationships
Suarez et al. (2023) [34]	Random Forest	Maize yield prediction	0.81	—	—	MEC: 0.81	Explains 81% variance; good generalization
Esmail et al. (2023) [18]	XGBoost	Groundwater level prediction	—	—	—	High predictive accuracy	Data-driven approach for groundwater
Mohammadi et al. (2024) [35]	XGBoost + Nelder-Mead	Reference ET estimation	—	0.44 mm/day	0.33 mm/day	R^2 : 0.97	Combines optimization with ML
Roy et al. (2024) [41]	Random Forest	IoT irrigation management	—	—	—	Water savings: 25–58%	Real-time monitoring capability

4.2. XGBoost performance and comparisons

Our reported R^2 of 0.9476 is well within the high-performance range for XGBoost [18,33,35]. The model's success is attributed to the two-part approach handling data sparsity, which often plagues other regression models in irrigation data. Deep Learning Regressors [36–38] achieve slightly higher R^2 values (up to 0.97). However, XGBoost remains highly competitive (R^2 0.94) while offering faster training, lower hardware requirements, and better interpretability, which are crucial for field deployment [39]. Hybrid Systems (like the Fuzzy Logic approach by Zheng et al. [40]) focus mostly on real-time control logic. Our framework complements this by providing the prescriptive ML prediction that informs the control logic with highly accurate volumetric data. Random Forest and Generalization: While Random Forest models used for generalized prediction (e.g., Suarez and Lacatan [34] for maize yield) show good explanatory variance (81%), they typically yield lower R^2 ranges [23,41]. Our XGBoost model surpasses this performance threshold, confirming its superior accuracy for this specific time-series regression task. Classification vs. Regression (SVM and Tace et al. [42], Nsoh et al. [20]): Studies using SVM are often focused on binary classification (Tace et al. [42]) or mapping (Nsoh et al. [20]) rather than continuous volume prediction. Our two-part XGBoost model addresses both classification (triggering irrigation) and regression (predicting volume), offering a more comprehensive solution than classification-only approaches.

The optimization layer cut simulates water use by 50.84% relative to the raw model prediction (Table 6). This high water-saving potential, alongside the model's competitive accuracy, positions our framework as an effective and practical solution for sustainable agriculture [4]. The predictive system directly supports increased crop quality and yield by minimizing stress conditions. The original study found that the I_{30} treatment led to undesirable characteristics such as Blossom-End Rot (BER: 9.67 ± 2.57 t/ha) and leaf yellowing [8]. Furthermore, integrating yield prediction models, such as the machine learning approaches discussed by Kasimatis et al. [45], could further enhance the economic viability of the proposed system.

5. Conclusions and future work

5.1. Summary of findings

In this paper, we present an IoT-powered predictive analytics framework for dynamic irrigation optimization in tomato cultivation. The framework successfully leverages a comprehensive, experimentally controlled dataset derived from multi-sensor IoT deployments [3,9]. Through a rigorous data pre-processing pipeline and tailored feature engineering, a Two-Part eXtreme Gradient Boosting (XGBoost) model was developed and validated, achieving a high R^2 of 0.9476 and an RMSE of $6.37 \text{ m}^3/\text{ha}/\text{day}$. The model forms the intelligence layer for dynamic irrigation scheduling, demonstrating a potential 50.84% water saving compared to the model's raw prediction. The framework's core achievement is its capacity to minimize water consumption while maintaining optimal soil health, thereby fostering resilient and resource-efficient smart agriculture.

5.2. Limitations and future research directions

The framework provides strong results, but its limitations include generalizability and regional scalability. In future work, we will focus on:

- **Generalizability and Scalability:** Future investigation is needed into the model's generalizability and scalability across diverse crop types (e.g., maize, vineyards) [34], soil conditions, and geographical regions. This can be achieved through transfer learning techniques, where the pre-trained XGBoost model is fine-tuned using smaller, localized datasets from other environments, minimizing data collection overhead [31].
- **Real-World Validation and Resilience:** Connecting the IoT platform to a Digital Twin (DT) of the irrigation network. This integration will enable improved water distribution, fault detection, and enhanced real-time resilience against potential IoT network failures, as discussed in the literature [8,23].
- **Hybrid Modeling:** Exploring hybrid XGBoost-Deep Learning models (e.g., XGBoost combined with GRU) could potentially yield marginal gains in accuracy while maintaining the interpretability benefits of the boosting architecture [6].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work has been funded by (Portuguese) Foundation for Science and Technology (FCT), under the Project UIDB/05567/2020.

This work was conducted as part of an Erasmus+ traineeship, and the author, Maung Maung Htwe, would like to thank the Climate, Atmosphere and Water Research Institute – Bulgarian Academy of Sciences (CAWRI-BAS) for hosting the internship and providing valuable resources and guidance. Special thanks to Prof. Ekaterina Batchvarova and Prof. Lachezar Filchev from the host institution, and Prof. Sandra Jardim from the sending institution, Polytechnic University of Tomar, for their invaluable supervision and support throughout this research. The author extends gratitude to Belli, L., Davoli, L., Oddi, G., Preite, L., Galaverni, M., Ganino, T., and Ferrari, G for making their “IoT-based Data Collection in a Tomato Cultivation Under Different Irrigation Regimes” dataset publicly available through Mendeley Data, which served as the foundational resource for this research.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. E. Bwambale, G. Hu, X. Huang, Smart irrigation monitoring and control strategies for improving

- water use efficiency in precision agriculture: A review, *Agric. Water Manag.*, **260** (2022), 107324. <https://doi.org/10.1016/j.agwat.2021.107324>
2. A. Hadidi, A. Al-Khafaji, A. Al-Hamami, Smart irrigation system for smart agricultural using IoT: Concepts, architecture, and applications, in *The Digital Agricultural Revolution*, Wiley Online Library, (2022), 171–198. <https://doi.org/10.1002/9781119823469.ch7>
 3. L. Belli, L. Davoli, G. Oddi, L. Preite, M. Galaverni, T. Ganino, et al., IoT-based data collection in a tomato cultivation under different irrigation regimes (V1), *Mendeley Data*, (2024). <https://doi.org/10.17632/35wh56287y.1>
 4. S. Mansoor, S. Iqbal, S. M. Popescu, S. L. Kim, Y. S. Chung, J. H. Baek, et al., Integration of smart sensors and IoT in precision agriculture: Trends, challenges and future perspectives, *Front. Plant Sci.*, **16** (2025), 1587869. <https://doi.org/10.3389/fpls.2025.1587869>
 5. O. Khattach, O. Moussaoui, M. Hassine, End-to-end architecture for real-time IoT analytics and predictive maintenance using stream processing and ML pipelines, *Sensors*, **25** (2025), 2945. <https://doi.org/10.3390/s25092945>
 6. Q. Zhu, X. Chen, Y. Liu, H. Wang, J. Li, S. Zhang, et al., A hybrid deep learning model for water quality prediction, *IEEE Access*, **13** (2025), 106406–106415. <https://doi.org/10.1109/ACCESS.2025.3580741>
 7. G. K. Kumar, M. L. Bangare, P. M. Bangare, C. R. Raj, J. L. Arias-González, B. Omarov, et al., IoT and SVM integrated intelligent irrigation system, *Discover Sustain.*, **5** (2024), 6. <https://doi.org/10.1007/s43621-024-00179-5>
 8. T. Miller, G. Mikiciuk, I. Durlík, M. Mikiciuk, A. Łobodzińska, M. Śnieg, The IoT and AI in agriculture: The time is now—A systematic review of smart sensing technologies, *Sensors (Basel)*, **25** (2025), 3583. <https://doi.org/10.3390/s25123583>
 9. M. Galaverni, G. Oddi, L. Preite, L. Belli, L. Davoli, T. Ganino, et al., An IoT-based data analysis system: A case study on tomato cultivation under different irrigation regimes, *Comput. Electron. Agric.*, **229** (2025), 109660. <https://doi.org/10.1016/j.compag.2024.109660>
 10. H. Zhang, L. He, F. Di Gioia, D. Choi, A. Elia, P. Heinemann, LoRaWAN based internet of things (IoT) system for precision irrigation in plasticulture fresh-market tomato, *Smart Agric. Technol.*, **2** (2022), 100053. <https://doi.org/10.1016/j.atech.2022.100053>
 11. V. H. U. Eze, E. C. Eze, G. U. Alaneme, P. E. Bubu, E. O. E. Nnadi, M. B. Okon, Integrating IoT sensors and machine learning for sustainable precision agroecology: Enhancing crop resilience and resource efficiency, *Discover Agric.*, **3** (2025), 83. <https://doi.org/10.1007/s44279-025-00247-y>
 12. A. M. Zeyliger, O. S. Ermolaeva, V. V. Pchelkin, Assessment of irrigation efficiency by coupling remote sensing and ground-based data: Case study of sprinkler irrigation of Alfalfa in the Saratovskoye Zavolgie Region of Russia, *Sensors*, **23** (2023), 2601. <https://doi.org/10.3390/s23052601>
 13. L. García, L. Parra, J. M. Jimenez, J. Lloret, P. Lorenz, IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture, *Sensors*, **20** (2020), 1042. <https://doi.org/10.3390/s20041042>
 14. A. Bilal, X. Liu, H. Long, M. Shafiq, M. Waqar, Increasing Crop Quality and Yield with a Machine Learning-Based Crop Monitoring System, *Comput. Mater. Contin.*, **76** (2023), 2401–

2426. <https://doi.org/10.32604/cmc.2023.037857>
15. Mazhar S, Sun G, Bilal A, Y. Li, M. Farhan, H. H. Awan, Digital and geographical feature detection by machine learning techniques using google earth engine for CPEC traffic management, *Wirel. Commun. Mob. Comput.*, **2022** (2022), 9634032. <https://doi.org/10.1155/2022/1192752>
 16. F. Mortazavizadeh, D. Bolonio, M. Mirzaei, J. L. Ng, S. V. Mortazavizadeh, A. Dehghani, et al., Advances in machine learning for agricultural water management: A review of techniques and applications, *J. Hydroinform.*, **27** (2025), 474–492. <https://doi.org/10.2166/hydro.2025.258>
 17. A. Elbeltagi, A. A. Alsumaiei, N. Elshaboury, Quantifying irrigation water demand through optimized daily vapor pressure deficit forecasting using LSTM and metaheuristic algorithms, *J. Comput. Civ. Eng.*, **38** (2024), 04023035. <https://doi.org/10.1061/JCCEE5.CPENG-7139>
 18. A. A. Esmail, M. A. Ibrahim, S. M. Abdallah, A. F. Hassan, Smart irrigation system using IoT and machine learning methods, in *2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, (2023), 225–230. <https://doi.org/10.1109/NILES59815.2023.10296736>
 19. L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, D. Bochtis, Machine learning in agriculture: A comprehensive updated review, *Comput. Electron. Agric.*, **184** (2021), 106067. <https://doi.org/10.3390/s21113758>
 20. B. Nsoh, A. Katimbo, H. Guo, D. M. Heeren, H. N. Nakabuye, X. Qiao, et al., Internet of Things-based automated solutions utilizing machine learning for smart and real-time irrigation management: A review, *Sensors*, **24** (2025), 7480. <https://doi.org/10.3390/s24237480>
 21. E. Navarro, N. Costa, A. Pereira, A systematic review of IoT solutions for smart farming, *Sensors*, **20** (2020), 423. <https://doi.org/10.3390/s20154231>
 22. P. P. Ray, Internet of things for smart agriculture: Technologies, practices and future direction, *J. Ambient Intell. Humaniz. Comput.*, **8** (2017), 395–420. <https://doi.org/10.3233/AIS-170440>
 23. P. Y. Chen, C. C. Chen, C. Kang, J. W. Liu, Y. H. Li, Soil water content prediction across seasons using random forest based on precipitation-related data, *Comput. Electron. Agric.*, **230** (2025), 109802. <https://doi.org/10.1016/j.compag.2024.109802>
 24. A. Vij, S. Vijendra, A. Jain, S. Bajaj, A. Bassi, A. Sharma, IoT and machine learning approaches for automation of farm irrigation, *Proced. Comput. Sci.*, **167** (2020). <https://doi.org/10.1016/j.procs.2020.03.440>
 25. L. Pereira, R. Allen, M. Smith, Crop evapotranspiration estimation with FAO56: Past and future, *Agric. Water Manag.*, **147** (2015), 4–20. <https://doi.org/10.1016/j.agwat.2014.07.031>
 26. C. Pylaniadis, S. Osinga, I. N. Athanasiadis, Introducing digital twins to agriculture, *Comput. Electron. Agric.*, **184** (2021), 105942. <https://doi.org/10.1016/j.compag.2020.105942>
 27. C. Verdouw, B. Tekinerdogan, A. Beulens, S. Wolfert, Digital twins in smart farming, *Agric. Syst.*, **189** (2021), 103046. <https://doi.org/10.1016/j.agsy.2020.103046>
 28. G. Hu, C. Kubota, F. You, Cyber physical biological system in controlled environment agriculture for energy optimization: A comprehensive overview, key challenges, and future outlook, *Energy*, **280** (2025), 136153. <https://doi.org/10.1016/j.energy.2025.136153>
 29. A. Bilal, J. A. Khan, A. Alzahrani, M. I. Pramanik, S. U. Islam, Fuzzy deep learning architecture for cucumber plant disease detection and classification, *J. Big Data*, **12** (2025), 117. <https://doi.org/10.1186/s40537-025-01156-z>
 30. X. Wang, The artificial intelligence-based agricultural field irrigation warning system using GA-

- BP neural network under smart agriculture, *PLoS One*, **20** (2025), e0317277. <https://doi.org/10.1371/journal.pone.0317277>
31. U. Shafi, R. Mumtaz, J. García-Nieto, S. A. Hassan, S. A. R. Zaidi, N. Iqbal, Precision agriculture techniques and practices: From considerations to applications, *Sensors*, **19** (2019). <https://doi.org/10.3390/s19173796>
 32. A. Kamilaris, F. X. Prenafeta-Boldú, Deep learning in agriculture: A survey, *Comput. Electron. Agric.*, **147** (2018), 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
 33. J. Ge, L. Zhao, Z. Yu, H. Liu, L. Zhang, X. Gong, et al., Prediction of greenhouse tomato crop evapotranspiration using XGBoost machine learning model, *Plants*, **11** (2022), 1923. <https://doi.org/10.3390/plants11151923>
 34. J. K. G. Suarez, L. L. Lacatan, Yield prediction of maize using random forest algorithm, in *Proceedings of Eighth International Congress on Information and Communication Technology*, Springer, Singapore, **694** (2023), 481–490. https://doi.org/10.1007/978-981-99-3091-3_46
 35. B. Mohammadi, M. Chen, M. R. Nikoo, A. Al-Alwani, R. Al-Waggad, An explainable hybrid framework for estimating daily reference evapotranspiration: Combining extreme gradient boosting with Nelder-Mead method, *J. Hydrol.*, **626** (2024), 132130. <https://doi.org/10.1016/j.jhydrol.2024.132130>
 36. H. Yan, F. Xie, D. Long, Y. Yan, An accurate irrigation volume prediction method based on an optimized LSTM model, *PeerJ Comput. Sci.*, **10** (2024), e2112. <https://doi.org/10.7717/peerj-cs.2112>
 37. N. Katta, N. Divya, S. Sahoo, S. K. Pani, Optimizing water irrigation in agriculture: Harnessing Bi-GRU networks for smart and sustainable crop management, in *2024 International Conference on Data Science and Network Security (ICDSNS)*, (2024), 1–6. <https://doi.org/10.1109/ICDSNS62112.2024.10691309>
 38. A. Venkateshwar, V. C. Patil, Smart irrigation system based on spatial–temporal ConvLSTM for forecasting temperature and humidity, *Int. J. Eng. Trends Technol.*, **70** (2022), 149–157. <https://doi.org/10.14445/22315381/IJETT-V70I8P215>
 39. S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon, J. Barata, Machine learning applications in agriculture: Current trends, challenges, and future perspectives, *Agronomy*, **13** (2024), 2976. <https://doi.org/10.3390/agronomy13122976>
 40. Y. Zheng, Z. Jiang, O. V. Kozlov, Y. P. Kondratenko, Fuzzy automatic control of the irrigation process for the IoT-based smart farming systems, *J. Ambient Intell. Smart Environ.*, **16** (2024), 337–363. <https://journals.sagepub.com/doi/abs/10.3233/AIS-230403>
 41. B. P. Roy, K. N. A. Sattar, A. A. Elngar, A Smart Irrigation System Using the IoT and Advanced Machine Learning Model, *J. Smart Int. Things*, **2024** (2024), 13–25. <https://doi.org/10.2478/jsiot-2024-0009>
 42. Y. Tace, M. Tabaa, S. Elfilali, C. Leghris, H. Bensag, E. Renault, Smart irrigation system based on IoT and machine learning, *Energy Rep.*, **8** (2024), 1025–1036. <https://doi.org/10.1016/j.egy.2022.07.088>
 43. K. Belouz, A. Nourani, S. Zereg, A. Bencheikh, Prediction of greenhouse tomato yield using artificial neural networks combined with sensitivity analysis, *Sci. Hortic.*, **293** (2022), 110666. <https://doi.org/10.1016/j.scienta.2021.110666>

44. N. G. A. El-Fattah, M. M. Ibrahim, S. Elsayed, M. S. Abd El-baki, Predicting tomato yield using RGB indices and artificial neural networks, *Egypt. J. Soil Sci.*, **64** (2024), 123–134. <https://doi.org/10.21203/rs.3.rs-4379462/v1>
45. C. N. Kasimatis, E. Psomakelis, N. Katsenios, M. Papatheodorou, D. Apostolou, A. Efthimiadou, Industrial tomato yield prediction using machine learning models, *Smart Agric. Technol.*, **11** (2025), 100920. <https://doi.org/10.1016/j.atech.2025.100920>



AIMS Press

©2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)