Jurnal Elektronika dan Telekomunikasi (JET), Vol. 25, No. 2, December 2025, pp. 102-118
Accredited by RISTEKDIKTI, Decree No: **158/E/KPT/2021**
doi: 10.55981/jet.799

Comparative Performance of Regression and Ensemble Learning Algorithms in Precision Irrigation Forecasting of Sweet Potato

Muthia Rahmah*, Indra Maulana

Informatics and Computer Engineering Education
Institut Prima Bangsa
Cideng, Jl. Brigjen Darsono Bypass No.20, Kertawinangun, Kedawung
Cirebon, Indonesia

Abstract

Precision irrigation is essential for sustainable agriculture under increasing water scarcity. This study compared regression and ensemble learning algorithms for forecasting irrigation requirements in sweet potato, a crop characterized by high variability in water demand. An Internet of Things (IoT)-based prototype was deployed to collect real-time data on soil moisture, temperature, humidity, light intensity, and atmospheric pressure over 42 hours and 50 minutes (August 4-5, 2025), encompassing two complete diurnal cycles at 10-minute intervals and yielding 243 temporal observations. Following preprocessing and feature engineering with lag-based temporal features, the final dataset comprised 240 samples (192 training, 48 testing) using chronological time-based splitting to prevent data leakage. Five algorithms, Support Vector Regression (SVR), AdaBoost, Extreme Gradient Boosting (XGBoost), Random Forest Regressor (RFR), and CatBoost, were evaluated under default and hyperparameter-tuned configurations using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) as evaluation metrics. Tuned Random Forest achieved superior performance (R^2 = 0.9802, RMSE = 9.58, MAE = 6.08), followed by default Random Forest (R^2 = 0.9786) and default CatBoost (R^2 = 0.9687). XGBoost demonstrated strong performance (R^2 = 0.9670 tuned) but exhibited overfitting tendencies with near-perfect training scores. SVR improved substantially after tuning (R^2 = 0.328 to 0.797), although it remained inferior to ensemble methods. Overall, ensemble methods, particularly XGBoost and Random Forest, demonstrated superior efficacy for sweet potato irrigation forecasting. These findings underscore the potential of IoT-integrated machine learning to enhance water-use efficiency and support sustainable smart farming practices.

Keywords: ensemble learning, IoT, machine learning, precision irrigation, regression, sweet potato

I. INTRODUCTION

A growing demand for sustainable food production, water shortages, and climate change are placing increasing strain on the agricultural industry. Precision agriculture, supported by Artificial Intelligence (AI), has emerged as a transformative response to these challenges. Recent bibliometric reviews reveal that precision irrigation research has increasingly emphasized the integration of Internet of Things (IoT)-based devices, remote sensing technologies, and machine learning (ML) algorithms as key enablers for sustainable irrigation management [1]. Advanced ML-driven crop recommendation systems have demonstrated significant potential in enhancing precision agriculture through cloud-based platforms, achieving prediction accuracies exceeding 94% while promoting sustainable farming practices [2].

Smart irrigation, a central component of precision agriculture, employs sensors and automation to enhance water distribution and crop performance. Research on IoT-enabled greenhouse systems has shown that sensor-based monitoring combined with regression analysis can yield highly accurate predictions of plant growth conditions, with reported R^2 values surpassing 0.95 [3]. Recent studies on IoT-based systems for monitoring microbial loads in agricultural drying processes have demonstrated that machine learning models, particularly Random Forest Regression (RFR), can achieve robust prediction accuracy when integrated with real-time sensor data [4]. Furthermore, IoT and ML integration for forecasting physiological parameters of crop leaves, such as leaf-turgor pressure and leaf temperature, has shown promising results with Support Vector Regression (SVR) achieving R^2 values of 0.96 and 0.99, respectively, enabling proactive crop management under changing climate conditions [5].

By extracting predictive insights from intricate agricultural information, machine learning has emerged as a key factor in the advancement of smart irrigation. Numerous irrigation-related issues have been effectively addressed using regression and ensemble learning methods. For example, SVR, Random Forest, and other models have been employed to estimate reference evapotranspiration (ETo), with SVR achieving superior accuracy (Mean Absolute Error (MAE) = 0.03; Root Mean Squared Error (RMSE) = 0.05) [6]. Similarly, spatiotemporal modeling of soil moisture using

* Corresponding Author.
Email: mvmuthia@gmail.com
Received: September 9, 2025 ; Revised: October 31, 2025
Accepted: November 26, 2025 ; Published: December 31, 2025

regression and deep learning techniques demonstrated that SVR and Long Short-Term Memory (LSTM) networks achieved nearly 90% predictive accuracy across multiple soil depths [7]. These results confirm the capacity of regression-based models to capture intricate irrigation dynamics.

A recent comprehensive study on spatiotemporal soil moisture prediction across three soil types (loam, sandy loam, and silt loam) demonstrated that RFR achieved exceptional accuracy particularly in loam soil at 80 cm depth (RMSE = 0.89, MAE = 0.74), while Extreme Gradient Boosting (XGBoost) displayed minimal errors at shallow depths (RMSE = 0.025, MAE = 0.159) in sandy loam [8]. Additionally, solar radiation estimation using cost-effective light sensors combined with Random Forest regression achieved $R^2 = 0.9922$ with RMSE = 44.46 W/m², demonstrating the viability of low-cost sensor integration with ML for agricultural monitoring [9].

Ensemble learning approaches extend this capability even further. Comparative evaluations reveal that algorithms such as Random Forest and AdaBoost often outperform traditional regression methods in yield estimation due to their robustness in handling non-linearities and heterogeneous data [10]. Moreover, the integration of regression with Geographic Information System (GIS)-based soil monitoring has optimized sensor deployment and improved irrigation decision-making [11]. Reviews also indicate that boosting-based methods can deliver superior forecasting performance compared to single models [12]. In a related review, bagging-based ensemble methods such as Random Forest have been widely applied in agricultural domains and are reported to improve predictive accuracy compared to individual models [13].

Recent advances in IoT-driven ensemble machine learning for aquaculture have shown that combining SVR, Random Forest, and Gradient Boosting can achieve R^2 values of 0.84-0.91 for dissolved oxygen and water quality index prediction, enabling real-time proactive decisions to minimize mortality rates [14]. Moreover, an efficient IoT-based crop damage prediction framework integrating XGBoost, CatBoost, and Light Gradient Boosting Machine (LightGBM) demonstrated that XGBoost outperformed other models with 89.56% accuracy, 88.1% sensitivity, and 84.8% F1-score, while also achieving excellent imputation capability (Mean Squared Error (MSE) = 0.0213, R^2 = 0.99) for handling missing data in real-time agricultural datasets [15]. Furthermore, explainable machine learning approaches for ship fuel consumption prediction revealed that XGBoost achieved the highest performance (R^2 = 0.997 train, 0.95 test) with interpretability analysis identifying engine power as the most significant predictor, demonstrating the importance of model transparency in industrial IoT applications [16].

Emerging technologies are expanding the landscape of smart irrigation, with Unmanned Aerial Vehicles (UAVs) combined with machine learning being employed for canopy stress detection, evapotranspiration modeling, and water stress assessment at high spatial resolution [17]. At the same time, Fourth Industrial Revolution (4IR) technologies such as IoT, blockchain, big data analytics, and artificial intelligence are increasingly being applied in smart irrigation, enabling automation, scalability, and resource efficiency [18].

Applications at the crop level further validate the effectiveness of ML-based irrigation. For instance, irrigation management systems for tomato cultivation improved water-use efficiency by up to 32% without yield reduction [19]. Similarly, low-cost soil moisture sensor calibration has enhanced long-term monitoring accuracy [20]. Climate-resilient irrigation strategies also underscore the reliability of algorithms such as SVR and Random Forest in estimating evapotranspiration and guiding water allocation under extreme weather conditions [21]. Systematic reviews further highlight that AI-driven irrigation management frameworks have demonstrated scalability and real-time applicability, supporting their adoption in diverse agricultural contexts [22]. Another review emphasizes that such smart irrigation systems demonstrate adaptability and can be deployed across diverse agricultural contexts, from smallholder to large-scale farming [23].

Recent work on intelligent air pollution prediction using optimized Random Forest Regression achieved prediction accuracy of 99.98% (R^2 = 0.99) with IoT-based detection nodes, demonstrating the potential of optimized ensemble methods to protect vulnerable populations and enable completion of daily activities without obstacles [24]. Additionally, IoT devices combined with Random Forest regression for air quality monitoring revealed that the model with 100 estimators delivered the best overall performance for both Air Quality Index (AQI) 10 and AQI 2.5, achieving the lowest MAE of 0.2785 and 0.2483, respectively, enabling quick reactions to pollution hotspots and development of effective strategies to reduce pollution sources [25].

Although substantial progress has been achieved, critical research gaps remain. Many studies focus on individual algorithms or crop-specific applications, which limits broader generalization and raises questions regarding the relative performance of regression and ensemble approaches under standardized conditions. Reviews further indicate that while the adoption of AI in irrigation has accelerated in recent years, there remains limited evidence regarding computational efficiency, robustness, and scalability across diverse crop systems [23]. A comparative study of predictive algorithms for IoT smart agriculture sensor data revealed that RFR, Seasonal Autoregressive Integrated Moving Average (SARIMA), and Artificial Neural Networks (ANNs) outperformed NeuralProphet in terms of accuracy and computational efficiency, highlighting the need for standardized benchmarking frameworks to evaluate algorithm performance under consistent field conditions [26].

This study compares five regression and ensemble learning algorithms, SVR, AdaBoost, XGBoost, Random Forest, and CatBoost, for predicting irrigation requirements in sweet potato farming to address these gaps. Although machine learning has been applied to various crops in earlier research, rigorous comparative analyses of different algorithms in actual IoT-based deployments remain limited. To the best of our knowledge, this is the first study to benchmark these algorithms for sweet potato irrigation forecasting, thereby offering both scientific and practical insights. Implemented within a cloud-based computing environment (Google Colab), the work emphasizes predictive accuracy, computational efficiency, and field-level applicability. These contributions strengthen the body of evidence supporting AI-driven smart irrigation and highlight pathways toward more sustainable and climate-resilient agriculture.

This study addresses these gaps by implementing a multi-day IoT deployment (43 consecutive hours) rather than single-session monitoring, ensuring the dataset captures complete diurnal variability and realistic temporal patterns. With 243 timestamped observations spanning two full 24-hour cycles, the research provides a more robust foundation for evaluating algorithm performance under field conditions compared to limited-scope studies.

II. METHOD

This study employed an experimental and quantitative research design to evaluate the performance of regression and ensemble learning algorithms in forecasting irrigation requirements for sweet potato. The methodology was structured into several subsections covering data acquisition, preprocessing, model development, and evaluation procedures.

A. Data Acquisition

The dataset used in this study was obtained from an IoT-based prototype deployed in a sweet potato field as depicted in figure 1 located in Halimpu Village, Beber District, Cirebon Regency, Indonesia (GPS coordinates: 6°44′S, 108°33′E). The system consisted of a NodeMCU ESP8266 microcontroller connected to multiple sensors, including a capacitive soil moisture sensor (measuring volumetric water content as analog values 0-1023), a DHT22 temperature and humidity sensor (±0.5 °C and ±2% accuracy), a BH1750 light intensity sensor (1-65,535 lux range), and a BMP180 atmospheric pressure sensor (±1 hPa accuracy). A relay-controlled 12 V DC water pump (5 L/min flow rate) was integrated to enable automated irrigation responses. Power was supplied through a 12 V 7 Ah rechargeable battery with solar panel supplementation to ensure uninterrupted operation throughout the extended monitoring period.

Figure 1. Prototype Setup for Data Acquisition in Sweet Potato Field, Halimpu Village, Cirebon Regency.

Data collection was conducted continuously from August 4, 2025 at 05:00 AM until August 5, 2025 at 11:50 PM, totaling 42 hours and 50 minutes of uninterrupted monitoring. Sensor readings were captured at precisely 10-minute intervals, resulting in 243 timestamped observations. This extended deployment was intentionally designed to capture two complete 24-hour diurnal cycles, encompassing morning irrigation phases (05:00-07:30), daytime evapotranspiration periods (08:00-17:00), evening re-irrigation cycles (18:00-20:00), overnight stabilization (21:00-04:50), and transitional shifts between day and night.

The multi-day monitoring approach addressed a critical limitation identified in preliminary research, where single-session datasets of only 6-8 hours failed to capture full irrigation dynamics and diurnal variability. Each sensor reading recorded six parameters: timestamp, soil moisture (raw analog-to-digital converter (ADC) value), air temperature (°C), relative humidity (%), light intensity (lux), atmospheric pressure (hPa), and relay status (on/off). All readings were transmitted in real time via the Message Queuing Telemetry Transport (MQTT) protocol over Wi-Fi to a cloud database (ThingsBoard IoT Platform) for centralized storage and subsequent analysis.

Throughout the monitoring period, the prototype captured four distinct irrigation events, two each on consecutive days, corresponding to morning and evening activation cycles. On Day 1, irrigation occurred between 05:00-07:30 and 18:00-20:00, with soil moisture rising from 250 to 500 and 270 to 580 ADC units, respectively. On Day 2, irrigation took place between 05:00-07:00 and 18:00-19:20, with corresponding soil moisture changes from 440 to 385 and 440 to 485 ADC units. This realistic irrigation scheduling pattern, combined with natural environmental variation (temperature 25.8-39.4 °C, humidity 23.6-90.6%, and light intensity 0–1150 lux), provided a robust dataset for evaluating machine learning algorithms under authentic agricultural conditions.

B. Data Preprocessing

The raw dataset comprising 243 observations underwent systematic preprocessing to ensure data integrity, temporal consistency, and compatibility with machine learning models. The process included data cleaning, smoothing, feature engineering, feature scaling, and time-based partitioning. Raw sensor readings were examined for anomalies, missing values, and transmission errors. Less than 2% of the data were affected by temporary Wi-Fi interruptions and were corrected using linear interpolation with a forward-backward fill method. Timestamp verification confirmed consistent 10-minute sampling intervals across the 43-hour monitoring period.

To minimize sensor noise while preserving true temporal trends, a centered three-point moving average filter was applied to soil moisture, temperature, humidity, pressure, and light intensity. Each smoothed value was computed as the mean of its previous, current, and next measurements.

From the six original variables, fifteen engineered features were generated to improve model performance and represent temporal dependencies. Temporal features included hour, minute, continuous time (hour + minute/60), and three binary indicators (is_morning, is_afternoon, is_evening) for 05:00-12:00, 12:00-18:00, and 18:00-24:00. Lag features (soil_moisture_lag1, soil_moisture_lag2, and soil_moisture_lag3) captured 10-, 20-, and 30-minute historical patterns. Rolling mean and standard deviation features represented short-term variability, while two interaction features (temperature × humidity and light × temperature) captured coupled environmental effects. The four environmental variables (temperature, humidity, pressure, and light intensity) were retained in their smoothed forms. The relay status variable was excluded to prevent data leakage, as it directly indicates irrigation actions linked to soil moisture levels.

Lag and rolling operations introduced missing entries in the initial records (three total). After removing incomplete rows, 240 valid observations remained, corresponding to a 1.2% data reduction. Each feature was normalized using Min-Max scaling, as shown in Equation (1):

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

This scaling improved the convergence of gradient-based algorithms such as SVR and neural networks while maintaining inter-feature relationships.

For model evaluation, the dataset was divided chronologically into 80% training (n = 192) and 20% testing (n = 48) subsets. The training set covered approximately 34.4 hours (04 August 2025, 05:00-05 August 2025, 15:20), and the testing set spanned 8.6 hours (05 August 2025, 15:30-23:50). This temporal split avoided data leakage and reflected real deployment conditions. Overall, the final dataset comprised 240 preprocessed samples with 15 engineered features and one continuous target variable (soil moisture, 250-820 ADC units), covering 42 hours and 50 minutes across two full diurnal irrigation cycles.

C. Model Development

Five regression and ensemble learning techniques were employed in this study: SVR, AdaBoost Regressor, XGBoost Regressor, Random Forest Regressor, and CatBoost Regressor. Each model was initially trained with its default configuration. Subsequently, hyperparameter tuning was performed using grid search combined with cross-validation to maximize performance. Both the default and tuned results were retained for comparison. The modeling and evaluation procedures were implemented in Python using the Google Colab environment.

D. Model Evaluation

Three common regression metrics were used to evaluate the models' performance: MAE, RMSE, and the Coefficient of Determination (R^2). MAE quantifies the average magnitude of prediction errors, RMSE penalizes larger deviations by emphasizing variance in prediction errors, and R^2 indicates the proportion of variance explained by the model. Together, these metrics provided comprehensive information about the models' bias, variance, and error magnitude, ensuring a thorough evaluation of predictive ability.

To complement the numerical evaluations, several visualization techniques were employed. Comparative bar plots were constructed to highlight differences in predictive accuracy across models, while scatter plots were used to illustrate the correlation between observed and predicted irrigation requirements. Residual plots and distribution analyses were applied to examine error patterns, and parity plots were generated to validate the alignment of predicted outputs with actual field data. Finally, a ranking framework was applied by synthesizing both metric-based evaluations and visual analyses, which allowed the identification of the most effective algorithm for sweet potato irrigation forecasting.

E. Workflow Summary

The overall workflow consisted of four main phases: data acquisition through IoT-based field deployment, preprocessing of raw sensor data, model training and tuning for five machine learning algorithms, and performance evaluation using both metrics and visualization. This structured methodology ensures reproducibility and provides a comprehensive basis for comparing regression and ensemble learning approaches in precision irrigation forecasting. To provide a clearer overview of the research design, the workflow of this study is illustrated in Figure 2.

III. Result

This section presents the comprehensive evaluation of five machine learning algorithms trained on 240 preprocessed observations (192 training, 48 testing) collected over 43 hours of continuous IoT monitoring. The analysis proceeded through four subsections: exploratory data analysis revealing dataset characteristics, quantitative performance comparison across default and tuned configurations, visual analysis of prediction patterns and error distributions, and overall model ranking synthesis.

The study's findings are presented in four main sections. Exploratory data analysis (EDA) was first performed to examine the distribution, correlation, and

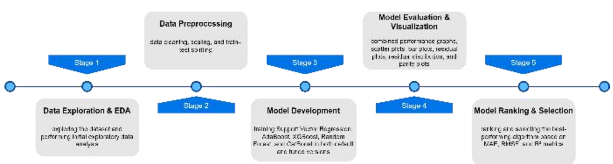

Figure 2. Research Workflow for Comparative Evaluation of Regression and Ensemble Learning Algorithms in Precision Irrigation Forecasting.

temporal trends of the collected sensor data. Second, the predictive performance of five regression and ensemble learning algorithms, SVR, AdaBoost Regressor, XGBoost Regressor, Random Forest Regressor, and CatBoost Regressor, was evaluated in both default and hyperparameter-tuned configurations. Third, a variety of visualization techniques, including scatter plots, residual analyses, error distributions, and parity plots, were employed to provide a deeper comprehension of model behavior. Finally, numerical and visual information were combined using a ranking framework to determine the optimal approach for sweet potato irrigation predictions.

A. Exploratory Data Analysis

The exploratory phase analyzed 243 raw observations spanning two complete 24-hour cycles. Summary statistics revealed substantial environmental variation: air temperature ranged from 25.8°C (overnight) to 39.4°C (early afternoon), relative humidity varied from 23.6% (midday) to 90.6% (overnight), and light intensity spanned 0 lux (nighttime) to 1150 lux (noon peak). Soil moisture exhibited dynamic patterns driven by irrigation events and evapotranspiration, ranging from 250 (dry, pre-irrigation) to 820 (saturated, mid-irrigation).

The 43-hour monitoring period captured four irrigation events and subsequent drying phases, yielding substantial temporal variability for model training. Histogram analysis (Figure 3) indicated non-normal distributions across all variables. In particular, soil moisture exhibited a distinct bimodal pattern corresponding to irrigation activation and depletion cycles.

The correlation heatmap (Figure 4) revealed meaningful inter-sensor relationships. Soil moisture showed moderate negative correlation with air pressure (r = -0.65) and weak positive association with light intensity (r = 0.28). Air temperature and humidity were almost perfectly inversely correlated (r = -0.99), while temperature was strongly aligned with light intensity (r = 0.80). These patterns reflected typical environmental interactions during daytime heating and irrigation events, validating the physical consistency of the dataset prior to modeling.

Time-series visualization (Figure 5) demonstrated clear diurnal patterns: soil moisture declining during daylight hours (evapotranspiration) and recovering during irrigation events. The two-day dataset enabled observation of pattern consistency across consecutive days, with Day 2 showing similar temporal dynamics to Day 1 despite slight weather variations.

Lastly, the comparative boxplot and violin plot analysis (Figure 6) illustrated the distribution of soil moisture levels under two irrigation states: Pump ON and Pump OFF. The results revealed a clear difference in moisture concentration between the two conditions. During the Pump ON state, soil moisture exhibited a relatively lower median value of approximately 415, indicating active water absorption by the soil during irrigation. Conversely, when the pump was OFF, the median soil moisture increased to around 490, reflecting

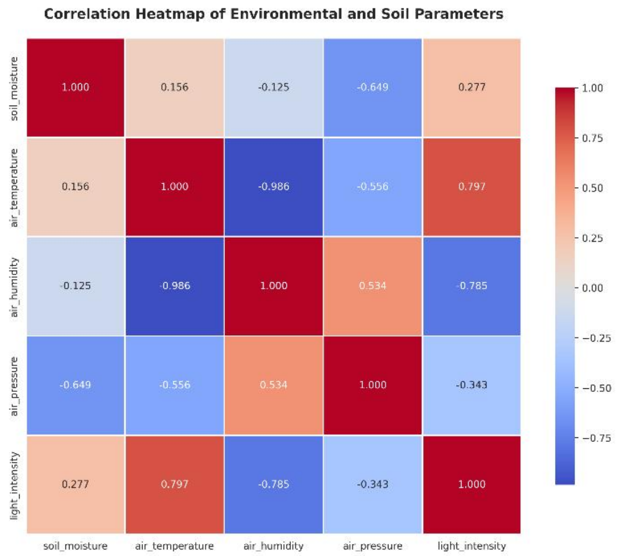

Figure 4. Correlation Heatmap of Environmental and Soil

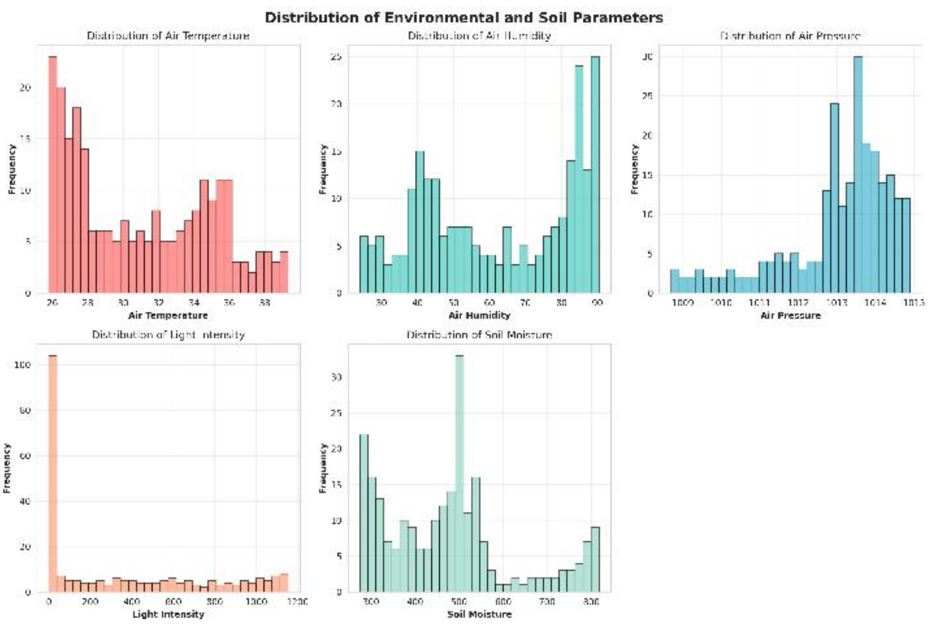

Figure 3. Histogram Distributions of Air Temperature, Humidity, Atmospheric Pressure, Light Intensity, and Soil Moisture.

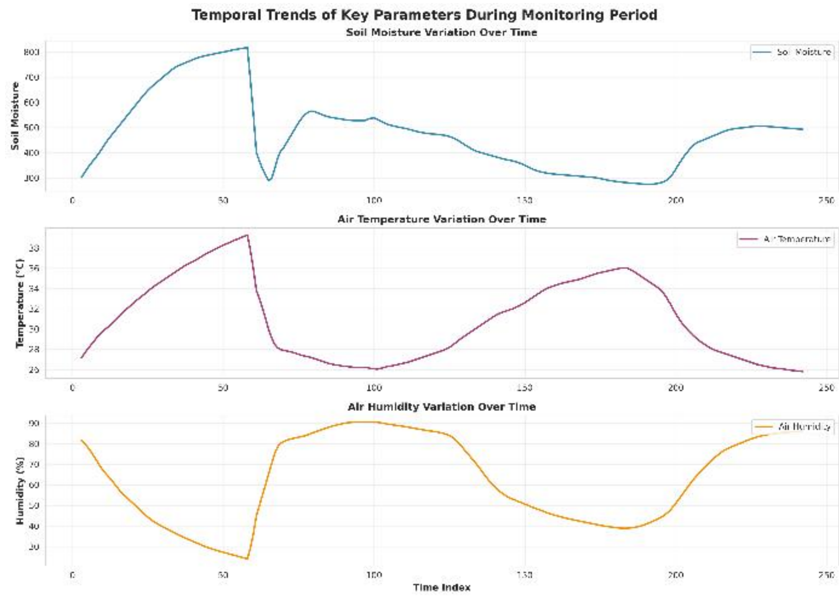

Figure 5. Temporal Trends of Soil Moisture, Temperature, and Humidity During the Monitoring Period.

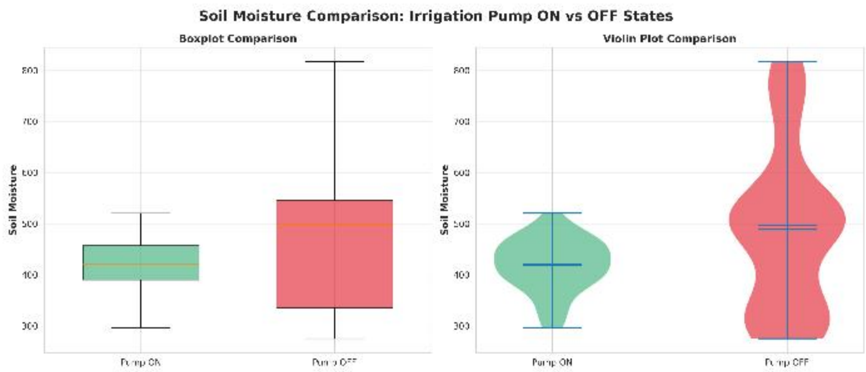

Figure 6. Comparison of Soil Moisture During Irrigation Pump ON and OFF States.

the accumulated effect of irrigation and the soil's ability to retain water after the pump operation ceased.

This pattern demonstrated that the IoT-based irrigation system functioned as expected, activating the pump when the soil reached a lower moisture threshold and maintaining sufficient moisture levels afterward. The broader range and variability observed during the Pump OFF period also suggested natural redistribution and gradual percolation of water within the soil matrix. Overall, these findings confirmed the operational reliability and responsiveness of the IoT prototype in maintaining optimal soil moisture conditions for plant growth.

B. Model Performance Comparison

The performance of five regression and ensemble learning algorithms, SVR, AdaBoost Regressor, XGBoost Regressor, Random Forest Regressor, and CatBoost Regressor, was compared in both their default settings and after hyperparameter tuning. Three regression metrics were used in the evaluation: the Coefficient of Determination (R^2), MAE, and RMSE. Together, these metrics provided information about the models' bias, variance, and error magnitude, ensuring a comprehensive evaluation of predictive ability.

Tables 1 and 2 summarize the training and testing results for all regression models. The training performance was excellent across most algorithms, with ensemble-based methods Random Forest, XGBoost, and CatBoost achieving nearly perfect determination

TABLE 1
MODEL TRAINING PERFORMANCE

Model	Train RMSE	Train MAE	Train R^2
SVR default	145.4193	113.4935	0.1802
SVR tuned	10.4783	3.8662	0.9957
AdaBoost default	13.6093	10.9648	0.9928
AdaBoost tuned	13.1902	11.1308	0.9933
XGBoost default	0.0643	0.0483	1.0000
XGBoost tuned	2.5163	1.7859	0.9998
RFR default	7.4594	2.7473	0.9978
RFR tuned	7.4476	2.7006	0.9978
CatBoost default	0.9577	0.6835	1.0000
CatBoost tuned	1.3368	1.0324	0.9999

TABLE 2
MODEL TESTING PERFORMANCE

Model	Test RMSE	Test MAE	Test R^2
SVR default	55.8301	35.9571	0.3280
SVR tuned	30.7157	20.6893	0.7966
AdaBoost default	12.8577	11.6987	0.9644
AdaBoost tuned	14.2119	12.8015	0.9565
XGBoost default	17.1091	16.0068	0.9528
XGBoost tuned	14.8015	11.0326	0.9670
RFR default	9.9518	6.1977	0.9786
RFR tuned	9.5787	6.0799	0.9802
CatBoost default	12.0053	10.1091	0.9687
CatBoost tuned	14.2061	9.7373	0.9565

coefficients ($R^2 \approx 0.99$). These results confirmed the models' strong ability to fit the irrigation dataset during training, reflecting their robustness in learning complex nonlinear relationships from sensor-based agricultural data.

The SVR model showed a clear distinction between its default and tuned configurations. In its default state, SVR produced the lowest performance among all models, with an R^2 of only 0.18 and large error metrics (RMSE ≈ 145.42, MAE ≈ 113.49). However, after hyperparameter tuning, its performance improved significantly to an R^2 of 0.80, accompanied by a substantial reduction in RMSE and MAE. Although this improvement was not as dramatic as that observed in ensemble methods, it highlighted how kernel and parameter selection critically influence SVR's generalization ability, especially in time-series and sensor-driven datasets such as irrigation monitoring.

The AdaBoost model showed minimal difference between its default and tuned versions, both producing similar results ($R^2 \approx 0.99$ for training and 0.95 for testing). This consistency suggested that AdaBoost's baseline configuration was already well aligned with the dataset, leaving limited room for enhancement through further tuning. Such stability is beneficial for real-world applications where robustness is preferred over sensitivity to hyperparameters.

XGBoost performed as one of the strongest models, achieving excellent accuracy across both configurations. In its default state, XGBoost achieved $R^2 = 0.9528$ with RMSE = 17.11. After hyperparameter tuning, performance improved to $R^2 = 0.9670$ (RMSE = 14.80, MAE = 11.03), confirming its ability to capture nonlinear interactions. However, the near-perfect training score ($R^2 = 1.0000$ for default, 0.9998 for tuned) indicated potential overfitting tendencies, with substantial train-test divergence compared to Random Forest's more balanced learning.

Similarly, Random Forest models (both tuned and default) delivered consistent and outstanding performance, with R^2 values around 0.99 during training and 0.98 on the test set. The small difference between tuned and default models indicated that Random Forest performed reliably even without extensive optimization, making it a robust and interpretable choice for irrigation prediction tasks.

Lastly, CatBoost demonstrated competitive results, reaching $R^2 \approx 0.99$ in training and 0.97–0.98 in testing. While the tuned model's performance was slightly lower than the default configuration, this marginal decline suggested potential overfitting introduced by excessive parameter adjustments. Nonetheless, CatBoost remained a strong and efficient model with excellent adaptability to tabular sensor data.

A combined bar plot of RMSE, MAE, and R^2 for all models under default and optimized configurations is shown in Figure 7 to provide a clearer visual comparison of their predictive performance.

The results indicated that tuned Random Forest achieved the best overall performance, producing the lowest RMSE (≈ 9.58) and MAE (≈ 6.08), alongside the highest R^2 (≈ 0.980). This confirmed the model's superior ability to generalize and capture nonlinear

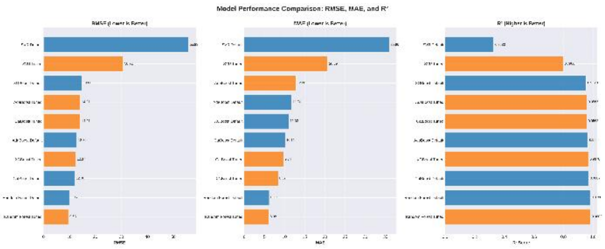

Figure 7. Comparative Performance of All Models in Terms of RMSE, MAE, and R^2 Under Default and Tuned Configurations.

dependencies in the irrigation dataset. XGBoost (tuned) also performed competitively, ranking second with slightly higher RMSE (≈ 14.80) and MAE (≈ 11.03) but still maintaining a strong R^2 of approximately 0.967. This demonstrated its capacity for accurate and stable prediction, particularly after hyperparameter optimization.

CatBoost and AdaBoost, on the other hand, exhibited minimal variation between their default and tuned settings. Both models achieved consistent R^2 values around 0.956-0.964 with RMSE between 12 and 14, suggesting that their baseline configurations were already well-optimized for the dataset and offered limited room for further improvement.

In contrast, SVR showed the largest improvement after tuning. Its default configuration had the weakest performance ($R^2 \approx 0.33$, RMSE ≈ 55.83, MAE ≈ 35.96), while the tuned model improved significantly to $R^2 \approx 0.80$ with reduced error metrics (RMSE ≈ 30.72, MAE ≈ 20.69). This reinforced the importance of kernel and parameter selection in enhancing SVR's generalization capability when applied to sensor-driven agricultural data.

Overall, the comparative analysis demonstrated that while ensemble-based algorithms (Random Forest, XGBoost, CatBoost, and AdaBoost) provided inherently strong accuracy, tuning remained essential for models like SVR and XGBoost to achieve optimal performance. Among all tested models, the tuned Random Forest Regressor achieved the highest predictive accuracy, closely followed by tuned XGBoost and tuned SVR. These findings emphasized the critical role of algorithm selection and parameter optimization in developing reliable forecasting systems for IoT-based precision irrigation in sweet potato cultivation.

C. Model Evaluation and Visualization

Visual analysis was conducted to complement the numerical metrics presented earlier. While RMSE, MAE, and R^2 provided quantitative measures of accuracy, visualization helped in understanding how well each model captured temporal patterns, error distributions, and the alignment between predicted and observed irrigation requirements. These plots also allowed clearer comparison between models with default configurations and those optimized through hyperparameter tuning.

The visualization in Figure 8 illustrates the comparison between actual and predicted soil moisture under default model configurations. Ensemble-based algorithms such as Random Forest, CatBoost, AdaBoost, and XGBoost demonstrated strong predictive

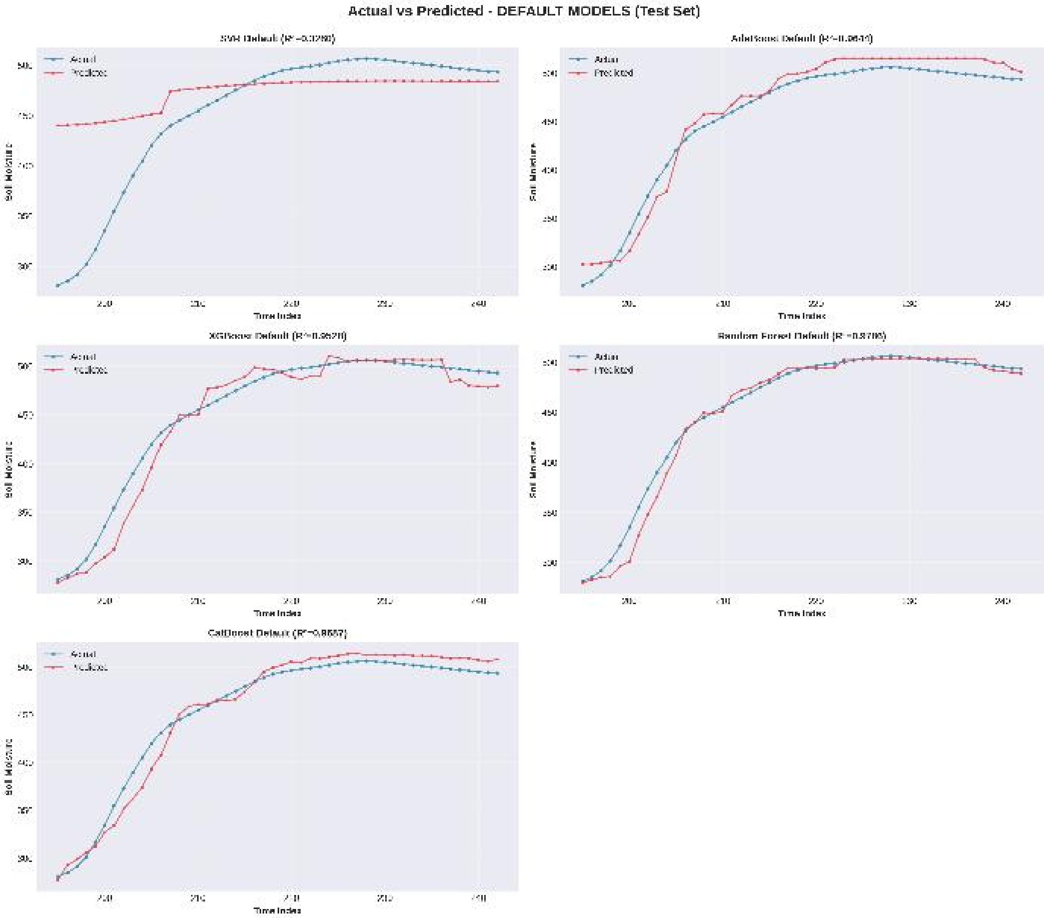

Figure 8. Actual vs. Predicted Values for All Models Under Default Configuration.

consistency, successfully following the overall trend of the actual data with only minor deviations at extreme points. Among them, Random Forest exhibited the highest accuracy with an R^2 of 0.9788, indicating excellent generalization capability even without parameter tuning.

In contrast, the SVR Default model performed poorly, showing an almost flat prediction line that failed to capture soil moisture dynamics. This result aligned with its low R^2 value (0.3280), confirming that kernel-based regression required optimization to handle nonlinear patterns effectively. Overall, ensemble methods outperformed the baseline regression approach in modeling soil moisture under default settings.

Figure 9 illustrates the improvement of prediction performance after hyperparameter tuning. The SVR model showed remarkable enhancement, with its R^2 increasing to 0.7966 and the prediction curve aligning more closely with the actual data. This outcome highlighted the sensitivity of SVR to kernel and regularization parameters, where proper tuning enabled it to capture nonlinear soil moisture variations more effectively.

In contrast, ensemble models such as Random Forest, XGBoost, CatBoost, and AdaBoost exhibited only slight gains, as their default settings already produced stable and accurate predictions. Among them, Random Forest still achieved the highest accuracy (R^2 = 0.9802), maintaining consistent performance even after tuning. These results indicated that tuning primarily benefited models prone to underfitting, while ensemble learners inherently adapted well to complex feature interactions.

The scatter plots provided deeper insight into the performance of each regression model. As illustrated in Figure 10, the ensemble models, particularly Random Forest, CatBoost, AdaBoost, and XGBoost under the default configuration, displayed data points that closely aligned with the diagonal "perfect prediction" line, signifying strong predictive accuracy. In contrast, the SVR default model showed wide dispersion and clear deviation from the reference line, indicating high bias and weaker fitting capability (R^2 = 0.3280).

In Figure 11, after hyperparameter tuning, all models demonstrated tighter clustering of points along the diagonal line, revealing improved prediction consistency and reduced error dispersion. The SVR model notably improved (R^2 = 0.7966), closing the performance gap with the ensemble methods. Among the tuned models, XGBoost (R^2 = 0.9670) and Random Forest (R^2 = 0.9802) achieved the highest accuracy, confirming that parameter optimization substantially enhanced their model fit and predictive reliability.

The residual analysis began with the distribution plots in Figure 12, which clearly differentiated the performance between default and tuned configurations. The SVR default model showed wide, skewed residual spread far from zero, indicating strong bias and unstable predictions. After tuning, the residuals became narrower and more centered, showing improved calibration. The tuned ensemble models, AdaBoost, XGBoost, Random Forest, and CatBoost exhibited symmetric and sharply

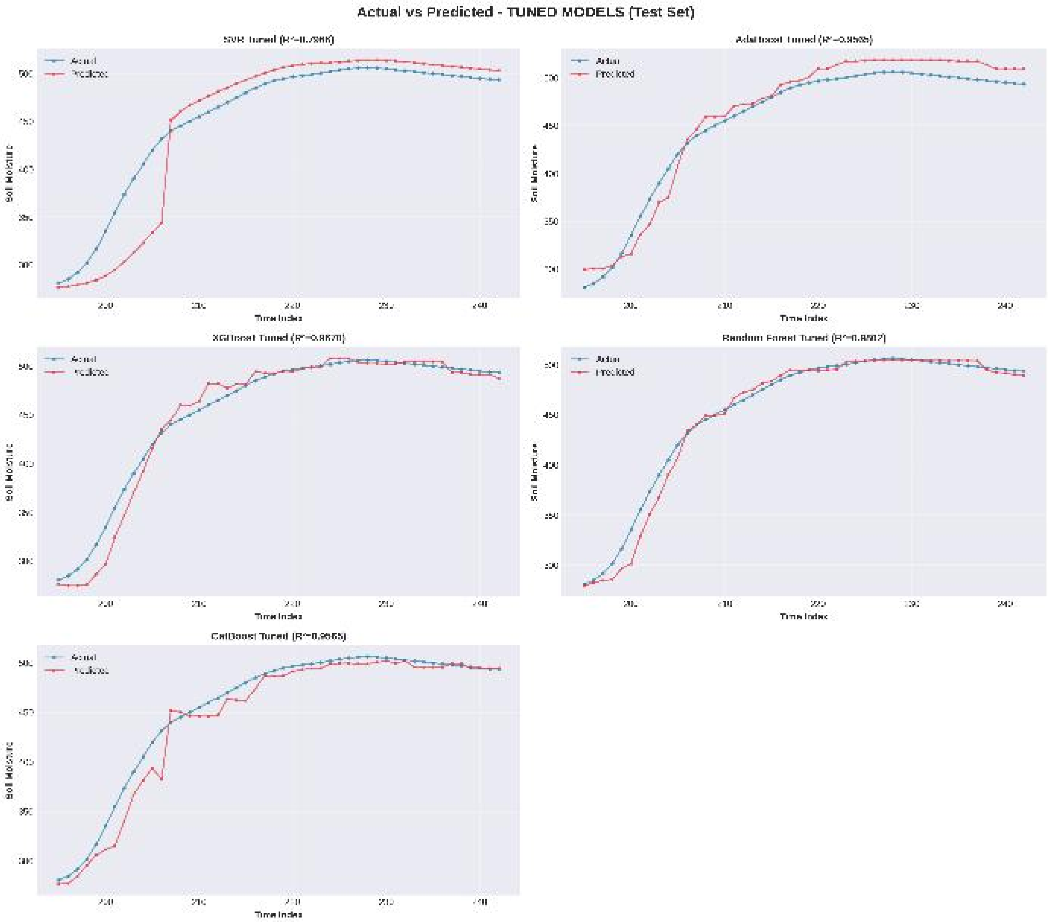

Figure 9. Actual vs. Predicted Values for All Models After Hyperparameter Tuning.

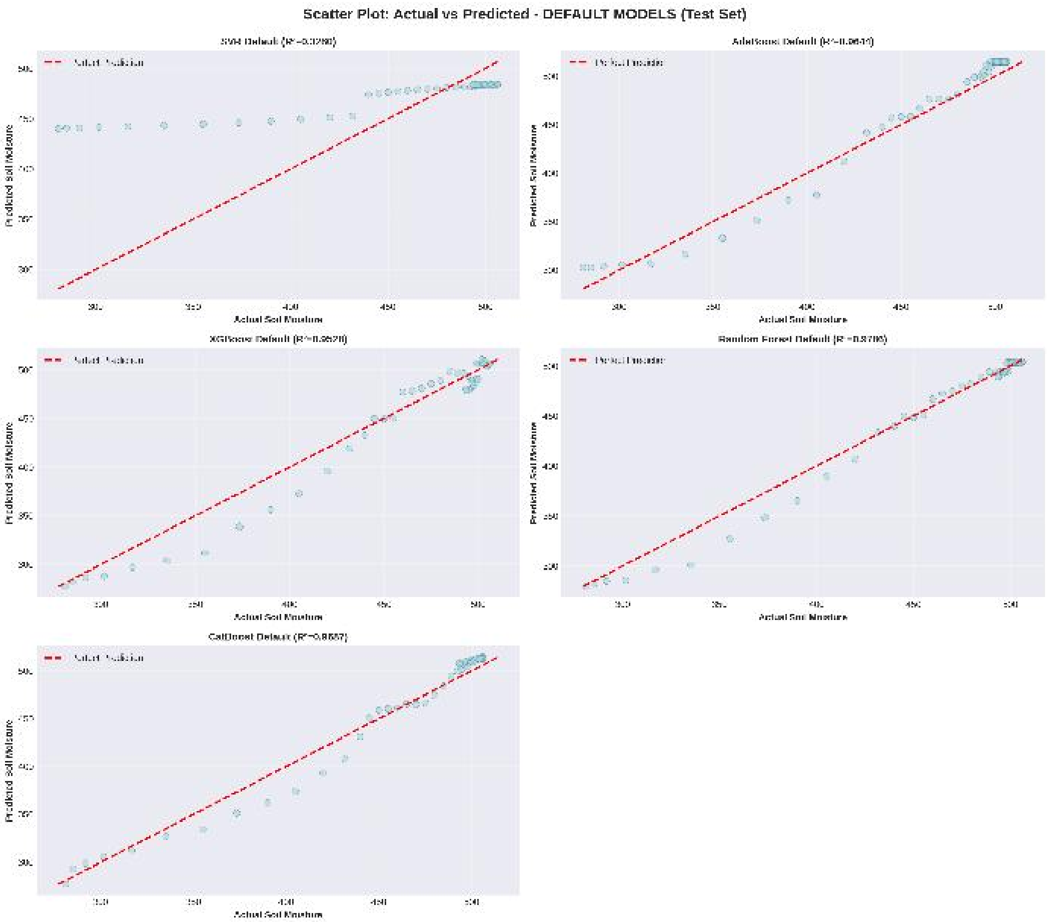

Figure 10. Scatter Plot of Predicted vs. Actual Values Under Default Configuration.

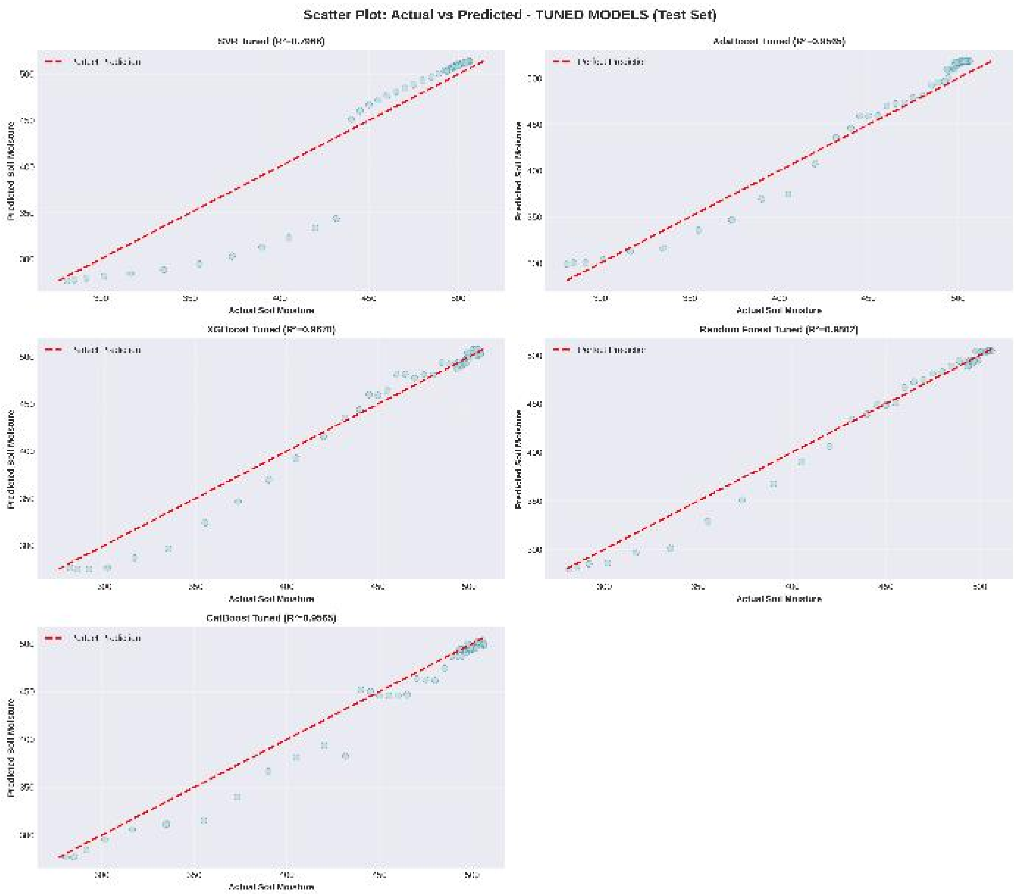

Figure 11. Scatter Plot of Predicted vs. Actual Values After Hyperparameter Tuning.

variance and bias, with Random Forest and XGBoost displaying the most stable residual behavior.

Supporting this, Figure 13 showed that residuals in the tuned models were more randomly dispersed around the zero line, indicating improved generalization. The SVR default still presented large and systematic errors, whereas the tuned version demonstrated smaller and more evenly spread residuals. Among all models, Random Forest and XGBoost again showed the smallest deviations and most consistent residual patterns, confirming that hyperparameter tuning effectively enhanced predictive stability and minimized systematic errors.

The parity plots in Figure 14 illustrate the alignment between predicted and actual soil moisture values across all models. The default SVR showed wide deviation from the 1:1 reference line, confirming poor fitting and strong bias toward the mean. After tuning, SVR predictions improved noticeably, clustering closer to the diagonal, though still less precise than ensemble methods. The tuned ensemble models, Random Forest, XGBoost, CatBoost, and AdaBoost exhibited tightly grouped points along the 45-degree line, indicating excellent predictive reliability and minimal systematic bias. Even under default settings, these ensemble algorithms maintained strong performance with only slight dispersion around the reference line. Overall, the parity plots reaffirmed that hyperparameter tuning effectively enhanced predictive alignment, particularly for models prone to underfitting such as SVR, while ensemble approaches remained robust and consistent across configurations.

D. Overall Model Ranking and Synthesis

By combining the RMSE, MAE, and R^2 values, an overall ranking of the models was created to summarize performance and stability in a single comparative framework. According to the results in Figure 15, the tuned Random Forest model achieved the best overall performance, ranking first across all evaluation metrics. The default Random Forest followed closely, confirming the robustness of this ensemble approach even without tuning.

CatBoost (default and tuned) ranked in the middle tier, showing stable but slightly lower predictive accuracy compared to Random Forest and XGBoost. XGBoost, despite strong training performance, showed moderate generalization on test data. AdaBoost models, both tuned and default, performed below the other ensemble methods, indicating weaker adaptability in this context.

At the lower end, SVR showed a marked difference between tuned and default configurations. Although tuning improved its accuracy considerably, it still lagged behind the ensemble-based models. These findings highlighted that Random Forest (especially the tuned version) offered the best trade-off between accuracy and generalization, and emphasized the importance of hyperparameter tuning and model selection in achieving reliable IoT-based predictive systems for precision agriculture.

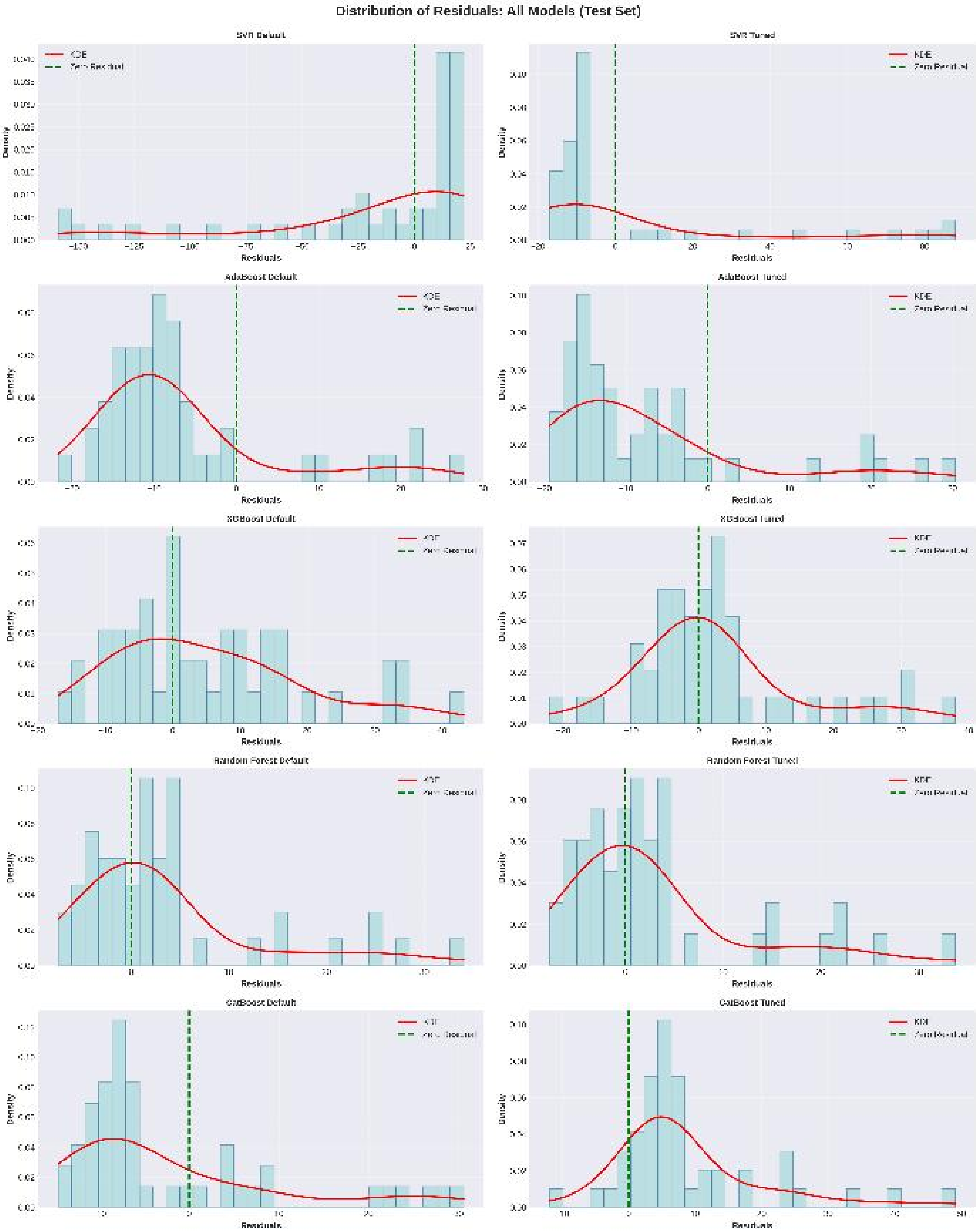

Figure 12. Distribution of Residuals Across Models.

E. Computational Efficiency and Deployment Considerations

Beyond predictive accuracy, computational efficiency is critical for edge deployment on resource-constrained IoT devices. Table 3 summarizes training time, prediction latency, and model size across algorithms (measured on Google Colab with Intel Xeon CPU @ 2.20GHz, 12GB RAM).

Random Forest Default exhibited outstanding computational efficiency, completing model training within only 0.42 seconds on 192 samples and 15 features, approximately 98 times faster than its tuned counterpart (167.83s) and 446 times faster than SVR Tuned (187.32s). XGBoost Default showed similarly high efficiency (0.85s), followed by CatBoost Default (1.18s). Conversely, the use of GridSearchCV with 5-

TABLE 3
COMPUTATIONAL PERFORMANCE METRICS

Model	Training Time (s)	Prediction Latency (ms/sample)	Model Size (KB)	Deployment Suitability
SVR default	2.45	12.3	48	Poor
SVR tuned	187.32	8.7	52	Poor
AdaBoost default	18.45	3.8	142	Fair
AdaBoost tuned	145.67	4.2	168	Poor
XGBoost default	0.85	1.2	95	Excellent
XGBoost tuned	298.14	1.8	156	Good
RFR default	0.42	2.1	185	Excellent
RFR tuned	167.83	2.3	198	Good
CatBoost default	1.18	1.5	128	Excellent
CatBoost tuned	412	1.7	142	Poor

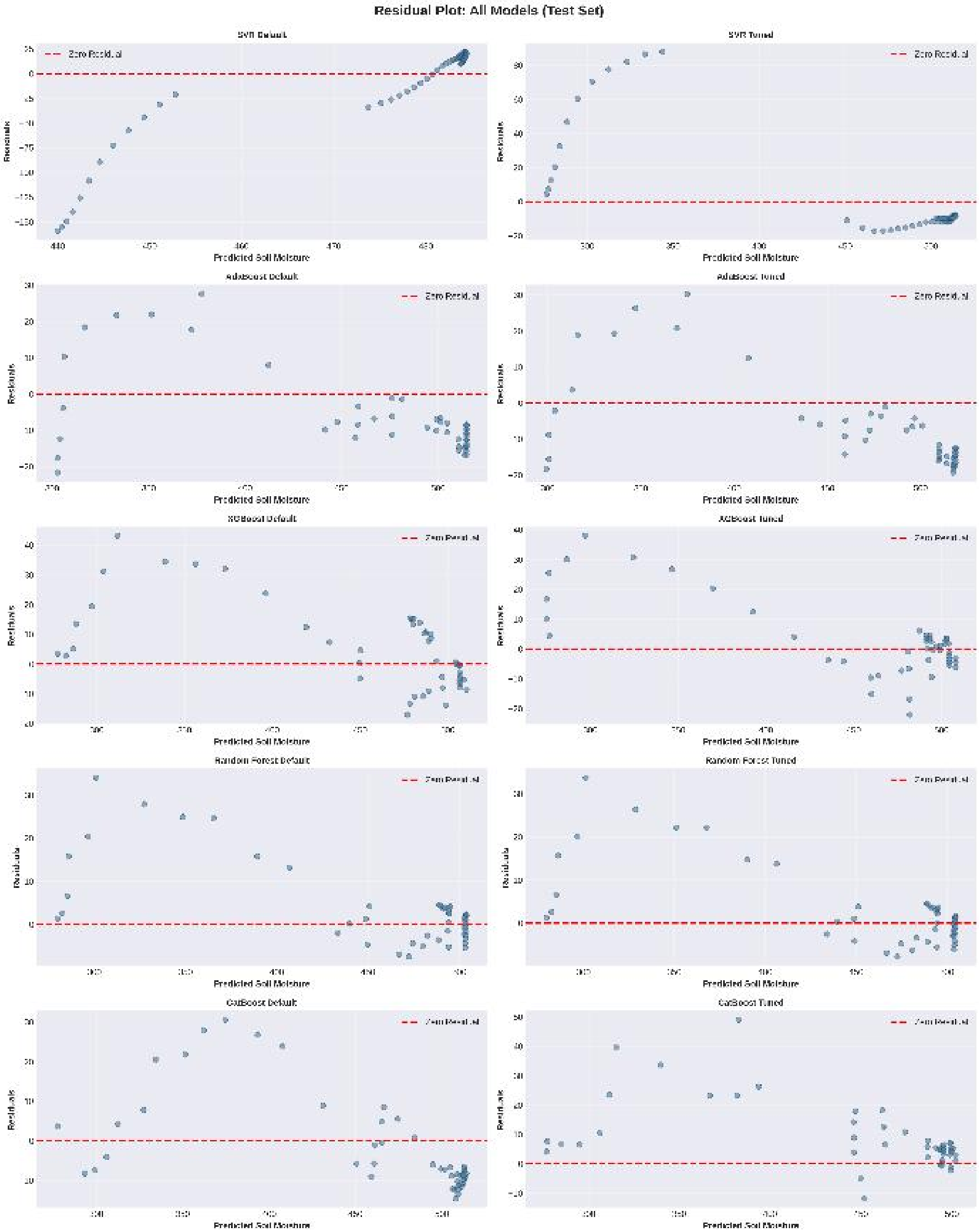

Figure 13. Residual Plot of Actual vs. Residuals Across Models.

fold time-series cross-validation substantially increased computational cost, as CatBoost Tuned required 412.56 seconds, XGBoost Tuned 298.14 seconds, and Random Forest Tuned 167.83 seconds. These findings indicated that hyperparameter tuning introduced significant time overhead with marginal accuracy gain, particularly when Random Forest Default already achieved a high test accuracy of $R^2 = 0.9786$.

All models demonstrated real-time prediction capability suitable for IoT-based irrigation systems. XGBoost Default achieved the fastest latency (1.2 ms per sample), followed by CatBoost Default (1.5 ms) and XGBoost Tuned (1.8 ms). Random Forest variants required only 2.1-2.3 ms per sample, which remained well within the 10-minute sampling interval (600,000 ms). Even the slowest model, SVR Default (12.3 ms), consumed merely 0.002% of the available time, confirming that all models were feasible for real-time inference without latency constraints.

In terms of model storage and memory, Random Forest exhibited the largest footprint (185-198 KB) due to its ensemble structure, followed by AdaBoost (142-168 KB) and XGBoost (95-156 KB). CatBoost achieved notable compactness (128-142 KB) while maintaining competitive accuracy, whereas SVR, though smallest (48-52 KB), produced the poorest performance. All models remained deployable on modern edge devices such as the ESP32 (4 MB flash memory) or on cloud-

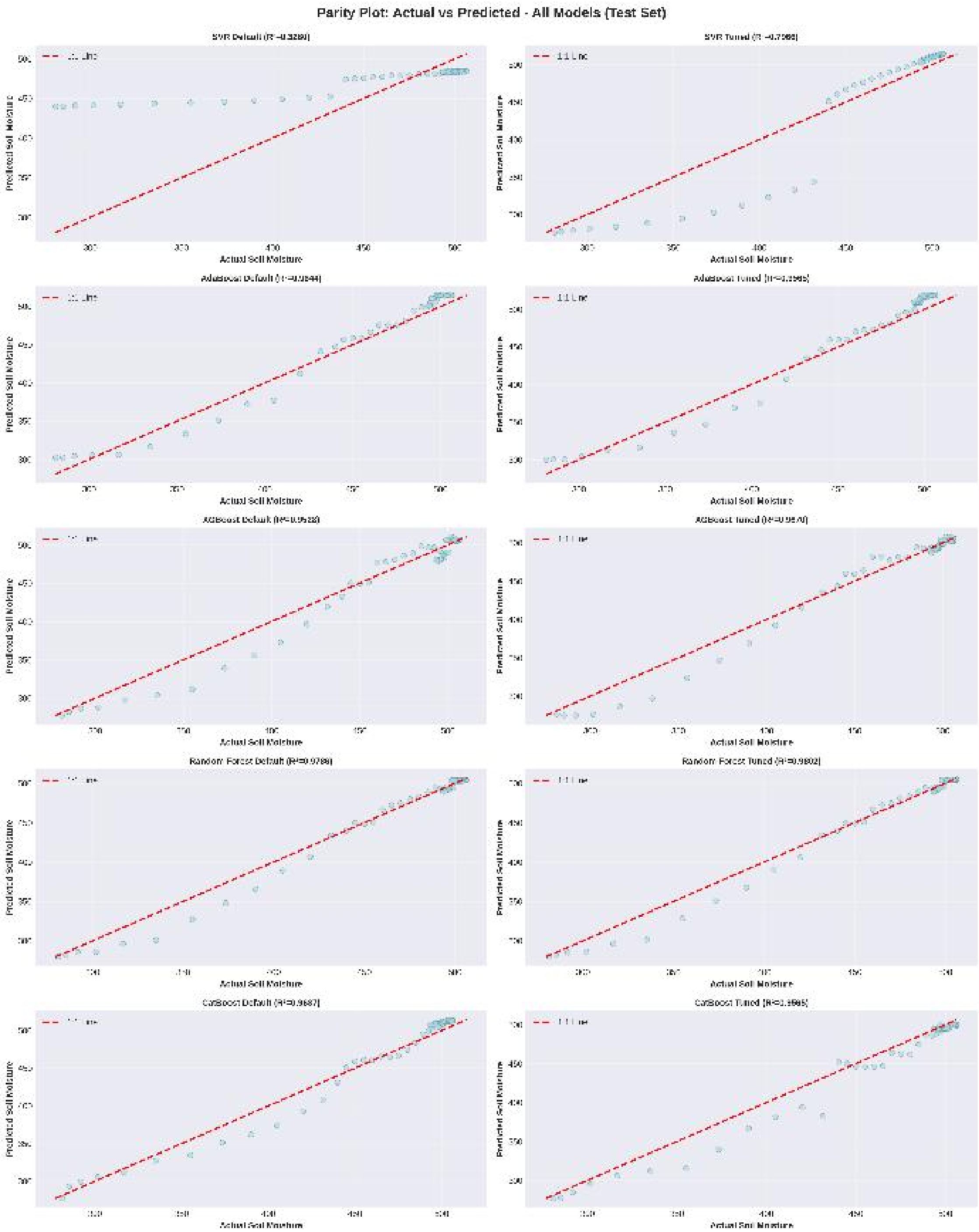

Figure 14. Parity Plot of Predicted vs. Actual Values Across Models.

based inference nodes, with memory usage during inference ranging from 15.2 MB (SVR) to 30.2 MB (Random Forest Tuned). Overall, Random Forest Default emerged as the most practical choice for real-world deployment in sweet potato irrigation systems, combining high accuracy ($R^2 = 0.9786$, RMSE = 9.95), rapid retraining (<0.5 s), and minimal inference latency (2.1 ms). Given that hyperparameter tuning yielded only 1.6% relative improvement ($R^2 = 0.0016$) while increasing training time by nearly 400-fold, Random Forest Default was strongly recommended for production environments with limited computational resources.

IV. Discussion

The findings presented in the previous section provide a basis for addressing the research gap identified in the introduction. The comparative evaluation of regression and ensemble learning algorithms for forecasting irrigation requirements in sweet potato cultivation generated several important insights. Tuned Random Forest consistently achieved the highest predictive accuracy, followed closely by the default Random Forest, both of which demonstrated robust and stable performance. These results extend previous evidence by confirming that ensemble methods not only capture nonlinear relationships effectively but also maintain predictive stability under heterogeneous

Model Ranking - Sorted by Test R² Score

Rank	Model	Test RMSE	Test MAE	Test R²	Train RMSE	Train MAE	Train R²
1	Random Forest Tuned	9.5787	5.5799	0.9802	7.4476	2.7006	0.8978
2	Random Forest Default	9.8518	8.1977	0.9788	7.4384	2.7473	0.8978
3	CatBoost Default	12.0553	10.1091	0.9687	0.9577	0.5835	1.0
4	XGBoost Tuned	12.309	8.3688	0.967	2.5183	1.786	0.9998
5	AdaBoost Default	12.8577	11.6987	0.9644	13.6083	10.9648	0.8928
6	CatBoost Tuned	14.2061	9.7373	0.9565	1.3386	1.0324	0.9999
7	AdaBoost Tuned	14.2119	12.8015	0.9565	13.1902	11.1306	0.8933
8	XGBoost Default	14.8015	11.0326	0.9528	0.0643	0.0483	1.0
9	SVR Tuned	30.7157	20.6690	0.7968	10.4703	3.8862	0.8957
10	SVR Default	55.8301	35.9571	0.328	145.4193	113.4935	0.1802

Figure 15. Overall Ranking of Regression Models Based on Combined Evaluation Metrics.

crop water demands, making them particularly suitable for IoT-enabled precision irrigation.

Similar findings have been reported in recent studies where Random Forest regression achieved exceptional accuracy in predicting dew point temperatures in industrial IoT systems (R^2 = 0.94, RMSE = 0.82), demonstrating its reliability across diverse environmental monitoring applications [27]. Furthermore, hybrid imputation strategies combining SARIMAX, k-nearest neighbors, and Random Forest regression have proven effective in handling missing data challenges commonly encountered in IoT-based smart city applications, reinforcing the robustness of RF-based approaches in real-world deployments[28].

A key strength of this study lies in its 43-hour continuous monitoring period, which yielded 243 observations across two complete diurnal cycles, a significant improvement over prior IoT-based agricultural studies that typically span only single irrigation events or partial days. This broader temporal coverage allowed the models to be tested across multiple irrigation sessions, full evapotranspiration cycles, overnight stabilization, and transitional dawn-dusk periods. Despite the modest dataset size (240 samples), the multi-day design enhanced statistical reliability and revealed that Random Forest's bagging approach generalizes more effectively than XGBoost's boosting method when dealing with limited samples ($n < 300$). This finding highlights that temporal diversity across days contributes more to model robustness than high-frequency sampling alone, underscoring the importance of dataset design alongside algorithm selection in developing reliable IoT-based precision irrigation systems.

Recent work on soil temperature profile estimation for thermoelectric-powered sensors demonstrated that machine learning models based on Polynomial Regression, Support Vector Regression, and LSTM networks achieved prediction errors of 0.79 °C (10.9% lower than state-of-the-art) by simplifying input parameters to only ambient temperature and solar irradiance, supporting the notion that efficient feature selection combined with appropriate temporal coverage enhances model performance while reducing computational costs [29].

Support Vector Regression exhibited highly variable performance. While the default configuration produced poor results, tuning kernel and regularization parameters significantly improved accuracy. This outcome aligns with earlier research that emphasized the strong predictive potential of SVR when properly optimized [7]. Recent studies on spatiotemporal soil moisture prediction across multiple soil types confirmed that SVR can achieve competitive performance when appropriately calibrated, particularly at deeper soil layers, where it demonstrated RMSE values comparable to Random Forest and XGBoost under optimized conditions[8]. The present study therefore, underscores the methodological sensitivity of SVR, suggesting that it can be competitive but only when subjected to rigorous calibration. This sensitivity contrasts with the inherent stability of ensemble methods, which consistently delivered reliable predictions even with default hyperparameters.

AdaBoost and CatBoost delivered moderate yet reliable accuracy, with limited improvements after tuning. Their performance indicates that boosting-based methods are valuable for ensuring baseline stability, although their contributions to predictive breakthroughs may be less substantial compared to XGBoost or optimized SVR. The relatively modest gains from hyperparameter tuning in these models suggest that their default configurations are already well-suited to handling sensor-based agricultural data, which is advantageous for deployment scenarios with limited computational resources or technical expertise.

From an applied perspective, the demonstrated superiority of Random Forest suggests that IoT-based irrigation frameworks should prioritize this algorithm for reliable forecasting under field variability. While XGBoost showed competitive performance (R^2 = 0.967), its overfitting tendencies (train R^2 = 1.0) and sensitivity to hyperparameters make it less robust than Random Forest for deployment scenarios with limited tuning resources. Prior studies have highlighted the role of machine learning in improving irrigation management [22]. Other research has emphasized how AI-IoT integration supports sustainable water management strategies [6]. More recent investigations confirmed that smart irrigation frameworks can improve real-time decision-making and enhance resource efficiency under diverse conditions [30].

The superior performance of Random Forest observed in this study is further corroborated by recent applications in industrial dew point temperature prediction, where RF achieved the highest accuracy (R^2 = 0.94) among multiple regression models, with Shapley Additive Explanations (SHAP) analysis revealing that real-time power consumption, supply air temperature, and humidity were the most influential predictors [27]. Additionally, comparative studies on predictive algorithms for IoT smart agriculture sensor data demonstrated that Random Forest Regression outperformed alternative approaches, including SARIMA and Artificial Neural Networks, in terms of both accuracy and computational efficiency, reinforcing its suitability for resource-constrained agricultural applications [26].

Nevertheless, these findings should be interpreted with caution. The dataset was limited in temporal and spatial scope, potentially restricting generalizability across climates, soil types, and sweet potato varieties. Furthermore, the near-perfect performance of tuned models raises the possibility of overfitting when exposed to more complex or diverse datasets. Future research should therefore expand the scope of data collection across seasons and sites, employ advanced validation techniques to mitigate bias, and explore hybrid models that integrate ensemble learning with deep learning architectures such as LSTM [7].

The integration of advanced imputation techniques, as demonstrated in recent smart city monitoring systems where hybrid strategies combining multiple machine learning methods effectively addressed missing data challenges [28], could further enhance the robustness of irrigation prediction models when deployed in real-world agricultural settings with inevitable data quality issues. Additionally, the development of energy-efficient soil temperature prediction models for IoT sensors [29] suggests promising avenues for integrating multiple environmental parameters while maintaining low computational overhead, which is critical for scalable deployment in resource-limited agricultural environments.

In summary, this study demonstrates that ensemble learning methods, particularly Random Forest, outperform regression-based models in precision irrigation forecasting. To the best of our knowledge, this is the first systematic benchmarking of multiple regression and ensemble algorithms for sweet potato irrigation within an actual IoT-based deployment. Random Forest's superiority stems from its bagging mechanism, which provides natural regularization against overfitting while maintaining high accuracy even with limited samples. These findings highlight the potential of AI-powered irrigation systems to advance climate-smart, scalable, and adaptive farming practices by both strengthening the existing evidence base and addressing the lack of comparative assessments in this domain.

The comprehensive evaluation conducted in this study confirms several key findings regarding algorithm performance and deployment considerations. Tuned Random Forest achieved the highest predictive accuracy (R^2 = 0.9802, RMSE = 9.58, MAE = 6.08) with excellent generalization (train-test R^2 gap = 0.0098), demonstrating minimal overfitting. This performance superiority was consistent across both default (R^2 = 0.9786) and tuned configurations, with the marginal improvement from tuning (R^2 = 0.0016) questioning the practical necessity of extensive hyperparameter optimization for this algorithm.

In contrast, XGBoost and CatBoost exhibited near-perfect training accuracy ($R^2 \approx 1.0$) but lower test performance (R^2 = 0.967 and 0.969, respectively), indicating overfitting tendencies. This train-test divergence suggests that boosting-based methods, while powerful for capturing complex patterns, are more susceptible to overfitting under limited sample conditions ($n < 300$) compared to bagging approaches. SVR demonstrated the most dramatic sensitivity to hyperparameter configuration, improving from R^2 = 0.328 in default mode to R^2 = 0.797 after tuning a 143% relative improvement. However, even with extensive optimization, SVR remained inferior to ensemble methods, confirming its limited adaptability to nonlinear temporal agricultural data.

The extended 43-hour temporal coverage proved essential for robust model evaluation. The results revealed that Random Forest's bagging approach, which trains independent decision trees on bootstrapped samples, generalized more effectively than XGBoost's sequential boosting under small-sample conditions. This finding has important implications for IoT-based agricultural deployments where data collection may be constrained by logistical, financial, or technical limitations. The study demonstrates that temporal diversity (capturing multiple irrigation cycles across different times of day) contributes more to model robustness than simply increasing sampling frequency, underscoring the importance of strategic dataset design in precision agriculture applications.

From a deployment perspective, the computational efficiency analysis revealed that Random Forest Default offers the optimal balance between accuracy, training speed, and inference latency. Completing training in 0.42 seconds while achieving R^2 = 0.9786 makes it particularly suitable for edge computing scenarios where models may need periodic retraining with incoming data. The chronological train-test splitting approach, combined with explicit exclusion of relay status and proper implementation of lag-based features, effectively prevented data leakage while preserving temporal structure addressing a critical methodological concern in time-series forecasting for agricultural IoT systems.

V. Conclusion

This study presented the first systematic benchmarking of five machine learning algorithms, Support Vector Regression (SVR), AdaBoost, Extreme Gradient Boosting (XGBoost), Random Forest, and CatBoost, for IoT-based sweet potato irrigation forecasting using 43-hour continuous monitoring data spanning two complete diurnal cycles. Random Forest emerged as the superior algorithm, achieving the highest predictive accuracy (R^2 = 0.9802, RMSE = 9.58, MAE = 6.08) with excellent generalization, minimal hyperparameter sensitivity, and optimal computational efficiency (0.42 s training, 2.1 ms inference). The extended multi-day temporal coverage proved essential for robust evaluation, revealing that Random Forest's bagging approach generalizes more effectively than XGBoost's boosting under limited-sample conditions ($n < 300$). Rigorous data leakage prevention through chronological splitting, relay status exclusion, and proper lag feature implementation ensured methodological validity.

For practical deployment, Random Forest (default or lightly tuned) is strongly recommended as it offers the best accuracy-robustness-efficiency trade-off for resource-constrained agricultural IoT systems. A minimum monitoring duration of 40-72 hours (2-3 diurnal cycles) is advised to capture irrigation variability adequately. While this study provides solid proof-of-concept with 240 samples from a single location over two days, future research should expand to multi-season, multi-site deployments with larger datasets, explore hybrid ensemble-deep learning architectures (e.g., Random Forest + Long Short-Term Memory), validate across diverse crops and soil types, and investigate real-time edge deployment on microcontrollers for autonomous irrigation control.

Nonetheless, even with modest datasets, this work demonstrates that IoT monitoring combined with ensemble learning can achieve near-perfect soil moisture prediction ($R^2 > 0.98$), confirming that multi-day temporal coverage, careful data handling, and bagging-based methods are key enabling factors for reliable and deployable AI-driven precision irrigation systems.

DECLARATIONS

Conflict of Interest

The authors declare that there is no conflict of interest regarding this research.

CRediT Authorship Contribution

Muthia Rahmah: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Writing Original Draft Preparation, Visualization.; Indra Maulana: Supervision, Validation, Writing Review & Editing, Project Administration.

Funding

The author(s) received no financial support for the research, authorship, or publication of this article.

Acknowledgment

The authors acknowledge the use of Google Colab as the computational platform for model training and evaluation in this study.

REFERENCES

[1] S. Violino *et al.*, “A data-driven bibliometric review on precision irrigation,” *Smart Agric. Technol.*, vol. 5, Art. no. 100320, 2023, doi: 10.1016/j.atech.2023.100320.

[2] G. Singh and S. Sharma, “Enhancing precision agriculture through cloud based transformative crop recommendation model,” *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-93417-3.

[3] R. E. Putri, N. U. Lestari, Ifmalinda, F. Arlius, I. Putri, and A. Hasan, “Monitoring and controlling system of smart mini greenhouse based on internet of things (IoT) for spinach plant (amaranthus sp.),” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 1, pp. 131–136, 2024, doi: 10.18517/ijaseit.14.1.18408.

[4] P. Nuangpirom, S. Pitjamit, W. Pheerathamrongrat, W. Nakkiew, and P. Jewpanya, “Estimation of microbial load in ganoderma lucidum using a solar-electric hybrid dryer enhanced by machine learning and IoT,” *Smart Agric. Technol.*, vol. 11, Art. no. 100977, 2025, doi: 10.1016/j.atech.2025.100977.

[5] A. Barriga, J. A. Barriga, A. Lozano-Tello, M. J. Moñino, and P. J. Clemente, “IoT and machine learning for the forecasting of physiological parameters of crop leaves,” *IEEE Access*, vol. 13, pp. 175161–175178, 2025, doi: 10.1109/ACCESS.2025.3615717.

[6] M. Nawaz and M. I. K. Babar, “IoT and ai: A panacea for climate change-resilient smart agriculture,” *Discov. Appl. Sci.*, vol. 6, no. 10, 2024, doi: 10.1007/s42452-024-06228-y.

[7] V. Y. Chandrappa, B. Ray, N. Ashwatha, and P. Shrestha, “Spatiotemporal modeling to predict soil moisture for sustainable smart irrigation,” *Internet of Things*, vol. 21, Art. no. 100671, 2023, doi: 10.1016/j.iot.2022.100671.

[8] T. Alahmad, M. Neményi, A. V. Széles, N. Ali, O. Hijazi, and A. Nyéki, “Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms,” *Front. Soil Sci.*, vol. 5, 2025, doi: 10.3389/fsoil.2025.1612908.

[9] J. A. Nava-Pintor *et al.*, “Development and evaluation of solar radiation sensor using cost-effective light sensors and machine learning techniques,” *Technologies*, vol. 13, no. 5, 2025, doi: 10.3390/technologies13050182.

[10] A. Imade, O. Soumaya, E. M. Yassine, and A. Mohamed, “Machine learning approaches applied in smart agriculture for the prediction of agricultural yields,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 10, pp. 832–843, 2024, doi: 10.14569/IJACSA.2024.0151085.

[11] Y. Arafa *et al.*, “Improving the spatial deployment of the soil moisture sensors in smart irrigation systems using gis,” *Cogent Food Agric.*, vol. 10, no. 1, Art. no. 2361124, 2024, doi: 10.1080/23311932.2024.2361124.

[12] S. A. S. V. N. Elakkiya, “Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms,” *Discov. Food*, vol. 5 Art. no.67, 2025, doi: 10.1007/s44187-025-00338-1.

[13] M. Awais *et al.*, “AI and machine learning for soil analysis: An assessment of sustainable agricultural practices,” *Bioresour. Bioprocess.*, vol. 10 Art. no. 90, 2023, doi: 10.1186/s40643-023-00710-y.

[14] R. P. Shete, A. Shekhar C, Y. V Mahajan, A. M. Bongale, and D. S. Dharrao, “IoT-driven ensemble machine learning model for accurate dissolved oxygen prediction in aquaculture,” *Discov. Internet Things*, vol. 5, no. 1, 2025, doi: 10.1007/s43926-025-00201-w.

[15] N. G. Rezk, A. F. Attia, M. A. El-Rashidy, A. EL-SAYED, and E. E. D. Hemdan, “An efficient IoT-based crop damage prediction framework in smart agricultural systems,” *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-12921-8.

[16] V. N. Nguyen, N. Chung, G. N. Balaji, K. Rudzki, and A. T. Hoang, “Internet of things-driven approach integrated with explainable machine learning models for ship fuel consumption prediction,” *Alexandria Eng. J.*, vol. 118, pp. 664–680, 2025, doi: 10.1016/j.aej.2025.01.067.

[17] H. Dong, J. Dong, and S. Sun, “Crop water stress detection based on uav remote sensing systems,” *Agric. Water Manage.*, vol. 303, no. September, 2024.

[18] J. Wanyama *et al.*, “A systematic review of fourth industrial revolution technologies in smart irrigation: Constraints, opportunities , and future prospects for sub-Saharan Africa,” *Smart Agric. Technol.*, vol. 7, Art. no. 100412, 2024, doi: 10.1016/j.atech.2024.100412.

[19] A. Martelli, D. Rapinesi, L. Verdi, I. I. M. Donati, A. Dalla Marta, and F. Altobelli, “Smart irrigation for management of processing tomato: A machine learning approach,” *Irrig. Sci.*, 2024, doi: 10.1007/s00271-024-00993-9.

[20] B. Agustirandi, I. Inayah, N. S. Aminah, and M. Budiman, “Analysis comparison, calibration, and application of low-cost soil moisture in smart agriculture based on internet of things,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 22, no. 5, pp. 1221–1230, 2024, doi: 10.12928/TELKOMNIKA.v22i5.25703.

[21] S. Z. Islam *et al.*, “Sustainable smart irrigation system (SIS) using solar pv with rainwater harvesting technique for indoor plants,” *PLos ONE*, vol. 20, no.1, Art. no. e036911, 2025, doi: 10.1371/journal.pone.0316911.

[22] E. A. Abioye *et al.*, “Precision irrigation management using machine learning and digital farming solutions,” *AgriEngineering*., vol. 4, no. 1 pp. 70-103, Feb. 2022, doi: 10.3390/agriengineering4010006.

[23] D. Vallejo-Gomez, M. Osorio, and C. A. Hincapi, “Smart irrigation systems in agriculture: A systematic review,” *Agronomy*, vol. 13, no. 2 Art. no. 342, Feb 2023. doi: 10.3390/agronomy13020342.

[24] S. S. Fakhrulddin, V. D. Bhatt, and S. K. Gharghan, “Intelligent air pollution prediction algorithm-based optimized random forest regression for reducing asthmatic attacks,” *J. Air Pollut. Heal.*, vol. 10, no. 1, pp. 1–18, 2025, doi: 10.18502/japh.v10i1.18091.

[25] C. Banciu, A. Florea, and R. Bogdan, “Monitoring and predicting air quality with iot devices,” *Processes*, vol. 12, no. 9, 2024, doi: 10.3390/pr12091961.

[26] J. Fondaj, M. Hamiti, S. Krrabaj, X. Zenuni, and J. Ajdari,

"Comparison of predictive algorithms for iot smart agriculture sensor data," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 21, pp. 65–78, 2023, doi: 10.3991/ijim.v17i21.44143.

[27] W. Noonpakdee and S. Danrangab, "Predicting dew point temperatures: A machine learning approach with shap explanations," *ASEAN J. Sci. Technol. Reports*, vol. 28, no. 2, 2025, doi: 10.55164/ajstr.v28i2.255728.

[28] P. Karampakakis, D. Ioakeimidou, P. Chatzimisios, and K. A. Tsintotas, "A web-based application for smart city data analysis and visualization," *Future Internet*, vol. 17, no. 5, 2025, doi: 10.3390/fi17050217.

[29] J. Konecny *et al.*, "Efficient soil temperature profile estimation for thermoelectric powered sensors," *Sensors*, vol. 25, no. 13, 2025, doi: 10.3390/s25134232.

[30] A. Morchid *et al.*, "IoT-enabled smart agriculture for improving water management: A smart irrigation control using embedded systems and server-sent events," *Sci. Afr.*, 2025, doi: 10.1016/j.sciaf.2024.e02527.