

An IoT-Based Environmental Monitoring System with Machine Learning Soil Moisture Prediction for Polybag Rice Cultivation on Limited Land

Defitriana Fardiyanti*, Rendy Munadi, Helni Mutiarsih Jumhur

Universitas Telkom, Indonesia

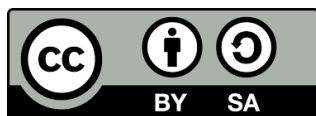
Email: defitrianafardiyanti@student.telkomuniversity.ac.id*,
rendymunadi@telkomuniversity.ac.id, helnimj@telkomuniversity.ac.id

Keywords:

Automated Irrigation; ESP32;
Firebase; Internet of Things;
Machine Learning.

Abstract

The continued decline in conventional farmland has created a need for cultivation methods that do not depend on paddy field infrastructure. This study presents an Internet of Things (IoT)-based environmental monitoring and automated irrigation system for polybag rice cultivation, developed using an ESP32 microcontroller integrated with a DHT22 temperature and humidity sensor, a BH1750 ambient light sensor, and an analog soil moisture sensor. Sensor readings were transmitted to the Firebase Realtime Database and visualized through a responsive web dashboard, while pump actuation followed a two-threshold control logic at soil moisture levels of 50% and 80%. Functional validation yielded a mean absolute percentage error (MAPE) of 0.53% for the DHT22 and 2.86% for the BH1750, an average transmission time of 409.8 ms to the cloud and 3.8 s to the dashboard, and pump response times of 1.2 s in automatic mode and 12 s in manual mode, with a preliminary dashboard usability rating of 4.2 out of 5. Beyond functional validation, a 79-day continuous field deployment generated 7,208 environmental samples, from which diurnal soil moisture patterns and the correlation structure of the polybag growing medium were quantitatively characterized. A machine learning model was subsequently developed to predict soil moisture using only data obtained from the low-cost sensors. Among the four approaches evaluated using chronologically held-out data, the Random Forest model achieved an R^2 of 0.997 and reduced prediction error by 94.5% compared with a non-predictive baseline. Overall, the results demonstrate the feasibility of transitioning from reactive to anticipatory irrigation using a low-cost cultivation platform designed for land-constrained environments.



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2026 by Author. Published by
Riviera Publishing

INTRODUCTION

Rice is a staple food for more than half of the world's population and a strategic commodity whose availability directly affects national food security in Indonesia (Santos et al., 2024; Togneri et al., 2022; Suspidayanti & Rokhmana, 2021). Conventional rice cultivation, however, faces increasing pressure. In Karawang Regency, one of the country's major rice-producing areas, paddy fields continue to be converted to residential and industrial uses, driven by investor demand, repeated crop failures, and low agricultural productivity (Irawan et al., 2023). Alternative cultivation approaches that do not depend on the availability of conventional paddy fields are therefore becoming increasingly relevant. This situation

underscores the urgency of developing cultivation methods that can be implemented on limited land while maintaining productivity and sustainability.

Cultivating rice in polybags represents one such alternative. This method enables cultivation in backyards and other confined spaces without conventional paddy field infrastructure, and previous research has shown that rice grown in 40 cm × 40 cm polybags can achieve satisfactory growth and yield, provided that appropriate environmental conditions are maintained (Saraf et al., 2022). Maintaining these conditions, however, remains challenging. The limited capacity of the polybag growing medium makes rice plants more sensitive to fluctuations in soil moisture, temperature, and water availability than plants grown in open paddy fields, while environmental variables such as light intensity and temperature significantly influence plant growth and grain quality (Majumder, 2023). When monitoring is performed manually and periodically, critical environmental conditions may be detected only after they have occurred, with potential losses becoming particularly significant during the reproductive and ripening stages, when rice plants are highly sensitive to environmental stress.

Internet of Things (IoT) technology offers a potential solution by integrating microcontrollers, environmental sensors, and cloud platforms to enable the automatic collection, transmission, and visualization of environmental data (Qazi et al., 2022; Thilakarathne et al., 2023). Existing IoT-based irrigation systems for limited-land cultivation have demonstrated the feasibility of this approach (Naseer et al., 2024). However, three important limitations remain. First, their control mechanisms are fundamentally reactive: irrigation is triggered by the current state of the growing environment, whether through a fixed soil moisture threshold, a temperature limit, or a predetermined schedule. Consequently, the system responds only after a critical condition has been reached. Even when computational approaches such as fuzzy inference are used to determine irrigation duration (Farooq et al., 2022), the logic determines how the system should respond to current conditions rather than anticipating future changes in soil moisture. Second, validation is typically conducted over relatively short observation periods of one to two weeks, with only one or two measurements per day (Wardani, 2023). Consequently, the environmental behavior of the growing medium throughout an extended cultivation period has not been quantitatively characterized. Third, previous systems have generally focused on crops such as ornamental plants (Farooq et al., 2022) or potted chili plants (Wardani, 2023), whereas the soil moisture dynamics of confined polybag growing media used for rice, which may exhibit different drainage and drying characteristics from open soil, have received comparatively limited attention. Notably, Roziqin et al. (Wardani, 2023) identified the integration of machine learning to predict plant water requirements as a potential direction for future research.

Machine learning has been applied to soil moisture estimation and prediction in previous studies (Karel et al., 2022; Pan, 2024) however, such research has often been conducted under laboratory conditions or has relied on comparatively costly instrumentation. Therefore, it remains necessary to determine whether soil moisture can be reliably predicted under field conditions using only the low-cost sensors already incorporated into a small-scale cultivation system. This study addresses these research gaps by implementing a functional IoT-based monitoring and irrigation system for polybag rice cultivation, characterizing the

soil moisture dynamics of the polybag growing medium during a long-term field deployment, and evaluating whether soil moisture can be predicted using low-cost sensor data. In doing so, the study represents a step toward transitioning from reactive irrigation, which responds after a critical condition has been reached, to anticipatory irrigation, which responds to a predicted decline before the critical threshold is reached. The study makes two primary contributions. First, its empirical contribution consists of a quantitative characterization of soil moisture dynamics in polybag rice cultivation based on a field dataset comprising 7,208 samples collected over 79 days of continuous outdoor operation. This characterization encompasses descriptive statistics, diurnal soil moisture patterns, and interparameter correlations, thereby providing reusable empirical knowledge about the growing medium rather than merely demonstrating a functional system. Second, its methodological contribution lies in advancing from reactive, threshold-based irrigation toward machine learning-based predictive irrigation, in which soil moisture is predicted using data from existing low-cost sensors and the predictive performance is benchmarked against a non-predictive baseline.

The system was further examined in relation to the Indonesian telecommunications regulatory framework. Because the system operates as a wireless device in the 2.4 GHz frequency band and relies on a cloud-based data platform, regulatory compliance was incorporated into the system design considerations rather than treated solely as a post-development requirement.

The remainder of this paper is organized as follows. Section II reviews the related literature. Section III describes the system design. Section IV presents the evaluation methodology. Section V presents and discusses the results. Finally, Section VI provides the conclusions.

METHOD

The evaluation was conducted through four complementary approaches: functional validation of the system components, statistical analysis of the long-term dataset, development and evaluation of the prediction model, and regulatory analysis.

Functional Validation

Functional testing was carried out in the yard of a house in Sukaluyu Village, Karawang Regency, on 13 June 2026 under natural lighting and weather so that the results would reflect actual operating conditions. Each scenario was repeated five times. Sensor accuracy was assessed against reference instruments: a Taffware HTC-2 digital hygrothermometer for the DHT22 and a digital lux meter for the BH1750, both placed in the same position as the sensor under test. Deviation was quantified using the mean absolute percentage error (MAPE), which is widely adopted for evaluating sensor accuracy in IoT monitoring systems (ElBeheiry & Balog, 2023). The error of an individual measurement and the overall MAPE are given by where R_i and S_i denote the reference value and the sensor value at the i -th measurement.

$$\text{Error (\%)} = \frac{|R_i - S_i|}{R_i} \times 100\% \quad (1)$$

Connectivity testing measured the time required for the ESP32 to establish a Wi-Fi connection from power-on, and the time for sensor data to be stored in Firebase. Latency testing measured the interval from a change at the sensor to the updated value appearing on

the dashboard. Pump testing measured actuation delay in automatic mode, by drying the medium until the sensor read below 50%, and in manual mode, by alternately issuing ON and OFF commands from the dashboard. The system was considered to pass if the pump activated automatically within three seconds and executed a manual command within fifteen seconds. Dashboard testing compared displayed values against the ESP32 serial monitor, verified layout across a 14-inch laptop and a 6-inch smart-phone, and collected a preliminary usability rating on a five-point Likert scale (Chicaiza et al., 2024).

Dataset and Statistical Analysis

Environmental records were logged to Firebase at 15-minute intervals over a 79-day deployment from 1 May to 18 July 2026, spanning the vegetative through early reproductive phase. Rice seedlings had been transplanted on 25 April 2026. Each record contains air temperature, light intensity, and soil moisture with an NTP timestamp. An uninterrupted logging cycle would yield 7,584 records; the dataset contains 7,208 complete records, corresponding to 95% completeness, the shortfall arising from five brief interruptions caused by intermittent power and network conditions. Records were ordered chronologically and checked for completeness before analysis.

Descriptive statistics were computed for all 7,208 samples of each parameter, comprising the mean, median, standard deviation with Bessel's correction, coefficient of variation, quartiles, and extremes (Botta-Dukát, 2023). To examine relationships between parameters, the Pearson correlation coefficient was computed for each pair of variables across all samples. Samples were additionally grouped by hour of day to expose the diurnal pattern.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

Soil Moisture Prediction

Seven input features were derived from the raw readings and timestamps, as summarised in Table II. Two considerations guided their design. First, the hour of day was encoded cyclically through sine and cosine transformations, so that consecutive hours such as 23:00 and 00:00 are treated as adjacent rather than twenty-three units apart. Second, because the correlation analysis indicated that soil moisture responds to accumulated effects over time rather than to instantaneous readings, lag features were introduced: soil moisture recorded 15 minutes earlier ($t-1$) and one hour earlier ($t-4$), together with temperature 15 minutes earlier. These give the model a short-term memory of recent moisture behaviour. The prediction target is soil moisture at the current time step.

After lag construction, the first four records lacking a complete set of lag values were removed, leaving 7,204 usable samples. These were split 70:30 into training and test sets chronologically rather than at random: the earlier 5,042 samples (1 May to 23 June) were used for training and the later 2,162 (23 June to 18 July) held out for testing. The choice matters for a time series. Had the data been shuffled, the model could have been trained on moments occurring later in time than some test samples, leaking future information and

inflating apparent accuracy. Splitting by time removes that possibility and mirrors the way the system would actually operate, where only past data is available to predict future values.

Four approaches were compared. A naive baseline pre-dicting that the next value equals the most recent one represents the system without machine learning and anchors the comparison. Against it were set a Linear Regression, a Random Forest of 200 trees with maximum depth 12, and a Support Vector Regression with an RBF kernel. For the Linear Regression and SVR, whose behaviour is sensitive to feature scale, features were standardised to zero mean and unit variance, fitted only on the training data so that no test information influenced preparation. The random seed was fixed at 42 for reproducibility.

Performance was assessed on the held-out test set using the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R2). These provide complementary information: RMSE emphasises larger errors in the same unit as soil moisture, MAE reflects typical error magnitude with less sensitivity to outliers, and R2 indicates how much of the variance in soil moisture is explained (Mukasine et al., 2023). Random Forest feature importance scores were additionally examined.

Table 1. Engineered Input Features For Soil Moisture Prediction

Feature	Description
Temperature	Air temperature at time t ($^{\circ}\text{C}$)
Light intensity	Light intensity at time t (lux)
Hour (sine)	Cyclic encoding, ($\sin(2\pi h/24)$)
Hour (cosine)	Cyclic encoding, ($\cos(2\pi h/24)$)
Soil moisture ($t - 1$)	Soil moisture 15 minutes earlier
Soil moisture ($t - 4$)	Soil moisture one hour earlier
Temperature ($t - 1$)	Temperature 15 minutes earlier

Regulatory Analysis

(Khan, 2024) Because the system operates as a wireless IoT device, it must comply with applicable regulations rather than being designed in isolation and justified afterward. The analysis proceeded in four stages. First, the system was decomposed into its regulated aspects: use of the radio frequency spectrum, the telecommunications device itself, operation of the system as a telecommunications activity, and the agricultural cultivation method. Second, for each aspect the applicable regulations were traced hierarchically from the Telecommunications Law downward through implementing government and ministerial regulations, so that each technical provision could be linked to its higher legal basis and superseded regulations excluded. Third, requirements were examined at the level of individual articles and matched against the corresponding characteristic of the system, determining in each case whether and under what condition the system complies, the risk that would arise if the requirement were not met, and how the design or operating context mitigates that risk. Framing the analysis around risk rather than a binary compliant-or-not judgement ensures that the review reflects the actual consequences of each provision. Fourth, an internationally recognised standard applicable to the wireless hardware was included to position compliance within a broader frame of reference.

Table 2. Sensor Accuracy Against Reference Instruments

Sensor	Reference	Trials	MAPE
DHT22 (temperature)	Taffware HTC-2	5	0.53%
BH1750 (light)	Digital lux meter	5	2.86%
Soil moisture	Historical log ^a	15	—

^aNo calibrated reference instrument was available for the analog soil moisture sensor; readings were instead verified for range coverage and consistency against the historical log, so no MAPE is reported.

RESULTS AND DISCUSSION

Sensor Accuracy

Sensor accuracy is examined first because everything that follows depends on it: if the readings cannot be trusted, neither the automated irrigation decisions nor the long-term analysis would be reliable. Table III summarises the outcome.

The DHT22 achieved a MAPE of 0.53%, consistent with the manufacturer's stated accuracy of approximately 0.5 °C and in line with independent assessments of this sensor (Tran et al., 2021). Individual errors ranged from 0.00% to 1.32%, indicating that agreement with the reference was systematic rather than incidental. This resolution is sufficient to distinguish the temperature variations that matter for rice cultivation, where the difference between a favourable and a stressful microclimate can be only a few degrees.

The BH1750 achieved a MAPE of 2.86%, tracking the reference across 4,374–7,832 lux under full daylight and 299–305 lux under shade. The larger deviation relative to the temperature sensor is expected: light intensity fluctuates far more rapidly than temperature and the sensor is more sensitive at low illumination, which is reflected in the higher errors observed in the shaded afternoon readings (4.91% and 3.39%) compared with midday.

Soil moisture could not be assessed against a calibrated reference instrument, as none was available; the readings were therefore verified for range coverage and internal consistency rather than for absolute accuracy, and no MAPE is reported for this sensor. Values drawn from the historical log spanned 45–75%, covering the full range observed during deployment.

Communication and Actuation Performance

Table 3 reports the communication and actuation results. All trials succeeded.

Table 3. Communication and Actuation Performance (Five Trials Each)

Measurement	Average	Range
Wi-Fi connection time	4.45 s	4.34–4.54 s
Sensor to Firebase	409.8 ms	387–431 ms
Sensor to dashboard	3,824 ms	3,241–4,314 ms
Pump response, automatic	1,243 ms	1,187–1,312 ms
Pump response, manual	12,060 ms	11,843–12,341 ms

The ESP32 connected to Wi-Fi in an average of 4.45 s with a spread of only 0.20 s, indicating a stable and predictable connection process. Data reached Firebase in 409.8 ms on average, consistently under half a second, which is more than sufficient for an application in which conditions change gradually.

End-to-end latency from sensor to dashboard averaged 3,824 ms. The difference relative to the cloud transmission time reflects the additional interval required for the

dashboard to retrieve the updated value and refresh the interface. For monitoring plant cultivation, where environmental conditions change gradually rather than instantaneously, a delay of a few seconds is well within acceptable limits.

Automatic pump actuation averaged 1,243 ms, comfortably below the three-second tolerance, with soil moisture during testing ranging from 38% to 47%. Deactivation above the 80% threshold was observed qualitatively in each trial after the polybag was watered, confirming correct operation of the upper-threshold logic; precise measurement of the deactivation delay is left to future testing. Manual actuation averaged 12,060 ms, longer because the command traverses a longer path from dashboard to Firebase to ESP32 to relay, but still within the fifteen-second tolerance and acceptable for manual polybag irrigation.

Dashboard Validation

All six dashboard features, namely air temperature, light intensity, soil moisture, pump status, manual control, and stored history, displayed values matching the ESP32 serial monitor, confirming that the retrieval path from Firebase to the dashboard functions correctly. Every interface component rendered correctly on both a 14-inch Dell Latitude E7470 and a 6-inch iPhone 11, confirming responsiveness across screen sizes.

A preliminary usability assessment with two participants rating five aspects on a five-point Likert scale returned an overall average of 4.2. The highest score was for operating the pump ON/OFF button (5.0) and the lowest for understanding historical data (3.5), suggesting that the trend graphs are the clearest candidate for improvement. This assessment involved a limited number of participants who were not the target end users, and therefore provides an initial indication rather than a representative evaluation; a broader usability study including actual farmers is required before conclusions can be drawn about suitability for that population.

Table 4. Descriptive Statistics Of The Environmental Dataset (N = 7,208)

Parameter	Mean	Median	Std	Min–Max	CV
Temperature (°C)	27.88	27.0	3.09	24–34	11.1%
Soil moisture (%)	62.50	63.0	9.37	45–75	15.0%
Light intensity (lux)	2,410.20	1.0	2,676.93	0–8,000	111.1%

Table 5. Frequency Of Soil Moisture Conditions

Condition	Range	Samples	Share
Dry (pump activated)	<50%	930	12.9%
Normal	50–80%	6,278	87.1%
Saturated (pump stopped)	>80%	0	0.0%
Total	—	7,208	100.0%

Dataset Characterisation

Over 79 days of continuous outdoor operation the system recorded 7,208 complete samples at 95% completeness. The 376 missing samples arose from five short recording interruptions caused by intermittent power and network conditions, which are characteristic of real outdoor IoT deployment and confirm that the data originate from genuine field operation rather than simulation. Table V presents the descriptive statistics.

The three parameters exhibit markedly different variability. Air temperature is the most stable, with a coefficient of variation of 11.1% and an interquartile range of 5.2 °C around a

mean of 27.88 °C. Soil moisture shows moderate variability (CV 15.0%), with the middle 50% of readings concentrated between 55% and 71%, confirming that the primary control parameter remained within a controlled and predictable range; the close agreement between its mean (62.50%) and median (63.0%) indicates an approximately symmetric distribution. Light intensity, by contrast, shows extreme variability (CV 111.1%), its median of 1 lux against a mean of 2,410 lux reflecting the natural day–night cycle in which more than half the readings fall in darkness, and confirming a strongly right-skewed distribution.

Classifying the samples against the irrigation thresholds (Table VI) shows that soil moisture remained within the normal range for 87.1% of the deployment. Dry conditions occurred in 930 samples, roughly one reading in every eight, confirming that the automated pump served a genuine operational purpose rather than a merely theoretical one. No sample exceeded the upper threshold.

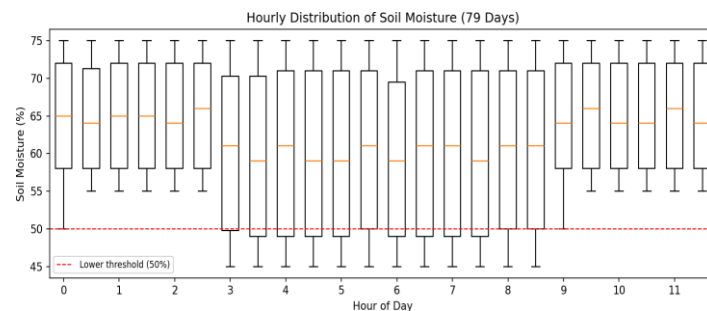


Figure 1. Hourly distribution of soil moisture across all 79 days of deployment.

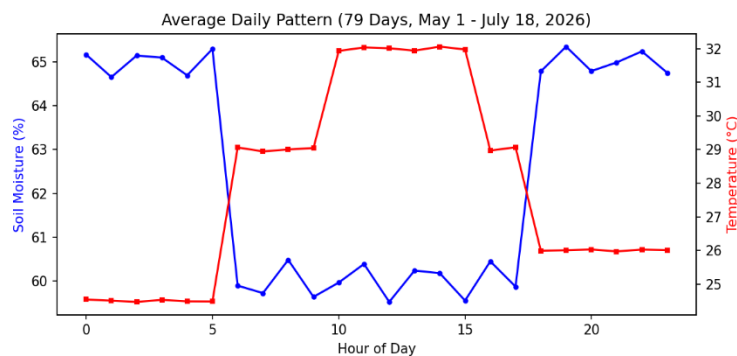


Figure 2. Average daily pattern of soil moisture and air temperature over the deployment.

Diurnal Pattern and Correlation

Grouping the samples by hour of day reveals a clear diurnal pattern (Fig. 4). Soil moisture is highest and most stable during the night and early morning, then declines and becomes more variable through the daytime period of approximately 06:00–17:00 WIB, coinciding with the higher temperature and light intensity that drive evapotranspiration. The widest spread and the lowest values occur consistently around midday, precisely when the system most frequently approaches the 50% irrigation threshold.

Quantitatively, mean daytime soil moisture was 60.0% with a standard deviation of 10.6%, against 65.0% with a standard deviation of 7.1% at night, a difference of 5.0 percentage points. The greater daytime variability reflects the stronger and more fluctuating influence of evapotranspiration when temperature and light intensity peak. The same

behaviour is evident at the scale of the full deployment (Fig. 5), where soil moisture follows an inverse relationship with temperature while remaining above the threshold on most days.

The Pearson correlation matrix (Fig. 6) reveals three relationships. Air temperature and light intensity are strongly and positively correlated ($r = 0.92$), as expected since both rise and fall together over the daily cycle. Soil moisture correlates weakly and negatively with air temperature ($r = -0.23$), consistent with the physical mechanism in which higher temperatures increase evapotranspiration and gradually reduce the water content of the medium. Soil moisture shows near-zero correlation with instantaneous light intensity ($r = -0.04$), indicating that it does not respond directly to momentary light levels.

This last finding is particularly informative for the modelling that follows. The weak instantaneous correlations suggest that soil moisture is driven not by the immediate value of any single parameter but by cumulative thermal effects accumulated over time, which explains why lagged moisture features proved substantially more informative than instantaneous environmental readings.

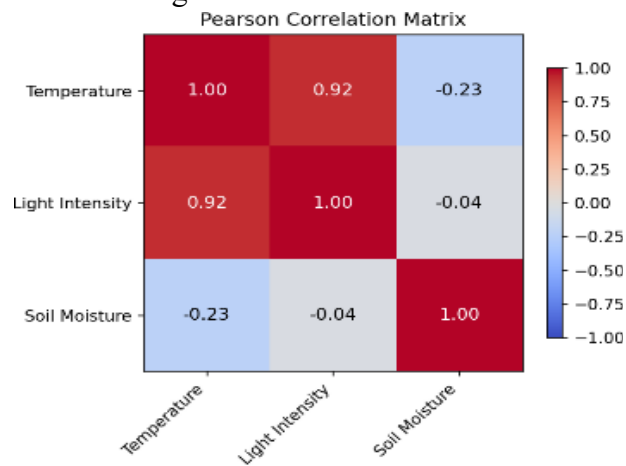


Figure 3. Pearson correlation matrix of the environmental parameters across all 7,208 samples.

Table 6. Model Performance On The Held-Out Test Set (N = 2,162)

Model	RMSE (%)	MAE (%)	R^2
Naive baseline ($t-1$)	8.717	7.636	0.137
Linear Regression	5.267	4.227	0.685
SVR (RBF kernel)	1.605	0.832	0.971
Random Forest	0.482	0.112	0.997

Soil Moisture Prediction

The functional testing established that the system operates reactively. The distribution analysis, however, showed that soil moisture in the polybag medium does not fluctuate randomly but follows a strong and regular diurnal rhythm. If the pattern is that predictable, it should be possible to anticipate the moisture level rather than merely react to it. Table VII reports the comparison.

The Random Forest emerged as the clear best performer, achieving an RMSE of 0.482%, an MAE of 0.112%, and an R^2 of 0.997 on unseen test data, reducing the baseline error by 94.5%. The progression across the four approaches is at least as informative as the winning figure. The naive baseline performed only marginally better than a random guess

($R^2 = 0.137$), because assuming that moisture remains unchanged from one step to the next ignores the gradual drift the data actually follow. The Linear Regression improved substantially ($R^2 = 0.685$) but was held back by its assumption of straight-line relationships that the data do not obey. Only the two non-linear models could bend to the true shape of the data, and of these the Random Forest fit it most closely.

The ratio between the RMSE and the MAE of the Random Forest is notably high at 4.3, well above the value of approximately 1.25 expected for normally distributed errors. This reflects the discrete nature of the sensor output: because the soil moisture readings are whole numbers confined to a narrow 45–75% band, the majority of predictions coincide exactly with the observed value and contribute no error at all, while a small number of transition points deviate by two to three percentage points and dominate the squared-error term. The MAE therefore represents typical predictive performance in this setting more faithfully than the RMSE, and the two figures should be read together rather than either one alone. Fig. 7 overlays the predictions of the three learning models on the actual soil moisture across the final three days of the test period.

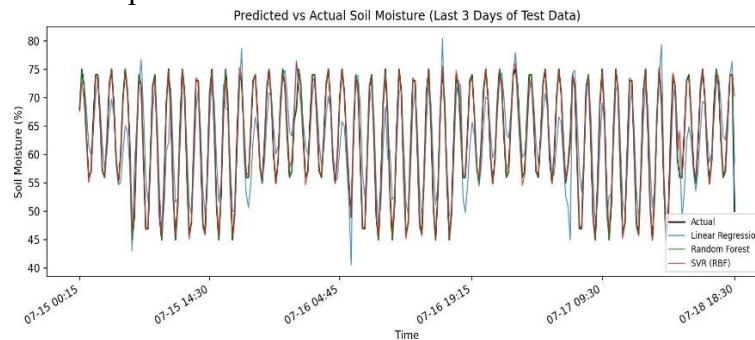


Figure 4. Predicted versus actual soil moisture over the final three days of the test period, for the three learning models.

Table 7. Comparison Of The Proposed System With Previous Studies

Study	Focus	Contribution
Karel et al., 2022	Fuzzy-logic watering duration for potted ornamental plants	Functional prototype
Rozijin et al.	Threshold-based IoT irrigation for chili cultivation in pots	Simple control logic
Sazzad et al.	Machine learning prediction of soil moisture and soil type	Regression benchmark
Atzal et al.	Adaptive watering with ARIMA-based soil moisture prediction	Model comparison
This work	Integrated IoT polybag monitoring with predictive soil moisture analysis and regulatory assessment	Sensor validation, latency testing, and regulatory mapping

Feature Importance and Validity of the Result

Fig. 8 and Table IX report the Random Forest feature importance scores. Two features together account for roughly two-thirds of the model's predictive power: soil moisture from one hour earlier (0.339) and light intensity (0.325). Both make physical sense in light of the earlier analysis. The lagged moisture matters because soil moisture drifts gradually and carries its recent state forward, so its value an hour ago is a strong clue to its present value. It is also telling that the cyclic hour features contribute very little: the temporal rhythm they

were intended to encode appears already to have been absorbed by the lag features, suggesting that a leaner model could omit them with little loss.

A result as strong as an R^2 of 0.997 rightly invites scrutiny, since such figures can signal a model that has memorised its training data. Several considerations argue against that reading. The most important is that the figure was earned on future data: the weeks from 23 June to 18 July, which the model never saw during training. Memorisation would manifest as strong training performance collapsing on the test set, whereas here it is the test performance itself that stands high. The remaining explanation lies in the data. Soil moisture follows so regular a daily pattern, and the fifteen-minute lag places so recent and relevant a value at the model's disposal, that the prediction task is genuinely more tractable than it might first appear, helped further by the readings being whole numbers confined to a fairly narrow 45–75% band.

None of this should be mistaken for a claim of universality. These results describe one location across one growing season, and whether they would hold for a different soil, a different climate, or a different time of year remains untested. That boundary is acknowledged as a limitation. What ultimately gives the result its weight is less the number than the capability it points to: a system that can only respond once the soil has already dried is by nature always a step behind, whereas one that can foresee the decline can act ahead of it.

Table 8. Comparison Of The Proposed System With Previous Studies

Study	Focus	Contribution
Karel <i>et al.</i>	Fuzzy-logic watering duration for potted ornamental plants	Functional prototype
Rozijin <i>et al.</i>	Threshold-based IoT irrigation for chili cultivation in pots	Simple control logic
Sazzad <i>et al.</i>	Machine learning prediction of soil moisture and soil type	Regression benchmark
Afzal <i>et al.</i>	Adaptive watering with ARIMA-based soil moisture prediction	Model comparison
This work	Integrated IoT polybag monitoring with predictive soil moisture analysis and regulatory assessment	Sensor validation, latency testing, and regulatory mapping

Comparison with Previous Studies

Table IX positions the proposed system against representative prior work. Previous studies generally concentrated on a single aspect. The irrigation-focused works (Karel *et al.*, 2022) demonstrated the feasibility of IoT-based irrigation but remained limited to irrigation functionality and basic monitoring, without cloud analytics, historical environmental analysis, or predictive modelling. The prediction-focused works (Al-Muqit, 2023) advanced modelling but offered limited treatment of complete IoT deployment, long-term operation, user interaction, or practical field conditions. Two design choices distinguish the present system. First, operational control is separated from predictive analysis: automatic irrigation uses predefined thresholds to ensure operational reliability, while machine learning predicts future conditions. This allows the irrigation system to remain fully functional if the prediction model is unavailable, requires retraining, or encounters uncertainty. Second, validation was conducted in a real outdoor cultivation environment for 79 consecutive days, yielding 7,208

observations, which provides stronger evidence of practical reliability than isolated laboratory or simulation experiments.

Regulatory Compliance

Because the system transmits data over the radio frequency spectrum, it was designed to operate within the applicable regulatory framework rather than reconciled with it afterward. Table X summarizes the assessment.

The risk associated with the spectrum provision is concrete rather than nominal: under Article 53(1) of Law 36/1999, using the radio frequency spectrum without a permit or outside its designated allocation is a criminal offence punishable by up to four years' imprisonment and a fine of up to IDR 400,000,000, while Article 45 provides for administrative revocation. That risk is mitigated by design. Because the 2.4 GHz band is class-licensed, the system operates under a general authorisation rather than an individual permit, and by confining operation to the standardised band with unmodified certified hardware it remains within the permitted technical envelope. Across all regulated aspects, compliance rests on deliberate choices, namely a class-licensed frequency band, an internationally certified low-power module, and a non-commercial research scope. The analysis also identifies the specific points, chiefly device certification and business licensing, at which additional obligations would arise were the system developed toward commercial deployment.

Table 9. Regulatory Assessment of The Proposed System

Regulated aspect	Applicable regulation	Implication
Radio frequency spectrum	Law 36/1999 Art. 33(1); Gov. Reg. 46/2021 Art. 45; Minister Reg. 2/2025 Art. 2, 3, 11	2.4 GHz band is class-licensed
Telecommunications device	Law 36/1999 Art. 32(1); Minister Reg. 3/2024 Art. 10; ESP32-WROOM-32 datasheet	Certified module used as a research prototype
Operation and licensing	Minister Reg. 5/2021; Gov. Reg. 6/2021	Not applicable to non-commercial use
International standard	ETSI EN 300 328 under RED 2014/53/EU	Module conformity reference
Sustainable agriculture	Law 22/2019 Art. 99	Supports efficient resource use

CONCLUSION

This paper presented an Internet of Things (IoT)-based environmental monitoring and automated irrigation system for rice cultivation in polybags on limited land, enhanced with machine learning-based soil moisture prediction and evaluated against the Indonesian telecommunications regulatory framework. Functionally, the system performed within acceptable tolerances across all tested components. The DHT22 achieved a mean absolute percentage error (MAPE) of 0.53%, while the BH1750 achieved 2.86%. The ESP32 connected to Wi-Fi in 4.45 s and transmitted data to the Firebase Realtime Database in 409.8 ms, with an end-to-end dashboard latency of 3.8 s. The pump response times were 1.2 s in automatic mode and 12 s in manual mode. The dashboard also displayed all features correctly across different screen sizes and achieved a preliminary usability rating of 4.2 out of 5.

Beyond functional validation, the 79-day deployment enabled a quantitative characterization of the polybag growing medium. From 7,208 samples with 95% data

completeness, a clear diurnal soil moisture pattern emerged, with daytime mean soil moisture 5.0 percentage points lower than the nighttime mean and greater variability during the daytime. The correlation analysis showed that air temperature and light intensity were strongly correlated ($r = 0.92$), whereas soil moisture was weakly correlated with air temperature ($r = -0.23$) and showed negligible correlation with instantaneous light intensity ($r = -0.04$). Using this dataset, the Random Forest model predicted soil moisture from the available low-cost sensor data with a root mean squared error (RMSE) of 0.482%, a mean absolute error (MAE) of 0.112%, and an R^2 of 0.997 on chronologically held-out data, reducing prediction error by 94.5% compared with the non-predictive baseline. These findings demonstrate the study's primary contribution: advancing from reactive, threshold-based irrigation toward anticipatory irrigation using a low-cost polybag cultivation platform.

Several directions for future research are recommended. The prediction model could be extended to support closed-loop anticipatory irrigation, in which predicted declines in soil moisture automatically trigger irrigation before the critical threshold is reached. Future evaluations should include a larger number of polybags under more diverse environmental and network conditions and cover the entire rice growth cycle, including the reproductive and maturation stages, to determine whether the observed soil moisture patterns and predictive performance can be generalized beyond the vegetative-to-early-reproductive period examined in this study. The static soil moisture threshold could also be adapted according to the growth stage, while a water-level sensor could be incorporated to monitor irrigation water availability. Regarding communication protocols, Message Queuing Telemetry Transport (MQTT) could be investigated as an alternative communication mechanism to reduce latency, while a mobile application could improve field accessibility and more stringent Firebase security rules should be implemented for production-scale deployment. Finally, integrating camera-based visual detection with environmental sensor data and incorporating soil pH and nutrient sensors could support more comprehensive data-driven cultivation decisions.

REFERENCES

- Al-Muqit, M. M. (2023). Design and implementation of ESP32-based IoT devices. *Sensors*, 23(15), 6739.
- Botta-Dukát, Z. (2023). Quartile coefficient of variation is more robust than CV for traits calculated as a ratio. *Scientific Reports*, 13(1), 4671.
- Chicaiza, K., Paredes, R., Sarzosa, I. M., Yoo, S. G., & Zang, N. (2024). Smart farming technologies: A methodological overview and analysis. *IEEE Access*, 12, 164922–164941.
- ElBeheiry, N., & Balog, R. S. (2023). Technologies driving the shift to smart farming: A review. *IEEE Sensors Journal*, 23(3), 1752–1769.
- Farooq, M. S., Riaz, S., Helou, M. A., Khan, F. S., Abid, A., & Alvi, A. (2022). Internet of Things in greenhouse agriculture: A survey on enabling technologies, applications, and protocols. *IEEE Access*, 10, 53374–53397.
- Irawan, A., Noor, T. I., & Karyani, T. (2023). Faktor-faktor yang berkaitan dengan alih fungsi lahan sawah di Kecamatan Purwasari Kabupaten Karawang Provinsi Jawa Barat. *Mimbar Agribisnis*, 9(1), 277–290.
- Karel, F. N., Fauziah, & Soepriyono, G. (2022). Smart agriculture: Pengendalian kelembapan dan suhu pada penyiraman otomatis tanaman berbasis IoT. *Jurnal Teknik Informatika Dan Sistem Informasi (JATISI)*, 9(2), 839–854.

- Khan, M. M. H. H. (2024). Internet of Things-based smart agriculture irrigation and monitoring system using Ubidots server. *Engineering Proceedings*, 82(1), 99.
- Majumder, S. (2023). Assessing low-cost capacitive soil moisture sensors: Accurate, affordable, and IoT-ready solutions for soil moisture monitoring. *International Journal of Environment and Climate Change*, 13(11), 2233–2242.
- Mukasine, A., Sibomana, L., Jayavel, K., Nkurikiyeyezu, K., & Hitimana, E. (2023). Correlation analysis model of environment parameters using IoT framework in a biogas energy generation context. *Future Internet*, 15(8), 265.
- Naseer, A., Shmoon, M., Shakeel, T., Rehman, S. U., Ahmad, A., & Gruhn. (2024). A systematic literature review of the IoT in agriculture—global adoption, innovations, security, and privacy challenges. *IEEE Access*, 12, 60986–61021.
- Pan, S. (2024). Using the Pearson’s correlation coefficient as the sole metric to measure the accuracy of quantitative trait prediction: Is it sufficient? *Frontiers in Plant Science*, 15, 1480463.
- Qazi, S., Khawaja, B. A., & Farooq, Q. U. (2022). IoT-equipped and AI-enabled next generation smart agriculture: A critical review, current challenges and future trends. *IEEE Access*, 10, 21219–21235.
- Santos, E. P. D., Martini, D. K. T., Lele, O. K., & Kiuk, Y. (2024). Rice plant cultivation technology (*Oryza sativa* L.) based on organic fertilization in realizing food security in the RI–RDTL border area. *Jurnal Penelitian Pertanian Terapan*, 24(3), 434–442.
- Saraf, P. R., Jadhao, S. M., Wanjari, S. J., Kolwate, S. G., & Patil, A. D. (2022). A review on Firebase (backend as a service) for mobile application development. *International Journal for Research in Applied Science and Engineering Technology*, 10(1), 967–971.
- Suspidayanti, L., & Rokhmana, C. A. (2021). Identifikasi fase pertumbuhan padi menggunakan citra SAR Sentinel-1. *Elipsoida: Jurnal Geodesi Dan Geomatika*, 4(1), 9–15.
- Thilakarathne, D. N., Bakar, M. S. A., Abas, P. E., & Yassin, H. (2023). Towards making the fields talks: A real-time cloud enabled IoT crop management platform for smart agriculture. *Frontiers in Plant Science*, 13, 1030168.
- Togneri, R., dos Santos, D. F., Camponogara, G., Nagano, H., Custódio, G., Prati, R., Fernandes, S., & Kamienski, C. (2022). Soil moisture forecast for smart irrigation: The primetime for machine learning. *Expert Systems with Applications*, 207, 117653. <https://doi.org/10.1016/j.eswa.2022.117653>
- Tran, L. G., Tran, N. M., & Tran, T. V. (2021). IoT system for monitoring a large-area environment sensors and control actuators using real-time Firebase database. In *Intelligent Human Computer Interaction*.
- Wardani, I. K. (2023). The feasibility study: Accuracy and precision of DHT22 in measuring the temperature and humidity in the greenhouse. *IOP Conference Series: Earth and Environmental Science*, 1230(1), 12146.