Low-cost AIoT System for Real-Time Power Estimation in Small and Medium-Scale Rooftop Solar Power Systems

Kim-Anh Nguyen*
Faculty of Electrical Engineering
University of Science and Technology, The University of Danang
Da Nang, Vietnam
* Corresponding author: nkanh@dut.udn.vn

Quang Linh Nguyen
Faculty of Electrical Engineering
University of Science and Technology, The University of Danang
Da Nang, Vietnam
105230486@hv.dut.udn.vn

Thai Anh Au Tran
Faculty of Electrical Engineering
University of Science and Technology, The University of Danang
Da Nang, Vietnam
ttaau@dut.udn.vn

Minh Hien Ngo
Faculty of Electrical Engineering
University of Science and Technology, The University of Danang
Da Nang, Vietnam
105200404@sv1.dut.udn.vn

Abstract*—The rapid growth of rooftop solar power (RSP) systems in Vietnam and many other countries has created an urgent demand for reliable real-time monitoring and power estimation solutions, particularly for small- and medium-scale installations. However, existing commercial systems often involve high deployment costs, strong dependence on network infrastructure, and limited flexibility in remote areas. This paper proposes a low-cost AIoT-based monitoring and power estimation system that integrates Internet of Things (IoT) technologies with machine learning models to improve scalability, resilience, and adaptability under real-world conditions. The system is implemented on low-cost hardware using a Raspberry Pi as an edge computing node, enabling synchronized acquisition of electrical data from inverters and environmental data from sensors via industrial communication protocols. To estimate the DC power output, two machine learning models including eXtreme Gradient Boosting (XGBoost) and Support Vector Regression (SVR) are trained using real measurement data collected from a 1.2 MW rooftop PV system in Da Nang, Vietnam. The estimated results are compared with actual measurements to evaluate model performance under varying weather conditions. Experimental results show that both models achieve high accuracy, while XGBoost demonstrates superior performance under rapidly fluctuating conditions. In addition, the system integrates a real-time alert mechanism to detect abnormal deviations between estimated and measured power, enabling early fault identification. By combining edge computing and cloud-based services through the open-source MING stack, the proposed solution ensures low latency, stable operation during network interruptions, and efficient remote monitoring. Overall, the proposed AIoT framework provides a practical, cost-effective, and scalable solution for monitoring and optimizing rooftop PV systems.

Keywords- Rooftop photovoltaic system; AIoT; Edge computing; XGBoost; Real-time power estimation; Anomaly detection

I. INTRODUCTION

Solar energy has become a key pillar in the transition toward sustainable energy systems, particularly as the cost of photovoltaic (PV) electricity has significantly declined over the past decade. Rooftop solar power (RSP), which utilizes available roof space in residential and commercial buildings, plays an increasingly important role in reducing dependence on the utility grid, lowering electricity costs, and mitigating greenhouse gas emissions [1]. According to the International Energy Agency (IEA), RSP accounted for over 30% of newly installed global solar capacity in 2022, with strong growth in the Asia–Pacific region [2]. However, the operational efficiency and longevity of RSP systems strongly depend on accurate real-time monitoring and power estimation.

The performance of PV systems is influenced by multiple factors, including solar irradiance, ambient and module temperatures, humidity, soiling, and the condition of power electronic components [3], [4]. The failure to detect performance degradation—caused by aging, hotspots, or dust accumulation—can lead to significant energy losses, particularly under harsh environmental conditions [5]. This issue is particularly critical for small- and medium-scale systems, where advanced SCADA platforms are often omitted due to cost constraints, and monitoring relies mainly on aggregated inverter data, which is insufficient for detailed diagnostics and accurate forecasting.

Traditionally, PV power estimation has relied on physical or semi-empirical models, such as diode-based models or irradiance–temperature relationships. Although these methods are simple and computationally efficient, they have limited capability to capture nonlinear behaviors and rapid environmental fluctuations [6]. Moreover, their dependence on ideal conditions or difficult-to-acquire inputs reduces their effectiveness in real-world applications.

In recent years, machine learning (ML) techniques have emerged as powerful alternatives for PV power estimation. Models such as Support Vector Regression (SVR), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) have demonstrated superior accuracy by learning complex nonlinear relationships from data [7]. In particular, XGBoost achieves high predictive performance while maintaining robustness against overfitting and suitability for deployment on embedded platforms. In addition, SVR remains effective for smaller datasets and near-linear patterns due to its ε-insensitive loss function.

However, despite their high accuracy, most ML-based approaches for PV systems [8] have primarily been designed for offline analysis and forecasting. They rely on pre-collected datasets and do not sufficiently address practical challenges such as real-time data acquisition, synchronization between heterogeneous data sources, and integration into a complete monitoring architecture. Consequently, their applicability in real-world PV systems, especially small and medium-scale installations, remains limited.

In parallel, the development of the Internet of Things (IoT) and edge computing has enabled on-site data acquisition and processing, reducing reliance on cloud infrastructure. Low-cost platforms such as Raspberry Pi, Jetson Nano, and ESP32 facilitate synchronized data collection from inverters and environmental sensors, thereby reducing latency and improving system autonomy [9]. This approach is particularly beneficial in regions with unstable network connectivity.

Several studies have combined ML and IoT to improve PV system performance. For example, Özer et al. [10] applied a YOLO-based deep learning model for real-time fault detection using UAV imagery, while Aghenta et al. [11] developed a low-cost IoT-based SCADA system using Arduino and Raspberry Pi for monitoring and control. These approaches demonstrate the potential of integrating data-driven models with IoT platforms.

Nevertheless, existing IoT-based monitoring solutions [12] still face several limitations, including limited scalability, lack of data synchronization, and reliability issues. In many cases, additional sensors are used to estimate electrical parameters that are already available from inverters, increasing system complexity and cost while introducing potential inconsistencies. Furthermore, heavy reliance on cloud-based storage without robust local backup mechanisms may result in data loss during network interruptions, which is particularly problematic for remote installations.

Therefore, a significant research gap remains in the development of a low-cost, scalable, and reliable monitoring solution tailored for small and medium-scale RSP systems. Such a system should (i) integrate seamlessly with existing infrastructure, (ii) ensure robust operation under harsh environmental conditions, and (iii) support real-time edge processing to minimize latency and network dependency.

In this study, we propose a low-cost RSP monitoring and power estimation system that integrates IoT and AI to achieve high efficiency, scalability, and adaptability. The system employs edge devices to collect real-time data from inverters and environmental sensors, combined with advanced ML models for accurate power estimation under uncertain conditions. An intelligent alert mechanism is incorporated to detect anomalies and ensure stable operation. By leveraging open-source platforms, the proposed solution enables low-latency processing, continuous operation during network outages, and efficient remote monitoring. The system is experimentally validated on a real RSP installation in Hoa Vang, Da Nang, Vietnam.

The remainder of this paper is organized as follows. Section II presents the proposed AIoT-based monitoring and power estimation approach, including the IoT architecture, AI estimation models, and the integration framework between IoT and AI. Section III describes the experimental setup and discusses the experimental results under different operating conditions. Finally, Section IV concludes the paper and outlines future research directions.

II. Proposed Approach

A. IoT Architecture

The proposed IoT system is designed to synchronously collect multi-source data, perform on-site processing, and support real-time power estimation. It is organized into three main layers: (i) sensing and data acquisition, (ii) communication and edge processing, and (iii) application and visualization, as illustrated in Fig. 1.

At the sensing and data acquisition layer, two primary data sources are considered: environmental and electrical data. Environmental parameters including humidity, ambient temperature, solar irradiance, and cell temperature are measured via dedicated sensors communicating through the Modbus RTU protocol over an RS-485 bus. This approach enables reliable long-distance transmission, improved noise immunity, and system scalability. Meanwhile, electrical parameters are directly obtained from the PV inverter via Modbus TCP, including DC/AC voltage and current, frequency, power, and energy. The use of Modbus TCP enables high-speed and stable communication, which meets real-time data acquisition requirements.

At the communication and edge-processing layer, a Raspberry Pi 4 Model B serves as both a local server and an edge computing unit. It simultaneously communicates with sensors (Modbus RTU) and the inverter (Modbus TCP), while assigning timestamps to ensure data synchronization. The collected data are subsequently preprocessed through noise filtering, outlier removal, and normalization for machine learning compatibility. An AI model is deployed directly on the Raspberry Pi to perform real-time power estimation. This edge-based implementation significantly reduces latency, enhances data security, and ensures continuous operation during Internet outages. In addition, all data are locally stored on an internal storage device to prevent data loss.

Finally, at the application and visualization layer, the system utilizes an open-source toolchain to deliver user services. Specifically, Mosquitto MQTT acts as the communication

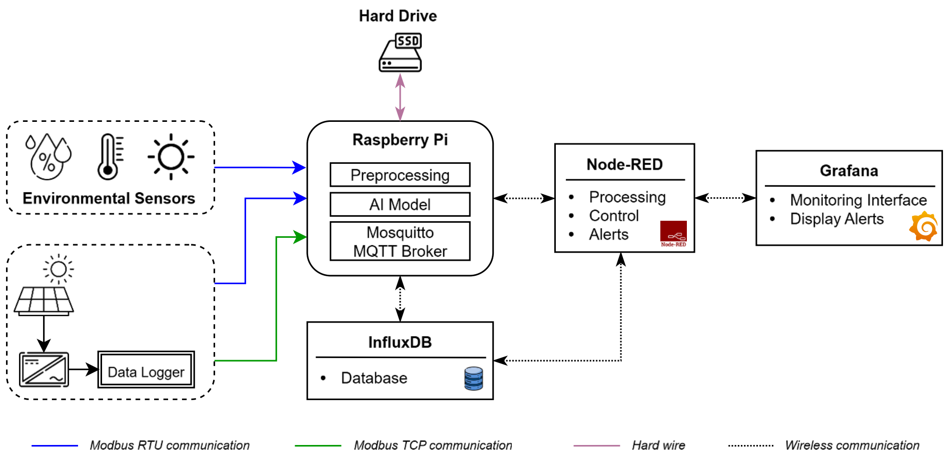

Figure 1. Block diagram of the proposed AIoT solution.

broker, enabling lightweight and efficient data exchange. Node-RED manages data flow, executes control logic, and triggers alerts when deviations between estimated and measured power exceed predefined thresholds. Grafana provides real-time visualization, including power monitoring, comparison between predicted and actual values, and long-term trend analysis. Furthermore, when network connectivity is available, data are synchronized to an InfluxDB database to support advanced analytics and future model retraining.

B. AI Estimation Model

The primary objective of the AI model in this study is to estimate the DC power output of the rooftop photovoltaic (PV) system in real-time using synchronously collected environmental and electrical data. Accurate DC power estimation not only enables effective performance monitoring but also supports operational optimization and proactive maintenance strategies.

1) Data Preprocessing and Dataset Construction

After data acquisition from the IoT layer, preprocessing is performed to ensure data quality and synchronization. Outliers are removed based on predefined physical operating thresholds of the PV system. The input feature vector is constructed as:

$$X(t)=\left[G(t),T_{amb}(t),T_{cell}(t),H(t)\right] \tag{1}$$

where $G(t)$ is the irradiance, $T_{amb}(t)$ is the ambient temperature, $T_{cell}(t)$ is the PV cell temperature, and $H(t)$ is the relative humidity at time *t*. The target output is the DC power:

$$y(t)=P_{DC}(t) \tag{2}$$

All variables are normalized using Min–Max scaling to ensure consistent ranges, thereby improving model convergence and mitigating feature imbalance.

2) XGBoost Model

XGBoost is adopted due to its superior capability in modeling nonlinear relationships under varying environmental conditions. At iteration *k*, the prediction is updated as:

$$\hat{y}_i^{(k)}=\hat{y}_i^{(k-1)}+f_k\left(x_i\right) \tag{3}$$

The objective function is defined as:

$$J^{(k)}=\sum_{i=1}^{n} l\left(y_i,\hat{y}_i^{(k)}\right)+\sum_{j=1}^{k}\Omega\left(f_j\right) \tag{4}$$

where $l(\bullet)$ is the loss function and $\Omega\left(f_j\right)$is the regularization term used to control the complexity. Moreover, XGBoost supports parallel training, enabling high accuracy with reasonable computational cost.

3) SVR Model

SVR is employed as a complementary model, particularly suitable for small and noisy datasets commonly found in rooftop PV systems. It aims to determine a regression function such that data points lie within an ε - insensitive tube:

$$\left|y_i-\left(\omega^T\phi\left(x_i\right)+b\right)\right|\le\varepsilon \tag{5}$$

where $\phi\left(x\right)$ is a nonlinear mapping induced by the RBF kernel. The SVR optimization function is defined as:

$$\min\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{n}\left(\xi_i+\xi_i^*\right) \tag{6}$$

where C is the penalty parameter and ξ_i and ξ_i^* are the slack variables.

4) Evaluation Metrics

Model performance is evaluated using three standard metrics:

- MAE (Mean Absolute Error)
- RMSE (Root Mean Square Error)
- R^2 (Coefficient of Determination)

These metrics provide complementary insights: MAE reflects average error magnitude, RMSE penalizes larger deviations, and R^2 indicates the proportion of variance explained by the model. Together, they offer a comprehensive assessment of estimation accuracy and reliability.

C. *Integration AI-IoT*

The integration of IoT and AI represents a key step in transforming the system from a conventional data acquisition platform into an intelligent real-time power estimation tool. This integration is implemented through a layered architecture that enables synchronized data collection, processing, and forecasting, as illustrated in Fig. 2.

After initializing platform services (Mosquitto, Node-RED, InfluxDB, and Grafana) and synchronizing time via NTP, the system operates in a periodic sampling cycle with a predefined interval. At each cycle, environmental data (ambient temperature, humidity, irradiance, and PV cell temperature) and inverter electrical data are collected via Modbus RTU/TCP. Subsequently, a data quality check is performed to detect CRC errors, timestamp mismatches, and out-of-range values. In cases where anomalies are detected, the system performs limited retries before temporarily buffering the data.

The validated data are then preprocessed, including noise filtering, outlier removal, normalization, and feature vector construction $x(t)$. This feature vector is then fed into the machine learning model deployed on the Raspberry Pi edge device to estimate the instantaneous DC power. The predicted values are compared with actual inverter measurements to compute absolute and relative errors. If the error exceeds a predefined threshold over multiple consecutive cycles, an alert is triggered via Node-RED, displayed on the Grafana dashboard, and sent via email notification.

Meanwhile, all data are stored both locally and in the InfluxDB database to ensure data integrity during network interruptions. Once connectivity is restored, the system synchronizes data to the cloud for long-term analysis and model retraining. Finally, the results are presented through dashboard visualization, enabling operators to simultaneously monitor real-time power, estimation accuracy, environmental conditions, and overall system status.

III. EXPERIMENTAL RESULTS AND EVALUATION

To evaluate the effectiveness of the proposed AIoT-based monitoring and power estimation system, a synchronized workflow was implemented, including: (i) acquisition of electrical data and system status monitoring; (ii) collection and processing of environmental data; (iii) local and cloud-based data storage; (iv) data synchronization, querying, and time-series alignment; and (v) real-time analysis and visualization via a Human–Machine Interface (HMI).

The experiment was conducted on a grid-connected rooftop PV system in Hoa Vang District, Da Nang, Vietnam, under tropical monsoon conditions characterized by high humidity and variable irradiance. The system has a total capacity of 1.2 MWp, consisting of 2700 JA Solar JAM78S10 monocrystalline modules (445 Wp each), connected to nine Solis-100K-5G string inverters (100 kW each). Environmental data were collected using RIKA sensors, including RK330-02 (temperature and humidity), RK200-04 (irradiance), and RK220-01 (panel temperature), communicating via Modbus RTU over RS-485. Meanwhile, electrical parameters (DC/AC voltage, current, frequency, power, and energy) were obtained directly from the inverters via Modbus TCP through a data logger.

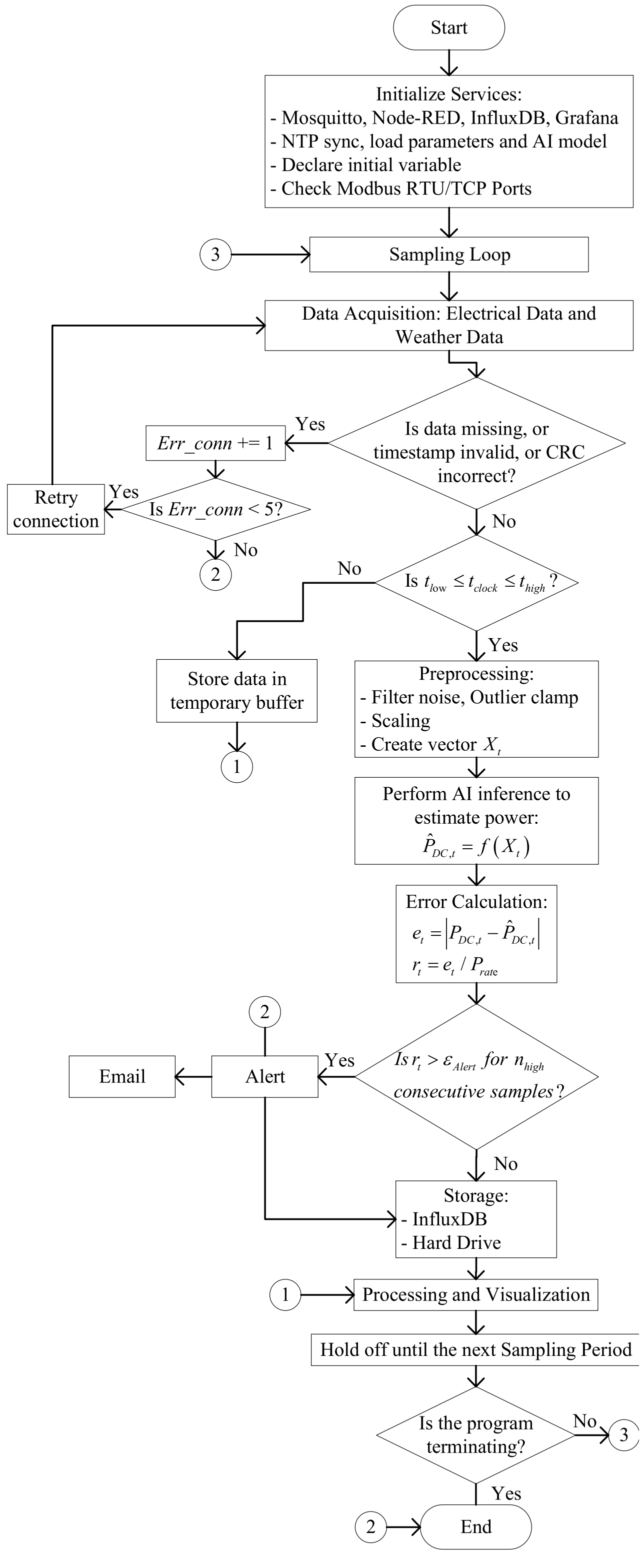

Figure 2. Algorithm flowchart of the proposed AIoT-based solution.

Data were collected at one-minute intervals from January to December 2024, covering both clear-sky and cloudy conditions. During preprocessing, duplicate records and invalid samples (e.g., zero AC power or system downtime) were removed to ensure data quality. The final dataset consisted of 256,588 samples, which were split into 70% for training, 15% for testing, and 15% for evaluation. To ensure consistency, data were collected under clean-panel conditions with bi-weekly maintenance.

1) AI Model Evaluation

The results, presented in Figs. 3 and 4, compare actual DC power with estimates from SVR and XGBoost under different weather conditions. Under cloudy conditions (Fig. 3), both models successfully captured the overall trends; however, errors increased during rapid irradiance fluctuations. SVR tended to smooth variations, resulting in delayed responses and higher local errors. In contrast, XGBoost more effectively captured nonlinear dynamics, resulting in improved tracking and reduced error.

Under stable, sunny conditions (Fig. 4), both models achieved high accuracy, with predicted curves closely matching actual power and correlation coefficients of $R^2 = 0.9994$ (SVR) and $R^2 = 0.9998$ (XGBoost). The results demonstrate that XGBoost performs better under highly variable conditions, while both models remain highly accurate in stable environments. The evaluation metrics (MAE, RMSE, R^2) summarized in Table I further confirm the advantage of XGBoost in complex conditions, whereas SVR remains a reliable choice for smaller-scale systems with limited data.

TABLE I. PERFORMANCE METRICS OF PREDICTING MODELS UNDER DIFFERENT WEATHER CONDITIONS

Conditions	Model	MAE (kW)	RMSE (kW)	R²
Cloudy	SVR	3.21	3.65	0.9982
	XGBoost	1.635	2.309	0.9997
Clear and Sunny	SVR	1.21	1.91	0.9994
	XGBoost	0.93	1.23	0.9998

2) Evaluation of the Integrated AIoT System.

Fig. 5 presents the Grafana dashboard, which integrates multiple layers of system information, including: (i) actual versus estimated power (XGBoost and SVR); (ii) environmental parameters (ambient temperature, cell temperature, humidity, and irradiance); (iii) system resource usage (CPU, memory, and storage); and (iv) alert notifications with different priority levels. The dashboard operates with a refresh interval of 1 second, enabling real-time monitoring of system status.

A notable feature of the system is its proactive alert mechanism, which is triggered when the deviation between measured and estimated power exceeds a predefined threshold. To validate this function, an inverter was intentionally disconnected at 11:00 while suppressing its communication loss alert, thereby simulating a power discrepancy. As shown in Fig. 5, the system generated an alert when the deviation exceeded the configured 10% threshold, indicated by a color transition from green (normal) to red (abnormal). At the same time, an email notification was automatically sent to the administrator. Fig. 6 illustrates a sample alert email, including error type, threshold value, and deviation magnitude.

This mechanism enables timely fault detection, even without continuous operator supervision. Moreover, the system provides operational flexibility through customizable thresholds and operating time windows (e.g., 05:30–18:00), allowing adaptation to different PV system requirements.

In addition, system reliability was verified under network interruption conditions. During connectivity loss, the Raspberry Pi continued to collect data, perform real-time estimation, and store data locally in InfluxDB. Once the connection was restored, all data were successfully synchronized to the cloud without loss, demonstrating the robustness and autonomous operation of the proposed system.

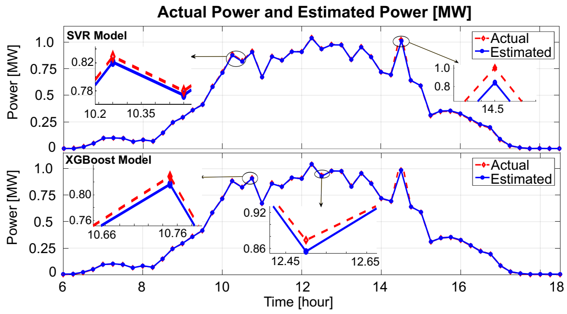

Figure 3. Estimated power and actual power from the XGBoost and SVR model on a cloudy day.

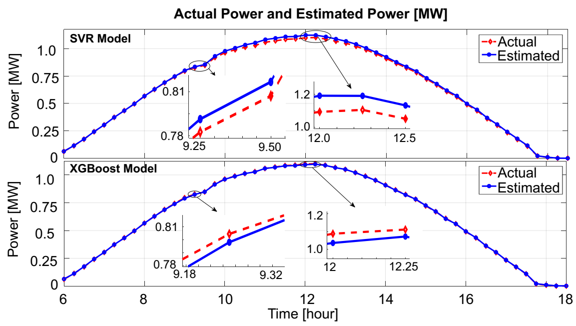

Figure 4. Estimated power and actual power from the XGBoost and SVR models on a clear and sunny day.

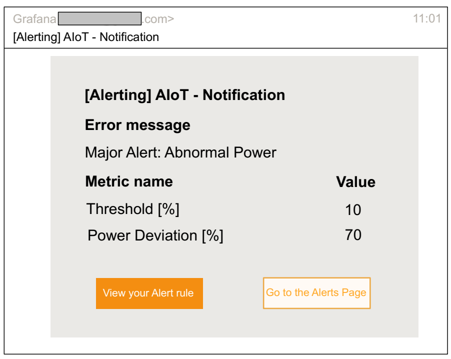

Figure 6. Email Alert.

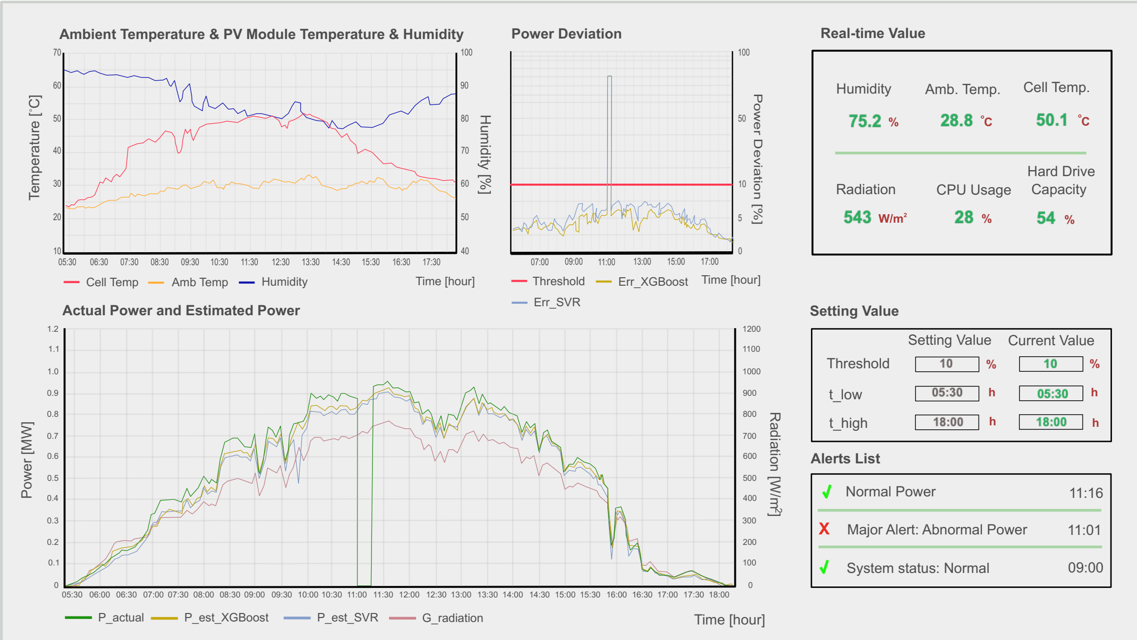

Figure 5. Web–based HMI of PV power system.

IV. CONCLUSION

This study presents a low-cost AIoT framework that integrates machine learning models (SVR and XGBoost) with an open-source IoT platform for real-time DC power estimation in small- to medium-scale rooftop PV systems. The proposed implementation utilizes low-cost hardware (Raspberry Pi, RIKA sensors) with Modbus RTU/TCP communication and the MING stack (Mosquitto, InfluxDB, Node-RED, Grafana) for data acquisition, processing, and visualization.

Experimental results on a 1.2 MWp system under tropical conditions show that both models achieve high accuracy under stable irradiance ($R^2 > 0.99$), while XGBoost outperforms SVR under rapidly fluctuating conditions due to its superior nonlinear modeling capability.

In addition to accuracy, the system demonstrates strong practical feasibility through real-time monitoring and proactive alerting. The dashboard enables simultaneous visualization of power, environmental parameters, and system status, while alerts are triggered when estimation errors exceed predefined thresholds, accompanied by email notifications for early fault detection. Moreover, stable operation during network outages is ensured through local data storage and automatic cloud synchronization.

Overall, the results demonstrate that integrating AI and IoT on a low-cost platform is both feasible and effective for RSP applications, improving monitoring capability, reducing cost, and enhancing operational reliability. Future work will focus on expanding the dataset under diverse conditions, incorporating lightweight deep learning models (e.g., LSTM, CNN), and integrating with Energy Management Systems (EMS) for advanced optimization.

REFERENCES

[1] D. C. Jordan, S. R. Kurtz, K. VanSant, and J. Newmiller, “Compendium of photovoltaic degradation rates,” Prog. Photovolt.: Res. Appl., vol. 24, no. 7, pp. 978–989, 2016.

[2] International Energy Agency, World Energy Outlook, Paris, France: IEA, 2023.

[3] Z. A. Darwish, K. Sopian, and A. Fudholi, “Reduced output of photovoltaic modules due to different types of dust particles,” J. Cleaner Prod., vol. 280, Art. no. 124317, 2021.

[4] T. M. A. Alnasser, A. M. J. Mahdy, K. I. Abass, M. T. Chaichan, and H. A. Kazem, "Impact of dust ingredient on photovoltaic performance: An experimental study," Sol. Energy, vol. 195, pp. 651–659, 2020.

[5] J. Alonso-Montesinos, F. R. Martínez, J. Polo, N. Martín-Chivelet, and F. J. Batlles, “Economic effect of dust particles on photovoltaic plant production,” Energies, vol. 13, no. 23, Art. no. 6376, 2020.

[6] A. Elhammoudy, M. Elyaqouti, D. Ben Hmamou, E. H. Arjdal, D. Saadaoui, S. Lidaighbi, and I. Choulli, "A novel numerical method for estimation the photovoltaic cells/modules parameters based on dichotomy method," Results Opt., vol. 12, Art. no. 100445, 2023.

[7] M. Singla, R. Chauhan, K. S. Gill, and A. Kapruwan, "Usage of a support vector machine learning classifier to identify and sort electrical faults," in Proc. 2024 OPJU Int. Technol. Conf. (OTCON) Smart Comput. Innov. Adv. Ind. 4.0, Raigarh, India, 2024, pp. 1–5.

[8] K. A. Nguyen, D. T. H. Do, H. K. Trinh, N. K. Nguyen, and N. B. Doan, “Machine learning-based estimation of power output in solar photovoltaic systems under real-world conditions,” J. Sci. Technol., Univ. Danang, vol. 23, no. 9D, pp. 79–85, 2025. Doi: 10.31130/ud-jst.2025.23(9D).569E.

[9] S. H. Sajjad, M. E. H. Sajib, M. N. Hasan, and M. A. Razzak, “Design and implementation of an IoT-based solar power monitoring system,” in Proc. 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, pp. 768–773.

[10] T. Özer and Ö. Türkmen, “An approach based on deep learning methods to detect the condition of solar panels in solar power plants,” Comput. Electr. Eng., vol. 116, Art. no. 109143, 2024.

[11] L. O. Aghenta and M. T. Iqbal, “Development of an IoT-based open-source SCADA system for PV system monitoring,” in Proc. 2019 IEEE Canadian Conf. Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 2019, pp. 1–4.

[12] T. A. A. Tran, M. H. Ngo, and K. A. Nguyen, “An IoT– and mobile network–based architecture for monitoring of a solar PV system,” in Proc. 2025 10th Int. Conf. Applying New Technology in Green Buildings (ATiGB), Da Nang, Vietnam, Jul. 2025, pp. 187–192. Doi: 10.1109/ATiGB66719.2025.11142214.