Journal of Solar Energy Research (JSER)

Journal homepage: www.jser.ut.ac.ir

Smart Maintenance with Regression Analysis for Efficiency Improvement in Photovoltaic Energy Systems

İlker Ay[a], Murat Kademli[a], Serkan Savaş[b,c,*], Sotirios Karellas[d], Angelos Markopoulos[d], Christina-Stavroula Hatzilau[d], Philip Devlin[e], Hüseyin Duşbudak[b], Ali Samet Arslan[f], Mustafa Koç[g], Kazım Duraklar[h], Kamil Sunal[i], Mathieu Mehmet Ozer[j]

[a] *Department of Alternative Energy Resources Technology Program, Hacettepe Ankara Chamber of Industry 1st Organized Industrial Zone Vocational School, Hacettepe University, Ankara, Türkiye.*
[b] *Sincan District Directorate of National Education, Ankara, Türkiye.*
[c] *Department of Computer Engineering, Kırıkkale University, Kırıkkale, Türkiye.*
[d] *National Technical University of Athens, Athens, Greece.*
[e] *North West Regional College, Londonderry, Northern Ireland.*
[f] *Impektra IT Software, Ankara, Türkiye.*
[g] *Yenikent Ahmet Çiçek Vocational and Technical Anatolian High School, Ankara, Türkiye.*
[h] *Private Ankara Chamber of Industry Technical College Vocational and Technical Anatolian High School, Ankara, Türkiye.*
[i] *Ankara Chamber of Industry 1st Organized Industrial Zone Management, Ankara, Türkiye.*
[j] *Oryx-Data Incubator EURL, Paris, France.*

ARTICLE INFO

Article Type:

Research Article

Received:04.08.2023
Accepted:08.12.2023

Keywords:
Photovoltaic
Efficiency
Solar energy system
Maintenance and repair
Regression analysis

ABSTRACT

This research had the overarching goal of optimizing maintenance intervals and reducing the maintenance workload by enhancing accessibility for individuals lacking technical expertise in the upkeep of photovoltaic systems, with a particular focus on rooftop applications. The study achieved this objective by employing a linear regression algorithm to analyse climatic parameters such as wind speed, humidity, ambient temperature, and light intensity, collected from the installation site of a photovoltaic solar energy system. Simultaneously, the current and voltage values obtained from the system were also examined. This analysis not only facilitated the determination of power generation within the system but also enabled real-time detection of potential issues such as pollution, shadowing, bypass, and panel faults on the solar panels. Additionally, an artificial intelligence-supported interface was developed within the study, attributing any decline in power generation to specific causes and facilitating prompt intervention to rectify malfunctions, thereby ensuring more efficient system operation.

*Corresponding Author Email:serkansavas@kku.edu.tr

Cite this article: Ay, Kademli, M., Savaş, S., Karellas, S., Markopoulos, A., Hatzilau, C., Devlin, P., Duşbudak, H., Arslan, A., Koç, M., Duraklar, K., Sunal, K., & Ozer, M. (2023). Smart Maintenance with Regression Analysis for Efficiency Improvement in Photovoltaic Energy Systems. Journal of Solar Energy Research, 8(4), 1663-1679. doi: 10.22059/jser.2023.363200.1335

DOI: 10.22059/jser.2023.363200.1335 DOR: 20.1001.1.25883097.2023.8.4.2.4

1. Introduction

In contemporary times, a growing imperative stemming from the decline in fossil fuel resources and an augmented environmental consciousness has instigated the pursuit of more eco-friendly means to meet the escalating global demand for electricity. One prominent response to this challenge is the burgeoning adoption of Photovoltaic Solar Energy Systems (PV-SES). These systems have gained prominence due to their capacity to harness solar energy directly from the sun and their demonstrated trend of declining costs. However, it is essential to recognize that the electrical output generated by PV-SES is intrinsically contingent upon climatic parameters and is subject to instantaneous fluctuations. Consequently, this variability can introduce uncertainties in capacity calculations and potentially impede the consistent attainment of the desired electricity supply [1]. To mitigate these challenges and enhance the precision of electricity production estimations in PV-SES, it becomes imperative to leverage advanced technologies. In this regard, Artificial Intelligence (AI), characterized by its capacity to endow machines with data-driven automatic decision-making capabilities, simulate human cognition, and replicate cognitive functions such as learning, emerges as a pivotal tool. Within the domain of AI, the subsets of machine learning (ML) and deep learning (DL) algorithms assume a central role in improving the accuracy of electricity production estimations within the context of photovoltaic solar energy systems.

Many algorithms such as random forest [2], support vector machines [3], artificial neural networks [4], and tree-based ensemble methods [5] are carried out for prediction of PV power and/or identification processes through data analysis.

In the realm of AI applications for PV-SES, a noteworthy and established domain pertains to fault diagnosis, an area that has garnered attention for a minimum of a decade. Notably, the application of AI in the context of fault diagnosis and maintenance for photovoltaic solar energy systems has witnessed a discernible surge in both discourse and scholarly investigations in recent years [6].

For PV-SES in particular, AI has been applied at least since the year 2000 for applications mainly on PV-SES's performance and on sizing as well as on the prediction of solar radiation data and energy production planning. Solar radiation data is especially necessary in isolated areas where meteorological data are not measurable and/or available. Many studies have been carried out in the literature. For example Sfetsos and Coonick aimed univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques in [7]. Dorvlo et al. used artificial neural networks for solar radiation estimation in [8]. Mellit et al. provided a literature review in their study for AI techniques for sizing PV systems in [9]. A solution was provided with AI for managing a PV energy production unit by Gligor et al., in [10]. Another technique for short-term PV power forecasting is genetic algorithm-based support vector machine which is used by VanDeventer et al., in [11]. ML methods were also used for prediction of full load electrical power output of a base load operated combined cycle power plant by Tüfekçi in [12]. Innovative and smart maintenance opportunities and techniques were explained in [13] by Savaş et al.

Within the spectrum of AI applications in PV-SES, fault diagnosis assumes a pivotal role. An overarching perspective on predictive maintenance for PV-SES reveals three fundamental approaches. The first method encompasses direct visual periodic inspections of all system components, alongside an analysis of the I-V characteristics of the entire PV-SES plant, infrared thermography assessments, and a comparison of present generation data against the plant's actual generation capacity. The second approach relies on the utilization of ML and AI-based forecasting techniques. The third approach centers on the deployment of smart remote monitoring and control systems, facilitated by wireless sensor networks. Recognizing the paramount objective of promptly detecting deviations in the performance of PV-SES, particularly in close proximity to the affected location, remote-control systems that monitor critical parameters such as frequency and voltage stability stand out as the most efficacious means of predictive fault detection [14]. Especially in areas not interconnected to the grid, the option of AI fault detection is of great significance and such opportunities have been investigated for example in [15].

In addition, many studies have been carried out to determine the reasons for the parameters that affect the efficiency, such as pollution and shading in PV-SES's, their occurrence time and the precautions to be taken. An example study [16] was held in California for direct monitoring of energy lost due to soiling on first solar modules by Caron and Littmann. The effect of soiling on the performance of on-site PV characterization was investigated by Kalogirou et al.

in [17]. Guo et al. [18] studied the effect of dust and weather conditions on PV performance in Qatar.

Panel contamination is one of the most important parameters affecting the efficiency of a PV-SES. Detection of panel contamination and cleaning the panels when this contamination reaches a certain level will increase efficiency directly. However, researchers say that panel contamination, which has always been considered homogeneous, is actually not homogeneous [19] and this creates difficulties in determining the contamination index. Many analytical calculations have been made to solve this problem.

As is widely acknowledged, numerous interface programs are available for the remote management of PV-SES. These programs serve to monitor the instantaneous power generation of the PV-SES through the inverter, thus enabling the detection of system errors. However, they typically fall short of providing insights into the underlying causes of these errors. This limitation, in turn, results in efficiency losses, particularly in rooftop PV-SES owned by individuals lacking technical expertise, due to inadequate maintenance or extended maintenance intervals.

In the context of this research, a novel approach is undertaken. The contrubutions of this study can be explained as follow.

- Utilizing the Linear Regression algorithm, real-time data acquired from the PV-SES are meticulously analyzed employing an appropriate comparative technique.
- To identify potential issues such as contamination, shadowing, bypass, and panel failures affecting the PV-SES.
- Establishing both the hardware and software prerequisites necessary to relay this diagnostic information to an online platform in real-time.
- To streamline the operations of PV-SES systems and enhance the accessibility of critical diagnostic data, catering not only to professional operators but also to non-technical users.

The remainder of the paper was structured as follow: In the second section, the material and method used in the study are explained, and information about hardware installations is given. In the third section, the experimental results obtained from the study and the interface created are explained. In the fourth section, previous studies are discussed and in the fifth section, the study is concluded.

2. Literature Review

The integration of AI into solar energy management systems has ushered in a profound transformation in the monitoring, management, and optimization of solar energy generation. These AI-enabled systems possess the capability to convert raw data into actionable insights by leveraging data analytics and machine learning algorithms. Consequently, this integration yields enhanced operational efficiency and facilitates well-informed decision-making within the realm of solar energy management [20].

Yongho [21] harnessed the potential of big data and sensor networks in a comprehensive approach to augment the operational efficiency of solar power plants. His endeavors encompassed the development of an expert system tailored for power generation prediction, the formulation of module unit fault detection techniques, the prediction of the lifespan of inverter components, and the creation of reporting mechanisms. Furthermore, he embarked on the optimization of maintenance procedures and executed the development of a smart monitoring system, which encompasses algorithmic innovations and economic analysis. The culmination of these efforts converged towards the realization of optimal functionality and efficiency in PV power plants.

In recent years, especially thanks to the Internet of Things (IoT), monitoring, maintenance, and management activities have started to be realized.

Kingsley-Amaehul et al. [22] have introduced a model for the monitoring, maintenance, and management of solar PV systems, incorporating the principles of the IoT. This initiative involved the formulation of a mathematical model to represent solar PV systems and the subsequent implementation of algorithms designed to facilitate this model. Additionally, the authors developed an embedded expert system as a proof of concept. An essential facet of this study involved the acquisition of real-time data at both fog and cloud levels, thereby underscoring the robustness of the control topology that was employed. The result analysis revealed that the expert embedded system exhibited an impressive overall accuracy rate of 98.95%. Moreover, a comparative analysis of data collected at fog and cloud levels indicated that the system demonstrated 100% integrity in data communication, accompanied by a remarkable 98% availability.

Rani et al. [23] proposed a project which is based on the use of the up-to-date and cost-effective method for remotely monitoring a solar plant performance by the inclusion of IoT. They specified that the system can assist with plant maintenance, problem diagnostics, and real-time monitoring. Another IoT-based study was held by Sharma et al. in [24]. The IoT is notably recognized for its capacity to enable the automation of hydraulic systems, streamline troubleshooting and maintenance processes, provide electrical support, and enable efficient and effective work monitoring, including in remote and challenging geographical locations.

Chen et al. [25] have also presented an IoT-based framework for an energy-efficient and intelligent street lighting system. This system replaces traditional metal halide lamps with LED lamps designed according to human eye sensitivity, resulting in significant energy savings. The IoT sensors collect data on traffic flow and occupancy, allowing a smart decision-making module to adjust light intensity via Pulse Width Modulation dimming. Moreover, sustainable energy resources, including PV solar panels, battery storage, and smart grids, are employed for optimal power usage, complemented by a maximum power point tracking algorithm for battery charging. Experimental results confirm substantial energy savings not only on highways but also in residential and suburban pedestrian areas, ultimately reducing energy consumption and carbon emissions, with a noteworthy focus on battery performance through the dynamic charging algorithm.

Hema et al. [26] delved into the realm of IoT-enabled applications spanning electricity generation, transmission, distribution, and utilization. Their comprehensive study encompassed the implementation of the physical layer, utilization of models, operating systems, standards, protocols, and architecture of the IoT-enabled Smart Solar Grid system. They addressed various facets of the system, including configuration, solar power system design, IoT device integration, backend systems, workflow and procedures, implementation specifics, test results, and system performance. Additionally, the study delved into the real-time implementation of the smart solar grid system and discussed experimental findings, shedding light on the encountered challenges in the process.

In Shakya's proposed study [27], the primary objective is to establish a maintenance alert system contingent upon the analysis of generated current and voltage data from solar panels. The solar panel system's performance is monitored by comparing its observed values to pre-established calibrated values, which are determined based on various solar radiation conditions. The proposed model is designed to trigger maintenance alerts when significant deviations in power generation by the solar panels are detected, thereby facilitating timely maintenance interventions.

Panda and Das's study [28] provides a comprehensive review of the components within the smart grid architectural model (SGAM). It focuses on exploring the information and communication layers and their integration with power networks through co-simulation. The study also involves mapping smart grid components across various layers, zones, and domains within the SGAM framework. Additionally, the paper addresses cybersecurity challenges, particularly in the context of machine learning, and delves into considerations of interoperability. Finally, the authors lay out future research directions.

Hasankhani and Hakimi present a stochastic management algorithm in their study [29], aiming to minimize the overall cost in a multi-Microgrid (MG) setting. The study encompasses modeling the interactions between multiple MGs, their connections to upstream networks, and participation in the electricity market. The modeling takes into account the impact of renewable resource intermittencies on the market clearing price. The paper also addresses the optimization of renewable resource capacities within the MGs, both before and after their engagement with the electricity market. Additionally, the method's resilience and effectiveness are validated through sensitivity analysis, underscoring its robustness in diverse scenarios.

In the context of renewable energy systems, digital twins and deep learning techniques have found application. Steindl et al. [30] undertook an analysis of digital twin concepts, architectures, and frameworks in existing literature. Their objective was to devise a technology-independent Generic Digital Twin Architecture (GDTA) that aligns with the information technology layers defined in the Reference Architecture Model Industry 4.0. This alignment serves to establish a unified nomenclature and understanding of the proposed architectural structure. To exemplify the application of the GDTA, a proof-of-concept was developed, employing Semantic Web technologies in the instantiation process for a specific use case involving Packed-Bed Thermal Energy Storage. In the work by Shihavuddin et al. [31], the utilization of drone images for the health inspection of PV and wind installations is explored. Their approach introduces a novel method for detecting damage in PV and wind turbines through object detection techniques. The study includes a performance comparison of contemporary object detection methods for assessing damage. Notably, the authors develop a trained model that amalgamates both Infrared and drone imagery to achieve precise damage detection results. To support their research, a dataset containing annotated instances of damages in solar and wind power installations is also introduced.

Upon an analysis of the aforementioned studies, certain limitations in the existing literature become evident, including:

- Several applications remain in the simulation phase, without practical real-world implementation.
- Some applications are confined to laboratory settings and do not address real-world issues.
- Many studies primarily take the form of field surveys and speculative discussions regarding future directions.
- A limited number of studies effectively combine real-time data with practical applications.

This current study has been meticulously designed and executed with the intention of mitigating these prevalent limitations in the literature. The outcomes of this research are anticipated to serve as pioneering and exemplary contributions, offering insights and solutions that can inspire and guide future investigations in the field.

3. Materials and Methods

There are basically two reasons that affect the efficiency of a PV panel: The panel surface temperature (Tp) and the amount of perpendicular radiation to the panel surface (E) [32]. Knowing how these parameters change the maximum operating current (Impp) and maximum operating voltage (Vmpp) values of the PV-SES will enable the detection of contamination, shadowing, and panel malfunctions in the system [33].

3.1. Hardware Setup

In the study context, a meteorology station is established next to the 2.4 kWp fixed PV-SES within our structure. In the installed system, the vertical radiation intensity on the panels was taken with the Kipp Zonen CMP11 pyranometer, the wind speed with the NRG 40C anemometer, and the ambient temperature and humidity values with the Campbell Scientific HC2S3 sensor. While the panel surface temperature was taken with the QM42VT2 temperature sensor placed under the panel, the Impp and Vmpp values were read with the EPA242-RSI amperemeter and EPV242-RSI voltmeter placed at the inverter input. All data were read simultaneously with the MODBUS communication system and sent to our online platform via GSM and evaluated in our AI-based infrastructure (see Figure 1).

The hardware configuration used to create this system is as follows:

Outdoor Temperature: M12FTH3Q

- Slave ID: 3, Register Value: 1, Read Holding Register, Byte: 1, Multipliers: 0.01, 0.

Panel Temperature: QM42VT2Slave ID: 2

- Register Value: 5204, Read Holding Register, Bytes: 1, Multipliers: 0.01, 0.

Sunlight Intensity Measurement: CMP 11

- Slave ID: 1, Register Value: 4001 (2 Words, 4 Bytes, 32 bit Floating Big-endian casting), Read Holding Register, Bytes: 2, Multipliers: 1, 0.

Ampermeter: EPA242-RSI-230VAC

- Slave ID: 5, Register Value: 1, Read Input Registers, Bytes: 1, Multipliers: 0.039, 0 (Calibrated according to the value read from the ammeter in the laboratory).

Voltmeter: EPV242-RSI-230VAC

- Slave ID: 4, Register Value: 1, Read Input Registers, Bytes: 1, Multipliers: 1, 0.

Outdoor Humidity: M12FTH3Q

- Slave ID: 3, Register Value: 2, Read Holding Register, Bytes: 1, Multipliers: 0.01, 0, Shunt Resistor: 60 mV - 20A.

CMP 11, NRG 40C (Anemometer) and NRG 200P sensors are planned to be converted to MODBUS in order to be compatible with our system, since their outputs are different analog signals. In this context, we converted the very low voltage information, which is the output data of the CMP 11 pyranometer device, into MODBUS using Klemsan's ASCON 352 signal converter. After the installation of the system was discovered in the residential area, it was planned to be installed indoors where it would not be affected by external factors, not where the panels are located. In this context, it was decided and carried out to extend the sensor cables by approximately 10 meters. MODBUS inputs were collected in a box and combined as a single MODBUS output. Voltmeter and Ammeter were also put in the same box, 220 VAC connections were made. The hardware system installed and components are shown in Figure 2.

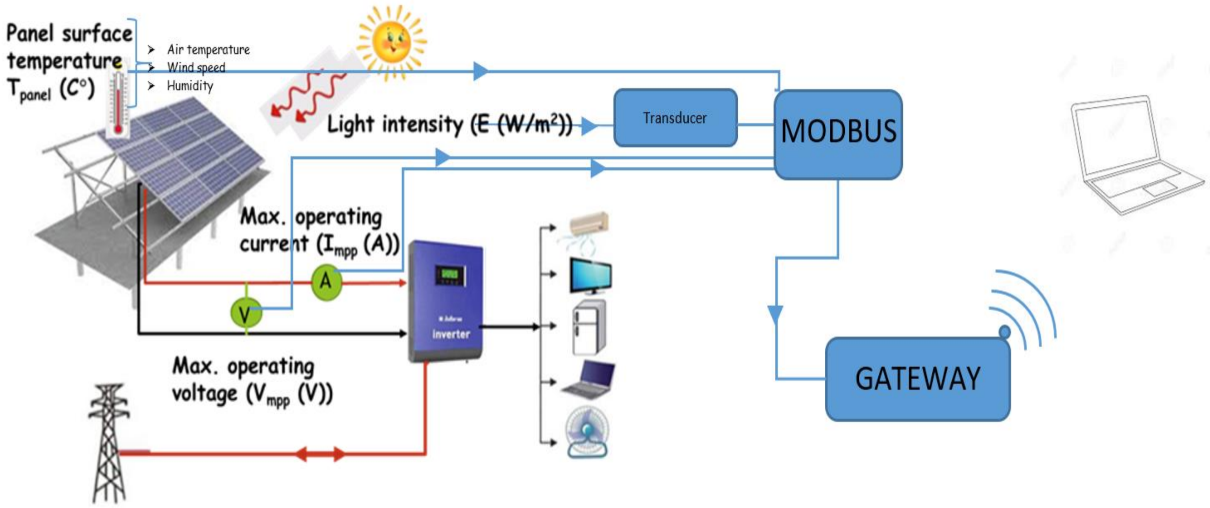


Figure 1. Schematic representation of data acquisition and evaluation

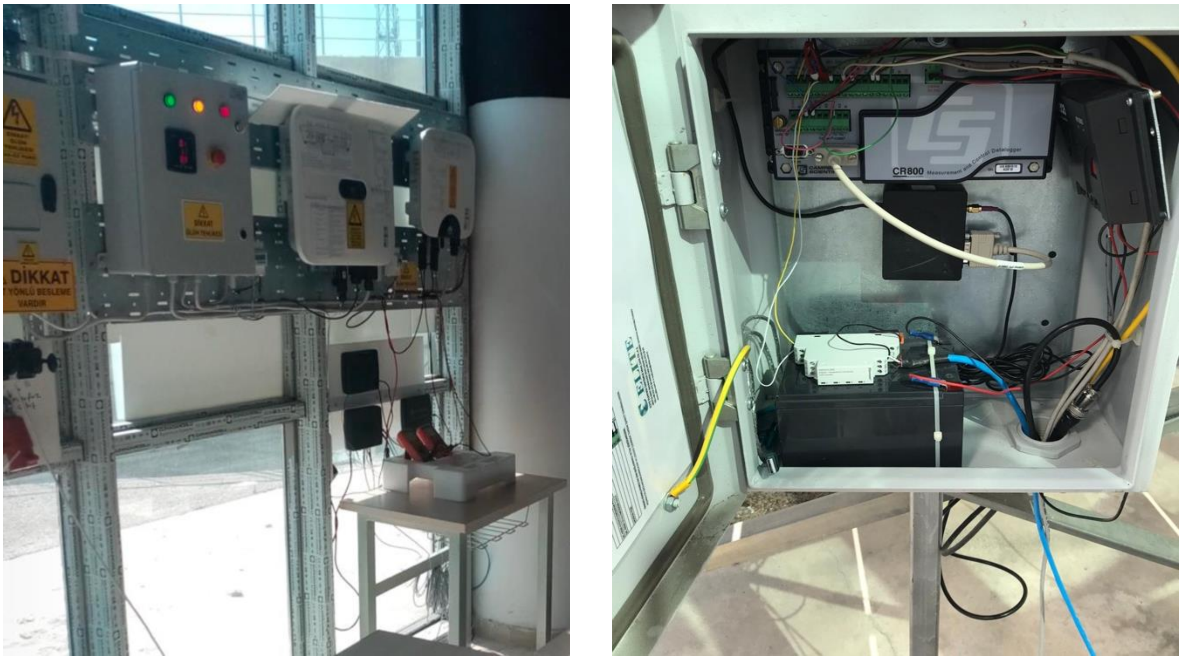


Figure 2. The installed hardware system and components

In order for the AI program, we designed to make the right decisions, the simultaneous data must be in thermal equilibrium. There is no more than a 2% difference between the light intensity data taken consecutively for one minute and the panel surface temperature data as well. When this condition is met, the average of the E(W/m2), Impp(A), Tp(C°), and Vmpp(V) data read for one minute is saved to the file for analysis. Thus, sudden and unrealistic decreases/fluctuations in surface temperature and Impp value caused by sudden changes in cloud, rain, and wind speed have been prevented from entering the data. In addition, the latitude, and longitude information of the place where the system was installed and the settlement angle information of the consumption were loaded into the program as input data, and the calculated azimuth and elevation angles of the sun were compared with this information (Figure 3). Thus, data acquisition has been prevented without sunlight falling on the panels.

For the software to make accurate and appropriate comparisons for installed PV-SES, reference data showing the change of light intensity coming perpendicular to the panels with Impp (Impp-E) and change between panel surface temperature with Vmpp (Vmpp-Tp), should be created for once by operator. Before taking these data, which will be accepted as a reference, it is necessary to ensure that the surfaces of the panels used in the PV-SES are clean and that there is no shading on the system. When these conditions are met and the weather is clear, a controlled measurement is taken once for a certain period of time.

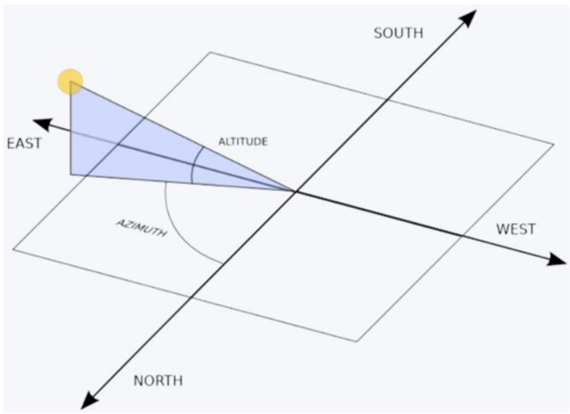


Figure 3. Hardware presets and Azimuth for analysis and calculation

3.2. Feature Extraction

After the reference data and deviation values for current and voltage are determined, these data are loaded into the software once. After that, all the data received are compared with the Linear Regression algorithm by the software in accordance with the conditions given in Table 1, and the errors and their causes are shown to the operator via the interface prepared on the infrastructure.

Table 1. Comparison conditions of the received data (En, In, Tpn, Vn) with reference data

Impp-E graph	Condition	Vmpp-Tp graph	Error / Explanation
for En = Er values	AND	for Tpn = Tpr values	-
In = Ir ± Isd	AND	Vn = Vr ±Vsd	Added to reference data
Ir-Isd> In > Ir-Ir*%10	AND	Vn = Vr ± Vsd	Pollution started
In < Ir -Ir*%10	AND	Vn = Vr ± Vsd	Clean the panels
In < Ir-Isd	AND	Vn > Vr ± Vsd	Power drawing decreased or there is a reflection
In-1 < In	AND	Vn-1 > Vn	Power drawing increased
In-1 ≥ In AND In ≤ Ir - Isd	AND	$V_{n-1} - \frac{V_{n-1}}{n_d} > V_n$	$\left(\frac{V_{n-1} - V_n}{V_{n-1}}\right).n_d$ .x. Bypass activated
If the condition above does not change for 3 hours			Panel failure
En = En-1 ± En-1 * 0,02 AND In >>>> In-1 + In-1*0,5 OR In > 10A AND E < 1000 W/m2			In this situation, the received data is not plotted
If this situation repeats 2 times in a row			Short circuit

In Table 1, *Er* means light intensity in the reference graph and *Ir* means current value corresponding to *Er* in the reference graph. *Tpr* is panel surface temperature on the reference graph, and *Vr* is voltage value corresponding to *Tpr* in the reference graph. *Isd* and *Vsd* are deviation in current and voltage, respectively. Finally, *nd* is total number of bypass diodes in the system.

3.3. Model and AI Infrastructure

The study employed the Linear Regression algorithm, which is a statistical technique used to investigate and model the relationships between variables. In this method, it assumes a connection between the "y" variable that is being estimated and the predictor variables "X_1, X_2,...,X_n." It is further categorized into two types like simple and multiple linear regression. In cases where there is only one independent variable, simple regression is utilized, while multiple regression analysis is employed when there are more than one independent variables. Equations 1 and 2 represent the calculations for simple and multiple regression, respectively [34].

$$y = \beta_0 + \beta_1 X_{1i} + \in_i \quad (1)$$

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \in_i \quad (2)$$

where $i = 1, 2, \ldots, n$ and β_0, β_1, β_2, and β_p are weights.

Regression analysis is a versatile statistical method employed in various domains including energy, finance, production, marketing, health, and agriculture for both descriptive and predictive purposes. When conducting regression analysis, several key concepts should be considered, including outliers, multicollinearity, variance, underfitting, and overfitting. Properly addressing these factors and establishing the most accurate model is crucial for enhancing performance in identification and estimation processes. To determine the correct model, various methods such as evaluating the R-squared value, examining P-values for independent variables, implementing stepwise regression, and employing best subset regression are commonly utilized. These techniques aid in selecting the most appropriate model that best fits the data and facilitates meaningful inferences [35].

Solarpanel application has been developed as a web-based software with layered architecture in a client-server architecture. Layered architecture basically consists of three layers. Data transfer/communication between the client and the server is carried out using application units with Application Programming Interface (API). Different API types are used in software applications depending on the need. Rest API is used as API type in SolarPanel application. It is done by using Hyper Text Transfer Protocol protocol for communication between the client and server with Rest API. Clients are structures that make requests from the server and can use the data on the server. This structure can use the data it receives from the server.

In the layered architecture, database connection, operations such as adding, deleting, updating, and extracting data from the database are performed in the data layer. NoSql, that is, a non-relational database, MongoDB, was used as the database system. The business layer contains business codes and rules are written there. The data, pulled from the data layer, is processed in this layer. The development of the services in the business layer was made with the Nest.JS framework as a Node.js application. In addition, Python 3.7.9, Docker 20.10.5, Postman, and Swagger can be listed in used languages and tools. The processes that interact with the user are performed in the presentation layer. In this layer, the data from the user is directed to other layers. Angular 14.20, HTML 5.0, and CSS 3.0 are the technologies used here.

4. Results and Discussion

In this section, the results of the analysis performed based on the data obtained through the hardware setup established for the research are presented. In addition, the AI-based infrastructure developed for the evaluation of these results is also explained. The reason for choosing this technology and the benefits of the system were explained also in this section.

The current–voltage (I-V) changes of the panels we use, showing the change of the panel surface temperature under constant light intensity and the light intensity falling perpendicularly on the panel under constant temperature, are given in Figure 4.

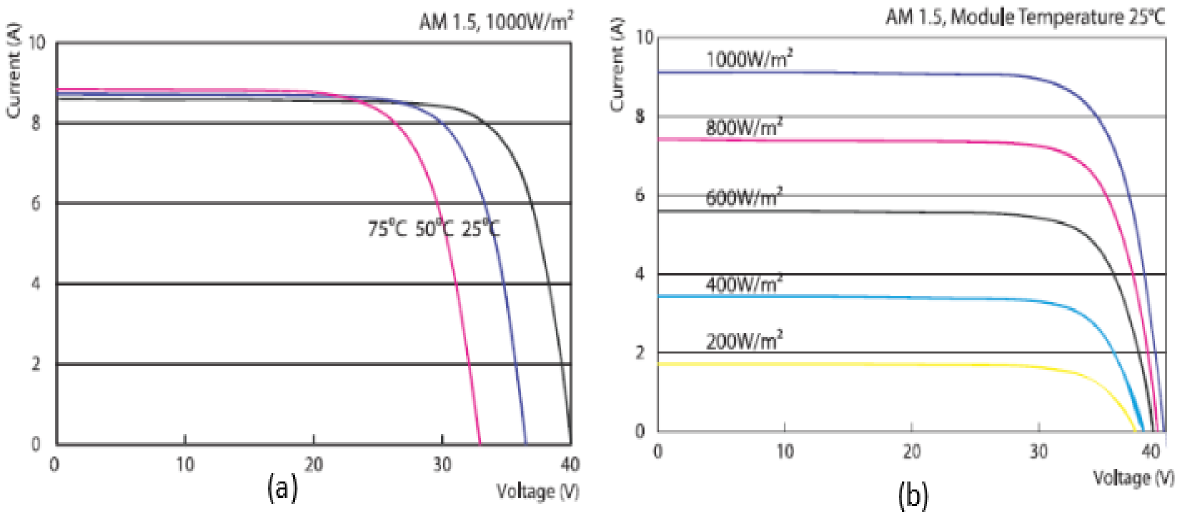
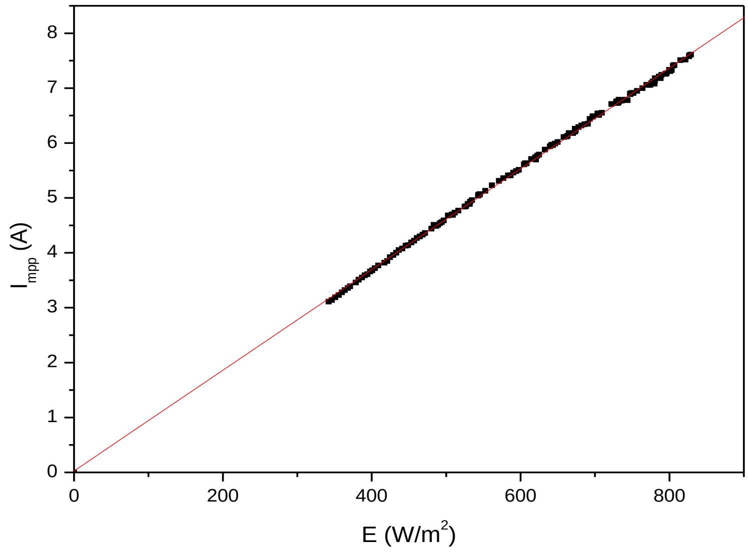


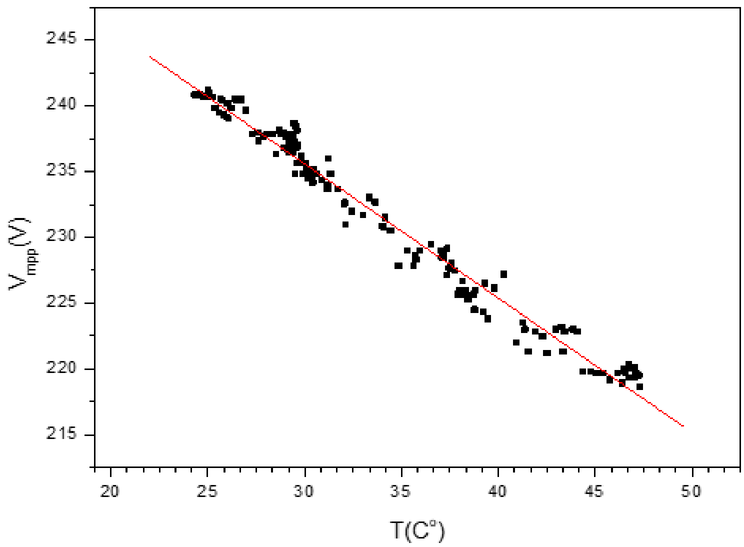
Figure 4. I-V graphs showing the variation of panels with temperature (a) and light intensity (b) under standard test conditions [33]

In the panels, Vmpp=40.88V and Impp=9.55A, an under the standard test conditions of the modules used. The temperature variation coefficient given for the open circuit voltage is βVoc=0.32/°C, while the temperature variation coefficient given for the short circuit current is αIsc =0.05/°C.

The reference data used for the research infrastructure were meticulously arranged, and clean reference data, depicted in Figure 5.a and Figure 5.b, were generated. These reference data pertain to the time frame between 9:00 and 13:00 on October 27, 2022. Throughout the operation of the system, data of reference quality were continuously incorporated into the reference data repository, adhering to the feature selection criteria outlined in Table 1. This iterative process culminated in the determination of optimal value ranges, which were dynamically updated to refine the system's performance.



(a)



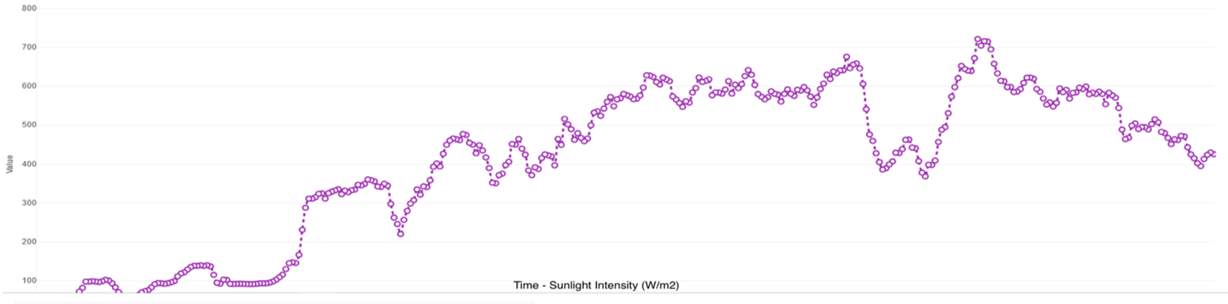
(b)

Figure 5. Reference (a) Impp-E and (b) Vmpp-Tp graphs

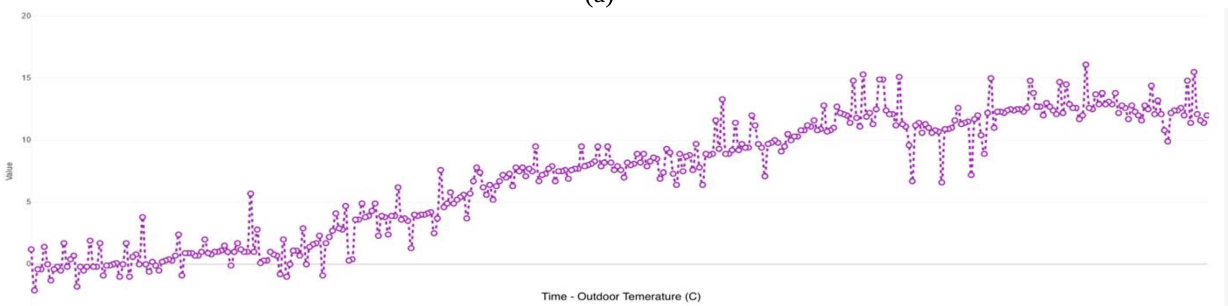
In the process of analyzing the collected measurements, the operator assesses the data's suitability by generating two graphs: the Impp-E graph (Figure 5.a) and the Vmpp-Tp graph (Figure 5.b) using the measured data. During this examination, variations in Impp with temperature and Vmpp with light intensity are disregarded, given that the coefficients of variation for these relationships are minimal, as evident in Figure 4.

As anticipated, Figure 5.b reveals a linear increase in Impp with rising light intensity, while Figure 5.b illustrates a linear decrease in Vmpp as the panel surface temperature increases.

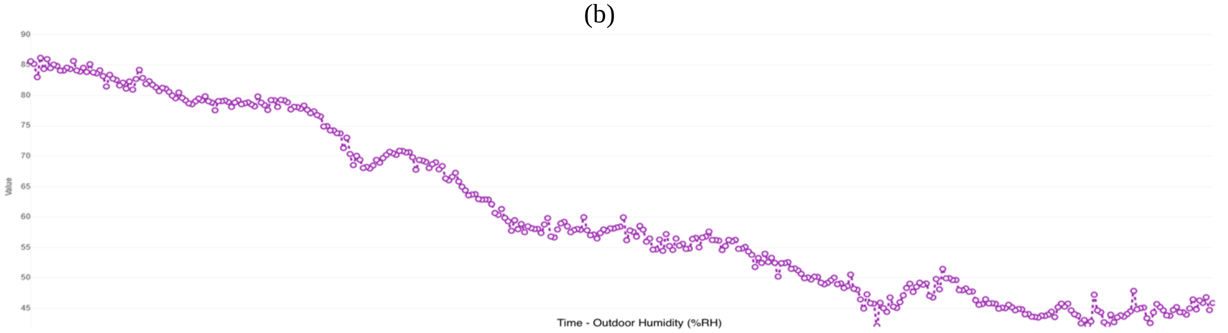
The data collected from the installed hardware system were systematically recorded for subsequent processing using the reference data, and an in-depth analysis was conducted through the application of a linear regression algorithm. The system's functionality is bolstered by preserving the resultant values within the database. The graphics representing the data obtained following the hardware setup, as detailed in the Materials and Methods section, are depicted in Figure 6.


800
700
600
500
400
300
200
100
Value
Time - Sunlight Intensity (W/m2)

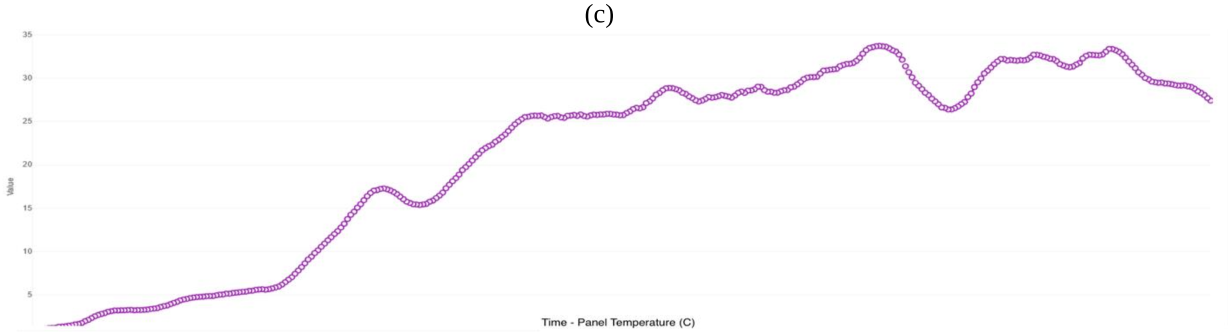
(a)


20
15
10
5
0
Value
Time - Outdoor Temerature (C)

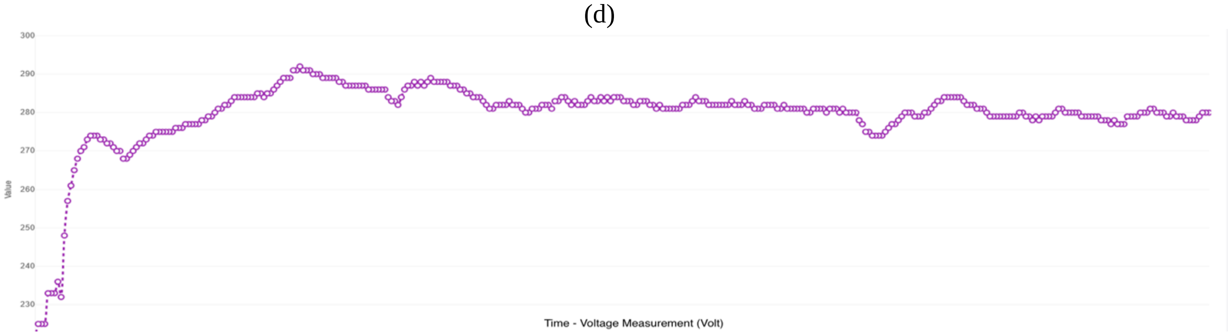
(b)


90
80
75
70
65
60
55
50
45
Value
Time - Outdoor Humidity (%RH)

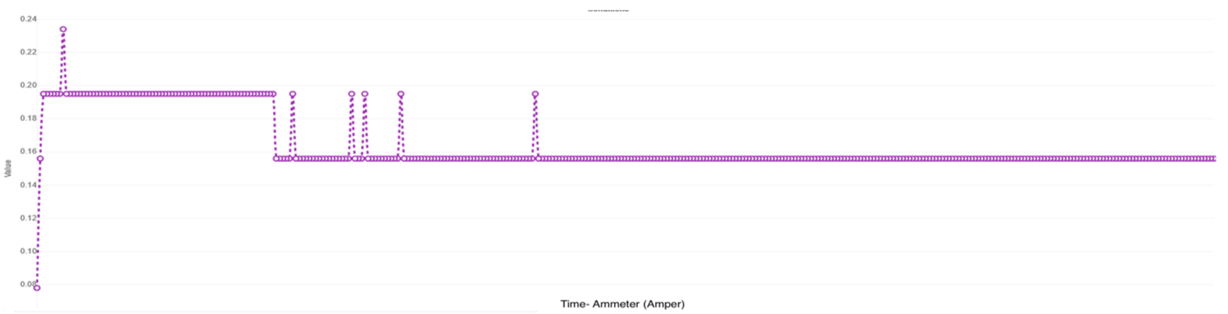
(c)


35
30
25
20
15
10
5
Value
Time - Panel Temperature (C)

(d)


300
290
280
270
260
250
240
230
Value
Time - Voltage Measurement (Volt)

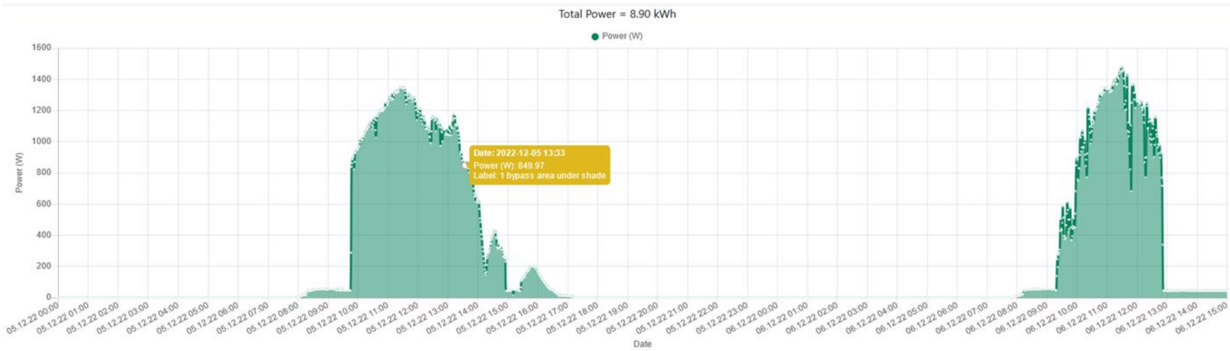
(e)



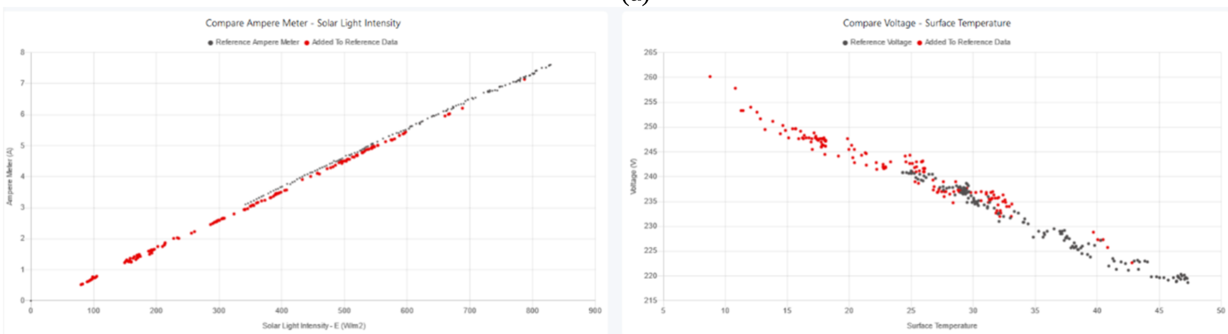
(f)

Figure 6. Sensor data (a) sunlight intensity, (b) outdoor temperature, (c) outdoor humidity, (d) panel temperature, and results obtained with (e) voltmeter and (f) ammeter

The research infrastructure established within this study has been designed with the primary objectives of simplifying the interpretation of the analytical findings and offering early warnings to stakeholders involved in SES. Through the development of these warning systems, the process of maintaining and repairing SES installations has been significantly enhanced. The AI-based interface functioned in alignment with the specified rules for feature extraction and reference data values, as demonstrated in Figure 7. In Figure 7.a, the top part of the screen displays a power-time graph. Users can interact with this graph, allowing them to not only monitor the current power generation but also receive alerts regarding any system errors at any point on the graph. This real-time feedback and diagnostic capability significantly improves the management and performance of SES.



(a)



(b)

Figure 7. (a) Power time sample graph and (b) reference graphics and instant data

At the bottom of the page (Figure 7.b) the Impp-E graph and Vmpp-Tp graph drawn from the reference data (black dots) are given. The AI-based infrastructure compares each instantaneous measurement with the Linear Regression algorithm following the conditions determined in Table 1 and shows the results on the same line (Figure 7.b). It also labels the power time curve. In addition, over time, by adding the data within the deviation range determined from the reference curve to the reference data (red dots), it has Vmpp and Impp values in a wide range of temperature and light intensity.

Thus, this infrastructure will be able to calculate the real efficiency loss of the panels over time from the change in the reference data over the years and compare it with the values given in the catalogue. In addition, in case of a decrease in power generation in the PV-SES, it can detect whether this decrease is due to panel contamination or shadowing.

Depending on the situation;

- It detects the pollution caused by reasons such as dust, sand, and mud accumulating on the panels and gives a warning for cleaning the panels according to the intensity of the pollution.
- In case the pollution on the panel disappears due to meteorological reasons (wind, rain, etc.), it can give a warning that the pollution has disappeared.
- When inhomogeneous contamination occurs on the panels, it perceives this situation as shading, and if the shading is not removed for a certain period of time (3 hours), it gives a panel error and directs the operator to the panel.
- When shading occurs on the panels, it can determine how many bypass diodes are activated, shading time and time.
- It can detect short circuits and panel faults that may occur in the PV-SES and give warnings.

Examples of some of the above errors and warnings received from the panels after installation are presented in Figure 8.

The designed infrastructure, while not explicitly calculating the intensity of contamination on the panels, offers significant advantages. It achieves this by enabling swift maintenance and repairs to panels immediately upon any occurrence. Furthermore, it detects and mitigates situations such as pollution caused by meteorological events, preventing undue increases in workload. This is particularly valuable in the current landscape where rooftop applications are on the rise. The system serves as a valuable guide, even for users lacking technical knowledge about photovoltaic systems. It efficiently shortens the time needed for maintenance and repair from the moment a failure occurs, thus enhancing overall system performance and reducing the associated workload.

Moreover, the system rapidly identifies factors responsible for shadows falling on the PV system. In cases where these factors are within human control, the infrastructure helps in eliminating or mitigating these shadowing issues, further enhancing efficiency. Additionally, by recognizing the factors causing system losses within a year, it allows for the precise scheduling of maintenance cycles, thereby reducing the chances of errors and ensuring smooth operation of the system.

Date	Solar Light Intensity(W/m2)	Ampere Meter(A)	Voltage (V)	Surface Temperature	Outdoor Temperature	Outdoor Humidity	Label
2022-11-29 11:50	0	5.13	212.17	26.84	12.12	56.13	Not Valid
2022-11-29 11:49	0	3.54	213.33	26.53	12.15	57.31	Not Valid
2022-11-29 11:48	0	5.44	203.33	26.47	12.38	56	Not Valid
2022-11-29 11:47	0	5.92	197.5	26.31	12.92	56.82	Not Valid
2022-11-29 11:46	679.3	5.82	197.17	26.18	11.85	56.57	3 bypass area under shade
2022-11-29 11:45	0	5.22	204.33	25.96	12.3	55.84	Not Valid
2022-11-29 11:44	682.59	5.55	209.17	25.74	12.23	57.02	2 bypass area under shade
2022-11-29 11:43	673.22	5.76	238.83	25.79	12.47	57.3	pollution started on the panels
2022-11-29 11:42	669.38	5.74	238.83	25.84	11.85	57.47	pollution started on the panels
2022-11-29 11:41	677.8	5.85	240	25.99	11.93	57.52	pollution started on the panels
2022-11-29 11:40	0	5.65	240.5	26.14	12.45	56.73	Not Valid
2022-11-29 11:39	658.82	5.65	238.5	26.03	12.23	56.98	pollution started on the panels
2022-11-29 11:38	660.16	5.72	239.5	26.13	11.68	57.76	pollution started on the panels
2022-11-29 11:37	676	5.84	240.5	26.37	11.97	56.27	pollution started on the panels
2022-11-29 11:36	704.92	6.15	240	26.32	12.48	54.2	pollution started on the panels

Figure 8. Example system alerts

Especially in the last decade, with the widespread use of deep learning studies, various estimation studies have been carried out for the repair and maintenance of SES. The purpose of these research is to prevent malfunctions before they occur and preventive maintenance and repair.

The study of Sundaram et al. [6] provides insight into the fault detection techniques implemented for PV panels and includes studies related to predictive maintenance needed to improve the performance of the solar PV systems using AI techniques. Also they give detailed overview of fundamental concepts of fault diagnosis algorithm for solar PV system in their study. Mellit [9] and Abubakar et al. [14] outlined the role of AI in PV operation and maintenance, and reviewed the literatures in their studies.

Colmenares-Quintero et al. [36] presented a methodology for automatic fault detection in PV arrays. In the study, nine possible faults are detected by artificial neural networks, caused by malfunction in the bypass and blocking diodes. The obtained models were trained from simulation data and the evaluation shows that the prediction system has a total accuracy of 92.95%. Sfetsos and Coonick [7] also indicated in their study that the developed AI models predict the solar radiation time series more effectively compared to the conventional procedures based on the clearness index.

Tovilović and Đurišić [19] presented direct regression models for forecasting PV system output power based on ML methods such as random forests, extra trees and gradient boosting, in which the Cleanness Index (CI) was introduced as an indicator of the PV panel soiling level. The mean average error of forecasts of the best model, which contained CI, and the model created by excluding CI from the set of input variables were 0.22% and 1.24%, respectively, as related to the nominal power of the PV panel.

Garud et al. [4] carried out a studie to model solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models and to research AI applications on Grid-Connected Solar PV Systems.

The studies discussed here collectively underscore the increasing momentum of predictive and preventive maintenance and repair efforts within the renewable energy sector, particularly driven by the integration of AI algorithms. However, a common limitation in many of these studies is their confinement to simulations or theoretical analyses, with a limited number progressing to practical application.

This present study seeks to overcome this limitation by conducting real-time AI applications. It not only issues real-time system warnings but also dynamically updates the application. Furthermore, it enhances the process by continually integrating relevant data based on feature selection criteria into the reference dataset. As a result, the reference data grows incrementally and becomes increasingly representative of the data over time, reflecting the true conditions and behavior of the renewable energy system, ultimately bridging the gap between theoretical studies and practical application.

5. Conclusions

This work presents a comprehensive monitoring and maintenance prediction framework for PV-SES that leverages multisensory data to streamline the maintenance process and enhance its effectiveness. This system not only accurately estimates power generation but also possesses the capability to predict a range of potential system errors based on continuous signal monitoring.

The experimental setup involved the collection of various data points, including vertical radiation intensity, panel surface temperature, maximum operating current, and voltage, alongside wind speed, air temperature, and humidity values. These data were meticulously recorded, taking into account the avoidance of unrealistic variations due to sudden weather changes. Considering the substantial influence of panel surface temperature and perpendicular radiation intensity on maximum operating current and voltage, reference data were formulated to establish correlations between maximum operating current and light intensity, as well as maximum operating voltage and panel surface temperature.

Subsequently, the obtained measurement results were scrutinized by applying the linear regression method and multiple criteria, leading to the identification of potential flaws in the PV system, such as pollution, shading, variations in power production, panel failure, short circuits, and more.

The monitoring and maintenance prediction system provides a visual representation of the power production's temporal evolution and any identified errors according to predefined criteria. Additionally, it calculates the efficiency loss of the panels over time by comparing them to reference data. This comprehensive approach empowers maintenance teams to take timely corrective actions, ensuring the sustained high efficiency of the PV-SES.

The most important limitation of the study is that the data required for accurate analysis are obtained from panels. For this reason, additional hardware arrangements were made for the study and data were obtained during the project process.

In future studies, meteorological information can be obtained and added to the AI-supported infrastructure. The results can be extended with different ML techniques. Thus, the scope of the study can be expanded and made more widespread.

Acknowledgements

This study was supported by the Erasmus+ Program (KA202 - Strategic Partnerships for vocational education and training) of the European Union under grant number 2020-1-TR01-KA202-093257 (Project name: INNOVATIVE AND SMART MAINTENANCE IN SOLAR ENERGY SYSTEMS). However, the European Commission and the Turkish National Agency cannot be held responsible for any use that may be made of the information contained herein.

Nomenclature

Tp	Panel surface temperature (C°)
E	The amount of perpendicular radiation to the panel surface (W/m^2)
Impp	Maximum operating current (A)
Vmpp	Maximum operating voltage (V)
Er	Light intensity
Ir	Current value
Tpr	Panel surface temperature on the reference graph
Vr	Voltage value in the reference graph
Isd	Deviation in current
Vsd	Deviation in voltage
nd	Total number of bypass diodes
βVoc	Temperature variation coefficient given for the open circuit voltage (/°C)
αIsc	Temperature variation coefficient given for the short circuit current (/°C)

References

[1] Sobri, S., Koohi-Kamali, S. and Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review, Energy Convers Manag, 156, 459–497. DOI: 10.1016/J.ENCONMAN.2017.11.019.

[2] Lahouar, A., Mejri, A. and Ben Hadj Slama, J. (2017). Importance based selection method for day-ahead photovoltaic power forecast using random forests, International Conference on Green Energy and Conversion Systems, GECS 2017. DOI: 10.1109/GECS.2017.8066171.

[3] Shi, J., Lee, W. J., Liu, Y., Yang, Y. and Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines, IEEE Trans Ind Appl, 48(3), 1064-1069. DOI: 10.1109/TIA.2012.2190816.

[4] Garud, K. S., Jayaraj, S. and Lee, M. Y. (2021). A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models, Int J Energy Res, 45(1), 6–35. DOI: 10.1002/ER.5608.

[5] Ahmad, M. W., Mourshed, M. and Rezgui, Y. (2018). Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression, Energy, 164, 465–474. DOI: 10.1016/J.ENERGY.2018.08.207.

[6] Sundaram, K. M., Padmanaban, S., Holm-Nielsen, J. B. and Pandiyan, P. (2022). Photovoltaic Systems: Artificial Intelligence-based Fault Diagnosis and Predictive Maintenance, 1st ed. CRC Press. DOI: 10.1201/9781003202288.

[7] Sfetsos A. and Coonick, A. H. (2000). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, Solar Energy, 68(2), 169–178. DOI: 10.1016/S0038-092X(99)00064-X.

[8] Dorvlo, A. S. S., Jervase, J. A. and Al-Lawati, A. (2002). Solar radiation estimation using artificial neural networks, Appl Energy, 71(4), 307–319. DOI: 10.1016/S0306-2619(02)00016-8.

[9] Mellit, A., Kalogirou, S. A., Hontoria, L. and Shaari, S. (2009). Artificial intelligence techniques for sizing photovoltaic systems: A review, Renewable and Sustainable Energy Reviews, 13(2), 406–419. DOI: 10.1016/J.RSER.2008.01.006.

[10] Gligor, A., Dumitru, C. D. and Grif, H. S. (2018). Artificial intelligence solution for managing a photovoltaic energy production unit, Procedia Manuf, 22, 626–633. DOI: 10.1016/J.PROMFG.2018.03.091.

[11] VanDeventer W. et al., (2019). Short-term PV power forecasting using hybrid GASVM technique, Renew Energy, 140, 367–379. DOI: 10.1016/J.RENENE.2019.02.087.

[12] Tüfekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning

methods, International Journal of Electrical Power & Energy Systems, 60, 126–140. DOI: 10.1016/J.IJEPES.2014.02.027.
[13] Savaş S. et al., (2022). Innovative and Smart Maintenance in Solar Energy Systems, Journal of Information Systems and Management Research, 4(2), 35–49.
[14] Abubakar, A., Almeida, C. F. M. and Gemignani, M. (2021). Review of Artificial Intelligence-Based Failure Detection and Diagnosis Methods for Solar Photovoltaic Systems, Machines 9(12), 328. DOI: 10.3390/MACHINES9120328.
[15] Mellit A. and Kalogirou, S. (2022). Handbook of Artificial Intelligence Techniques in Photovoltaic Systems: Modeling, Control, Optimization, Forecasting and Fault Diagnosis, Handbook of Artificial Intelligence Techniques in Photovoltaic Systems: Modeling, Control, Optimization, Forecasting and Fault Diagnosis, 1–358. DOI: 10.1016/C2019-0-00960-0.
[16] Caron J. R. and Littmann, B. (2013). Direct monitoring of energy lost due to soiling on first solar modules in California, IEEE J Photovolt, 3(1), 336–340. DOI: 10.1109/JPHOTOV.2012.2216859.
[17] Kalogirou, S. A., Agathokleous, R. and Panayiotou, G. (2013). On-site PV characterization and the effect of soiling on their performance, Energy, 51, 439–446. DOI: 10.1016/J.ENERGY.2012.12.018.
[18] Guo, B., Javed, W., Figgis, B. W. and Mirza, T. (2015). Effect of dust and weather conditions on photovoltaic performance in Doha, Qatar, 2015 1st Workshop on Smart Grid and Renewable Energy, SGRE 2015. DOI: 10.1109/SGRE.2015.7208718.
[19] Tovilović D. M. and Đurišić, Ž. R. (2022). Tree-based machine learning models for photovoltaic output power forecasting that consider photovoltaic panel soiling, International Journal of Sustainable Energy, 41(9), 1279–1302. DOI: 10.1080/14786451.2022.2045989.
[20] Mohammad, A. and Mahjabeen, F. (2023). Revolutionizing solar energy: The impact of artificial intelligence on photovoltaic systems. International Journal of Multidisciplinary Sciences and Arts, 2(1), 117-127. DOI: 10.47709/ijmdsa.v2i1.2599
[21] Yoon, Y. (2019). Smart Monitoring System to Improve Solar Power System Efficiency. The Journal of The Institute of Internet, Broadcasting and Communication, 19(1), 219–224. DOI: 10.7236/JIIBC.2019.19.1.219
[22] Kingsley-Amaehule, M., Uhunmwangho, R., Nwazor, N. and Okedu, K. E. (2022). Smart Intelligent Monitoring and Maintenance Management of Photo-voltaic Systems. International Journal of Smart Grid, 6(4), 110-122. DOI: 10.20508/ijsmartgrid.v6i4.260.g246
[23] Rani, D. P., Suresh, D., Kapula, P. R., Akram, C. M., Hemalatha, N. and Soni, P. K. (2023). IoT based smart solar energy monitoring systems. Materials Today: Proceedings, 80, 3540-3545. DOI: 10.1016/j.matpr.2021.07.293
[24] Sharma, M., Singla, M. K., Nijhawan, P., Ganguli, S. and Rajest, S. S. (2020). An application of IOT to develop concept of smart remote monitoring system. Business Intelligence for Enterprise Internet of Things, 233-239. DOI: 10.1007/978-3-030-44407-5_15
[25] Chen, Z., Sivaparthipan, C. B. and Muthu, B. (2022). IoT based smart and intelligent smart city energy optimization. Sustainable Energy Technologies and Assessments, 49, 101724. DOI: 10.1016/j.seta.2021.101724
[26] Hema, N., Krishnamoorthy, N., Chavan, S. M., Kumar, N. M. G., Sabarimuthu, M. and Boopathi, S. (2023). A Study on an Internet of Things (IoT)-Enabled Smart Solar Grid System. In Handbook of Research on Deep Learning Techniques for Cloud-Based Industrial IoT, 290-308. IGI Global. DOI: 10.4018/978-1-6684-8098-4.ch017
[27] Shakya, S. (2021). A self monitoring and analyzing system for solar power station using IoT and data mining algorithms. Journal of Soft Computing Paradigm, 3(2), 96-109. DOI: 10.36548/jscp.2021.2.004
[28] Panda, D. K. and Das, S. (2021). Smart grid architecture model for control, optimization and data analytics of future power networks with more renewable energy. Journal of Cleaner Production, 301, 126877. DOI: 10.1016/j.jclepro.2021.126877
[29] Hasankhani, A. and Hakimi, S. M. (2021). Stochastic energy management of smart microgrid with intermittent renewable energy resources in electricity market. Energy, 219, 119668. DOI: 10.1016/j.energy.2020.119668
[30] Steindl, G., Stagl, M., Kasper, L., Kastner, W. and Hofmann, R. (2020). Generic digital twin architecture for industrial energy systems. Applied Sciences, 10(24), 8903. DOI: 10.3390/app10248903
[31] Shihavuddin, A. S. M., Rashid, M. R. A., Maruf, M. H., Hasan, M. A., ul Haq, M. A., Ashique, R. H., & Al Mansur, A. (2021). Image based surface damage detection of renewable energy installations using a unified deep learning approach. Energy Reports, 7, 4566-4576. DOI: 10.1016/j.egyr.2021.07.045
[32] Raza, M. W., Amin, R., Malik, A. S., Kasi, M., Kasi, B. and Muhammad, F. (2017), Analysis of The Impact of Environmental Factors on Efficiency of Different Types of Solar Cells, Journal of Applied

and Emerging Sciences, 7(1), 76-90. DOI: 10.36785/JAES.71219.
[33] Ay, İ., Kademli, M., Karabulut, Ş. and Savaş, S. (2022). Affecting Factors of Efficiency in Photovoltaic Energy Systems and Productivity-Enhancing Suggestions, 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), 1–6. DOI: 10.1109/ASYU56188.2022.9925271.
[34] Yıldıran, Y. (2022). Doğrusal Regresyon Modeli, in Teori ve Uygulamada Makine Öğrenmesi, Savaş S. and Buyrukoğlu, S. Eds., 1.Ankara: Nobel Akademik Yayıncılık Eğitim Danışmanlık TİC. LTD. ŞTİ., 21–36.
[35] Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2005). Applied linear statistical models, 5, McGraw-Hill Irwin Boston.
[36] Colmenares-Quintero, R. F., Rojas-Martinez, E. R., Macho-Hernantes, F., Stansfield, K. E. and Colmenares-Quintero, J. C. (2021). Methodology for automatic fault detection in photovoltaic arrays from artificial neural networks, Cogent Eng, 8(1). DOI: 10.1080/23311916.2021.1981520.