

Method for Monitoring and Identifying PV (Photovoltaic) System Failures Using Machine Learning

Meenakshi A. Thalor¹, Domale Rutuja^{2*}
AISSMS IOIT Pune, India

Corresponding Author: Domale Rutuja; domalerutuja10@gmail.com

ARTICLE INFO

Keywords: Artificial intelligence, Green energy supplies, Observing automated Intelligential, Sustainable power sources, Control

Received : 5, August

Revised : 18, September

Accepted: 20, October

©2023Thalor, Rutuja: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

Artificial intelligence techniques have been utilized to address intricate practical challenges in various domains and are gaining popularity in the contemporary era. The principal aim of this article is to assess the prediction of power generation in three distinct photovoltaic configurations and the surveillance of measurement sensors, employing artificial intelligence and data extraction, to conform to the behavior of environmental factors in the examined region. Additionally, it encompasses the incorporation of the resulting models into the SCADA system using benchmarks, allowing the operator to actively monitor the power grid. Furthermore, it provides a method for real-time simulation and anticipation of photovoltaic systems and measurement detector within the framework of intelligent system.

INTRODUCTION

The incorporation of green Energies (RE) into the electrical network increases the intricacy of network control, requiring a focal point on maintaining provider reliability and the balance between production and intake because of the sporadic and uncertain nature of RE sources. As a end result, it is vital to intensify studies and improvement efforts at numerous authorities and administrative ranges to harness RE resources and deal with global electricity needs. within this framework, the thing explores the usage of tracking and failure detection strategies primarily based on device mastering for Photovoltaic (PV) structures. The number one objective of this take a look at is to evaluate strength manufacturing predictions for three awesome photovoltaic structures and the oversight of measurement sensors. system getting to know and records mining are hired to reply to the varying climatic conditions in the studied region. The three photovoltaic systems underneath research are: 1) Polycrystalline, 2) Monocrystalline, and three) Axis-tracking. moreover, this research encompasses the incorporation of the consequent models into the SCADA (Supervisory manage And information Acquisition) system thru overall performance indicators, facilitating energetic control of the energy grid via operators.

THEORETICAL REVIEW

Many challenges await in the realm of Big Data (BD) technology for intelligent grids. These obstacles encompass facets such as data consolidation, storage, real-time data handling, data compression, advanced data representation, and concerns regarding privacy and data security. It illustrates various existing applications of BD and "Machine learning customized for Renewable Energy (RE) administration in smart cities. Analytics and extensive data possess the capability to enhance the precision of wind, solar, and load predictions by efficiently processing vast volumes of historical data. These subjects can all be examined through a database utilizing machine learning techniques and their derivatives. As a result, advancements in various energy management applications, including ER resources and distributed storage systems in intelligent grids, are currently underway. Artificial Intelligence (AI) methods find applications across a wide spectrum of RE sources, encompassing wind, solar, geothermal, hydroelectric, oceanic, and hydrogen energy. AI plays a pivotal role in design, optimization, control, estimation, management, distribution, and economics. The current and future of RE primarily involve the development of innovative technology to maximize production from available natural resources, promote environmental awareness, and establish advanced management and distribution systems, as previously detailed in research.

Similar to other sectors (health, education, business, technology, industry, security, etc.), AI holds promise for attaining future RE objectives and, within that, contributes to the field of machine learning. Machine learning techniques have been instrumental in solving intricate real-world challenges across various domains and are gaining increasing popularity in today's landscape.

PV System

The solar power system contains 136 Ateresia brand sun panels, each with a energy score of 250Wp. those panels are divided into three instructions: 60 monocrystalline panels (15kW), 60 polycrystalline panels (15kW), and sixteen polycrystalline panels with single-axis tracking (5kW). each machine is associated with an unbiased DC/AC inverter that interfaces with most people grid. The PVS set up is geared up with 136 sun panels, every rated at 250Wp, all synthetic with the aid of way of Atersa. the ones panels are grouped into three wonderful sorts: 60 monocrystalline panels with a cumulative ability of 15kW, 60 polycrystalline panels with the equal 15kW capability, and a further sixteen polycrystalline panels supplying single-axis tracking, totaling 5kW. every of these subsystems connects to its devoted DC/AC inverter, that's connected to most of the people grid. in the Photovoltaic device (PVS), there are a whole of 136 solar panels, with each panel rated at 250Wp and artificial thru Atersa. those panels are further classified into 60 monocrystalline panels, offering a collective electricity output of 15kW, 60 polycrystalline panels with an identical 15kW functionality, and 16 polycrystalline panels organized with unmarried-axis tracking, providing a combined capability of 5kW. every of those subsystems connects to an impartial DC/AC converter for connection to the public grid.

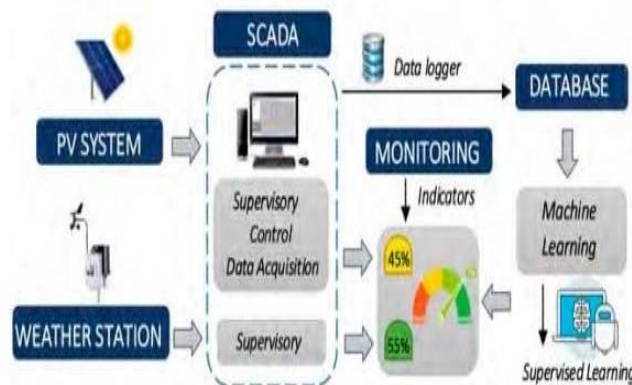


Figure 1: Illustration of machine learning implementation in a photovoltaic (PV) system.

Database

The data repository is generated using the variables mentioned earlier and saved in the server's data recorder through the Data Recording and Supervisory Control Module, along with the Data Connection Tool of the LabVIEW 2015 software, utilizing the communication interface via National Instruments (NI) OPC servers. This approach enables the acquisition of variable values and their transformation from analog to digital signals, preparing them for subsequent observation and archival purposes. The database used in this investigation corresponds to a one-year data history.

Item	Description	Panels (series-parallel)	Type	Max. Power
1 st	PVS1 Photovoltaic System 1	60 (15 × 4)	Polycrystalline	15 kW
2 nd	PVS2 Photovoltaic System 2	60 (15 × 4)	Monocrystalline	15 kW
3 th	PVS3 Photovoltaic System 3	16 (16 × 1)	Polycrystalline (tracking)	5 kW

Table 1: Characteristics of photovoltaic systems

Surveillance

The primary focus of this study centers on this aspect. It puts forward an application blueprint for overseeing this photovoltaic setup. Additionally, it scrutinizes the data stored in the database and the execution of the models within the SCADA system, serving as a resource or assistance in instances of malfunctions in the PV electrical system or errors in sensor measurements when they surpass predetermined thresholds. This approach, involving signals associated with the photovoltaic system's variables and real-time weather station data, empowers the electric grid operator to enhance the safety and dependability of the PV system and its integration with the grid.

Supervisory Control and Data Acquisition Device

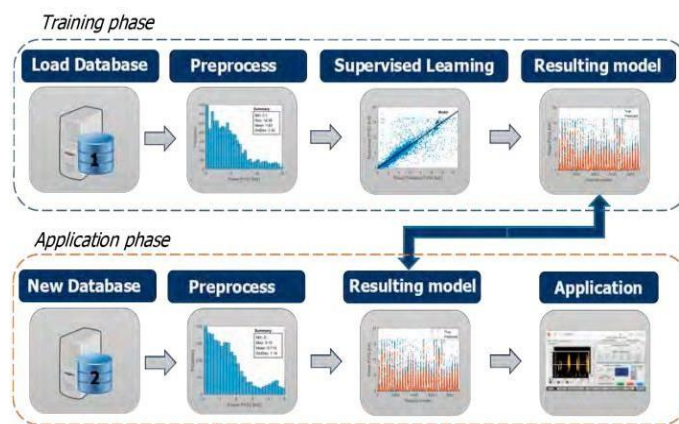
The supervisory control and data acquisition (SCADA) device provides the capability to manage the activation and deactivation of the DC/AC inverters within the photovoltaic (PV) system, including their connection to or disconnection from the electrical grid. Furthermore, it offers oversight of the electrical parameters encompassing current, voltage, and power in both DC and AC for each PV system. Data collection is achieved through the utilization of measurement sensors and network analyzers, which are connected via a protocol through Modbus communication. This enables the examination, recording, and retrieval of data within the local server.

Machine learning

Recent advancements in computing and the reduction in hardware costs have paved the way for the development of novel methods for extracting information. Machine learning, a subfield of artificial intelligence, is dedicated to creating and studying systems capable of analyzing data without the need for explicit programming. It specializes in recognizing patterns and extracting knowledge from data, making it one of the most valuable techniques for data extraction today. Machine learning algorithms leverage computational techniques to autonomously glean insights from data, bypassing the need for predefined equations as models. Generally, there exist nine widely-used machine learning algorithms, including K-means, Linear Support Vector Machines (LSVM), Logistic Regression (LR), Locally Weighted Linear Regression (LWLR), Gaussian Discriminant Analysis (GDA), Back-propagation Neural Networks (BPNN), Expectation Maximization (EM), Naive Bayes (NB), and Value-Added Tax (VAT). Each of these algorithms possesses its unique characteristics, making them suitable for various scenarios.

Various artificial intelligence methodologies, such as machine learning, genetic algorithms (GA), and neural networks (NN), have been suggested and employed for modeling and predicting solar irradiance. It's worth noting that neural networks, while effective, have the drawback of prolonged training times and the requirement for substantial user intervention. Additionally, it's crucial to acknowledge that a larger volume of data and relevant information (data mining) related to a specific study subject enhances the chances of identifying an accurate application model.

Considering the objectives of this case study, data analysis is carried out through supervised learning, entailing the creation of a predictive model based on known input and response data. Two primary categories of machine learning algorithms are employed: classification, geared toward databases containing qualitative values (words), and regression, suitable for quantitative (numerical) databases. Based on the numerical characteristics of the databases obtained from the SCADA system, the regression category is employed for analysis. The workflow for establishing an optimal model consists of two phases: the training phase, aimed at effectively applying a learning approach to achieve a performance feature. This phase is further divided into four stages, including initial database input, preprocessing (involving filters, statistical summaries, and cluster analysis), classification of supervised learning (categorized or regression), and ultimately model acquisition. In the application phase, a new database is introduced, preprocessing is repeated, and the model obtained in the training phase is utilized to identify essential features and architectural parameters for each model to define the predictive model. Lastly, the model's performance is evaluated.



Software for SCADA

Accurate forecasting of the electrical output from photovoltaic (PV) systems stands as a critical requirement for the proper operation of the electric grid and the optimal management of power flows within the PV system. Consequently, it becomes imperative to incorporate a real-time monitoring system to ensure operational safety and oversight of the electrical systems supervised by the Supervisory Control and Data Acquisition (SCADA) system. Under this framework, the integration of linear regression models, developed within this research, is envisaged. It's worth noting that to establish a real-time monitoring system, a comprehensive evaluation of the

corresponding database is indispensable. By executing all the steps to approximate actual values with predictions, once the model is formulated to address a wide array of scenarios and circumstances, its implementation becomes viable.

A power utility operator must consistently maintain a delicate balance between electricity generation and consumption. This task often proves challenging when using conventional and manageable power production systems, especially in small or non-interconnected systems. Hence, the imperative need for real-time capabilities. However, to effectively deploy a real-time predictive model, a preliminary examination of databases is obligatory. Subsequently, the application of the resulting models within the SCADA system is possible, entailing the input of real-time measurements. In this context, the SCADA system performs measurements, discretization, and calculations within the corresponding model framework.

Fault Detection Methodology

A method for the detection and analysis of failures, as put forward by Garoudja, encompasses four fundamental stages: (i) the extraction of parameters from the photovoltaic (PV) module, (ii) validation of the model, (iii) the development of pertinent data sets, and ultimately (iv) the identification and analysis of faults. In this investigation, the panel configuration conforms to 15x4 (parallel series) for both the PVS1 and PVS2 systems. Consequently, if a fault occurs within one of the branches, the fuse protection mechanism is triggered, and the total power output of the PVS1 system is reflected at the measurement point. In this manner, when scrutinizing the photovoltaic production during two typical days, one without any failure and the other with a failure, it becomes feasible to assess the performance of both systems in comparison to the predictive model. In the first scenario, the values fall within the established 20% limit. For example, a data point such as (186, 7.18) and (187, 5.57) indicates a variation of $7.18 - 5.57 = 1.61$ kW, corresponding to a 10.73% deviation. In contrast, the second case exceeds the limit by $12.85 - 6.96 = 5.89$ kW, amounting to a 39.27% discrepancy.

Moreover, the precise moment of branch failure within the PVS1 system is scrutinized, as depicted in Figure 20 of the report, approximately at 1,140 seconds. The reference photovoltaic production experiences a significant surge over a broad range of data, triggering the activation of alarm 2a at a high value. Alternatively, if the power level falls within the 20% threshold, alarm 2a registers a low value. The application of these equations is executed within the SCADA system, and through modifications in the LabVIEW Software, real-time calculations have been integrated. This empowers the operator to attain a more precise reference of the PV production and enhance the detection of failures for optimal operational management.

METHODS

The incorporation of Renewable Energies (RE) into the electrical grid adds complexity to grid management, aiming to ensure service continuity and balance in production and consumption. This is due to the intermittent and unpredictable nature of RE sources. Hence, there is a growing necessity to direct more attention toward research and development at governmental and other levels to explore RE resources and meet global energy demands. Accurate prediction of photovoltaic (PV) production is a crucial aspect of seamlessly integrating this technology into existing power systems, particularly for electric grid operators. However, there are two primary concerns associated with high penetration rates of PV systems: their intermittent nature and the uncertainty of their availability.

Furthermore, poorly performing photovoltaic panels can lead to gradual or rapid declines in energy output. A study indicates that it's feasible to predict the daily energy curve of a PV panel based on the energy curves of neighboring panels, employing neural networks, which facilitates monitoring for optimal performance. Precise forecasting of PV production can alleviate the effects of power quality issues that are prevalent in distributed systems, enhancing the active management of the electric grid. It serves as a vital function that aids organizations and operators in energy management and economic dispatch planning. The electricity generated by a PV system at any given time is directly proportional to the sunlight it receives. However, sunlight levels fluctuate with seasons, times of day, geographical location, and panel orientation. Therefore, accurate forecasting of solar radiation and its corresponding power output is essential, enabling operators to take the necessary measures and manage intermittency effectively.

The approach advocated in this research is based on implementing sensors within Renewable Energy Systems (RES) operations and leverages big data technology for data processing and analysis to predict PV system performance. In their work [8], the authors provide a comprehensive overview of PV power generation forecasting methods using machine learning and metaheuristic techniques, encompassing classifiers such as (i) persistence methods, (ii) statistical techniques, (iii) machine learning approaches, and (iv) hybrid methods. Robust data analysis tools with capabilities and flexibility are employed to handle power system data for analysis and information extraction, aimed at improving power quality, enhancing distribution efficiency, optimization, and machine learning, among other objectives. Following this same principle, the models are adjusted as new data becomes available. However, achieving an accurate model necessitates the analysis of a substantial volume of data for training, which can be challenging for novel systems and applications where such data might be scarce. Notably, the generation of "big data" (BD) in the power system is still in its early stages, with a long road ahead. An intelligent fault diagnosis approach for PV arrays, founded on an advanced rotation forest algorithm, has been proposed. This approach involves feature selection and classification based on two rotation forest (RoF) algorithm classifiers combined with Extreme Learning Machine (ELM) for PV array fault analysis. Additionally, a 9.54 kWp PV system has introduced a novel method for detecting and diagnosing PV system failures, centered around a Pink

Probabilistic Neuronal Classifier (PNN). Several related studies have examined solar photovoltaic systems from the standpoint of modeling and simulation. However, the performance of these systems varies as climatic conditions differ across regions.

Hence, it's vital to conduct further investigations with real-world data to assess their behavior under normal operating conditions.

In this study, we present a predictive model of electrical power (kW) generated by three distinct photovoltaic systems: polycrystalline, monocrystalline, and single-axis tracking. Machine learning and data mining techniques are applied to SCADA databases (Supervisory Control and Data Acquisition) and climatic data obtained from a weather station. The primary objective is to establish a model using real-time data that enables a comparison between actual PV production and the model-predicted PV production. Moreover, the research delves into fault detection techniques using the derived equations. The results are implemented in the SCADA system, empowering the operator to attain better monitoring, control, and fault detection capabilities in photovoltaic power production. Lastly, this article is an extension of the report presented at the ICSC-cities 2019 conference, titled "Machine Learning Data Applied to Monitoring PV Systems: A Case Study." Some of the novel aspects of this article revolve around the application of the PVS1, PVS2, and PVS3 models for fault detection in photovoltaic systems, involving a comparative analysis between real-time photovoltaic generation and model-predicted values, with an established 20% allowable range to determine measurement failures or errors through alarm signals.

RESULT & DISCUSSION

The implementation of equations derived in this study has enabled the acquisition of reference data for the photovoltaic power generated by PVS1, PVS2, and PVS3. Consequently, they can serve as a system for monitoring and fault detection, especially when comparing real-time photovoltaic power generation with values calculated based on meteorological variables such as radiation and temperature. The equations for temperature measurement modeling and radiation measurements also provide a means to observe sensor behavior and eliminate potential measurement errors over time. By training on 75% of randomly collected data, we have been able to adapt to varying values during both winter and summer months. When comparing the PVS1 and PVS2 systems, the coefficients of the equations displayed some similarity, given their identical installed power capacity of 15 kW. However, minor variations in parameter values were observed due to the differences in the monocrystalline and polycrystalline cell types. Through data training, they were fine-tuned to achieve optimal performance. In contrast, the PVS3 system, with its monitoring function and lower installed power capacity of 5 kW, exhibited significantly different parameter values. This is primarily because the parameters needed to be adjusted differently for this specific configuration.

CONCLUSIONS AND RECOMMENDATIONS

As industrial progress advances, automation and processes generate increasing volumes of data that necessitate analysis, interpretation, and communication. Therefore, this research has demonstrated the utilization of machine learning methods in the analysis of real-world data and the creation of predictive models. Consequently, this study has proven that it is viable to forecast the photovoltaic power output of the three examined systems using regression models, achieving a high degree of accuracy. The importance of monitoring variables through measurement sensors has afforded us effective control over the photovoltaic system. The application of a fault detection approach has been validated through predictive model techniques in PV systems, allowing us to monitor PV systems by comparing real-time photovoltaic power generation with calculated values based on meteorological variables like radiation and temperature.

Correlation coefficients of 83.27%, 82.36%, and 85.76% were obtained in the model results for the PVS1, PVS2, and PVS3 systems, respectively. An allowance of 20% has been established, permitting the comparison of calculated values with real-time measurements. The equations pertaining to temperature measurement models and radiation measurements also enable the monitoring of sensor performance and the exclusion of potential measurement errors over time. In this manner, an additional means of monitoring and controlling systems using these parameters has been acquired. The implementation of predictive models for PV systems within the SCADA system enables optimal monitoring by the electric grid operator. Ultimately, the significance of applying machine learning techniques and their broad utility in the field of energy management and their role in smart grids has been affirmed.

ACKNOWLEDGMENT

Achievement of any project hinges significantly on the motivation and directions provided by numerous individuals. This research endeavor would have been unattainable without their assistance. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. First and foremost, we wish to record our sincere gratitude to the mentor of our team and to our Respected HOD Mrs. Meenakshi Thalor, for her constant support and encouragement in the preparation of this report and for the availability of library facilities needed to prepare this report. Our numerous discussions were extremely helpful. We are highly indebted to her for her guidance and constant supervision as well as for providing necessary information regarding the project & also for her support in completing the project. We hold her in esteem for guidance, encouragement and inspiration received from her.

REFERENCES

[http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S012062302022000100026#:~:text=A%20method%20of%20detection%20and,iv\)%20fault%20detection%20and%20diagnosis](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S012062302022000100026#:~:text=A%20method%20of%20detection%20and,iv)%20fault%20detection%20and%20diagnosis)

https://www.researchgate.net/publication/342760289_Method_of_monitoring_and_detection_of_failures_in_PV_system_based_on_machine_learning_techniques

https://www.researchgate.net/publication/342760289_Method_of_monitoring_and_detection_of_failures_in_PV_system_based_on_machine_learning_techniques

<https://jeas.springeropen.com/articles/10.1186/s44147-023-00200-0>

- C. Voyant and et al, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, May 2017. [Online]. Available: <https://doi.org/10.1016/j.renene.2016.12.095> .
- S. Theocharides, G. Makrides, G. E. Georghiou, and A. Kyprianou, "Machine learning algorithms for photovoltaic system power output prediction," in 2018 IEEE International Energy Conference (ENERGYCON), Limassol, Cyprus, 2018.
- C. Kurien and A. K. Srivastava, "Scope of artificial intelligence techniques for exhaust emission prediction of CI engines and renewable energy applications," *International Journal of Engineering Research in Computer Science and Engineering*, vol. 5, no. 2, pp. 456–461, Feb 2018.
- S. Preda, S. Vasilica, A. Bâra, and A. Belciu, "PV forecasting using support vector machine learning in a big data analytics context," *Symmetry*, vol. 10, no. 12 December 2018. [Online]. Available: <https://doi.org/10.3390/sym10120748>