Sensor fault detection in photovoltaic systems using ensemble learning-based statistical monitoring chart

Fouzi Harrou, Ying Sun
King Abdullah University of Science and Technology
CEMSE Division
Thuwal, Saudi Arabia
fouzi.harrou.2016@ieee.org

Abdelhakim Dorbane
SSL Lab, Department of Mechanical Engineering
University of Ain Temouchent
Ain Temouchent, Algeria
abdelhakim.dorbane@univ-temouchent.edu.dz

Benamar Bouyeddou
LESM Lab, Faculty of Technology
University of Saida-Dr Moulay Tahar
Saida, Algeria
benamar.bouyeddou@univ-saida.dz

***Abstract*—Photovoltaic systems have become increasingly popular as a source of renewable energy due to their environmental benefits and cost-effectiveness. However, sensor faults can significantly impact the performance of photovoltaic systems, resulting in reduced energy output and increased maintenance costs. This paper presents an effective approach for detecting sensor faults in photovoltaic (PV) systems using ensemble learning and the exponentially weighted moving average (EWMA) chart with nonparametric threshold estimation. The proposed approach trains the ensemble models using data collected during normal operating conditions of the PV system and detects any sensor faults by analyzing the residuals generated from the ensemble models. The EWMA chart is then applied to track changes in the residuals over time and detect any abnormalities. The flexibility of the chart is enhanced by computing the detection threshold using kernel density estimation (KDE). This approach improves the accuracy and reliability of the fault detection process. The proposed approach is assessed based on simulated data from a PV system using PVGIS. The results of the study demonstrate that the proposed method effectively detects sensor faults in photovoltaic systems, and the baggeed trees-based EWMA scheme outperforms the Boosted trees-based scheme in detecting faults in the pyranometer.**

***Index Terms*—Anomaly detection, Photovoltaic system, sensor faults, machine learning, monitoring charts.**

I. Introduction

Photovoltaic (PV) systems are an essential renewable energy source that converts sunlight into electricity. These systems have gained significant popularity worldwide due to their environmental benefits and potential cost savings [1]. The PV systems have several components, including solar panels, inverters, batteries, and sensors. Solar panels are the main component that captures sunlight and converts it into direct current (DC) electricity. The inverters convert the DC electricity into alternating current (AC) electricity compatible with the grid or appliances. Batteries store excess energy generated by the system during peak hours for use during low-demand periods or emergencies. On the other hand, sensors provide crucial information about the system's performance, including the amount of energy produced, system efficiency, and environmental conditions [2].

The performance of PV systems depends on various factors, including the quality of the components, the system's design, and environmental conditions. Faults in PV systems can occur due to various factors, such as aging, manufacturing defects, installation errors, and environmental factors, such as shading or dust. Sensor readings play a crucial role in optimizing the operation of photovoltaic systems [3]. However, sensor faults can occur due to various factors, such as aging, installation errors, and environmental factors. These faults can lead to performance degradation, reduced efficiency, and even system failures. Therefore, continuously monitoring and diagnosing the system's behavior is essential to detect and prevent faults [4].

Sensor faults can significantly impact the performance and reliability of photovoltaic (PV) systems. PV systems rely on sensors to measure various parameters, such as solar irradiance, temperature, current, and voltage. These measurements are used to control and optimize the operation of the PV system, such as adjusting the angle of the solar panels or regulating the output power. If a sensor is faulty, it can provide inaccurate or unreliable measurements, leading to suboptimal or even dangerous operation of the PV system. For example, a faulty temperature sensor can cause the system to overheat or underperform, while a faulty pyranometer can cause the system to overestimate or underestimate the available solar irradiance. Faults in current and voltage sensors can lead to incorrect power output, affecting the system's efficiency and safety. In addition, sensor faults can also lead to increased maintenance and repair costs for PV systems. If a fault goes undetected, it can cause further damage to the system or even lead to system failure. Therefore, detecting sensor faults in a timely and accurate manner is essential for ensuring PV systems' performance, reliability, and safety.

Several techniques have been introduced for identifying faults in photovoltaic systems, including rule-based, model-based, and data-driven methods [5]. The rule-based approaches rely on predefined rules to detect system malfunctions based on anticipated sensor behavior [6]. Although easy to implement, these approaches have limitations, including a lack of flexibility and an inability to identify complex faults. Model-based methods rely on mathematical models to represent the behavior of photovoltaic systems and their sensors, enabling the prediction of expected behavior and the comparison with actual sensor data. However, obtaining precise models for large-scale PV plants can be challenging. As a result, data-driven monitoring methods are typically employed in such situations. Data-driven monitoring relies on historical data from a fault-free process to create an empirical model for detecting faults in future data. Data-driven methods have gained popularity recently due to their ability to handle large and complex data sets [7]–[9]. Data-driven methods for fault detection in photovoltaic systems can be categorized into statistical methods and machine learning methods. Statistical methods use statistical techniques to analyze the data and detect any deviations from the expected behavior. These methods include control charts, regression analysis, and time-series analysis. On the other hand, machine learning methods use algorithms to learn patterns and relationships in the data and use them to detect faults. These methods include neural networks, decision trees, support vector machines, and random forests. Machine learning methods are often combined with statistical methods to improve their performance.

Several data-driven methods have been developed in the last twenty years to enhance the detection of faults in PV systems [10], [11]. For instance, in [12], Taghezouit et al. proposed a monitoring methodology for detecting anomalies in photovoltaic systems using principal component analysis (PCA) and multivariate monitoring schemes. An assumption-free PCA-based detection method is introduced using Kernel Density Estimation (KDE) to set nonparametric thresholds for decision statistics. The proposed method is applied to real measurements from a 9.54 kWp grid-connected PV system, and six case studies are investigated to verify its detection efficiency. Results indicate that the proposed method with nonparametric thresholds achieved promising detection performance and can be used as an automatic tool to detect anomalies in PV systems' DC and AC sides. In [13], Maleki et al. introduced a statistical fault detection algorithm to analyze the waveshape of superimposed PV array power using the kurtosis function. The algorithm can detect light faults, discriminate them from severe partial shading, and work for open-circuit faults. The algorithm's effectiveness is demonstrated through case studies on a test PV array simulation model with parameter uncertainty and signal noise. The study in [14] presented a statistical procedure to analyze the operation of a PV plant without using environmental data as input. The method compares the statistical distributions of the energy dataset of different arrays and can detect and locate abnormal operating conditions before they become failures. In [15], a fault detection method for solar energy production systems using the least squares method has been proposed. An analytical model based on the Bishop model is presented and implemented in MATLAB to simulate healthy and faulty PV cases. The simulation results show that the least squares method can improve fault detection and is easy to implement. Also, a fuzzy logic approach is used for decision-making and provides encouraging overall results for this approach. The study in [16] proposes a decision tree algorithm (C4.5) for fault classification and diagnosis in a PV plant. A non-parametric model is used to predict the state of the GCPVS by learning from a dataset collected under different weather conditions. Results show high prediction performance for detection and diagnosis, with an accuracy of 99.80% in discriminating between string fault, short circuit fault, line-line fault, and fault-free data. In [17], Chen et al. propose a method for monitoring photovoltaic (PV) systems that utilizes a vector autoregressive (AR) model to model the post-change signal and the generalized local likelihood ratio test to detect faults. Multiple meters are used to measure various output signals of the PV system, and the method exploits the time correlation of the faulty signal and signal correlation among different meters. The method is evaluated through extensive simulations and is shown to achieve fast detection and satisfactory performance for various types of faults in PV systems.

Although fault detection in photovoltaic systems has been studied in the literature, sensor faults detection is a specific subfield that has not received as much attention. Sensor faults can lead to misinterpretation of the data, causing false alarms or missed detections of actual faults. Therefore, it is essential to develop reliable methods to detect sensor faults in photovoltaic systems. The contribution of this paper is the use of ensemble learning methods, specifically boosted trees and bagged trees, combined with the EWMA chart for sensor fault detection in photovoltaic systems. Ensemble learning methods are efficient in modeling input-output data, and by combining several weak learners, they can reduce prediction errors [18]. Ensemble models are employed to predict solar power production based on environmental and electrical inputs. The EWMA chart is sensitive to small changes and is applied to the residuals generated from the ensemble models to detect sensor faults. In addition, the detection threshold of the EWMA chart is computed in a nonparametric way using kernel density estimation to extend its flexibility. The reference model is obtained by training the algorithm using only fault-free data in the proposed sensor fault detection approach. This means that the algorithm learns the system's normal behavior by observing its output under normal conditions. By doing so, the algorithm can establish a threshold for detection based on the system's expected behavior. This approach has the advantage of not requiring any data labeling, which can be time-consuming and costly. Moreover, using only fault-free data for training ensures that the reference model is based on the actual behavior of the system under normal conditions, which increases its accuracy in detecting faults. Therefore, this approach offers a simple and efficient solution for sensor fault detection that can be applied in a wide range of systems without requiring extensive labeling or prior knowledge of the system's faults. The proposed approach was evaluated by simulating different sensor faults pyranometers to assess its performance using four statistical metrics. The results showed that the proposed approach achieved high detection performance for all simulated sensor faults. This approach can help improve the reliability and performance of photovoltaic systems by enabling the timely detection of sensor faults, leading to prompt maintenance and repair actions.

The paper is organized as follows: Section II presents an overview of Bagged trees, EWMA, and the proposed ensemble learning-based EWMA anomaly detection approach. In Section IV, the performance of the proposed method is evaluated using simulated PV data with various sensor faults. Finally, Section V concludes the paper with a summary of the findings and suggestions for future research directions.

II. Preliminary materials and methods

A. Bagged regression trees (BT)

Bagged regression trees (BTs) are a machine-learning approach that can be used for solar power prediction. The method involves training multiple decision trees on different subsets of the training data, which helps to reduce overfitting and improve the robustness of the model [19]. The final prediction is then obtained by averaging the predictions of all the trees in the ensemble. The bagging algorithm involves constructing k bootstrap datasets $D_1, D_2, \ldots, D_k$ by sampling from the original dataset D with replacement, where each sample has an equal probability of being selected. A regression tree is then fit to each bootstrap sample D_i to obtain a set of models $f_1(x), f_2(x), \ldots, f_k(x)$. The final prediction is then given by the average of the predictions of each model:

$$\hat{y}(x) = \frac{1}{k}\sum_{i=1}^{k} f_i(x). \tag{1}$$

For solar power prediction, the input x could be weather data such as temperature, humidity, and wind speed, and the output $\hat{y}$ would be the predicted solar power output. Bagging helps to reduce overfitting and improve the stability and accuracy of the predictions [20]. By randomly selecting subsets of the data, the bagged trees are exposed to different patterns in the data, which helps to reduce overfitting [21]. In this work, we will utilize Bayesian optimization for the purpose of optimizing the hyperparameters of BT.

Overall, the bagging technique enhances the model's ability to handle noise and outliers by sub-sampling the data and reducing their impact on the final predictions. Additionally, bagging is highly flexible as it can be applied to a wide range of base models including decision trees, neural networks, and support vector machines (SVMs) to capture complex relationships within the data. Its implementation is also relatively simple, and it can be scaled to handle large datasets by parallelizing the computation. Overall, bagging is a useful technique for improving the accuracy and robustness of machine learning models, particularly in scenarios where the data is noisy or contains outliers.

B. EWMA monitoring chart

The EWMA monitoring chart is a widely used statistical process control technique that effectively identifies shifts and anomalies in time series data. It involves computing a weighted average of previous observations, with the weights decreasing exponentially as the observations become older. The EWMA statistic at time t is calculated using a formula given by [22]:

$$S_t = \begin{cases} X_1, & t = 1 \\ \lambda X_t + (1-\lambda)S_{t-1}, & t > 1, \end{cases} \tag{2}$$

where x_t is the observation at time t, λ is the smoothing parameter, and S_t is the EWMA statistic at time t. The proposed approach utilizes the kernel density estimation (KDE) method to set the threshold of KD-based anomaly detectors by utilizing data without anomalies, which increases flexibility. The probability density function (PDF) of the charting statistics is estimated using the KDE method, and the detection threshold is determined based on the estimated distribution of charting statistics, with the $(1-\alpha)$-th quantile being used as the threshold.

III. Proposed BT-based EWMA chart for sensor fault detection in PV systems

The proposed BT-EWMA anomaly detection method is a statistical process control approach that combines the benefits of BT and the EWMA control chart. The BT algorithm is used to model the underlying relationships between the input variables and the output variable, while the EWMA control chart is used to monitor the residuals of the BT model. This approach involves several steps, including data preprocessing, training a BT model using 5-fold cross-validation, computing the EWMA thresholds, and online monitoring of the test data. During the training stage, the BT model is trained on anomaly-free data using Bayesian optimization for hyperparameter tuning. We evaluate the model using three commonly used metrics, namely R^2 (coefficient of determination), RMSE (root mean squared error), and MAPE (mean absolute percentage error). These metrics indicate the accuracy and performance of the model in predicting the expected values of the solar power output. Moreover, to avoid overfitting and ensure the model's robustness, we perform a 5-fold cross-validation during the training process. This approach allows us to evaluate the model's performance on multiple subsets of the training data, thereby reducing the risk of overfitting to a particular training data set. The trained BT model is then used to predict the expected values of the test data. The residuals between the predicted and actual values are used to compute the EWMA control statistic. Using the KDE approach, the detection threshold for fault detection is computed based on the estimated distribution of the charting statistics. Finally, the EWMA control statistic is compared to the previously computed threshold, and if the control statistic exceeds the threshold, the corresponding sample is labeled as an anomaly. This BT-EWMA approach effectively detects faults in PV systems, reducing the need for manual inspection and maintenance. Algorithm 1 presents an overview of the fundamental procedures involved in the BT-EWMA approach.

Algorithm 1 BT-EWMA Anomaly Detection

Input: Training data D_{train} and test data D_{test}
Output: Anomaly labels for D_{test}
Step 1: Data Preprocessing
- Discard outliers and impute missing values in D_{train}
- Impute missing values in D_{test} using Amelia package

Step 2: Train BT model
- Train BT model on D_{train} with Bayesian optimization for hyperparameter tuning

Step 3: Compute EWMA Thresholds
- Compute the EWMA statistic based on the residuals from the BT model
- Compute the detection threshold for fault detection using KDE approach

Step 4: Online Monitoring
- For each sample $x_t \in D_{test}$, predict its expected value using the trained BT model
- Calculate the residual error between the predicted value and the actual value
- Compute the EWMA control statistic for the residual error

Step 5: Anomaly Detection
- Compare the EWMA control statistic with the previously computed threshold
- If the control statistic exceeds the threshold, label x_t as an anomaly

IV. Results and Discussions

The section discusses assessing a proposed fault detection approach based on simulated data from a PV system. Several tools are available for evaluating and estimating PV power production, including PVGIS, PVWatts, and RETScreen [23]. The study utilized PVGIS to simulate a PV system with an optimal slope of 23 degrees and an optimal azimuth of 7 degrees (Figure 1). The system had a nominal power output of 9.5 kWp and was based on crystalline silicon technology with a system loss of 14%. Hourly data were collected for one year, including PV power output (P), global in-plane irradiance (G(i)) in watts per square meter, sun height (Hsun) in degrees, air temperature (T2m) in degrees Celsius, and wind speed at 10 meters (WS10m) in meters per second.

Figure 2 presents the pairwise correlation coefficients between the active power (P) and the four input variables. We observe a strong linear relationship between the PV power output, global in-plane irradiance, and sun height. This is because the amount of solar irradiance directly affects the power output, and the sun height changes during the day, affecting the angle at which the solar irradiance strikes the panels. We also observe a moderate correlation between the power output and air temperature, while wind speed at 10m has a relatively weak effect on PV power output.

The RReliefF algorithm [24] was used to determine the most significant features for solar power prediction, as shown

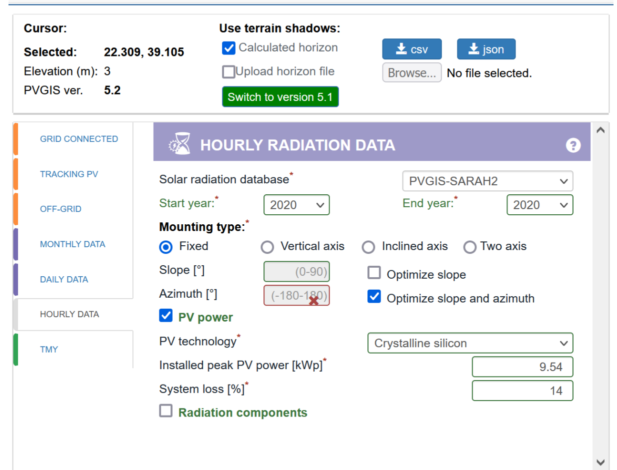

Fig. 1. PV system simulation via the PVGIS web-interface.

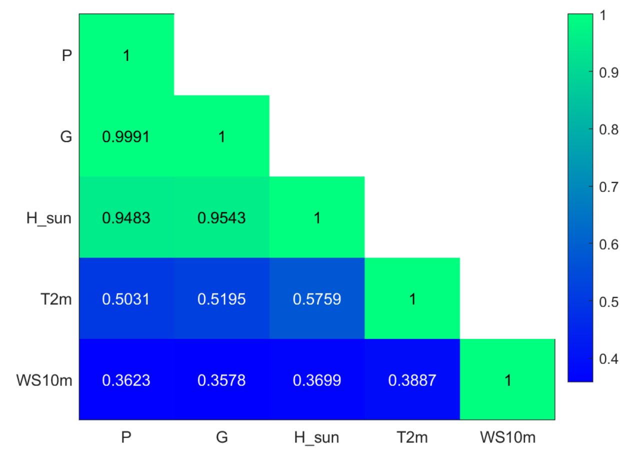

Fig. 2. Correlation matrix of input-output variables in a simulated PV system.

in Figure 3. The results indicate that the global in-plane irradiance is the most important input variable, followed by the sun's height. The impact of air temperature and wind speed at 10 meters on solar power output is relatively lower. These feature importance scores can be helpful in optimizing the performance of a solar energy system.

To evaluate the performance of two methods, this study utilized one year of hourly data. The data was split into 80% for training and the remaining for testing. To prevent overfitting, a fivefold cross-validation procedure was employed during the training. The BT and BST models were used with default parameters of 30 learners, a minimum leaf size of 8, and a learning rate of 0.1. The actual and predicted DC power from the trained models are presented in Figure 4, indicating that the ensemble models can accurately capture the trend in the DC power data. Visual analysis shows that the BT model provides better prediction performance than the BST model.

The MAE and RMSE values presented in Table I indicate that the BT model has lower errors than the BST model, indicating its potential as a better predictor of PV power output. Moreover, the high R-squared value of 0.999 indicates

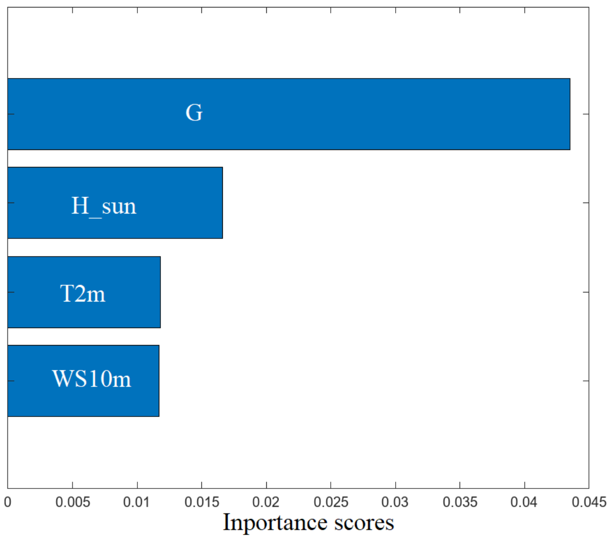

Fig. 3. Feature importance scores sorted using RReliefF algorithm.

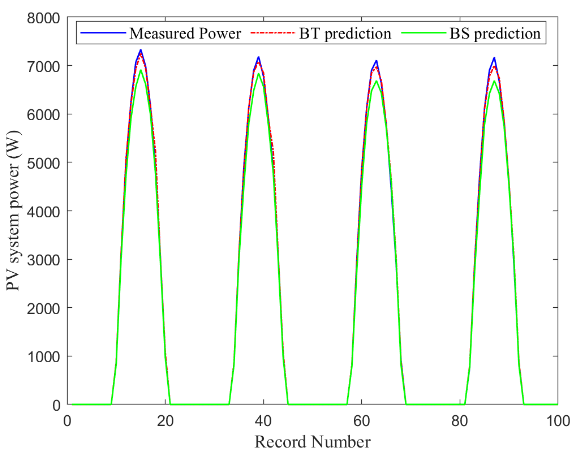

Fig. 4. Power prediction results of the two machine learning models based on test data.

that the BT model explains most of the variability in the data.

TABLE I
COMPARISON OF PV POWER PREDICTION RESULTS.

Model	MAE	RMSE	R^2
Bagged Trees	25.125	61.349	0.999
Boosted Trees	90.047	160.951	0.996

The trained models will be combined with the EWMA chart to detect sensor faults in simulated data from a PV system. One case study will be considered: faults in the pyranometer. To evaluate the performance of the proposed approach for anomaly detection, we utilized four standard metrics: Accuracy, Precision, Area Under the Curve (AUC), and F1-score [25]. These metrics are commonly used in machine learning and data mining to assess the effectiveness of classification and detection methods. Accuracy is the ratio of correctly predicted instances to the total number of instances, while precision measures the ratio of true positives to the total number of positive instances. AUC is a measure of the detector's ability to distinguish between positive and negative instances, and F1-score is a harmonic mean of precision and recall. These metrics provide a comprehensive evaluation of the proposed approach's effectiveness in detecting anomalies.

1) Case Study 1: Sensor Faults in Pyranometer: A pyranometer is a measurement device to determine the amount of solar irradiance received on a planar surface. It comprises a thermopile sensor that generates a voltage proportional to the quantity of incident solar radiation. Pyranometer sensor faults can occur due to various factors such as aging, environmental factors, and mechanical damage. These faults can lead to inaccurate measurements and adversely affect the performance of solar power systems. Thus, detecting and diagnosing pyranometer sensor faults is essential for ensuring reliable and efficient solar power system operation. In this case study, two examples of sensor faults in pyranometers were considered, and a bias fault was introduced between sampling time instants 55 and 65. In the first example, the bias had a constant amplitude of 70% of the total variation in solar irradiance measurements, making it relatively easy to detect using the BT and BST-based EWMA schemes, as shown in Figure 5.

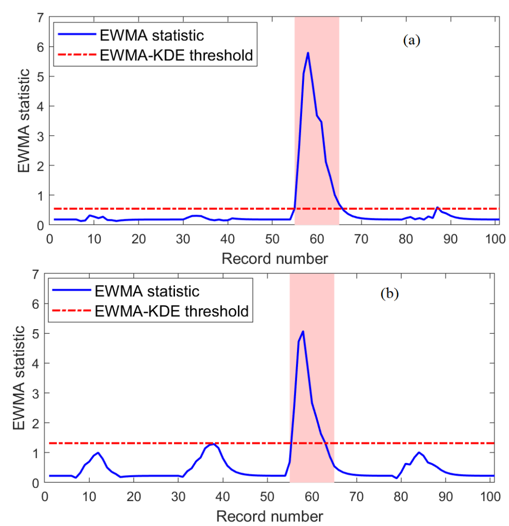

Fig. 5. Detection results of (a) BT-EWMA, and (b) BS-EWMA schemes in the presence of Sensor fault in pyranometer.

Table II presents the detection performance of the BT and BS-based EWMA schemes for a constant bias sensor fault in a pyranometer. The results indicate that both methods can detect the fault with high accuracy and precision. However, the BT-EWMA scheme outperforms the BS-EWMA in terms of AUC and F1-score when the bias is at 50%. Both methods achieve perfect detection accuracy and precision when the bias is increased to 80%. These results demonstrate the effectiveness of the proposed approach in detecting constant bias sensor faults in pyranometers. It is worth noting that the performance of the BS-based scheme is relatively lower than the BT-based scheme. This may be attributed to the fact that the BT model is better at capturing the underlying patterns and trends in the data, which are essential for detecting faults accurately.

TABLE II
DETECTION RESULTS OF THE BT AND BS-BASED EWMA SCHEMES FOR THE PRESENCE OF A CONSTANT BIAS SENSOR FAULT IN A PYRANOMETER.

Method	Accuracy	Precision,	AUC	F1-score
		Bias=50%		
BT-EWMA	0.99	0.92	0.99	0.96
BST-EWMA	0.96	1	0.82	0.78
		Bias=80%		
BT-EWMA	1	1	1	1
BS-EWMA	0.96	0.89	0.86	0.8

V. CONCLUSION

In this study, we proposed a method for detecting sensor faults in photovoltaic systems using ensemble learning models and the EWMA chart. We evaluated the performance of two ensemble models, the BT and BS, using one year of hourly data. The results showed that the BT model performed better than the BST model in predicting PV power output, as indicated by lower MAE and RMSE values and a higher R^2 value. We then combined the trained models with the EWMA chart to detect sensor faults in simulated data from a PV system. Using the KDE approach to determine the detection threshold also increases the flexibility of the method, making it suitable for different applications. One case study was considered: sensor faults in the pyranometer. The results showed that the proposed method effectively detected the introduced sensor faults. The BT-based EWMA scheme outperformed the BS-based scheme in detecting faults in the pyranometer. Overall, the proposed method can be a useful tool for ensuring the reliable and efficient operation of photovoltaic systems by detecting sensor faults in a timely manner.

Future work includes applying the proposed method to real-world photovoltaic systems and comparing its performance with other methods for sensor fault detection. Additionally, exploring the applicability of the proposed method to other types of sensors, such as temperature and humidity sensors, can be beneficial. Future research could also focus on exploring the application of deep learning techniques, such as transformers, for detecting sensor faults in PV plants. Additionally, integrating the proposed method with advanced control and optimization algorithms could improve the performance and efficiency of PV plants.

REFERENCES

[1] C. Lupangu and R. Bansal, "A review of technical issues on the development of solar photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 950–965, 2017.
[2] R. Rajesh and M. C. Mabel, "A comprehensive review of photovoltaic systems," *Renewable and sustainable energy reviews*, vol. 51, pp. 231–248, 2015.
[3] D. S. Pillai and N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in pv systems," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 18–40, 2018.
[4] F. Harrou, B. Taghezouit, S. Khadraoui, A. Dairi, Y. Sun, and A. Hadj Arab, "Ensemble learning techniques-based monitoring charts for fault detection in photovoltaic systems," *Energies*, vol. 15, no. 18, p. 6716, 2022.
[5] Y.-Y. Hong and R. A. Pula, "Methods of photovoltaic fault detection and classification: A review," *Energy Reports*, vol. 8, pp. 5898–5929, 2022.
[6] F. Arévalo, C. Tito, M. R. Diprasetya, and A. Schwung, "Fault detection assessment using an extended fmea and a rule-based expert system," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1, pp. 740–745, IEEE, 2019.
[7] S. R. Madeti and S. Singh, "Modeling of pv system based on experimental data for fault detection using knn method," *Solar Energy*, vol. 173, pp. 139–151, 2018.
[8] J. Sun, F. Sun, J. Fan, and Y. Liang, "Fault diagnosis model of photovoltaic array based on least squares support vector machine in bayesian framework," *Applied Sciences*, vol. 7, no. 11, p. 1199, 2017.
[9] F. Harrou, B. Taghezouit, and Y. Sun, "Improved k nn-based monitoring schemes for detecting faults in pv systems," *IEEE Journal of Photovoltaics*, vol. 9, no. 3, pp. 811–821, 2019.
[10] F. Harrou, A. Saidi, Y. Sun, and S. Khadraoui, "Monitoring of photovoltaic systems using improved kernel-based learning schemes," *IEEE Journal of Photovoltaics*, vol. 11, no. 3, pp. 806–818, 2021.
[11] F. Harrou, A. Dairi, B. Taghezouit, and Y. Sun, "An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class support vector machine," *Solar Energy*, vol. 179, pp. 48–58, 2019.
[12] B. Taghezouit, F. Harrou, Y. Sun, A. H. Arab, and C. Larbes, "Multivariate statistical monitoring of photovoltaic plant operation," *Energy Conversion and Management*, vol. 205, p. 112317, 2020.
[13] A. Maleki, I. Sadeghkhani, and B. Fani, "Statistical sensorless short-circuit fault detection algorithm for photovoltaic arrays," *Journal of Renewable and Sustainable Energy*, vol. 11, no. 5, p. 053501, 2019.
[14] S. Vergura, "A statistical tool to detect and locate abnormal operating conditions in photovoltaic systems," *Sustainability*, vol. 10, no. 3, p. 608, 2018.
[15] T. Andrianajaina, E. J. R. Sambatra, C. B. Andrianirina, T. D. Razafimahefa, and N. Heraud, "Pv fault detection using the least squares method," in *2016 International Conference and Exposition on Electrical and Power Engineering (EPE)*, pp. 846–851, IEEE, 2016.
[16] R. Benkercha and S. Moulahoum, "Fault detection and diagnosis based on c4. 5 decision tree algorithm for grid connected pv system," *Solar Energy*, vol. 173, pp. 610–634, 2018.
[17] L. Chen, S. Li, and X. Wang, "Quickest fault detection in photovoltaic systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1835–1847, 2016.
[18] J. Lee, W. Wang, F. Harrou, and Y. Sun, "Reliable solar irradiance prediction using ensemble learning-based models: A comparative study," *Energy Conversion and Management*, vol. 208, p. 112582, 2020.
[19] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, pp. 123–140, 1996.
[20] F. Harrou, A. Saidi, and Y. Sun, "Wind power prediction using bootstrap aggregating trees approach to enabling sustainable wind power integration in a smart grid," *Energy Conversion and Management*, vol. 201, p. 112077, 2019.
[21] C. D. Sutton, "Classification and regression trees, bagging, and boosting," *Handbook of statistics*, vol. 24, pp. 303–329, 2005.
[22] J. S. Hunter, "The exponentially weighted moving average," *Journal of quality technology*, vol. 18, no. 4, pp. 203–210, 1986.
[23] C. S. Psomopoulos, G. C. Ioannidis, S. D. Kaminaris, K. D. Mardikis, and N. G. Katsikas, "A comparative evaluation of photovoltaic electricity production assessment software (pvgis, pvwatts and retscreen)," *Environmental Processes*, vol. 2, no. 1, pp. 175–189, 2015.
[24] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relieff and rrelieff," *Machine learning*, vol. 53, pp. 23–69, 2003.
[25] W. Wang, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "Cyber-attacks detection in industrial systems using artificial intelligence-driven methods," *International Journal of Critical Infrastructure Protection*, vol. 38, p. 100542, 2022.