ELKHA : Jurnal Teknik Elektro, Vol. 18 No.1, April 2026, pp. 56 - 64
ISSN: 1858-1463 (print), 2580-6807 (online)

XAI Implementation for Inverter-Level Underperformance Analysis of the PV Power Plant from Actual Operational Data

Reza Adisetia Saputra[1], Dwi Joko Suroso[2*)], and Mohammad Kholid Ridwan[3)]

[1, 2,3)] Department of Nuclear Engineering and Engineering Physics, Universitas Gadjah Mada, Indonesia
Corresponding Email: *) dwi.jokosuroso@ugm.ac.id

***Abstract*- Photovoltaic (PV) power plants installed on uneven terrain often experience spatially non-uniform operating conditions that lead to performance disparities among inverters, which may not be detected through conventional system-level monitoring. This study presents an inverter-level performance analysis of the Nusa Penida PV power plant using one year of operational data at 30-minute resolution. A data-driven framework integrating Extreme Gradient Boosting (XGBoost) classification and explainable artificial intelligence was developed to detect underperforming inverters and interpret the factors affecting system performance. The analysis identified significant performance variations among the 18 inverters, with seven units categorized as underperforming based on a relative performance ratio threshold of 0.80 compared to the highest-performing inverter at the same timestamp. The proposed XGBoost classification model achieved an AUC of approximately 0.79 on the test dataset, indicating reliable discrimination between normal and underperforming inverter conditions. Further analysis shows that the detected underperformance corresponds to annual energy losses ranging from approximately 81,000-119,000 kWh per inverter when compared with the best-performing reference unit. Explainable analysis using SHapley Additive exPlanations (SHAP) reveals that irradiance and temporal variables are the dominant contributors affecting inverter performance. In contrast, persistent negative feature contributions across several inverters indicate location-related constraints beyond natural environmental variability. These results demonstrate that inverter-level monitoring combined with interpretable machine learning provides deeper diagnostic insight than aggregated performance indicators and can support more effective identification of structural performance limitations in PV power plants installed on non-uniform terrain.**

***Keywords— photovoltaic power plant, inverter-level analysis, energy loss quantification, Extreme Gradient Boosting, explainable artificial intelligence*.**

I. INTRODUCTION

The increasing deployment of PV power plants has positioned solar energy as an important contributor to national energy portfolios, particularly in tropical countries with high solar irradiance such as Indonesia [3]. Although climatic conditions are generally favorable, the actual operational performance of PV systems often deviates from design expectations.

Environmental variability, gradual component degradation, and operational constraints can introduce energy losses that are not always apparent during routine evaluation [1], [9], [10], [20]. Over time, these effects may accumulate and lead to persistent differences between predicted and realized energy production.

PV system performance is commonly assessed using aggregated indicators such as performance ratio, capacity factor, and annual energy yield [2]. While these metrics provide a useful overview of plant behavior, they are often insufficient to capture localized performance differences within the system. In large multi-inverter PV plants, degradation and efficiency decline rarely occur uniformly across all components. As a result, inverter-level underperformance may remain concealed within system-level averages [4], [17]. This limitation becomes more significant in plants installed on uneven terrain, where irradiance exposure and shading conditions vary spatially across inverter zones [19], [30].

For this reason, inverter-level monitoring has received increasing attention in recent years. Localized issues such as power-electronic aging, string mismatch, partial shading, and thermal stress may persist without triggering alarms, yet still lead to measurable energy losses over extended operating periods [10], [11], [16]. Detecting such conditions requires analytical approaches that go beyond aggregated performance indicators and consider subsystem-level variability. Surface characteristics, including terrain-induced irradiance variation and albedo differences, further complicate this assessment in geographically non-uniform installations [19], [29].

Machine learning (ML) techniques have been widely adopted in PV forecasting and performance modeling due to their ability to capture nonlinear relationships between environmental inputs and electrical output [5], [12], [15]. Among these techniques, XGBoost has demonstrated strong predictive capability and robustness when applied to large operational datasets [6], [24]. In several renewable energy studies, tree-based ensemble models have achieved competitive or even superior performance compared with more complex deep learning architectures while maintaining stability in tabular data environments typical of utility-scale PV plants [22].

Despite their predictive strength, many ML models remain difficult to interpret. In engineering applications,

Manuscript received 2026-02-10; revised 2026-03-16; accepted 2026-03-27

predictive accuracy alone is insufficient; system operators must also understand the factors influencing model predictions to support maintenance planning and operational decision-making. Explainable artificial intelligence (XAI) methods address this challenge by quantifying each input variable's contribution to model predictions [7], [13], [14], [18]. Recent studies have applied SHAP-based explanations to PV power forecasting [8], [21], solar radiation modeling [24], temperature prediction [25], and long-term energy yield analysis considering degradation mechanisms [27]. However, most of these studies focus primarily on system-level prediction accuracy or short-term datasets.

Consequently, there remains limited research that combines long-term inverter-level operational data with interpretable machine learning to investigate persistent structural performance differences within a single PV power plant. This limitation is particularly relevant for PV installations located on uneven terrain, where spatial variability in irradiance and thermal conditions can significantly influence inverter operating behavior [19], [29], [30].

The Nusa Penida PV power plant provides a relevant case study due to its installation on non-uniform terrain. Spatial differences in irradiance exposure, shading patterns, and operating temperature are expected across inverter zones. Simulation-based assessments indicate that the plant operates at approximately 81% of its design reference under normal conditions [2], suggesting persistent energy losses that warrant further investigation. These characteristics make the site suitable for evaluating inverter-level performance disparities using a data-driven and interpretable modeling approach.

This study addresses the identified gap by analyzing one year of operational data from the Nusa Penida PV power plant. The contributions of this research are threefold. First, underperforming inverters are identified using long-term operational data rather than short-term anomaly detection. Second, annual energy losses are quantified at the inverter level to assess the magnitude of structural performance disparities within the plant. Third, an interpretable machine learning framework combining XGBoost and SHAP is developed to support diagnostic analysis and condition-based maintenance for PV systems installed on uneven terrain.

II. METHODOLOGY

A. Overview of the Methodological Framework

This study adopts an inverter-level operational data approach to investigate performance disparities within the Nusa Penida PV power plant. Unlike conventional system-level evaluations that rely on aggregated indicators such as performance ratio or annual energy yield, the proposed framework focuses on identifying localized inverter underperformance using operational measurements and machine learning techniques.

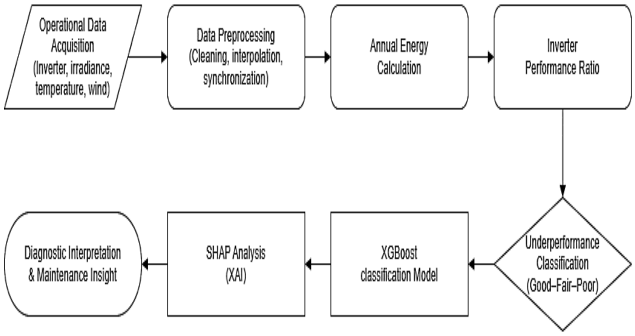

Figure 1. Methodological framework for inverter-level underperformance analysis using ML and XAI

The methodological workflow, illustrated in Fig. 1, consists of four main stages. First, inverter-level operational data are collected and preprocessed to ensure consistency and reliability. Second, inverter performance ratios are calculated to detect deviations in power output relative to the best-performing inverter within the plant. Third, a machine learning classification model based on XGBoost is trained to identify underperforming inverter conditions using environmental and temporal features. Finally, SHAP are applied to interpret model predictions and identify the dominant factors influencing inverter performance.

This approach enables the integration of operational data analysis with interpretable machine learning to support diagnostic evaluation and performance monitoring in PV power plants installed on complex terrain.

B. Description of the Nusa Penida PV Power Plant

The case study examined in this research is the Nusa Penida PV power plant, a grid-connected hybrid solar generation system integrated with battery energy storage and conventional generation units. The PV installation has a total installed capacity of 4.2 MWp. It is supported by a battery energy storage system rated at 3 MW / 1.84 MWh, which helps stabilise the Nusa Penida island power network.

The PV system consists of approximately 7,780 PV modules rated at 540 Wp and 18 grid-connected inverters (Huawei SUN2000-215KTL) with a combined AC capacity of approximately 3.5 MW. The electrical configuration includes PV modules, DC combiner boxes, inverters, low-voltage collectors, step-up transformers, medium-voltage switchgear, and connection to the 20 kV distribution network.

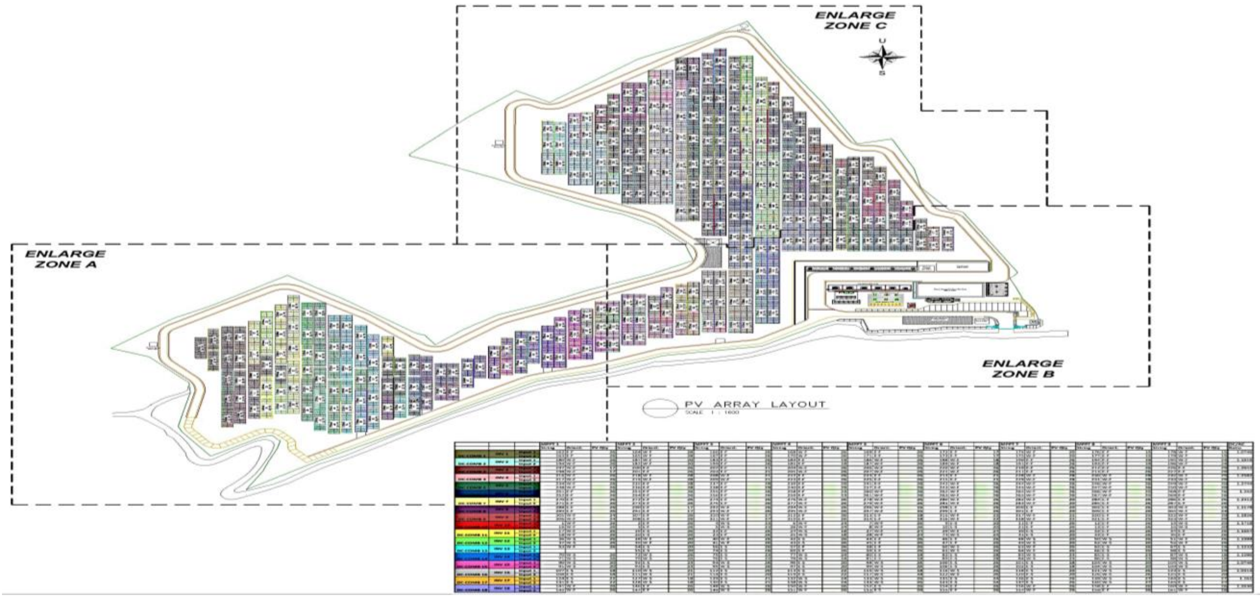

Figure 2. Site layout map of the Nusa Penida PV power plant area.

As shown in Fig. 2, the PV plant is installed on uneven terrain, resulting in variations in module orientation, tilt angle, and environmental exposure across inverter zones. These site conditions create spatially non-uniform irradiance distribution, localized shading effects, and thermal variability that directly influence inverter operating conditions and energy production. Each inverter is connected to multiple PV strings and is equipped with monitoring sensors that record operational parameters at regular intervals, including power output and environmental conditions. These inverter-level measurements serve as the primary dataset used in this study.

C. Data Operational and Pre-Processing

The dataset consists of inverter-level operational measurements collected throughout 2024 with a 30-minute temporal resolution. The recorded variables include inverter output power (kW), solar irradiance (W/m²), ambient temperature (°C), wind speed (m/s), and temporal features such as hour of the day and month of the year. To ensure that the dataset represents actual PV generation conditions, the analysis focuses on effective operating periods between 08:00 and 17:00 local time, which correspond to typical solar production hours. Measurements outside this interval are excluded from the analysis.

Data preprocessing is conducted to improve the quality and consistency of the dataset before modelling. Missing values with small proportions are handled using linear interpolation, while intervals containing significant data gaps are removed to avoid bias in model training. Temporal alignment between the inverter output and environmental variables is verified with a synchronization tolerance of ±5 minutes. Outlier values are retained in the dataset because they may represent meaningful indicators of operational disturbances or inverter underperformance rather than measurement errors.

D. Energy Calculation and Inverter Performance Ratio

Inverter performance analysis is conducted by calculating the annual energy output of each inverter by integrating the output power over all measurement intervals [6], [24]. The annual energy of the i-th inverter is computed as:

$$E_i = \sum_{t=1}^{T} P_{i,t} \, . \, \Delta t \; \dots\dots\dots\dots\dots\dots \quad (1)$$

where E_i represents the annual energy of the i-th inverter (kWh), $P_{i,t}$ denotes the output power of the i-th inverter at time interval t(kW), Δt is the measurement time interval (hours), and N is the total number of time intervals within the observation period.

To evaluate performance differences among inverters, a relative performance ratio is calculated by comparing each inverter's output power to that of the highest-performing inverter at the same timestamp. The performance ratio is defined as

$$R_{i,t} = \frac{p_{i,t}}{p_{max,t}} \; \dots\dots\dots\dots \quad (2)$$

where $P_{i,t}$ represents the output power of inverter $i$at time interval t, and $P_{max,t}$ denotes the maximum inverter output observed among all inverters at the same timestamp. An inverter observation is labelled as underperforming when the performance ratio satisfies $R_{i,t} < 0.80$.

This threshold corresponds to a 20% deviation from the best-performing inverter. It is used to identify significant performance disparities while minimizing misclassification due to natural environmental variability. Observations with $R_{i,t} \geq 0.80$ are categorized as normal operating conditions.

The resulting binary labels are used as the target variable for the machine learning classification model.

E. Definition and Classification of Inverter Underperformance

Inverter underperformance is defined based on the deviation of an inverter's performance ratio relative to the best-performing inverter within the PV plant. The performance ratio reflects the relative output of each inverter compared to the maximum inverter output observed at the same time interval.

Based on the performance ratio, each inverter is classified into one of three categories. An inverter is categorized as good when $Ri \geq 0.90$, fair when $0.80 \leq Ri < 0.90$, and poor when $Ri < 0.80$. The threshold of 0.80 represents a deviation of approximately 20% from the best-performing inverter and is used to identify significant performance disparities within the plant. This classification framework enables the identification of inverters that experience persistent performance degradation relative to the reference unit.

F. XGBoost-Based ML Model

The XGBoost algorithm is employed as a classification model to detect inverter underperformance conditions based on operational variables. Tree-based ensemble methods such as XGBoost are well-suited to structured operational datasets and can capture complex nonlinear relationships between environmental variables and inverter behaviour.

The input features used in this study include irradiance, ambient temperature, wind speed, measurement time, and inverter identity. The target variable represents the inverter's operating condition, which is labelled as normal or underperforming based on the inverter performance ratio threshold.

The SHAP explanation model represents the prediction of the machine learning model as an additive contribution of input features, which can be expressed as:

$$f(x) = \phi 0 + \sum_{j=1}^{m} \phi j \quad \ldots\ldots\ldots \qquad (3)$$

where φ0 denotes the base value of the model and φj represents the contribution of feature j to the final prediction.

In the early stage of model development, a classification approach was evaluated to distinguish between underperforming and normal inverter conditions using operational data. The XGBoost classifier was trained with class imbalance adjustment through the scale_pos_weight parameter. The model achieved a validation AUC of approximately 0.81 and a test AUC of approximately 0.79, indicating satisfactory discriminative capability with relatively high recall.

Despite its strong predictive capability, XGBoost operates as a complex ensemble model whose internal decision structure is not inherently transparent. To improve model interpretability, an XAI approach is incorporated into the analysis. XAI aims to reveal how input variables influence model predictions, allowing machine learning results to be interpreted in relation to the physical behaviour of the PV system rather than treated solely as numerical outputs.

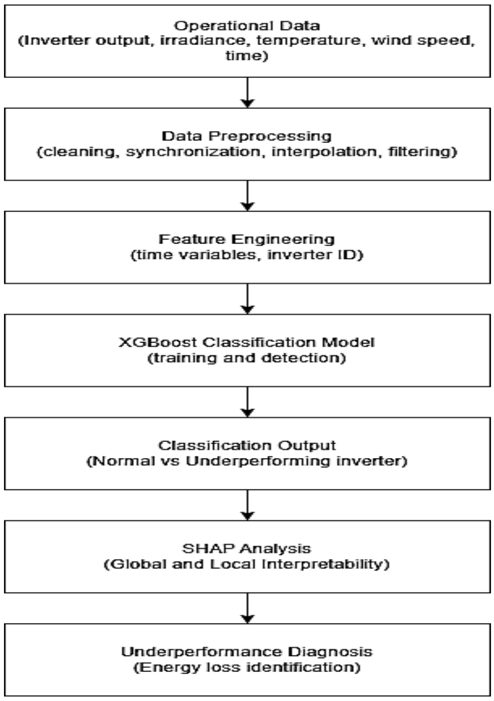

Figure 3. Workflow of the XGBoost-based machine learning model integrated with SHAP analysis for inverter-level underperformance diagnosis.

G. SHAP-Based Model Interpretation

Although ensemble machine learning models such as XGBoost provide strong predictive capability, their internal decision processes are often difficult to interpret. To address this limitation, XAI techniques are incorporated into the analysis.

In this study, SHAP are used to quantify the contribution of each input feature to the model predictions. SHAP values are computed using the Tree Explainer algorithm, which is specifically designed for tree-based models.

Two levels of interpretation are applied. Global SHAP analysis is used to identify the overall importance of input variables across the dataset, providing insight into the dominant factors affecting inverter performance. Local SHAP analysis is performed on selected inverter observations to examine feature contribution patterns associated with persistent underperformance.

Through this approach, the machine learning model functions not only as a predictive tool but also as a diagnostic instrument that reveals environmental and operational factors influencing inverter performance within the PV power plant. Although ensemble machine learning models such as XGBoost provide strong predictive performance, their internal decision mechanisms are often difficult to interpret. To address this limitation, XAI techniques are incorporated into the analysis.

III. RESULTS AND DISCUSSION

A. Annual Energy Distribution and Inverter Performance Variation

Based on inverter-level operational data collected throughout 2024, noticeable variations in annual energy production are observed among the inverters installed at the Nusa Penida PV power plant. The annual energy of each inverter is calculated by integrating the inverter's output power across all measurement intervals as defined in Equation (1).

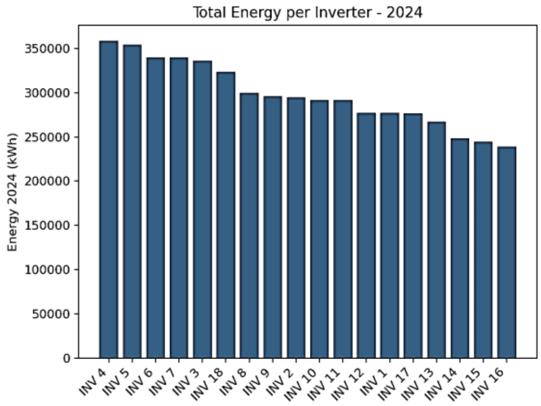

Figure 4. The annual energy distribution of the Nusa Penida solar power plant was inverted in 2024.

Figure 4 presents the distribution of annual energy generation across the eighteen inverters in the plant. The results show that inverter INV-4 produces the highest annual energy output of 358,519 kWh, which is used as the reference inverter representing 100% performance. In contrast, inverter INV-16 records the lowest annual production at 238,838 kWh, corresponding to approximately 66.7% of the reference inverter's output.

This substantial difference indicates a clear imbalance in inverter performance within the same PV installation. Because all inverters operate under the same plant infrastructure, these disparities suggest spatially dependent operational conditions rather than random fluctuations.

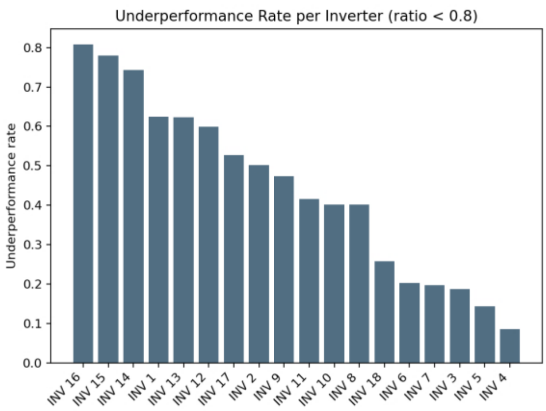

Figure 5. Inverter underperformance levels based on the energy output ratio relative to the reference inverter.

Figure 5 illustrates the relative inverter performance ratios compared with the reference unit. The clustering of several low-performing inverters suggests that performance deviations are associated with systematic factors, such as terrain-induced irradiance variability, localized shading, or differences in subsystem configurations across inverter zones.

B. Performance Ratio and Classification of Inverter Underperformance

To evaluate relative inverter performance, the inverter performance ratio defined in Equation (2) is used to compare the energy output of each inverter with that of the best-performing unit.

Based on the performance ratio distribution, the inverters are categorized into three groups.

Table 1. Classification criteria of inverter performance based on the performance ratio

Category	Performance Ratio
Good	$Ri \geq 0.90$
Fair	$0.80 \leq Ri < 0.90$
Poor	$Ri < 0.80$

Using this classification framework, six inverters fall within the good category, five inverters are classified as fair, and seven inverters are categorized as poor.

Table 2. Inverters with the highest underperformance based on 2024 operational data

Inverter	Energy (kWh)	Ratio to best	Energy loss (kWh)
INV-16	238,838	0.667	119,681
INV-15	244,854	0.683	113,666
INV-14	248,415	0.693	110,105
INV-13	267,375	0.746	91,144
INV-17	276,498	0.771	82,021
INV-1	277,433	0.774	81,086

Table 2 summarizes the inverters that exhibit the greatest underperformance. Among them, **INV-16**, **INV-15**, and **INV-14** represent the most significant deviations, with performance ratios of **0.667**, **0.683**, and **0.693**, respectively.

The clustering of these underperforming units suggests that inverter performance disparities are not randomly distributed but are influenced by localized site conditions. Such factors may include differences in terrain elevation, shading patterns, module mismatch, or environmental exposure affecting specific inverter zones.

C. Annual Energy Loss Analysis

The observed annual energy losses reach up to 119,681 kWh per inverter, corresponding to performance ratio reductions exceeding 20% relative to the reference inverter. For comparison, long-term degradation studies indicate that PV systems typically experience performance losses of approximately 0.5–1% per year due to module aging and environmental stress [1], [9]. The magnitude of the deviations observed in this study, therefore, significantly exceeds expected degradation-related losses.

Previous inverter-level monitoring studies report mismatch and localized shading losses typically ranging between 5–15%, depending on plant configuration and array layout [4], [17]. However, the performance ratio reductions below 0.80 observed in seven inverters in this study indicate more pronounced disparities than those commonly reported for PV installations on relatively uniform terrain.

Furthermore, under high irradiance conditions, the deviation between the reference inverter and underperforming units remains consistent. This behavior suggests that the performance reduction is structural rather than due to transient environmental variability. These findings highlight the importance of inverter-level performance evaluation, particularly in PV plants installed on uneven terrain where aggregated performance metrics may conceal localized energy losses.

D. Results of XGBoost-Based ML Modeling

The XGBoost classification model is applied to detect inverter underperformance patterns based on operational and environmental variables. Tree-based ensemble methods such as XGBoost are well-suited for structured operational datasets because they can capture nonlinear relationships among irradiance, temperature, temporal variables, and inverter behavior.

To evaluate the effectiveness of the proposed approach, the performance of the XGBoost classifier is compared with two commonly used baseline models in PV data analysis: logistic regression and decision tree classification.

All models are trained using identical input features, including solar irradiance, ambient temperature, wind speed, temporal variables (hour and month), and inverter identity. The dataset is divided using the same time-based training and testing split to ensure a fair comparison.

The annual inverter energy deviation is quantified as:

$$\Delta E_i = E_{ref} - E_i \; \dots\dots\dots\dots \quad (5)$$

where E_{ref} denotes the annual energy of the best-performing inverter and E_i represents the annual energy of the i-th inverter. The systematic deviations in prediction observed in selected inverters correspond closely with these quantified energy gaps, indicating residual behavior that cannot be fully explained by irradiance and temperature variables alone.

Table 3. Model performance evaluated

Model	AUC
Logistic Regression	0.71
Decision Tree	0.74
XGBoost (Proposed Method)	0.79

The results indicate that the XGBoost classifier provides improved discrimination capability compared with the baseline models. This improvement can be attributed to the ability of gradient boosting ensembles to capture nonlinear interactions among environmental variables and inverter operating conditions. Such interactions are particularly relevant in PV systems, where complex relationships between irradiance, temperature, and system configuration influence energy output.

E. Comparison with PVsyst Baseline Simulation

To assess the significance of the detected inverter-level disparities, the operational results are compared with the original PV system design simulation performed using PVsyst.The PVsyst simulation predicts an annual energy production of approximately 6.90 GWh with a system performance ratio of approximately 81% under standard operating conditions.

While this simulated result reflects the expected performance under relatively uniform operating conditions, the inverter-level analysis conducted in this study reveals significant disparities not captured by system-level simulations. Several inverters operate well below the expected performance range, with performance ratios below 0.80 relative to the best-performing inverter.

The annual energy losses for these inverters range from 81,000 kWh to 119,000 kWh per inverter. These deviations demonstrate that although the overall plant performance ratio remains close to the simulated value, substantial localized losses occur within specific inverter zones.

This comparison highlights the added value of the proposed inverter-level machine learning framework. While PVsyst simulations provide a reliable reference for expected plant performance, they cannot capture spatially localized operational disparities arising from terrain-induced shading, module mismatch, or subsystem constraints.

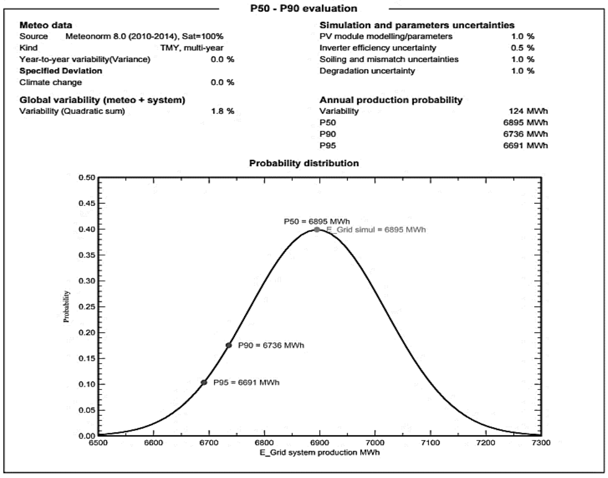

Figure 6. PVSyst Simulation Project

These results indicate that the proposed approach provides improved diagnostic capability compared with simpler classification models.

Similar deviation patterns have been reported in inverter-level fault detection and performance assessment studies, where persistent discrepancies suggest location-specific constraints or degradation in power-electronic efficiency [16], [31], [32]. These findings validate the suitability of XGBoost not only as a forecasting model but also as a diagnostic-support tool capable of revealing structural performance limitations at the subsystem level.

F. SHAP Interpretation of Inverter Performance Drivers

The SHAP analysis provides quantitative insight into the factors influencing inverter performance classification. As illustrated in Fig. 8, irradiance and temporal variables dominate the global SHAP importance. This result reflects the physical operation of PV systems, where solar radiation is the primary driver of electrical power generation. This observation is consistent with the physical characteristics of PV systems, where electrical power generation is strongly dependent on incident solar radiation. In practical terms, the relationship can be illustrated as a simple chain:

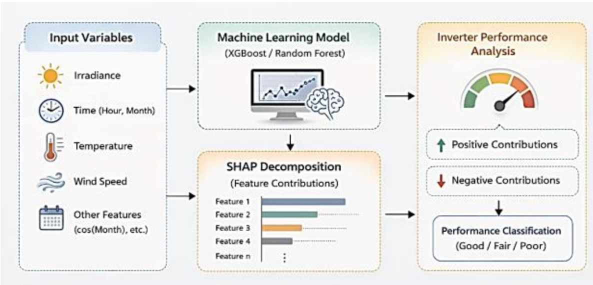

Figure 7. Conceptual illustration of the SHAP-based interpretation process

Consequently, variations in irradiance directly propagate through the PV conversion process and become the dominant driver of inverter energy output.

Temporal variables such as hour and month also exhibit strong SHAP contributions. These variables capture the diurnal and seasonal dynamics of solar energy availability. The hour variable reflects the daily solar trajectory, which determines the angle of incidence and effective irradiance received by the PV modules. Meanwhile, the month variable represents seasonal variations in solar geometry and atmospheric conditions that influence the long-term distribution of irradiance throughout the year.

More importantly, the SHAP results reveal that several underperforming inverters consistently exhibit negative feature contributions even under comparable irradiance conditions. This behavior suggests that the observed performance degradation cannot be explained solely by environmental variability. Instead, it indicates the presence of persistent subsystem-level constraints affecting specific inverter zones.

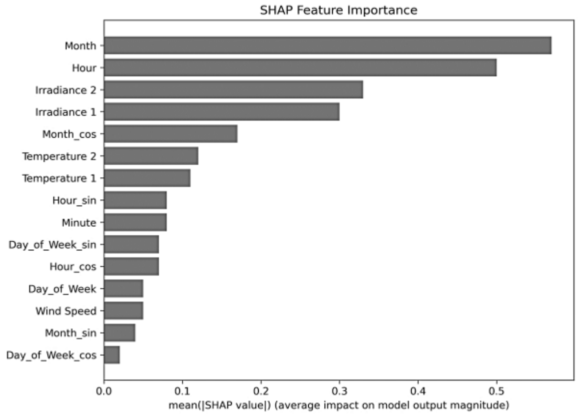

Figure 8. Relative contribution of input variables to inverter performance classification

In PV plants installed on uneven terrain, spatial variations in irradiance exposure, module orientation, and localized shading can lead to systematic differences in inverter operating conditions. These factors may lead to string mismatch, partial shading losses, or thermal imbalance between inverter zones. The clustering of negative SHAP contributions in specific inverters, therefore, supports the hypothesis that terrain-induced irradiance heterogeneity significantly contributes to the observed underperformance patterns.

Table 4. Dominant features that affect inverter energy output based on SHAP analysis

No	Features	Mean (SHAP)
1	Month	0.567
2	Hour	0.491
3	Irradiance 2	0.321
4	Irradiance 1	0.295
5	Month. Cos	0.161
6	Temperature 2	0.111
7	Temperature 1	0.101

Unlike conventional feature-importance metrics, SHAP values provide a consistent additive explanation of each model prediction. This capability enables both global interpretation of dominant environmental drivers and local analysis of inverter-specific behavior. As a result, integrating SHAP with machine learning transforms the predictive model into a diagnostic tool capable of distinguishing between normal environmental variability and persistent structural performance limitations within the PV system.

From an operational perspective, this insight is particularly valuable because it enables plant operators to identify inverter zones where localized constraints disproportionately contribute to annual energy losses. Such information can support targeted inspections, shading assessments, and condition-based maintenance strategies to improve overall plant performance.

G. Spatial Zoning of the PV Array and Inverter Performance

To further interpret the inverter underperformance patterns, the spatial layout of the PV array at the Nusa Penida power plant is analysed. As shown in Figure 2, the PV installation can be divided into three main zones (Zone A, Zone B, and Zone C) based on the physical arrangement of the module arrays and their corresponding inverter connections. Due to the uneven terrain of the installation site, these zones experience slightly different irradiance exposure and shading conditions.

When the inverter performance results are interpreted in relation to this spatial layout, several underperforming inverters are associated with arrays located in zones where terrain irregularities and potential shading interactions between rows may occur. This spatial clustering supports the SHAP analysis results, which indicate that inverter underperformance cannot be explained solely by environmental variability but is likely influenced by location-dependent factors such as terrain-induced irradiance differences, module mismatch, or localized shading effects.

H. Technical and Operational Implications

The integration of inverter-level operational analysis with machine learning provides a powerful tool for identifying specific inverter zones that contribute disproportionately to annual energy losses.

Compared with conventional system-level performance assessment, the proposed approach not only detects underperforming inverters but also provides interpretable insight into the underlying factors influencing performance degradation. This capability supports condition-based maintenance strategies, allowing operators to prioritize inspection and corrective actions in areas where persistent performance limitations are detected.

Consequently, the proposed framework can support improved operational management and long-term performance optimization for PV systems installed on complex terrain.

I. Study limitations

Despite the insights obtained from this analysis, several limitations should be acknowledged.

First, the study is based on operational data from a single PV power plant, which may limit the generalization of the results to other PV installations with different site characteristics or system configurations.

Second, the analysis relies on available inverter-level operational variables and does not incorporate module-level electrical measurements or detailed shading maps, which could provide additional diagnostic resolution.

Finally, although the XGBoost model demonstrates strong classification capability, its performance may vary when applied to datasets from different climatic conditions or plant configurations. Future work may extend this approach by incorporating multi-plant datasets, higher-resolution monitoring data, and additional diagnostic features to improve model robustness and applicability.

Table 5. Comparison of the proposed study with previous photovoltaic monitoring studies

Study	Analysis Level	Method	XAI
Marion et al.	System-level	Performance indicators	No
Padmavathi & Daniel	Plant-level	Statistical monitoring	No
Sharadga et al.	System-level	ML forecasting	No
Harrou et al.	Subsystem-level	Fault detection models	No
Song et al.	Environmental modelling	ML + SHAP	Yes
Fezzani et al.	Plant-level	Operational performance analysis	No
This study	Inverter-level	XGBoost + SHAP	Yes

IV. CONCLUSION

This study analysed inverter underperformance at the Nusa Penida PV power plant using inverter-level operational data collected throughout 2024. The results revealed significant performance disparities among 18 inverters, with seven units experiencing more than 20% reduction in annual energy production compared to the reference inverter, corresponding to energy losses ranging from 81,000 kWh to 119,000 kWh per inverter. Unlike previous studies that primarily rely on aggregated system-level indicators or short-term anomaly detection, this research integrates long-term inverter-level evaluation with quantitative annual energy loss estimation and explainable machine learning. The XGBoost classification model successfully captured nonlinear relationships between environmental variables and inverter output, while SHAP analysis provided a transparent interpretation of dominant factors influencing performance variation. The findings confirm that topographical variations and site conditions significantly affect inverter performance and demonstrate that the proposed framework offers practical advantages for condition-based maintenance prioritization, system performance evaluation, and optimization of PV power plants installed on uneven terrain.

ACKNOWLEDGEMENT

The authors would like to thank the operations and maintenance team of the Nusa Penida PV power plant for their support in providing operational data and valuable technical discussions that contributed to this research.

REFERENCES

[1] D. C. Jordan and S. R. Kurtz, "PV degradation rates—An analytical review," Prog. Photovolt.: Res. Appl., vol. 21, no. 1, pp. 12–29, 2013.

[2] B. Marion et al., "Performance parameters for grid-connected PV systems," Solar Energy, vol. 86, no. 9, pp. 2679–2694, 2012.

[3] G. Makrides et al., "Potential of photovoltaic systems in tropical regions," Energy, vol. 36, no. 2, pp. 1202–1210, 2011.

[4] A. Pillai et al., "Inverter-level performance monitoring of large-scale photovoltaic plants," Solar Energy, vol. 240, pp. 1–12, 2022.

[5] A. Mellit, S. Kalogirou, S. Shaari, H. Salhi, and A. Hadj Arab, "Artificial intelligence techniques for photovoltaic applications: A review," Renew. Sustain. Energy Rev., vol. 33, pp. 44–64, 2014.

[6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), 2016, pp. 785–794.

[7] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. 31st Conf. Neural Information Processing Systems (NeurIPS), 2017.

[8] H. Li et al., "Explainable machine learning for photovoltaic power prediction," Appl. Energy, vol. 336, p. 120796, 2023.

[9] M. G. Deceglie, "Perspective: Performance loss rate in PV systems," Solar RRL, vol. 7, no. 3, 2023.

[10] A. Dolara, R. Faranda, and S. Leva, "Aging effects on inverter efficiency in photovoltaic plants," IEEE Trans. Power Electron., vol. 30, no. 7, pp. 3983–3991, 2015.

[11] S. B. Kouro et al., "Power electronic reliability in photovoltaic inverters," IEEE Trans. Ind. Electron., vol. 68, no. 6, pp. 5412–5422, 2021.

[12] H. Sharadga, S. Hajimirza, and R. Balog, "Time series forecasting of solar power generation," Renewable Energy, vol. 151, pp. 115–124, 2020.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), 2016, pp. 1135–1144.

[14] A. Molnar, Interpretable Machine Learning. 2nd ed., 2022.

[15] M. A. Hossain et al., "Hybrid deep learning model for renewable energy forecasting," J. Cleaner Prod., vol. 299, p. 126900, 2021.

[16] S. M. Hossain, M. Mekhilef, and R. Saidur, "Machine learning-based fault detection in photovoltaic systems: A comprehensive review," Renew. Sustain. Energy Rev., vol. 168, p. 112786, 2022.

[17] Y. Wang, L. Zhao, and H. Chen, "Data-driven performance assessment of large-scale photovoltaic plants using inverter-level monitoring," Appl. Energy, vol. 336, p. 120796, 2023.

[18] F. R. Islam, K. A. Mamun, and M. Hossain, "Explainable artificial intelligence for renewable energy systems: A review," Energy Reports, vol. 9, pp. 4760–4780, 2023.

[19] J. Li and H. Yang, "Impact of terrain-induced irradiance variability on photovoltaic system performance," Solar Energy, vol. 246, pp. 1–12, 2022.

[20] A. Kumar and B. Singh, "Long-term operational performance evaluation of utility-scale photovoltaic plants," IEEE Trans. Sustain. Energy, vol. 14, no. 2, pp. 1234–1245, 2023.

[21] M. Kuzlu, U. Cali, V. Sharma, and Ö. Güler, "Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools," IEEE Access, vol. 8, pp. 187033–187045, 2020.

[22] M. Al-Shammari et al., "An interpretable deep learning model for solar power generation forecasting in a grid-connected hybrid solar system," IEEE J. Photovolt., vol. 12, no. 3, pp. 827–838, 2022.

[23] X. Zhang et al., "CEEMDAN-LSTM-XGBoost based multi-step short-term power forecasting for offshore photovoltaic systems," 2026.

[24] Z. Song, S. Cao, and H. Yang, "An interpretable framework for modeling global solar radiation using tree-based ensemble machine learning and Shapley additive explanations methods," Appl. Energy, vol. 364, p. 123238, 2024.

[25] B. Ren, Q. Wang, N. Zhou, and S. Mekhilef, "Development and validation of interpretable machine learning models for photovoltaic panel temperature prediction," Expert Syst. Appl., vol. 296, p. 128952, 2026.

[26] A. Dinakar et al., "Explainable machine learning models for predicting current and voltage in photovoltaic systems," Energy Convers. Manage.: X, vol. 30, p. 101627, 2026.

[27] M. Huang et al., "Materials-informed long-term photovoltaic energy yield prediction: Unveiling the dynamic degradation mechanisms of anti-reflective coatings via explainable machine learning," Renewable Energy, vol. 260, p. 125172, 2026.

[28] A. Abdellatif et al., "SHapley Additive exPlanations-guided rule-based energy management: Bridging machine learning interpretability and adaptive control strategies," Renewable Energy Focus, vol. 56, p. 100779, 2026.

[29] X. Gao et al., "Surface albedo evaluation in an arid-region photovoltaic power plant through field spectral radiometry and explainable machine learning," Solar Energy, vol. 299, p. 113761, 2025.

[30] J. Liu, C. Peng, and J. Zhang, "Understanding the relationship between rural morphology and photovoltaic (PV) potential using SHAP values," Appl. Energy, vol. 380, p. 125091, 2025.

[31] S. Bordihn et al., "Machine learning based identification and classification of field-operation caused solar panel failures observed in electroluminescence images," IEEE J. Photovolt., vol. 12, no. 3, 2022.

[32] P. Xi et al., "Online fault diagnosis for photovoltaic arrays based on Fisher discrimination dictionary learning for sparse representation," IEEE Access, vol. 9, pp. 12345–12358, 2021.

[33] A. Apaza-Pinto et al., "Supervised machine learning techniques for the prediction of the state of charge of batteries in photovoltaic systems," IEEE Access, vol. 10, pp. 45678–45690, 2022.

[34] M. Alagarsamy et al., "XAI-based photovoltaic energy management framework for smart cities," IEEE Access, vol. 13, 2025.