Data-Driven Predictive Modeling and Anomaly Detection for Photovoltaic System

Ka Tai LAU[1,*], Sammy, Sau Kuen YEUNG[1], Pok Man SO[1], Ka Chuen YIP[1]

[1]Electrical and Mechanical Services Department, the Government of the HKSAR, Hong Kong, China.

ABSTRACT

This paper introduces a cost-effective, data-driven framework for the maintenance of photovoltaic (PV) systems in Hong Kong (HK), addressing the surge in installations following the 2017 FiT Scheme. Although PV systems are generally low-maintenance, regular upkeep is vital due to defects that can impair output and safety. Current reactive maintenance practices focus only on electrical components and overlook comprehensive monitoring, which is insufficient to given challenges such as degradation and environmental impacts. Our framework utilizes existing network infrastructure and open data from the HK Observatory to offer a scalable solution for urban areas. By integrating local and regional weather data, and PV parameters, we achieve centralized cloud monitoring and anomaly detection without extra installation. A GRU model, optimized through tailored by prediction timeframes, feature engineering, and hyperparameter tuning. This model reduces the mean absolute error by 30% and maintains a maximum weekly error of 1.85%, surpassing traditional models. A dynamic alarm method enhances fault detection, ensuring rapid, accurate anomaly identification. Validation in real-world scenarios confirms its effectiveness. The cloud-based design allows quick deployment across similar urban sites, supporting sustainable PV management and maximizing solar energy generation in HK.

KEYWORDS

Anomaly Detection, Predictive Analytics, Data-Driven Optimization, Building energy system, Solar Panel System

INTRODUCTION

Climate change is a pressing global issue requiring urgent action. In line with the Paris Agreement, governments are setting zero-carbon targets. HK's Climate Action Plan 2050 aims to cut carbon emissions and achieve net-zero electricity by phasing out coal, boosting renewable energy, and exploring new sources. The city targets a renewable energy share of 7.5% to 10% by 2035, increasing to 15% thereafter, with a goal of net-zero electricity before 2050. Solar energy is critical to this transition, with PV installations expected to rise due to the Feed-in Tariff Scheme and government

* Corresponding author email: ktlau@emsd.gov.hk

initiatives. Launched in 2017, the Scheme encourages renewable energy by allowing individuals and businesses to sell renewable power at premium rates. This has resulted in over 23,000 applications for PV installations, highlighting solar energy's growing importance.

NEED FOR MAINTENANCE OF PV SYSTEM IN HK

PV systems typically last over 25 years, but regular maintenance is crucial due to degradation and external damage. Issues include power degradation, electrical corrosion, cell breakage, and delamination. (Di Tommaso et al., 2022; Bendale et al., 2023). External factors like shading, soiling, and rooftop slope can reduce power generation, while internal issues may lead to hotspots, causing overheating and potential damage, including fires. Research indicates over 40% of systems experience degradation (Mgonja & Saidi, 2017), with energy losses around 10% (Drif et al., 2008), underscoring the need for continuous monitoring and maintenance. In HK, PV systems currently use a reactive maintenance approach due to their small and distributed nature. Although preventive or predictive maintenance tools like thermal imaging and proprietary monitoring exist, they are not preferred due to high resource requirements and HK-specific challenges such as urban density and shading from tall buildings. To address these issues, cost-effective monitoring solutions are needed. This study proposes a framework using existing data and an AI model tailored to HK's PV installations to enhance operation and maximize power generation.

LITERATURE OF ADVANCE MAINTENANCE APPROACH

There are several advanced approaches address performance evaluation and fault detection which can be categorized into three dimensions: Electrical Parameter Monitoring, Visual Analysis, and Power Generation Comparison (Bosman et al., 2020).

Electrical Parameter Monitoring involves analyzing real-time electrical data like voltage and current, providing immediate insights into system performance. It enables quick response to issues such as module or connection failures. However, the high cost of sensing infrastructure can be prohibitive for the use in HK. While effective for short-term issues, it might not detect long-term degradation or complex problems. Innovations include a simulation program for I-V curve analysis under partial shading (Gallardo-Saavedra & Karlsson, 2018) and AC impedance spectroscopy for fault detection beyond traditional I-V characteristics (Bendale et al., 2023). A low-cost, real-time supervision method using Voc-Isc curves for early fault detection has also been proposed (García et al., 2022).

Visual Analysis process image to detect physical defects in PV modules, such as cracks or hotspots, revealing issues not apparent through electrical monitoring. For example, thermal imaging software enhances fault detection and system efficiency (Constantin et al., 2023). However, regular site visits are time-consuming and costly, especially for distributed sites. Manual inspections also risk human error and may not fully assess system performance. To overcome these limitations, advanced techniques integrate

CNN and UAVs to automate image analysis (Bendale et al., 2023). This reduces manual inspections and efficiently accesses hard-to-reach areas. Another study uses YOLOv3 for similar purposes (Di Tommaso et al., 2022). Research correlating image data with power loss offers deeper insights (Cavieres et al., 2022). These advancements provide valuable insights but still face limitations in HK due to the distributed nature of systems.

Power Generation Comparison approach leverages historical data and weather conditions, to estimate expected PV output. Analyzing deviations between predicted and actual outputs identifies suboptimal performance or potential faults. The capital investment and OPEX are minimal when compared to the others while offering desirable alarm, allowing operators to proactively address performance issues. Recent studies use AI models like ANN and LSTM to predict solar irradiation, temperature, and corresponding power generation as they can handle learn complex data patterns, (De Benedetti et al., 2018). LSTM models, capturing long-term dependencies in time series data, are valuable for accurate estimation. Studies have combined LSTM with K-means clustering for daily power output predictions and employed edge-computing systems for city-level applications (Vicente-Gabriel et al., 2021). After all, the power generation comparison approach is getting more and more attention as it is more suitable for the city-level application utilizing the public data.

APPLICATION OF ADVANCE MAINTENANCE APPROACH IN HK

Given the feasibility of AI models for power generation comparison, their suitability for PV monitoring in HK requires investigation due to unique geographical challenges. This study develops an approach tailored to these conditions. The goal is to identify the most suitable model for PV monitoring, enhancing performance evaluation and maintenance practices, and supporting sustainable energy.

The Electrical and Mechanical Services Department (EMSD) maintains government buildings, 30 numbers of PV systems remotely connected to the Regional Digital Control Centre (RDCC) at Kowloon Bay via Government Wide IoT Network (GWIN). This paper proposes an AI model utilizing the existing infrastructure and cloud-based RDCC to detect anomalies without additional investment. A specific site was selected for detailed analysis due to its extensive dataset, serving as a representative case study of the anomaly detection of PV system in HK. There are 30 thin-film PV panels arranged in three strings, each linked to the same inverter with a total rated output of 9750W. According to minimal requirement in government building, sensors for electrical signals, ambient temperature and solar irradiance are installed.

***Figure 1**. PV system installed in the site and related sensor*

In addition to local data, regional meteorological data from the HK Observatory (HKO) is considered to enhance solar energy generation predictions. HK's meteorological network includes 50 automatic weather stations, with a density of 22.29 stations/m^2. Solar radiation data is available at King's Park and HK International Airport (HKO, 2024), emphasizing the need for strategic data integration.

DATA PREPARATION AND EXPLORATION

Data cleansing ensures integrity using domain-specific and mathematical methods to identify anomalies. Approximately 0.2043% of the data showed issues such as excess power generation and missing points. Surplus data was removed, and interpolation filled gaps to ensure dataset accuracy. After cleansing, detailed exploration enhanced understanding of dataset characteristics and revealing other abnormalities. Correlation heatmap indicated a strong correlation between solar irradiation and power output (coefficient 0.98), consistent with previous research (Arshi et al., 2019). This underscores solar irradiation's significant influence on PV system performance.

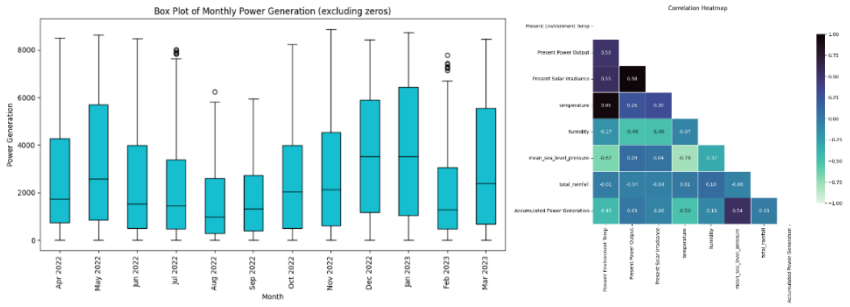

***Figure 3**. Boxplot of monthly energy generation and correlation map of all features*

Baseline Model and Selection of Model

A baseline model, done by Castillo-Rojas et al. (2023), employs a Hybrid Model of Recurrent and Shallow Neural Networks using weather data, achieving an MSE of 0.03, MAE of 0.09, and R^2 of 0.96, serving as a normalized benchmark. Five models are selected for further review and development which are Random Forest (RF), Gradient Boosting (GB), Long Short-Term Memory (LSTM), Artificial Neural Networks (ANN), Gated Recurrent Units (GRU).

Evaluation of Model Performance

To assess model performance, we used several evaluation metrics to understand accuracy. Standard metrics such as MSE, MAE, and R^2 serve as primary benchmarks. Additionally, Peak Absolute Percentage Error (PAPE) is introduced to capture the largest absolute difference between predicted and actual values, highlighting worst-case prediction scenarios. The study utilized 14 months of historical data, divided into an 80% training set and a 20% testing set for model development. An additional 4 months of unseen data were reserved for independent validation, and 2 months for real-life application testing. Model predictions were compared with actual daily, weekly, and monthly power generation metrics.

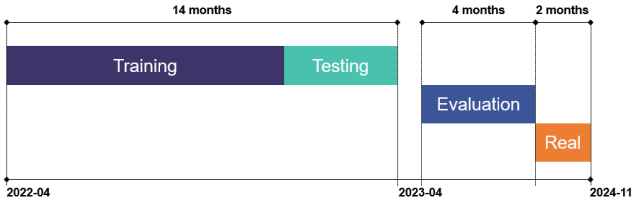

***Figure 4**. Data adopted for the study*

Enhancements to the Model

Enhancement to the model was carried out to improve the accuracy of the five models to achieve the real-world applicability. Data analysis was first conducted to find the optimal time frame for generating alarms and the minimum sensor sampling frequency. The dataset, initially collected at 1-minute intervals, was resampled to 10 minutes, 1 hour, and 1 day. During training stage, GB and RF outperformed deep learning models, with MSE, MAE, and R^2 values of 0.0016, 0.0191, and 0.9613, respectively, for the 10-minute interval. However, transitioning to a 1-hour interval increased MSE by 31.25% and 13.09%. In validation stage, models trained on 1-hour data significantly outperformed those using 10-minute intervals. RF was the top performer, achieving daily, weekly, and monthly MAE values of 4.11%, 2.75%, and 2.15%, showing improvements of 55.58%, 72.39%, and 78.74% over 10-minute data. No further enhancements were observed in resampling to daily. The 10-minute time frame offered a larger training dataset, enhancing performance in training stage but increasing overfitting risk during validation. Without regularization, 10-minute intervals may lack robustness with unseen data. To ensure real-world applicability, 1-hour time frame is adopted, balancing trend capture with robustness.

Feature Engineering

Feature engineering was employed to streamline input variables and identify key predictors to improve model accuracy. Findings indicated that local irradiance and regional rainfall were the most impactful inputs for predicting PV output power. Interestingly, these results diverged from the typical inverse relationship with panel temperature. This anomaly may stem from the use of a single installed temperature sensor inadequately representing actual site conditions. Regional weather data significantly affected MAE, leading to the exclusion of most data from the HKO, except rainfall. This exclusion improved model performance, likely due to the over 2-kilometer distance between the observatory and the test site. Greater improvements are anticipated with more localized data. In cases lacking local irradiance data, substituting with regional radiation and rainfall maintained acceptable performance, with an MAE of 12.61%. The best input features are local irradiance and rainfall.

***Table 2.** Top three performing input features*

Input Features	*Average MAE (%)*
Local Irradiance + Regional Rainfall	4.33
Local Irradiance + Local Temperature	4.58
All Local and Regional Data	7.45

Regularization and Hyperparameter Tuning

Deep learning models often face overfitting due to their complexity. To counter this, regularization techniques were implemented to improve generalization. Results showed slight performance enhancements. Specifically, L1 and L2 regularization improved the LSTM and GRU models during both training and validation, enhancing their generalization capabilities. However, improvements were not observed in the ANN. Hyperparameters are crucial for optimizing model performance, as they are not learned from the data. This study employed hyperparameter tuning to enhance predictive capabilities. Optuna, a popular optimization library, was used for machine learning models, while Random Search was applied to deep learning models. The tuning process involved systematically exploring various hyperparameter combinations to identify the optimal configuration for each model.

Result and Discussion

In the final stage, the PAPE was used as a direct measure of prediction accuracy, complementing the MAE. Initially, MSE outperformed MAE as a loss function, with improvements of 51.40%, 12.11%, and 20.69% for daily, weekly, and monthly time frames, respectively. After hyperparameter tuning, MSE generally enhanced model performance, notably in most models. MAE showed less consistent improvements, with only the LSTM model benefiting further. Ultimately, the GRU model, fine-tuned with Random Search and using MSE as lost function, proved most effective, achieving PAPE values of 23.04%, 1.85%, and 1.14% for daily, weekly, and monthly comparisons. The overall weekly result is very accurate for perform the task in detection of possible fault in individual PV panels.

Alarm Settings and Generation

Based on the best model developed with high accuracy in the weekly timeframe, any significant discrepancy between the actual measured and the predicted output will trigger an alarm to notify the Operation & Maintenance staff. Two alarm-setting approaches are compared. The first uses the weekly PAPE with a 50% safety margin, leading to delayed alarms, especially if faults occur later in the week. The second, more systematic approach, uses the 25th and 75th percentiles with the same offset, resulting in narrower limits.

Table 3. *Upper and Lower Alarm Setting of the two Approaches*

Alarm Limit	Simple off set	Off-set from percentiles
Upper	+2.78%	1.26%
Lower	-2.78%	-2.96%

Testing and Application of Models in Real-Case Scenarios

The two alarm settings were tested on an additional two-month dataset, including a control fault case where output was lost in 1 of 30 panels, starting November 2. The simple offset approach generated an alarm on November 12, one-week post-fault. Conversely, the percentile offset approach triggered an alarm on November 5, within the same week as the fault, offering a more timely response.

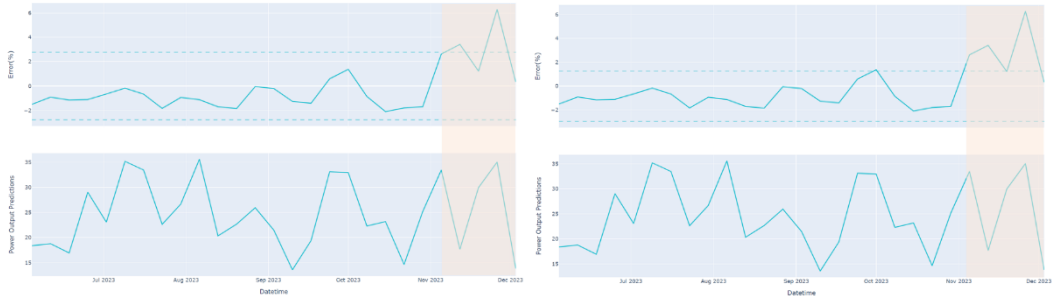

***Figure 5**. Alarm Generation By The Two Approach*

Conclusion

This paper introduces a comprehensive framework for detecting abnormalities in PV systems using existing infrastructure, eliminating the need for additional installations. It proactively identifies and addresses potential operational issues, ensuring optimal performance and efficiency. The proposed GRU model demonstrates effective hourly prediction capabilities by leveraging key features such as local solar irradiation and regional rainfall data. Enhancements in time domain analysis, feature selection, metrics, regularization, and hyperparameter tuning have significantly improved model performance, outperforming the baseline model.

Additionally, a statistical method for setting alarms enhances the detection of abnormal energy patterns, enabling timely intervention and maintenance to ensure smooth PV system operation. The trained model is stored on the RDCC cloud server, receiving live data from the site and HKO, and processes it weekly to generate alarms, notifying O&M staff of any abnormalities. The framework shows significant potential for expansion to other locations with minimal costs. Its versatility and scalability make it a promising tool for improving PV system operation and maintenance globally. By optimizing green energy generation, it aids the transition towards carbon neutrality, contributing to a sustainable and environmentally friendly future.

Limitation and Future Work

While the model shows promise, it has not been extensively tested on larger systems with over 100 panels. The reliance on a single local irradiation sensor suggests that a more sophisticated model is necessary for larger-scale applications. Additionally, the model struggles to provide accurate predictions for systems lacking a local solar irradiation sensor, a common issue in cost-constrained residential PV systems. Future work will focus on improving prediction accuracy for these cases. One approach is to embed the algorithm in a compact, cost-effective device for integration into systems without local sensors, enhancing their performance with predictive capabilities. Another strategy is to refine the model by interpolating regional weather data to estimate local conditions more accurately. This would allow for more precise predictions without direct local sensor data. Ongoing research in these areas will advance the model's applicability and accuracy across diverse PV systems and operational scenarios.

REFERENCES

Di Tommaso, A., Betti, A., Fontanelli, G., & Michelozzi, B. (2022). A multi-stage model based on YOLOv3 for defect detection in PV panels based on IR and visible imaging by unmanned aerial vehicle. Renewable Energy, 193, 941–962.

Bendale, H., Aswar, H., Bamb, H., Desai, P., & Aher, C. N. (2023) Deep learning for solar panel maintenance: detecting faults and improving performance.

Mgonja, C., & Saidi, H. (2017). Effectiveness on implementation of maintenance management system for off-grid solar pv systems in public.

Drif, M., Pérez, P., Aguilera, J., & Aguilar, J. (2008). A new estimation method of irradiance on a partially shaded PV generator in grid-connected photovoltaic systems. Renewable Energy, 33(9), 2048–2056.

Bosman, L. B., Leon-Salas, W. D., Hutzel, W., & Soto, E. A. (2020). PV System Predictive Maintenance: challenges, current approaches, and opportunities. Energies, 13(6), 1398.

Osawa, S., Nakano, T., Matsumoto, S., Katayama, N., Saka, Y., & Sato, H. (2016) Fault diagnosis of photovoltaic modules using AC impedance spectroscopy.

García, E., Ponluisa, N., Quiles, E., Zotovic-Stanisic, R., & Gutiérrez, S. C. (2022). Solar panels string predictive and parametric fault diagnosis using Low-Cost sensors. Sensors, 22(1), 332.

Constantin, A., Iosif, G., Chihaia, R., Marin, D., Shehadeh, G. U. A., Karahan, M., Gerikoglu, B., & Stavrev, S. (2023). Preventive Maintenance in Solar Energy Systems and Fault Detection for Solar Panels based on Thermal Images. Electrotehnica Electronica Automatica, 71(1), 1–12.

Cavieres, R., Barraza, R., Estay, D., Bilbao, J., & Valdivia-Lefort, P. (2022). Automatic soiling and partial shading assessment on PV modules through RGB images analysis. Applied Energy, 306, 117964.

De Benedetti, M., Leonardi, F., Messina, F., Santoro, C., & Vasilakos, A. (2018). Anomaly detection and predictive maintenance for photovoltaic systems. Neurocomputing, 310, 59–68.

Vicente-Gabriel, J., Gil-González, A., Luis-Reboredo, A., Chamoso, P., & Corchado, J. M. (2021). LSTM Networks for Overcoming the Challenges Associated with Photovoltaic Module Maintenance in Smart Cities. Electronics, 10(1), 78.

RSamara, S., & Natsheh, E. (2019) Intelligent Real-Time Photovoltaic Panel monitoring System using artificial neural networks.

Maitanova, N., Telle, J., Hanke, B., Grottke, M., Schmidt, T., Von Maydell, K., & Agert, C. (2020). A machine learning approach to Low-Cost photovoltaic power prediction based on publicly available weather reports. Energies, 13(3), 735.

Hong Kong Observatory. (2024). Information of Weather station. https://www.hko.gov.hk/en/cis/stn.htm, last accessed on 3 September 2024.

Arshi, S., Zhang, L., & Strachan, R. (2019). Weather based photovoltaic energy generation prediction using LSTM networks. ResearchGate.

Castillo-Rojas, W., Quispe, F. M., & Hernández, C. (2023). Photovoltaic Energy Forecast Using Weather Data through a Hybrid Model of Recurrent and Shallow Neural Networks. Energies, 16(13), 5093.