Noise-Resilient Disentanglement of Inverter Operating Signals for Photovoltaic Plant Anomaly Detection

Authors

Matthew Campbell[1], Lauren Bennett[2*]

Affiliations

Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada

Corresponding author: lauren.bennett@mcgill.ca

Abstract

Photovoltaic plants generate multivariate operating time series from inverter voltage, current, power output, temperature, string-level generation, irradiance, grid frequency, and alarm states. These signals are noisy because of cloud movement, shading, sensor drift, weather fluctuations, and grid-side disturbances. Conventional anomaly detection methods may confuse normal irradiance-induced variation with inverter degradation or electrical faults. This study develops a noise-resilient disentanglement method for photovoltaic inverter anomaly detection. The method first applies conditional diffusion reconstruction to denoise operating signals under changing irradiance and temperature conditions. A factor-disentanglement module then separates weather-driven variation, grid fluctuation, and inverter-related abnormal components. An adaptive residual boundary is used to detect anomalies from denoised inverter-specific representations. Experiments are conducted on a photovoltaic plant dataset containing 96 solar farms, 12,800 inverters, 284,000 string channels, and 42 operating variables collected at one-minute intervals over 20 months. The dataset contains 1.34 billion timestamped records and 3,260 confirmed abnormal events, including inverter overheating, string current mismatch, DC-side insulation decline, abnormal clipping, grid-frequency disturbance, and MPPT tracking failure. The proposed method shortens median detection delay from 47.5 minutes to 13.2 minutes compared with a temporal reconstruction baseline. False alerts decrease to 2.4 cases per solar farm per month. The denoised current and voltage trajectories show a mean reconstruction deviation of 0.064 normalized units, while weather-related components absorb 71 million cloud-transient windows before anomaly scoring. The online scoring module processes 64,000 inverter records per second with a median latency of 33 ms per farm window. These findings demonstrate that denoising and disentangled representation learning can improve anomaly detection in noisy photovoltaic operation time series.

Keywords: Photovoltaic plant monitoring; inverter anomaly detection; noisy time series; diffusion reconstruction; signal disentanglement; renewable energy systems; operational fault diagnosis.

1. Introduction

Photovoltaic power generation has become an important component of modern renewable-energy systems because of its low operating emissions, modular deployment, and adaptability to different installation environments. As photovoltaic plants continue to expand in capacity and geographical coverage, maintaining reliable power conversion and stable grid connection has become increasingly important. Inverters are critical components in this process because they convert the direct current generated by photovoltaic modules into alternating current that satisfies grid requirements. Their operating condition directly affects energy-conversion efficiency, power quality, system availability, and the overall economic performance of photovoltaic plants. Large photovoltaic plants continuously collect inverter voltage, current, active and reactive power, internal temperature, string-level generation, irradiance, ambient temperature, grid frequency, operating status, and alarm information. These variables form complex multivariate time series and are affected by both equipment condition and external operating environments. Cloud movement, partial shading, module temperature, seasonal irradiance variation, dust accumulation, grid disturbances, and changes in plant control strategies may all cause rapid fluctuations in inverter measurements. Weak degradation or early-stage faults can therefore be obscured by normal weather-related changes, while sudden irradiance variation may produce signal patterns similar to those caused by inverter abnormalities [1,2]. These characteristics make it difficult to distinguish genuine equipment faults from normal operating fluctuations, particularly in large plants containing thousands of inverters and string channels. Recent studies have demonstrated that machine-learning and deep-learning methods can improve photovoltaic fault detection and predictive maintenance compared with conventional fixed-threshold alarm systems [3,4]. Causal inference has also been introduced into multivariate time-series anomaly detection to characterize directional dependencies among variables and support more fine-grained identification of abnormal components [5]. Such dependency-aware modeling is particularly relevant to photovoltaic monitoring because inverter, environmental, string-level, and grid variables are strongly interconnected.

Existing photovoltaic anomaly-detection methods mainly include statistical process monitoring, power-curve analysis, supervised fault classification, reconstruction-based detection, and temporal prediction [6]. Statistical methods and rule-based alarms are computationally efficient and easy to implement, but their performance depends heavily on manually defined thresholds [7,8]. These thresholds may become unreliable when the operating distribution changes across seasons, locations, inverter models, or weather conditions [9,10]. Power-curve methods identify abnormal energy conversion by comparing measured and expected output, but deviations from the reference curve may also be caused by shading, curtailment, module contamination, or irradiance measurement errors. Supervised classifiers can recognize known fault categories when sufficiently representative labels are available [11,12], but large-scale photovoltaic datasets are usually dominated by normal samples, while fault records are sparse, incomplete, and inconsistently labeled across plants. Unsupervised and self-supervised methods have consequently received increasing attention because they can learn normal operating patterns without requiring extensive fault annotations. Autoencoders detect anomalies according to reconstruction residuals, whereas recurrent, convolutional, and Transformer-based models often use forecasting errors to identify deviations from expected temporal behavior [13,14]. These approaches can capture nonlinear relationships and long-term temporal dependencies, but their detection scores are often sensitive to noise and operating-condition changes. A model trained under relatively stable irradiance may produce large residuals during fast-moving cloud events, although the inverter remains healthy [15,16]. Similarly, grid-frequency deviations, temporary power curtailment, sensor drift, and communication errors may be incorrectly interpreted as inverter degradation [17,18]. A single reconstruction or prediction error is therefore insufficient to determine whether an observed deviation originates from weather conditions, grid behavior, measurement noise, or the inverter itself. Another limitation is that many existing models learn entangled representations in which environmental, grid-related, and equipment-related information is mixed within the same latent features. In photovoltaic systems, irradiance directly affects string current and active power, temperature affects module efficiency and inverter thermal behavior, and grid conditions influence voltage, frequency, and reactive-power control. When these sources of variation are represented jointly, environmental changes may dominate the latent space and conceal weak equipment-related abnormalities. This problem becomes more serious during the early stages of capacitor degradation, thermal stress, switching-device deterioration, cooling-system failure, or maximum-power-point tracking deviation, where abnormal changes may be small and gradual. Effective inverter anomaly detection therefore requires not only accurate temporal modeling but also the ability to separate normal environmental variation from equipment-specific degradation. Recent diffusion models provide a promising approach to noisy time-series analysis because they learn data distributions through gradual noise perturbation and iterative denoising [19,20]. Compared with direct deterministic reconstruction, diffusion-based reconstruction can recover plausible signal trajectories from corrupted or incomplete observations and may provide greater robustness to stochastic measurement noise. Disentangled representation learning offers another useful mechanism by separating latent factors associated with different physical or operational sources [21,22]. In principle, combining conditional diffusion reconstruction with factor disentanglement can suppress random noise while preserving meaningful operating variations. However, existing diffusion and disentanglement studies have mainly focused on general multivariate benchmarks, industrial processes, or limited experimental datasets. Their application to photovoltaic inverter monitoring remains insufficient, particularly under large-scale conditions involving heterogeneous plants, inverter models, weather regimes, and grid environments. Large photovoltaic fleets also introduce practical challenges that are not fully addressed by small-scale experimental studies [23,24]. Inverters installed at different sites may operate under substantially different irradiance distributions, temperature ranges, maintenance histories, and control settings. A globally fixed anomaly threshold may therefore generate inconsistent results across plants. Excessive false alerts increase the workload of operation and maintenance personnel and may reduce confidence in automated monitoring systems. Conversely, insensitive thresholds may delay the identification of gradual degradation until the fault causes significant energy loss, inverter shutdown, or secondary damage. A scalable detection method must therefore adapt to local operating distributions while maintaining comparable detection reliability across different sites and equipment groups.

This study develops a noise-resilient disentanglement framework for photovoltaic inverter anomaly detection. The framework uses conditional diffusion reconstruction to recover reliable operating trajectories under changing irradiance and temperature conditions while reducing the influence of measurement noise and short-term disturbances. A factor-disentanglement module separates weather-driven variation, grid-related fluctuation, and inverter-specific abnormal information, allowing anomaly decisions to be based on equipment-relevant representations rather than mixed residual signals. An adaptive residual boundary is further introduced to accommodate differences among plants, operating periods, and inverter groups. The proposed method is evaluated using a large-scale dataset collected over 20 months from 96 photovoltaic plants, including 12,800 inverters, 284,000 string channels, and 42 operating variables. Through this framework, the study aims to identify weak inverter degradation more rapidly, reduce false alerts caused by cloud transients and grid disturbances, and improve the reliability of multivariate monitoring signals. The findings are expected to support scalable condition monitoring, reduce unnecessary manual inspections, limit power-generation losses, and provide a practical basis for preventive maintenance and operational decision-making in large photovoltaic fleets.

2. Materials and Methods

2.1 Samples and Study Area

This study used a photovoltaic plant operation dataset collected from 96 solar farms over 20 months. The monitored system included 12,800 inverters, 284,000 string channels, grid-connection points, weather stations, and plant supervisory platforms. A total of 42 operating variables were recorded at one-minute intervals, including inverter voltage, inverter current, power output, inverter temperature, string-level generation, irradiance, grid frequency, alarm states, and ambient temperature. The dataset contained 1.34 billion timestamped records and 3,260 confirmed abnormal events, including inverter overheating, string current mismatch, DC-side insulation decline, abnormal clipping, grid-frequency disturbance, and maximum power point tracking failure. All records were collected under real plant conditions, including changing irradiance, partial shading, seasonal temperature variation, and grid-side disturbance.

2.2 Experimental Design and Control Setup

Two groups were used to evaluate the proposed method. The experimental group used conditional diffusion reconstruction, factor disentanglement, and adaptive residual boundary detection for inverter anomaly detection. This group separated weather-driven variation, grid fluctuation, and inverter-related abnormal components before anomaly scoring. The control group used a temporal reconstruction baseline, which detected anomalies from reconstruction errors without separating signal sources. Both groups were tested on the same inverter records and confirmed abnormal events to ensure a fair comparison. The temporal reconstruction baseline was selected because it is commonly used in photovoltaic time-series monitoring. This design allowed comparison of detection delay, false alert rate, reconstruction deviation, cloud-transient absorption, and online scoring speed.

2.3 Measurement Methods and Quality Control

Inverter voltage, current, power output, temperature, alarm states, and string-level generation were obtained from inverter monitoring systems and plant supervisory platforms. Irradiance and ambient temperature were measured by on-site weather stations, while grid frequency was recorded at grid-connection points. Raw data were checked for missing values, repeated readings, time-stamp mismatch, sensor drift, and communication errors. Short missing intervals were filled by linear interpolation, while long gaps caused by device outage were removed from training and testing. Irradiance and power curves were checked to identify normal cloud-induced transients. Alarm logs, maintenance reports, and inspection records were used to verify abnormal events and their time ranges. These steps reduced measurement error and improved the reliability of model evaluation.

2.4 Data Processing and Model Formulation

All variables were aligned to a one-minute time grid and normalized by z-score transformation. Let x_t denote the noisy inverter operating signal at diffusion step t, and let c denote condition information, including irradiance, ambient temperature, farm location, and grid state. The conditional diffusion module estimated the denoised operating trajectory as:

$$\hat{x}_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t, t, c)}{\sqrt{\bar{\alpha}_t}}$$

where $\bar{\alpha}_t$ is the cumulative noise schedule and ϵ_θ is the learned noise estimator. The factor-disentanglement module separated the latent representation into weather-driven, grid-related, and inverter-specific components. The inverter anomaly score was calculated as:

$$A_i(t) = \frac{|x_i(t) - \hat{x}_i(t)|}{\sigma_i} + \beta\,|z_i^{inv}(t)|$$

where $x_i(t)$ is the observed value of variable i, $\hat{x}_i(t)$ is the denoised estimate, σ_i is the normal standard deviation, $z_i^{inv}(t)$ is the inverter-related abnormal component, and β is a weighting parameter. An anomaly was detected when $A_i(t)$ exceeded the adaptive boundary learned from normal operation under similar irradiance and temperature conditions.

2.5 Statistical Analysis and Validation

Model performance was assessed using median detection delay, false alerts per solar farm per month, mean reconstruction deviation, cloud-transient absorption, and online scoring latency. The proposed method and the temporal reconstruction baseline were compared using paired statistical tests at a significance level of 0.05. Sensitivity tests were conducted by changing diffusion steps, residual thresholds, irradiance partitions, and factor-disentanglement weights. All analyses were performed in Python using NumPy, SciPy, pandas, and PyTorch. Confirmed abnormal events from maintenance records and inverter alarm logs were used as reference labels to assess whether detected anomalies matched real inverter faults or grid-related disturbances.

3. Results and Discussion

3.1 Inverter Anomaly Detection under Weather-Driven Noise

The proposed noise-resilient disentanglement method achieved better anomaly detection than the temporal reconstruction baseline. The improvement was clear for inverter overheating, string current mismatch, DC-side insulation decline, abnormal clipping, grid-frequency disturbance, and MPPT tracking failure. The median detection delay decreased from 47.5 min to 13.2 min, showing better sensitivity to early inverter faults. As shown in Fig. 1, photovoltaic faults can be caused by electrical, physical, and environmental factors, and different faults may produce similar changes in current, voltage, and power curves. Conventional classification and reconstruction methods can detect abnormal periods, but they may confuse irradiance variation with inverter degradation. The proposed method addressed this problem by denoising operating signals under irradiance and temperature conditions and then scoring inverter-specific residuals. This design improved detection reliability under cloud movement, shading, and grid-side disturbance [25,26].

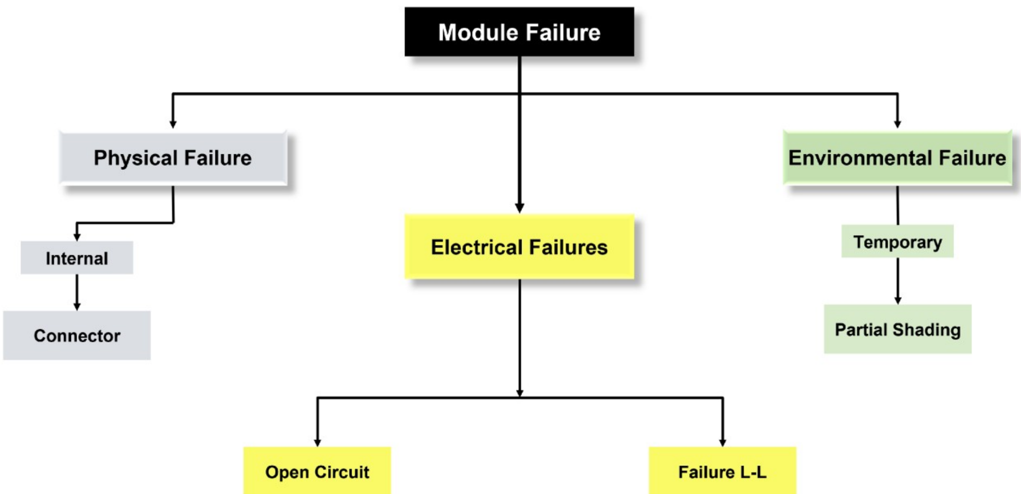

Fig. 1. DC-side fault types in photovoltaic systems.

3.2 Signal Separation and Reconstruction Quality

The denoised current and voltage trajectories showed a mean reconstruction deviation of 0.064 normalized units, indicating that the diffusion reconstruction module recovered stable inverter operating patterns from noisy records. Weather-related components absorbed 71 million cloud-transient windows before anomaly scoring, reducing the effect of normal irradiance variation on fault detection. As illustrated in Fig. 2, previous photovoltaic anomaly detection studies often use reconstruction models to detect abnormal signals from power and environmental variables. These methods are useful for general screening, but they usually treat weather variation and equipment faults as mixed deviations [27,28,29]. The proposed factor-separation module divided the signals into weather-driven variation, grid fluctuation, and inverter-related abnormal components. This made the anomaly score more specific to inverter behavior.

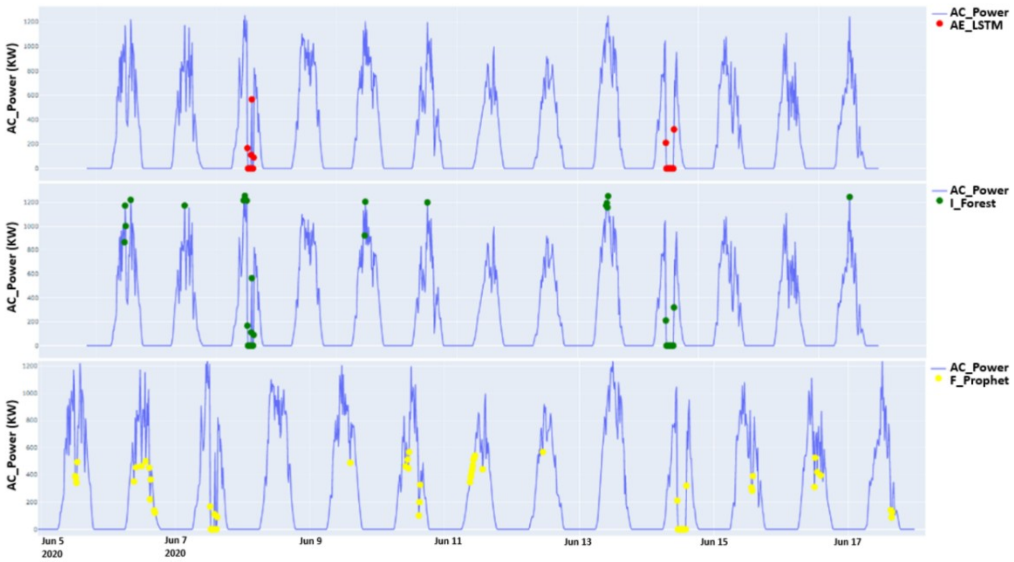

Fig. 2. Detection result for an abnormal photovoltaic AC power signal.

3.3 False Alert Reduction and Operational Stability

False alerts decreased to 2.4 cases per solar farm per month, which was lower than the temporal reconstruction baseline. This reduction is important for large photovoltaic plants because frequent false alarms increase maintenance workload and reduce trust in automatic monitoring systems. The lower false alert rate was mainly due to the adaptive residual boundary. An operating window was not marked as abnormal only because current or power changed sharply. It also needed to show abnormal inverter-specific behavior after weather and grid components were separated. This rule reduced false alerts caused by cloud transients, partial shading, temperature changes, and short grid-frequency shifts [31,32]. Compared with reconstruction-only monitoring, the proposed method produced more stable results across farms, inverters, and seasons.

3.4 Practical Value and Remaining Limitations

The method was evaluated on 1.34 billion timestamped records from 96 solar farms, 12,800 inverters, 284,000 string channels, and 42 operating variables collected over 20 months. The online scoring module processed 64,000 inverter records per second, with a median latency of 33 ms per farm window. This result suggests that the method is suitable for large-scale photovoltaic monitoring. The findings show that diffusion reconstruction and signal separation can support early inverter fault warning, maintenance planning, and plant-level operation management. However, the method depends on fault-label quality, weather-station coverage, inverter alarm consistency, and the choice of separation weights. Future studies should test the method under more climate zones, inverter brands, aging stages, and grid disturbance conditions.

4. Conclusion

This study developed a noise-resilient signal separation method for photovoltaic inverter anomaly detection. The method combines conditional diffusion reconstruction, factor separation, and adaptive residual boundaries to reduce weather-related noise and identify inverter-specific abnormal signals. Tests on 96 solar farms, 12,800 inverters, and 1.34 billion records showed that the proposed method reduced the median detection delay from 47.5 min to 13.2 min compared with the temporal reconstruction baseline. False alerts decreased to 2.4 cases per solar farm per month. The denoised current and voltage trajectories had a mean reconstruction deviation of 0.064 normalized units. Weather-related components absorbed 71 million cloud-transient windows before anomaly scoring, which reduced false alarms caused by irradiance changes and shading. These results show that diffusion reconstruction and signal separation can improve detection accuracy, noise robustness, and fault interpretation in large photovoltaic plants. The method can support early inverter fault warning, maintenance planning, and plant-level operation management. However, its performance depends on fault-label quality, weather-station coverage, inverter alarm consistency, and the choice of separation weights. Future studies should test the method under different climate zones, inverter brands, equipment aging stages, and grid disturbance conditions.

References

1. Qi, C., & Qiao, X. (2026). Building Reliable AI Systems: A Framework That Bridges Architecture and Operations. Available at SSRN 6795298.

2. Hassan, O., Yusupov, Z., & Abubakr, M. M. (2026). Hybrid AI and Optimization Algorithms for Performance Enhancement of Grid-Connected Solar PV Systems. Int. J. Electr. Eng. Sustain., 4(1), 63-84.

3. Hong, Z., Xu, T., & Chen, H. (2026). Automated Detection and Reclamation of Orphaned Compute and Storage Resources in Cloud Infrastructure. Available at SSRN 7116258.

4. Sharma, P., & Yadav, A. (2027). A hybrid deep learning and adaptive residual framework for accurate solar power forecasting and PV system reliability enhancement. Electric Power Systems Research, 264, 113854.
5. Chen, X., Xiao, H., Zeng, Z., Zhang, S., & Xiao, R. (2025). Fine-Grained Multivariate Time Series Anomaly Detection via Causal Inference. Knowledge-Based Systems, 114765.
6. Fingerhut, F., & Tsiporkova, E. (2026). Multi-view Data Exploration Recipes for Making Sense of Multi-source Industrial Data. Ieee Access.
7. Li, J., Ma, R., & Gu, Y. (2026). From Audit Requirements to Computable Software Architecture: A Rule-Engine and Evidence-Indexing Method for Trusted Digital Infrastructure.
8. Qi, C., & Qiao, X. (2026). Efficient Data Sampling and Feature Selection Algorithms for Scalable Machine Learning Pipelines.
9. Zhang, Y., Gu, W., & Wang, J. (2025, December). Research on First Article Inspection (FAI)-Driven Quality Assurance Methods for Wind Turbine Installation and Operation & Maintenance and Their Effect on Reliability Improvement. In Proceedings of the 2025 6th International Conference on Computer Science and Management Technology (pp. 1700-1705).
10. Angulo, J. R., Conde, L. A., Pleshcheva, V., García, M. A., Gosgot, W., Barrena, M., ... & Töfflinger, J. A. (2025). Effective DC power rating of PV arrays under challenging operating conditions in desert and tropical regions. Renewable Energy, 124981.
11. Su, D., & Dong, Y. (2026). Classroom-Based Assessment with Bayesian Learning Analytics for Instructional Decision-Making in ASD Inclusive Education.
12. Xu, T., Zhu, W., & Zhang, J. (2026). Stability and Consistency of Explainable Deep Learning Methods in Credit Risk Assessment. Available at SSRN 6893938.
13. Kage, P., Rothenberger, J., Andreadis, P., & Diochnos, D. (2026). A review of pseudo-labeling for computer vision. Journal of Artificial Intelligence Research, 85.
14. Du, Y., Dong, Y., Liu, Q., & Chen, Z. (2026). Modal Reliability Assessment and Drift Early Warning in Visible-Infrared Target Tracking. Available at SSRN 7132381.
15. Zhang, Z., Tong, Y., & Gao, Y. (2026). Retrieval-Augmented Generation with Low-Latency Deployment for Vertical Domains Question Answering: A Case Study on Economic Resource Platforms.
16. Bustamante, P. (2026). Bayesian Deep Learning for Photovoltaic Power Forecasting: Probabilistic Framework for Uncertainty Quantification and Grid Reliability Optimization (Doctoral dissertation, The University of Texas at El Paso).
17. Zhang, Y., & Wang, J. (2026, February). Design and Implementation of a Computer-Aided Full Lifecycle Quality Management System for Wind Farms in Upgrades, Renovations, and Subcontractor Supervision. In Proceedings of the 2026 International Conference on Big Data and Informatization Education (pp. 1309-1314).
18. Su, D., & Zhang, H. Developing a Data-Driven Computational Framework for Scalable Special Education Practices for Autism and Intellectual Disabilities in the US Public School

System.
19. Xiong, W., & Zeng, Y. (2026). Multi-Layer Coupled Diagnosis of Energy Efficiency Degradation in Complex Building MEP Systems and Identification of Optimization Boundaries. Available at SSRN 7121782.
20. Kull, K., Asad, B., Khan, M. A., Naseer, M. U., Kallaste, A., & Vaimann, T. (2025). Faults, failures, reliability, and predictive maintenance of grid-connected solar systems: A comprehensive review. Applied Sciences, 15(21), 11461.
21. Solís-Martín, D., Galán-Páez, J., & Borrego-Díaz, J. (2026). D3a-ts: Denoising-driven data augmentation in time series. International Journal of Data Science and Analytics, 22(1), 48.
22. Hong, Z., Chen, H., & Xu, T. (2026). Long-Horizon Load Stability Testing and Degradation Forecasting for High-Concurrency Cloud Storage. Available at SSRN 7115859.
23. Li, J., Zhang, Y., & Gu, Y. (2026). Chain-Hash Audit Framework with Trusted Anchoring for Tamper-Evident Evidence Lifecycle Management.
24. Ganguli, A., Ramachandra, N., Bessac, J., & Constantinescu, E. (2025). Enhancing interpretability in generative modeling: statistically disentangled latent spaces guided by generative factors in scientific datasets. Machine Learning, 114(9), 197.
25. Du, Y., Chen, Z., Liu, Q., & Dong, Y. (2026). Co-identification of Vibration, Structural Looseness, and Surface Damage in Industrial Equipment Operation Videos.
26. Xiong, W., Zeng, Y., & Guo, Y. (2026). A Study on the Characterization of Building MEP System States and the Learning of Behavioral Patterns Driven by Large-Scale Operational Data. Available at SSRN 7121883.
27. Guanghua, L., Shah, S. H. H., Zaidi, A. A., Nawaz, F., & Khan, S. Z. (2026). A Review of State-of-the-Art Developments in Large-Scale Photovoltaic Solar Renewable Energy Systems. Journal of Engineering, 2026(1), 4799038.
28. Fu, Y., Gui, H., Li, W., & Wang, Z. (2020, August). Virtual Material Modeling and Vibration Reduction Design of Electron Beam Imaging System. In 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 1063-1070). IEEE.
29. Yin, J., Huang, Y., & Rao, H. (2026). A Study on the Dynamic Evolution of Learning Behavior and Outcome Prediction in Digital Educational Environments. Available at SSRN 6607139.
30. Kijanowski, K., Barszcz, T., & Dao, P. B. (2025). A cluster-based filtering approach to SCADA data preprocessing for wind turbine condition monitoring and fault detection. Energies, 18(22), 5954.
31. Kull, K., Asad, B., Khan, M. A., Naseer, M. U., Kallaste, A., & Vaimann, T. (2025). Faults, failures, reliability, and predictive maintenance of grid-connected solar systems: A comprehensive review. Applied Sciences, 15(21), 11461.
32. Zhang, Z. (2026). A Study on Multimodal Product Understanding and Assembly Guidance Optimization in e-commerce for Complex Consumer Goods. Available at SSRN 6734559.