SOURCE REVIEW / PV PERFORMANCE DIAGNOSTICS

Physics-guided ML for PV performance deviation assessment

Residual-based diagnosis, probabilistic severity classification, and explainable machine learning using SCADA and a calibrated PVsyst baseline.

DOCUMENT	DATE	PREPARED BY	STATUS
TP-004	OCT 2026	SPECTRA GROUP S.A.S.	TECHNICAL SUMMARY

Technical scope

Mohajon et al. present a physics-guided diagnostic framework for grid-connected photovoltaic systems that combines a bias-corrected PVsyst baseline with SCADA measurements, residual analysis, probabilistic fault-severity classification, and explainable machine learning. The method is designed to detect when measured plant behavior departs from physically expected performance while reducing baseline bias, avoiding temporal leakage, and preserving interpretability of severity predictions.

01 Framework & diagnostics

Baseline. PVsyst expected generation is calibrated by linear bias correction before residual calculation, reducing systematic simulation error.

Residual & FSI. Normalized residuals quantify deviation from expected power; GMM clustering with BIC derives a data-driven severity threshold of Rnorm = -0.161.

Classification. A random forest uses seven physics-aware variables. Residual-derived variables are excluded from the main predictive feature set to avoid circular validation.

Explainability. SHAP/MDI are used to interpret feature influence, while KDE characterizes how performance-deviation probability changes with operating load.

02 Validation & applicability

Field data. Two grid-connected 56.32 kWp PV plants in Karnataka, India, are evaluated using 34 days of SCADA measurements.

Validation. Time-series walk-forward cross-validation is used rather than random train/test splitting, limiting temporal data leakage.

Performance. The random forest reaches 92.0% accuracy and 0.907 macro-F1, with benchmarking against XGBoost, SVM, and decision tree classifiers.

Scope boundary. The 34-day dataset and site-specific severity threshold constrain generalization; longer records with verified faults and maintenance history are required for broader deployment.

SOURCE Mohajon, R.; Sarker, S.C.; Abid, M.A.S.; Ronobir, S.D.; Rumi, A.K.; Chakraborty, S.; Juoairia, A.; Mohammad, N. Clean Energy 2026, 10(3), 143-163. Published 29 May 2026.
DOI: 10.1093/ce/zkag026 | https://academic.oup.com/ce/article/10/3/143/8698220

SOURCE TITLE A physics-guided explainable machine learning framework for residual-based performance deviation detection and probabilistic severity assessment in photovoltaic systems