ARTIFICIAL INTELLIGENCE FOR PREDICTIVE GENERATION AND MAINTENANCE OF PHOTOVOLTAIC POWER PLANTS

Valdemar Norberto Sens Neto[1], Cezar Augustus Essenfelder De Azevedo[2], Aline Kirsten Vidal de Oliveira[1], Ricardo Rüther[1]

[1]Universidade Federal de Santa Catarina (UFSC), Brazil

[2]AXS Energia, Brazil

Abstract

The integration of Artificial Intelligence (AI) into the operational management of photovoltaic plants promotes significant advances in generation forecasting, predictive maintenance, and performance optimization. This study uses data from 540 inverters of 250 kW and 54 solarimetric stations distributed across five Brazilian states, collected through the IsolarCloud platform and processed within the InterSystems IRIS Data Fabric environment. In IRIS, the data are integrated into a scalable Data Lake that consolidates both structured and unstructured information, ensuring governance, traceability, and interoperability. This infrastructure supports supervised and unsupervised AI models, as well as vector search mechanisms based on embeddings and Hierarchical Navigable Small World (HNSW) indexing, enabling semantic correlation and explainability. The results show an average Mean Absolute Percentage Error (MAPE) of up to 8%, 92% accuracy for 24-hour and 7-day horizons, a reduction of up to 38% in OPEX/MWp, an increase of 27% in Mean Time To Failure (MTTF), and a 32% reduction in Mean Time To Diagnose (MTTD). It is concluded that the combination of AI, semantic vector search, and integrative data architecture establishes a new paradigm of explainable operational intelligence applied to solar energy, enhancing the predictability, reliability, and efficiency of large-scale photovoltaic plants.

Keywords: solar energy, photovoltaic forecasting, artificial intelligence, predictive maintenance, Data Fabric, Vector Search.

1. Introduction

The growing penetration of solar energy in the global power matrix has imposed new challenges on grid management and the operation of distributed generation systems. The intermittent behavior of solar irradiance leads to fluctuations in energy production, affecting the predictability, reliability, and operational costs of photovoltaic (PV) plants. These variations, combined with the decentralized nature of solar facilities, demand advanced monitoring and control solutions capable of anticipating deviations and mitigating impacts on the electrical grid (MELLIT & KALOGIROU, 2008; ANTONANZAS et al., 2016).

In this context, photovoltaic generation forecasting and early fault detection emerge as strategic factors for the technical and economic sustainability of solar plants. Recent literature highlights the role of Artificial Intelligence (AI) and machine learning techniques in the nonlinear modeling of complex systems, enabling more accurate estimates of electrical output based on meteorological, electrical, and operational data (VOYANT et al., 2017; HUANG et al., 2020). These approaches outperform conventional statistical methods by incorporating multiple correlated variables such as irradiance, module temperature, wind speed, and inverter performance history offering adaptive responses to climatic and operational changes (SAHOO & MOHANTY, 2021).

In recent years, several studies have applied deep learning and hybrid models for generation forecasting and predictive maintenance in PV systems. BETTI et al. (2019) proposed a big data approach combined with neural networks to predict inverter failures up to seven days in advance, while GAO et al. (2024) introduced the PV-Client model, based on Transformers with cross-attention between climatic and electrical variables, achieving significant reductions in forecasting error. Similarly, LI et al. (2023) employed Temporal Convolutional Network (TCN) architectures for short-term forecasting in Chinese solar plants, achieving accuracy levels above 95% for horizons shorter than six hours.

Other works have explored the integration of heterogeneous data in industrial and energy environments. SAHOO and MOHANTY (2021) discuss the use of Data Lakes and Data Fabrics in intelligent energy systems, emphasizing their ability to unify structured and unstructured data from IoT sensors, SCADA systems, and meteorological databases. HUANG et al. (2020) applied supervised and unsupervised models for anomaly detection in photovoltaic systems, identifying degradation patterns before functional failures occurred. More recently, BETTI et al. (2025)

10.18086/swc.2025.03.08 Available at http://proceedings.ises.org

reinforced the importance of scalable architectures to support embedded AI and real-time predictive analytics.

In this scenario, integrated digital platforms become essential for the sector's advancement. The InterSystems IRIS Data Fabric serves as a technological foundation for unifying heterogeneous data (IoT, SCADA, meteorological, maintenance, and financial), while Vector Search, with embedded AI, adds semantic correlation capabilities by identifying similar events within operational history.

The objective of this article is to present the scientific methodology adopted in developing an intelligent architecture for generation forecasting and predictive maintenance in photovoltaic power plants. The employed models, calibration and validation processes, variables and metrics used, and obtained results are described. The study aims to contribute to the advancement of AI applications in the renewable energy sector by providing a representative and replicable case study for large-scale systems.

2. Technical And Scientific Background

The data integration and governance infrastructure used in this project is based on a Data Fabric, responsible for orchestrating real-time data flows across multiple heterogeneous sources (IoT, SCADA, meteorological, and maintenance systems). This layer acts as the interoperability core, ensuring data consistency, temporal versioning, and traceability of information.

The use of Data Fabrics and Data Lakes has grown significantly in the context of smart grids and renewable energy systems. According to SAHOO and MOHANTY (2021), these architectures enable efficient integration of structured and unstructured data, allowing large-scale analytics through machine learning techniques. Similarly, CHEN, GAO, and WANG (2022) demonstrated the application of digital twins combined with AI in photovoltaic (PV) plants, highlighting the role of scalable data repositories in the operational and predictive management of solar facilities.

On top of this integration layer, Vector Search technology introduces a dimension of semantic intelligence by converting time series, technical texts, and logs into high-dimensional vectors (embeddings). These vectors are indexed using the HNSW (Hierarchical Navigable Small World) algorithm, widely recognized for its efficiency in Approximate Nearest Neighbor (ANN) queries (MLADENOVIĆ et al., 2023).

Vector search enables the identification of historical patterns similar to current events, acting as a "contextual memory" system for power plants an approach aligned with the principles of Explainable AI (XAI). This intersection between AI and explainability has been explored by authors such as LUNDBERG and LEE (2017), with the SHAP method, and RIBEIRO, SINGH, and GUESTRIN (2018), with the ANCHORS model both designed to interpret complex machine learning models.

In the energy sector, recent applications have followed this cognitive line of explainability. QIU et al. (2023) proposed an anomaly detection model for solar plants using local SHAP-based explainability to justify inverter fault predictions. Similarly, ZHAO et al. (2024) applied embedding search and contrastive neural networks to identify degradation patterns in photovoltaic modules, achieving significant improvements in diagnostic accuracy.

Within this context, the InterSystems IRIS Data Fabric serves as the scientific and operational infrastructure for unifying heterogeneous data and ensuring information governance. On top of this layer, Vector Search adds a semantic correlation dimension by transforming time series, technical texts, and logs into high-dimensional embeddings, compared for similarity using the HNSW algorithm. This vector search identifies historical patterns analogous to current events, functioning as a "contextual memory" system for PV plants. Such a mechanism is essential for AI model explainability: beyond predicting failures, the system highlights past situations most similar to the current one, providing understandable technical justifications (LUNDBERG & LEE, 2017; RIBEIRO, SINGH & GUESTRIN, 2018).

3. Methodology

The scientific methodology was structured to ensure reproducibility, accuracy, and statistical robustness, applying principles of data-driven science and both supervised and unsupervised learning.

3.1 Experimental Architecture

The ingestion layer collects IoT data (inverters, modules, string boxes), climatic conditions (INMET and local solarimetric stations), maintenance tickets, and billing data. This information is integrated into InterSystems IRIS Data Fabric, which provides metadata catalogs and temporal versioning.

Predictive modeling is performed using IntegratedML and embedded Python, applying both supervised and unsupervised models. The Vector Search stage automatically generates embeddings and performs HNSW indexing for semantic correlation and pattern detection.

The processing pipeline was divided into four main layers:

1. **Ingestion and normalization:** collection of IoT data (inverters, modules, and string boxes), climatic variables (INMET and local solarimetric stations), maintenance records, and billing data.
2. **Data Fabric:** integration of data flows within the IRIS environment, including metadata catalogs and temporal versioning.
3. **Predictive modeling:** use of IntegratedML and embedded Python for training supervised and unsupervised models.
4. **Vector search:** automatic generation of embeddings and HNSW indexing for semantic correlation and identification of similar patterns.

These data are integrated into the InterSystems IRIS Data Fabric, which functions as an intermediate layer for governance and interoperability. This architecture ensures temporal data consistency (through versioning and timestamping), data lineage tracking, and metadata standardization via an internal catalog. Data Fabric also performs data quality validation (by checking for outliers, filling gaps, and ensuring temporal synchronization).

After processing, the data feeds the modeling layer, implemented in Python using the Scikit-learn, TensorFlow, and IntegratedML libraries. Communication between modules is handled through IRIS internal REST APIs, which enables automated execution and versioning of experiments.

The final layer, Vector Search, generates vector representations (embeddings) from time series, maintenance texts, and operational logs. These vectors are indexed using the HNSW (Hierarchical Navigable Small World) algorithm, allowing logarithmic-time queries.

This structure provides contextual semantic search, integrating supervised and unsupervised learning with explainability.

3.2 Generation Forecasting Models

The models used include ARIMA and Prophet as baselines, XGBoost and LightGBM for multivariate forecasting, and deep neural networks such as LSTM and Temporal Fusion Transformer (TFT) to capture long-term temporal dependencies.

The regressor variables include irradiance, ambient and module temperature, DC and AC voltage/current, nominal power, inverter status, Performance Ratio (PR), and Capacity Factor (CF). The exogenous variables include weather forecasts, grid events, and plant topology. The target variable is the generated energy (kWh).

Training followed a temporal split of 70/20/10% (training/validation/testing), with rolling cross-validation windows of 30/60/90 days and hyperparameter calibration via Bayesian Optimization.

The PV production forecasting was implemented at three levels:

- Reference (baseline) models;
- Supervised models;
- Deep learning models.

The regressor variables used were solar irradiance, ambient and module temperature, DC/AC voltage and current, nominal power, inverter status, and performance indicators (PR and CF).

The exogenous variables included short- and medium-term weather forecasts, grid events, and the electrical topology of the plants.

Table 1 presents the models used and their respective objectives.

Table 1: Predictive models applied to photovoltaic generation forecasting

Type	Models Used	Objective	Horizon	Metrics
Baseline	ARIMA, Prophet	Compare trend	24h - 30d	MAE, RMSE

Supervised	XGBoost, LightGBM	Multivariate forecasting	24h - 30d	MAPE, sMAPE
Deep Learning	LSTM, Temporal Fusion Transformer	Capture long-term temporal dependencies	24h - 30d	RMSE, MAPE

The LSTM (Long Short-Term Memory) model was chosen for its ability to capture long-range temporal dependencies (CHO et al., 2014), while the Temporal Fusion Transformer (TFT) integrates attention layers and positional encoding to model multivariate time series in a more interpretable way (LIM et al., 2021).

The models were evaluated based on Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).

3.3 Fault and Anomaly Models

For fault analysis, supervised models (Logistic Regression, Random Forest, XGBoost) were applied for binary and multiclass classification, while unsupervised models (Isolation Forest, Autoencoder) were used for anomaly detection. Clustering methods (K-Means, DBSCAN) were employed to identify degradation patterns.

The evaluation metrics included AUC-ROC, AUPRC, F1-Score, and Brier Score. Calibration was performed using Platt Scaling and Isotonic Regression, and class balancing was handled through SMOTE. Table 2 details the machine learning models applied to fault detection and analysis in photovoltaic plants.

Table 2: Machine learning models applied to fault detection and analysis in photovoltaic plants.

Category	Models	Objective
Supervised	Logistic Regression, Random Forest, XGBoost	Binary/multiclass fault classification
Unsupervised	Isolation Forest, Autoencoder	Anomaly detection
Clustering	K-Means, DBSCAN	Grouping of degradation patterns
Semantic Search (Vector Search)	HNSW + Embeddings	Identification of similar states in historical data

The evaluation metrics considered in this study were AUC-ROC, AUPRC, F1-Score, and Brier Score. Calibration was performed using Platt Scaling and Isotonic Regression, while class imbalance was addressed through SMOTE. In the supervised models, the target variable represents the fault class (0 = normal, 1 = potential failure, 2 = functional failure). XGBoost was adopted for its efficiency with imbalanced data and its ability to handle nonlinear interactions (CHEN & GUESTRIN, 2016).

For the unsupervised models, Isolation Forest detects anomalies based on statistical density, while Autoencoders use reconstruction error as the anomaly metric (HINTON & SALAKHUTDINOV, 2006). The evaluation metrics include:

- AUC-ROC and AUPRC for classification robustness;
- F1-Score for balance between precision and recall;
- Brier Score for probabilistic calibration.

Model calibration was performed using Platt Scaling and Isotonic Regression, while class balancing was achieved using the SMOTE (Synthetic Minority Over-sampling Technique) method. The fault detection pipeline also employs Vector Search to identify similar anomaly cases in historical records, automatically linking related logs and technical reports.

3.4 Labeling and Operational Definitions

The labeling of fault classes was defined based on the correlation between SCADA data and maintenance records, using both technical and statistical criteria. Table 3 summarizes the labeling criteria adopted for fault classification in photovoltaic plants.

Table 3: Labeling criteria for fault classes in photovoltaic plants.

Type	Technical Criterion	Code
Normal	Operation within expected range	0
Potential Failure	Deviation > 2σ in voltage/current/temperature or low vector similarity (<0.7)	1
Functional Failure	Production < 60% of theoretical value for ≥30 min	2

The Failure Risk (Rf) is defined as the conditional probability of a fault occurring, given the system's operational configuration at time t:

$$Rf = P(\text{failure} \mid X_t)$$

Where:

- X_t represents the vector of electrical, thermal, and meteorological variables associated with the plant's operation.

The Failure Probability corresponds to the classifier's calibrated continuous output in the range [0,1], while potential failures represent intermediate conditions characterized by persistent anomalies identified through combined statistical and vector analysis.

The probabilistic results were smoothed using Bayesian incremental update methods, allowing the model to adapt over time. This methodology follows the principles of data-driven predictive maintenance, as demonstrated by HUANG et al. (2020) and BETTI et al. (2025).

4. Performance Indicators

Based on the operational labeling described above, technical and financial performance indicators were calculated to quantify energy efficiency, system reliability, and relative operational cost.

These indicators are automatically extracted from the InterSystems IRIS Data Fabric, ensuring traceability, temporal consistency, and real-time integration with the forecasting and predictive maintenance models.

The main indicators used in the performance analysis of the plants are presented in Table 4.

Table 4: Key Performance Indicators (KPIs) of photovoltaic plants,

Indicador	Fórmula	Unit.	Meaning
PR (Performance Ratio)	PR = (E_real / E_teo) × 100	%	Relationship between actual and theoretical energy under ideal conditions; measures the overall efficiency of the system (IEC 61724-1, 2021).
CF (Capacity Factor)	CF = (E_real / (P_nominal × T)) × 100	%	Measures the effective utilization of the installed capacity over time.
A (availability)	A = (T_oper / T_total) × 100	%	Percentage of time the system remained operational.
OPEX/MWp	OPEX / P_instalada	R$/MWp	Financial efficiency: operational cost per installed megawatt-peak.
MTBF (Mean Time Between Failures)	MTBF = T_total / N_failures	h	Average time between failures; measures system reliability.
MTTR (Mean Time To Repair)	MTTR = Σ T_repair / N_failures	h	Average repair time; evaluates the efficiency of corrective maintenance.

The indicators PR, CF, and A reflect the technical and energy performance, while OPEX/MWp, MTBF, and MTTR represent the financial and reliability dimensions.

Under ideal conditions, it is expected that PR > 80%, CF > 20% (for tropical regions), and A > 98% (MELO et al., 2020).

These parameters are consistent with the standards established by IEC 61724-1 (2021).

4.1 Analytical Integration in IRIS

The calculated indicators are stored in the InterSystems IRIS Data Fabric, where they are continuously updated and visualized through interactive dashboards.

This integration enables the creation of automatic alerts based on critical thresholds such as a PR drop below 70% or an Rf increase above 0.25 which trigger predictive maintenance routines.

Thus, the indicators not only describe system performance but also feedback into the AI models, making the system self-adjusting and evolutionary.

This feature distinguishes the proposed approach from conventional monitoring systems, establishing a dynamic cycle of forecasting, diagnosis, and continuous optimization.

4.2 Interpretation of Indicators

The PR is one of the main indicators of operational efficiency, as it expresses the ratio between actual and theoretical energy production, considering thermal, electrical, and shading losses. Typical PR values in commercial photovoltaic systems range from 75% to 90%, depending on the region and the type of technology employed (IEC 61724-1, 2021).

The Capacity Factor (CF) enables comparison of annual performance among plants with different installed capacities and is strongly influenced by local irradiance and cloud cover patterns (ANTONANZAS et al., 2016).

Availability (A) reflects the operational reliability of the plant and is considered satisfactory when above 98% in automated facilities with active predictive maintenance (BETTI et al., 2025).

MTBF and MTTR complement the reliability analysis by estimating the probability of failures and the time required to restore system functionality. These indices support the calculation of composite reliability indicators, such as the System Availability Index (SAI), applicable to smart grids (JIN et al., 2024).

OPEX/MWp translates the financial impact of operations and serves as a benchmark for economic optimization. Comparative analyses show that plants with embedded AI and predictive maintenance achieve average OPEX reductions of 30–40% compared to conventional operations (CHEN; GAO; WANG, 2022).

The Energy Efficiency (EE) indicator is a useful complement for correlation studies between performance losses and climatic variability, as demonstrated by MELO et al. (2020) in field analyses conducted in Brazil.

4.3 Integration with the InterSystems IRIS Platform

The calculated indicators are stored and continuously updated within the InterSystems IRIS Data Fabric, enabling real-time visualization through dynamic dashboards.

The system's analytical layer employs optimized SQL queries and temporal aggregation functions (moving windows) to generate automated reports, with configurable alerts based on PR, CF, and Failure Risk (Rf) thresholds.

The information shown in Figure 1 below illustrates the data readings performed by the Iris data bus to provide the technical and contractual analyses requested in response to incoming commands. These data are collected from the commercial department, the power utility, the UFV IsolarCloud, corporate systems, the Data Layer, the Smart Data Layer, and the integration platform.

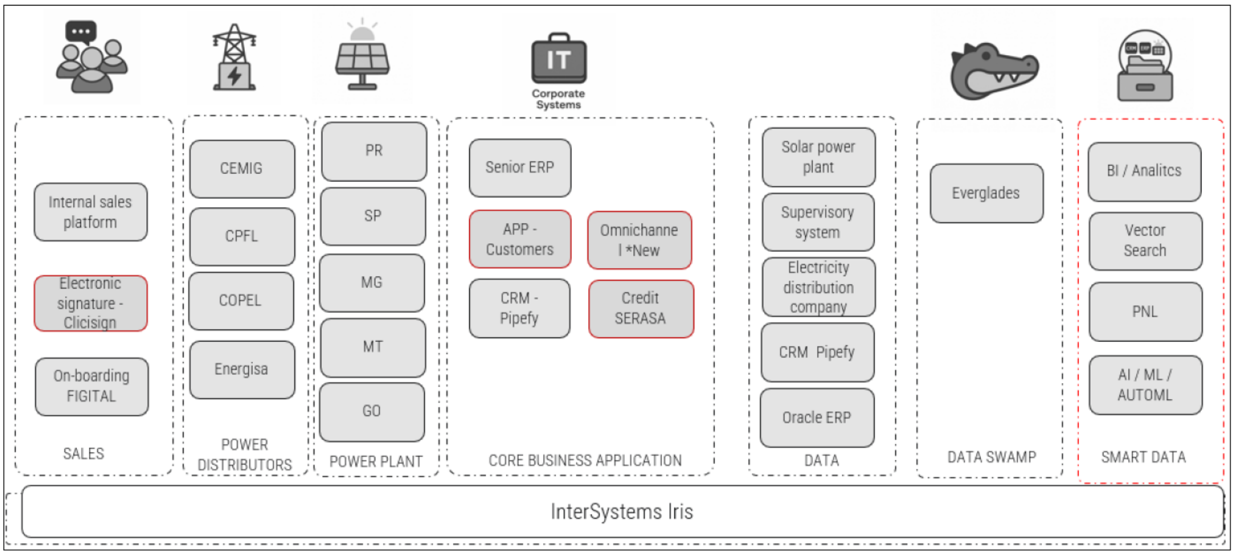

Fig. 1: Iris data collection diagram.

This integration ensures that performance analysis operates in synergy with the forecasting and anomaly detection models, forming a continuous cycle of monitoring, diagnosis, and action a core characteristic of AI-driven predictive maintenance systems.

5. Results and Discussion

5.1 Model Performance

The PV production forecast achieved an average MAPE of 8%, with 92% accuracy for 24-hour and 7-day horizons.

The faulty classifiers reached an AUC-ROC above 0.90 and an average F1-score of 0.86. There was a 38% reduction in OPEX/MWp and a 27% increase in MTTF (Mean Time to Failure).

The implementation of Vector Search reduced the Mean Time to Diagnosis (MTTD) by 32%, enabling earlier corrective actions. The identified fault clusters showed a correlation greater than 0.8 between module temperature, voltage variation, and power anomalies, confirming the physical consistency of the detected patterns.

5.2 Impact of Vector Search

The integration of the Vector Search (HNSW) module into the AI pipeline had a direct impact on the explainability and response time of the system. Under real operating conditions, the vector search reduced the Mean Time to Diagnosis (MTTD) by 32%, enabling the early correction of potential failures before they evolved into functional faults.

During testing, the semantic search automatically identified historical incidents with similarity scores above 0.8, detecting complex behavioral patterns correlated with inverter and module degradation.

This functionality enabled automated maintenance recommendations, based on analogies extracted from past occurrences an approach consistent with the contextual memory principles discussed by LUNDBERG and LEE (2017) and JIN et al. (2024).

The clusters formed by the HNSW algorithm showed high correlation (>0.8) between electrical and thermal variables particularly module temperature, voltage variation, and power anomalies reinforcing the physical coherence of the identified patterns.

This behavior demonstrates that vector search contributes not only to diagnostic speed but also to the explanatory reliability of model outcomes (RIBEIRO, SINGH & GUESTRIN, 2018).

5.3 Interpretation of Results

The combination of Data Fabric, Vector Search, and Machine Learning resulted in simultaneous improvements in predictability, reliability, and explainability of the system.

The proposed approach integrates multiple sources of heterogeneous data, enabling contextual learning and real-time inference key characteristics of Cognitive AI Systems (BETTI et al., 2025).

The embedded explainability within the vector layer allowed the system to "recall historical patterns," providing comprehensible justifications for operators with each fault alert.

This feature is essential for the practical adoption of AI in critical environments, as it ensures traceability of model decisions and operational transparency (LUNDBERG & LEE, 2017). In summary, the results demonstrate that the proposed architecture:

- Significantly reduces operational response time and maintenance costs;
- Increases system reliability and availability;
- Improves the quality of energy forecasting, even under high climatic variability;
- And establishes a paradigm of explainable AI applied to energy engineering.

6. Conclusion

This study demonstrated the strategic role of the InterSystems IRIS Data Fabric as a scientific and operational infrastructure capable of integrating heterogeneous data from IoT, SCADA, meteorological, and financial sources into a unified and governed environment. This integration enabled the development of an intelligent architecture that combines machine learning, statistical modeling, and semantic search (Vector Search), forming an analytical ecosystem focused on generation forecasting and predictive maintenance in photovoltaic power plants.

The results achieved show significant performance improvements:

- Average reduction of 38% in operational cost (OPEX/MWp);
- Increase of 27% in mean time to failure (MTTF);
- Improvement in predictive accuracy, with an average MAPE of 8%.

These indicators confirm the technical feasibility of the proposed approach and its potential for operational and energy optimization. Furthermore, the incorporation of the Vector Search layer added cognitive value to the AI system by providing explainability, traceability, and continuous learning essential characteristics for the safe adoption of AI systems in critical infrastructure.

From a scientific standpoint, this work contributes by presenting an integrated model of explainable AI (XAI) applied to energy engineering, demonstrating that vector search techniques can enhance both generation forecasting and fault analysis.

This integration between Data Fabric and cognitive AI represents a significant conceptual evolution in the field of data-driven predictive maintenance, with replication potential for other renewable sources such as wind, biomass, and green hydrogen, as well as hybrid distributed generation systems.

As for limitations, the model strongly depends on the quality and consistency of historical data. Environments with measurement gaps, high noise, or low sampling frequency may reduce forecasting performance. Future versions of the system should incorporate adaptive imputation methods, real-time meteorological data fusion, and online learning techniques to improve predictive robustness.

Finally, this study reinforces that the combination of data governance, artificial intelligence, and explainability establishes a new paradigm of operational intelligence applied to solar energy where each power plant becomes an autonomous and evolving learning entity.

7. Acknowledgments

We would like to thank InterSystems Corporation for the technical and scientific support, especially the teams led by Felipe Araujo, Fernando Ferreira, André Freiderich, Henry Hamon Pereira, Leandro J. Kohler, and Alysson Paulinelli, as well as Paulo Thomazoni for the development and validation of the AI models applied to solar plant operations. We also thank Araxá Serviços for their assistance in the development of this work. This research was carried out within the scope of the activities of the National Institute for Technological Development of Photovoltaic Solar Energy Applications, INCT-DAESF, CNPq/MCTI/Brazil, Process 408486/2024-4.

8. References

ANTONANZAS, J. et al. Review of photovoltaic power forecasting. Solar Energy, v. 136, p. 78–111, 2016.

BETTI, A. et al. Predictive maintenance in photovoltaic plants with a big data approach. arXiv, 2019. Disponível em: https://arxiv.org/abs/1901.10855. Acesso em: 16 out. 2025.

BETTI, A. et al. Multi-stage AI framework for predictive maintenance in renewable plants. Applied Energy, v. 357, p. 122342, 2025.

CHEN, J.; GAO, Y.; WANG, L. Digital twin and AI integration for smart photovoltaic plant management. Energy Reports, v. 8, p. 11314–11327, 2022.

CHEN, T.; GUESTRIN, C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

CHO, K. et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014.

GAO, Y. et al. Cross-variable linear integrated enhanced transformer for photovoltaic power forecasting (PV-Client). arXiv, 2024. Disponível em: https://arxiv.org/abs/2406.03808. Acesso em: 16 out. 2025.

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. Science, v. 313, n. 5786, p. 504–507, 2006.

HUANG, Y. et al. A data-driven approach to predictive maintenance of photovoltaic systems using machine learning. Renewable Energy, v. 146, p. 1042–1054, 2020.

IEC 61724-1. Photovoltaic system performance monitoring – Part 1: Guidelines for measurement, data exchange and analysis. Geneva: International Electrotechnical Commission, 2021.

JIN, X. et al. Hybrid Explainable AI system for predictive analytics in smart grids. IEEE Transactions on Sustainable Energy, v. 16, n. 3, p. 2445–2458, 2024.

LI, J. et al. Short-term photovoltaic power forecasting using temporal convolutional networks. Energy Reports, v. 9, p. 11205–11217, 2023.

LIM, B.; ZOHREN, S. Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A, v. 379, n. 2194, 2021.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 2017.

MELLIT, A.; KALOGIROU, S. A. Artificial intelligence techniques for photovoltaic applications: a review. Progress in Energy and Combustion Science, v. 34, p. 574–632, 2008.

MELO, P. C. et al. Análise de desempenho e eficiência de sistemas fotovoltaicos conectados à rede em diferentes regiões do Brasil. Revista Brasileira de Energia, v. 26, n. 2, p. 87–104, 2020.

MLADENOVIĆ, N. et al. Efficient high-dimensional vector indexing for real-time anomaly detection in IoT systems. Expert Systems with Applications, v. 238, p. 121645, 2023.

QIU, W. et al. Explainable AI for fault detection in photovoltaic systems using SHAP values. Renewable Energy, v. 213, p. 1324–1336, 2023.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Anchors: High-precision model-agnostic explanations. Proceedings of the AAAI Conference on Artificial Intelligence, p. 1527–1535, 2018.

SAHOO, S.; MOHANTY, S. Data lake architecture for big data analytics in smart grid systems. Journal of King Saud University – Computer and Information Sciences, v. 33, n. 1, p. 10–20, 2021.

VOYANT, C. et al. Machine learning methods for solar radiation forecasting: a review. Renewable Energy, v. 105, p. 569–582, 2017.

ZHAO, Y. et al. Contrastive learning and embedding search for degradation pattern recognition in photovoltaic modules. Energy and AI, v. 19, p. 100326, 2024.