Detecting Crossover Anomalies in Renewable-Powered PEM Electrolyzers Using Electrical Sensor Analytics

Leonardo Andrade Santos[a], Caio Souto Maior[b], and Márcio das Chagas Moura[c]

[a] Federal University of Pernambuco, Recife, Brazil, leonardo.las@ufpe.br

[b] Federal University of Pernambuco, Recife, Brazil, caio.maior@ufpe.br

[c] Federal University of Pernambuco, Recife, Brazil, marcio.cmoura@ufpe.br

---

Abstract: Green hydrogen production via PEM electrolysis is a cornerstone of the energy transition; however, its integration with highly dynamic renewable sources imposes severe challenges to the system's structural integrity. Extreme load variability makes it difficult to distinguish between normal transient responses and early indicators of catastrophic failures, such as gas crossover or external leakages. This work proposes a monitoring methodology based on "soft sensors" to enhance operational reliability. Utilizing the large-scale dataset from the National Laboratory of the Rockies (NLR), supervised machine learning models were applied to predict hydrogen mass flow in real-time, conditioned on process variables and oscillating input power. Residual analysis between the model's prediction and the physical Coriolis meter measurement allows for the identification of performance deviations and sensor failures. The results demonstrate that this data-driven monitoring layer enables a more dynamic Probabilistic Safety Assessment (PSA), reducing false alarms caused by intermittency and providing quantitative metrics for risk management and predictive maintenance of green hydrogen (GH2) systems.

---

1. INTRODUCTION

The global urgency to decarbonize hard-to-abate industrial sectors has elevated green hydrogen — produced via water electrolysis powered exclusively by renewable energy sources — as a strategic pillar of long-term net-zero decarbonization roadmaps [1]. Among available electrolysis technologies, Proton Exchange Membrane (PEM) water electrolysis has emerged as the most technically mature platform for coupling with variable renewable inputs, owing to its high current density capability, rapid dynamic response, and compact form factor [2].

Ocean wave energy represents one of the most energy-dense yet underexplored renewable resources. Its direct integration with PEM electrolyzers introduces highly transient power profiles — characterized by rapid current and voltage excursions driven by the stochastic nature of wave hydrodynamics — that significantly exceed the operational stress imposed by conventional grid-connected generation [3]. These intermittency-driven loading cycles accelerate a particularly critical membrane degradation mechanism: hydrogen gas crossover, defined as the diffusive permeation of molecular H2 through the polymer electrolyte membrane from the cathode into the anode compartment [4]. When the anode-side hydrogen concentration surpasses the 2% Lower Flammability Limit (LFL) threshold, the risk of ignition constitutes an unacceptable safety hazard that demands immediate protective intervention [4].

Conventional crossover monitoring relies on downstream gas analyzers positioned at the anode outlet, which inherently introduce a measurement latency that may render corrective interventions ineffective under fast-transient operating regimes. Furthermore, the predictive exploitation of upstream electrical signatures — wave input power, stack terminal voltage, and generated current — as early precursors of crossover onset has received limited systematic investigation in the context of wave-coupled PEM systems, representing a critical gap in the probabilistic health monitoring literature.

This work directly addresses that gap by proposing a non-invasive, machine learning-based predictive framework that detects the operational transition from stable conditions to hydrogen crossover risk before the critical chemical threshold physically manifests at the downstream analyzer. Two ensemble classifiers are implemented and benchmarked: Random Forest (RF) [5] and Extreme Gradient Boosting (XGBoost) [6]. An ablation study systematically evaluates three distinct feature engineering strategies derived exclusively from upstream electrical telemetry: pure rolling statistical windows, pure multi-scale transient gradients, and a hybrid combination of both. The experimental foundation is provided by the dynamic PEM electrolyzer dataset generated by the National Renewable Energy Laboratory (NREL) under a wave-energy-coupled operational scenario [3].

2. DATASET DESCRIPTION

The experimental data utilized in this study was developed by the National Renewable Energy Laboratory (NREL) in collaboration with the U.S. Department of Energy (DOE). The primary objective of this dataset is to evaluate the operational behavior and degradation dynamics of a commercial Proton Exchange Membrane (PEM) electrolyzer when coupled to an intermittent ocean wave energy source. The hydrodynamic power profiles were generated using a dedicated wave-to-wire simulation framework developed by Lettenmaier et al. [2022].

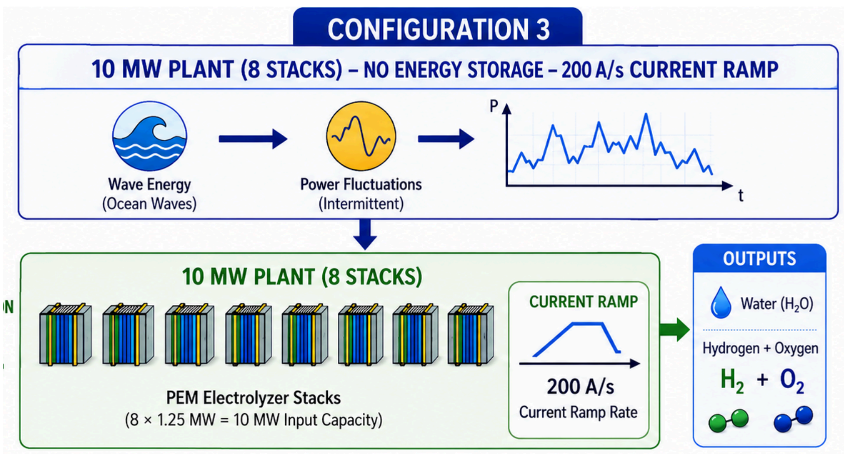
The original NREL/DOE testing campaign comprises eight distinct plant configurations, which vary based on:

- The presence or absence of an integrated energy storage buffer (accumulator);
- The maximum current ramp rate limits (configured at either 200 A/s or 400 A/s);
- The number of active electrolyzer stacks coupled to the system (either 4 or 8 stacks, each rated at a nominal power capacity of 1.25 MW).

Under the intermittent regime, each of these structural configurations was subjected to highly dynamic wave-induced power profiles for a continuous duration of 25 minutes. Additionally, the campaign included a steady-state baseline regime where the power source was held constant; during this baseline test, the operational current started at 3000 A and was stepped down by 300 A every 30 minutes over a total duration of 3 hours, concluding at 300 A.

To ensure a rigorous evaluation of the gas crossover phenomenon under highly transient stress, this work explicitly focuses on the configuration featuring 8 stacks, no energy storage buffer, and a current ramp rate of 200 A/s (noacc_4_200). This specific operational profile was selected because its data distribution offers the most comprehensive and unbuffered representation of transience-induced degradation, which is highly conducive to tracking crossover anomalies. Figure 1 illustrates the schematic architecture of the specific system configuration chosen for this investigation.

Figure 1: H2 Plant Configuration



The resulting dataset is composed of continuous multivariate time-series measurements from a comprehensive sensor manifold, capturing electrical, fluidic, and thermal parameters across the cell stack, the deionized water supply loop, and the auxiliary Balance of Plant (BoP) subsystems. In total, 29 raw features were evaluated prior to the feature engineering phase. A detailed breakdown and classification of these physical variables are summarized in Table 1.

Table 1: Sensor Categories

Category	Variables Included	Measurement Purpose
Electrical Power	Current, Voltage, Power	Monitoring of real-time power consumption and stack electrical response.
Production	H2 Flow Rate (Measured/Calc)	Quantification of hydrogen mass yield and system efficiency.
Pressure (PT)	Stack, Cooling, and Delivery	Monitoring of hydraulic balance and gas delivery conditions.
Temperature (TE/TT)	Stack, Water, and Dryers	Thermal management and monitoring of the gas drying process.
Fluid Management	Water Level & Resistivity	Control of water replenishment and monitoring of deionized water quality.
Gas Quality	Dew Point & H2/O2 Crossover	Assessment of hydrogen purity and membrane safety/integrity.

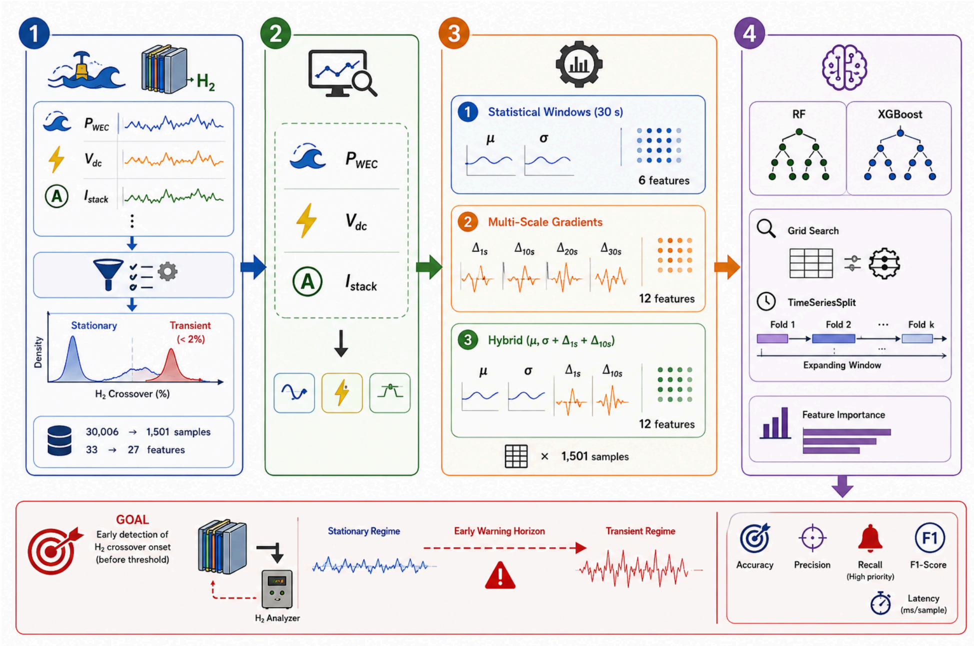
3. METHODOLOGY

The predictive framework developed in this study aims to detect the operational transition of a Proton Exchange Membrane (PEM) electrolyzer from a stable, stationary regime to a volatile transient state characterized by hydrogen crossover. This is achieved by non-invasively monitoring upstream electrical signatures directly tied to green hydrogen production: renewable input power (derived from Wave Energy Converters - WECs), the actual stack voltage (Vdc), and the resulting generated current (A). Furthermore, to systematically isolate and quantify the physical impact of intermittency-driven rates of change, a comprehensive ablation study was executed. The core machine learning architecture was evaluated across three distinct feature spaces:

- Purely Transient Features: Utilizing strictly the multi-scale derivatives of the raw electrical signals.
- Purely Statistical Windows: Utilizing rolling averages and standard deviations to capture historical operational stress.
- A Hybrid Feature Space: Combining statistical windows and transient gradients (1s and 10s) into a unified input matrix.

The ultimate objective of this methodology is to model, measure, and predict the onset of H2 permeation into the anode chamber, establishing an early-warning horizon before the dangerous chemical threshold physically manifests at the downstream gas analyzer. As illustrated in Figure 2, the proposed methodological pipeline is structured into four sequential phases: (i) Data pre-processing and conditioning, (ii) temporal feature engineering, (iii) rolling statistical windowing and multi-scale gradients, and (iv) machine learning modeling and benchmarking.

Figure 2: Methodology Summary

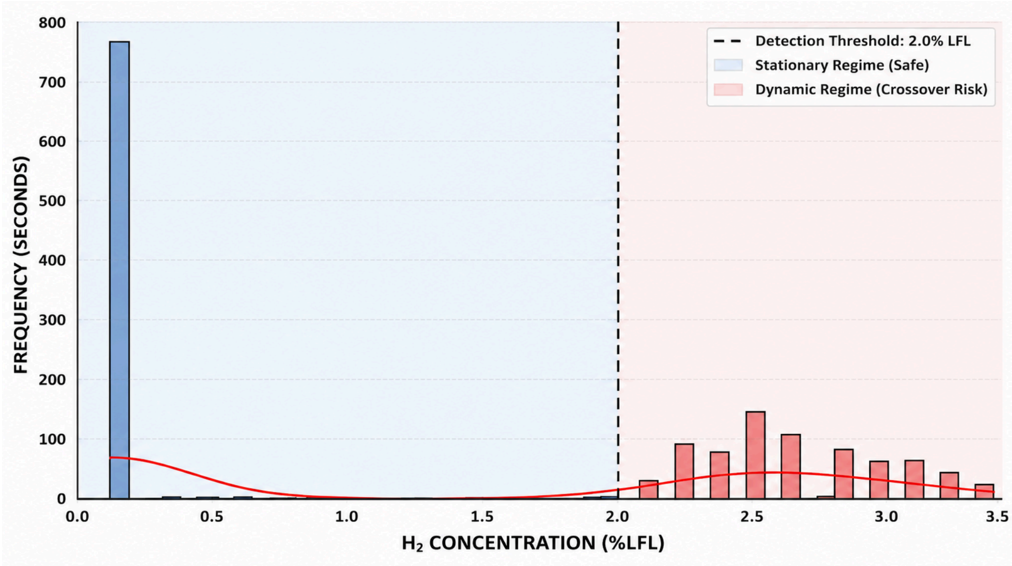


3.1. Data pre-processing and conditioning

In this phase, raw datasets were intentionally utilized without implementing digital noise filters. This approach ensures that the machine learning models learn from real-world noise profiles and operational disturbances, thereby evaluating whether high-frequency peaks and valleys in the sensor signals possess a causal relationship with the hydrogen crossover mechanism. The conditioning stage was strictly dedicated to handling missing data (NaNs), detecting frozen sensor telemetry, and isolating irrelevant variables.

Upon analyzing the crossover target variable, a distinct and peculiar distribution behavior was observed. The data revealed a clear bimodal distribution divided into two primary groups: a concentrated near-zero baseline cluster and a dispersed dynamic group with values fluctuating above 2%. A representative example of this phenomenon is illustrated in Figure 3, which displays the crossover density distribution within the selected intermittent configuration (Noacc_8_200). This specific configuration represents an un-buffered system (operating without an electrical battery accumulator) equipped with an 8-cell stack profile (equivalent to a 1.25 MW scaling) and subjected to a maximum ramp rate of 200A/s.

Figure 3: H2 Crossover Density Distribution (%LFL)



Based on these physical distribution patterns, the observations were categorized into either a "stationary regime" or a "transient regime". This scenario was selected for model development due to its highly representative and balanced distribution of hydrogen crossover events. Through this targeted filtering process, the global dataset was scaled down from its initial 30,006 observations across 33 columns to a focused subset of 1,501 rows. No frozen sensor readings were detected within this target subset, maintaining a clean sample size of 1,501 instances. Regarding the feature space, two columns were completely removed due to a critical lack of data density (excessive missing values), while others were excluded because they exerted a statistically negligible impact on the phenomenon under analysis. This reduction successfully narrowed the dataset down to 27 candidate features for the subsequent methodological stages.

3.2. Temporal Feature Engineering

The selection of predictive features was strategically designed to enable the machine learning algorithms to map the dynamic behavior of electrical intermittency curves and directly capture their impact on accelerated hydrogen crossover rates. Consequently, out of the 27 available candidate features, only three primary upstream variables were selected: the raw wave power input, the electrolyzer stack terminal voltage, and the resulting generated current. These variables directly capture the high-frequency physical disruptions introduced by wave energy before they propagate through the system.

3.3. Rolling statistical windowing and Multi-Scale Gradients

To systematically analyze the temporal impact of these electrical fluctuations, three distinct feature space inputs were engineered to feed the predictive models, establishing an ablation study framework:

- Input Scenario 1 (Pure Statistical Windows): Designed to evaluate the impact of sustained energetic baselines. This matrix utilizes a 30s rolling window to calculate the moving average and moving standard deviation for each of the three primary sensors, resulting in a matrix of 1,501 observations and 6 features.
- Input Scenario 2 (Pure Multi-Scale Gradients): Designed to capture the absolute velocity of the electrical shocks across different time horizons. This matrix computes the transient gradients across four scales: 1s (instantaneous derivative), as well as 10s, 20s, and 30s moving averages of the gradients. This configuration yields a matrix of 1,501 observations and 12 features.
- Input Scenario 3 (Hybrid Feature Space): Designed to provide both structural context and instantaneous reflex by combining two gradient scales 1s and 10s with the 30s rolling

statistics (averages and standard deviations). This configuration maintains the same 1,501 observations across 12 highly uncoupled features.

3.4. Machine Learning Modeling and Benchmarking

The selection of the machine learning algorithms was fundamentally driven by the requirement for high predictive sensitivity (Recall) paired with structural model interpretability. Given the safety-critical nature of green hydrogen production and the explosive hazards associated with membrane degradation, prioritizing high recall is paramount; the operational and safety costs of a false negative—failing to detect an imminent hydrogen crossover event—drastically outweigh the minor operational inconvenience of a false positive. Within this context, two ensemble-based algorithms were selected to establish a robust comparative benchmarking framework: Random Forest (RF) and Extreme Gradient Boosting (XGBoost).

The Random Forest architecture is a bagging-based ensemble method that constructs a multitude of parallel decision trees during the training phase. Each individual tree is trained on a distinct bootstrap sample derived randomly from the engineered training space, and the final classification output is obtained through a majority voting mechanism across all decision trees. This approach significantly reduces model variance and enhances robustness against localized noise.

The implementation was executed utilizing the scikit-learn framework, with the hyperparameter optimization specifically targeting the number of decision trees (n_estimators), the maximum tree depth (max_depth), and the minimum sample size required to form a terminal leaf node (min_samples_leaf).

In contrast, Extreme Gradient Boosting represents a highly efficient and scalable implementation of the gradient boosting paradigm. Instead of building trees in parallel, XGBoost constructs decision trees sequentially, where each consecutive tree is explicitly trained to correct the residual errors committed by the preceding trees through the optimization of a specialized objective loss function. Furthermore, the algorithm incorporates intrinsic L1 (Lasso) and L2 (Ridge) regularization terms directly into its objective function to penalize model complexity and mitigate overfitting.

Modeling was performed via the xgboost open-source package, systematically tuning the learning rate (learning_rate), the total number of boosting rounds (n_estimators), and the maximum depth of the sequential estimators (max_depth).

Unlike completely uninterpretable black-box deep learning architectures, both chosen ensemble models offer innate transparency through the calculation of feature importance scores. This analytical capability is vital to the core objectives of this research, as it shifts the model's value from a purely reactive warning system to an explainable tool capable of identifying the underlying physical and electrical drivers behind membrane degradation.

To ensure optimal hyperparameter selection and systematically evaluate model performance across the three input scenarios described in the ablation study, an exhaustive Grid Search strategy was implemented. This optimization pipeline systematically evaluated all possible hyperparameter combinations to isolate the exact configuration that maximizes the Recall metric on the validation sets. Crucially, because the operational telemetry of the PEM electrolyzer exhibits a strict sequential time-series nature, standard randomized K-Fold cross-validation protocols were mathematically discarded. Applying a randomized split would breach chronological integrity and induce data leakage, allowing future information to artificially contaminate past training steps.

To circumvent this vulnerability, a rigorous temporal cross-validation protocol was enforced using the TimeSeriesSplit method from the scikit-learn package. This technique employs an expanding training window that progresses chronologically through successive folds, ensuring that the evaluation sets always strictly succeed the training vectors in time. This methodology guarantees that the fine-tuned hyperparameter profiles yield models with high generalization capacity, fully shielded against temporal overfitting and data leakage.

Finally, a comprehensive performance benchmarking phase was executed using the isolated testing partition. The models were evaluated across five standard classification and computational dimensions: Accuracy, Precision, Recall, F1-Score, and absolute inference execution latency measured in milliseconds per sample. This multidimensional benchmarking allows for a rigorous comparison, evaluating whether the addition of multi-scale gradients and statistical windows under different feature spaces enhances the predictive fidelity of the models while remaining computationally efficient enough for real-time edge deployment.

4. RESULTS AND DISCUSSION

The calibration of hyperparameters via structured Grid Search with TimeSeriesSplit aimed not only to maximize accuracy and recall metrics but also to adapt the topology of tree-based models to the mathematical properties of each feature space. The optimal parameters selected for the Random Forest and XGBoost models reveal distinct learning behaviors according to the nature of the information provided, as summarized in Table 2.

Table 2: Hiperparameters results

Input	**Model**	**Hiperparameters**	**Values**
Pure Statistical Windows	Random Forest	n_estimators	100
		max_depth	15
		min_samples_leaf	5
	XGboost	n_estimators	100
		max_depth	8
		learning rate	0.1
Pure Multi-Scale Gradients	Random Forest	n_estimators	200
		max_depth	8
		min_samples_leaf	5
	XGboost	n_estimators	100
		max_depth	4
		learning rate	0.05
Hybrid Feature Space	Random Forest	n_estimators	200
		max_depth	15
		min_samples_leaf	5
	XGboost	n_estimators	200
		max_depth	6
		learning rate	0.1

In the Pure Statistical Windows scenario, both algorithms converged toward deeper tree structures (max_depth of 15 for Random Forest and 8 for XGBoost). Because moving averages and rolling standard deviations smooth out the high-frequency noise induced by wave intermittency, the models required greater depth to segment the subtle interactions and steady-state thresholds of the upstream variables that precede the degradation or crossover events. The min_samples_leaf = 5 parameter acted consistently as a regularization agent, preventing overfitting at the terminal nodes of these deeper trees.

Conversely, introducing the Pure Multi-Scale Gradients space radically altered the learning strategy. The models converged to significantly shallower topologies (max_depth reduced to 8 for Random Forest and 4 for XGBoost), accompanied by a more conservative learning rate in XGBoost (learning rate = 0.05). Physically, gradients and temporal derivative rates are highly volatile predictors that capture the transient kinematics of the system. Deeper trees in this domain would prone the models to memorize high-frequency fluctuations and power command noises. Imposing shallower structures forced the algorithms to operate as mathematical filters, extracting only the macro acceleration trends of the upstream process.

Finally, in the Hybrid Feature Space, the algorithms sought a structural equilibrium to reconcile both static features (window levels) and dynamic features (gradient derivatives). Both models expanded their ensemble size to n_estimators = 200, indicating the need for greater diversity to accurately map the higher-dimensional space. XGBoost settled its depth at an intermediate level (max_depth = 6), demonstrating that the synergy between statistical and kinematic components allows for a more balanced convergence, eliminating the need for extreme depths while maintaining robust generalization capabilities for real-time inference.

To evaluate the diagnostic capacity and real-time viability of the trained models, their predictive performance and inference costs were benchmarked across all three feature spaces. Table 3 summarizes the global metrics alongside the Relative Prediction Time (RPT) and absolute test execution times.

Table 3: Final Results

Input	Model	Accuracy	Precision	Recall	F1-Score	RPT (ms/sampling)	Test Time (s)
Pure Statistical Windows	**Random Forest**	86.03%	85.84%	85.45%	85.65%	0,0598	59,15
	XGBoost	**88.91%**	88.64%	**88.64%**	**88.64%**	**0,0107**	12,32
Pure Multi-Scale Gradients	**Random Forest**	58.98%	56.55%	68.64%	62.01%	0,1212	22,66
	XGBoost	59.20%	56.87%	67.73%	61.83%	0,0461	**9,37**
Hybrid Feature Space	**Random Forest**	86.70%	88.46%	83.64%	85.98%	0,0662	58,9
	XGBoost	87.80%	**90.24%**	84.09%	87.06%	0,0243	20,53

The Pure Statistical Windows space yielded the highest overall diagnostic performance, with XGBoost achieving the absolute peak Accuracy (88.91%), Recall (88.64%), and F1-Score (88.64%). Moving averages and rolling standard deviations act as robust low-pass filters that preserve macro-level operational regimes. Conversely, performance sharply deteriorated within the Pure Multi-Scale Gradients space, with accuracy dropping to approximately 59% for both algorithms. This severe drop confirms that while temporal derivatives (rates of change) are crucial for spotting localized transience and command fluctuations, they lack the structural context of absolute magnitudes. A rapid change at a low operating power profile can exhibit an identical gradient to a change at peak capacity, confusing tree-based splits. Interestingly, this space yielded the absolute lowest total training and validation test time for XGBoost (9.37 seconds). This rapid execution is a direct consequence of the highly regularized, shallow topologies (max_depth = 4) forced during optimization to avoid memorizing high-frequency noise.

The fusion of both domains into the Hybrid Feature Space unlocked a powerful synergetic effect, driving XGBoost to its maximum Precision of 90.24%. From an industrial health monitoring perspective, this is a vital outcome. High precision indicates a low false-alarm rate; when the model flags a system state change, it does so with mathematical confidence.

The inclusion of multi-scale gradients served as a verification layer for the statistical windows. The model learned to only trigger a true state change if the moving baseline shift was accompanied by a corresponding kinematic acceleration trend. This stricter decision criteria caused a slight trade-off in recall (84.09%), but provided a more dependable tool for field deployment where false alarms lead to costly, unnecessary maintenance downtime.

Regarding computational overhead, XGBoost consistently outperformed Random Forest across all input spaces, demonstrating an inference latency that is roughly 4x to 5x faster. Within the Pure Statistical Windows, XGBoost recorded an outstanding RPT of 0.0107 ms per sample.

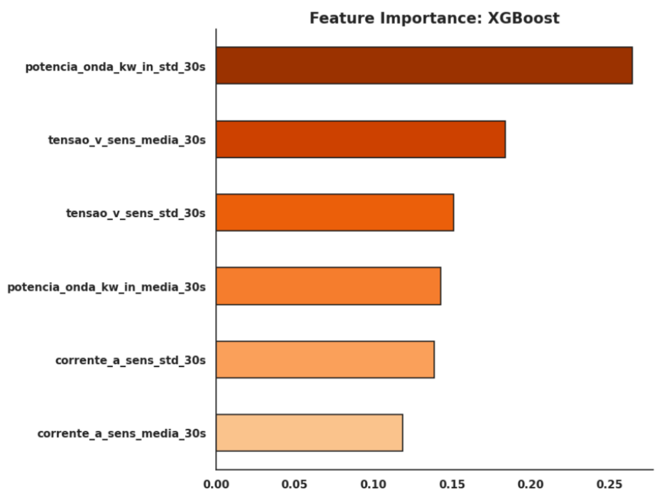
Random Forest suffers from a heavier computational footprint because it evaluates large comitês of independent, deeply grown trees (max_depth = 15) in parallel, requiring extensive memory lookups. XGBoost's sequential boosting process, combined with its optimized gradient tree-splitting mechanism, generates ultra-lightweight decision logic. An RPT of such low magnitude establishes XGBoost as a prime candidate for edge computing deployment, capable of executing real-time, sub-millisecond anomaly diagnostics directly on localized microcontrollers attached to the upstream sensor manifold.

4.3. Feature Importance and Physical Interpretation

To uncover the underlying decision-making mechanisms of the top-performing XGBoost model, the relative feature importance was evaluated across the three distinct input spaces. Analyzing the weight assigned to each predictor clarifies how the model translates upstream power, voltage, and current signatures into early indicators of system health changes.

When the model is restricted exclusively to steady-state rolling metrics (Figure 4), a single variable emerges as the dominant predictive driver: potencia_onda_kw_in_std_30s (accounting for over 26% of total importance).

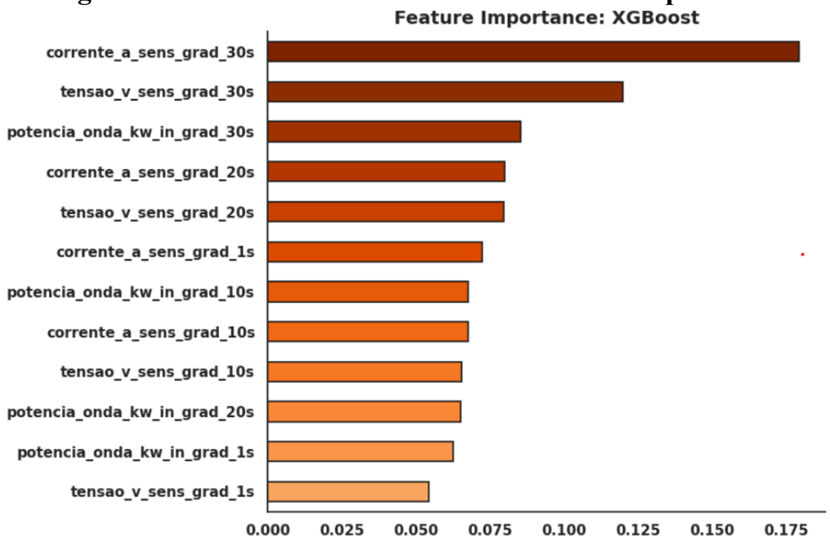
Figure 4: Pure Statistical Windows Feature Importance



From a physical standpoint, the 30-second standard deviation of the incoming wave power profiles directly quantifies operational instability. The model relies heavily on this metric, alongside tensao_v_sens_media_30s and tensao_v_sens_std_30s, to establish a baseline threshold for degradation. This reliance explains the high baseline recall achieved in this space; macroscopic shifts in rolling voltage averages and power volatility act as reliable markers of impending operational transience.

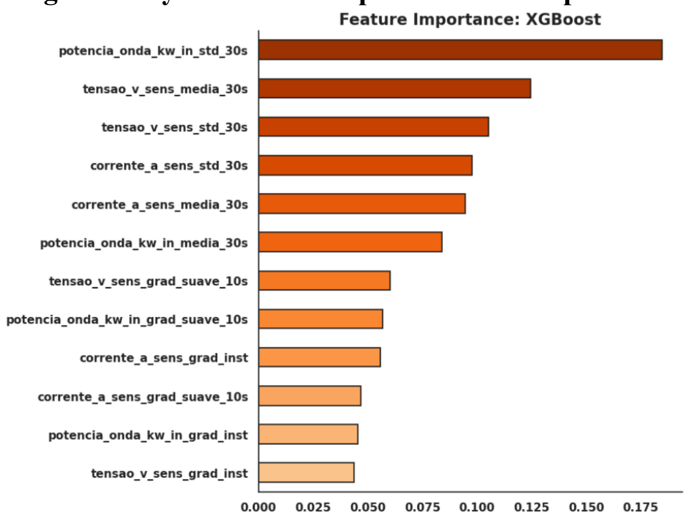
In the pure derivative domain (Figure 5), where absolute values are omitted, the model shifts its attention toward longer integration windows rather than short-term variations. The primary features selected are corrente_a_sens_grad_30s and tensao_v_sens_grad_30s. Interestingly, instantaneous or ultra-short derivatives, such as tensao_v_sens_grad_1s and potencia_onda_kw_in_grad_1s, settle at the very bottom of the importance hierarchy. This distribution explains the poor performance observed earlier for this feature space.

Figure 5: Pure Multi-Scale Gradients Feature Importance



The Hybrid Feature Space chart (Figure 6) demonstrates how the model reconciles both worlds to maximize precision. The top tier is controlled by rolling statistical windows, led by potencia_onda_kw_in_std_30s and tensao_v_sens_media_30s. However, the intermediate tier is populated by multi-scale gradient features, such as tensao_v_sens_grad_suave_10s and potencia_onda_kw_in_grad_suave_10s.

Figure 6: Hybrid Feature Space Feature Importance



This structural distribution validates the high-precision mechanism discussed previously:

- Primary Layer (Windows): The rolling means and standard deviations identify broad operational shifts, acting as a coarse filter.
- Secondary Layer (Gradients): The smoothed 10-second gradients function as a verification gate.

The model only flags a true anomaly or health change when a shift in the absolute power or voltage baseline coincides with an accelerating kinematic trend. By relying on this two-tiered verification, the hybrid model eliminates false positives.

In summary, the benchmarking of these distinct input spaces demonstrates that there is no "one-size-fits-all" configuration for upstream hydrogen plant diagnostics; rather, the optimal selection depends entirely on the specific industrial deployment objective:

- For Maximum Anomaly Capture and Edge Computing: If the primary engineering goal is to maximize the early detection rate of impending degradation events (Recall) while minimizing computational latency, the Pure Statistical Windows space paired with XGBoost is the undisputed superior choice. It achieves the highest overall Accuracy (88.91%), Recall (88.64%), and F1-Score (88.64%), while maintaining an ultra-low inference footprint (0.0107 ms/sample) that is ideal for localized microcontrollers.
- For False Alarm Suppression: Conversely, if the overriding priority of the plant operators is to prevent costly false-positive triggers that lead to unnecessary maintenance downtime, the Hybrid Feature Space becomes the preferred paradigm due to its peak Precision (90.24%), though this choice requires accepting a 4.5% drop in recall and a doubling of the computational overhead.
- Regarding Pure Kinematics: Finally, the standalone use of Pure Multi-Scale Gradients is fundamentally unviable for reliable health monitoring due to the severe loss of absolute baseline context, rendering both algorithms highly susceptible to signal volatility.

Ultimately, these trade-offs underscore that feature engineering for intermittent renewable-powered systems must be intentionally aligned with either fault sensitivity or alarm suppression priorities.

5. CONCLUSION

This study presented and benchmarked a machine learning-based predictive framework for the early detection of hydrogen crossover anomalies in PEM electrolyzers directly coupled to intermittent ocean wave energy. Exploiting exclusively upstream electrical telemetry, two ensemble classifiers, Random Forest and XGBoost, were evaluated across three distinct feature engineering strategies through a rigorous ablation study.

The results consistently demonstrated that the Pure Statistical Windows feature space, combining 30-second rolling means and standard deviations, yielded the strongest overall diagnostic performance. XGBoost achieved the highest Accuracy (88.91%), Recall (88.64%), and F1-Score (88.64%), while maintaining an ultra-low inference latency of 0.0107 ms/sample. This configuration establishes a compelling case for edge-computing deployment directly on localized microcontrollers attached to the upstream sensor manifold, enabling sub-millisecond anomaly diagnostics without reliance on centralized data processing infrastructure.

The Hybrid Feature Space, fusing rolling statistical windows with smoothed 10-second transient gradients, enabled XGBoost to achieve its peak Precision of 90.24% through a two-tiered verification mechanism: rolling statistics act as a coarse operational filter identifying macroscopic regime shifts, while smoothed kinematic gradients function as a confirmation gate. This configuration is recommended for industrial deployments where false-alarm suppression is the primary operational priority — such as plants where unplanned maintenance downtime carries significant economic penalties — accepting a modest 4.5% reduction in recall and a doubling of computational overhead relative to the pure statistical approach.

Conversely, the standalone use of Pure Multi-Scale Gradients proved fundamentally unviable for reliable health monitoring, with classification accuracy collapsing to approximately 59% for both algorithms. This outcome confirms that transient derivatives, while physically meaningful for characterizing intermittency kinematics, lack the absolute magnitude context required for stable decision boundaries under tree-based architectures.

Across all feature spaces, XGBoost demonstrated a consistent computational advantage over Random Forest, with inference latencies approximately 4 to 5 times lower, reinforcing its suitability for real-time diagnostics in resource-constrained and latency-sensitive environments.

From a safety standpoint, the proposed framework establishes a viable early-warning horizon that precedes physical crossover manifestation at the downstream gas analyzer, directly supporting proactive membrane integrity monitoring under highly transient wave-coupled operation — a critical contribution for probabilistic safety assessment of next-generation renewable-powered hydrogen facilities.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Human Resources Program of the Brazilian National Agency of Petroleum, Natural Gas and Biofuels – PRH-ANP, specifically through the PRH 38.1 program entitled 'Risk Management, Reliability, and Sustainability in the Exploration and Production of Petroleum, Natural Gas, and Renewable Energies,' funded by ANP and managed by Fapesp (São Paulo Research Foundation), under grant/process number 2025/03497-0.

References

[1] International Energy Agency. *The Future of Hydrogen: Seizing Today's Opportunities*. IEA, Paris, 2019.

[2] M. Carmo, D. L. Fritz, J. Mergel, and D. Stolten. "A comprehensive review on PEM water electrolysis," *International Journal of Hydrogen Energy*, vol. 38, no. 12, pp. 4901–4934, 2013.

[3] Abel, Riley, Marty Schwarz, Daniel Leighton, Vahan Gevorgian, and Nicholas Wimer. 2025. "Public Reference Data for Megawatt-Scale Hydrogen Electrolysis – Simulated Wave." NLR Data Catalog. Golden, CO: National Laboratory of the Rockies. Last updated: March 12, 2026. DOI: 10.7799/3007836.

[4] P. Trinke, P. Haug, J. Brauns, B. Bensmann, R. Hanke-Rauschenbach, and T. Turek. "Hydrogen crossover in PEM and alkaline water electrolysis: mechanisms, direct comparison and mitigation strategies," *Journal of The Electrochemical Society*, vol. 165, no. 7, p. F502, 2018.

[5] L. Breiman. "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] T. Chen and C. Guestrin. "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.