Prediction of Power Outages in Hybrid Renewable Energy Systems Using Supervised Learning

Ilkhomjon Siddikov[1], Slobodan Bojanic[2], Akmaljon Abdumalikov[3] and Javokhir Sherbaev[3]
[1]*Department of Power Supply and Renewable Energy Sources, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers National Research University, Kari Niyazov Str. 39, 100000 Tashkent, Uzbekistan*
[2]*Universidad Politecnica de Madrid, Calle Ramiro de Maeztu 7, 28040 Madrid, Spain*
[3]*Jizzakh Branch of the National University of Uzbekistan, Sh. Rashidov Shoh Str. 259, 130100 Jizzakh, Uzbekistan*
isiddikov.1954@gmail.com, slobodan.bojanic@upm.es, akmalabdumalikov6@gmail.com, javokhirsherbaev@gmail.com

Keywords: Power Outage Prediction, Hybrid Renewable Energy, IoT Monitoring, Supervised Learning, Logistic Regression, Decision Tree.

Abstract: Hybrid renewable energy systems (HRES) combining photovoltaic generation, battery storage, and backup sources are increasingly used in microgrids and remote installations. Despite improved sustainability, these systems remain vulnerable to short-term supply interruptions caused by irradiance variability, load changes, battery depletion, or backup unavailability. Conventional threshold-based alarms often fail to capture the multi-factor nature of outage formation, motivating data-driven prediction approaches. This paper presents an IoT-based monitoring and supervised learning framework for near-term outage-risk prediction in HRES. An Arduino-based edge device collects key electrical measurements and transmits time-stamped telemetry to a cloud platform via GSM/GPRS. Using a power-balance perspective, features are engineered to represent system adequacy, including irradiance conditions, battery state-of-charge (SoC), depletion trends, and backup availability. Horizon-based labeling is applied so that samples are marked as outage-risk when a failure occurs within a defined future window, enabling early warning rather than post-event detection. Three supervised classifiers-logistic regression, decision tree, and random forest-are trained and evaluated on a synthetic one-year dataset sampled at five-minute intervals. Performance is assessed using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Results show that tree-based models, especially random forests, provide stronger recall and risk ranking performance, enabling earlier detection of outage-prone conditions. The proposed approach demonstrates that combining IoT telemetry with probabilistic prediction can support proactive actions such as generator start, load shedding, and maintenance planning, while providing a practical foundation for future validation using real operational data.

1 INTRODUCTION

Hybrid renewable energy systems (HRES) integrate PV generation with battery storage and a backup source such as the grid or a generator. While this architecture improves sustainability and can reduce fuel usage, it also creates reliability challenges when renewable output drops, loads rise, or the backup source becomes unavailable. In practice, outages often occur from a combination of factors rather than a single threshold.

Low-cost IoT sensing makes continuous monitoring feasible, and machine learning offers a way to turn logged telemetry into actionable predictions. Instead of reacting after a failure, operators can use early warning to start backup supply, shift or shed non-critical loads, and schedule targeted maintenance. This paper develops an outage-risk prediction workflow for HRES and compares common supervised classifiers using a consistent labeling strategy and evaluation protocol.

Related work and motivation. Power outage prediction in microgrids and hybrid renewable energy systems (HRES) has been approached from both reliability-based modeling and data-driven forecasting. Reliability and adequacy studies provide useful long-horizon risk estimates and planning indicators, but operational early warning remains challenging because outages often arise from a combination of short-term factors such as irradiance drops, load variations, and backup unavailability [1], [2]. In parallel, recent outage prediction research increasingly uses machine learning to capture nonlinear interactions among weather, demand, and system constraints, producing probabilistic risk estimates rather than only deterministic threshold alarms [3], [4]. Recent studies have also emphasized the importance of IoT communication reliability and weather-related outage prediction using deep learning and severity-based event classification [5]-[7]. These results motivate framing outage prediction as a near-term decision-support problem where lead time and false-alarm control are as important as overall accuracy.

IoT telemetry for operational analytics. Low-cost IoT monitoring has enabled continuous acquisition of electrical and context signals (e.g., voltage/current measurements, battery state indicators, irradiance proxies, and backup status), which supports cloud- or edge-based analytics for distributed energy systems [8], [9]. For supervised prediction on such telemetry, interpretable baselines (e.g., logistic regression) and tree-based models are commonly used because they can model feature interactions and operate under practical deployment constraints (limited compute, noisy sensors, intermittent connectivity) [10]-[12]. However, many studies evaluate prediction at the failure instant or report aggregate metrics without explicitly defining a forward-looking labeling strategy and alert policy (e.g., persistence thresholds and lead time), which are necessary to convert model outputs into actionable operational warnings [13]-[15].

Gap and contribution positioning. Building on these directions, this paper targets operational early warning by combining IoT-to-cloud monitoring with horizon-based labeling and deployment-oriented alerting. The proposed pipeline defines “imminent outage” using a fixed future horizon, constructs power-balance-guided features reflecting PV scarcity, battery depletion, and backup availability, and maps predicted probabilities to persistence-based warning rules that align with practical actions such as generator start, load shedding, and maintenance dispatch.

Contributions:
1) IoT monitoring architecture using Arduino + GSM telemetry;
2) power-balance-guided labeling and feature set for near-term outage prediction;
3) supervised model training and comparison (logistic regression, decision tree, random forest);
4) operational alert logic based on probability and persistence.

2 METHODOLOGY

The target HRES consists of PV generation connected to a DC bus, a battery storage unit, and a backup source (grid and/or generator). To observe the system state, voltage and current sensors measure PV output, battery terminal behavior, and load demand. An Arduino-based controller aggregates these signals, attaches timestamps, and transmits telemetry to a cloud server using a GSM/GPRS module. When risk conditions persist (for example, rapid battery discharge during a period without backup power), the edge device can also issue simple operator notifications such as SMS alerts. The overall IoT-based monitoring architecture of the proposed hybrid renewable energy system is shown in Figure1.

To describe the operating condition, we use a power balance relationship. Let P_{solar} represent PV power, P_{grid} the backup supply (grid/generator) contribution, P_{batt} battery power (positive when discharging to support the load), and P_{load} the load demand. A simplified instantaneous balance can be written as:

$$P_{solar} + P_{grid} + P_{batt} \approx P_{load}.$$

An outage risk increases when P_{pv} decreases (due to weather/irradiance drop), P_{grid} becomes unavailable (resulting from a grid-off event), and the battery approaches depletion.

To model battery dynamics, the SoC update can be approximated by:

$$SoC_{t+1} = SoC_t + \frac{\eta_c P_c(t) - \frac{1}{\eta_d} P_d(t)}{E_{batt}} \Delta t,$$

where $P_{c,t}$ and $P_{d,t}$ are charge/discharge power components, E_{batt} is battery energy capacity, η_c,η_d are efficiencies, and Δt is the sampling interval (5 minutes).

We construct a synthetic one-year operational dataset sampled at 5-minute intervals. Each record contains electrical measurements and contextual indicators that influence HRES adequacy, including PV conditions, battery state, load level, and backup availability.

Reproducibility. A one-year synthetic HRES dataset was generated at 5-minute resolution using a simplified power-balance model with stochastic PV/load profiles, battery SoC dynami cs, and backup availability. Key parameters (sampling interval, battery capacity, SoC minimum threshold, and prediction horizon) are fixed and reported in the Methodology. Dataset generation scripts and preprocessing code can be shared by the authors upon reasonable request.

Instead of labeling only the exact failure instant, we define a forward-looking label to support early warning. For a chosen prediction horizon 30 minutes, we label a data point as "imminent outage" if an outage occurs within the next 30 minutes. In the simulation, an outage event is triggered when available supply cannot meet demand and the battery is at or below a minimum SoC threshold:

$$Outage \Leftrightarrow (P_{pv} + P_{grid} < P_{load})^{\wedge}(SoC \leq SoC_{min})$$

This definition aligns the ML target with operational needs: predicting risk early enough to take corrective action.

To capture dynamics that precede failure, we additionally use engineered variables such as:

- $P_{pv} = V_{pv} \cdot I_{pv}$ (PV power proxy).
- $P_{load} = V_{load} \cdot I_{load}$ (Load power proxy).
- $\Delta SoC = SoC_t - SoC_{t-1}$ (depletion/charge trend).
- Rolling statistics (mean/variance) for irradiance and ΔSoC over recent windows.
- "Backup-off duration" (time since grid/generator became unavailable). The complete feature engineering and horizon-based labeling workflow is illustrated in Figure 2. The input variables used for outage-risk prediction are summarized in Table 1.

The prediction objective is binary classification: determine whether the system is likely to enter an outage state within the selected future horizon. We evaluate three supervised approaches:

- Logistic Regression as an interpretable baseline that estimates outage probability from a linear combination of features.
- Decision Tree to capture nonlinear thresholds and feature interactions that reflect engineering logic.
- Random Forest as an ensemble of trees to improve generalization and reduce variance relative to a single tree.

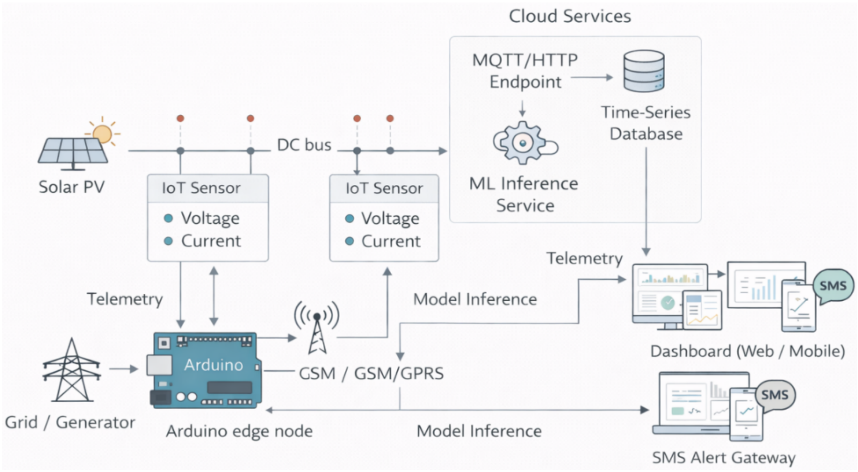

Figure 1: System architecture of the hybrid renewable energy monitoring setup.

Table 1: Input features for outage prediction.

Feature	Description
Solar Voltage	PV panel voltage (volts)
Solar Current	PV panel current (amperes)
Irradiance	Solar irradiance (W/m²)
Battery SoC	Battery state-of-charge (%)
Battery Voltage	Battery terminal voltage (volts)
Load Current	Load demand current (amperes)
Grid Power	Power from grid or generator (watts)
Time OfDay	Hour or categorical time indicator
Weather Condition	Clear/Cloudy/Rainy (categorical)

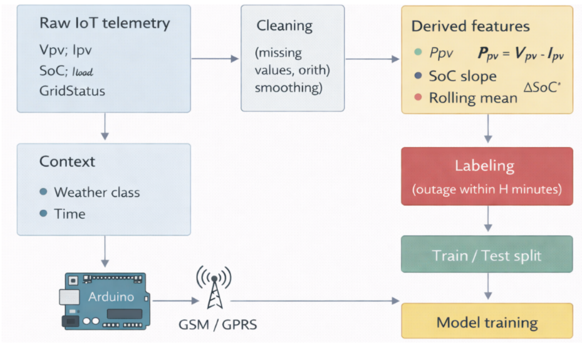

Figure 2: Feature engineering and labeling pipeline.

Training protocol. The dataset is split into 70% for training and 30% for testing using stratified sampling to preserve the label distribution. Continuous features are scaled as needed; categorical variables (e.g., weather condition) are encoded (e.g., one-hot). Hyperparameters are selected using grid search with cross-validation.

Example hyperparameters:

- Logistic regression: L2 regularization, C=1.0, balanced class weights;
- Decision tree: max depth 6, minimum leaf size 20, Gini impurity;
- Random forest: 200 estimators, max depth 10, minimum leaf size 10.

Evaluation metrics**.** We report accuracy along with precision, recall, and F1-score to reflect class imbalance sensitivity. ROC-AUC is also included to assess ranking quality across thresholds. Because outage events are typically rare, threshold-independent and imbalance-aware metrics such as ROC-AUC and F1-score are recommended for reliable comparison [16].

In deployment, the classifier output probability is treated as an outage-risk score and converted into alerts using a fixed probability threshold with short persistence to reduce noise. We also report lead time as the delay between the first threshold crossing and the outage event to reflect operational usefulness.

The proposed monitoring concept is also consistent with previous work on signal conversion and measurement processes in multiphase power measurement and control devices [17].

3 RESULTS

We trained and tested the models on the simulated dataset using horizon-based labels. Figure 3 summarizes performance across accuracy, precision, recall, F1-score, and ROC-AUC for logistic regression, decision tree, and random forest.

Overall, the random forest provides the strongest results, particularly in recall and ROC-AUC, suggesting better sensitivity to multi-factor risk scenarios (e.g., simultaneous irradiance reduction, increased load, and backup unavailability). Logistic regression remains competitive and is valuable when interpretability is required. Decision trees offer human-readable rules that often align with engineering expectations, such as elevated risk under combined low SoC and reduced supply conditions.

Rule-based baseline for comparison. To contextualize the supervised models, we additionally evaluated a simple rule-based early-warning baseline that mimics a practical engineering alarm. The baseline triggers an “outage-risk” alert if (i) the estimated net deficit persists for k consecutive samples ($P_{pv} + P_{backup} < P_{load}$), and (ii) the battery state-of-charge is at or below SoC_{min} We set $k = 2$ (10 minutes) to reduce single-sample noise alarms. The baseline is evaluated on the same test split and with the same prediction horizon H = 30 min as the learning-based models. This comparison clarifies the value added by probabilistic prediction (risk scoring) beyond deterministic thresholding, especially under fluctuating PV and load conditions. The numerical accuracy, precision, recall, and F1-score achieved by this rule-based baseline, alongside the corresponding metrics for the three learning-based models shown in Figure 3.

Feature importance insight. For tree-based models, feature importance rankings commonly highlight battery SoC, backup availability/power, irradiance, and ΔSoC as dominant predictors. This supports the interpretation that outages typically emerge from combined resource scarcity rather than a single threshold.

Probability timeline behavior. A representative daily plot (Fig. 4) illustrates how predicted risk increases ahead of outage events, especially during extended backup-off intervals coupled with low irradiance and ongoing battery depletion. This demonstrates the early-warning behavior targeted by horizon-based labeling.

We also reviewed error types (false alarms vs missed outages) to assess operational suitability. The random forest reduces missed outages compared to logistic regression while keeping false alarms at an acceptable level under the simulated assumptions. The experimental results indicate that supervised learning can provide effective outage-risk prediction for HRES under the simulated operating assumptions. From a practical perspective, the primary advantage is risk-based early warning, which can enable actions such as starting the generator, shedding non-critical loads, or scheduling maintenance before loss of supply occurs.

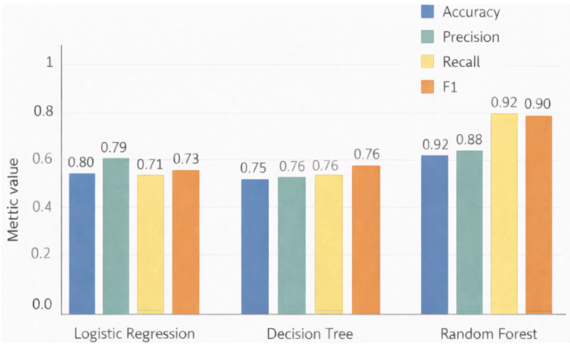

Figure 3: Multi-metric model comparison.

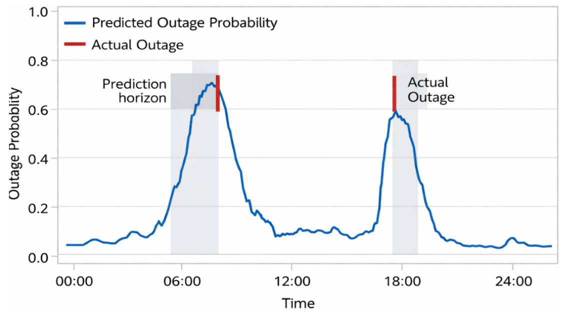

Figure 4: Outage-risk timeline with event markers.

Deployment considerations:

- Measurement quality. Real installations may experience sensor noise, drift, and missing data. Practical mitigation includes filtering, validity checks, and fallback behavior when telemetry is unavailable.
- Class imbalance. Outages are typically rare relative to normal operation; therefore, F1-score and ROC-AUC are more informative than accuracy. Alert thresholds should be tuned based on the operational cost of false alarms versus missed events.
- Edge vs cloud inference. Cloud inference simplifies model updates and fleet management, while edge inference reduces latency and dependence on connectivity. A hybrid approach is realistic: rules at the edge for immediate safety triggers, and probabilistic ML scoring in the cloud.
- Security. Telemetry links and APIs should be protected with authentication and secure transport where feasible to reduce spoofing and tampering risk.

Limitations. Because the dataset is simulated, reported metrics likely represent optimistic performance under controlled assumptions. Field behavior may include additional effects such as battery aging, inverter faults, intermittent communications, and unobserved loads. These limitations suggest the next step is collecting real operational logs for retraining and calibration, rather than treating simulation results as final.

4 CONCLUSIONS

This paper presented an IoT-monitored, supervised learning-based framework for near-term outage prediction in hybrid renewable energy systems. By combining low-cost sensing hardware, cloud telemetry, and machine learning models, the proposed approach enables predictive risk assessment instead of reactive failure handling. Horizon-based labeling and power-balance-guided feature engineering allow models to generate operationally meaningful early warnings. Evaluation on a synthetic one-year dataset showed that supervised classifiers can effectively capture interactions between renewable variability, load demand, battery state, and backup availability. Among the tested models, random forest achieved the strongest overall performance, particularly in recall and ROC-AUC, indicating improved sensitivity to multi-factor outage scenarios. Logistic regression provided interpretable probabilistic estimates, while decision trees offered engineering-consistent decision rules.

From an operational perspective, probability-based risk scoring supports proactive actions such as generator activation, load management, and maintenance scheduling. The framework is compatible with low-cost hardware and can be deployed using edge-cloud architectures depending on latency and connectivity constraints. The main limitation of the study is the use of simulated data, which may not fully represent real-world factors such as sensor noise, aging components, or communication interruptions. Future work will focus on validation using real HRES operational logs, model recalibration, and adaptive deployment strategies. Overall, the proposed method provides a practical and scalable basis for improving reliability and resilience in renewable-based hybrid energy systems.

REFERENCES

[1] B. Lei, Y. Ren, H. Luan, R. Dong, X. Wang, J. Liao, S. Fang, and K. Gao, "A review of optimization of microgrid," Mathematics, vol. 11, no. 4, art. 822, 2023, [Online]. Available: https://doi.org/10.3390/math11040822.

[2] A. Nargeszar, A. Ghaedi, M. Nafar, and M. Simab, "Reliability evaluation of the renewable energy-based microgrids considering resource variation," IET Renewable Power Generation, vol. 17, no. 3, pp. 507-527, 2023, [Online]. Available: https://doi.org/10.1049/rpg2.12611.

[3] T. Adefarati and R. C. Bansal, "Reliability and economic assessment of a microgrid power system with the integration of renewable energy resources," Applied Energy, vol. 206, pp. 911-933, 2017, [Online]. Available: https://doi.org/10.1016/j.apenergy.2017.08.228.

[4] A. H. Fathima and K. Palanisamy, "Optimization in microgrids with hybrid energy systems-A review," Renewable and Sustainable Energy Reviews, vol. 45, pp. 431-446, 2015, [Online]. Available: https://doi.org/10.1016/j.rser.2015.01.059.

[5] B. Jesus et al., "An approach to assess robustness of MQTT-based IoT systems," Internet of Things, vol. 31, art. 101590, 2025, [Online]. Available: https://doi.org/10.1016/j.iot.2025.101590.

[6] X. Wang et al., "Deep learning-based weather-related power outage prediction with socio-economic and power infrastructure data," in Proc. IEEE Power & Energy Society General Meeting (PESGM), 2024, pp. 1-5, [Online]. Available: https://doi.org/10.1109/PESGM51994.2024.10688596.

[7] F. Yang et al., "Enhancing weather-related power outage prediction by event severity classification," IEEE Access, vol. 8, pp. 60029-60042, 2020, [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2983159.

[8] N. Rouibah et al., "Smart monitoring of photovoltaic energy systems: An IoT-based prototype approach," Scientific African, vol. 30, art. e02973, 2025, [Online]. Available: https://doi.org/10.1016/j.sciaf.2025.e02973.

[9] M. N. Razali et al., "Development of IoT based smart photovoltaic energy monitoring system," AIP Conference Proceedings, vol. 2991, art. 050048, pp. 050048-1-050048-10, 2024, [Online]. Available: https://doi.org/10.1063/5.0200747.

[10] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001, [Online]. Available: https://doi.org/10.1023/A:1010933404324.

[11] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Hoboken, NJ, USA: Wiley, 2013, [Online]. Available: https://doi.org/10.1002/9781118548387.

[12] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861-874, 2006, [Online]. Available: https://doi.org/10.1016/j.patrec.2005.10.010.

[13] P. L. Watson et al., "Improved quantitative prediction of power outages caused by extreme weather events," Weather and Climate Extremes, vol. 37, art. 100487, 2022, [Online]. Available: https://doi.org/10.1016/j.wace.2022.100487.

[14] P. Gyasi et al., "Design of alarm thresholds and delay timers for non-IID process variables based on alarm durations," Process Safety and Environmental Protection, vol. 170, pp. 1173-1187, Feb. 2023, [Online]. Available: https://doi.org/10.1016/j.psep.2022.12.069.

[15] N. A. Adnan et al., "Study of generalized delay-timers in alarm configuration," Journal of Process Control, vol. 23, no. 3, pp. 382-395, 2013, [Online]. Available: https://doi.org/10.1016/j.jprocont.2012.12.013.

[16] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLOS ONE, vol. 10, no. 3, art. e0118432, 2015, [Online]. Available: https://doi.org/10.1371/journal.pone.0118432.

[17] I. Kh. Siddikov and A. A. Abdumalikov, "Modeling and research signals conversion proceses of multiphase power measure and control devices," in Proc. 2020 Int. Conf. Information Science and Communications Technologies (ICISCT), Nov. 2020, pp. 1-4, [Online]. Available: https://doi.org/10.1109/ICISCT50599.2020.9351458.