AI-Driven Optimization and Fault Intelligence for Grid-Connected Photo Sensor-Based Battery Balancing (GCPBBB) Systems

Technical Report

December 2025

Abstract

This report presents an advanced AI framework for optimizing Grid-Connected Photo Sensor-Based Battery Balancing (GCPBBB) systems. The framework integrates three machine learning subsystems: (i) an Energy Predictor for short-term demand forecasting, (ii) a Balancing Signal Classifier for intelligent battery scheduling, and (iii) a Grid Stability Monitor for real-time fault detection and classification.

A hybrid data strategy is employed, combining real PV fault signatures with high-fidelity synthetic operational data that simulate one month of minute-level GCPBBB operation. The system is grounded in explicit physical models for PV power, battery state-of-charge (SoC) dynamics, and grid power balance, and is implemented as a cyber–physical architecture with a FastAPI backend and a React "Control Room" dashboard.

Experimental results demonstrate high-quality forecasting performance (MAE ≈ 42 W), near-perfect classification accuracy ($> 99.7\%$) for both balancing and fault detection tasks, and strong interpretability via feature importance, PR curves, ROC curves, and confusion matrices. The resulting framework achieves industry-ready robustness for deployment in smart microgrids.

1 Introduction

The rapid penetration of solar photovoltaics (PV) introduces volatility and uncertainty into distribution networks. Grid-Connected Photo Sensor-Based Battery Balancing (GCPBBB) architectures mitigate these issues by combining PV generation, grid-connected inverters, and battery storage. However, truly optimal and safe operation requires:

- accurate short-term forecasting of grid consumption,
- intelligent charge/discharge decisions that respect SoC and energy constraints,
- fast and reliable detection of abnormal and fault conditions.

This work augments GCPBBB systems with an AI layer that provides predictive, prescriptive, and protective capabilities. Three non–deep-learning models are developed and tightly integrated:

1. **Grid Consumption Predictor** (Energy Planner),
2. **Balancing Signal Classifier** (Battery Planner),
3. **Grid Stability Monitor** (Guardian / Protection module).

The focus is on industry-grade reliability, interpretability, and real-time performance rather than maximum algorithmic complexity.

2 Physical and System-Level Modeling

2.1 PV Generation

PV array output is modeled as

$$P_{\mathrm{pv}}(t) = \eta A G(t) \left[1 - \beta\big(T(t) - T_{\mathrm{ref}}\big)\right], \tag{1}$$

where $G(t)$ is solar irradiance, $T(t)$ is cell temperature, η is nominal efficiency, A the effective area, and β the temperature coefficient. This expression is used both to calibrate the synthetic data generator and to ensure that produced operating points are physically feasible.

2.2 Battery SoC Dynamics

Battery State of Charge (SoC) evolves according to

$$\mathrm{SoC}(t+1) = \mathrm{SoC}(t) + \frac{\Delta t}{C_{\mathrm{bat}}}\left(\eta_c P_{\mathrm{ch}}(t) - \frac{1}{\eta_d} P_{\mathrm{dis}}(t)\right), \tag{2}$$

with C_{bat} the usable battery capacity, η_c and η_d the charge/discharge efficiencies, and $P_{\mathrm{ch}}(t)$, $P_{\mathrm{dis}}(t)$ the charge and discharge power. Operational constraints enforce

$$0 \le \mathrm{SoC}(t) \le 1. \tag{3}$$

2.3 Grid Power Balance

At each time step,

$$P_{\mathrm{grid}}(t) = P_{\mathrm{load}}(t) - P_{\mathrm{pv}}(t) - P_{\mathrm{bat}}(t), \tag{4}$$

where $P_{\mathrm{bat}}(t) = P_{\mathrm{dis}}(t) - P_{\mathrm{ch}}(t)$. This identity constrains the AI policies and is used in synthetic data generation and simulation.

3 Data Acquisition and Synthesis

3.1 Real-World Calibration Data

Real PV fault data were obtained from the GPVS-Faults dataset, consisting of 16 CSV files containing high-frequency voltage and current measurements under operating states F0 (stable) and F1–F7 (distinct fault modes). Calibration statistics such as:

- peak power (95^{th} percentile) $\approx 215.7\,\mathrm{W}$,
- mean PV voltage $\approx 88.1\,\mathrm{V}$,

were used to parameterize the synthetic generator and bound the simulated operating regime.

3.2 Synthetic Operational Dataset

A month-long synthetic dataset (43,200 samples at 1-minute resolution) was generated to emulate realistic GCPBBB operation:

- **Solar irradiance** was modeled as a diurnal sine wave with stochastic cloud-cover noise.
- **Grid consumption** followed a mixture-of-Gaussians pattern with morning and evening peaks, plus random intra-day fluctuations.
- **Battery SoC** was computed by integrating net energy flow using the dynamics of Section 2.

- **Balancing labels** were derived via simple rule-based logic:

$$a_t = \begin{cases} 0, & \text{if } P_{\text{pv}}(t) + P_{\text{grid}}(t) < P_{\text{load}}(t) \quad \text{(deficit / discharge)}, \\ 2, & \text{if } P_{\text{pv}}(t) > P_{\text{load}}(t) \ \wedge \text{SoC}(t) < \text{SoC}_{\text{max}}, \\ 1, & \text{otherwise (hold)}. \end{cases} \tag{5}$$

3.3 Empirical Distributions and Correlations

Histograms and kernel density estimates show:

- Irradiance is right-skewed, dominated by near-zero values at night.
- Consumption exhibits a bimodal distribution reflecting morning/evening peaks.
- SoC distribution is U-shaped, indicating frequent deep cycling between low and high SoC.

A correlation matrix for irradiance, temperature, grid consumption, and SoC reveals:

$$r(G, \text{SoC}) \approx 0.80, \quad r(T, G) \approx 0.87, \quad r(G, P_{\text{load}}) \approx 0.33, \tag{6}$$

confirming strong solar–battery coupling and weak dependence between irradiance and demand.

4 Machine Learning Subsystems

4.1 Energy Predictor (Grid Consumption Forecasting)

4.1.1 Problem Formulation

Given feature vector

$$\boldsymbol{x}_t = \{P_{\text{load}}(t-k:t),\ G_t,\ T_t,\ \sin(\omega t),\ \cos(\omega t)\}, \tag{7}$$

the objective is to estimate one-step-ahead consumption:

$$\widehat{P}_{\text{load}}(t+1) = f_\theta(\boldsymbol{x}_t). \tag{8}$$

Autocorrelation analysis yields $\rho_1 \approx 0.91$ and $\rho_5 \approx 0.76$, supporting the use of lagged consumption features.

4.1.2 Model

A Random Forest Regressor is employed due to its robustness, nonlinearity, and interpretability. Feature importance analysis shows:

- grid consumption at time t and $t-1$ dominate (combined importance > 0.9),
- exogenous features (irradiance, temperature, time-of-day) contribute marginally.

This confirms that short-term load is highly autoregressive.

4.2 Balancing Signal Classifier

4.2.1 Formulation

The balancing action $a_t \in \{0, 1, 2\}$ is modeled as a multiclass classification problem with policy

$$\pi^*(\boldsymbol{x}_t) = \arg\max_a p(a|\boldsymbol{x}_t), \tag{9}$$

where

$$\boldsymbol{x}_t = \{\text{SoC}(t),\ G_t,\ \widehat{P}_{\text{load}}(t+1),\ P_{\text{pv}}(t),\ P_{\text{grid}}(t)\}. \tag{10}$$

4.2.2 Model

An XGBoost classifier is trained with regularization and early stopping. Feature importance and mutual information analysis indicate:

- Irradiance accounts for $\sim 80\%$ of importance.
- SoC accounts for $\sim 12\%$.
- Remaining features (predicted load, grid consumption, temperature) collectively contribute $< 10\%$.

Thus, the classifier learns an intuitive policy: "charge when solar is abundant and SoC is low; discharge when solar is limited and demand is high."

4.3 Grid Stability Monitor

4.3.1 Feature Engineering

High-frequency PV-side voltage and current signals are aggregated over a sliding window to produce:

$$\boldsymbol{x}_t = \{\mu_v, \sigma_v, v_{\max}, v_{\min},\ \mu_i, \sigma_i, i_{\max}, i_{\min}\}. \tag{11}$$

4.3.2 Model

A Random Forest Classifier predicts fault state $F_t \in \{F0, \ldots, F7\}$. Bhattacharyya distances between classes exceed 3, indicating strong separability of feature distributions across fault types.

5 Cyber–Physical Architecture

5.1 Control Room Dashboard

A React-based single-page application serves as the operator-facing "Control Room". It:

- visualizes real-time solar irradiance, grid consumption, and SoC,
- displays current fault state and balancing action,
- provides manual controls for simulation and scenario testing.

5.2 Backend Services

A FastAPI backend orchestrates:

1. data ingestion from sensors (or synthetic simulator),
2. preprocessing and feature engineering in real time,
3. invocation of the three ML models,
4. stateful simulation of SoC and grid power flows.

5.3 Decision Loop

The decision logic is structured as:

Algorithm 1 AI-Driven GCPBBB Decision Loop

Acquire sensor vector $\boldsymbol{s}_t$
fault $\leftarrow$ StabilityMonitor($\boldsymbol{s}_t$)
if fault $\neq F0$ **then**
    Trigger protection relay and isolate PV/battery
else
    $\widehat{P}_{\text{load}}(t+1) \leftarrow$ EnergyPredictor($\boldsymbol{s}_t$)
    $a_t \leftarrow$ BalancingClassifier($\boldsymbol{s}_t, \widehat{P}_{\text{load}}(t+1)$)
    Apply battery action a_t and update SoC
end if

6 Experimental Evaluation

6.1 Energy Predictor

The Random Forest forecaster achieves:

- Mean Absolute Error (MAE): 42.11 W,
- Mean Absolute Percentage Error (MAPE): $\sim 3.9\%$,
- R^2: 0.98.

Given that typical residential loads range from 200 W to 2000 W, this error is negligible for operational scheduling.

6.2 Balancing Classifier Performance

Table 1 shows the class-wise metrics.

Class	Precision	Recall	F1-Score	Support
0 (Discharge)	1.00	1.00	1.00	4473
1 (Hold)	0.71	0.43	0.54	23
2 (Charge)	1.00	1.00	1.00	4141

Table 1: Balancing Signal Classifier performance on the test set.

The rare "Hold" class is the only challenging case. Precision–Recall curves show:

- $\text{AP}_0 \approx 1.00$,
- $\text{AP}_2 \approx 1.00$,
- $\text{AP}_1 \approx 0.78$.

Overall accuracy is 99.77%, and ROC curves for all classes achieve AUC ≈ 1.00.

6.3 Grid Stability Monitor Performance

Table 2 summarizes fault classification metrics.

Fault Type	Precision	Recall	F1-Score	Support
F0 (Stable)	1.00	1.00	1.00	47
F1	1.00	0.98	0.99	58
F2	1.00	1.00	1.00	54
F3–F7	1.00	1.00	1.00	274

Table 2: Grid Stability Monitor performance on fault dataset.

Confusion matrices demonstrate near-perfect diagonal dominance, with only isolated misclassifications (e.g., one F1 event classified as F6). ROC and PR curves for each class show almost ideal behavior (AUC and AP close to 1.00), reflecting strong separability of electrical signatures.

6.4 Model Interpretability

Feature-importance plots corroborate domain intuition:

- For the Energy Predictor, current and lagged consumption dominate, confirming strong autocorrelation in load.
- For the Balancing Classifier, irradiance and SoC are primary drivers, indicating that the battery strategy is predominantly solar-driven.
- For the Stability Monitor, PV-side voltage and current statistics (means and maxima) are most informative, matching expectations for fault discrimination.

7 Reliability, Safety, and Industry Alignment

7.1 Safety Layer

The Stability Monitor acts as a front-line protective layer. Upon detection of non-F0 states, it immediately triggers an isolation command, preventing unsafe power flows during faults.

7.2 Industry Standards

The architecture and control policies are designed to be compatible with:

- IEEE 1547: Standard for Interconnection and Interoperability of Distributed Energy Resources,
- IEC 61727: PV system characteristics and grid interface,
- IEC 62116: Test procedure for anti-islanding.

7.3 Deployment and DevOps

The complete system is containerized using Docker and orchestrated via Docker Compose on a Google Cloud Platform (GCP) Compute Engine instance. An Nginx reverse proxy manages routing and TLS termination. GitHub Actions provide CI/CD, running unit tests, building backend and frontend images, and validating deployment configurations before release.

8 Conclusion and Future Work

This report has presented a cohesive AI-driven framework for GCPBBB systems that combines:

- physically grounded models for PV, battery, and grid power flows,
- high-performing ML models for forecasting, balancing, and fault detection,
- a cyber–physical deployment architecture suitable for industrial environments.