

Generalized versus Panel-specific Machine Learning Models for Photovoltaic Power Estimation Under Limited Multi-Module Field Data

Giovanni Dimas Prenata*, Ahmad Ridho'i

Department of Electrical Engineering, Faculty of Intelligent Electrical and Informatics Technology, Universitas 17 Agustus 1945 Surabaya, INDONESIA

Article Info

Abstract

Article history:

Received
02-08-2026

Revised
29-08-2026

Accepted
12-09-2026

Available Online
19-10-2026

Keywords:

Cross-day Validation,
Generalized Modeling,
Panel-specific
Modeling, Photovoltaic
Power Estimation,
Random Forest,
Ridge Regression

Limited multi-module field data creates a trade-off between generalized models that pool observations from several photovoltaic modules and panel-specific models that preserve module dependent behavior but use fewer training samples. This study compares three modeling scopes for photovoltaic power estimation using 63 observations from three nominally 100 W modules measured over four days. Generalized-Shared, Generalized-Panel-Aware, and Panel-Specific scopes were evaluated using Ridge Regression, K-Nearest Neighbor Regression, Radial Basis Function Network Regression, and Random Forest Regression under nested leave-one-day-out validation. A datasheet physical power estimate was also evaluated as a non-machine-learning baseline. Panel-Specific Ridge achieved the best overall performance with an RMSE of 7.3236 W, MAE of 5.6750 W, and R2 of 0.8788, reducing RMSE by 60.23% relative to the physical baseline (18.4161 W). Generalized-Panel-Aware Random Forest achieved an RMSE of 7.7876 W overall and became the best configuration for P3 (6.3310 W) and for the Tuesday left-out day (4.9565 W). Generalized KNN and panel-aware RBF also benefited from pooled training data, whereas Panel-Specific Ridge remained the strongest regularized linear configuration. Wilcoxon signed-rank tests with Holm correction found no statistically significant scope differences across the 12 panel-day units; the largest scope effect was observed for panel-aware Random Forest (rank-biserial = -0.4872, adjusted $p = 0.3027$). The results show that the benefit of pooling or specialization is algorithm and operating condition dependent and should be interpreted within the limitations of the small single-site field dataset.



© 2026 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license <https://creativecommons.org/licenses/by/4.0/>.

Published by Universitas Katolik Indonesia Atma Jaya.

1. INTRODUCTION

Photovoltaic (PV) power varies with irradiance, module temperature, electrical operating conditions, and module-specific characteristics. Even modules with the same nominal rating can exhibit different open-circuit voltage, short-circuit current, maximum-power-point behavior, and field response. Consequently, a model trained on one module may not generalize directly to another module of the same rated capacity [1].

Machine learning methods can represent nonlinear PV relationships that are difficult to capture with one fixed analytical equation, but their performance depends on data coverage, feature selection, model complexity, and validation design [2]-[6]. In multi-module data, a key modeling choice is whether observations should be pooled into a generalized model or separated into panel-specific models. Pooling increases sample size and operating-space coverage, whereas panel-specific fitting preserves module-dependent behavior at the cost of substantially fewer training observations. Related transfer and multi-location studies likewise show that local and pooled models provide different advantages under heterogeneous data [7]-[9].

The preferred scope can also depend on the learning algorithm. Ridge Regression offers a regularized linear representation that is comparatively stable under correlated predictors and limited samples [10]. K-Nearest Neighbor (KNN) Regression depends on local training density [11], while Radial Basis Function (RBF) networks use localized nonlinear responses around representative centers [12]. Random Forest

*Corresponding author : Giovanni Dimas Prenata
Email address : gprenata@untag-sby.ac.id

provides a modern tree-based ensemble that can capture nonlinear interactions and is widely used as a robust comparator in PV prediction studies [13],[14]. These algorithms therefore have different requirements for sample size and feature-space coverage.

Evaluation under limited data requires particular care. Hyperparameter selection and performance estimation on the same folds can produce optimistic error estimates; nested cross-validation separates inner model selection from outer testing [15]. Because measurements collected within the same day share environmental and operating context, day-blocked validation is also preferable to random row-wise splitting [16]. Statistical comparisons should likewise use grouped evaluation units rather than treating repeated measurements as fully independent observations [17].

Despite recent PV forecasting studies under limited-data and data-scarce conditions [18]-[20], direct evidence remains limited on generalized versus panel-specific PV power estimation using the same multi-module field dataset and identical cross-day partitions. Methodologically related electrical-engineering studies compared KNN and SVR for reliability prediction [21] and small-sample regression with nested grouped validation [22], while structured cross-validation [23] and gradient-versus-metaheuristic training comparisons [24] have been used to assess robustness under restricted data. These precedents motivate the present multi-algorithm and leakage-resistant cross-day evaluation. Three scopes are evaluated: Generalized-Shared, Generalized-Panel-Aware, and Panel-Specific. Each scope is implemented using Ridge, KNN, RBF Network, and Random Forest Regression. A datasheet-informed physical power estimate is additionally evaluated as a non-machine-learning benchmark. Loaded voltage and current are used only to construct the target and are excluded from model inputs to prevent direct target leakage.

The study uses 63 observations from three nominally 100 W PV modules measured over four field days. Its contributions are: (1) a controlled three-scope comparison; (2) evaluation across linear, local, nonlinear, and tree-based algorithms; (3) nested leave-one-day-out validation with training-only preprocessing; (4) comparison against a physical benchmark; and (5) overall, panel-level, day-level, and panel-day analyses with exact Wilcoxon-Holm testing and rank-biserial effect sizes. The aim is not to identify a universally best scope, but to determine how pooling versus specialization interacts with algorithm complexity and the operating conditions of an unseen day.

2. LITERATURE REVIEW AND METHODOLOGY

2.1. Research Design, Field Measurements, and Model Scopes

The computational experiment was implemented in C++ without an external machine-learning library. Figure 1 summarizes the computational sequence from field-data preparation and model-scope definition to nested cross-day validation, final testing, physical-baseline comparison, statistical analysis, and output generation.

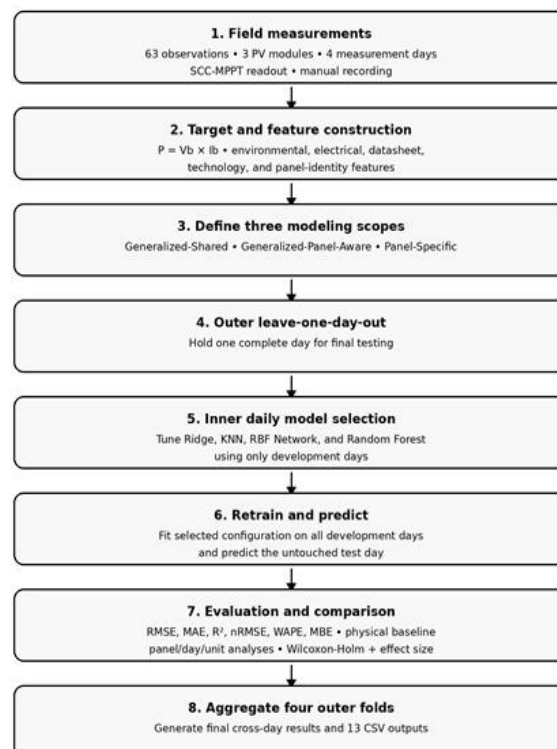


Figure 1. Research workflow for generalized versus photovoltaic panel power estimation.

Field measurements were conducted at the Renewable Energy Laboratory (*Laboratorium Energi Baru Terbarukan*), Electrical Engineering Study Program, Universitas 17 Agustus 1945 Surabaya, Surabaya, Indonesia. The dataset contained 63 observations from three nominally 100 W modules measured during the same week on 20, 21, 23, and 24 April 2026. P1 and P3 were polycrystalline and P5 was monocrystalline. Measurements were taken at approximately one-hour intervals. The modules were operated with an MPPT solar charge controller (SCC-MPPT), and the electrical and environmental parameters used in this study were read directly from the device interface. All values were manually recorded at each measurement time, so there is no separate automatic data logger being used. The campaign was conducted during the dry season under generally clear and rain-free conditions. Separate manufacturer specifications for the embedded irradiance and temperature sensing elements, including accuracy and resolution, were not documented and are therefore treated as a measurement limitation rather than inferred.

Measurement time in *HH.MM* format was converted into decimal hours. The prediction target was the measured loaded power, where $V_{b,i}$ and $I_{b,i}$ denote the loaded voltage and current of observation i :

$$P_{\text{actual},i} = V_{b,i} I_{b,i} \quad (1)$$

The standard test condition fill factor was calculated from the rated maximum power P_{max} and the datasheet open circuit and short circuit values:

$$FF_{\text{STC},i} = \frac{P_{\text{max},i}}{V_{oc,\text{STC},i} I_{sc,\text{STC},i}} \quad (2)$$

A datasheet-informed physical power estimate was then constructed as:

$$P_{\text{phy},i} = FF_{\text{STC},i} V_{oc,i} I_{sc,i} \quad (3)$$

The quantity in Equation (3) is not treated as a measured maximum power point value. It combines field measured V_{oc} and I_{sc} with the STC fill factor and is used both as a normalized physics feature and in its abnormalized form, as a non machine learning physical benchmark. Loaded voltage and loaded current are excluded from the model inputs because they directly form the target in Equation (1). The program also retains a diagnostic flag when

$$I_{b,i} > I_{sc,i} \quad (4)$$

or

$$V_{b,i} \geq V_{oc,i} \quad (5)$$

Flagged observations were retained in the principal experiment and were not treated as confirmed faults. Twelve candidate features were constructed in Equation (6), the ratio terms denote measured or datasheet quantities normalized by their corresponding STC references, T_{ech_i} is the monocrystalline indicator, and P3/P5 are panel dummy variables with P1 as the reference category.

$$x_i = [G_i, T_i, t_i, R_{Voc,i}, R_{Isc,i}, R_{P,i}, F, FF_{\text{STC},i}, R_{Vmp,i}, R_{Imp,i}, T_{ech_i}, D_{P3,i}, D_{P5,i}] \quad (6)$$

Three model scopes were compared as summarized in Table 1. Generalized-Shared uses pooled observations without explicit panel identity; Generalized-Panel-Aware adds panel dummy variables; Panel-Specific fits separate models for P1, P3, and P5. Datasheet ratios, technology, and panel indicators that are constant within one module are omitted from Panel-Specific models because they contain no within-panel predictive variation.

Table 1. Model scopes and input representations

Scope	Training structure	Features	Description
Generalized-Shared	One model trained using P1, P3, and P5	10	Environmental, electrical, datasheet, and technology features without explicit panel identity
Generalized-Panel-Aware	One model trained using P1, P3, and P5	12	Generalized-Shared features plus P3 and P5 dummy variables
Panel-Specific	Separate models for P1, P3, and P5	6	Irradiance, temperature, time, normalized open circuit and short circuit quantities, and normalized physical power proxy

2. Machine Learning Models and Hyperparameter Selection

Each scope was implemented using Ridge Regression, K-Nearest Neighbor Regression, Radial Basis Function Network Regression, and Random Forest Regression, producing 12 scope algorithm configurations. All predictors were normalized using only the corresponding training set according to Equation (7).

$$x'_{i,j} = \frac{x_{i,j} - x_{j,\min}^{(\text{train})}}{x_{j,\max}^{(\text{train})} - x_{j,\min}^{(\text{train})}} \quad (7)$$

Validation and test observations used the minimum and maximum obtained from their associated training set. If a feature had zero training range, its normalized value was set to zero. Ridge and RBF targets were normalized using the training-target range in Equation (8) and transformed back to watts using Equation (9), KNN and Random Forest retained target power directly in watts. Final predictions were constrained to non-negative values.

$$y'_i = \frac{y_i - y_{\min}^{(\text{train})}}{y_{\max}^{(\text{train})} - y_{\min}^{(\text{train})}} \quad (8)$$

and returned to watts through

$$\hat{y}_i = y_{\min}^{(\text{train})} + \hat{y}'_i (y_{\max}^{(\text{train})} - y_{\min}^{(\text{train})}) \quad (9)$$

Ridge Regression. Ridge uses an L2-regularized linear objective [10]. In Equation (10), y_i is the normalized target, x_i is the normalized feature vector, β_0 is the unpenalized bias, β is the coefficient vector, and λ controls the regularization strength.

$$\min_{\beta_0, \beta} [\sum_{i=1}^{N_{\text{train}}} (y'_i - \beta_0 - (x'_i)^T \beta)^2 + \lambda \|\beta\|_2^2] \quad (10)$$

The coefficients were obtained by solving the regularized normal-equation system using Gauss-Jordan elimination with pivot selection. The tested λ values are listed in Equation (11).

$$\lambda \in \{10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\} \quad (11)$$

K-Nearest Neighbor Regression. KNN represents a local nonparametric relationship [11]. Equation (12) defines Euclidean distance between normalized observations across D features.

$$d(i, l) = \sqrt{\sum_{j=1}^D (x'_{i,j} - x'_{l,j})^2} \quad (12)$$

The program evaluated uniform averaging and inverse-distance weighting in Equation (13), with prediction defined by Equation (14).

$$w_l = \frac{1}{d(i, l) + 10^{-8}} \quad (13)$$

with prediction

$$\hat{P}_i = \frac{\sum_{l \in \mathcal{N}_k(i)} w_l P_l}{\sum_{l \in \mathcal{N}_k(i)} w_l} \quad (14)$$

For uniform KNN, the weight is set to one. When a query point is numerically identical to a training point (distance $< 10^{-10}$), the associated training target is returned directly. Neighborhood sizes $k = 1, \dots, 7$ in Equation (15) were selected only through inner validation, rather than fixed a priori.

$$k \in \{1, 2, 3, 4, 5, 6, 7\} \quad (15)$$

RBF Network Regression. The RBF Network uses Gaussian radial responses [12]. In Equation (16), c_j is the j -th center and γ controls radial width. Equation (17) combines the M radial responses through regularized output weights.

$$\phi_m(\mathbf{x}'_i) = \exp(-\gamma \|\mathbf{x}'_i - \mathbf{c}_m\|^2) \quad (16)$$

with normalized output

$$\hat{y}'_i = w_0 + \sum_{m=1}^M w_m \phi_m(\mathbf{x}'_i) \quad (17)$$

The first center is the training observation closest to the normalized feature centroid, providing a deterministic representative starting prototype. Subsequent centers are selected by the farthest point procedure that iteratively maximizes the minimum distance to existing centers, thereby increasing training coverage. This deterministic implementation is explicitly treated as a hybrid RBF design rather than a canonical gradient trained RBF network. Output weights are obtained through Ridge Regression in RBF space, with the candidate center counts, gamma values, and regularization values defined in Equations (18)-(20).

$$M \in \{2,3,4,5,6,7,8\} \quad (18)$$

$$\gamma \in \{0.25, 0.50, 1, 2, 4, 8, 16\} \quad (19)$$

and

$$\lambda_{\text{RBF}} \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\} \quad (20)$$

When the requested number of RBF centers exceeded the available inner-training observations, it was capped at the training sample count. For Ridge, KNN, RBF, and Random Forest, the selected hyperparameter set was the candidate with the lowest pooled inner validation RMSE. Tuning was repeated independently for every outer fold and, for Panel-Specific scope, for every module.

Random Forest was added as a modern tree-based comparator [13],[14]. Each candidate used 80 regression trees trained from bootstrap samples. At every node, a random subset of predictors was considered and the split giving the largest reduction in squared error was selected. *Maximum depth* {2,3,4,5}, *minimum leaf size* {2,3}, and *mtry* {2,4,6} were tuned by the same inner daily folds used for the other algorithms. Forest prediction was the arithmetic mean of all tree outputs, and deterministic seeds were used so that repeated execution with the same data and parameter grid produced identical results.

2.2 Nested Leave-One-Day-Out Validation and Training Procedure

A nested leave-one-day-out strategy was used to separate hyperparameter selection from final performance evaluation [15], [16]. The four outer test days were Monday, Tuesday, Thursday, and Friday. In each outer fold, one complete day formed the untouched test set and the remaining three days formed the development set, as defined in Equations (21) and (22).

$$\mathcal{D}_{\text{test}}^{(h)} = \{(\mathbf{x}_i, y_i) : \text{day}_i = h\} \quad (21)$$

$$\mathcal{D}_{\text{dev}}^{(h)} = \frac{\mathcal{D}}{\mathcal{D}_{\text{test}}^{(h)}} \quad (22)$$

The outer fold sample composition is summarized in Table 2. Monday contained 18 test observations and 45 development observations; each other test day contained 15 test observations and 48 development observations.

Table 2. Outer leave-one-day-out configuration

Fold	Test day	Generalized training data	Generalized test data
1	Monday	45	18
2	Tuesday	48	15
3	Thursday	48	15
4	Friday	48	15

For Panel-Specific models, Monday contained 15 training and six test observations per panel, whereas each fold contained 16 training and five test observations per panel. Inside each outer development set, one of the three development days served as inner validation and the other two as inner training; the process rotated until every development day had served once as validation.

Candidate hyperparameters were evaluated by pooled inner-validation RMSE in Equation (23). After selection, the model was retrained on all outer development observations and evaluated on the untouched test day. Feature and target normalization were recalculated using outer training data only. Generalized scopes pooled panels during tuning and training; Panel-Specific models repeated the complete procedure independently for each panel.

$$RMSE_{\text{inner}} = \sqrt{\frac{\sum_{f=1}^3 \sum_{i \in \mathcal{V}_f} (\hat{P}_i - P_i)^2}{\sum_{f=1}^3 |\mathcal{V}_f|}} \quad (23)$$

Each of the 63 observations was predicted once by every combination of three scopes and four machine-learning algorithms, producing $63 \times 3 \times 4 = 756$ outer test prediction records.

2.3 Performance Evaluation, Statistical Analysis, and Outputs

Predictions from the four outer folds were combined for each scope algorithm configuration. RMSE, MAE, R2, range-normalized RMSE, WAPE, and MBE were calculated using Equations (24)-(29), where P_i and $\hat{P}_i$ are the measured and predicted powers, $\bar{P}$ is the evaluation mean, and N is the number of evaluation observations. The physical estimate in Equation (3) was evaluated with the same metrics as a non-ML benchmark.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{P}_i - P_i)^2} \quad (24)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{P}_i - P_i| \quad (25)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_i - \hat{P}_i)^2}{\sum_{i=1}^N (P_i - \bar{P})^2} \quad (26)$$

$$nRMSE = 100 \frac{RMSE}{P_{\text{max}}^{(\text{eval})} - P_{\text{min}}^{(\text{eval})}} \quad (27)$$

$$WAPE = 100 \frac{\sum_{i=1}^N |\hat{P}_i - P_i|}{\sum_{i=1}^N |P_i|} \quad (28)$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (\hat{P}_i - P_i) \quad (29)$$

Positive MBE indicates overall overestimation and negative MBE indicates underestimation. Negative R2 values are retained because they identify cases that perform worse than the corresponding test set mean predictor. Metrics were computed overall, by left-out day, by panel, and for each of the 12 panel day units.

For each of the four algorithms, Generalized-Shared and Generalized-Panel-Aware were compared separately with Panel-Specific modeling. For panel day unit u , the paired error difference in Equation (30) was defined as generalized error minus panel error, so a negative value favored the generalized scope.

$$\Delta_u = E_{\text{generalized},u} - E_{\text{panel-specific},u} \quad (30)$$

Exact two-sided Wilcoxon signed-rank tests were applied to non-zero RMSE and MAE differences across the 12 panel-day units. The analysis reports W^+ and W^- , raw p -values, Holm adjusted p -values, and the rank biserial effect size $rrb = (W^+ - W^-)/(W^+ + W^-)$. Holm correction was applied to the two planned generalized-versus-panel-specific comparisons within each algorithm-metric family [17]. Statistical significance was assessed at 0.05, while effect size was reported because $N = 12$ limits statistical power.

The panel-day combination, rather than each individual measurement, was used as the statistical unit to avoid treating repeated observations from the same panel and day as independent replications. The implementation is deterministic for a fixed dataset and parameter grid: Ridge uses a direct linear-system solution, KNN deterministic distance ordering, RBF deterministic center selection, and Random Forest deterministic pseudo-random seeds. Model generated 13 CSV files containing individual predictions, fold and panel metrics, overall results, selected hyperparameters, all inner-validation candidates, panel-day

scope comparisons, scope summaries, Wilcoxon-Holm statistics with effect size, daily environmental ranges, physical-baseline results, and training-testing sample counts.

3. RESULT AND DISCUSSION

3.1 Overall Comparison of Generalized and Panel-Specific Models

The nested leave-one-day-out experiment produced one out-of-day prediction for every observation, scope, and algorithm. Each of the 12 scope algorithm configurations therefore used the same 63 outer-test observations, while each test day remained excluded from normalization, hyperparameter selection, and training. Table 3 also includes the non-ML physical benchmark.

Table 3. Overall cross-day performance of machine-learning configurations and physical baseline

Scope / benchmark	Model	RMSE (W)	MAE (W)	R2	nRMSE (%)	WAPE (%)	MBE (W)
Generalized-Shared	Ridge	7.7204	6.1052	0.8653	10.9432	18.8307	0.0129
Generalized-Shared	KNN	8.4619	6.4524	0.8381	11.9943	19.9014	0.4471
Generalized-Shared	RBF Network	7.9730	5.8987	0.8563	11.3013	18.1935	0.7550
Generalized-Shared	Random Forest	8.2709	6.3682	0.8454	11.7235	19.6419	1.0810
Generalized-Panel-Aware	Ridge	7.7423	6.1188	0.8645	10.9743	18.8725	0.0154
Generalized-Panel-Aware	KNN	8.4619	6.4524	0.8381	11.9943	19.9014	0.4471
Generalized-Panel-Aware	RBF Network	7.5831	5.8278	0.8700	10.7486	17.9750	-0.7297
Generalized-Panel-Aware	Random Forest	7.7876	5.8662	0.8629	11.0384	18.0934	0.3383
Panel-Specific	Ridge	7.3236	5.6750	0.8788	10.3808	17.5037	-0.0932
Panel-Specific	KNN	9.0820	6.7739	0.8135	12.8733	20.8931	0.5367
Panel-Specific	RBF Network	8.7025	6.7572	0.8288	12.3352	20.8415	1.0674
Panel-Specific	Random Forest	8.8221	6.3875	0.8241	12.5048	19.7013	0.8751
Physical benchmark	FF_STC Voc Isc	18.4161	14.4523	0.2333	26.1038	44.5761	13.0098

Panel-Specific Ridge remained the best overall machine-learning configuration, with RMSE = 7.3236 W, MAE = 5.6750 W, and R2 = 0.8788. Its RMSE was 5.14% lower than Generalized-Shared Ridge and 5.41% lower than Generalized-Panel-Aware Ridge. The physical benchmark produced RMSE = 18.4161 W and R2 = 0.2333. Therefore, the best ML configuration reduced RMSE by 60.23% relative to the transparent datasheet-informed baseline.

The strongest generalized configuration remained Generalized-Panel-Aware RBF (RMSE = 7.5831 W, R2 = 0.8700). The newly added Generalized-Panel-Aware Random Forest was also competitive (RMSE = 7.7876 W, R2 = 0.8629), outperforming its Generalized-Shared and Panel-Specific counterparts. These results show that adding a modern tree-based model does not overturn the best overall result, but it does provide additional evidence that pooled data and explicit panel identity can benefit nonlinear models.

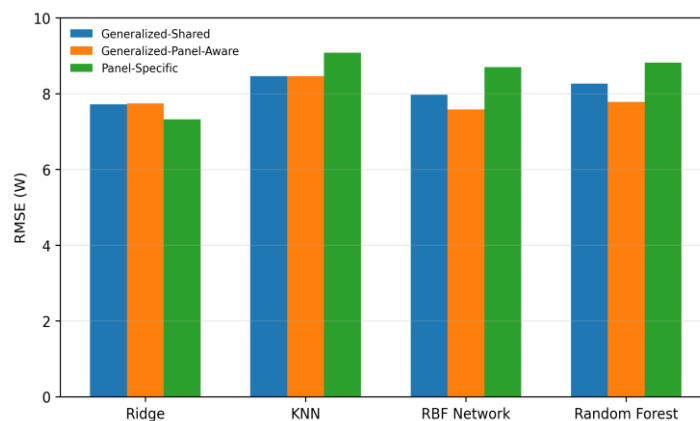


Figure 2. Overall cross-day RMSE across three model scopes and four machine-learning algorithms.

Figure 2 confirms an algorithm-dependent scope effect. Panel-specific fitting improved Ridge, whereas pooled generalized data improved KNN, RBF, and Random Forest relative to their panel-specific versions. Generalized-Panel-Aware was especially useful for RBF and Random Forest, indicating that explicit module identity can help nonlinear pooled models separate panel-dependent response patterns.

The result also illustrates the limited-data bias-variance trade-off. The regularized linear Ridge model can remain stable with only 15-16 development observations per panel, while KNN, RBF, and Random Forest require broader local or partition-based coverage and therefore benefit more often from pooled observations.

3.2 Scope Performance across Panel-Day Units

Overall metrics weight individual observations, whereas the panel-day analysis gives equal weight to each of the 12 panel-day units. Table 4 therefore evaluates whether aggregate scope advantages are distributed consistently across modules and days.

Table 4. RMSE distribution across the 12 panel-day evaluation units

Model	Scope	Mean RMSE (W)	SD (W)	Wins vs. panel-specific	Losses
Ridge	Generalized-Shared	7.0667	3.3495	6	6
Ridge	Generalized-Panel-Aware	7.0896	3.3531	6	6
Ridge	Panel-Specific	6.8609	2.6511	-	-
KNN	Generalized-Shared	7.7098	3.0933	8	4
KNN	Generalized-Panel-Aware	7.7098	3.0933	8	4
KNN	Panel-Specific	7.9838	3.7952	-	-
RBF Network	Generalized-Shared	7.3846	3.2586	7	5
RBF Network	Generalized-Panel-Aware	6.9802	2.9828	6	6
RBF Network	Panel-Specific	8.1728	3.0271	-	-
Random Forest	Generalized-Shared	7.8553	2.8187	5	7
Random Forest	Generalized-Panel-Aware	7.2370	3.0532	8	4
Random Forest	Panel-Specific	8.1797	3.1393	-	-

Panel-Specific Ridge had the lowest mean unit-level RMSE within the Ridge family (6.8610 W), but generalized and panel-specific Ridge each won six of 12 comparisons. KNN showed the opposite pattern: both generalized scopes reduced mean RMSE from 7.9838 W to 7.7098 W and won eight of 12 units. The identical Generalized-Shared and Generalized-Panel-Aware KNN predictions indicate that the added panel indicators did not alter the selected local neighborhoods.

Generalized-Panel-Aware RBF reduced mean unit-level RMSE from 8.1728 W to 6.9802 W. Generalized-Panel-Aware Random Forest similarly reduced mean unit-level RMSE from 8.1797 W to 7.2370 W and won eight of 12 units. Thus, pooled observations were most useful for the data-demanding nonlinear and local methods, whereas the regularized linear model benefited more from panel-specific calibration.

3.3 Panel-Level Performance

Panel-level evaluation revealed different preferred configurations for P1, P3, and P5 after adding Random Forest. Table 5 reports the lowest RMSE configuration for each module across its four outer-test days.

Table 5. Best-performing configuration for each PV module

Panel	Technology	Best configuration	RMSE (W)	MAE (W)	R ²
P1	Polycrystalline	Generalized-Shared RBF Network	7.3015	5.1733	0.8786
P3	Polycrystalline	Generalized-Panel-Aware Random Forest	6.3310	4.9421	0.8330
P5	Monocrystalline	Panel-Specific Ridge	7.1809	5.4894	0.9160

Panel-level preferences were module-dependent. P1 favored Generalized-Shared RBF (RMSE = 7.3015 W), whereas P3 favored the newly added Generalized-Panel-Aware Random Forest (RMSE = 6.3310 W, R² = 0.8330), which improved on Panel-Specific Ridge (6.7600 W). P5 continued to favor Panel-Specific Ridge (RMSE = 7.1809 W, R² = 0.9160), with Generalized-Shared RBF remaining close at 7.2225 W. These outcomes demonstrate panel-dependent model requirements, not a general

technological ranking. P5 is the only monocrystalline module, so the results cannot support a broad monocrystalline-versus-polycrystalline conclusion.

3.4 Day-Specific Generalization and Operating Conditions

Cross-day performance varied with the environmental regime of the left-out day. Table 6 reports the measured irradiance and module-temperature ranges together with the lowest RMSE configuration for each outer fold.

Table 6. Operating conditions and best configuration for each left-out test day

Test day	Operating range (G; T)	Best model	RMSE (W)	MAE (W)	R ²
Monday	G 581.1-1009.0 W/m ² ; T 42.3-52.9 °C	Generalized-Shared Ridge	6.8725	5.6959	0.6330
Tuesday	G 89.4-965.0 W/m ² ; T 28.9-56.9 °C	Generalized-Panel-Aware Random Forest	4.9565	4.1655	0.9113
Thursday	G 93.2-1049.0 W/m ² ; T 30.4-58.9 °C	Panel-Specific KNN	6.7343	5.8583	0.9013
Friday	G 206.5-1052.0 W/m ² ; T 33.1-57.9 °C	Panel-Specific RBF Network	4.7645	3.2369	0.9490

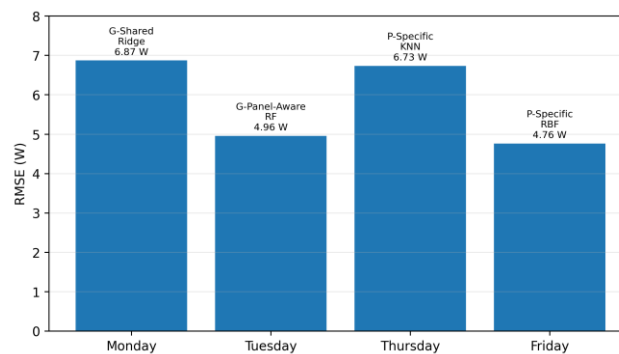


Figure 3. Lowest RMSE configuration for each left-out measurement day.

The best configuration changed across all four days: Generalized-Shared Ridge on Monday, Generalized-Panel-Aware Random Forest on Tuesday, Panel-Specific KNN on Thursday, and Panel-Specific RBF on Friday. The result directly demonstrates that algorithm and scope selection depended on the operating conditions represented by the held-out day.

Monday was particularly difficult for KNN. Generalized KNN produced RMSE = 11.8079 W and R² = -0.0834, while Panel-Specific KNN reached RMSE = 13.9014 W and R² = -0.5016. Monday itself was concentrated at moderate-to-high irradiance (581.1-1009.0 W/m²; mean 843.2 W/m²) and temperatures of 42.3-52.9 °C. Because the other development days collectively spanned these extrema, the negative KNN result should not be attributed to a simple out-of-range irradiance value. A more defensible interpretation is limited local multivariate density and neighborhood representativeness in the normalized feature space.

Tuesday covered the broadest low-to-high irradiance transition (89.4-965.0 W/m²). Generalized-Panel-Aware Random Forest achieved the best fold RMSE of 4.9565 W. Thursday also contained a broad irradiance range (93.2-1049.0 W/m²), where Panel-Specific KNN was best at 6.7343 W. Ridge predictions overestimated Tuesday by more than 2.6 W on average but underestimated Thursday by approximately 4.2-4.7 W, showing that near-zero pooled Ridge bias partly reflects cancellation between days.

Friday favored nonlinear panel-specific modeling: Panel-Specific RBF achieved RMSE = 4.7646 W and R² = 0.9490, while Generalized-Panel-Aware RBF reached 5.3457 W. Together with the Random Forest result on Tuesday, this confirms that nonlinear model suitability changed with both panel identity and the left-out environmental regime.

3.5 Hyperparameter Selection, Model Complexity, and Bias-Variance Considerations

Nested inner validation selected relatively stable generalized-model parameters. Generalized Ridge chose $\lambda = 1$ in three of four folds and 0.1 once; generalized KNN used inverse-distance weighting throughout and usually selected small neighborhoods. Both generalized RBF scopes selected eight centers

in all folds, while Random Forest selected moderate tree depths and leaf sizes that varied with the held-out day.

Panel-specific hyperparameters were more variable because each development set contained only 15-16 observations per module. This pattern is consistent with a bias-variance trade-off: pooling stabilizes local/nonlinear estimation through wider feature-space coverage, while panel-specific fitting can be advantageous when a simpler panel-level relationship is sufficient.

3.6 Statistical Comparison and Diagnostic Observations

Exact Wilcoxon signed-rank tests compared each generalized scope with its panel-specific counterpart across the 12 panel-day units. Table 7 reports W^+ , W^- , rank-biserial effect size, and Holm-adjusted p -values for RMSE and MAE.

Model	Metric	Shared vs. Panel-Specific	Panel-Aware vs. Panel-Specific
Ridge	RMSE	$W^+=42; W^-=36; r_{rb}=0.077; p=0.850; p_H=1.000$	$W^+=43; W^-=35; r_{rb}=0.103; p=0.791; p_H=1.000$
Ridge	MAE	$W^+=42; W^-=36; r_{rb}=0.077; p=0.850; p_H=1.000$	$W^+=42; W^-=36; r_{rb}=0.077; p=0.850; p_H=1.000$
KNN	RMSE	$W^+=31; W^-=47; r_{rb}=-0.205; p=0.569; p_H=1.000$	$W^+=31; W^-=47; r_{rb}=-0.205; p=0.569; p_H=1.000$
KNN	MAE	$W^+=31; W^-=47; r_{rb}=-0.205; p=0.569; p_H=1.000$	$W^+=31; W^-=47; r_{rb}=-0.205; p=0.569; p_H=1.000$
RBF Network	RMSE	$W^+=30; W^-=48; r_{rb}=-0.231; p=0.519; p_H=0.848$	$W^+=28; W^-=50; r_{rb}=-0.282; p=0.424; p_H=0.848$
RBF Network	MAE	$W^+=30; W^-=48; r_{rb}=-0.231; p=0.519; p_H=1.000$	$W^+=30; W^-=48; r_{rb}=-0.231; p=0.519; p_H=1.000$
Random Forest	RMSE	$W^+=35; W^-=43; r_{rb}=-0.103; p=0.791; p_H=0.791$	$W^+=20; W^-=58; r_{rb}=-0.487; p=0.151; p_H=0.303$
Random Forest	MAE	$W^+=37; W^-=41; r_{rb}=-0.051; p=0.910; p_H=0.910$	$W^+=20; W^-=58; r_{rb}=-0.487; p=0.151; p_H=0.303$

No comparison reached the 0.05 significance threshold. The Ridge effects were very small ($|r_{rb}| \leq 0.103$), KNN effects were small-to-moderate in favor of generalized training ($r_{rb} = -0.205$), and panel-aware RBF produced $r_{rb} = -0.282$ for RMSE. The strongest effect occurred for Generalized-Panel-Aware Random Forest versus Panel-Specific Random Forest ($r_{rb} = -0.487$ for both RMSE and MAE), although its Holm adjusted p -value remained 0.3027.

Accordingly, non-significance should not be interpreted as numerical equivalence. The sample contains only 12 statistical units, so the evidence supports algorithm and condition dependent tendencies rather than universal superiority of one scope.

3.7 Diagnostic Observations

Five observations were flagged because loaded current exceeded short-circuit current or loaded voltage was not lower than open-circuit voltage. They remained in the main experiment and were not treated as confirmed faults. For the algorithm-specific best scopes, flagged-observation RMSE was markedly higher than non-flagged RMSE: 13.0684 versus 6.5982 W for Panel-Specific Ridge, 14.7478 versus 7.6829 W for Generalized-Shared KNN, 12.9011 versus 6.9363 W for Generalized-Panel-Aware RBF, and 15.2195 versus 6.7754 W for Generalized-Panel-Aware Random Forest. The flagged rows generally produced approximately twice the error of the remaining observations and tended to be underestimated. Because only five observations were flagged and no independent fault verification was available, this result is treated only as a diagnostic sensitivity finding.

3.8 Practical Implications and Limitations

Generalized and panel-specific modeling provided different advantages. Panel-Specific Ridge was strongest overall, whereas KNN, RBF, and Random Forest generally benefited from pooled observations. Explicit panel identity was useful mainly for the nonlinear models: it produced little change for Ridge, no prediction change for KNN, but improved both RBF and Random Forest relative to their shared generalized counterparts. The conclusions are bounded by 63 observations, three modules, four clear dry-season days, one site, and one monocrystalline module. Panel-specific development sets contained only 15-16 observations per panel, and the Wilcoxon analysis used only 12 panel-day units. Measurements were manually recorded from the SCC-MPPT display; separate sensor accuracy/resolution and automatic data logging were not available. The physical benchmark is therefore a simple datasheet-informed reference rather than a fully identified single-diode model.

Feature structures were adapted to each scope because module-level descriptors that are constant within one panel provide no within-panel predictive information. Broader validation should therefore include more modules, longer seasonal and multi-site campaigns, independently calibrated environmental measurements, and direct I-V characterization. Within the present limits, the evidence supports panel-specific regularized models for stable panel-level relationships and generalized training for more data-demanding nonlinear/local models.

4. CONCLUSION

Nested leave-one-day-out validation of 63 observations from three 100 W PV modules showed that modeling scope interacts strongly with algorithm type. Panel-Specific Ridge achieved the best overall machine-learning performance (RMSE = 7.3236 W, MAE = 5.6750 W, $R^2 = 0.8788$) and reduced RMSE by 60.23% relative to the datasheet-informed physical benchmark (18.4161 W). Pooled data benefited the more data-demanding methods: generalized KNN outperformed Panel-Specific KNN, Generalized-Panel-Aware RBF reached RMSE = 7.5831 W, and Generalized-Panel-Aware Random Forest reached 7.7876 W and became the best model for P3 and Tuesday. The best day-specific configuration changed across all four outer folds, confirming operating-condition dependence. Wilcoxon-Holm tests found no statistically significant universal scope advantage across 12 panel-day units, although the largest observed effect favored Generalized-Panel-Aware Random Forest. These findings support a conditional strategy rather than a universally preferred scope. Generalization remains limited by the small single-site, four-day dataset; future work should include more modules, longer seasonal and multi-site measurements, calibrated sensors, and independent I-V characterization.

AUTHOR CONTRIBUTIONS

G.D.P.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing. A.R.: Investigation, Validation, Resources, Supervision, Writing – review & editing. All authors have read and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICAL CLEARANCE STATEMENT

This study did not involve human participants, personal data, or animals; therefore, ethical clearance was not required.

DECLARATION OF AI TOOL USAGE

OpenAI ChatGPT (GPT-5 series, OpenAI) was used as an AI-assisted tool for language refinement, editorial assistance, manuscript organization, and improving the clarity of several sections during manuscript preparation and revision. It was not used to generate experimental data or replace the authors' scientific analysis and interpretation. All AI-assisted outputs were critically reviewed, verified, and revised by the authors. The authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

REFERENCE

- [1] L.-G. Hua, Q. A. Memon, M. F. Shaikh, S. A. Shaikh, R. A. Rahimoon, S. H. H. Shah, and A. Qadir, "Comparative analysis of power output, fill factor, and efficiency at fixed and variable tilt angles for polycrystalline and monocrystalline photovoltaic panels – The case of Sukkur IBA University," *Energies*, vol. 15, no. 11, Art. no. 3917, 2022, doi: 10.3390/en15113917.
- [2] A. Mellit, A. M. Pavan, E. Ogliari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," *Applied Sciences*, vol. 10, no. 2, Art. no. 487, 2020, doi: 10.3390/app10020487.
- [3] S. Theocharides, M. Theristis, G. Makrides, M. Kynigos, C. Spanias, and G. E. Georghiou, "Comparative analysis of machine learning models for day-ahead photovoltaic power production forecasting," *Energies*, vol. 14, no. 4, Art. no. 1081, 2021, doi: 10.3390/en14041081.
- [4] E. Sauter, M. A. Mughal, and Z. Zhang, "Evaluation of machine learning methods on large-scale spatiotemporal data for photovoltaic power prediction," *Energies*, vol. 16, no. 13, Art. no. 4908, 2023, doi: 10.3390/en16134908.
- [5] M. Alrashidi and S. Rahman, "Short-term photovoltaic power production forecasting based on novel hybrid data-driven models," *Journal of Big Data*, vol. 10, Art. no. 26, 2023, doi: 10.1186/s40537-023-00706-7.
- [6] S. M. Radhi, S. D. Al-Majidi, M. F. Abbod, and H. S. Al-Raweshidy, "Machine learning approaches for short-term photovoltaic power forecasting," *Energies*, vol. 17, no. 17, Art. no. 4301, 2024, doi:

- 10.3390/en17174301.
- [7] Y. Nie, Q. Paletta, A. Scott, L. M. Pomares, G. Arbod, S. Sgouridis, J. Lasenby, and A. Brandt, "Sky image-based solar forecasting using deep learning with heterogeneous multi-location data: Dataset fusion versus transfer learning," *Applied Energy*, vol. 369, Art. no. 123467, 2024, doi: 10.1016/j.apenergy.2024.123467.
 - [8] J. Schreiber and B. Sick, "Model selection, adaptation, and combination for transfer learning in wind and photovoltaic power forecasts," *Energy AI*, vol. 14, Art. no. 100249, 2023, doi: 10.1016/j.egyai.2023.100249.
 - [9] A. Hooshmand and R. Sharma, "Energy predictive models with limited data using transfer learning," in *Proc. 10th ACM Int. Conf. Future Energy Systems*, 2019, pp. 12–16, doi: 10.1145/3307772.3328284.
 - [10] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970, doi: 10.1080/00401706.1970.10488634.
 - [11] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.
 - [12] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1988.
 - [13] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
 - [14] R. Banik and A. Biswas, "Improving solar PV prediction performance with RF-CatBoost ensemble: A robust and complementary approach," *Renewable Energy Focus*, vol. 46, pp. 207–221, 2023, doi: 10.1016/j.ref.2023.06.009.
 - [15] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, Art. no. 91, 2006, doi: 10.1186/1471-2105-7-91.
 - [16] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Aroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017, doi: 10.1111/ecog.02881.
 - [17] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
 - [18] D. E. C. Barragán, B. A. A. Acurio, J. López, H. Morais, C. P. Guzman and L. Silva, "Day-Ahead Photovoltaic Power Forecasting with Limited Data," 2024 *IEEE URUCON*, Montevideo, Uruguay, 2024, pp. 1-5, doi: 10.1109/URUCON63440.2024.10850063.
 - [19] Y. Wang, W. Fu, J. Wang, Z. Zhen, and F. Wang, "Ultra-short-term distributed PV power forecasting for virtual power plant considering data-scarce scenarios," *Applied Energy*, vol. 373, Art. no. 123890, 2024, doi: 10.1016/j.apenergy.2024.123890.
 - [20] W. Zheng, H. Xiao, and W. Pei, "Distributed-regional photovoltaic power generation prediction with limited data: A robust autoregressive transfer learning method," *Applied Energy*, vol. 380, Art. no. 125058, 2025, doi: 10.1016/j.apenergy.2024.125058.
 - [21] G. D. Prenata and A. Ridhoi, "Analisis perbandingan kinerja KNN Regression dan Support Vector Regression dalam prediksi kehandalan sistem tenaga listrik berdasarkan indeks SAIDI-SAIFI," *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 8, no. 1, pp. 139–150, 2026, doi: 10.32528/elkom.v8i1.5213.
 - [22] G. D. Prenata and A. Ridho'i, "Breakdown voltage prediction of transformer oil under copper/iron contamination using SVR and KNN regression: A baseline study on small experimental data," *El Sains: Jurnal Elektro*, vol. 8, no. 1, pp. 81–87, 2026, doi: 10.30996/elsains.v8i1.133785.
 - [23] G. D. Prenata, "Evaluasi komparatif DFFNN, DFFNN yang dioptimasi GA, DFFNN yang dioptimasi GWO, dan LSTM untuk prediksi konsumsi energi listrik menggunakan sliding window cross-validation," *Techné: Jurnal Ilmiah Elektroteknika*, vol. 25, no. 1, pp. 91–104, 2026, doi: 10.31358/techné.v25i1.656.
 - [24] G. D. Prenata, "Predicting breakdown voltage of transformer oil under copper/iron contamination: A comparative study of gradient vs metaheuristic training," *ELKHA: Jurnal Teknik Elektro*, vol. 18, no. 1, pp. 65–73, 2026, doi: 10.26418/elkha.v18i1.101026.