

Operational Data Quality Outweighs Algorithm Choice in Machine-Learning Solar Power Prediction

Deepak Painuli*, Divya Mishra

Department of Computer Science and Engineering, COER University, Roorkee, Uttarakhand, India.
Email's: prof.deepak.painuli@gmail.com; divya19aug@gmail.com

Abstract

Machine-learning models for photovoltaic (PV) power prediction are usually compared on a single cleaned dataset, and the resulting rankings are read as properties of the algorithms. This study tests that reading on two grid-connected PV plants in India observed over the same 34 days (15 May to 17 June 2020), using 15-min inverter-level AC output and plant weather sensors from a public dataset. Plant-level targets were built from fixed inverter sets without imputation, and a data-quality audit flagged timestamps at which any inverter reported zero output while irradiance exceeded 0.1 kW m^{-2} . Two tasks were evaluated on the final 10 days after training on the first 24: nowcasting power from same-time sensor readings, and forecasting 15 to 120 min ahead from information available at issue time. Multiple linear regression, support vector regression, random forest, XGBoost, LightGBM and a multilayer perceptron were tuned by day-blocked forward-chaining cross-validation and compared with persistence and smart persistence using normalized RMSE (nRMSE), skill scores and Harvey–Leybourne–Newbold-corrected Diebold–Mariano tests. In nowcasting, the six learners reached 4.7–6.2% nRMSE on Plant 1 ($R^2 \geq 0.989$) but 29.2–37.8% on Plant 2, where 35.0% of daylight timestamps carried the zero-output flag against 1.7% on Plant 1. Excluding flagged test timestamps lowered Plant 2 nowcasting error to 22.6–27.8% without closing the gap. At 1 h ahead the learners improved on persistence by 30.0–32.1% (Plant 1) and 21.3–27.3% (Plant 2); only one pairwise difference from the best learner was significant at the 5% level, and at 15 min the gain over persistence fell to 3.3–10.8%. SHAP attribution and feature ablation placed most nowcasting information in irradiance and most 1-h forecasting information in the diurnal profile. Models transferred between plants were biased by +24% and –29% in nowcasting. In this record, learner choice moved nRMSE by a few points while inverter operating condition moved it by more than 20. Reporting data-quality audits and persistence-referenced skill alongside model rankings would make such comparisons more transferable across plants.

Index Terms—Photovoltaic power forecasting, machine learning, gradient boosting, support vector regression, data quality, inverter anomalies, skill score, Diebold–Mariano test, SHAP

Introduction

Short-term prediction of photovoltaic (PV) output supports unit commitment, reserve scheduling and market bidding, and its value rises as PV takes a larger share of supply [1], [2]. Horizons of interest run from seconds to days, and the method families differ accordingly: sky imagery and cloud tracking for intra-hour horizons, satellite and numerical weather prediction (NWP) inputs for hours to days, and statistical or machine-learning (ML) models that map measured or forecast predictors to plant output across all of these ranges [1], [3]–[5]. Operational use also brings requirements beyond accuracy, including standardized verification and forecast submission rules set by grid operators [6], [7].

Supervised ML has been applied to PV power with support vector machines [8]–[10], tree ensembles and gradient boosting [11], [12], neural networks and recurrent architectures [13]–[16], and hybrids of physical model chains with learned components [17]–[19]. Several reviews catalogue these families and their reported accuracies [20]–[24]. A recurring observation in that literature is that rankings obtained on one dataset do not generalize well. Voyant et al. noted that differences in data set, time step, horizon and error measures make performance ranking difficult and that the prediction errors of the main ML families are broadly similar [25]. Large controlled comparisons point the same way: on 16 Hungarian plants, the choice of physical model chain changed RMSE by about 12% [26], and among 24 ML models driven by NWP, input selection alone changed RMSE by 13.1% [27].

Two aspects receive less attention in single-plant ML comparisons. First, verification practice varies: many studies report R^2 or percentage errors without a naïve reference forecast or a test of whether differences between models exceed sampling noise, although both are recommended for solar forecasts [28]–[30]. Second, the operating condition of the plant is usually handled by cleaning the data before modeling, so the effect of inverter outages, communication gaps or curtailment on the reported error is not measured. When such events are frequent, the error a model attains may reflect the plant record more than the algorithm.

This paper addresses both points with a controlled experiment on a public dataset from two grid-connected PV plants in India [31]. The plants share the observation period and data format but differ in operating condition, which allows the contribution of data integrity to be separated from that of the learning algorithm. The contributions are as follows.

- 1) A data-quality audit of the two plant records that identifies a scale error in one DC channel, incomplete inverter reporting, and frequent zero-output events under high irradiance, together with a target construction that avoids imputation.
- 2) A comparison of six learners and two naïve references on nowcasting and on 15–120 min forecasting, with day-blocked tuning, five-seed repetition, skill scores and HLN-corrected Diebold–Mariano tests.
- 3) Evidence, from anomaly-screened evaluation, retraining and cross-plant transfer, that the between-plant difference in error exceeds the between-learner difference by roughly an order of magnitude in this record.
- 4) Feature attribution and ablation that identify which inputs carry the predictive information for each task, and the released analysis code that regenerates every table and figure.

Section II reviews related work and sets out the evaluation design. Section III describes the data and the audit, Section IV the methods, Section V the results, and Section VI discusses implications and limitations. Section VII concludes.

TABLE I ABBREVIATIONS AND SYMBOLS

Term	Meaning
PV	Photovoltaic
AC / DC	Alternating / direct current
IRR	Plant irradiation sensor reading (kW m^{-2})
$T_{\text{amb}}, T_{\text{mod}}$	Ambient and module temperature ($^{\circ}\text{C}$)
Task N	Nowcasting: $P(t)$ from same-time sensors
Task F	Forecasting: $P(t+h)$ from data known at t
h	Horizon in 15-min steps (1–8)
MLR	Multiple linear regression
SVR	Support vector regression
RF	Random forest
XGB	XGBoost gradient boosting
LGBM	LightGBM gradient boosting
MLP	Multilayer perceptron
RMSE / MAE / MBE	Root mean square / mean absolute / mean bias error
nRMSE	RMSE as % of mean observed power
SS	Skill score relative to persistence (%)
DM / HLN	Diebold–Mariano test / Harvey–Leybourne–Newbold correction
SHAP	Shapley additive explanations
$A(t)$	Zero-output anomaly flag

II. RELATED WORK AND EVALUATION DESIGN

A. Physical, statistical and machine-learning approaches

Physical approaches convert forecast irradiance and temperature to power through a model chain of decomposition, transposition, temperature and inverter models; their accuracy depends on the chain chosen and on plant metadata such as orientation and inverter sizing [26]. Data-driven approaches learn the mapping from historical records and

do not need that metadata, which suits plants where it is unavailable. Hybrid schemes that feed physical model outputs to an ML regressor have been reported to combine the strengths of the two [17], [18]. For intra-hour horizons, reviews of cloud-motion methods indicate that local telemetry and sky imagery carry information that NWP cannot resolve at that scale [3].

B. Machine-learning comparisons for PV power

Support vector regression (SVR) with weather classification was an early data-driven approach to day-ahead PV power [8], weather-based hybrid methods were proposed for day-ahead hourly output [32], and later work paired SVR with satellite images [9] or with wavelet decomposition and particle-swarm tuning on SCADA data [10]. Tree ensembles appear in intra-hour and hourly settings, often with a smart-persistence component that normalizes power by a clear-sky or climatological profile [12], [33], [34]; gradient boosting with nearest-neighbor post-processing performed well in the GEFCom2014 solar track [11], [35]. Deep recurrent and convolutional models have been compared for day-ahead and short-term forecasting [13]–[16]. Case studies of individual solar farms report accurate one-hour-ahead forecasts from tree-based models [36], [37], and hyperparameter optimization is repeatedly identified as a condition for realizing a learner's accuracy [27], [38]. Irradiance-forecasting studies add that the preferred model can change with weather variability [39], [40].

C. Verification practice

Solar forecast verification guidance recommends reporting errors against a reference such as persistence or a persistence–climatology combination, so that skill can be compared across sites with different variability [28], [29]. The choice of normalizing quantity for smart persistence, clear-sky index or clearness index, changes the reference itself [41]. Scale-dependent and relative measures answer different questions [42], and pairwise tests of equal predictive accuracy [43], [44] indicate whether a ranking is distinguishable from sampling variation. Random cross-validation of time series leaks future information into training, and blocked or forward-chaining schemes are preferred [45]. Post-processing and probabilistic formulations extend these ideas [46], [47] but are outside the scope of this study. Table II lists the design choices adopted here and their basis.

TABLE II EVALUATION DESIGN CHOICES AND THEIR BASIS

Design choice	Basis in literature	Implementation here
Reference forecast	Skill against a naïve reference [28], [29]	Persistence and smart persistence; skill score (13)
Time-ordered validation	Leakage in shuffled CV of series [45]	Day-blocked forward-chaining CV; chronological test
Error measures	Scale-dependent vs. relative measures [42]	RMSE, MAE, MBE in MW; nRMSE; R^2
Significance of differences	Equal-accuracy tests [43], [44]	HLN-corrected DM on squared errors
Hyperparameter tuning	Tuning needed to realize model accuracy [27], [38]	Grid search per plant and task
Weather dependence	Model skill varies with sky conditions [39]	Errors by daily insolation class
Attribution	Tree SHAP [48], [49]	Mean SHAP share per feature
Data integrity	Rarely quantified in model comparisons	Inverter zero-output flag; screened evaluation

III. DATA AND PROBLEM FORMULATION

A. Dataset

The Kaggle Solar Power Generation Data [31] contain records from two grid-connected PV plants in India, labeled Plant 1 (identifier 4135001) and Plant 2 (4136001), for 15 May to 17 June 2020. Each plant has a generation file with DC power, AC power, daily yield and total yield for each of 22 inverters at 15-min resolution, and a weather file with ambient temperature, module temperature and an irradiation reading from a single plant-level sensor array. Row counts of the retrieved files (68,778 and 67,698 generation rows; 3,182 and 3,259 weather rows) match the published release. The source does not document plant location, capacity, module technology or units; power values are used as recorded and interpreted as kW, and irradiation values (maximum 1.22) are interpreted as kW m^{-2} . Table III summarizes the processed data and Fig. 1 shows daily energy and insolation.

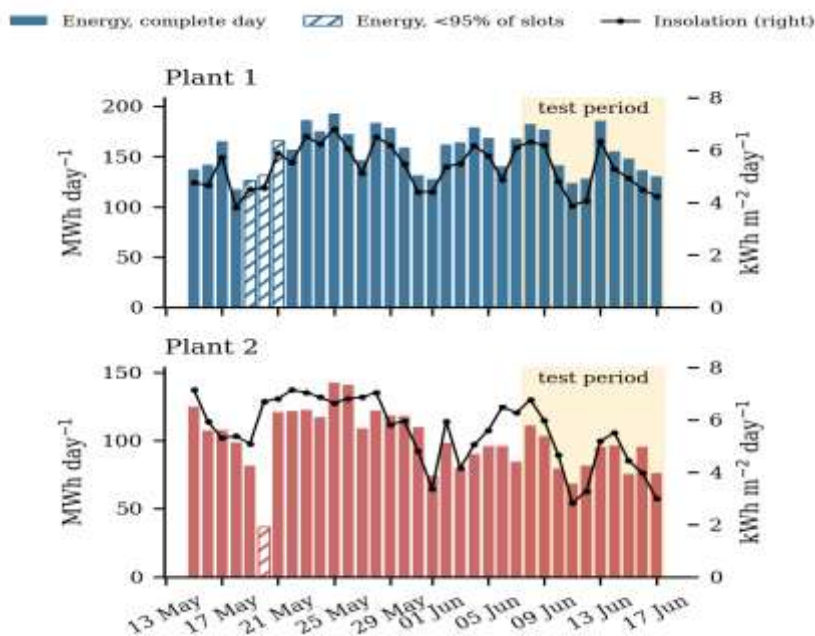


Fig. 1. Daily energy and insolation for both plants.

TABLE III DATASET SUMMARY AFTER TARGET CONSTRUCTION

Quantity	Plant 1	Plant 2
Inverters in source / used	22 / 22	22 / 18
Complete 15-min timestamps	3,057 of 3,158	3,195 of 3,259
Daylight slots per day	49	51
Daylight samples (train / test)	1,134 / 488	1,187 / 510
Mean / max daylight power (MW)	13.03 / 29.15	8.02 / 21.16
Mean / max IRR (kW m^{-2})	0.447 / 1.222	0.435 / 1.099
T_{amb} range ($^{\circ}\text{C}$)	20.8–35.3	21.6–39.2
T_{mod} range ($^{\circ}\text{C}$)	18.7–65.5	21.0–66.6
Mean daily insolation (kWh m^{-2})	5.34	5.58
Pearson r (power, IRR)	0.991	0.832
Median DC/AC ratio	10.22	1.02

Quantity	Plant 1	Plant 2
Flagged daylight samples: all / train / test (%)	1.7 / 1.1 / 3.1	35.0 / 39.7 / 24.1
Unflagged AC per inverter per unit IRR (kW per kW m ⁻²)	1,343	1,392

Power is the AC sum over the fixed inverter set, as recorded in the source files.

B. Data-quality audit

Three issues were found. In Plant 1 the median ratio of DC to AC power during generation is 10.22, against 1.02 in Plant 2; since inverter efficiency cannot exceed unity, the Plant 1 DC channel carries a scale error and only AC power is used in this study. Inverter reporting is complete for 3,057 of 3,158 Plant 1 timestamps, but in Plant 2 four inverters report at only 72.3% of timestamps, while the other 18 report at 98% or more. Finally, individual inverters report zero AC output during periods of high irradiance. Define $G(t)$ as the irradiation reading and $\mathcal{I}$ as the fixed inverter set; the flag

$$A(t) = \mathbf{1}[\exists i \in \mathcal{I}: P_i(t) = 0 \wedge G(t) > 0.1] \quad (1)$$

marks timestamps at which at least one inverter in the set is at zero while irradiance exceeds 0.1 kW m⁻². The flag holds for 1.7% of Plant 1 daylight samples and 35.0% of Plant 2 daylight samples. Across all 22 Plant 2 inverters, 263 of 716 inverter-days at irradiance above 0.3 kW m⁻² show zero output in 10–90% of intervals, a pattern of intermittent rather than day-long outages. The record does not indicate whether these events are equipment faults, curtailment or communication loss, so the flag is treated as an operational anomaly indicator rather than a fault label. Fig. 2 plots inverter-mean AC power against irradiance: unflagged samples in both plants follow a tight, nearly linear relation, and flagged Plant 2 samples form a dispersed cloud below it.

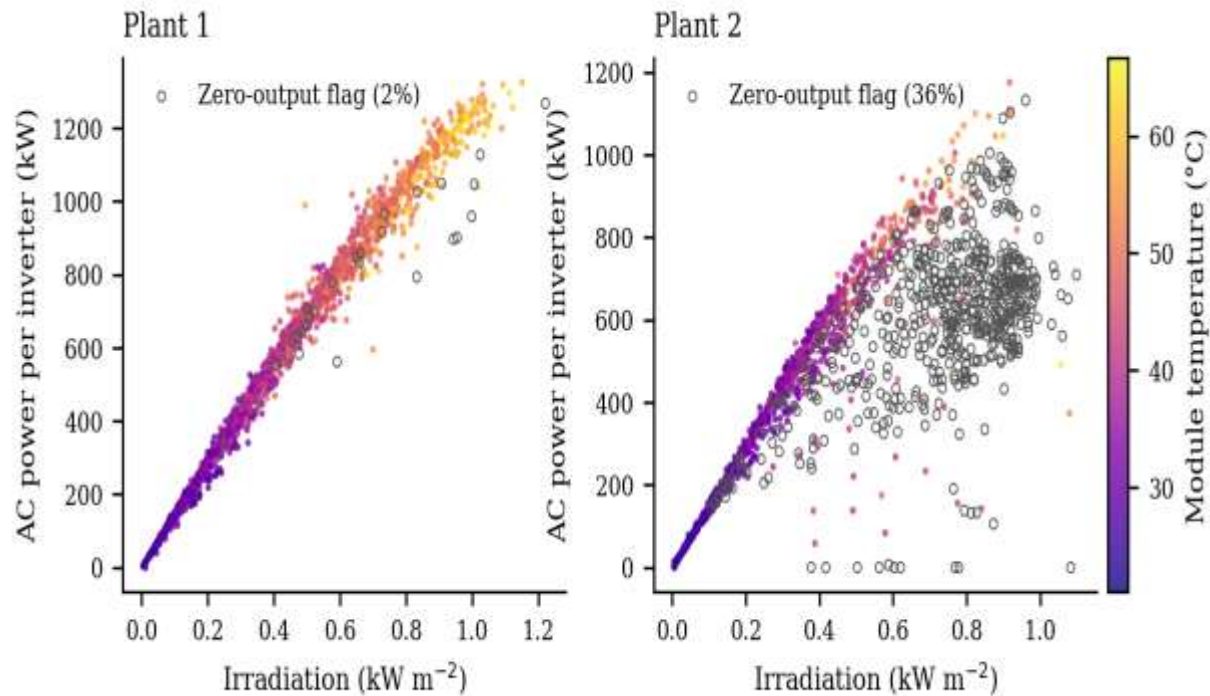


Fig. 2. Inverter-mean AC power against irradiance.

When flagged samples with irradiance above 0.2 kW m⁻² are excluded, the median AC output per inverter per unit irradiance is 1,343 kW per kW m⁻² for Plant 1 and 1,392 for Plant 2. The two plants therefore convert irradiance at similar rates when all inverters operate; the difference between their records lies mainly in how often they do not.

C. Target construction and tasks

The inverter set $\mathcal{I}$ contains all inverters reporting at 95% or more of timestamps: 22 in Plant 1 and 18 in Plant 2. The

plant target is

$$P(t) = \sum_{i \in I} P_i(t), \text{ defined only if all } i \in I \text{ report} \quad (2)$$

placed on a regular 15-min grid; timestamps where any member of I is missing are left undefined and no value is imputed. Evaluation is restricted to daylight slots, defined as the 15-min slots whose mean training-period power exceeds 1% of the maximum slot mean (49 slots per day for Plant 1, 51 for Plant 2). The first 24 days (15 May to 7 June) form the training period and the last 10 days (8 to 17 June) the test period. Two tasks are defined. Task N estimates power from same-time sensor readings,

$$\hat{P}(t) = f(\mathbf{x}(t)) \quad (3)$$

which corresponds to estimating plant output where metering is delayed or unavailable, and to establishing an expected-output reference. Task F forecasts power h steps ahead using only information available at issue time t ,

$$\hat{P}(t+h|t) = g_h(\mathbf{z}(t)), \quad h = 1, \dots, 8 \quad (4)$$

for horizons of 15 to 120 min. No weather forecasts are available in the dataset, so Task F relies on persistence of the recent state and on the diurnal cycle.

IV. METHODOLOGY

Fig. 3 summarizes the workflow from acquisition to diagnostics.

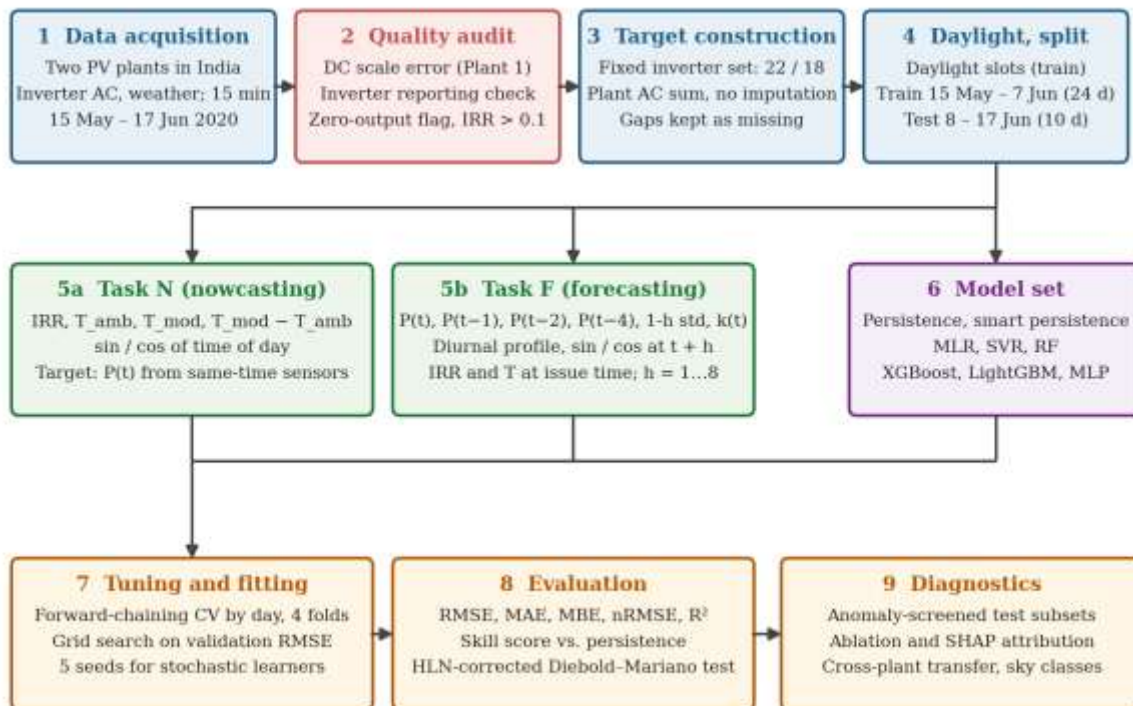


Fig. 3. End-to-end experimental workflow.

A. Features

Table IV lists the inputs. For Task N, the weather readings are complemented by the module-to-ambient temperature difference, which varies with irradiance and wind cooling, and by a cyclical encoding of time of day. For Task F, the diurnal profile $\bar{P}_s$ is the mean training power in 15-min slot s , evaluated at the target slot. Lagged power, its one-hour standard deviation as a variability measure, and the profile-normalized index $k(t)$ describe the state at issue time. $k(t)$ is set to 1 where the profile is below 5% of its maximum and clipped to $[0, 2]$.

TABLE IV INPUT FEATURES BY TASK

Group	Features	Task
Weather	IRR, T_{amb} , T_{mod} , $T_{mod}-T_{amb}$	N
Time	sin and cos of time of day	N
Time / profile	$\bar{P}_{s(t+h)}$; sin and cos of time of day at $t+h$	F
Power lags	$P(t)$, $P(t-1)$, $P(t-2)$, $P(t-4)$; 1-h standard deviation of P ; $k(t)$	F
Weather at issue	IRR(t), IRR($t-1$), $T_{mod}(t)$, $T_{amb}(t)$	F

$\bar{P}_s$: mean training power in 15-min slot s ; $k(t)$ defined in (5).

B. Reference forecasts

Persistence repeats the latest observation,

$$\hat{P}_{pers}(t+h|t) = P(t) \quad (5)$$

and smart persistence carries the current profile-normalized index forward to the target slot,

$$k(t) = \frac{P(t)}{\bar{P}_{s(t)}}, \quad \hat{P}_{sp}(t+h|t) = k(t) \bar{P}_{s(t+h)} \quad (6)$$

which preserves the diurnal shape. The profile is empirical because plant location and orientation are not given, so a clear-sky model could not be configured; this makes the reference a climatological-profile variant of smart persistence [12], [41].

C. Learners

Six regressors from established families were implemented with scikit-learn [50], XGBoost [51] and LightGBM [52]. MLR is ordinary least squares on standardized inputs and serves as the linear benchmark. SVR with a radial basis function kernel fits an ϵ -insensitive loss in a reproducing-kernel space [53]. RF averages 300 decorrelated regression trees grown on bootstrap samples [54]. XGB and LGBM are gradient-boosted tree ensembles [55],[56]; XGB grows level-wise trees with second-order gradient information, and LGBM grows leaf-wise trees on histogram-binned features. The MLP has two hidden layers with ReLU activations, trained with Adam and early stopping on a 15% validation split. SVR and MLP used standardized targets. Predictions were clipped at zero [57],[58].

D. Tuning and repetition

Hyperparameters were selected by grid search within the training period using day-blocked forward-chaining cross-validation [45],[59]: the 12 training days after the first 12 were divided into four consecutive blocks, and each block was predicted by a model trained on all earlier days. The configuration with the lowest mean validation RMSE was refitted on the full training period. For Task F, tuning was done at $h = 4$ and the selected values reused for $h = 1, 2$ and 8 to limit computation. Table V gives the grids and the selected values. RF, XGB, LGBM and MLP were fitted with five random seeds; tables report seed-mean metrics and the standard deviation of RMSE across seeds. Tests and subset analyses use the seed-0 predictions.

TABLE V HYPERPARAMETER SEARCH SPACE AND SELECTED VALUES

Model	Search space	P1, Task N	P1, Task F	P2, Task N	P2, Task F
MLR	—	—	—	—	—
SVR	$C \in \{1, 10, 100\}$; $\epsilon \in \{0.05, 0.1\}$; $\gamma \in \{\text{scale}, 0.05\}$	$C=10$, $\epsilon=0.05$, $\gamma=0.05$	$C=1$, $\epsilon=0.1$, $\gamma=0.05$	$C=1$, $\epsilon=0.1$, $\gamma=\text{scale}$	$C=1$, $\epsilon=0.05$, $\gamma=0.05$
RF	300 trees; $\text{depth} \in \{\infty, 10\}$; $\text{leaf} \in \{1, 5\}$; $\text{feat} \in \{1.0, 0.5\}$	$\text{depth}=\infty$, $\text{leaf}=5$, $\text{feat}=1.0$	$\text{depth}=10$, $\text{leaf}=5$, $\text{feat}=0.5$	$\text{depth}=\infty$, $\text{leaf}=5$, $\text{feat}=0.5$	$\text{depth}=\infty$, $\text{leaf}=5$, $\text{feat}=0.5$

Model	Search space	P1, Task N	P1, Task F	P2, Task N	P2, Task F
XGB	trees $\in\{300,600\}$; $\eta\in\{0.03,0.1\}$; depth $\in\{3,5\}$; subsample 0.8	trees=600, $\eta=0.03$, depth=3	trees=300, $\eta=0.03$, depth=3	trees=300, $\eta=0.03$, depth=3	trees=300, $\eta=0.03$, depth=3
LGBM	trees $\in\{300,600\}$; $\eta\in\{0.03,0.1\}$; leaves $\in\{15,31\}$; mcs $\in\{10,20\}$; bagging 0.8	trees=300, $\eta=0.03$, leaves=31, mcs=20	trees=300, $\eta=0.03$, leaves=15, mcs=20	trees=300, $\eta=0.03$, leaves=15, mcs=20	trees=300, $\eta=0.03$, leaves=15, mcs=10
MLP	layers $\in\{64-32,128-64\}$; $\alpha\in\{10^{-4},10^{-3}\}$; Adam, early stopping	layers=128-64, $\alpha=0.0001$	layers=64-32, $\alpha=0.0001$	layers=64-32, $\alpha=0.0001$	layers=64-32, $\alpha=0.0001$

Task F values were selected at $h = 4$ and reused for $h = 1, 2$ and 8 . SVR and MLP used standardized inputs and target. leaf: min. samples per leaf; feat: fraction of features per split; mcs: min. child samples.

E. Error measures

With n test samples, observed power P_j , prediction $\hat{P}_j$ and test-mean power $\bar{P}$, the measures are

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{P}_j - P_j)^2} \quad (7)$$

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |\hat{P}_j - P_j|, \quad \text{MBE} = \frac{1}{n} \sum_{j=1}^n (\hat{P}_j - P_j) \quad (8)$$

$$\text{nRMSE} = 100 \frac{\text{RMSE}}{\bar{P}}, \quad R^2 = 1 - \frac{\sum_j (\hat{P}_j - P_j)^2}{\sum_j (P_j - \bar{P})^2} \quad (9)$$

and the skill score relative to persistence is

$$\text{SS} = 100 \left(1 - \frac{\text{RMSE}_{\text{model}}}{\text{RMSE}_{\text{pers}}} \right) \quad (10)$$

RMSE, MAE and MBE are reported in MW. nRMSE allows the two plants, which differ in size, to be compared.

F. Statistical comparison

For each plant and task, the learner with the lowest RMSE was compared with every other method using the Diebold–Mariano test on the squared-error loss differential d_j [43], with the Harvey–Leybourne–Newbold small-sample correction [44]:

$$\text{DM} = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}}, \quad \text{DM}^* = \text{DM} \sqrt{\frac{T+1-2h+h(h-1)/T}{T}} \quad (11)$$

where $\hat{V}(\bar{d})$ is estimated from the autocovariances of d up to lag $h-1$ and T is the number of test samples. The corrected statistic is referred to a Student t distribution with $T-1$ degrees of freedom. The test samples come from 10 days with overnight gaps, so they are treated as a single series; this is an approximation.

G. Diagnostics

Five further analyses were run. (i) Anomaly-screened evaluation recomputes errors separately on flagged and unflagged test timestamps, with A evaluated at the target time. (ii) For Plant 2 Task N, MLR, SVR and LGBM were retrained on unflagged training samples only. (iii) Feature-group ablation refits LGBM (five seeds) on subsets of the inputs. (iv) Tree SHAP [48], [49] attributes each LGBM prediction additively to the inputs,

$$\hat{f}(\mathbf{x}) = \phi_0 + \sum_{m=1}^M \phi_m(\mathbf{x}) \quad (12)$$

and the mean $|\phi_m|$ over the test period, expressed as a share of the total, summarizes feature importance. (v) For

cross-plant transfer, targets and power lags were divided by the number of inverters, and an LGBM model trained on one plant was tested on the other. Finally, each test day was assigned to a sky class by its daily insolation relative to the maximum in the record: clear (≥ 0.85), mixed (0.65–0.85) or overcast (< 0.65).

V. RESULTS

A. Task N: nowcasting

Table VI and the left panel of Fig. 4 report the nowcasting results. On Plant 1 all six learners reached nRMSE between 4.74% (SVR) and 6.24% (MLR), with R^2 between 0.989 and 0.994. On Plant 2 the same learners gave 29.22% (LGBM) to 37.77% (MLR), with R^2 between 0.611 and 0.767. The five nonlinear learners fell within 1.5 points of each other on Plant 1 and within 2.0 points on Plant 2. MLR was the least accurate in both plants, by 1.5 points on Plant 1 and by more than 8 points on Plant 2, where the flagged samples introduce structure that a linear model cannot follow. Seed-to-seed standard deviation of RMSE was at most 0.014 MW on Plant 1 and 0.087 MW on Plant 2 (MLP), small relative to the differences between plants.

TABLE VI TASK N RESULTS ON THE 10-DAY TEST PERIOD

Plant	Model	RMSE (MW)	MAE (MW)	MBE (MW)	nRMSE (%)	R^2	SD of RMSE (MW)
Plant 1	MLR	0.771	0.560	-0.142	6.24	0.9893	—
	SVR	0.585	0.400	-0.013	4.74	0.9939	—
	RF	0.691	0.478	-0.120	5.59	0.9914	0.001
	XGB	0.719	0.527	-0.110	5.82	0.9907	0.014
	LGBM	0.665	0.456	-0.070	5.39	0.9921	0.005
	MLP	0.634	0.422	-0.053	5.13	0.9928	0.010
Plant 2	MLR	2.615	1.965	-0.126	37.77	0.6109	—
	SVR	2.049	1.142	0.307	29.59	0.7611	—
	RF	2.067	1.156	0.118	29.86	0.7568	0.003
	XGB	2.161	1.292	0.051	31.21	0.7344	0.016
	LGBM	2.023	1.134	0.026	29.22	0.7671	0.010
	MLP	2.108	1.240	-0.002	30.45	0.7467	0.087

Stochastic learners: mean over 5 seeds; SD across seeds. Bold: lowest nRMSE per plant.

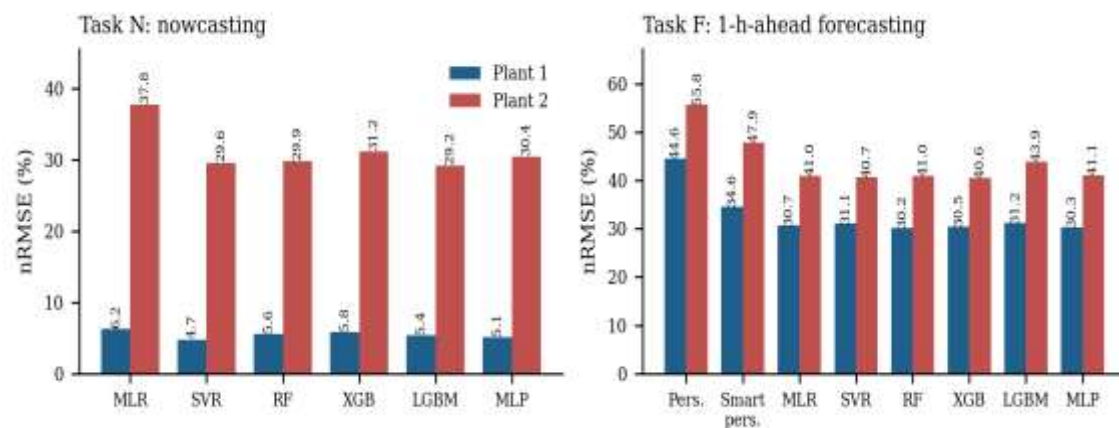


Fig. 4. Test nRMSE by model, task and plant.

Table VIII gives the significance tests. On Plant 1, SVR was more accurate than MLR, RF, XGB and LGBM ($p \leq 0.0014$), but its difference from MLP was not significant at the 5% level ($p = 0.074$). On Plant 2, LGBM differed significantly

from MLR and XGB but not from SVR, RF or MLP ($p \geq 0.27$). Differences of about one point of nRMSE between nonlinear learners were therefore detectable on the more regular Plant 1 record but not on Plant 2.

B. Task F: forecasting

Table VII and Fig. 5 report the forecasting results. Error grew with horizon for every method, but persistence degraded much faster: on Plant 1 its nRMSE rose from 24.9% at 15 min to 66.9% at 120 min, while the learners rose from 23.1–24.0% to 31.6–34.7%. Skill over persistence therefore increased with horizon. At 60 min, the learners improved on persistence by 30.0–32.1% on Plant 1 and by 21.3–27.3% on Plant 2; smart persistence improved on it by 22.4% and 14.1%. At 15 min the learners' skill was 3.3–7.2% on Plant 1 and 7.7–10.8% on Plant 2, and on Plant 1 the best learner (XGB) did not differ significantly from smart persistence ($p = 0.097$).

TABLE VII TASK F RESULTS: NRMSE (%) / SKILL SCORE (%) BY HORIZON

Plant	Model	15 min	30 min	60 min	120 min
Plant 1	Persistence	24.9 / 0	33.3 / 0	44.6 / 0	66.9 / 0
	Smart persistence	24.4 / 1.7	30.5 / 8.4	34.6 / 22.4	38.9 / 41.8
	MLR	23.3 / 6.2	27.8 / 16.5	30.7 / 31.2	33.1 / 50.5
	SVR	23.5 / 5.6	28.3 / 14.9	31.1 / 30.2	34.7 / 48.1
	RF	23.1 / 7.2	27.9 / 16.4	30.2 / 32.1	31.6 / 52.7
	XGB	23.1 / 7.2	28.4 / 14.9	30.5 / 31.6	32.3 / 51.7
	LGBM	23.8 / 4.1	30.0 / 10.1	31.2 / 30.0	33.8 / 49.5
	MLP	24.0 / 3.3	28.0 / 16.0	30.3 / 31.9	33.9 / 49.3
Plant 2	Persistence	36.6 / 0	47.6 / 0	55.8 / 0	73.5 / 0
	Smart persistence	36.1 / 1.4	45.2 / 5.1	47.9 / 14.1	53.0 / 27.9
	MLR	32.8 / 10.4	39.0 / 18.2	41.0 / 26.4	43.5 / 40.9
	SVR	32.6 / 10.8	37.6 / 21.0	40.7 / 27.1	43.3 / 41.0
	RF	33.2 / 9.2	38.1 / 19.9	41.0 / 26.4	43.4 / 41.0
	XGB	33.3 / 9.0	38.5 / 19.2	40.6 / 27.3	43.4 / 41.0
	LGBM	33.8 / 7.7	39.2 / 17.6	43.9 / 21.3	45.0 / 38.7
	MLP	33.6 / 8.3	38.1 / 19.9	41.1 / 26.3	44.8 / 39.0

Skill score relative to persistence. Bold: lowest nRMSE per plant and horizon.

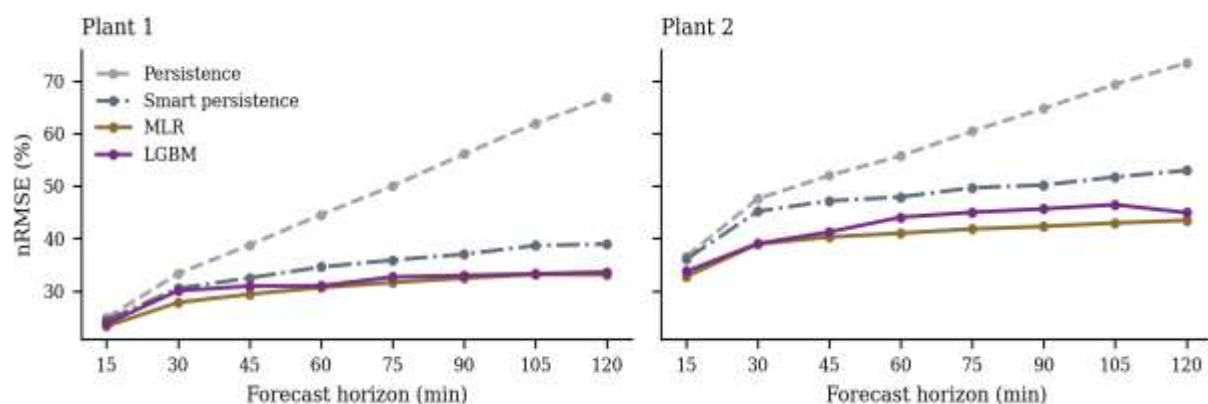


Fig. 5. Forecast error against horizon.

At 60 min, the learners' nRMSE spanned 30.25–31.21% on Plant 1 and 40.57–43.92% on Plant 2. On Plant 1 no learner differed significantly from the best (RF, all $p > 0.13$). On Plant 2 only LGBM was significantly less accurate than the best (XGB, $p < 0.001$); the other four were indistinguishable from it ($p > 0.25$). All learners had positive mean bias at

60 min (0.43–0.84 MW), whereas the two references were nearly unbiased. Fig. 6 shows three test days: on the clear day the Task N predictions follow the observed curve closely for Plant 1, and the 60-min forecasts are smoother than the observations and miss short ramps. On Plant 2, abrupt drops in observed output that are absent from the Task N prediction coincide with flagged timestamps.

TABLE VIII HLN-CORRECTED DIEBOLD–MARIANO TESTS AGAINST THE BEST LEARNER

Case	Best	Compared with	DM	p
P1, N	SVR	MLR	-4.17	<0.001
		RF	-4.54	<0.001
		XGB	-5.31	<0.001
		LGBM	-3.21	0.001
		MLP	-1.79	0.074
P1, F, 15 min	XGB	Persistence	-2.28	0.023
		Smart persistence	-1.66	0.097
		MLR	-0.81	0.421
		SVR	-1.38	0.168
		RF	-0.17	0.868
		LGBM	-2.52	0.012
		MLP	-3.78	<0.001
P1, F, 60 min	RF	Persistence	-6.91	<0.001
		Smart persistence	-2.65	0.008
		MLR	-0.67	0.505
		SVR	-1.21	0.228
		XGB	-0.61	0.541
		LGBM	-1.51	0.132
		MLP	-0.82	0.412
P2, N	LGBM	MLR	-7.15	<0.001
		SVR	-0.48	0.629
		RF	-1.10	0.272
		XGB	-3.37	<0.001
		MLP	0.78	0.436
P2, F, 15 min	SVR	Persistence	-3.92	<0.001
		Smart persistence	-3.49	<0.001
		MLR	-0.33	0.744
		RF	-0.91	0.366
		XGB	-1.05	0.296
		LGBM	-1.95	0.052
		MLP	-1.00	0.316
P2, F, 60 min	XGB	Persistence	-5.38	<0.001
		Smart persistence	-3.01	0.003
		MLR	-0.61	0.545

Case	Best	Compared with	DM	p
		SVR	-0.11	0.909
		RF	-0.70	0.481
		LGBM	-5.43	<0.001
		MLP	-1.15	0.252

Squared-error loss; negative DM favors the best learner; two-sided p from Student t with T-1 degrees of freedom. Seed-0 predictions.

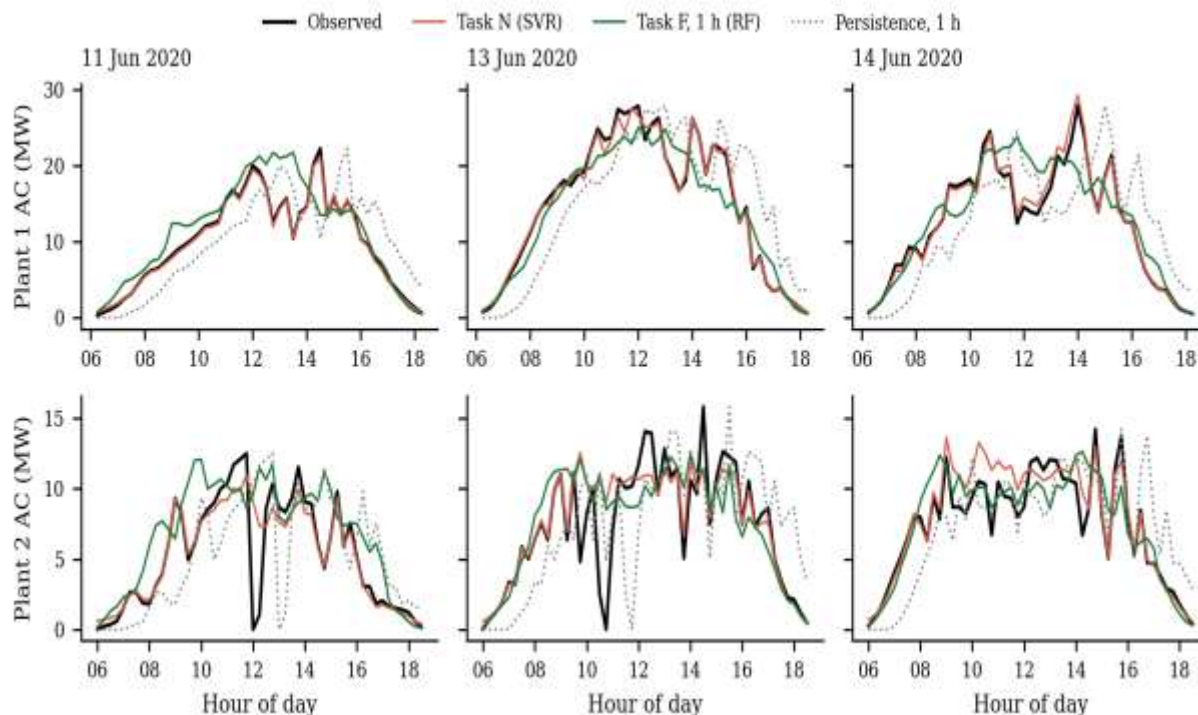


Fig. 6. Observed and predicted power on three test days.

C. Effect of the anomaly flag

Table IX and Fig. 7 separate the Task N errors by the flag. For Plant 2, removing the 123 flagged timestamps from the 510 test timestamps lowered nRMSE from 28.8–31.5% to 22.6–27.8% for the nonlinear learners (MLR: 37.8% to 35.3%). Errors on flagged timestamps were 32.9–39.7%, and the MBE on those timestamps was +1.30 to +2.46 MW for every learner, meaning that models overpredicted when part of the plant was not producing. The effect on Plant 1 was small because only 15 test timestamps were flagged. The unflagged Plant 2 error remained four to five times the Plant 1 error, so the flag explains part of the gap but not all of it; partial derating, missing inverters outside the fixed set and sensor placement are possible contributors that the data cannot separate.

TABLE IX TASK N ERROR ON FLAGGED AND UNFLAGGED TEST TIMESTAMPS

Plant	Model	All	Unflagged	Flagged	MBE flagged (MW)
Plant 1	MLR	6.2	6.0	9.3	0.97
	SVR	4.7	4.5	7.8	1.15
	RF	5.6	5.5	6.6	0.90
	XGB	5.9	5.7	8.3	1.08
	LGBM	5.3	5.0	8.7	1.13

Plant	Model	All	Unflagged	Flagged	MBE flagged (MW)
Plant 2	MLP	5.2	4.6	10.5	1.51
	MLR	37.8	35.3	39.7	2.46
	SVR	29.6	22.6	36.5	2.00
	RF	29.8	24.8	34.7	1.64
	XGB	31.5	27.8	34.9	1.78
	LGBM	29.2	25.2	33.2	1.37
	MLP	28.8	24.6	32.9	1.30

nRMSE (%) normalized by the mean of each subset. Flagged test timestamps: 15 of 488 (Plant 1), 123 of 510 (Plant 2). Seed-0 predictions.

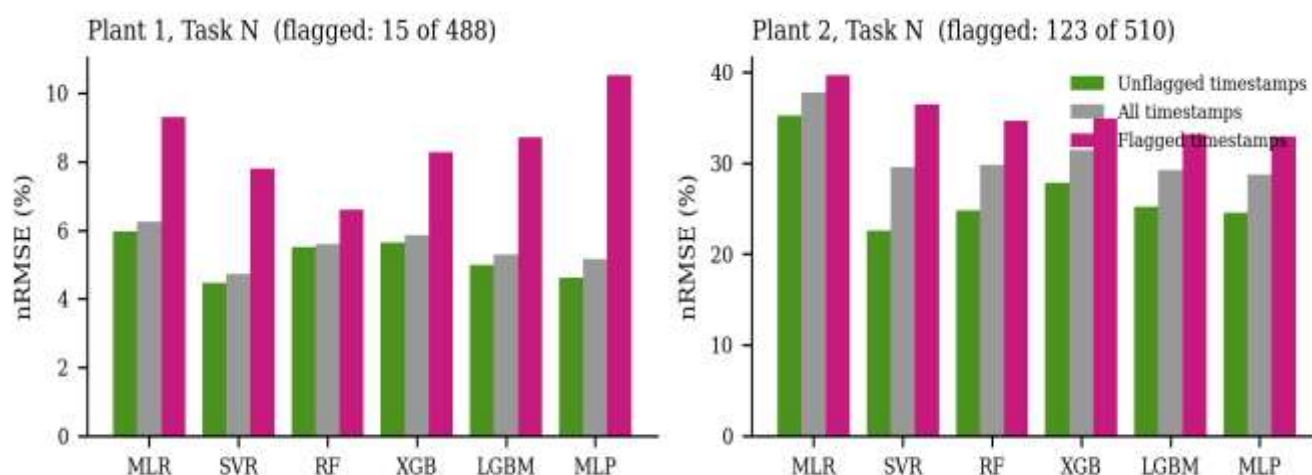


Fig. 7. Task N error on flagged and unflagged timestamps.

Retraining on unflagged samples (Table X) reduced the error on unflagged test timestamps for MLR from 35.3% to 25.2% and for LGBM from 25.2% to 24.1%, and left SVR unchanged at 22.5%. The same models were much less accurate on the full test set (40.5–52.4%), because they no longer anticipated the reduced output that flagged periods carry. The choice between the two training sets thus depends on whether the goal is expected output under normal operation or realized output.

TABLE X PLANT 2 TASK N: EFFECT OF TRAINING ON UNFLAGGED SAMPLES

Model	Training samples	nRMSE, all test (%)	nRMSE, unflagged test (%)
MLR	All	37.8	35.3
MLR	Unflagged only	52.4	25.2
SVR	All	29.6	22.6
SVR	Unflagged only	46.2	22.5
LGBM	All	29.2	25.2
LGBM	Unflagged only	40.5	24.1

For Task F the flag had little effect on the 60-min errors (all test timestamps 40.6–44.1%, unflagged 41.2–46.3% for Plant 2). Lagged power already reflects the reduced output of an ongoing event, and the per-subset normalization raises nRMSE on the unflagged subset, whose mean power is lower.

D. Feature attribution and ablation

Fig. 8 shows that irradiance accounted for 81.4% of mean $|\text{SHAP}|$ in Plant 1 nowcasting and 67.1% in Plant 2, where ambient temperature (11.1%) and time of day (11.4% for the sine and cosine terms together) took larger shares. For 60-min forecasting, the diurnal profile at the target slot accounted for 48.6% (Plant 1) and 42.1% (Plant 2), and the four power lags together for 18.1% and 11.6%.

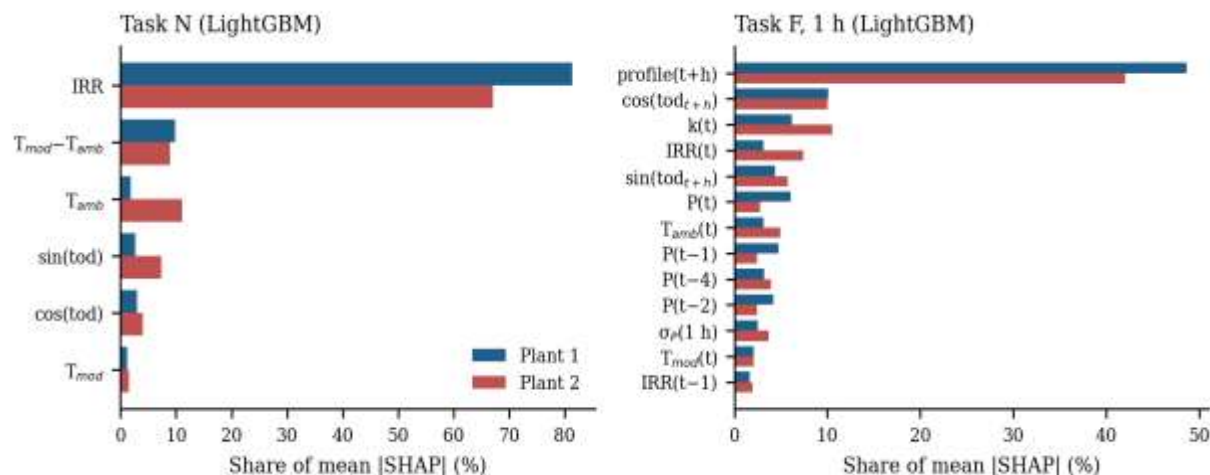


Fig. 8. Feature attribution by mean absolute SHAP value.

The ablation in Fig. 10 (right) agrees with the attribution. For Plant 1 nowcasting, irradiance alone gave 6.15% nRMSE; adding temperatures did not improve it (6.41%), adding time of day reduced it to 5.39%, and removing irradiance raised it to 18.56%. For Plant 1 forecasting, the profile and time terms alone gave 32.25% (skill 27.8%), and power lags brought this to 31.06% (skill 30.3%); issue-time weather added nothing measurable. For Plant 2 forecasting, the best subset excluded the power lags (43.06%, skill 22.8%), consistent with lags that carry anomaly-driven fluctuations of limited persistence [60].

E. Cross-plant transfer

Table XI and Fig. 9 report transfer between plants on per-inverter targets. A Task N model trained on Plant 1 gave 57.2% nRMSE on Plant 2 with a +24.5% bias; on unflagged Plant 2 timestamps its error was 24.2%, below the 25.2% of the model trained on Plant 2 itself. A model trained on Plant 2 gave 47.2% on Plant 1 with a -28.8% bias. The Plant 1 model thus describes normal Plant 2 operation about as well as the Plant 2 model does, while the Plant 2 model has learned a lower expected output that includes the flagged events. For Task F the transfer penalty was smaller: 44.3% (Plant 1 $\rightarrow$ Plant 2) against 43.9% in-plant, and 38.4% (Plant 2 $\rightarrow$ Plant 1) against 31.1%.

TABLE XI CROSS-PLANT TRANSFER OF LIGHTGBM MODELS

Task	Trained on	Tested on	nRMSE (%)	Unflagged nRMSE (%)	Bias (%)
N	Plant 1	Plant 1	5.3	5.0	-0.6
	Plant 1	Plant 2	57.2	24.2	+24.5
	Plant 2	Plant 1	47.2	47.3	-28.8
	Plant 2	Plant 2	29.2	25.2	+0.5
F, 60 min	Plant 1	Plant 1	31.1	30.8	+5.5
	Plant 1	Plant 2	44.3	44.5	+12.5
	Plant 2	Plant 1	38.4	38.6	-16.6
	Plant 2	Plant 2	43.9	46.2	+7.3

Targets and power lags normalized per inverter before training. Bias = MBE as % of mean observed power.

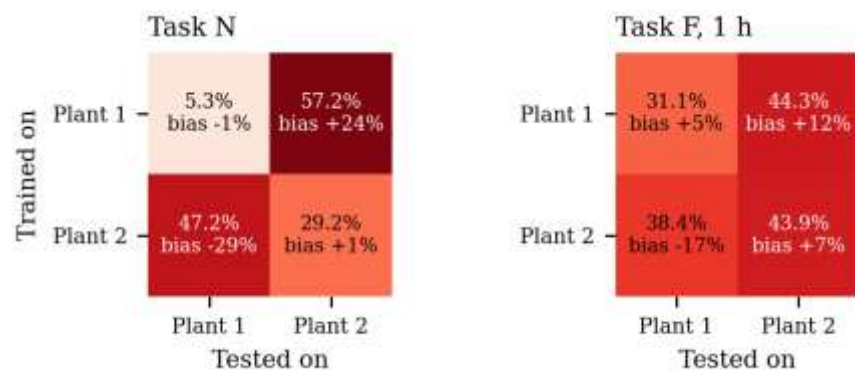


Fig. 9. Cross-plant transfer error and bias.

F. Sky conditions

For Plant 1 at 60 min (Fig. 10, left), the learners' nRMSE was 21.4–22.5% on the three clear days, 36.0–36.7% on the four mixed days and 32.9–34.4% on the three overcast days. Persistence was least accurate in every class (38.8%, 49.6% and 44.3%). On overcast days smart persistence (34.4%) matched the learners, whereas on clear and mixed days the learners were 4–5 points more accurate. For Plant 2, only one test day was clear; on its five overcast days the learners gave 47.5–51.6% against 54.6% for smart persistence. With one to five days per class, these figures describe the test period and are not estimates of long-run class-conditional error.

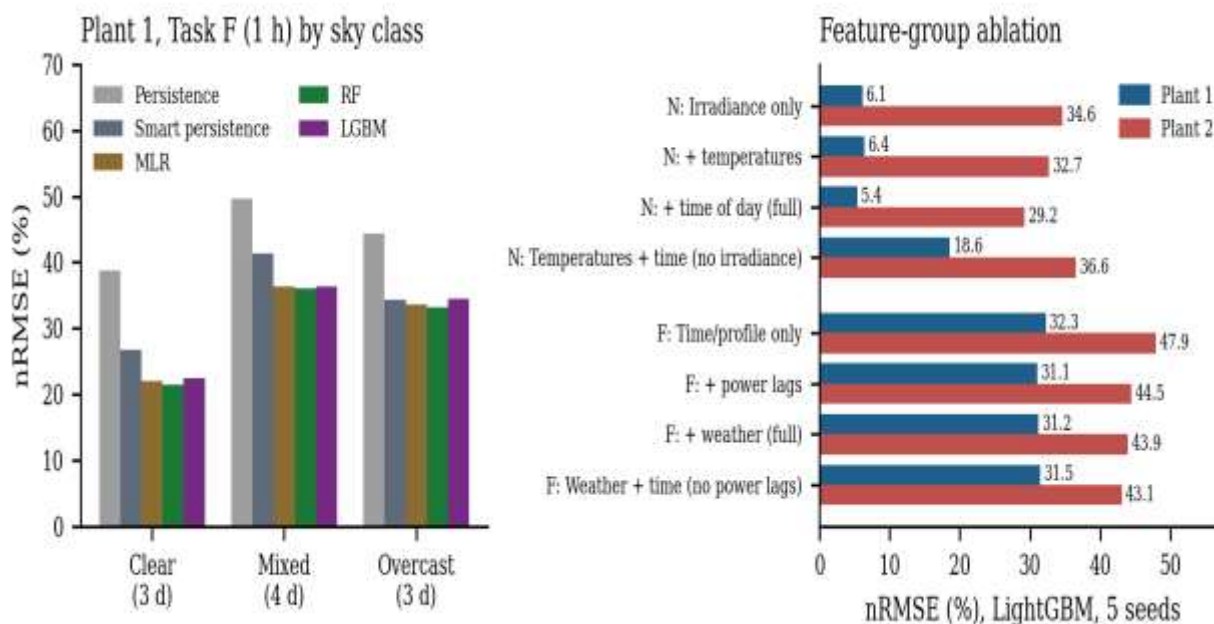


Fig. 10. Error by sky class and by feature group.

VI. DISCUSSION

A. Plant condition versus algorithm choice

Across both tasks, the spread of error attributable to the learner was smaller than the spread attributable to the plant. In nowcasting the nonlinear learners fell within 1.5–2.0 points of nRMSE of each other within a plant, while the same learners differed by about 24 points between plants. In 60-min forecasting, the learners fell within about 1 point on Plant 1 and 3.4 points on Plant 2, and most pairwise differences were not significant. The observation that the main ML families give broadly similar errors [25] is thus reproduced here under controlled conditions, and the present data add that a single change in operating condition can outweigh it. Unflagged Plant 2 samples convert irradiance at a rate close to that of Plant 1, and a model trained on Plant 1 reproduced unflagged Plant 2 output about as well as a model trained on Plant 2. This indicates that the between-plant gap arises mainly from how often and how much output departs from the irradiance-driven expectation, rather than from a different irradiance-to-power relation.

Two practical implications follow within the limits of this record. First, a ranking of learners on one plant says little about which learner will be best on another, and reporting the plant's anomaly rate and a persistence-referenced skill alongside the ranking would allow readers to judge how far it transfers [28]. Second, a Task N model trained on a well-behaved reference, such as Plant 1 here, yields an expected-output estimate whose residual on a second plant was positive and large during flagged periods. That residual could serve as an input to performance monitoring; this is a suggestion that the present data do not test, because no ground-truth fault labels are available.

B. Where learning adds value

Without weather forecasts, 60-min prediction relied mainly on the diurnal profile and recent power, and the learners improved on smart persistence by 3–8 points of nRMSE. At 15 min, persistence was hard to beat and the gain over smart persistence was not significant on Plant 1. These results agree with the view that intra-hour skill depends on inputs describing cloud motion [3] and that NWP or satellite predictors are needed for larger gains at longer horizons [9], [27]. The positive bias of all learners at 60 min, which the references did not show, suggests that the learners regressed toward the training mean in a test period that was cloudier than the training period on average; bias correction in post-processing [46] is a direct remedy.

The ablation also cautions against feature engineering without checking its effect. Temperatures, which are strongly correlated with power through their dependence on irradiance ($r = 0.918$ for module temperature on Plant 1), added nothing to irradiance alone for Plant 1 nowcasting, although they took a larger SHAP share on Plant 2. Attribution values describe what a fitted model uses, not what causes output, and on correlated inputs they can shift between features without a change in accuracy.

C. Limitations

The record spans 34 days in May and June at two plants, with a 10-day test period. Seasonal effects, soiling and degradation are therefore not represented, and the sky-class analysis rests on one to five days per class. Plant location, capacity, orientation and units are not documented, which precluded a clear-sky model and physical model chains; the smart-persistence reference is profile-based. The anomaly flag is a heuristic that cannot distinguish faults from curtailment or communication loss, and it does not capture partial derating. Task F hyperparameters were tuned at one horizon and reused. The Diebold–Mariano tests treat 10 daylight segments as one series. No NWP, satellite or sky-image inputs were available, so the forecasting results characterize what can be achieved from on-site history alone. Deep sequence models were not included; given the small sample and the absence of exogenous forecasts, they were not expected to change the conclusions, but this was not tested.

VII. CONCLUSION

On two Indian PV plants observed over the same 34 days, six machine-learning regressors reached 4.7–6.2% nRMSE in nowcasting on the plant with regular inverter behavior and 29.2–37.8% on the plant where 35% of daylight samples showed at least one inverter at zero output under irradiance. In 60-min forecasting, the learners improved on persistence by 21–32%, and their differences from one another were mostly not statistically significant. Screening flagged timestamps, retraining on unflagged samples and transferring models between plants all indicated that the gap between plants stems mainly from operating condition rather than from a different irradiance-to-power relation. For studies comparing ML methods for PV power, these results support reporting a data-quality audit, a persistence-referenced skill score and a significance test with every ranking. Extending the analysis to longer records with documented plant metadata, weather forecasts and labeled fault events would test whether the pattern holds across seasons and sites.

REFERENCES

- [1] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review

- of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016, doi: 10.1016/j.solener.2016.06.069.
- [2] D. Yang, E. Wu, and J. Kleissl, "Operational solar forecasting for the real-time market," *Int. J. Forecast.*, vol. 35, no. 4, pp. 1499–1519, 2019.
 - [3] F. Barbieri, S. Rajakaruna, and A. Ghosh, "Very short-term photovoltaic power forecasting with cloud modeling: A review," *Renew. Sustain. Energy Rev.*, vol. 75, pp. 242–263, 2017.
 - [4] S. Sobri, S. Koohi-Kamali, and N. Abd. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Convers. Manage.*, vol. 156, pp. 459–497, 2018, doi: 10.1016/j.enconman.2017.11.019.
 - [5] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in PV output power forecast," *Solar Energy*, vol. 136, pp. 125–144, 2016.
 - [6] D. Yang, W. Li, G. M. Yagli, and D. Srinivasan, "Operational solar forecasting for grid integration: Standards, challenges, and outlook," *Solar Energy*, vol. 224, pp. 930–937, 2021.
 - [7] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *WIREs Energy Environ.*, vol. 9, no. 2, 2020.
 - [8] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Trans. Ind. Appl.*, vol. 48, no. 3, pp. 1064–1069, 2012.
 - [9] H. S. Jang, K. Y. Bae, H.-S. Park, and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1255–1263, 2016, doi: 10.1109/TSTE.2016.2535466.
 - [10] Painuli, D., Bhardwaj, S., & Kose, U. (2025). Efficient feature selection and hyperparameter tuning for improved speech signal-based parkinson's disease diagnosis via machine learning techniques. *Health Technology Assessment in Action*, 9(1).
 - [11] J. Huang and M. Perry, "A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1081–1086, 2016.
 - [12] J. Tato and M. Brito, "Using smart persistence and random forests to predict photovoltaic energy production," *Energies*, vol. 12, no. 1, Art. no. 100, 2019.
 - [13] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Appl. Energy*, vol. 251, 2019.
 - [14] M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast," *IEEE Access*, vol. 8, 2020.
 - [15] Painuli, D., Bhardwaj, S., & Kose, U. (2023). Optimized diagnosis of central nervous system (cns) cancer using gene expression microarray & machine learning (ml) methods. *Eur. Chem. Bull*, 12, 9757-9771.
 - [16] E. Sarmas, E. Spiliotis, E. Stamatopoulos, V. Marinakis, and H. Doukas, "Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent long short-term memory models," *Renew. Energy*, vol. 216, Art. no. 118997, 2023.
 - [17] E. Ogliari, A. Dolara, G. Manzolini, and S. Leva, "Physical and hybrid methods comparison for the day ahead PV output power forecast," *Renew. Energy*, vol. 113, pp. 11–21, 2017.
 - [18] M. J. Mayer, "Benefits of physical and machine learning hybridization for photovoltaic power forecasting," *Renew. Sustain. Energy Rev.*, vol. 168, Art. no. 112772, 2022, doi: 10.1016/j.rser.2022.112772.
 - [19] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, 2020.
 - [20] U. K. Das et al., "Forecasting of photovoltaic power generation and model optimization: A review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 912–928, 2018, doi: 10.1016/j.rser.2017.08.017.
 - [21] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renew. Sustain. Energy Rev.*, vol. 124, Art. no. 109792, 2020, doi: 10.1016/j.rser.2020.109792.
 - [22] Painuli, D., Kumar, S., Bhardwaj, S., & Aggarwal, M. (2023). Machine learning based study on aphid infestation in agriculture crop at Uttarakhand.
 - [23] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. Mohamed Shah, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renew. Power Gener.*, vol. 13, pp. 1009–1023, 2019, doi: 10.1049/iet-rpg.2018.5649.
 - [24] K. J. Iheanetu, "Solar photovoltaic power forecasting: A review," *Sustainability*, vol. 14, no. 24, Art. no. 17005, 2022.
 - [25] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning

- methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, 2017, doi: 10.1016/j.renene.2016.12.095.
- [26] M. J. Mayer and G. Gróf, "Extensive comparison of physical models for photovoltaic power forecasting," *Appl. Energy*, vol. 283, Art. no. 116239, 2021, doi: 10.1016/j.apenergy.2020.116239.
- [27] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renew. Sustain. Energy Rev.*, vol. 161, Art. no. 112364, 2022, doi: 10.1016/j.rser.2022.112364.
- [28] D. Yang et al., "Verification of deterministic solar forecasts," *Solar Energy*, vol. 210, pp. 20–37, 2020, doi: 10.1016/j.solener.2020.04.019.
- [29] D. Yang, "Making reference solar forecasts with climatology, persistence, and their optimal convex combination," *Solar Energy*, vol. 193, pp. 981–985, 2019, doi: 10.1016/j.solener.2019.10.006.
- [30] R. Blaga, A. Sabadus, N. Stefu, C. Dughir, M. Paulescu, and V. Badescu, "A current perspective on the accuracy of incoming solar energy forecasting," *Prog. Energy Combust. Sci.*, vol. 70, pp. 119–144, 2019, doi: 10.1016/j.pecs.2018.10.003.
- [31] A. Kannal, "Solar power generation data," Kaggle, 2020. [Online]. Available: <https://www.kaggle.com/datasets/anikannal/solar-power-generation-data>
- [32] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, 2014.
- [33] H. T. C. Pedro, C. F. M. Coimbra, M. David, and P. Lauret, "Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts," *Renew. Energy*, vol. 123, pp. 191–203, 2018.
- [34] G. M. Yagli, D. Yang, and D. Srinivasan, "Automatic hourly solar forecasting using machine learning models," *Renew. Sustain. Energy Rev.*, vol. 105, pp. 487–498, 2019.
- [35] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *Int. J. Forecast.*, vol. 32, no. 3, pp. 896–913, 2016.
- [36] L. Cabezón, L. G. B. Ruiz, D. Criado-Ramón, E. J. Gago, and M. C. Pegalajar, "Photovoltaic energy production forecasting through machine learning methods: A Scottish solar farm case study," *Energies*, vol. 15, no. 22, Art. no. 8732, 2022, doi: 10.3390/en15228732.
- [37] C. Scott, M. Ahsan, and A. Albarbar, "Machine learning for forecasting a photovoltaic (PV) generation system," *Energy*, vol. 278, Art. no. 127807, 2023, doi: 10.1016/j.energy.2023.127807.
- [38] M. F. Tahir, M. Z. Yousaf, A. Tzes, M. S. El Moursi, and T. H. M. El-Fouly, "Enhanced solar photovoltaic power prediction using diverse machine learning algorithms with hyperparameter optimization," *Renew. Sustain. Energy Rev.*, vol. 200, Art. no. 114581, 2024, doi: 10.1016/j.rser.2024.114581.
- [39] A. Fouilloy et al., "Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability," *Energy*, vol. 165, pp. 620–629, 2018.
- [40] L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene, "Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components," *Renew. Energy*, vol. 132, pp. 871–884, 2019.
- [41] P. Lauret et al., "Solar forecasts based on the clear sky index or the clearness index: Which is better?" *Solar*, vol. 2, no. 4, pp. 432–444, 2022.
- [42] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006.
- [43] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, no. 3, pp. 253–263, 1995.
- [44] D. Harvey, S. Leybourne, and P. Newbold, "Testing the equality of prediction mean squared errors," *Int. J. Forecast.*, vol. 13, no. 2, pp. 281–291, 1997.
- [45] Mishra, D. (2016). Rule Based Expert System for Medical Diagnosis-A Review.
- [46] D. Yang and D. van der Meer, "Post-processing in solar forecasting: Ten overarching thinking tools," *Renew. Sustain. Energy Rev.*, vol. 140, 2021.
- [47] D. W. van der Meer, J. Widén, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1484–1512, 2018, doi: 10.1016/j.rser.2017.05.212.
- [48] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 4765–4774.
- [49] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, 2020.
- [50] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–

2830, 2011.

- [51] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794.
- [52] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 3146–3154.
- [53] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Statist. Comput., vol. 14, no. 3, pp. 199–222, 2004.
- [54] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
- [55] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Statist., vol. 29, no. 5, pp. 1189–1232, 2001.
- [56] A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and meteorological information," Renew. Energy, vol. 118, pp. 357–367, 2018.
- [57] S.-C. Lim, J.-H. Huh, S.-H. Hong, C.-Y. Park, and J.-C. Kim, "Solar power forecasting using CNN-LSTM hybrid model," Energies, vol. 15, Art. no. 8233, 2022, doi: 10.3390/en15218233.
- [58] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," Inf. Sci., vol. 191, pp. 192–213, 2012.
- [59] A. Mellit, A. Massi Pavan, E. Oglari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," Appl. Sci., vol. 10, no. 2, Art. no. 487, 2020, doi: 10.3390/app10020487.
- [60] Mishra, D., Agrawal, A. & Mittal, A. Early diagnosis of cardiovascular disease using stacking ensemble augmented by optimized machine learning framework. Mach. Learn. Comput. Sci. Eng **2**, 16 (2026). <https://doi.org/10.1007/s44379-026-00065-z>