The International Journal of Engineering and Science (IJES)
|| Volume || 15 || Issue 3 || March || Pages || PP 07-17 || 2026 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Short term PV power forecasting at a 15-minute horizon, a comparative evaluation of machine learning models using operational data from the Ma'an solar plant in Jordan

Nor Alden ELSHAWEESH[1], Özgür İNANÇ[2]
[1]Department of Energy Engineering Systems, Karabük University, Turkey
[2]Department of Energy Engineering Systems, Karabük University, Turkey

Abstract
Accurate short term photovoltaic power forecasts support grid operation and market decisions because PV output responds rapidly to irradiance and weather variability. This study evaluates machine learning models for one step ahead PV plant power prediction using operational measurements from the Ma'an solar power plant in Ma'an, Jordan. The workflow integrates synchronized electrical monitoring and onsite meteorological measurements recorded at a nominal 15-minute interval from September 1, 2025, to November 2, 2025. After time alignment and quality screening, the merged dataset includes about 5948 records, and listwise deletion yields about 5860 clean records. The target is total plant AC power computed as the sum of nine inverter active power channels in W, and the model predicts the value at t plus 15 minutes using predictors available at time t. Inputs include horizontal irradiance, ambient temperature, module temperature, relative humidity, wind speed, wind bearing, and cyclic time features, and the protocol uses an 80 percent training and 20 percent test chronological split with leakage control by excluding inverter level target components from predictors. The study compares persistence, KNN, SVR, random forest, gradient boosting, XGBoost, LightGBM, and CatBoost using RMSE, MAE, and R2 on the same held out test period. Persistence delivers the best overall accuracy with R2 of 0.9838, RMSE of 390,829.831 W, and MAE of 169,002.015 W. Among learning-based models, CatBoost performs best with R2 of 0.9620, RMSE of 599,102.249 W, and MAE of 269,402.428 W, yet all learning models remain behind persistence on RMSE and MAE with negative skill scores. The results show that persistence sets a stringent benchmark at 15-minute resolution, and they support baseline referenced reporting and careful tuning disclosure when comparing competitive tree ensemble models.

Keywords: PV power forecasting, 15-minute ahead, machine learning, persistence baseline, CatBoost, Ma'an Jordan

---

Date of Submission: 05-03-2026 Date of acceptance: 17-03-2026

I. Introduction

Fossil fuels still supply a large share of global electricity generation. Oil, coal, and natural gas increase carbon emissions and contribute to environmental pollution, and finite reserves create sustainability and supply risks [1]. Rising energy demand, climate change concerns, and volatility in fossil fuel prices have encouraged countries to expand renewable energy sources [2,3]. Solar energy has become a leading option because it offers broad accessibility, low operating costs, and environmental advantages [4]. PV generation remains variable because irradiance, temperature, and other meteorological conditions change over time, which creates operational challenges for grid management, balancing, and market operations [5].

Rapid changes in irradiance and weather conditions can cause fluctuations in PV output. When forecasts do not achieve sufficient accuracy over short time intervals, producers and market participants can incur financial losses [6]. Accurate PV power forecasting therefore supports operational planning and economic decision making in electricity systems [7,8]. PV output depends on multiple environmental and temporal drivers, which creates nonlinear and time varying relationships that often motivate machine learning approaches for prediction [4,5].

Solar resource conditions vary across geography and climate, which shapes PV performance and forecasting context. Solar engineering and solar resource references describe higher irradiation potential in low latitude and arid to semi-arid regions, with decreasing resource levels at higher latitudes, and mapping tools illustrate how irradiation varies spatially across regions and countries [2,9–12]. Jordan provides a relevant context due to strong solar resource levels and expanding PV deployment. Studies report an annual average global horizontal irradiation in the 5.5 to 6.4 kWh per square meter per day range, and a 2024 analysis reports Jordan as third globally by PVOUT in kWh per kWp per day [13]. Renewables exceeded 26 percent of total

electricity generation by 2023 [14]. Policy mechanisms, including the Renewable Energy and Energy Efficiency Law enacted in 2012 and the implementation of net metering and wheeling, have supported PV investment [15]. Jordan maintains a target to raise the renewable share in electricity generation above 31 percent by 2030 [14,16]. Global Solar Atlas data report annual average GHI values around 1970 to 2100 kWh per square meter per year in the north and increasing toward the south to about 2300 kWh per square meter per year, with Ma'an and Aqaba standing out as high radiation areas [17]. Figure 1 summarizes the reported spatial pattern for Jordan. High and regionally consistent GHI can support short term PV power forecasting and model learning under consistent measurement conditions [17,18].

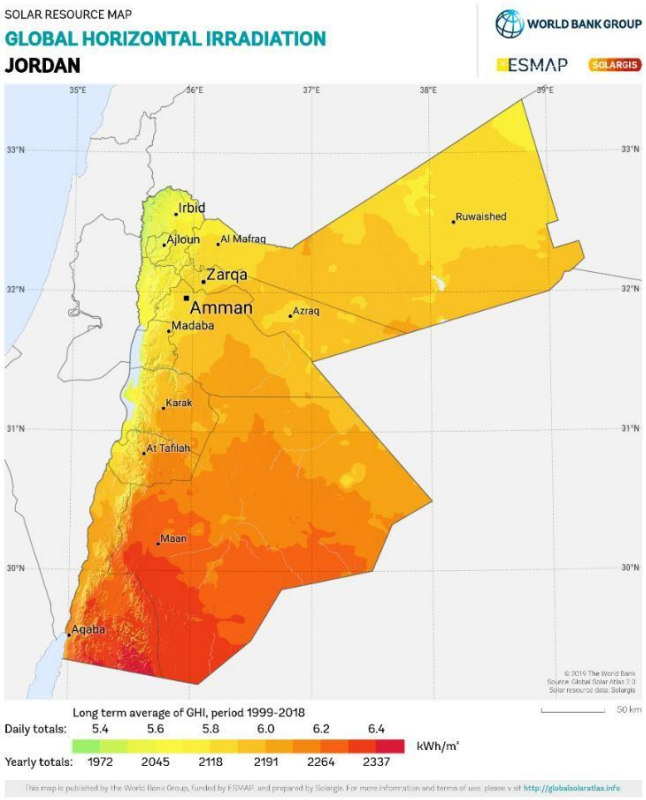

Figure 1. Global horizontal irradiation distribution in Jordan.

This study conducts a structured comparative evaluation of machine learning models developed to predict total PV plant power. The study trains multiple algorithms under a consistent protocol and evaluates them using common performance measures to identify a model that achieves strong predictive performance, robust generalization, and practical usability. The remainder of this paper reviews prior work, presents the study context and data, describes the training and evaluation approach, and reports results and implications for operational PV power forecasting.

II. Related work

PV power forecasting supports grid operation because PV output responds quickly to changes in irradiance and meteorological conditions. Reviews describe PV forecasting as a problem where performance depends on site conditions, sampling resolution, input design, and validation strategy. Studies therefore report inconsistent model rankings across datasets and horizons [21,22,32].

Machine learning methods attract attention in PV forecasting because they learn nonlinear mappings between multivariable inputs and PV output. Supervised learning dominates this setting because researchers typically train models on historical weather and time features with measured PV power as the labeled target. Background machine learning sources describe this learning setup and common modeling assumptions [22,25].

Comparative PV forecasting studies commonly report MAE, RMSE, and R2. These metrics quantify complementary aspects of error and fit, which enables comparison across model classes under a shared test set. Reviews of PV forecasting methods and comparative NWP based benchmarking studies highlight this metric set as a standard reporting choice [21–23].

Short horizon PV forecasting requires a strong baseline because PV power time series can show strong autocorrelation at fine time steps. Studies that benchmark methods without exogenous inputs show that

persistence can remain competitive when consecutive measurements change smoothly. This result motivates using persistence as a mandatory reference when the evaluation targets very short horizons [19].

Classical regressors remain common baselines in PV forecasting experiments. KNN offers a simple instance-based benchmark that predicts from similar historical conditions in feature space. Its behavior and limitations follow directly from the nearest neighbor mechanism described in introductory treatments [24]. Reviews also describe frequent use of kernel based regressors such as SVR in PV forecasting, and they note sensitivity to scaling and hyperparameters, which can affect robustness across periods with different distributions [21,22].

Tree based ensembles often perform strongly in PV forecasting because they model nonlinear interactions and heterogeneous feature effects with limited feature preprocessing. Random forest reduces variance through bagging across decorrelated trees. Reviews of random forest applications describe this variance reduction logic and its practical stability advantages [28]. Boosting based methods build additive models sequentially to reduce residual error, and tutorial and foundational sources describe the gradient boosting learning principle [29,30]. Comparative analyses of gradient boosting implementations report meaningful differences among modern boosted tree systems, which motivates direct comparisons among XGBoost, LightGBM, and CatBoost under a consistent protocol [31].

Evidence syntheses in PV forecasting and broader energy forecasting emphasize that study design choices can drive reported rankings. Input availability, split strategy, and horizon selection can shift whether learning based models outperform persistence, and can shift which ensemble variant leads among tree based methods. This variability supports controlled evaluations that hold preprocessing, leakage control, and metric reporting constant across models [21,22,32].

III. Study area and data context

III.1. Site and measurement sources

The empirical analysis uses operational measurements from the Ma'an solar power plant in Ma'an, Jordan. The workflow integrates two synchronized sources. The first source is the plant electrical monitoring stream labeled Maan B2. The second source is an integrated meteorological station labeled Maan B4, WS1. Both sources were provided as Excel files and recorded data at a nominal 15-minute interval. This structure aligns with standard PV monitoring practice that pairs AC power measurements with key meteorological drivers for performance assessment and modeling [33].

III.2. Data coverage and sampling resolution

Both datasets span September 1, 2025, to November 2, 2025. The study obtained the files on December 1, 2025. The 15-minute sampling interval defines the operational time scale of the study and frames the modeling objective as short horizon prediction at one time step resolution.

III.3. Electrical power data

The electrical file df1 includes a timestamp and nine inverter level AC power measurements. The Power AC fields represent real active power for each inverter. The file contains more than 5985 rows, which is consistent with continuous collection over roughly two months at 15-minute granularity. Several inverter columns contain missing observations. The missingness magnitude is about 88 rows per inverter column. The workflow treats these gaps as data quality limitations rather than physical behavior, and it addresses them explicitly in preprocessing to prevent models from learning artifacts.

Table 1. Time series structure of df1 and inverter power measurements.

Time	B2 AC Power	Inv 1.1 AC	Inv 1.2 A	Inv 2.1 AC	Inv 2.2 AC	Inv 3.1 AC	Inv 4.1 AC	Inv 4.2 AC	Inv 5.1 AC	Inv 5.2 AC
2025-09-01T05:15:00	-30000	0	0	0	0	0	0	0	0	0
2025-09-01T05:30:00	-30000	0	0	0	0	0	0	0	0	0
2025-09-01T05:45:00	-30000	0	0	0	0	0	0	0	0	0
2025-09-01T06:00:00	10000	13395	13315	11280	11925	21825	13220	16180	10260	12975
2025-09-01T06:15:00	380000	44050	43970	43650	41300	55380	60940	65590	58520	60800
2025-09-01T06:30:00	1250000	152415	156820	158950	156920	172280	124055	129450	122275	122430
2025-09-01T06:45:00	2660000	305160	310510	317890	318890	332410	265000	271550	264310	265680
2025-09-01T07:00:00	4285000	481805	494265	505715	504050	515750	454255	460520	451260	454770
2025-09-01T07:15:00	5870000	651890	667950	668190	678110	687720	633950	639080	628920	639050
2025-09-01T07:30:00	7125000	776085	787250	794220	802720	809945	795980	801555	792590	805615
2025-09-01T07:45:00	7610000	831620	828110	852940	853120	859940	855470	860820	861930	867930
2025-09-01T08:00:00	7870000	858705	855085	879800	881580	888805	884060	888430	891025	897455

2025-09-01T08:15:00	8080000	880350	876450	902460	906790	911900	908440	912360	914260	920560
2025-09-01T08:30:00	8240000	899615	894945	923200	926220	930070	928575	930280	933435	940010

III.4. Meteorological data

The meteorological file df2 includes a timestamp and variables that represent the local operating environment of the PV plant. The recorded fields include horizontal irradiance, ambient temperature, PV module surface temperature, relative humidity, wind speed, and wind bearing. The file contains about 5948 rows. Several meteorological variables contain missing observations in the range of about 80 to 90 rows. These variables act as physically meaningful drivers for PV output modeling, but their usefulness depends on correct time alignment and quality screening.

Table 2. Time series structure of the meteorological measurements in df2.

Time	Humidity	Irradiance Horizontal	Temperature Ambient	Temperature PV Module	Wind Bearing	Wind Speed
2025-09-01T12:30:00	7.76	951.7358333	33.66	57.745	260.975	1.85
2025-09-01T12:45:00	7.38	955.9808333	34.63	58.13	229.22	2.11
2025-09-01T13:00:00	6.71	953.0725	34.44	58.09	152.545	1.44
2025-09-01T13:15:00	6.16	943.4808333	35.48	60.68	262.74	1.68
2025-09-01T13:30:00	5.78	927.8491667	35.37	61.11	139.765	0.92
2025-09-01T13:45:00	6.51	914.3675	35.54	60.77	97.79	2.38
2025-09-01T14:00:00	6.375	904.76	35.12	56.065	99.84	2.38
2025-09-01T14:15:00	6.2	894.1233333	35.06	54.06	41.2	2.51
2025-09-01T14:30:00	5.925	877.065	35.545	58.56	48.765	1.305
2025-09-01T14:45:00	5.88	854.5283333	35.85	59.98	61.58	1.91
2025-09-01T15:00:00	5.81	824.2283333	35.87	58.685	164.785	1.39
2025-09-01T15:15:00	5.83	782.6583333	36.34	60.11	173.04	2.53
2025-09-01T15:30:00	6.385	731.0016667	35.8	56.56	141.42	2.74

IV. Methods

4.1 Time parsing, ordering, and interval integrity

The workflow converts the Time column in both files to datetime in Python using pandas. It then sorts records in ascending time order to preserve temporal causality and prevent sequencing errors during model training and evaluation. An interval check using time differences confirms the nominal 15-minute sampling interval for both sources. The workflow identifies rare irregular intervals. It removes their impact by restricting the analysis to the common overlap window used in the merged dataset, which restores a consistent time grid for time ordered modeling and evaluation [34].

4.2 Channel screening and exclusion of nonphysical power values

The column “Maan B2. Interconnection Facility, Power AC” contains negative values between - 30,000 and - 65,000 W. These values do not match a physically plausible generation output channel under the study context. The workflow treats this field as an unreliable operational indicator or an error coded signal and excludes it from subsequent processing and modeling. This step reduces the risk of contaminating the target definition and the learned relationships.

4.3 Target variable construction and unit handling

The workflow defines the prediction target as total plant AC power at each timestamp. It computes Power_Total as the sum of the nine inverter Power AC values. Because the inverter measurements are recorded in W, Power_Total remains in W as the modeling scale used for training, prediction, and metric computation. For descriptive plots only, the workflow converts to MW using Power_Total_MW = Power_Total / 1,000,000, and it preserves W for all model computations to keep units consistent and avoid rescaling errors.

4.4 Data integration across electrical and meteorological sources

The workflow merges df1 and df2 using an inner join on the Time column to form a single integrated dataset. The merge yields about 5948 rows, which represent the intersection time window of the two sources. Post merge checks confirm no duplicate timestamps and preservation of the nominal 15-minute interval. This integration design ensures that each model input vector corresponds to one timestamp with a unified set of meteorological and temporal features, subject to the subsequent missing data screen.

4.5 Distribution inspection of Power_Total for modeling implications

The workflow inspects Power_Total using a histogram and a boxplot. The histogram shows a strong concentration near zero, which reflects nighttime periods with no PV production. The distribution also shows a pronounced concentration in the high output region around 6 to 9 MW during daytime operation, with relatively sparse density between 1 and 5 MW. This pattern indicates rapid transitions between low and high production states at the 15-minute time scale, which can stress models that assume smoother response behavior. The boxplot shows a first quartile near zero and a third quartile around 6 to 7.5 MW, with an upper whisker near 9 MW. After removing the unreliable interconnection channel, the workflow identifies no clear outlier pattern that warrants additional automatic outlier deletion.

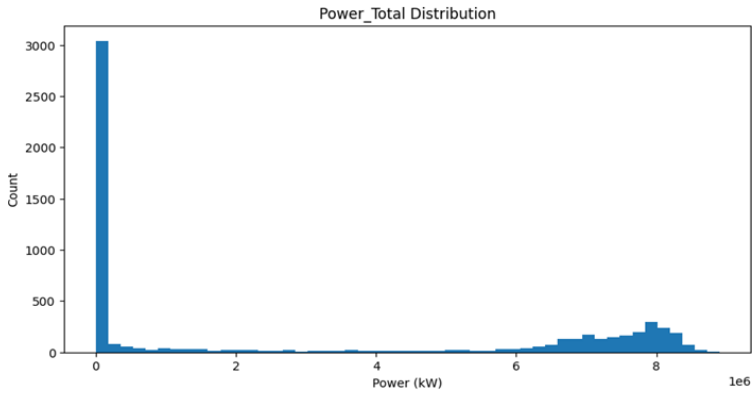

Figure 2. Distribution of Power_Total.

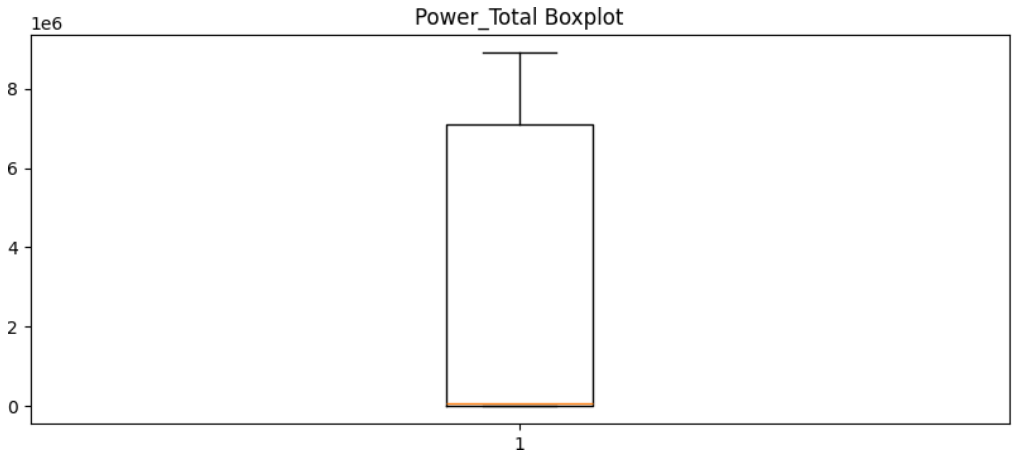

Figure 3. Boxplot analysis of Power_Total.

Table 3. Summary of key findings from the Power_Total boxplot analysis.

Explanation	Result
Represents nighttime hours or periods when production stops	High concentration at zero values
Corresponds to high production periods during the day	Quarter-hour values in the 6–7.5 MW range
Indicates the dataset is clean and physically consistent	No apparent outliers
Creates difficulty for simple models and increases the advantage of boosting models	Wide production range

4.6 Missing data treatment and rationale

The workflow addresses missing observations in both power and meteorological fields. It does not apply mean or median imputation because these approaches can introduce synthetic dynamics that do not correspond to PV behavior and can distort error statistics under short horizon evaluation. Instead, it applies listwise deletion and removes any row that contains a missing value in any modeling variable. This yields a cleaned dataset df_clean with about 5860 rows. This choice prioritizes internal validity of the learned relationships over maximal sample retention. The manuscript should acknowledge that deletion can change the effective data distribution if missingness is not completely random, and it can reduce statistical efficiency, which is a known limitation of listwise deletion in applied modeling [38].

4.7 Temporal feature engineering for periodic structure

The workflow derives time based features from the timestamp to represent systematic temporal variation in PV output. It includes direct calendar clock features, hour, minute, day_of_year, and month. It also encodes periodic effects using sine and cosine transforms to represent circular time structure without discontinuities at cycle boundaries. The resulting cyclic features include hour_sin, hour_cos, day_sin, and day_cos [37].

4.8 Feature set definition and leakage control

The workflow defines Power_Total as the target variable. It selects meteorological predictors, horizontal irradiance, ambient temperature, PV module temperature, humidity, wind speed, and wind bearing. It adds temporal predictors, hour_sin, hour_cos, day_sin, day_cos, and month. To prevent data leakage, it excludes the nine inverter power columns from the feature matrix because these variables are direct components of the target definition. This control avoids inflated performance estimates caused by target derived inputs in the predictors [36].

4.9 Train test split design for time series validity

Because the data form a time ordered series, the workflow uses a chronological split rather than random splitting. It assigns the oldest 80 percent of records to the training set and the newest 20 percent to the test set. This design aligns with operational deployment where models train on past observations and predict on later periods. It also reduces optimistic bias that can occur when random splits allow temporal dependence to leak across train and test partitions [34,35].

4.10 Reproducible cleaned dataset export

After preprocessing, the workflow saves df_clean in both CSV and XLSX formats. This step supports traceability and reduces the risk of divergence across modeling runs by establishing a single cleaned reference dataset.

4.11 Model training and evaluation

This study trains and tests a consistent set of baselines, classical machine learning models, and tree based ensemble models to predict total PV plant AC power, Power_Total, under one preprocessing and evaluation protocol. The task is one step ahead forecasting at a 15-minute horizon. The workflow defines the target as Power_Total at time t plus 15 minutes. The workflow uses predictors available at forecast issuance time, which include meteorological variables and engineered cyclic time features at time t. The workflow aligns inputs and targets by shifting the target forward by one 15-minute step and dropping the final timestamp that has no subsequent target value. This definition matches an operational setting where the forecaster issues a 15-minute ahead prediction from the latest available observations and time features. The workflow uses a chronological split to preserve temporal causality and reduce optimistic bias from temporal dependence across partitions [34,35]. The workflow keeps unit handling consistent by defining Power_Total in W for all training and evaluation, and by using MW only for descriptive plots.

4.11.1 Model families and training scope

The study evaluates four model families. First, a persistence baseline predicts the next 15-minute value using the most recent observed value available at the forecast issue time. In this study, the persistence forecast takes the form y_hat at time t plus 15 minutes given time t equals y at time t, where y denotes Power_Total. This baseline provides a strong reference in very short horizon PV forecasting because PV power series often show strong short term dependence at fine sampling intervals [19]. General forecasting references also treat persistence as a standard naive benchmark for one step ahead evaluation [34].

Second, classical machine learning regressors include KNN regression and support vector regression, plus tree based ensembles implemented as random forest regression and gradient boosting regression. These models learn mappings from meteorological and temporal features at time t to the 15-minute ahead target.

Third, the study reports tuned variants for selected classical models, including optimized KNN, optimized SVR

with standard scaling, optimized random forest, and an optimized gradient boosting configuration. Selective tuning can bias comparisons if tuning effort differs across model classes, so the manuscript states which models were tuned and which models used default or fixed settings [35].
Fourth, advanced boosted tree models include XGBoost, LightGBM, and CatBoost, which represent widely used gradient boosted decision tree frameworks that can differ in training scheme and regularization, and can differ in stability across datasets.

4.12 Training protocol and leakage control

The study uses a time ordered split rather than random splitting to match operational forecasting and to reduce optimistic bias when temporal dependence is strong. The workflow assigns the oldest 80 percent of observations to training and the newest 20 percent to testing [34,35]. The workflow keeps the target definition fixed as total plant AC power computed as the sum of the nine inverter active power measurements. It excludes individual inverter power columns from the input feature set because these fields form direct components of the target and would inflate performance through target derived inputs. The workflow uses meteorological variables and engineered cyclic time features as inputs, and it applies the same feature set to all learning based models to preserve comparability across algorithms.

4.13 Evaluation metrics and unit discipline

The study evaluates each model on the same held out test set using RMSE, MAE, and R2. PV forecasting studies commonly report these measures together because each captures a different aspect of error behavior. RMSE reacts strongly to large deviations. MAE summarizes typical absolute deviation. R2 summarizes variance explanation [21,22]. Power_Total remains in W throughout preprocessing, training, and evaluation, so the manuscript reports RMSE and MAE in W. The manuscript may report kW or MW conversions for interpretability, but it should keep metric computation on the original modeling scale to avoid unit ambiguity.

V. Results

5.1 Comparative performance on the test set

Table 4 compares models on the held-out test period for the 15-minute ahead, one step ahead task defined in Section 4.11. The persistence baseline achieves the lowest errors and the highest R2 under this setting. This outcome indicates strong short-term dependence in Power_Total at 15- minute resolution, and it matches prior evidence that persistence can remain difficult to beat on point forecast metrics at very short horizons, even when learning models use meteorological inputs [19]. Among learning-based models that use meteorological and temporal predictors at time t to forecast Power_Total at time **t** plus 15 minutes, CatBoost achieves the strongest performance, with optimized random forest and LightGBM close behind. This ranking aligns with review evidence that tree-based ensembles often perform well in PV forecasting under nonlinear interaction and rich feature effects [21,22,32]. SVR performs poorly under the reported feature representation and split, including a negative R2 for the scaled optimized variant, which indicates worse variance explanation than a mean-based predictor on the same test set.

Table 4. Comparative performance on the test dataset.

Model	R^2	RMSE	MAE
Baseline (Persistence)	0.9838	390,829.831	169,002.015
CatBoost	0.9620	599,102.249	269,402.428
Random Forest (Optimized)	0.9595	618,519.224	297,636.058
LightGBM	0.9566	639,967.381	292,204.242
Gradient Boosting (Original)	0.9523	671,386.054	305,011.408
Random Forest (Original)	0.9522	672,154.088	295,231.409
KNN (Optimized)	0.9489	694,853.414	355,082.160
Gradient Boosting (Optimized)	0.9437	728,986.362	399,518.350
KNN (Original)	0.9429	734,152.674	376,286.353
XGBoost	0.9383	763,551.849	314,495.354
SVR (Original)	0.0203	3,042,232.659	1,992,360.998
SVR (Optimized + Scaling)	–0.5101	3,777,037.658	2,428,424.179

The tuning outcomes show that optimization does not guarantee improvement. Optimized random forest improves relative to it untuned configuration, while optimized gradient boosting degrades relative to its untuned configuration. This divergence is consistent with sensitivity of boosting style learners to hyperparameter choices and training protocol details, where modest changes can shift the balance between occasional large errors and typical errors captured by RMSE and MAE [29– 31].

5.2 Performance relative to persistence

Table 5 shows that the persistence baseline achieves the lowest RMSE and MAE on the held-out test period for the one step, 15-minute ahead task. For this reason, baseline referenced reporting is required to keep performance claims aligned with the actual benchmark. Table 6 reports RMSE and MAE skill scores relative to persistence on the same test set and unit scale.

$$\mathrm{Skill}_{\mathrm{E}} = 1 - \frac{\mathrm{E}_{\mathrm{model}}}{\mathrm{E}_{\mathrm{persistence}}}$$

Where E is the error metric evaluated on the held-out test set for the 15-minute ahead, one step ahead task, E_model is the model error, and E_persistence is the persistence baseline error computed under the same split, horizon, and units. A value of 0 indicates parity with persistence. Positive values indicate improvement over persistence. Negative values indicate worse performance than persistence under the same split and units. All values in Table 5 are negative, which means all evaluated learning models have higher RMSE and MAE than the persistence baseline on the same test set and forecast horizon.

Table 5. RMSE and MAE skill scores relative to persistence on the test set.

Model	**RMSE skill vs persistence**	**MAE skill vs persistence**
CatBoost	-0.5327	-0.5941
Random Forest (Optimized)	-0.5824	-0.7611
LightGBM	-0.6373	-0.7290

Under this definition, a negative skill value directly quantifies the proportional error increase relative to persistence. For example, the CatBoost RMSE skill of -0.5327 implies RMSE_model / RMSE_persistence = 1 - (-0.5327) = 1.5327, which corresponds to a 53.27% higher RMSE than persistence. Likewise, the CatBoost MAE skill of -0.5941 corresponds to a 59.41% higher MAE than persistence. The optimized random forest and LightGBM also yield negative skill scores, indicating RMSE increases of 58.24% and 63.73% relative to persistence, and MAE increases of 76.11% and 72.90%, respectively.

Therefore, all reported learning models remain behind persistence on both RMSE and MAE for this 15-minute, one step ahead evaluation using the current chronological split. CatBoost remains the strongest learning model among those evaluated, but persistence remains best overall for point forecast error metrics under the present experimental design [19].

5.3 Temporal tracking in a representative test segment

Aggregate metrics summarize overall error behavior, but they do not show forecast behavior during ramps, peaks, and near zero nighttime periods. The study overlays observed and predicted power for a short, readable test segment. The plotted window uses the first 200 test observations to preserve key operating patterns while limiting visual congestion.

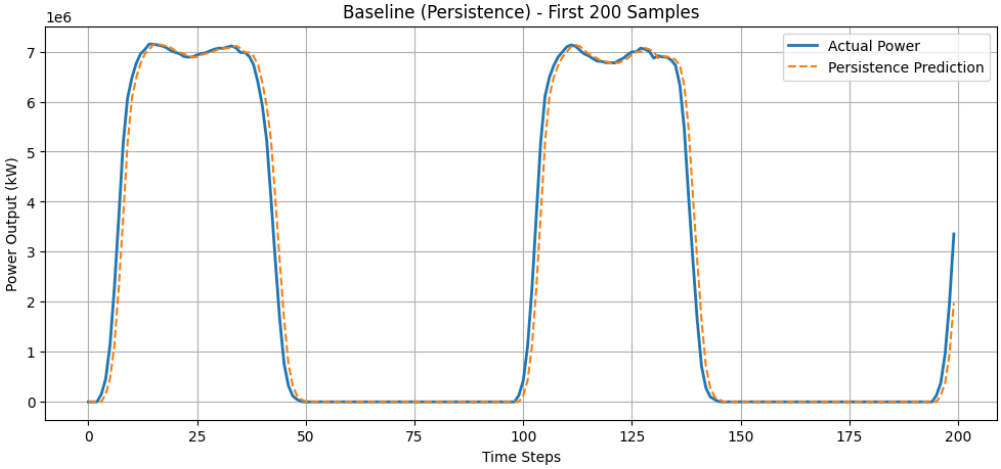

Figure 4. Persistence baseline, actual and predicted power for the first 200 test observations.

The persistence baseline follows stable periods closely by construction. It departs most during rapid transitions because it cannot anticipate ramps and it can only repeat the last observed value [19].

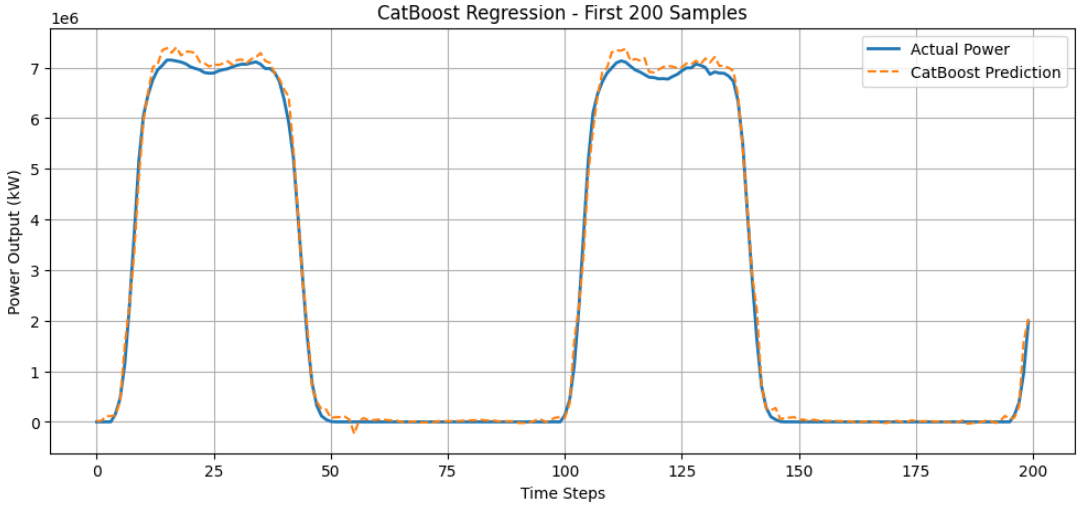

Figure 5. CatBoost, actual and predicted power for the first 200 test observations.

CatBoost tracks the overall trajectory using meteorological and temporal inputs. Deviations concentrate around rapid changes and near peaks. This qualitative pattern aligns with review evidence that short horizon PV forecast errors concentrate under fast variability and transition regimes where irradiance changes rapidly [21,22].

VI. Discussion

The results show that sampling resolution and horizon shape comparative conclusions. At 15- minute resolution with one step ahead evaluation, PV power exhibits strong short-term dependence, which makes persistence a stringent benchmark on point forecast metrics [19]. Persistence dominance does not mean persistence provides a complete operational solution. Persistence does not learn relationships between meteorological drivers and output, and it provides limited explanatory value for weather sensitive decision support.

Within the learning-based group, tree-based ensembles provide the most reliable performance under the reported setup. CatBoost, LightGBM, and random forest outperform KNN and SVR, which supports review evidence that ensemble learners often handle nonlinear interactions and heterogeneous feature effects effectively in PV power forecasting [21,22,32]. Differences among boosted tree frameworks can arise from differences in regularization and training scheme, which can affect stability across datasets, especially when models cluster closely in accuracy [31].

The tuning outcomes carry practical implications for transparent reporting. Random forests improve after optimization, yet gradient boosting degrades after optimization. This pattern indicates that tuning can change the balance between occasional large errors and typical errors, which RMSE and MAE capture

differently. It reinforces the need to report the tuning procedure and evaluation protocol clearly, since rankings can shift when competitive models differ by modest margins [29–31,35].

SVR fails under the current feature representation and scale. The negative R2 and large errors indicate poor generalization to the test period. This outcome aligns with PV forecasting reviews that note sensitivity of classical kernel methods to scaling, hyperparameter choices, and feature design under high variance regression settings [21,22].

Baseline referenced interpretation improves claim discipline. Table 6 shows that all learning models remain behind persistence on RMSE and MAE under the reported evaluation. Skill style framing prevents an implicit claim that the best learning model is best overall when a strong baseline leads [19].

VII. Conclusion

This study compares a persistence baseline, classical regressors, tuned classical models, and tree- based ensemble models for total PV plant power prediction using 15-minute data from Ma'an, Jordan under a single time ordered train test split and a consistent feature protocol. The persistence baseline achieves the best test performance, which indicates strong short term temporal dependence in the target series under the one step ahead setting.

Among learning-based models that use meteorological and temporal inputs, CatBoost achieves the strongest overall performance, followed closely by optimized random forest and LightGBM. SVR performs poorly under the reported setup, including a negative R2 for the scaled optimized variant.

These results support two implications for short horizon PV forecasting under strong autocorrelation. Authors should be a benchmark against persistence under the same split and units. Authors should interpret learning model rankings using baseline referenced performance to avoid overstating gains when the baseline remains dominant.

References

[1]. REN21. 2024. Renewables 2024 Global Status Report. REN21 Secretariat, Paris, France. Available at: https://www.ren21.net/gsr-2024/. Accessed: 2026-02-23.

[2]. International Energy Agency, IEA. 2023. World Energy Outlook 2023. IEA, Paris, France. Available at: https://iea.blob.core.windows.net/assets/ed1e4c42-5726-4269-b801- 97b3d32e117c/WorldEnergyOutlook2023.pdf. Accessed: 2026-02-23.

[3]. BP. 2023. BP Energy Outlook 2023. BP p.l.c. Available at: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy- economics/energy-outlook/bp-energy-outlook-2023.pdf. Accessed: 2026-02-23.

[4]. Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. 2016. Review of photovoltaic power forecasting. Solar Energy. 136, 78-111. doi:10.1016/j.solener.2016.06.093.

[5]. Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, Fouilloy A. 2017. Machine learning methods for solar radiation forecasting, a review. Renewable Energy. 105, 569-582. doi:10.1016/j.renene.2016.12.095.

[6]. Lari AJ, Sanfilippo AP, Bachour D, Perez-Astudillo D. 2025. Using machine learning algorithms to forecast solar energy power output. Electronics. 14(5), 866. doi:10.3390/electronics14050866.

[7]. Aouidad HI, Bouhelal A. 2024. Machine learning based short term solar power forecasting, a comparison between regression and classification approaches using extensive Australian dataset. Sustainable Energy Research. 11(1), 28. doi:10.1186/s40807-024-00115-1.

[8]. Jang SY, Oh BT, Oh E. 2024. A deep learning based solar power generation forecasting method applicable to multiple sites. Sustainability. 16(12), 5240. doi:10.3390/su16125240.

[9]. Duffie JA, Beckman WA. 2013. Solar Engineering of Thermal Processes. 4th edition. John Wiley and Sons, Hoboken, NJ, USA.

[10]. Perez R, Ineichen P, Seals R, Michalsky J, Stewart R. 1990. Modeling daylight availability and irradiance components from direct and global irradiance. Solar Energy. 44(5), 271-289. doi:10.1016/0038-092X(90)90015-S.

[11]. Global Solar Atlas. 2026. Global Solar Atlas. Available at: https://globalsolaratlas.info/. Accessed: 2026-02-23.

[12]. Solargis. 2026. Free solar maps and GIS data, high-resolution download. Available at: https://solargis.com/resources/free-maps-and-gis-data. Accessed: 2026-02-23.

[13]. Al Salaymeh A, Ficarella A, Al Smadi W, Daoud HS, Al Salaymeh B, Qutishat S, Komninos NP, Hauer M, Colangelo G, Aghazadeh Ardebili A, Longo A, Romano MP, Sakhrieh A. 2024. Solar Energy Resources in Jordan. Preprints.org. doi:10.20944/preprints202408.0866.v1.

[14]. Ministry of Energy and Mineral Resources, Jordan, MEMR. 2026. Maps, Maps of Renewable Projects. Available at: https://www.memr.gov.jo/en/pages/sitemap. Accessed: 2026-02-23.

[15]. International Renewable Energy Agency, IRENA. 2021. Renewable Readiness Assessment, The Hashemite Kingdom of Jordan. IRENA, Abu Dhabi, UAE. Available at: https://www.irena.org/-

[16]. /media/Files/IRENA/Agency/Publication/2021/Feb/IRENA_RRA_Jordan_2021.pdf. Accessed: 2026-02-23.

[17]. World Bank. 2024. Jordan Economic Monitor, Summer 2024, Strength Amidst Strain, Jordan's Economic Resilience. World Bank, Washington, DC, USA. Available at: https://documents1.worldbank.org/curated/en/099800110012498130/pdf/IDU19a9c4c7913b4f1442c1aa381b0d42eae64d3.pdf. Accessed: 2026-02-23.

[18]. Global Solar Atlas. 2026. Jordan, download maps and GIS data. Available at: https://globalsolaratlas.info/download/jordan. Accessed: 2026-02-23.

[19]. Owhaib W, Boretti A, Alkhalidi A, Al Kouz W, Hader M. 2022. Design of a solar PV plant for Ma'an, Jordan. IOP Conference Series, Earth and Environmental Science. 1008, 012012. doi:10.1088/1755-1315/1008/1/012012.

[20]. Pedro HTC, Coimbra CFM. 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy. 86(7), 2017-2028. doi:10.1016/j.solener.2012.04.004.

[21]. Mellit A, Pavan AM, Lughi V. 2021. Deep learning neural networks for short term photovoltaic power forecasting. Renewable Energy. 172, 276-288. doi:10.1016/j.renene.2021.02.166.
[22]. Ahmed R, Sreeram V, Mishra Y, Arif MD. 2020. A review and evaluation of the state of the art in PV solar power forecasting, techniques and optimization. Renewable and Sustainable Energy Reviews. 124, 109792. doi:10.1016/j.rser.2020.109792.
[23]. Gaviria JF, Narváez G, Guillen C, Giraldo LF, Bressan M. 2022. Machine learning in photovoltaic systems, a review. Renewable Energy. 196, 298-318.
[24]. doi:10.1016/j.renene.2022.06.105.
[25]. Markovics D, Mayer MJ. 2022. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renewable and Sustainable Energy Reviews. 161, 112364. doi:10.1016/j.rser.2022.112364.
[26]. Zhang Z. 2016. Introduction to machine learning, k nearest neighbors. Annals of Translational Medicine. 4(11), 218. doi:10.21037/atm.2016.03.37.
[27]. Qi C, Yilmaz E, Chen Q. 2024. Background of machine learning. In: Qi C, Chen Q, Yilmaz E, editors. Machine Learning Applications in Industrial Solid Ash. Elsevier, Amsterdam, The Netherlands. p. 93-130. doi:10.1016/B978-0-443-15524-6.00015-7.
[28]. Mahdavinejad MS, Rezvan M, Barekatain M, Adibi P, Barnaghi P, Sheth AP. 2018. Machine learning for internet of things data analysis, a survey. Digital Communications and Networks. 4(3), 161-175. doi:10.1016/j.dcan.2017.10.002.
[29]. Chakraborty D, Mondal J, Barua HB, Bhattacharjee A. 2023. Ensemble learning methods for prediction of solar power generation based on meteorological parameters in Eastern India. Renewable Energy Focus. 44, 277-294. doi:10.1016/j.ref.2023.01.007.
[30]. Belgiu M, Drăguţ L. 2016. Random forest in remote sensing, a review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 114, 24-31. doi:10.1016/j.isprsjprs.2016.01.011.
[31]. Natekin A, Knoll A. 2013. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics. 7, 21. doi:10.3389/fnbot.2013.00021.
[32]. Friedman JH. 2001. Greedy function approximation, a gradient boosting machine. The Annals of Statistics. 29(5), 1189-1232. doi:10.1214/aos/1013203451.
[33]. Bentéjac C, Csörgő A, Martínez-Muñoz G. 2021. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review. 54(3), 1937-1967. doi:10.1007/s10462-021-09996-0.
[34]. Mosavi A, Salimi M, Ardabili SF, Rabczuk T, Shamshirband S, Varkonyi-Koczy AR. 2019. State of the art of machine learning models in energy systems, a systematic review. Energies. 12(7), 1301. doi:10.3390/en12071301.
[35]. International Electrotechnical Commission, IEC. 2017. IEC 61724-1:2017. Photovoltaic system performance. Part 1: Monitoring. IEC, Geneva, Switzerland.
[36]. Hyndman RJ, Athanasopoulos G. 2021. Forecasting: Principles and Practice. 3rd edition. OTexts, Melbourne, Australia. Available at: https://otexts.com/fpp3/. Accessed: 2026-02-23.
[37]. Bergmeir C, Benítez JM. 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences. 191, 192-213. doi:10.1016/j.ins.2011.12.028.
[38]. Kapoor S, Narayanan A. 2023. Leakage and the reproducibility crisis in machine-learning- based science. Patterns. 4(9), 100804. doi:10.1016/j.patter.2023.100804.
[39]. Fisher NI. 1993. Statistical Analysis of Circular Data. Cambridge University Press, Cambridge, UK.
[40]. Pepinsky TB. 2018. A note on listwise deletion versus multiple imputation. Political Analysis. 26(4), 480-488. doi:10.1017/pan.2018.20.