Baltic J. Modern Computing, Vol. 14 (2026), No. 2, 356-375
https://doi.org/10.22364/bjmc.2026.14.2.05

Hybrid Physics-Informed Temporal Attention Model with External Physical Constraints for PV Power Forecasting

Dezdemona GJYLAPI, Alketa HYSO, Astrit DENAJ

University "Ismail Qemali" of Vlora, Vlora, Albania

dezdemona.gjylapi@univlora.edu.al, alketa.hyso@univlora.edu.al, astrit.denaj@univlora.edu.al

ORCID 0009-0005-5186-3636, ORCID 0009-0000-6256-9831, ORCID 0009-0000-7594-5058

Abstract: This paper introduces the Hybrid Physics-Informed Temporal Attention (H-PITA) model, a novel framework for accurate, physically consistent daily photovoltaic (PV) energy (kWh) forecasting. By integrating a temporal-attention encoder with external physical priors via gated fusion and a lightweight residual pathway, H-PITA balances data-driven learning with domain-specific constraints. The model incorporates physical knowledge through symmetric and asymmetric penalties, non-negativity constraints, and dynamic loss weighting, as well as an optional inference-time safety cap. Evaluated on a real-world PV system in Southeast Europe using a longitudinal dataset (2014–2025), H-PITA demonstrates superior performance over baselines including LSTM, XGBoost, persistence, climatology, and linear/ridge models, achieving a test R^2 of 0.78. Results indicate that data filtering and physical grounding significantly enhance generalization, reducing MAPE and RMSE. Overall, H-PITA provides a robust, modular, and transferable solution for risk-aware PV forecasting, making it highly suitable for operational grid integration and solar asset management.

Keywords: Photovoltaic energy forecasting; Physics-informed neural networks; Temporal attention; Hybrid deep learning; Physical constraints; Renewable energy prediction; PV systems;

1. Introduction

Modern power systems increasingly rely on solar photovoltaic (PV) generation, making accurate day-ahead PV power forecasting (here formulated as daily energy in kWh) crucial for grid stability and efficient energy management. Integrating large shares of intermittent PV energy into the electricity grid is challenging due to the variability of solar irradiance (e.g., rapid weather changes causing supply–demand imbalances and voltage fluctuations) (Di Leo et al., 2025).

Day-ahead forecasts (covering 24–48 hours ahead) enable grid operators and energy markets to better plan dispatch, schedule reserves, and minimize balancing costs. In fact, precise PV output predictions are considered one of the most cost-effective tools for accommodating higher renewable penetration without compromising reliability. By

anticipating fluctuations in solar generation, system operators can optimize resource allocation, maintain network stability, and reduce operational costs. Consequently, day-ahead PV forecasting has become indispensable for modern smart grids and energy markets striving to integrate renewables at scale (Di Leo et al., 2025; Barhmi et al., 2024).

However, forecasting PV power output remains a complex task due to several limitations of traditional methods and pure machine learning approaches. Statistical and time-series models often struggle to capture the nonlinear relationships between meteorological variables (e.g., irradiance, cloud cover, temperature) and PV output (Di Leo et al., 2025; Vázquez Pombo et al., 2022). The stochastic nature of weather can lead to model mis-specification and poor accuracy for conventional predictors. Meanwhile, data-driven machine learning (ML) models have achieved high predictive accuracy but come with their own challenges. Many ML-based PV predictors act as "black boxes" with limited interpretability, offering little insight into how input features (e.g., weather forecasts, irradiance, ambient temperature, cloud cover, and wind speed) influence the output (Vázquez Pombo et al., 2022). This lack of transparency is problematic for energy planners who require confidence that forecasts align with physical behaviour.

Additionally, ML models can be sensitive to noisy data and outliers in weather inputs; without special handling, their performance may degrade under unusual weather events (Barhmi et al., 2024). Purely data-driven models also risk learning spurious correlations, especially when training data are limited or unbalanced. Perhaps most critically, traditional ML approaches often do not incorporate known physics of PV systems, meaning they might violate energy-balance principles or perform inconsistently outside the training-data regime.

The absence of physical constraints or domain knowledge in the model can reduce generalizability and trustworthiness (Vázquez Pombo et al., 2022). Researchers have identified the integration of physical knowledge into ML as a promising avenue, yet comparatively fewer works have explored such hybrid models for PV power forecasting (Vázquez Pombo et al., 2022). These limitations motivate physics-guided approaches that retain ML's pattern-learning strength while enforcing domain plausibility.

In this paper, we propose a Hybrid Physics-Informed Temporal Attention (H-PITA) model for day-ahead PV power forecasting. H-PITA integrates data-driven learning with an external physics prior of daily PV energy production. The architecture couples a temporal-attention encoder with a gated fusion mechanism that combines the learned temporal representation with a physics-based auxiliary input x_{phys}, corresponding to the physically estimated daily PV energy. In addition, a lightweight residual path from x_{phys} is included to stabilize training and preserve physical consistency in the final prediction.

Attention mechanisms allow the network to focus on the most relevant portions of daily sequences/lags, dynamically highlighting time steps that strongly influence PV output (e.g., rapid cloud transitions) (Lei, 2024; Zhong et al., 2024).

The external physics prior (e.g., a clear-sky or simplified performance estimate) injects domain knowledge directly into the forecasting process, improving interpretability and helping predictions remain consistent with known physical relationships (Vázquez Pombo et al., 2022).

In this paper, training employs a composite objective with dynamic weighting of the physics term to balance accuracy and physical consistency; specifically, the prior enters through four complementary mechanisms: (i) a symmetric relative penalty around physical benchmark E_{phys}, (ii) a one-sided hinge that activates when predictions exceed

(1+margin)·E_{phys}, (iii) an explicit non-negativity penalty, and (iv) adaptive (dynamic) weighting of the physics term during training.

For operation, an optional +8% hard cap at inference provides a tunable safety mode that enforces a physically conservative upper bound with a controlled accuracy trade-off. The 8% value was adopted as a conservative heuristic safety margin for the optional inference-time cap. To examine whether the observed behaviour depended strongly on this choice, an additional 12% cap was also evaluated in the sensitivity analysis as a less restrictive alternative. This design aligns with evidence that attention improves the extraction of salient temporal patterns in weather sequences and with advances in physics-informed learning and dynamic loss balancing (Vázquez Pombo et al., 2022; Lei, 2024; Zhong et al., 2024).

Why this model? A recent meta-survey of PV-forecasting papers shows that regression/ANN/instance-based methods dominate the literature, while hybrid approaches are comparatively under-represented—signaling scope for architectures that combine deep temporal modeling with explicit physics (Figure 1) (Alcañiz et al., 2023). H-PITA is positioned precisely in this hybrid, physics-guided regime, aiming to deliver strong accuracy, interpretability, and trustworthy behaviour under atypical conditions.

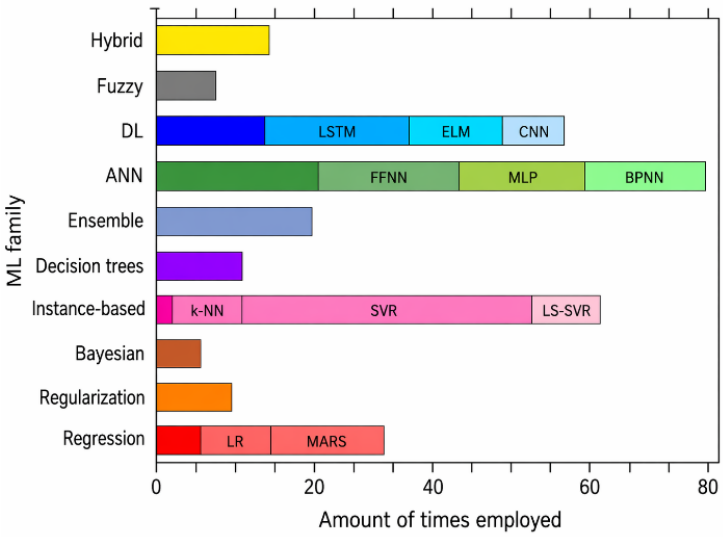

Figure 1. Distribution of ML families employed in PV-forecasting studies; hybrid methods are less common than ANN/instance-based/regression. Adapted from (Alcañiz et al., 2023), CC BY 4.0.

2. Recent developments in PV energy forecasting and physics-integrated models

2.1. Machine learning models for PV forecasting

Modern machine-learning (ML) techniques have markedly improved the accuracy of photovoltaic (PV) power forecasting across settings ranging from small rooftop systems to utility-scale plants. Among recurrent architectures, Long Short-Term Memory (LSTM) networks remain widely adopted for capturing temporal dependencies in PV generation; in a recent meta-survey, Alcañiz et al. report frequent use of LSTMs and generally strong performance across studies (Alcañiz et al., 2023).

Building on recurrence, attention mechanisms have been introduced to emphasize the most informative portions of multivariate weather sequences. For example, Hu et al. combine LSTM with self-attention—ingesting weather forecasts—to enhance multistep PV prediction, reporting substantial gains over a plain LSTM on a building-scale PV system in Japan (Hu et al., 2024).

Attention has also been incorporated to improve feature extraction prior to sequence modeling. Zhong et al. propose an attention-based DiPLS–BiLSTM framework, where Dynamic Inner Partial Least Squares is used to extract latent features from high-dimensional inputs, an attention layer assigns adaptive importance to these features, and a BiLSTM captures bidirectional temporal dependencies. Their results demonstrate high predictive accuracy; however, the approach remains purely data-driven and does not incorporate explicit physical constraints or physics-based priors (Zhong et al., 2024).

Attention has further been embedded within hybrid deep architectures: Song et al. pair a CNN–BiLSTM feature extractor with an NGBoost probabilistic head and attention, showing notable improvements over strong deep-learning baselines across multiple Australian sites (Song et al., 2025).

Separately, transformer-style attention has been adapted to solar time series within CNN–LSTM–Transformer hybrids, where attention layers help fuse spatial–temporal features and improve accuracy on public PV datasets (Lei, 2024). Overall, recent evidence indicates that mixing convolutional encoders, sequence models, and attention modules can yield state-of-the-art results on real PV deployments (Song et al., 2025; Al-Ali et al., 2023).

Similar conclusions regarding the effectiveness of machine-learning approaches for complex forecasting tasks have also been reported outside the photovoltaic domain, where comparative studies show that ensemble and tree-based models often outperform classical statistical time-series methods when nonlinear temporal patterns dominate (Dwivedi et al., 2025).

Beyond offline validation, ML predictors are increasingly tested in live settings. Nithya et al. demonstrate a sequence-to-sequence LSTM operating in real time on a 100 kW campus PV plant, using streaming measurements to generate rolling forecasts and illustrating practical reliability under operational constraints (Nithya et al., 2024).

The importance of explicitly modeling temporal dependencies has likewise been highlighted in broader machine-learning applications, where recurrent and sequential neural architectures consistently outperform static models by capturing long-term patterns in complex time-series data (Sneiders et al., 2025).

Taken together, the literature since ~2019 shows a maturation from single-family models (e.g., plain LSTMs) toward attention-enhanced hybrids that better exploit spatiotemporal structure while maintaining deployability (Alcañiz et al., 2023; Nithya et al., 2024). In the broader renewable-energy decision-support literature, Özkurt et al. further show that combining machine learning with explainability-oriented methods can support more transparent and interpretable energy-related decisions (Özkurt et al., 2025).

2.2. Use of physical models within the forecasting pipeline

In parallel with purely data-driven models, many PV forecasting pipelines explicitly integrate physics at one or more stages. A common practice is to reference a clear-sky model of irradiance to form the clear-sky index, thereby detrending strong diurnal/seasonal effects and improving stationarity for downstream learning; comparative analyses generally favor clear-sky index formulations over simpler normalizations for forecast accuracy (Lauret et al., 2022). Physics-derived features are also fed directly into deep models—for instance, including Ineichen–Perez clear-sky Global Horizontal Irradiance (GHI) as an input signal that anchors forecasts to a physically plausible envelope and lets the network learn cloud-induced deviations (Maciel et al., 2024).

Related evidence from broader energy-forecasting applications further suggests that incorporating physically meaningful and context-aware variables—such as environmental and temporal factors—can substantially improve predictive accuracy, particularly under data scarcity or noisy measurement conditions (Sanfilippo et al., *2025*).

At the post-processing level, indirect pipelines first predict irradiance and then convert to power via a PV performance model. Urhuerhi et al. illustrate such a two-stage approach—ML weather-to-irradiance followed by PVLib-based power conversion—leveraging known device physics to stabilize power outputs under varying conditions (Urhuerhi et al., 2025).

In day-ahead horizons, numerical weather prediction (NWP) provides a physical baseline that can be post-processed by ML to correct systematic biases; Buonanno et al. show that even compact linear models can yield consistent error reductions over raw NWP PV forecasts when data are scarce (e.g., new plants) (Buonanno et al., 2024). These physics–ML ensembles—where a physics model supplies a first-principles estimate, and ML performs bias-correction—are increasingly prevalent across PV forecasting workflows (Lauret et al., 2022, Buonanno et al., 2024).

Finally, simple physics-aware baselines such as persistence or smart persistence (clear-sky-index persistence) remain indispensable sanity checks for operators and continue to set a performance bar for advanced ML systems (IEA PVPS Task 16, 2025).

2.3. Physics-informed neural networks (PINNs) and physics-integrated learning

Physics-Informed Neural Networks (PINNs) embed governing equations directly into the training objective so that predictions honor known physics while fitting data (Raissi et al., 2019). In power systems, PINNs have been explored for fast, stable surrogates of grid dynamics: e.g., Stiasny et al. develop PINNSim to accelerate transient simulations by

enforcing differential–algebraic power-system equations in the loss, enabling larger time steps without sacrificing physical fidelity on benchmark networks (Stiasny et al., 2024).

For state estimation and security, physics-guided neural formulations that enforce power-flow constraints have shown improved robustness to bad data and cyber-attack scenarios relative to purely data-driven baselines (Falas et al., 2025).

In renewables, physics-integrated approaches have improved extrapolation under extremes; for wind power, Liu et al. incorporate analytical turbine power curves within a physics-informed reinforcement learning framework to enhance performance during rare events and data-sparse regimes (Liu et al., 2024).

Collectively, this line of work demonstrates that explicit physics—be it through constraints, priors, or loss terms—can reduce data demands, improve robustness, and support trustworthy behavior in safety-critical energy applications (Raissi et al., 2019; Liu et al., 2024).

In parallel, broader machine-learning studies emphasize that explicitly modeling temporal dependencies is critical for reliable forecasting, with sequential neural architectures consistently outperforming static models by capturing long-term patterns in complex time-series data (Sneiders et al., 2025*).*

Implication for hybrid PV forecasting. Taken together, the developments above motivate hybrid designs that combine temporal attention with external physical signals or constraints. From a broader methodological perspective, comparative forecasting studies show that machine-learning models—particularly ensemble and tree-based approaches—often outperform classical statistical time-series methods when nonlinear temporal effects dominate, further supporting hybrid designs that balance data adaptivity with structural priors (Dwivedi et al., 2025).

Such designs can retain the accuracy of modern deep models while aligning forecasts with domain knowledge—an approach our H-PITA model operationalizes by gating an external physics prior, adding a lightweight residual physical pathway, and employing a physics-aware training objective with dynamic weighting (see Section 3.2).

3. Methodology and data

The proposed model is a Hybrid Physics-Informed Neural Network with a temporal attention mechanism. The architecture is designed to combine information extracted from historical sequences of meteorological data with physics-based constraints, providing stable and interpretable forecasts for daily photovoltaic energy production.

Implementation and experiments were conducted in Python/PyTorch on Google Colab using an NVIDIA T4 GPU. NumPy and pandas handled preprocessing, scikit-learn provided standardization, and training employed AdamW with Lookahead and an exponential moving average (EMA) of weights. No mixed-precision training was used. To support reproducibility, deterministic random seeds were set, and the notebooks are structured to run end-to-end in Google Colab.

3.1. System description and dataset

3.1.1. Photovoltaic system

The case study considers a grid-connected rooftop PV system in Southeast Europe (exact coordinates withheld for privacy). The array comprises 42 SANYO HIT-235 modules (STC 235 W each) for a 9.87 kW DC nameplate capacity, mounted south at 1.0° tilt under low shading and coupled to the grid via a single SMA 10000TL inverter. A "100 W inverter size" field appears in the source metadata; this inconsistency is treated as a metadata anomaly and excluded from analysis. The reported commissioning date is 22 May 2012. Daily PV energy and static descriptors were obtained from PVOutput.org (publicly accessible). To protect privacy, precise coordinates are anonymized, and no personally identifiable information is disclosed.

3.1.2. Data preparation

Daily meteorology was retrieved from Visual Crossing Weather (max/avg temperature, humidity, cloud cover, solar-energy proxy, precipitation coverage, wind speed, qualitative conditions, and sunrise/sunset) and merged with PVOutput records on a unified daily timestamp (Visual Crossing Corporation, 2025), (PVOutput.org, 2025). The "solarenergy" variable was converted from MJ m^{-2} to kWh m^{-2}. Data quality control applied conservative hard bounds and de-duplication, followed by a physics-consistency screen against a daily benchmark $E_{phys} = G \cdot A \cdot \eta_{ref} \cdot \cos(1°)$ with G being Global Solar Irradiance, A=42×1.3 m^2, η_{ref}=0.185, and temperature coefficient 0.004°C; a soft ratio tolerance was used to avoid over-filtering edge cases.

Seasonal outliers were then removed using a monthly interquartile range (IQR) rule, where observations lying outside [Q1−1.5·IQR, Q3+1.5·IQR] were excluded for each calendar month. Overall, ~10% of rows were filtered, yielding 3 729 daily samples spanning 2014–2025.

Feature engineering encoded cyclic seasonality (month/DOY sine–cosine), added a one-day lag of energy and exponentially weighted moving averages EWM (3,7) trends for temperature, humidity, and solar-energy, and included simple interactions (e.g., humidity × wind) with optional numeric mapping of qualitative conditions. Inputs were organized as 14-day sequences (N,14,F) to predict next-day energy.

To avoid temporal leakage and reflect real-world forecasting conditions, the dataset was split chronologically, yielding 81.3% training data, 11.7% validation data, and the final 7.0% held out for testing.

All features were standardized using train-only statistics; the target used log1p followed by standardization. The unscaled physical benchmark E_{phys} was computed on the target index, standardized with a train-fit scaler, and provided as an auxiliary physics channel to the model.

3.2. Physics-informed temporal attention model

3.2.1. Model architecture

To capture temporal dependencies, we employ an attention mechanism that computes a weighted context vector across the input window of 14 days. This mechanism learns to emphasize days that are more informative for predicting photovoltaic generation.

The architecture of the proposed Hybrid Physics-Informed Temporal Attention (H-PITA) model (illustrated in Figure 2) combines temporal attention with a gated physics-aware fusion mechanism to integrate observational data and a physics-based prior of photovoltaic (PV) energy production. The multivariate temporal input sequence $x_{seq} \in \mathbb{R}^{B\times S\times F}$ where B denotes the batch size, S the sequence length (number of past days), and F the number of input features per day, is first projected into a latent embedding space through a linear transformation, followed by layer normalization and a SiLU activation.

Temporal dependencies within the input window (here, S=14 days) are captured using a temporal attention mechanism, which computes normalized attention weights across the sequence and produces a context vector as a weighted aggregation of the embedded features. This allows the model to emphasize the most informative days for PV energy prediction.

The physics-informed component introduces a gating mechanism that conditions the influence of the physical input x_{phys} (a scaled estimate of daily PV energy derived from a physical model) on the learned temporal context. Specifically, a sigmoid-based gate generates adaptive scaling coefficients, producing a gated physical signal that reflects the relevance of the physics prior under the current temporal conditions. The learned context vector, the original physical input, and its gated counterpart are then concatenated and normalized to form a joint representation.

This fused representation is processed by a multi-layer perceptron (MLP) to generate the primary prediction in the standardized target space. In addition, a lightweight residual branch from x_{phys} is included via a linear mapping scaled by a small learnable parameter initialized near zero. This residual pathway preserves a direct but tightly constrained influence of the physical prior, improving stability and physical consistency without dominating the data-driven prediction.

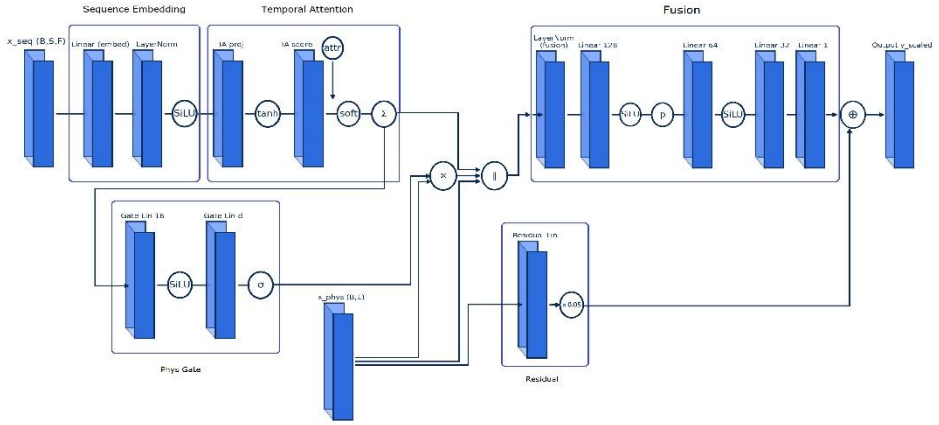

Figure 2. H-PITA Model Architecture.

3.2.2. Loss functions

The model is trained using a composite objective that balances fidelity to observed data with adherence to physical constraints. The primary data term is a weighted Huber loss, computed in real energy space (kWh) after inverting the log-scaling applied during preprocessing. To address outliers, higher weights are assigned to samples lying outside the interquartile range.

A physics loss is added, formulated as a relative Huber penalty between predicted and physically computed energy. This term is normalized by E_{phys} to preserve scale invariance, clipped to prevent domination by extreme values, and modulated differently depending on the magnitude of $E_{phys.}$ Soft hinge penalties further ensure non-negativity of predictions and prevent exceeding a dynamic upper bound defined as $1.05 \times E_{phys}$.

The total training objective is given by equation 1:

$$\mathcal{L} = \mathcal{L}_{data} + \lambda_{phys}(t) \cdot \mathcal{L}_{phys} + \lambda_{pos} \cdot \mathcal{L}_{pos} + \lambda_{cap} \cdot \mathcal{L}_{cap} \quad (1)$$

All constraint terms are applied in the real energy domain, where ŷ denotes the predicted daily PV energy in kWh.

Data loss ($\boldsymbol{\mathcal{L}_{data}}$): A weighted Huber loss computed in the real energy space (kWh), where larger weights are assigned to outlier samples to improve robustness.

Physics loss ($\boldsymbol{\mathcal{L}_{phys}}$): Measures deviation from E_{phys}, using a Huber-type formulation with clipping for stability and reduced weight for low-production days.

Non-negativity penalty ($\boldsymbol{\mathcal{L}_{pos}}$): Penalizes negative predictions given by equation 2:

$$\mathcal{L}_{pos} = \max(0, -\hat{y})^2 \quad (2)$$

Maximum capacity penalty ($\boldsymbol{\mathcal{L}_{cap}}$): Penalizes predictions exceeding $1.05 \cdot E_{phys}$ given by equation 3:

$$\mathcal{L}_{cap} = \max(0, \hat{y} - 1.05 \cdot E_{phys})^2. \quad (3)$$

Weighting coefficients: $\lambda_{phys}(t)$ increases gradually during early epochs (half-cosine ramp) and is updated dynamically within each batch to balance data and physics losses. λ_{pos} and λ_{cap} are small fixed values ($5 \times 10{-}4$) to softly enforce non-negativity and capacity limits without dominating the objective.

3.2.3. Training strategy

The model is trained with the AdamW optimizer wrapped in Lookahead for improved stability. A learning rate scheduler (ReduceLROnPlateau) adapts the step size based on validation performance, while exponential moving average (EMA) smoothing of model parameters further stabilizes training. Gradient clipping is employed to improve training stability.

Training proceeds for up to 3000 epochs with early stopping based on validation mean absolute percentage error (MAPE). Mini-batches of size 512 are used, and gradient accumulation ensures effective behaviour with larger batches. The dynamic physics weighting mechanism continuously adjusts the contribution of the physics loss per batch, constrained within ±10% relative changes to ensure smooth evolution.

This methodology yields a hybrid architecture that leverages physical priors while retaining the flexibility of deep learning, guided by carefully designed loss functions and stabilization techniques to ensure robustness and generalization.

The overall training procedure of the proposed H-PITA model is summarized in Algorithm 1. The pipeline includes data preparation and consistency checks, model construction, and an epoch–batch training loop with validation-based model selection. Physics-informed constraints are incorporated directly into the loss function, while exponential moving averages (EMAs) are used to stabilize metric tracking. Model selection is performed via validation-based early stopping, and the final checkpoint corresponds to the best validation performance.

Algorithm 1. Training of the H-PITA model

Input: Training and validation datasets (X_{seq}, X_{phys}, y, E_{phys}); maximum epochs E; patience P; target physics share τ.
Output: Trained H-PITA model.

Initialize H-PITA parameters, optimizer, learning-rate scheduler, and EMA.
Compute real-scale targets and sample weights using an IQR-based outlier rule.
Initialize dynamic physics weight w_{dyn} and base ramp schedule $w_{base}(t)$.
For each epoch t=1,…,E :
    Compute effective physics weight $w_{eff}=w_{base}(t)\cdot w_{dyn}$.
    For each mini-batch:
        Predict daily energy $\hat{y}$ and convert to real units (kWh).
        Compute data loss $\mathcal{L}_{data}$.
        Compute physics loss $\mathcal{L}_{phys}$ (relative Huber deviation from E_{phys}).
        Compute soft constraint penalties for non-negativity and capacity.
        Form total loss (Equation 1)

$$\mathcal{L} = \mathcal{L}_{data} + \lambda_{phys}(t) \cdot \mathcal{L}_{phys} + \lambda_{pos} \cdot \mathcal{L}_{pos} + \lambda_{cap} \cdot \mathcal{L}_{cap}$$

        Update model parameters and EMA.
        Adapt w_{dyn} to keep the physics loss close to the target share τ.
    End for
    Evaluate on validation set (EMA weights).
    Update learning rate and apply early stopping if needed.
End for
Restore the best model parameters and save the final model.

4. Results and analysis

4.1. Setup and metrics

We report MAE, RMSE, R^2, and MAPE. To avoid inflated percentage errors on very small-load days, MAPE is computed after excluding targets y<1 kWh; the number of excluded observations is explicitly reported. Performance metrics are first presented using a fixed train/validation/test split to reflect standard operational deployment. In addition,

rolling cross-validation is employed to assess temporal robustness, with confidence intervals used to quantify performance variability across folds. For a physics-aware sensitivity experiment, we optionally apply a physics cap to predictions given by equation 4:

$$\hat{y}_i^{cap} = \min\{\hat{y}_i\ , (1\ +\ \rho)E_{phys}, i\},\ \ \rho\ \in\ \{0.08, 0.12\} \tag{4}$$

The cap is not used in the main tables—only as an explanatory stress test.

Two data versions are considered: Unfiltered (Train=3368, Val=484, Test=291; MAPE drop counts: 4/0/1) and Filtered (Train=3019, Val=435, Test=260; MAPE drop counts: 0/0/0).

4.2. Model performance

This section presents model performance in two complementary stages. We first report results obtained using a fixed train/validation/test split, which reflect standard operational deployment and allow detailed analysis through multiple tables and figures. These results are used to assess absolute predictive accuracy, the impact of data filtering, and the effect of physics-aware capping.

We then extend the evaluation using rolling cross-validation to assess temporal robustness and statistical stability across different forecasting periods. Cross-validation results, reported with aggregated metrics, provide additional evidence that the observed performance trends are consistent over time rather than artifacts of a single data split.

Filtering substantially improves generalization. On the validation set, MAPE falls from 19.33% to 12.65% (−6.68 pp; −34.6% relative). On the test set, MAPE falls from 22.33% to 16.68% (−5.65 pp; −25.3% relative). RMSE and MAE also decrease, and R^2 improves slightly on both splits as shown in Table 1 and Figures 3 and 4.

Table 1. Baseline metrics (no cap) filtered vs unfiltered, fixed split

Dataset	Split	MAE	RMSE	R^2	MAPE	Removed
Filtered	Train	4.398	6.604	0.8372	14.35	0/3042
Unfiltered		5.106	8.046	0.8169	24.11	4/3393
Filtered	Val	4.171	5.633	0.8436	12.41	0/438
Unfiltered		5.016	7.526	0.7794	18.75	0/488
Filtered	Test	5.518	7.978	0.7833	16.68	0/262
Unfiltered		6.005	9.607	0.7539	22.42	2/293

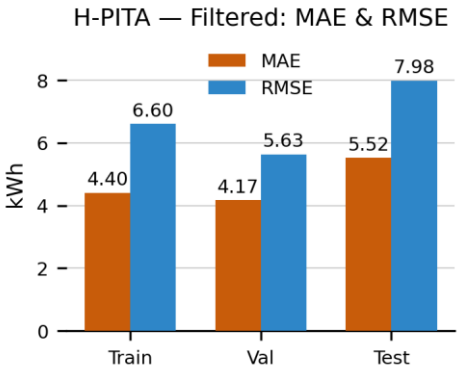

Figure 3. Results of H-PITA on filtered dataset

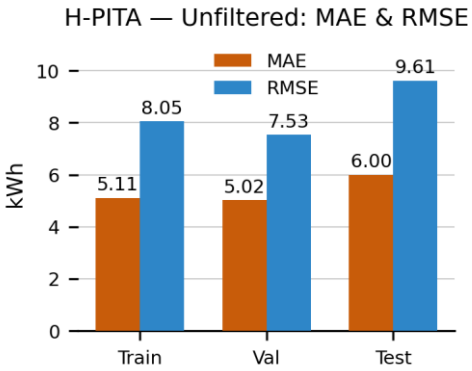

Figure 4. Results of H-PITA on unfiltered dataset

4.3. Physics-cap sensitivity (ρ = 0.08 & ρ = 0.12)

Applying the physics cap ($\hat{y} \leq (1 + \rho)E_{phys}$) decreases headline accuracy on both datasets. On the Unfiltered test set, MAPE increases by +1.99 pp (from 22.33% to 24.32%). On the Filtered test set, MAPE increases by +2.03 pp (from 16.68% to 18.71%). This suggests the physics proxy E_{phys} systematically underestimates some high-load days, forcing under-prediction. The detailed fixed-split results for the ρ=0.08 setting are reported in Table 2, while the broader effect of capping on RMSE and MAE across both datasets is illustrated in Figures 5 and 6, respectively. We therefore report uncapped metrics in the main tables and treat the cap as a diagnostic and safety constraint rather than a default operating mode.

Table 2. Metrics with physics cap (ρ = 0.08) on fixed split

Split	Dataset	MAE (kWh)	RMSE (kWh)	R^2	MAPE (%)
Train	Unfiltered + cap (8%)	6.522	9.236	0.7597	26.70
Val	Unfiltered + cap (8%)	5.877	8.246	0.7380	20.39
Test	Unfiltered + cap (8%)	6.520	8.960	0.7768	24.32
Train	Filtered + cap (8%)	5.971	8.184	0.7513	17.04
Val	Filtered + cap (8%)	5.066	6.593	0.7863	13.71
Test	Filtered + cap (8%)	6.292	8.275	0.7602	18.71

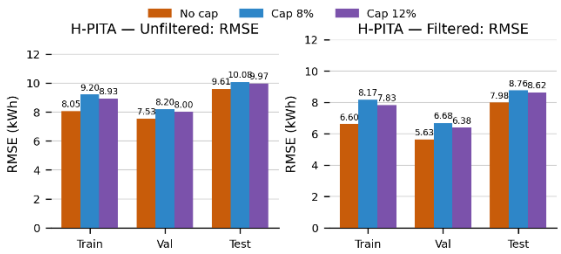

Figure 5. H-PITA RMSE unfiltered and filtered datasets, cap and no cap

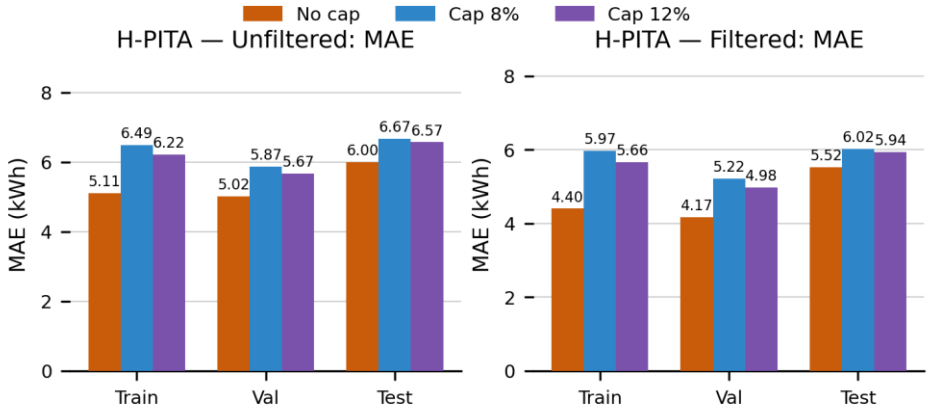

Figure 6. H-PITA MAE unfiltered and filtered datasets, cap and no cap

Imposing a physics cap at ρ = 8% or ρ = 12% consistently degrades predictive accuracy. On the filtered test set, MAE rises from 5.52 kWh (no cap) to 6.02 (+9.1%) with ρ = 8% and 5.94 (+7.7%) with ρ = 12%; RMSE increases from 7.98 to 8.76 (+9.8%) and 8.62 (+8.0%); MAPE from 16.68% to 17.63% / 17.55%; while R^2 drops from 0.783 to 0.739 / 0.747. The same pattern holds on the unfiltered test set: MAE 6.01 → 6.67 (+11.1%) / 6.57 (+9.5%), RMSE 9.61 → 10.08 (+5.0%) / 9.97 (+3.8%), MAPE 22.42% → 23.76% / 23.61%, and R^2 0.754 → 0.729 / 0.735. In short, the cap truncates high-load peaks—where much of the variance resides—biasing predictions downward and inflating error metrics; it is therefore best viewed as a safety constraint rather than a performance enhancer for H-PITA model.

To further assess the robustness of the proposed H-PITA model beyond a single train–validation–test split, we additionally evaluate predictive performance using rolling cross-validation. This evaluation protocol preserves temporal causality and enables a statistically sound assessment of performance stability across multiple forecasting horizons. The resulting MAE, RMSE, R^2, and MAPE metrics, averaged across all folds, are reported in Table 3, providing a complementary view of model behavior under realistic operational conditions.

Table 3. H-PITA rolling cross-validation performance

Physics cap (ρ)	MAE (kWh)	RMSE (kWh)	R^2	MAPE (%)
None (default)	5.55	7.59	0.741	20.58
0.08	5.98	7.80	0.738	16.93
0.12	5.74	7.54	0.754	16.69

Under rolling cross-validation, H-PITA achieves an average MAE of 5.55 kWh, RMSE of 7.59 kWh, R^2 of 0.74, and MAPE of 20.58% without physics capping. Applying a physics-aware cap improves MAPE (down to 16.69–16.93%) at the cost of a slight reduction in overall accuracy, confirming the cap's role as a robustness and safety mechanism rather than the default operating mode.

4.4. Baseline comparison

To benchmark the proposed H-PITA model against strong and representative baselines, we consider both classical forecasting methods and modern machine-learning approaches, including persistence (D–1), climatology (day-of-year), linear and ridge regression, LSTM with attention, and XGBoost.

4.4.1. Quantitative comparison

This section compares the proposed H-PITA model with the set of the baselines under two complementary evaluation settings. First, we report fixed train/validation/test split results for both the filtered and unfiltered datasets, using uncapped predictions (no physics cap) for all methods. This analysis reflects standard operational deployment and enables a direct comparison of absolute predictive performance across data quality conditions.

Second, to assess robustness and temporal stability, we evaluate all methods using rolling cross-validation on the filtered dataset only, again under the no-cap setting. This protocol preserves temporal causality and provides fold-averaged performance metrics that support a fair and statistically grounded comparison with H-PITA. Together, these two evaluation perspectives allow us to distinguish between absolute accuracy under fixed splits and consistency of performance over time.

Fixed train/validation/test split protocol (baselines vs. H-PITA): Under the fixed train/validation/test split protocol, H-PITA consistently outperforms all baseline models across both the filtered and unfiltered datasets. As reported in Table 4 (filtered data, no cap), H-PITA achieves the lowest errors and highest explanatory power on the test set (MAE = 5.518 kWh, RMSE = 7.978 kWh, R^2 = 0.7833), substantially improving upon classical baselines and advanced competitors such as XGBoost and LSTM with attention.

Table 4. Filtered data, all methods, no cap

Method	Train				Validation				Test			
	MAE	RMSE	R^2	MAPE%	MAE	RMSE	R^2	MAPE%	MAE	RMSE	R^2	MAPE%
Climatology DoY	8.300	10.811	0.5638	30.14	8.269	10.824	0.4225	30.08	9.036	12.084	0.5029	37.87
H-PITA	4.398	6.604	0.8372	14.35	4.171	5.633	0.8436	12.41	5.518	7.978	0.7833	16.68
LSTM+Attn	7.586	10.879	0.5582	30.26	6.845	10.558	0.4505	27.35	8.025	11.668	0.5365	34.64
Linear	9.244	11.369	0.5176	28.21	8.933	11.124	0.3901	28.19	9.106	11.439	0.5545	30.85
Persistence D-1	9.154	13.723	0.2965	31.17	7.654	12.136	0.275	27.23	7.985	12.184	0.4881	27.65
Ridge	9.217	11.321	0.5216	28.29	8.636	10.771	0.4282	27.49	9.037	11.281	0.5667	30.59
XGBoost	3.914	5.145	0.9012	13.31	6.605	9.513	0.5539	25.19	7.792	10.589	0.6183	31.98

Similar trends are observed for the unfiltered dataset in Table 5, where H-PITA maintains superior generalization despite increased noise and data variability. These quantitative results are further corroborated by the comparative visual analysis: Figure 7 and Figure 8 show consistently lower validation MAPE and RMSE for H-PITA relative

to all baselines, while Figure 9 and Figure 10 confirm that these advantages persist on the test set for both dataset variants.

Table 5. Unfiltered data, all methods, no cap

Method	Train				Validation				Test			
	MAE	RMSE	R^2	MAPE%	MAE	RMSE	R^2	MAPE%	MAE	RMSE	R^2	MAPE%
Climatology DoY	10.956	13.927	0.4513	64.00	10.365	13.796	0.2587	55.05	10.942	14.409	0.4465	66.99
H-PITA	5.106	8.046	0.8169	24.11	5.016	7.526	0.7794	18.75	6.005	9.607	0.7539	22.42
LSTM+Attn	10.122	14.014	0.4445	63.54	8.807	13.275	0.3137	50.51	9.160	13.510	0.5134	61.55
Linear	11.981	14.76	0.3837	52.11	11.590	14.242	0.2101	47.20	11.55	14.227	0.4604	51.21
Persistence D-1	11.505	17.033	0.1796	58.40	10.156	15.523	0.0618	51.16	10.02	15.242	0.3825	47.00
Ridge	11.970	14.698	0.3889	52.34	11.471	14.025	0.2339	46.86	11.501	14.137	0.4672	50.94
XGBoost	8.741	11.255	0.6416	50.00	8.736	12.089	0.4309	47.12	9.565	12.588	0.5775	55.74

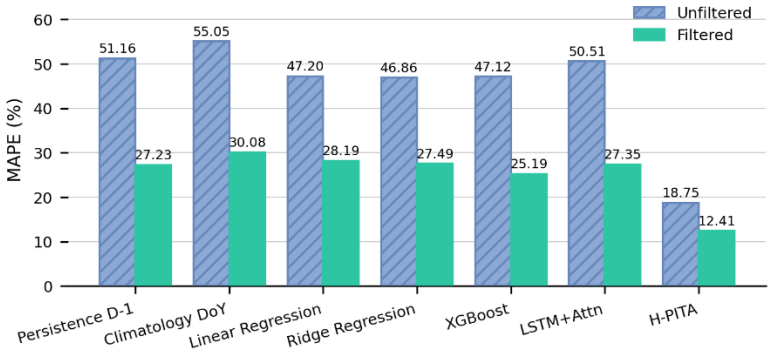

Figure 7. Validation MAPE - H-PITA vs baselines (unfiltered and filtered datasets)

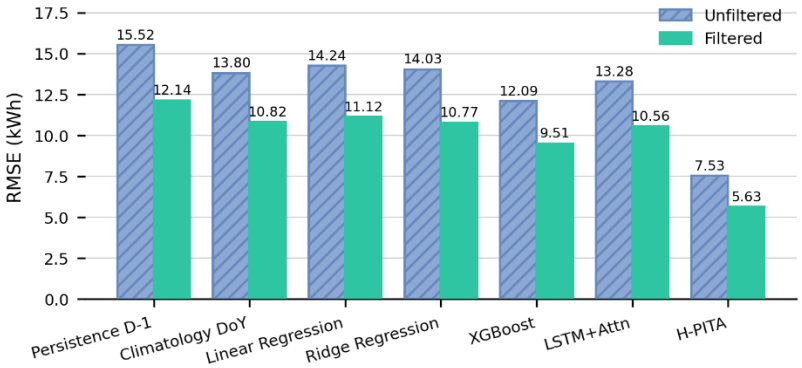

Figure 8. Validation RMSE - H-PITA vs baselines (unfiltered and filtered datasets)

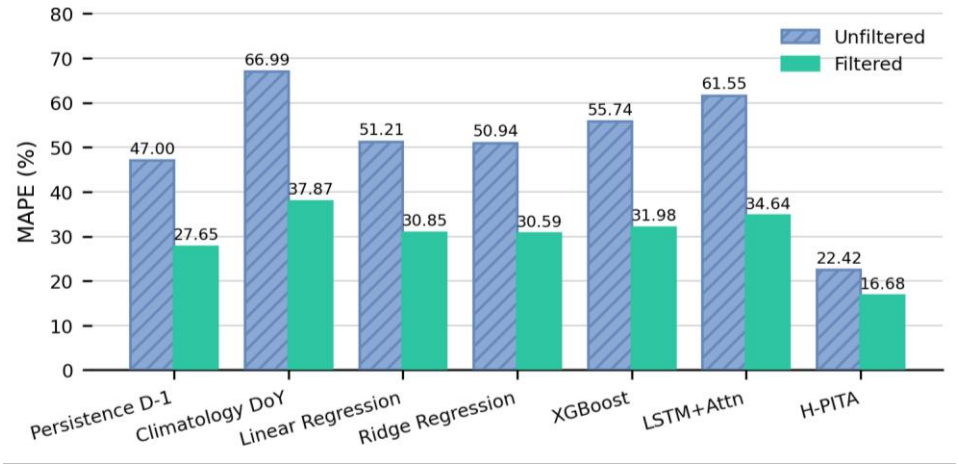

Figure 9. Test MAPE - H-PITA vs baselines (unfiltered and filtered datasets)

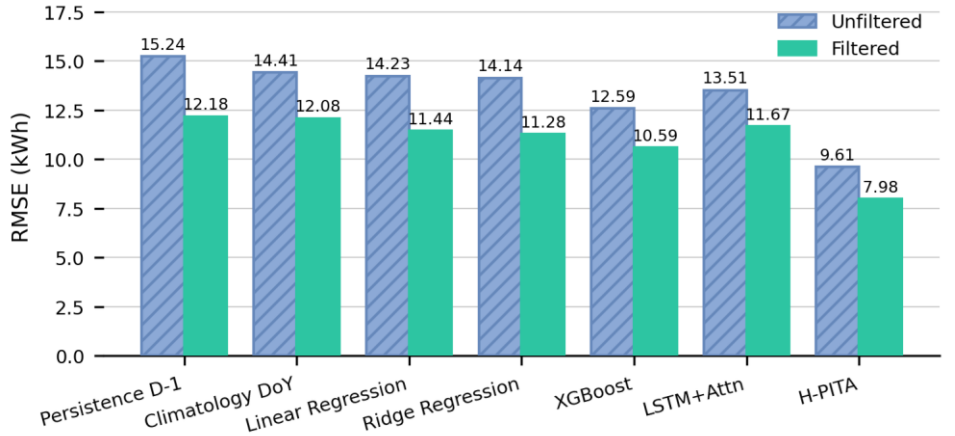

Figure 10. Test RMSE - H-PITA vs baselines (unfiltered and filtered datasets)

Overall, the agreement between tabular metrics and graphical trends demonstrates that the proposed H-PITA model delivers robust and stable performance gains under standard fixed-split evaluation, reinforcing its effectiveness compared to both traditional and data-driven baseline approaches.

Rolling cross-validation protocol (baselines vs. H-PITA): To ensure a fair and consistent comparison, rolling cross-validation is applied uniformly to H-PITA and all baseline models, including climatology, persistence, linear and ridge regression, LSTM with attention, and XGBoost. Each method is evaluated using the same five expanding-origin folds, identical input features, and the same temporal splits. Performance metrics

(MAE, RMSE, R^2, and MAPE) are averaged across folds and computed for all physics-cap settings. While capped variants are evaluated as part of the sensitivity analysis, we report uncapped (no-cap) results in Table 6 for clarity and comparability.

Importantly, H-PITA consistently outperforms all baseline models across both capped and uncapped settings, indicating that its superior performance is robust to the inclusion of physics-aware constraints. This evaluation protocol ensures that observed performance differences reflect genuine modeling capabilities rather than advantages arising from data access, tuning procedures, or temporal leakage, thereby enabling a balanced comparison between H-PITA and competing approaches.

Table 6. Rolling cross-validation results (baselines vs. H-PITA)

Method	MAE	RMSE	R2	MAPE
Climatology	8.7278	11.7009	0.412	32.4726
H-PITA	5.5509	7.5882	0.741	20.5754
Linear	8.2840	10.8618	0.493	29.6320
LSTM	8.0126	10.5507	0.522	27.2028
Persistence	8.3368	12.7965	0.295	28.9652
Ridge	8.0771	10.6347	0.514	29.3052
XGBoost	7.636183	10.54602	0.522	29.69312

4.4.2. Statistical significance analysis

While aggregate error metrics provide an overall view of predictive accuracy, they do not indicate whether observed performance differences between models are statistically meaningful. To assess the significance of the improvements achieved by H-PITA over baseline approaches, we employ Diebold–Mariano (DM) tests on rolling cross-validation forecasts. The DM test evaluates the null hypothesis of equal predictive accuracy using paired forecast errors, allowing a rigorous comparison across time while accounting for temporal dependence. By applying this analysis to all baseline models and physics-cap settings, we determine whether the observed gains of H-PITA reflect systematic improvements rather than random variation. Table 7 reports the p-values of the Diebold–Mariano test for the statistical comparison of predictive accuracy between H-PITA and baseline models. Diebold–Mariano tests conducted on rolling cross-validation forecasts consistently reject the null hypothesis of equal predictive accuracy between H-PITA and all baseline models.

Across all physics-cap settings, the DM statistics are negative and highly significant ($p < 10^{-6}$), indicating systematically lower forecast errors for H-PITA.

Table 7. p-values of the Diebold–Mariano test comparing the predictive accuracy of H-PITA against baseline models.

Physics cap (ρ)	Climatology	Linear	LSTM	Persistence	Ridge	XGBoost
0.08	9.48e-13	1.73e-13	5.35e-09	2.17e-20	6.23e-11	2.50e-06
0.12	7.80e-14	9.66e-15	4.40e-10	9.41e-21	3.86e-12	2.66e-07
None	8.17e-23	8.28e-19	2.93e-13	2.54e-19	1.26e-16	8.06e-16

The strongest statistical advantage is observed in the uncapped setting, where H-PITA outperforms LSTM- and tree-based baselines on more than 60% of test instances. Physics-aware caps reduce daily extremes and improve robustness, but slightly attenuate the relative advantage, confirming their role as a conservative safety mechanism rather than a default-operating mode.

5. Discussion

This study proposed H-PITA, a Hybrid Physics-Informed Temporal Attention model for daily photovoltaic energy forecasting, and evaluated it against a broad set of classical and machine-learning baselines under multiple evaluation protocols. The results demonstrate that H-PITA achieves superior predictive accuracy and robustness in both fixed-split evaluations and rolling cross-validation, highlighting the effectiveness of integrating temporal attention with physics-informed constraints.

Across both filtered and unfiltered datasets, H-PITA yields lower MAE, RMSE, and higher R^2 than all baselines, including strong competitors such as XGBoost and attention-enhanced LSTM models. The performance gains are most pronounced on the filtered dataset, where improved data quality allows the model to better exploit temporal dependencies and physical structure. This confirms that careful data curation is essential for enhancing generalization, particularly in hybrid models that combine data-driven and physics-based components.

The rolling cross-validation results further indicate that the observed improvements are stable over time and not driven by favorable fixed splits. Under expanding-origin evaluation, H-PITA consistently outperforms all baseline methods, achieving the best average performance across all reported metrics. Diebold–Mariano tests provide strong statistical evidence that these gains are systematic rather than coincidental, with extremely small p-values across all comparisons, even under different physics-cap settings.

A key characteristic of H-PITA is the explicit incorporation of physical knowledge through an external physics prior and adaptive loss weighting. Sensitivity analysis with physics-aware capping shows that while hard caps can enforce physical plausibility, they may slightly reduce accuracy on high-production days when the physics proxy underestimates peak output. This suggests that physics constraints are most effective when applied softly during training or used as a diagnostic and safety mechanism at inference, rather than as a strict operational limit. Accordingly, uncapped configurations offer the best balance between accuracy and flexibility, while capped variants provide an additional robustness option for risk-aware applications.

From an operational perspective, these findings are highly relevant for energy management and grid integration. Reliable PV forecasts support scheduling, market participation, and asset management, while physics-aware safeguards add confidence in safety-critical settings. The modular design of H-PITA—allowing plug-in physics estimates and dynamic loss weighting—facilitates adaptation to different PV systems, data availability conditions, and forecasting horizons.

This work does not aim to exhaustively isolate the contribution of each architectural component through ablation. Instead, robustness analyses across datasets, evaluation protocols, and physics-cap configurations provide practical evidence of the model's effectiveness. Future work may extend the framework toward probabilistic forecasting,

multi-site evaluation, and more detailed physical modeling to enhance uncertainty awareness and interpretability.

Overall, the results show that combining temporal attention with physics-informed learning yields a stable and effective forecasting framework. By embedding domain knowledge directly into the learning process, H-PITA advances beyond purely data-driven baselines and offers a compelling solution for reliable photovoltaic energy forecasting in real-world operational contexts.

6. Conclusion

This paper introduced H-PITA, a Hybrid Physics-Informed Temporal Attention model for daily photovoltaic energy forecasting, and validated it against a comprehensive set of baseline methods using both fixed-split and rolling cross-validation protocols. The results show that combining temporal attention with physics-informed learning leads to consistently improved accuracy and robustness compared to purely data-driven approaches. By embedding domain knowledge directly into the learning process, H-PITA provides a reliable and practically deployable forecasting framework, well suited for real-world photovoltaic energy management and operational decision support.

References

Al-Ali, E.M., Hajji, Y., Said, Y., Hleili, M., Alanzi, A.M., Laatar, A.H., Atri, M. (2023). Solar energy production forecasting based on a hybrid CNN–LSTM–Transformer model. Math. 11, 676. https://doi.org/10.3390/math11030676

Alcañiz, A., Grzebyk, D., Ziar, H., Isabella, O. (2023). Trends and gaps in photovoltaic power forecasting with machine learning. Energy Reports, 9, 447–471. https://doi.org/10.1016/j.egyr.2022.11.208

Barhmi, K., Heynen, C., Golroodbari, S. G. M., van Sark, W. (2024). A review of solar forecasting techniques and the role of artificial intelligence. Solar, 4(1), 99–135. https://doi.org/10.3390/solar4010005

Buonanno, A. et al. (2024). Machine learning and weather-model combination for PV production forecasting. Energies, 17(9), https://doi.org/10.3390/en17092203

Di Leo, P., Ciocia, A., Malgaroli, G., Spertino, F. (2025). Advancements and challenges in photovoltaic power forecasting: A comprehensive review. Energies, 18(8), 2108. https://doi.org/10.3390/en18082108

Dwivedi, V., Kadian, K., Gardiner, B., Pandey, R., Lathwal, A. (2025). Supply chain analytics: A performance evaluation of machine learning, statistical, and time-series models. *Baltic J. Modern Computing*, **13**(2), 453–485. https://doi.org/10.22364/bjmc.2025.13.2.09

Falas, S., Asprou, M., Konstantinou, C., Michael, M. K. (2025). Robust power system state estimation using physics-informed neural networks. https://arxiv.org/abs/2507.05874

Hu, Z., Gao, Y., Ji, S., Mae, M., Imaizumi, T. (2024). Improved multistep-ahead PV power prediction based on LSTM and self-attention with weather forecast data. Applied Energy, 359, https://doi.org/10.1016/j.apenergy.2024.122709

IEA PVPS Task 16 (2025). The added value of combining solar irradiance data and forecasts (Technical report). https://iea-pvps.org/

Lauret, P., Alonso-Suárez, R., Le Gal La Salle, J., David, M. (2022). Solar forecasts based on the clear-sky index or the clearness index: Which is better? Solar, 2(4), 432–444. https://doi.org/10.3390/solar2040026

Lei, X. (2024). A photovoltaic prediction model with integrated attention mechanism. Mathematics, 12(13), https://doi.org/10.3390/math12132103

Liu, Y., Wang, J., Liu, L. (2024). Physics-informed reinforcement learning for probabilistic wind power forecasting under extreme events. Applied Energy, 376, https://doi.org/10.1016/j.apenergy.2024.124068

Maciel, J. N., Gimenez Ledesma, J. J., Ando, O. H., Jr. (2024). Dataset for machine learning: Explicit all-sky image features to enhance solar irradiance prediction. Data, 9(10), https://doi.org/10.3390/data9100113

Nithya, C., Roselyn, J. P., Devaraj, D. (2024). Real-time solar PV generation in a building using LSTM-based time-series forecasting. Discover Electronics, 1, https://doi.org/10.1007/s44291-024-00023-0

Özkurt, C., Canay, Ö., Tunç, E. A., Aydın, E., Velioğlu, B. S. (2025). Renewable energy source ranking and analysis using fuzzy MCDM, ML, and XAI techniques. *Baltic J. Modern Computing*, **13**(3), 656–679. https://doi.org/10.22364/bjmc.2025.13.3.06

PVOutput.org. (2025). PVOutput—Free solar power monitoring for users and installers. https://pvoutput.org/

Raissi, M., Perdikaris, P., Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Sanfilippo, S., Hernandez-Cabrera, J. J., Kandler, C., Hernandez-Galvez, J. J., Evora-Gomez, J., & Roncal-Andres, O. (2025). A neural network-based causal model for electricity demand estimation in remote areas: A case study in El Espino, Bolivia. *Baltic J. Modern Computing*, **13**(2), 315–330. https://doi.org/10.22364/bjmc.2025.13.2.01

Sneiders, M., Urtans, E., Abu Saa, A. (2025). Predicting student performance on a novel Moodle dataset using GRU time series model. *Ba*ltic J. Modern Computing, 13(4), 885–893. https://doi.org/10.22364/bjmc.2025.13.4.07

Song, Z., Xiao, F., Chen, Z., Madsen, H. (2025). Probabilistic ultra-short-term solar PV power forecasting using NGBoost with attention-enhanced neural networks. Energy and AI, 20, Article 100496. https://doi.org/10.1016/j.egyai.2025.100496

Stiasny, J., Zhang, B., Chatzivasileiadis, S. (2024). PINNSim: A simulator for power system dynamics based on physics-informed neural networks. Electric Power Systems Research, 235, https://doi.org/10.1016/j.epsr.2024.110796

Urhuerhi, O. P., Udomboso, C., Sigauke, C. (2025). Forecasting solar power output in Ibadan: A machine learning approach leveraging weather data and system specifications. https://arxiv.org/abs/2508.07462

Vázquez Pombo, D., Bindner, H. W., Spataru, S. V., Sørensen, P. E., Bacher, P. (2022). Increasing the accuracy of hourly multi-output solar power forecast with physics-informed machine learning. Sensors, 22(3), Article 749. https://doi.org/10.3390/s22030749

Visual Crossing Corporation. (2025). Visual Crossing Weather—Historical & forecast weather data (API service). https://www.visualcrossing.com/

Zhong, Y., He, T., Mao, Z. (2024). Enhanced solar power prediction using attention-based DiPLS-BiLSTM model. Electronics, 13(23), https://doi.org/10.3390/electronics13234815

Received January 2, 2026, revised April 26, 2026, accepted May 15, 2026