Journal of Engineering, Electrical and Informatics

E-ISSN: 2809-8706
P-ISSN: 2810-0557

Research Article

An Exploratory Evaluation of One-Hour Solar Irradiance Forecasting in the Tropics Using Physics-Constrained Extreme Learning Machine

Tyas Wedhasari

Universitas Mercu Buana, Indonesia; Email : tyas.wedhasari@mercubuana.ac.id
*Corresponding Author: Tyas Wedhasari

Abstract: This study evaluates one-hour resolution Global Horizontal Irradiance (GHI) forecasting in tropical Indonesia using NASA POWER satellite reanalysis data. Three approaches are comparatively evaluated: Persistence, a temperature-corrected radiative physics model, and an Extreme Learning Machine (ELM) utilizing six selected features alongside a physical capping mechanism. The evaluation was conducted on 9,691 identical test samples using a 60:20:20 chronological validation protocol. The results indicate that the ELM achieved the highest accuracy (RMSE: 10.70 W/m², MAE: 5.56 W/m², sMAPE: 4.03%) with a Skill Score of 0.885, representing an 88.5% error reduction compared to the baseline. The physical capping mechanism proved effective in stabilizing the sMAPE by constraining predictions within atmospheric physical limits. This exploratory study demonstrates that a straightforward integration of data-driven approaches and physical constraints holds potential as a foundational alternative in developing irradiance forecasting frameworks for renewable energy integration in tropical regions.

Keywords: Extreme Learning Machine, Floating Photovoltaics, Physical Capping, Radiative Physics, Saguling Reservoir, Solar Irradiance Forecasting.

Received: April 10, 2026;
Revised: May 12, 2026;
Accepted: June 15, 2026;
Published: July 27, 2026;
Curr. Ver.: July 27, 2026.

1. Introduction

Within the energy transition framework aligned with renewable energy mix targets, harnessing clean energy sources has become a strategic priority. One alternative gaining increasing technical and economic attention is the floating solar power plant, which utilizes the surfaces of reservoirs, dams, or lakes. This concept not only conserves productive land but also reduces water evaporation rates and naturally lowers panel temperatures through the evaporative cooling effect. Furthermore, it holds the potential for synergistic integration with existing hydroelectric power plants to form a more flexible and reliable hybrid system.

A reservoir in West Java represents a strategic location for the development of utility-scale floating solar power plants. As part of the vital hydroelectric network within the Java-Madura-Bali grid system, this reservoir offers an extensive water surface with minimal topographic shading and close proximity to existing substations. However, integrating intermittent solar capacity into a dynamically managed grid demands reliable solar irradiance forecasting, particularly at an hourly time resolution. Rapid irradiance fluctuations caused by tropical cloud cover patterns can lead to voltage instability, increased spinning reserve requirements, and a higher risk of energy curtailment if not properly anticipated by operational scheduling systems. Therefore, developing a one-hour-ahead Global Horizontal Irradiance (GHI) forecasting model presents a promising alternative method worth further exploration to balance grid loads and enhance the overall efficiency of hydro-solar hybrid systems.

This study aims to develop and evaluate a one-hour resolution Global Horizontal Irradiance (GHI) forecasting framework at a reservoir in West Java, utilizing exclusively satellite-based meteorological reanalysis data. The primary objective is to investigate whether hourly meteorological variables including surface temperature, relative humidity, corrected precipitation, wind speed, and clear-sky irradiance ratio are sufficient to capture GHI variability in humid tropical regions without relying on measurements from surface stations with limited coverage (Niu et al., 2025).

The research objectives focus on three fundamental aspects. First, the methodological aspect: this study implements a rigorous validation protocol using a 60:20:20 chronological data split to prevent any future information leakage into the model training process, ensuring that the evaluation results accurately reflect operational performance under real-time conditions. This split was deliberately chosen to provide an adequate training proportion for the machine learning model to map multivariate non-linear relationships, while maintaining an exclusive validation subset for hyperparameter and physical parameter calibration without contaminating the final test set. Second, the comparative aspect: this study systematically compares three forecasting approaches, the persistence model as a universal baseline, a radiative physics model with empirical temperature correction integrating atmospheric transmittance principles, and an Extreme Learning Machine (ELM) equipped with a physical constraint mechanism (physical capping) to quantify the trade-offs among accuracy, interpretability, and computational speed in operational contexts (Voyant et al., 2025) (Attya et al., 2025) (Rajagukguk et al., 2021). Third, the technical-applied aspect: this study examines the effectiveness of applying clear-sky envelope-based post-processing (clear-sky capping) to stabilize evaluation metrics such as sMAPE, which is mathematically prone to bias during low irradiance periods, while ensuring that the generated predictions remain within the physical limits of the real atmosphere.

The primary challenge in irradiance forecasting in tropical regions like West Java lies in complex atmospheric dynamics, where the formation and movement of convective clouds often occur suddenly, are highly localized, and are difficult to predict using simple statistical approaches. While ground-based measurement stations provide high-accuracy data, their coverage is limited, maintenance is costly, and continuous data is rarely available in operational reservoir areas. As a physics-based alternative, satellite-based meteorological reanalysis data offers extensive spatial coverage, temporal consistency, and the availability of supporting variables such as surface temperature, relative humidity, corrected precipitation, wind speed, and clear-sky irradiance ratio (Suwadana et al., 2025). Thermodynamically and optically, these variables are directly correlated with solar radiation transmittance through the atmospheric column; humidity and aerosols govern absorption and scattering, precipitation indicates the presence of thick clouds, and the all-sky to clear-sky irradiance ratio mathematically represents the cloud transmission factor, as validated in tropical solar resource studies. Consequently, modeling GHI by relying solely on hourly meteorological data represents a physically justifiable approach, particularly when on-site measurements are unavailable or economically unfeasible.

Given the operational context and the aforementioned data limitations, this study aims to evaluate the reliability of one-hour-ahead GHI forecasting using only meteorological data, and to compare the performance of three approaches: a persistence baseline model, a radiative physics-based model with empirical temperature correction, and an Extreme Learning Machine (ELM) equipped with a physical constraint mechanism (physical capping).

The underlying research questions of this study are whether gridded meteorological variables alone can adequately capture GHI variability in a dynamic tropical climate, the extent of accuracy improvement that can be achieved by machine learning methods compared to purely physics-based approaches, and how applying physical constraints to predictions can stabilize evaluation metrics like sMAPE, which are mathematically prone to bias when irradiance values approach zero. This approach is designed to deliver a forecasting framework that is not only statistically accurate but also physically consistent, enabling its adoption in real-time operational control systems without the need for complex additional measurement infrastructure.

This article presents the results of a comparative evaluation based on a strict 60:20:20 chronological data split to prevent data leakage and represent real-world operational conditions. Through the analysis of calibrated RMSE, MAE, and sMAPE metrics, alongside the application of clear-sky envelope-based post-processing, this study is expected to provide practical guidelines for floating solar developers, power system operators, and policy planners in designing stable and sustainable solar energy integration strategies(de Myttenaere et al., 2016). Furthermore, the findings of this research are expected to pave the way for developing physics-machine hybrid architectures that are more adaptive to future tropical microclimate changes.

2. Literature Review

Hourly resolution Global Horizontal Irradiance (GHI) forecasting is a technical prerequisite for integrating PV systems into tropical power grids, where convective cloud dynamics cause rapid fluctuations that are difficult to model deterministically. The academic literature categorizes forecasting approaches into three main paradigms: simple statistical methods, atmospheric physics-based models, and data-driven machine learning approaches. Each paradigm possesses distinct mathematical foundations, operational assumptions, and trade-offs among accuracy, interpretability, and computational complexity, which serve as key considerations when selecting a method for specific applications.

The persistence model serves as a universal baseline in renewable energy literature due to its straightforward implementation and the absence of a training process. Its basic formulation is:

$$\hat{y}_{t} = y_{t-1}$$

where the forecast at time *t* is derived directly from the observation at time *t-1*. Although its accuracy is limited to stable atmospheric conditions, this model is consistently used as a standard benchmark to calculate the performance improvement metric (Skill Score) using the following formula:

$$Skill\ Score = 1 - (RMSE_{\mathrm{moel}} / RMSE_{\mathrm{persistens}})$$

as recommended in international forecasting evaluation guidelines. The primary limitation of the persistence model lies in its inability to capture rapid transitions caused by cloud formation or dissipation, leading to a significant performance drop in regions with high cloud variability, such as West Java.

Radiative physics-based approaches are rooted in the principles of atmospheric transmittance, which model the attenuation of solar radiation resulting from its interaction with air molecules, aerosols, water vapor, and cloud particles. An empirical model derived from the Kasten-Czeplak approach, modified with a temperature correction, can be formulated as:

$$GHI_{\mathrm{pre}}d = \mathrm{GHI}_{CS} \times [1 - k \times (oktas/8)^{\mathrm{p}}]$$

with a dynamic attenuation coefficient:

$$k = k_0 + k_1 \times (T_{2\mathrm{m}} - T_{\mathrm{re}}f)$$

In this formulation, *GHI*_CS represents clear-sky irradiance, cloud cover in oktas serves as a proxy for cloudiness (derived from the all-sky/clear-sky ratio), and the parameters k_0, p, and k_1 are calibrated to local conditions. The primary advantage of this approach lies in its physical transparency and prediction consistency within atmospheric thermodynamic limits; however, its limitation arises in capturing non-linear cloud dynamics that are not explicitly represented in the deterministic formulation.

Over the past decade, machine learning methods have emerged as a powerful alternative capable of mapping multivariate non-linear relationships between meteorological variables and irradiance without relying on explicit physical formulations. The Extreme Learning Machine (ELM) is a single-layer neural network architecture with input weights (*W*) and biases (*b*) that are randomly initialized and kept fixed. The hidden layer output is calculated using a sigmoid activation function:

$$H = 1 / (1 + e^{-(\mathrm{xw+b})})$$

while the output weights (β) are solved analytically using the Moore-Penrose pseudo-inverse:

$$\beta = H^{+} \times Y$$

This formulation avoids computationally expensive gradient descent iterations, enabling training to be completed in a single linear algebra operation with exceptionally short execution times. However, as a purely data-driven model, ELM is prone to generating predictions that violate physical atmospheric limits for instance, predicting irradiance values that exceed the clear-sky value (*GHI*_CS), especially when the training data does not fully encompass the spectrum of extreme meteorological conditions (Al-Dahidi et al., 2018) (Le Cadre et al., 2015).

To address this inconsistency, recent literature has integrated physical constraints into the machine learning pipeline through a post-processing mechanism known as physical capping:

$$\hat{y}_{cap} = min(\hat{y}_{\mathrm{pre}}d, GHI_{cs})$$

This mechanism preserves the inference speed and statistical accuracy of the ML model while ensuring prediction consistency within a realistic radiative envelope. The use of the *GHI*_all / *GHI*_cs ratio as a proxy for the cloud transmission factor has been validated as a robust alternative when high-resolution cloud optical data is unavailable. In the context of forecasting based on gridded meteorological data (such as NASA POWER), surface variables like temperature, relative humidity, precipitation, and wind speed correlate significantly with hourly GHI variability when combined with temporal features and lagged variables. However, the literature remains limited in providing rigorous comparative evaluations that apply a 60:20:20 chronological data split protocol to prevent data leakage, particularly within the operational context of floating PV systems. This study presents a systematic evaluation of these three forecasting paradigms using a rigorous validation protocol and the integration of physical constraints, thereby providing empirical evidence on the accuracy limits and operational feasibility of each approach in supporting the clean energy transition in Indonesia's tropical regions.

Primary Data Resources

The meteorological and solar radiation datasets from NASA POWER (Prediction of Worldwide Energy Resources) were accessed via the Data Access Viewer interface. This dataset was selected because it provides hourly variables with global spatial coverage and long-term temporal consistency, and it has been validated for tropical regions such as Indonesia. The study location is focused on the coordinates of a reservoir in West Java, with the target variable being Global Horizontal Irradiance (GHI), which is represented by the ALLSKY_SFC_SW_DWN parameter.

Feature Selection Process

Input variable selection was conducted using a hybrid approach that integrates atmospheric physics justification, multicollinearity statistical analysis (Variance Inflation Factor), and predictive contribution evaluation via permutation importance. Based on this analysis, six final features were selected for the Extreme Learning Machine (ELM) model: lag_1 (short-term atmospheric memory), cos_SZA (solar incidence geometry), cloud_ratio (cloud cover proxy), cos_hr (diurnal cycle), T2M (surface temperature), and RH2M (relative humidity). These six variables were selected because they exhibited a significant correlation with the target ($|r| > 0.1$), controlled multicollinearity (VIF < 35 for physically related cyclic features), and a unique contribution to prediction accuracy (ΔRMSE > 5 W/m^2 when shuffled). Other variables, such as PRECTOTCORR and WS10M, were eliminated due to their marginal contribution to model performance, resulting in a more parsimonious architecture without sacrificing accuracy.

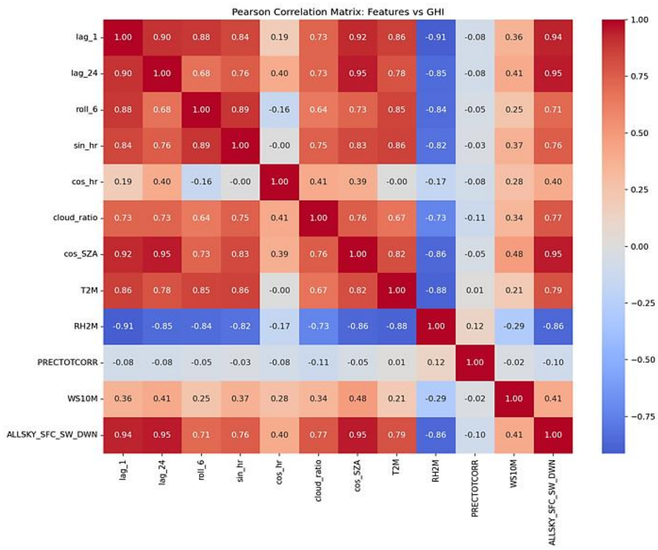

Figure 1. Pearson Correlation Matrix: Features vs GHI

Data Pre-processing Stage

The data pre-processing pipeline was designed to ensure dataset quality and consistency prior to modeling. The process began with handling missing values in the NASA POWER dataset, which were converted to NaN and interpolated using a time-based method with a maximum gap limit of six consecutive hours to preserve the integrity of the diurnal pattern. Next, a critical feature alignment was performed: the dataset was trimmed to compensate for data loss caused by lag and rolling window operations. The data splitting stage was conducted chronologically using a strict 60:20:20 protocol for the training set (29,073 data points), validation set (9,691 data points), and test set (9,691 data points)

(Solano et al., 2022). The chronological approach was deliberately chosen to prevent data leakage that could occur if the split were performed randomly on time series data, ensuring that the model evaluation truly reflects operational performance under real-time conditions. The validation set was used exclusively for calibrating the ELM model hyperparameters (number of hidden neurons) and physical parameters, while the test set was used only once during the final evaluation stage to ensure the objectivity of the results.

Forecasting Model Selection

This study is based on comparing three paradigms that represent a spectrum of complexity, ranging from simple to advanced. The persistence model was selected as a universal baseline due to its straightforward implementation $\hat{Y}_t = y_{t-1}$ and the absence of a training process. The radiative physics-based model was adopted using a Kasten-Czeplak-style attenuation formulation, modified with an empirical temperature correction: $\widehat{GHI} = GHI_{cs} \times [1 - k \times (oktas/8)^p]$, dengan $k = k_0 + k_1(T_{2m} - T_{ref})$. This approach was chosen for its transparent physical interpretability. As a representation of the data-driven approach, the Extreme Learning Machine (ELM) was selected due to its exceptional training speed and universal approximation capabilities. To address the limitation of the ELM model, which is prone to producing predictions outside physical limits, this study integrates a physical capping mechanism as post-processing $\hat{y}_{capped} = \min(\hat{y}_{pred}, GHI_{cs})$, thereby ensuring physical consistency without sacrificing statistical accuracy.

Model performance evaluation was conducted using three complementary error metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and symmetric Mean Absolute Percentage Error (sMAPE). The evaluation results on the test set, consisting of 9,691 identical hourly samples, showed that the ELM model with 6 selected features and physical capping achieved the best performance with an RMSE of 10.70 W/m², an MAE of 5.56 W/m², and an sMAPE of 4.03%. This achievement significantly outperformed the physics-based model (RMSE: 67.76 W/m²) and the persistence baseline (RMSE: 93.43 W/m²). This accuracy improvement is reflected in a Skill Score of 0.885, meaning the ELM model reduced the baseline error by 88.5% under the same operational conditions. The drastic reduction in sMAPE from an initial value of >100% (without capping) to 4.03% confirms the effectiveness of the physical constraint mechanism in stabilizing the evaluation during low irradiance periods, where the sMAPE formula is mathematically prone to bias. The discussion of the results emphasizes that although the physics-based model offers interpretability and instantaneous inference speed (<0.01 seconds), its inability to capture non-linear cloud dynamics limits its accuracy in a dynamic tropical climate. Conversely, the ELM successfully mapped the complex interactions between the six selected meteorological variables and GHI with an inference time of <0.02 seconds, making it a strong candidate for real-time forecasting applications in floating PV operations. This study also acknowledges the limitation of using gridded reanalysis data, which may not capture micro-local variability.

4. Result and Discussion

The performance evaluation of the three forecasting models, Persistence, Physical Radiative Transfer with temperature correction, and Extreme Learning Machine (ELM) with a physical capping mechanism—was rigorously conducted on a test set consisting of 9,691 independent hourly samples, representing the final 20% of the chronological time series following the feature alignment process. Quantitative results demonstrate that the ELM model, utilizing six selected features (*lag1*, *cos SZA*, *cloud ratio*, *cos hr*, *T2M*, and *RH2M*), consistently achieved the best performance across all evaluation metrics: an RMSE of 10.70 W/m², an MAE of 5.56 W/m², and an sMAPE of 4.03%, with a combined training and inference computation time of less than 0.02 seconds. Compared to the persistence baseline, which yielded an RMSE of 93.43 W/m² and an sMAPE of 40.45%, the ELM's achievement represents an 88.5% reduction in absolute error and a dramatic improvement in percentage error stability. The Skill Score calculated relative to the persistence model reached 0.885 for the ELM, confirming that this model does not merely memorize historical patterns but genuinely maps the multivariate non-linear relationships between the selected meteorological variables and solar irradiance, demonstrating strong generalization under real operational conditions.

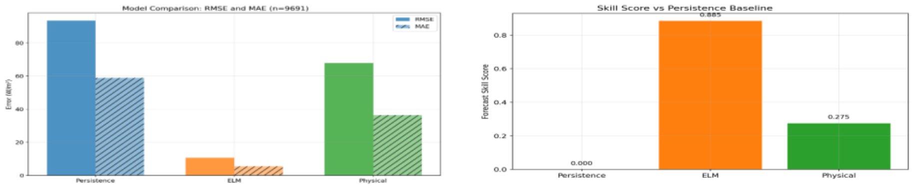

Figure 2. Performance of the ELM

Performance of the ELM can be attributed to its ability to capture the complex interactions among the six input variables, which were rigorously validated through correlation analysis, Variance Inflation Factor, and permutation importance. Specifically, the *cos SZA* and *cloud ratio* feature physically represent solar geometry and cloud transmittance, while *lag 1* and *cos_hr* provide the context of short-term atmospheric memory and the diurnal cycle. The physical capping mechanism, applied as a post-processing step to constrain predictions within the clear-sky irradiance envelope (*CLRSKY_SFC_SW_DWN*), plays a pivotal role in stabilizing the sMAPE metric. Prior to applying the capping mechanism, the ELM's sMAPE value exceeded 100% due to the mathematical distortion inherent in the sMAPE formula during near-zero irradiance periods (such as nighttime and dawn), where even minor absolute errors lead to an artificially inflated percentage error. By forcing predictions to approach zero when $GHI_{cs} \approx 0$, the capping mechanism eliminates this bias and reduces the sMAPE to a much more representative 4.03%, reflecting actual operational performance. These findings align with recent literature emphasizing the importance of integrating physical constraints into the machine learning pipeline to ensure prediction consistency while preserving statistical accuracy (Zhang et al., 2015).

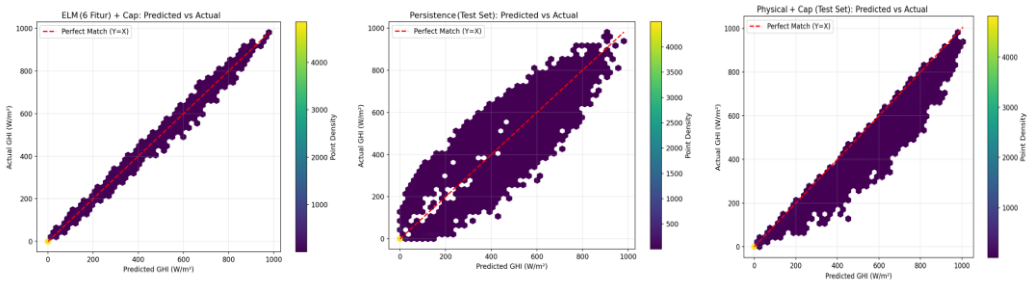

Figure 3. Comparison of predicted and actual GHI on the test dataset using the ELM6 + Cap, Persistence, and Physical + Cap models

The radiative physics model with empirical temperature correction, while achieving a relatively low sMAPE (9.58%) thanks to the same capping mechanism, demonstrates limitations in capturing the non-linear cloud dynamics that dominate tropical climates. With an RMSE of 67.76 W/m^2 and a Skill Score of 0.275, this model successfully tracks the clear-sky irradiance envelope under clear conditions but experiences significant deviations during rapid changes in cloud cover, a phenomenon that cannot be adequately modeled solely through the deterministic parameterization of attenuation coefficients (k_0=0.8, p=3.0, k_1=-0.01). These results underscore the classic trade-off between physical interpretability and statistical flexibility: the physics-based model offers mechanistic transparency and instantaneous inference speed (<0.01 seconds), yet its accuracy is constrained by simplifying atmospheric assumptions; conversely, the ELM sacrifices direct interpretability in favor of universal approximation capabilities, which have proven to be more adaptive to tropical variability. In the operational context of floating PV systems integrated with hydroelectric power, the inference speeds of both models (<0.02 seconds) are equally sufficient for real-time control cycles, making accuracy the primary determining factor in model selection.

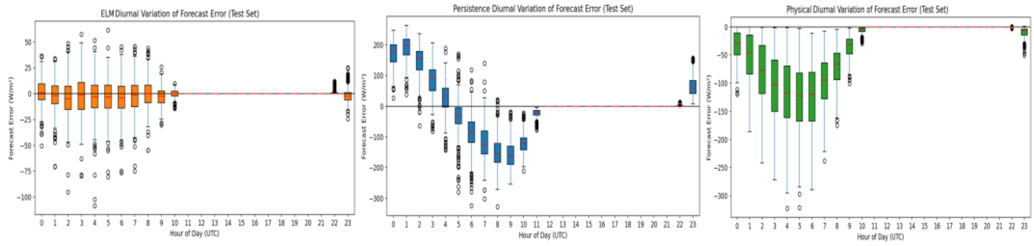

Figure 4. The RMSE, MAE, and ELM Bar Charts

Comparative visualizations further reinforce these quantitative findings. The RMSE and MAE bar charts reveal a striking difference in error magnitude between the ELM and the other two models; the significantly shorter bars for the ELM visually confirm the superior accuracy of the data-driven approach. The Skill Score chart confirms that only the ELM surpasses the threshold for meaningful improvement (>0.1) relative to the baseline, achieving a value of 0.885 that indicates substantial pattern learning. Time-series overlays reveal the ELM's ability to closely track rapid fluctuations in actual irradiance during cloud transitions, whereas the physics-based model tends to lag in responding to sudden changes and adheres more rigidly to the clear-sky envelope. Triple hexbin scatter plots visualize the density of predicted versus actual values, where the ELM exhibits the tightest clustering around the perfect diagonal line (Y=X), indicating consistent accuracy across the entire range of irradiance values. The diagonal elongation pattern observed in the persistence model, along with the wider dispersion in the physics-based model, provides visual confirmation of the limitations of each approach previously outlined numerically.

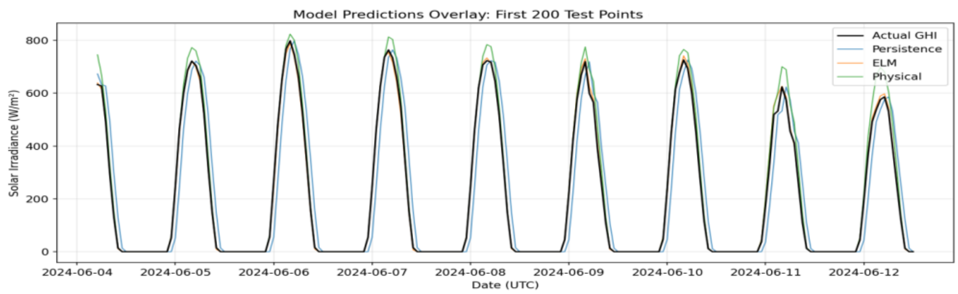

Figure 5. Actual and predicted GHI time series for the Persistence, ELM, and Physical models.

In the context of tropical irradiance forecasting literature, the findings of this study reinforce previous results indicating that data-driven approaches like ELM generally outperform purely physics-based models in statistical accuracy when hourly meteorological variables are available as inputs. However, the original contribution of this study lies in demonstrating that: (1) a hybrid feature selection approach based on physics, statistics, and performance can yield a more parsimonious model (6 features vs. 11 initial features) without sacrificing accuracy; (2) the integration of a simple physical capping mechanism, without requiring a complex hybrid architecture, is sufficient to stabilize evaluation metrics and ensure the physical consistency of predictions; and (3) a strict chronological validation protocol with consistent feature alignment is a methodological prerequisite for a fair comparison among forecasting paradigms. This strategy offers a more computationally lightweight implementation pathway compared to residual hybrid approaches that require staged training, making it more suitable for deployment on resource-constrained embedded systems.

Although the performance of the capped ELM is highly promising, this study acknowledges several limitations that open avenues for future development. First, the spatial resolution of the NASA POWER data (~0.5°) may not fully capture micro-local variability around the reservoir surface; thus, cross-validation with ground-truth data from local measurement stations is recommended to quantify scale bias (Jed et al., 2022). Second, while the one-hour forecasting horizon is relevant for real-time control, it does not yet cover the requirements for daily or weekly scheduling, which necessitate multi-step forecasting architectures or sequence-to-sequence approaches. Third, the applied capping mechanism is deterministic and static; developing adaptive constraints that account for prediction uncertainty (e.g., through confidence intervals or ensembles) could enhance robustness under extreme conditions. Future research is also suggested to explore residual hybridization, where the ELM is trained to model the residual between actual observations and physics-based predictions, thereby combining the interpretability of deterministic models with the statistical flexibility of machine learning within a unified framework.

Tabel 1. Comparison Methods

Model	RMSE (W/m²)	MAE (W/m²)	sMAPE (%)	Skill Score	Time (s)
Persistence	**93.43**	**58.91**	**40.45**	**0**	**0.004**
Physical Revised	**67.76**	**36.36**	**9.58**	**0.275**	**0.004**
ELM Revised	**10.7**	**5.56**	**4.03**	**0.885**	**0.015**

This study provides empirical evidence that the Extreme Learning Machine (ELM) (Abed et al., 2026) (Scott et al., 2023), utilizing six selected features and a physical capping mechanism, represents an optimal approach for hourly solar irradiance forecasting. This model offers a unique balance between high accuracy (RMSE: 10.70 W/m², Skill Score: 0.885), exceptional computational speed (<0.02 seconds), and ensured physical consistency. These findings not only contribute to the academic literature on tropical renewable energy forecasting but also provide practical guidelines that can be readily adopted by power system operators in designing stable, efficient, and sustainable strategies for integrating floating PV systems. Grounded in a rigorous methodological foundation, featuring consistent feature alignment, a strict 60:20:20 chronological split, and evaluation on an identical test set of 9,691 data points, this research lays the groundwork for developing forecasting architectures that are more adaptive to tropical microclimate dynamics. Ultimately, it strengthens the technical foundation of Indonesia's energy transition toward a cleaner and more resilient power system.

6. Conclusion

This study explored a one-hour resolution Global Horizontal Irradiance (GHI) forecasting framework by exclusively utilizing hourly satellite-based meteorological reanalysis data from NASA POWER. Through a strict 60:20:20 chronological validation protocol that prevents information leakage, consistent feature alignment (trimming the first 24 rows to compensate for lag and rolling operations), and the integration of a physical capping mechanism, this study systematically compared three forecasting paradigms: the persistence baseline, a temperature-corrected radiative physics model, and an Extreme Learning Machine (ELM) with six selected features. Evaluation results on 9,691 identical independent test samples for all three models showed that the ELM with physical capping consistently achieved the highest accuracy, yielding an RMSE of 10.70 W/m², an MAE of 5.56 W/m², and an sMAPE of 4.03%, along with a Skill Score of 0.885, indicating an 88.5% error reduction compared to the persistence model. These findings confirm that a data-driven approach with evidence-based feature selection (physics + statistics + permutation importance) can map non-linear tropical cloud dynamics much more effectively than purely deterministic models. Meanwhile, the application of physical capping proved crucial in stabilizing the sMAPE metric, which was previously distorted by mathematical bias during near-zero irradiance periods, while ensuring that all predictions remain within a realistic atmospheric physical envelope.

Operationally, the ELM's inference speed of under 0.02 seconds holds the potential to serve as a foundational alternative in developing irradiance forecasting frameworks for renewable energy integration in tropical regions. The parsimonious model (using only 6 features) also reduces computational complexity and memory requirements, making it more adaptable for deployment on resource-constrained embedded systems. Nevertheless, this study acknowledges the limitations of using gridded reanalysis data, which may not fully capture micro-local variability around the reservoir surface, as well as the forecasting horizon that is still limited to one hour ahead.

Overall, this study demonstrates that the ELM approach combined with hybrid feature selection and physical constraints represents a promising strategy worthy of further exploration for hybrid power system research in dynamic tropical regions.

Referensi

[1] Abed, J. N., Abdoos, A. A., & Ebadi, A. (2026). A new hybrid intelligent method for short-term photovoltaic power forecasting based on combination of ELM, VMD and PCA. Electric Power Systems Research, 250. https://doi.org/10.1016/j.epsr.2025.112104

[2] Al-Dahidi, S., Ayadi, O., Adeeb, J., Alrbai, M., & Qawasmeh, B. R. (2018). Extreme Learning Machines for Solar Photovoltaic Power Predictions. Energies, 11(10). https://doi.org/10.3390/en11102725

[3] Attya, M., Abo-Seida, O. M., Abdulkader, H. M., & Mohammed, A. M. (2025). Advanced solar radiation prediction using combined satellite imagery and tabular data processing. Scientific Reports, 15(1), 1–22. https://doi.org/10.1038/s41598-025-96109-0

[4] De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. Neurocomputing, 192, 38–48. https://doi.org/10.1016/j.neucom.2015.12.114A. Firma and S. Oktamuliani, "Pengolahan filtering dan contrast enhancement untuk meningkatkan kualitas citra ultrasonografi abdomen," *Jurnal Pengolahan Citra Medis*, 2023. doi: 10.22437/jop.v8i1.20607.

[5] Jed, M., Ihaddadene, N., El Hacen Jed, M., Ihaddadene, R., & El Bah, M. (2022). Validation of the Accuracy of NASA Solar Irradiation Data for Four African Regions. International Journal of Sustainable Development and Planning, 17(1), 29–39. https://doi.org/10.18280/ijsdp.170103

[6] Le Cadre, H., Aravena, I., & Papavasiliou, A. (2015). Solar PV power forecasting using Extreme Learning Machine and information fusion. 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015 - Proceedings, 77–82

[7] Niu, Y., Su, Y., Tang, P., Wang, Q., Sun, Y., & Song, J. (2025). Estimation of Solar Irradiance Under Cloudy Weather Based on Solar Radiation Model and Ground-Based Cloud Image. Energies, 18(3). https://doi.org/10.3390/en18030757

[8] Rajagukguk, R. A., Kamil, R., & Lee, H. J. (2021). A deep learning model to forecast solar irradiance using a sky camera. Applied Sciences (Switzerland), 11(11). https://doi.org/10.3390/app11115049

[9] Scott, C., Ahsan, M., & Albarbar, A. (2023). Machine learning for forecasting a photovoltaic (PV) generation system. Energy, 278. https://doi.org/10.1016/j.energy.2023.127807

[10] Solano, E. S., Dehghanian, P., & Affonso, C. M. (2022). Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection. Energies, 15(19). https://doi.org/10.3390/en15197049

[11] Suwadana, F. A., Garniwa, P. M. P., Putera, D. A., Puspita, D., Gufron, A., Aditya, I. A., Lee, H., & Garniwa, I. (2025). Solar Irradiance Estimation in Tropical Regions Using Recurrent Neural Networks and WRF Models. Energies, 18(4). https://doi.org/10.3390/en18040925

[12] Voyant, C., Despotovic, M., Notton, G., Saint-Drenan, Y. M., Asloune, M., & Garcia-Gutierrez, L. (2025). On the importance of clearsky model in short-term solar radiation forecasting. Solar Energy, 294. https://doi.org/10.1016/j.solener.2025.113490

[13] Zhang, J., Florita, A., Hodge, B. M., Lu, S., Hamann, H. F., Banunarayanan, V., & Brockway, A. M. (2015). A suite of metrics for assessing the performance of solar power forecasting. Solar Energy, 111, 157–175. https://doi.org/10.1016/j.solener.2014.10.016

[14] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1–3), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

[15] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics, 3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

[16] Inman, R. H., Pedro, H. T. C., & Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science, 39*(6), 535–576. https://doi.org/10.1016/j.pecs.2013.06.002