


Robustness of Transformer, Long Short-Term Memory, and Light Gradient Boosting Machine Models for Solar Power Forecasting Under Noisy Input Conditions



Nhat Tung Nguyen¹, Manh Hai Pham^{2*}

¹ Faculty of Electrical and Electronic Engineering, Thuyloi University, Hanoi 100000, Vietnam

² Faculty of New Energy, Electric Power University, Hanoi 100000, Vietnam

Corresponding Author Email: haipm@epu.edu.vn

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.590201>

ABSTRACT

Received: 2 December 2025

Revised: 5 February 2026

Accepted: 16 February 2026

Available online: 28 February 2026

Keywords:

solar power forecasting, robustness, Transformer, Long Short-Term Memory, Light Gradient Boosting Machine

Reliable solar power forecasting is important for stable grid operation, yet measurements from photovoltaic systems are often affected by sensor noise and calibration drift. This study investigates the robustness of three forecasting models—Transformer, Long Short-Term Memory (LSTM), and Light Gradient Boosting Machine (LightGBM)—under controlled input perturbations. To simulate measurement degradation, positive uniform noise is added to solar radiation data with intensities ranging from 0% to 50%. Experiments are conducted under two seasonal regimes representing different variability conditions: a relatively stable dry period (March) and a highly variable monsoon period (June). The results show a clear difference in noise tolerance among the models. The Transformer maintains relatively stable prediction accuracy even at high noise levels, whereas the LSTM exhibits noticeable performance deterioration as noise increases. In contrast, LightGBM shows the largest error growth when the input data are corrupted. Similar patterns are observed in both seasons, suggesting that the differences in robustness are related to model structure rather than seasonal conditions. These findings indicate that attention-based architectures are more suitable for photovoltaic power forecasting in situations where measurement quality is uncertain.

1. INTRODUCTION

The large-scale integration of renewable energy sources into the electrical grid has become a cornerstone of global strategies aimed at mitigating climate change and reducing greenhouse gas emissions. A critical global strategy to mitigate climate change and reduce carbon emissions. Among various renewable technologies, photovoltaic (PV) power generation plays a central role due to its vast resource availability, modular deployment, and rapidly declining installation costs. However, the inherent intermittency and instability of solar irradiance introduce substantial uncertainty into power generation, posing significant challenges for grid operation and energy management systems (EMS) [1]. As a result, accurate solar power forecasting has emerged as an essential tool for operators for planning, dispatching, and maintaining system stability [2].

Despite the advancements in forecasting algorithms, a substantial gap remains between research environments and real-world implementation. Most existing studies evaluate model performance based on high-quality, pre-processed datasets, assuming that the input data from sensors is reliable and clean. In practice, however, data acquisition systems are frequently subjected to various disturbances. Sensors deployed in outdoor environments are susceptible to degradation, dust accumulation, and electromagnetic interference, resulting in measurement noise and outliers [3].

Furthermore, the digitization of energy infrastructure exposes power plants to cybersecurity threats, where malicious actors could intentionally corrupt sensor readings (e.g., False Data Injection Attacks) to manipulate forecasting outcomes and inflict economic damage on the facility [4]. When solar power forecasting models trained on clean data are deployed in such volatile environments, their performance often deteriorates significantly, potentially leading to erroneous grid dispatch decisions.

In recent years, data-driven approaches have become the prevailing paradigm for solar power forecasting. Traditional machine learning (ML) methods, such as Gradient Boosting Machines (e.g., LightGBM), are widely adopted due to their computational efficiency training speed and interpretability on tabular data [5, 6]. In parallel, deep learning (DL) architectures, particularly recurrent neural networks (RNNs) such as Long Short-Term Memory (LSTM), have served as a standard choice for modeling handling dependencies in time-series [7]. More recently, the Transformer-based models, initially developed for natural language processing, have demonstrated strong potential in time-series analysis due to their self-attention mechanism [8-10]. Although the predictive accuracy of these models is well-documented under ideal conditions, their comparative robustness—the ability to maintain performance when input data is corrupted by noise has not been sufficiently investigated.

To address this limitation, we conduct a comprehensive

assessment of forecasting robustness under corrupted inputs. We compare three distinct architectures: LightGBM as a conventional machine-learning baseline, LSTM as a recurrent deep-learning model, and the Transformer as an attention-based deep-learning architecture. Specifically, we investigate how these models respond when uniform noise is introduced into the input features at varying intensities (from 0% to 50%). By systematically analyzing the degradation in prediction accuracy for each architecture, this study aims to identify the most robust modeling approach for scenarios where data integrity is compromised. The findings provide a comprehensive guideline for selecting forecasting models that ensure reliable operation even under adverse data conditions.

In this work, memory is defined strictly as an architectural property—internal mechanisms that store and propagate temporal information across time steps (e.g., recurrent cell states in LSTM or context aggregation through self-attention in Transformer models). In contrast, tree-based ensembles such as LightGBM do not maintain internal temporal states; thus, they operate without architectural temporal memory, irrespective of the particular input feature representation.

This study prioritizes radiation integrity because meteorological sensors are physically exposed and often lack the redundant security protocols found in revenue-grade power meters, making them the primary and most realistic 'attack surface' for False Data Injection Attacks (FDIA) in PV systems.

The remainder of this paper is organized as follows: Section 2 details the methodology, describing the theoretical background of the three forecasting models (LightGBM, LSTM, Transformer) and the noise injection mechanism employed. Section 3 presents the experimental setup, including the dataset and a description of evaluation metrics. Section 4 discusses the comparative results and analyzes the impact of noise intensity on each model. Finally, Section 5 concludes the paper with a summary of findings and suggestions for future research.

2. METHODOLOGY

This section outlines the theoretical framework of the three forecasting models employed in this study. We selected these architectures to represent distinct approaches to time-series regression: LightGBM (tree-based ensemble learning), LSTM (RNNs), and Transformer (attention-based DL).

2.1 Light Gradient Boosting Machine

LightGBM is a highly efficient gradient boosting framework that uses tree-based learning algorithms [11]. Unlike traditional implementations such as XGBoost, which utilize a level-wise tree growth strategy, LightGBM employs a leaf-wise growth strategy with depth constraints. This approach enables the model to converge faster and achieve higher accuracy by splitting the leaf with the maximum delta loss.

In this study, LightGBM is employed as a baseline model representing ML approaches without architectural temporal memory. By design, LightGBM performs a direct mapping from input features to output values and does not maintain any internal state that evolves across time. Each prediction is generated independently, without any mechanism to retain, update, or propagate information from previous time steps.

In contrast, LSTM and Transformer architectures explicitly encode temporal dependencies through recurrent cell states and self-attention mechanisms, respectively. This fundamental architectural difference motivates LightGBM's role in our comparative evaluation. By contrasting a model without intrinsic temporal memory against models equipped with architectural memory, this study aims to isolate and assess the contribution of internal memory mechanisms to robustness under noisy input conditions.

In this study, LightGBM is implemented as a non-sequential, memory-less baseline. Unlike LSTM and Transformer, which utilize a 288-step look-back window to capture temporal dependencies, LightGBM generates predictions based strictly on the meteorological features at the current time step. This design choice is intentional to isolate and assess the contribution of internal memory mechanisms to model robustness. By contrasting models that can 'anchor' their forecasts to clean historical power trends against a model that cannot, we demonstrate how architectural memory serves as a critical buffer against real-time sensor corruption.

2.2 Long Short-Term Memory

LSTM is a specialized variant of RNNs designed to overcome the vanishing gradient problem inherent in standard RNNs during long-term dependency learning [12-14]. LSTMs are particularly well-suited for solar power forecasting as they can model the sequential nature of irradiance data. The core of an LSTM unit is the cell state (C_t), which acts as a conveyor belt carrying information across time steps. The flow of information is regulated by three gates: the forget gate (f_t), the input gate (i_t), and the output gate (o_t). Mathematically, the transition functions at time step t , given input x_t and previous hidden state $h_{(t-1)}$ are defined as follows:

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C[h_{t-1}, x_t] + b_C) \\ C_t &= f_t \odot C_{t-1} \oplus i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \quad (1)$$

where, x_t is the input vector; h_t is the hidden state; C_t is the cell state; \tilde{C}_t is the candidate cell state, W and b are the weight matrices and bias vectors for the respective gates; σ and \tanh are the sigmoid and hyperbolic tangent activation functions; and \odot denotes the element-wise multiplication (Hadamard product).

In this study, we employ an LSTM model that receives the full autoregressive input sequence consisting of radiation, ambient temperature, cell temperature, and past power output. The network contains a single LSTM layer with 64 units, followed by a 0.2 dropout layer and a 32-unit ReLU dense layer. The model is trained using the Adam optimizer (learning rate 0.001) with MSE loss. This architecture enables the LSTM to capture temporal dependencies and leverage past power information for stabilizing predictions under noisy conditions.

2.3 Transformer

While LSTMs process data sequentially, the Transformer [15-17] architecture relies entirely on an attention mechanism to draw global dependencies between input and output.

Originally proposed for natural language processing, Transformers have shown superior performance in time-series forecasting due to their ability to parallelize computation and capture long-range interactions without the degradation seen in recurrent connections.

The fundamental component of the Transformer is the multi-head self-attention mechanism. For a given input sequence, the model computes three vectors: Query (Q), Key (K), and Value (V). The attention score determines the importance of other time steps relative to the current step and is calculated as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where,

- The dot product QK^T computes the similarity score between the query and all keys.
- The scaling factor $\frac{1}{\sqrt{d_k}}$ is applied to counteract the effect of large dot products pushing the softmax function into regions with extremely small gradients.
- The softmax function normalizes these scores to obtain attention weights, which are then used to compute a weighted sum of the values V.

Unlike RNNs, the Transformer does not inherently process data in order. To address this, positional encoding is added to the input embeddings to provide information about the relative or absolute position of the time steps in the sequence. This capability allows the Transformer to focus on relevant historical "signals" while ignoring "noise," a property that is central to the robustness analysis in this study.

In this study, we adopt a Transformer-based forecasting model that also utilizes the full 288-step autoregressive input sequence, including past power. The model employs a 64-dimensional projection layer, learned positional embeddings, a multi-head self-attention layer with four heads, and a feed-forward network with residual connections and layer normalization. Training is performed using the Adam optimizer (learning rate 0.0005) with MSE loss. The self-attention mechanism enables the Transformer to selectively emphasize relevant historical signals while suppressing noise injected into the radiation input.

3. EXPERIMENTAL SETUP

This section presents the experimental setup used to evaluate the robustness of the three forecasting architectures. We describe the dataset characteristics, the noise-injection mechanism, the model-input configurations, and the procedure for generating autoregressive sequences. The experimental design is structured to ensure that all models are trained under comparable conditions while isolating the intrinsic behaviors of each architecture when exposed to noisy inputs.

It is important to note that the LSTM and Transformer architectures inherently incorporate autoregressive memory through their recurrent and self-attention mechanisms, enabling them to internally capture temporal dependencies across the 288-step input window. In contrast, LightGBM has no built-in temporal modeling capability and can only access historical information through manually engineered lag features. To avoid granting LightGBM an external advantage

not originating from its native architecture, and to ensure that the robustness evaluation reflects the intrinsic characteristics of each model, we deliberately exclude lag-based feature engineering for the LightGBM baseline. This experimental design allows for a clean, unbiased comparison of the three forecasting architectures under noisy input conditions.

3.1 Dataset description

The experimental validation is conducted using real-world operational data acquired from a large-scale grid-connected PV power plant located in Dak Lak province, Vietnam. This facility operates with a total installed capacity of 50 MW.

Training dataset: The dataset spans approximately one year, consisting of 105,120 data points sampled at a 5 minutes resolution. This high-frequency sampling captures the rapid fluctuations in solar irradiance characteristic of the tropical climate in the Central Highlands of Vietnam. The data includes four primary meteorological and electrical variables: Radiation, temperature, temperature-cell, generate power. Prior to training, missing values in the raw sensor data were handled using linear interpolation to maintain time-series continuity. Subsequently, all input features were normalized into the [0, 1] range using MinMaxScaler. To prevent data leakage, the scaler was fit exclusively on the Training Set and then applied to the test set. To better characterize the inherent linear relationships between these input features and the generated power (kW) within this dataset, a Pearson correlation analysis was performed. The resulting coefficients, which quantify the influence of each feature on the power output, are presented in Table 1.

Table 1. Pearson correlation coefficient for power generation

Feature	Correlation Coefficient (r)
Radiation	0.987
Cell temperature	0.893
Ambient temperature	0.560

Pearson correlation analysis reveals a strong positive relationship between the environmental factors and the electrical power generation. Notably, radiation exhibits the closest to perfect correlation coefficient (0.987), confirming it as the most critical factor influencing power output. Cell temperature also demonstrates a very strong correlation (0.893), whereas ambient temperature shows only a moderate correlation (0.560).

Testing dataset: To evaluate the model's performance under realistic operational conditions, we selected a continuous 4 days period:

- **Look-back window:** Data from February 28, 2023, is used solely to initialize the autoregressive window (look-back = 288 steps).
- **Forecasting horizon:** Predictions are generated from March 1 to March 3, 2023, and June 4 to June 6, 2023, and these values are compared against the ground truth to calculate error metrics.

3.2 Noise injection protocol (simulated sensor drift and cyber-attacks)

Unlike conventional studies that assume clean inputs, this experiment systematically corrupts the input features of the test set to simulate two critical real-world scenarios: physical

sensor calibration drift and cybersecurity threats such as FDIA. In this study, the disturbance injection is specifically formulated as a biased data integrity attack (BDIA). Unlike zero-mean stochastic noise models, we implement a Positive-Bias Perturbation to mimic an over-reporting scenario where reported irradiance is consistently manipulated above the ground truth to mislead the EMS.

To model these conditions, we employ a uniform positive noise injection (UPNI) strategy. This approach tests the models' resilience against additive distortions where input data is consistently manipulated away from the ground truth.

Let $X = [x_1, x_2, \dots, x_N]$ denote the input feature vector for a specific variable (e.g., radiation) over N time steps. The corrupted feature vector \tilde{X} is generated by injecting noise into each element individually. For each time step i , the corrupted value \tilde{x}_i is calculated as:

$$\begin{aligned} \tilde{x}_i &= x_i + \delta_i \\ \delta_i &\sim U(0, \epsilon \max(X)) \end{aligned} \quad (3)$$

where,

- x_i is the original clean value at time step i .
- $\max(X)$ is the global maximum value of the feature vector X .
- ϵ is the noise intensity level, varying in the set $\{0\%, 5\%, 10\%, 20\%, 30\%, 50\%\}$.
- $U(a, b)$ denotes the continuous uniform distribution yielding positive values, simulating an additive drift.

This worst-case scenario serves to evaluate the intrinsic filtering capability of the Transformer's attention mechanism against structured, non-Gaussian malicious data, providing a more rigorous stress test than simple random noise.

The corruption is intentionally restricted to the radiation feature to isolate the models' intrinsic filtering capabilities from the confounding effects of recursive error propagation. This ensures a clean comparative analysis of how the Transformer's attention mechanism distinguishes between true physical drivers and structured malicious data, whereas corrupting the autoregressive power feedback would introduce a bias towards loop instability rather than feature-level robustness.

3.3 Experimental implementation framework

The overall experimental procedure employed in this study is visualized in Figure 1. The framework is designed to rigorously assess the sensitivity of forecasting models when transitioning from a controlled training environment to a volatile deployment environment. The process consists of three distinct phases:

Phase 1: Data preparation and model training

First, the raw solar power dataset is preprocessed and split chronologically to prevent future data leakage. The Training Set (comprising 12 months of historical data) is kept strictly clean (i.e. $\epsilon = 0\%$). This simulates the standard scenario where models are developed using high-quality historical records verified by engineers. All three models LSTM, Transformer, and LightGBM are trained on this clean dataset to learn the underlying physical correlations between irradiance, temperature, and power output.

Phase 2: Noise injection and simulation

To mimic real-world sensor malfunctions during the operation phase, the test set is subjected to the noise injection mechanism described in Section 3.2. We systematically vary

the noise intensity ϵ across six levels: $\{0\%, 5\%, 10\%, 20\%, 30\%, 50\%\}$. For each level, a specific Noisy Input Vector (\tilde{X}_{test}) is generated.

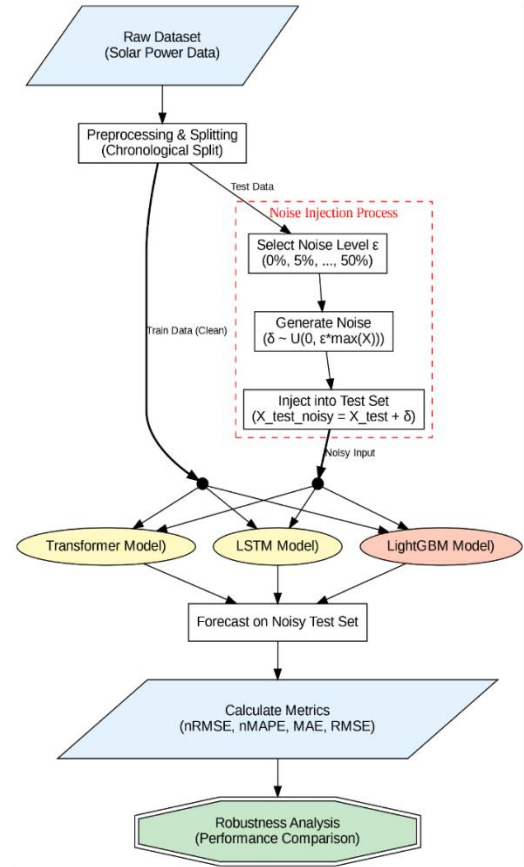


Figure 1. The proposed experimental framework

- For the purpose of evaluating model robustness, noise injection was applied exclusively to the input feature with the strongest influence on the output: radiation. This decision is justified by the fact that the radiation feature exhibits the highest correlation coefficient with the target variable, generated power. Applying perturbations solely to this key input ensures that the models' ability to maintain accuracy is rigorously tested against noise in the most influential component of the dataset. Crucially, throughout the testing phase, the ground truth target values (y_{test}) remained unchanged to serve as a reliable, unbiased benchmark for the calculation of all error metrics.
- The noise magnitude is scaled using the maximum radiation value computed from the training set only, ensuring no information leakage from the test data. While this scaling may amplify relative perturbations during low-irradiance periods, it intentionally models worst-case sensor corruption scenarios. To mitigate potential bias, additional noise models are also considered, as described below.

Phase 3: Comparative evaluation the trained models perform forecasting tasks on the noisy test sets

In this final phase, the trained models perform forecasting tasks on the noisy test sets. To rigorously evaluate model robustness, these forecasted outputs are directly compared against the uncorrupted target values (y_{test}), which remain unchanged to serve as a reliable benchmark. Finally, to quantify the degradation in forecasting accuracy across

varying noise intensities, four standard evaluation metrics, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Normalized Root Mean Square Error (nRMSE), and Normalized Mean Absolute Percentage Error (nMAPE), are calculated and analyzed.

3.4 Error metrics applied in the model

Mean Absolute Error (MAE): MAE measures the average magnitude of the absolute differences between forecasted and actual values [18], offering an intuitive and scale-dependent metric for evaluating prediction accuracy. The formula for calculating MAE is given as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i^f - P_i^r| \quad (4)$$

where, P_i^f is the forecasted power value (kW), P_i^r is the real power value (kW), n is the number of forecast points.

Normalized Mean Absolute Percentage Error (nMAPE): To address the instability of MAPE when actual values are close to zero, the NMAPE [19] is used. This metric provides a more reliable percentage error measure by normalizing the sum of absolute errors against the sum of plant power capacity.

$$nMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|P_i^f - P_i^r|}{P_{capacity}} \quad (5)$$

where, P_i^f is the forecasted power value (kW), P_i^r is the real power value (kW), $P_{capacity}$ is the rated power of the plant (kW), n is the number of forecast points.

RMSE: The RMSE [20] is a standard metric for quantifying the average magnitude of the error. RMSE is calculated by using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i^f - P_i^r)^2}{n}} \quad (6)$$

NRMSE: To express the RMSE as a percentage and facilitate easier comparison of model performance across different datasets, the NRMSE is calculated. It normalizes the RMSE against a reference value, which is the rated capacity of the plant [15].

$$nRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(P_i^f - P_i^r)^2}{P_{capacity}}} \quad (7)$$

4. RESULTS AND DISCUSSION

This section presents the comparative analysis of the three forecasting models. We evaluate their performance first under ideal conditions and subsequently under varying intensities of sensor noise. To rigorously quantify the degradation in forecasting accuracy, we evaluated the models across four standard metrics: RMSE, MAE, nRMSE, and nMAPE. Table 2 presents the detailed performance breakdown for each architecture under varying noise intensities ranging from 0%

to 50%.

Under ideal conditions ($\epsilon = 0\%$), both DL models consistently outperform the LightGBM baseline across the two evaluation periods. Notably, the June dataset exhibits substantially higher baseline errors (nRMSE \approx 5.6–6.0%) compared to March (nRMSE \approx 2.4–2.5%), reflecting a more challenging operating regime characterized by higher irradiance levels and increased power variability. Despite this increased difficulty, the relative robustness trends remain remarkably consistent across both periods. As noise intensity increases, the Transformer demonstrates the highest resilience, with only a modest increase in nRMSE from 2.45% to 3.01% in March and from 5.61% to 6.32% in June at $\epsilon = 50\%$. This stability indicates a strong ability to attenuate corrupted radiation signals by leveraging contextual and temporal information. In contrast, the LightGBM model exhibits a pronounced and nonlinear degradation in both periods. In March, its nRMSE escalates nearly fourfold from 3.90% to 15.34%, while in June it deteriorates even more severely, rising from 7.08% to 19.35%. This amplified sensitivity under high-power operating conditions highlights the vulnerability of memory-less models to input perturbations, particularly when radiation dominates the power generation process. The LSTM model consistently occupies an intermediate position. While it demonstrates robustness comparable to the Transformer at low noise levels, its performance degrades more noticeably beyond $\epsilon \geq 20\%$, especially during the high-variability June period. Similar trends are observed in nMAPE, where the Transformer maintains the lowest error growth across both datasets, whereas LightGBM experiences a dramatic escalation (from 1.80% to 13.29% in March and from 3.14% to 16.75% in June). Overall, the consistency of robustness rankings across distinct seasonal and operational regimes confirms that the observed performance differences are not scenario-specific but stem from fundamental architectural characteristics, particularly the presence or absence of internal temporal memory mechanisms.

While the quantitative results in Table 2 establish consistent robustness across both dry (March) and monsoon (June) regimes, the following visualizations utilize the March 01–03, 2023 horizon as a representative case study. This specific window is selected for qualitative analysis because it provides a clearer baseline to observe the models' instantaneous responses to input perturbations, without the confounding effects of the higher weather volatility observed in the June dataset. This representative focus ensures visual clarity, while comprehensive seasonal robustness is validated by the aggregated metrics.

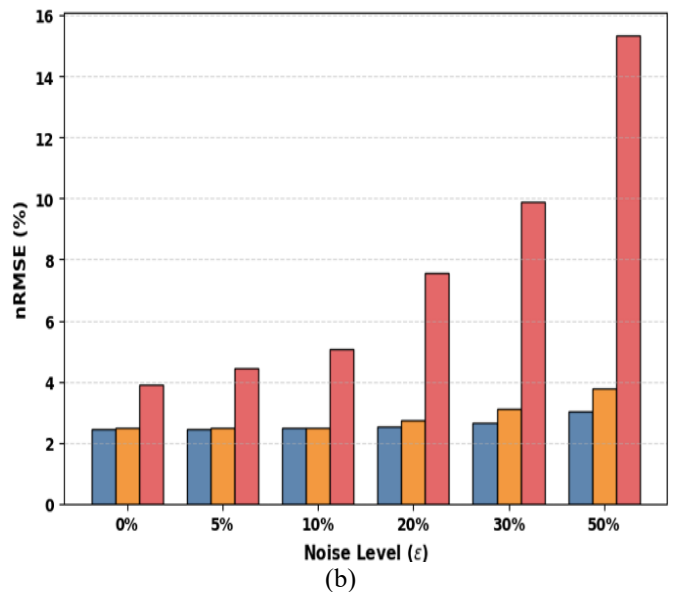
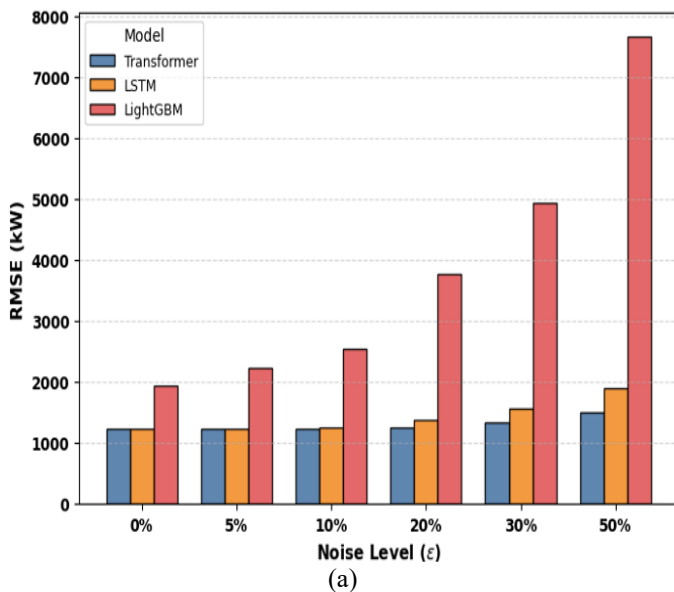
For a detailed comparison of the Transformer, LSTM, and LightGBM architectures, Figure 2 illustrates the error trends across varying noise intensities for the March 01–03, 2023 evaluation period. The Transformer model demonstrates the highest stability across increasing noise levels, while LightGBM exhibits a pronounced escalation in error magnitude. The bar charts in Figure 2 illustrate the comparative robustness of the Transformer, LSTM, and LightGBM models under varying noise levels (0–50%). While prediction errors increase for all models as noise intensity rises, the Transformer consistently exhibits the smallest performance degradation across all metrics. The LSTM model remains competitive at lower noise levels ($\epsilon < 20\%$) but shows more pronounced degradation at higher intensities. In contrast, LightGBM displays a sharp increase in error under severe noise conditions, indicating high sensitivity to corrupted

inputs.

Table 2. Quantitative performance comparison of forecasting models under varying noise intensities

March 01 to March 03, 2023					
Noise (ϵ) (%)	Model	RMSE (kW)	MAE (kW)	nRMSE (%)	nMAPE (%)
0	LSTM	1236.74	552.37	2.47	1.1
	Transformer	1225.61	595.87	2.45	1.2
	LightGBM	1949.75	897.41	3.9	1.8
5	LSTM	1235.51	552.6	2.47	1.11
	Transformer	1229.75	616.22	2.46	1.23
	LightGBM	2226.93	1431.78	4.45	2.864
10	LSTM	1246.34	613.82	2.49	1.23
	Transformer	1234.48	644.71	2.47	1.29
	LightGBM	2535.06	1965.64	5.07	3.93
20	LSTM	1374.08	866.85	2.75	1.73
	Transformer	1257.76	715.52	2.52	1.43
	LightGBM	3772.26	3161.14	7.54	6.32
30	LSTM	1563.53	1122.91	3.13	2.25
	Transformer	1335.9	830.72	2.67	1.66
	LightGBM	4945.3	4257.22	9.89	8.51
50	LSTM	1889.01	1440.91	3.77	2.88
	Transformer	1503.21	1054.26	3.01	2.11
	LightGBM	7667.64	6645.51	15.34	13.29

June 04 to June 06, 2023					
Noise (ϵ) (%)	Model	RMSE (kW)	MAE (kW)	nRMSE (%)	nMAPE (%)
0	LSTM	2977.25	1246.95	5.95	2.49
	Transformer	2804.88	1434.5	5.61	2.87
	LightGBM	3537.7	1571.16	7.07	3.14
5	LSTM	2975.71	1244.35	5.95	2.49
	Transformer	2816.87	1509.16	5.63	3.01
	LightGBM	4031.95	2297.86	8.06	4.59
10	LSTM	2983.28	1298.81	5.96	2.59
	Transformer	2844.75	1586.73	5.69	3.17
	LightGBM	4307.21	2925.29	8.61	5.85
20	LSTM	3021.28	1486.92	6.04	2.97
	Transformer	2923.81	1778.81	5.85	3.55
	LightGBM	5242.62	4189.57	10.49	8.38
30	LSTM	3105.61	1718.92	6.211	3.43
	Transformer	2990.21	1945.74	5.98	3.89
	LightGBM	6362.03	5409.19	12.75	10.82
50	LSTM	3272.84	2022.4	6.545	4.04
	Transformer	3157.91	2206.15	6.31	4.41
	LightGBM	9674.51	8374.24	19.35	16.75



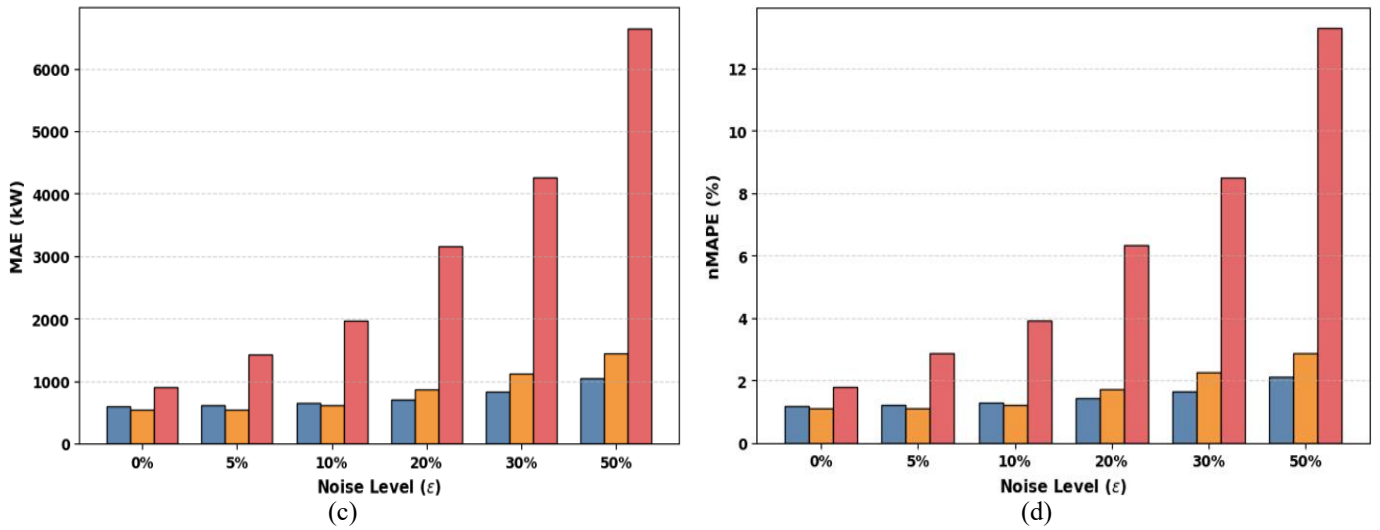


Figure 2. Impact of input noise intensity on prediction error metrics during the March 01–03, 2023 period. The grouped bar charts show the degradation trends for (a) RMSE, (b) MAE, (c) nRMSE, and (d) nMAPE

To establish the context of input data complexity, Figure 3 presents a comparison between the real (ground-truth) and 50% noisy radiation input over time steps 350 to 500. This visualization illustrates the magnitude and structure of the perturbations that the forecasting models must contend with.

Observation of Figure 3 shows that the noisy radiation data (pink line) oscillates significantly and chaotically around the underlying real trend (black line). This significant gap between

the two lines represents the primary challenge for accurate forecasting across the models.

To provide a granular analysis of model robustness, Figure 4 presents a zoomed-in comparison of the forecasting results under the highest noise condition ($\epsilon = 50\%$) over the interval from step 350 to 500. This focused window is selected to clearly illustrate model behavior under severe perturbations without the visual clutter of the full time series.

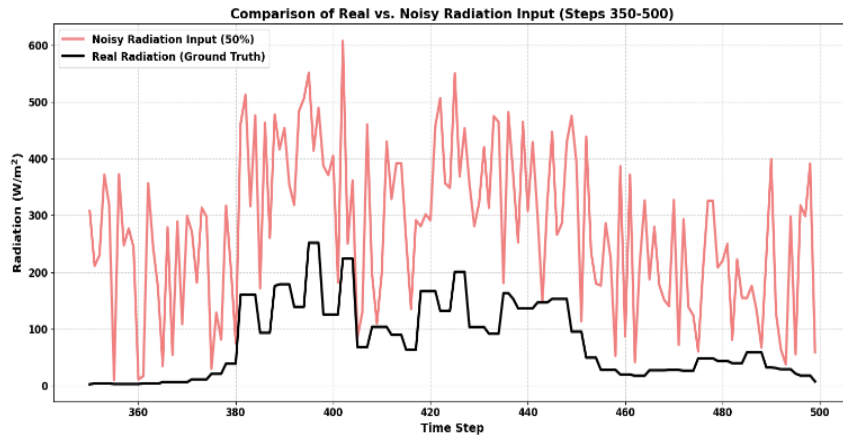


Figure 3. Comparison of real and noisy radiation (50%) input over the time interval from step 350 to 500

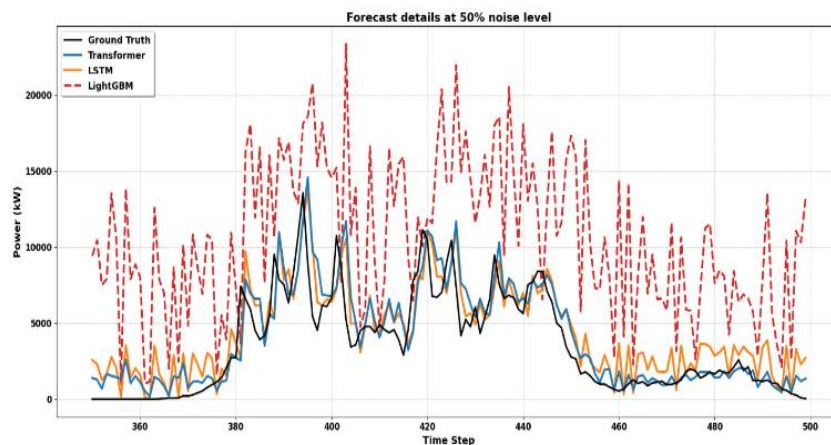


Figure 4. Detailed comparison of forecasted versus actual power at 50% noise level
Note: The plot focuses on the time interval from step 350 to 500 to illustrate model behavior under high uncertainty.

As shown in Figure 4, within this interval, the Transformer maintains predictions that closely track the ground-truth power output. In contrast, LightGBM exhibits pronounced oscillations and frequent over- and under-estimation, reflecting its limited ability to filter corrupted input signals. These visual patterns are consistent with the quantitative robustness trends reported in Table 2.

5. CONCLUSION

This study presented a systematic evaluation of architectural robustness in solar power forecasting by comparing LightGBM, LSTM, and Transformer models under simulated perturbations to data integrity. By extending the analysis across two contrasting seasonal regimes the dry season (March) and the high-variability monsoon season (June) the study provides insights relevant to the deployment of resilient forecasting models in practical EMS.

First, the results clearly demonstrate that predictive accuracy under ideal conditions ($\epsilon = 0\%$) does not necessarily translate to operational robustness. While all evaluated architectures exhibit competitive performance on clean data, distinct robustness hierarchies emerge as input noise increases. Notably, although the June dataset exhibits substantially higher baseline volatility, the relative robustness rankings remain consistent with those observed in March, indicating that the observed performance differences are primarily architectural rather than scenario-specific.

Second, the Transformer architecture consistently exhibits the highest resilience to input perturbations. Across both seasonal regimes, it maintains comparatively stable error levels ($nRMSE < 3.1\%$ in March and $< 6.4\%$ in June) even under severe radiation corruption. This robustness suggests that attention-based architectures are particularly effective at leveraging contextual information to mitigate the impact of corrupted input signals.

In contrast, the LightGBM baseline memoryless configuration shows pronounced performance degradation as noise increases, especially during the high-power June period. This behavior highlights the limitations of models without intrinsic temporal memory when exposed to systematic input corruption.

Overall, these findings suggest that forecasting architectures equipped with internal temporal memory particularly attention-based models are better suited for deployment in grid-connected PV systems where sensor degradation or data integrity issues may arise. Future work will extend this analysis to multi-feature corruption scenarios and explore hybrid modeling strategies that combine the computational efficiency of gradient boosting with the robustness benefits of attention mechanisms.

REFERENCE

- [1] AlSkaif, T., Dev, S., Visser, L., Hossari, M., van Sark, W. (2020). A systematic analysis of meteorological variables for PV output power estimation. *Renewable Energy*, 153: 12-22. <https://doi.org/10.1016/j.renene.2020.01.150>
- [2] Aslam, M., Lee, S.J., Khang, S.H., Hong, S. (2021). Two-stage attention over LSTM with Bayesian optimization for day-ahead solar power forecasting. *IEEE Access*, 9: 107387-107398. <https://doi.org/10.1109/ACCESS.2021.3100105>
- [3] Hayward, I., Martin, N.A., Ferracci, V., Kazemimanes, M., Kumar, P. (2024). Low-cost air quality sensors: Biases, corrections and challenges in their comparability. *Atmosphere*, 15(12): 1523. <https://doi.org/10.3390/atmos15121523>
- [4] Shees, A., Tariq, M., Sarwat, A.I. (2024). Cybersecurity in smart grids: Detecting false data injection attacks utilizing supervised machine learning techniques. *Energies*, 17(23): 5870. <https://doi.org/10.3390/en17235870>
- [5] Gaboitaolelwe, J., Zungeru, A.M., Yahya, A., Lebekwe, C.K., Vinod, D.N., Salau, A.O. (2023). Machine learning based solar photovoltaic power forecasting: A review and comparison. *IEEE Access*, 11: 40820-40845. <https://doi.org/10.1109/ACCESS.2023.3270041>
- [6] Qi, C., Liu, Y., Wang, Y., Dong, M. (2023). A classification method for electricity users based on the lightGBM algorithm. In 2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2), Hangzhou, China, pp. 4815-4819. <https://doi.org/10.1109/EI259745.2023.10512536>
- [7] Elsaraiti, M., Merabet, A. (2022). Solar power forecasting using deep learning techniques. *IEEE Access*, 10: 31692-31698. <https://doi.org/10.1109/ACCESS.2022.3160484>
- [8] Husein, M., Gago, E.J., Hasan, B., Pegalajar, M.C. (2024). Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies. *Heliyon*, 10(13): e33419. <https://doi.org/10.1016/j.heliyon.2024.e33419>
- [9] Kharkwal, V.S., Gupta, S.C. (2025). Recent Advancement in transformer based electric load forecasting: A review. In 2025 IEEE 1st International Conference on Smart and Sustainable Developments in Electrical Engineering (SSDEE), Dhanbad, India, pp. 1-6. <https://doi.org/10.1109/SSDEE64538.2025.10968772>
- [10] CFeng, C., Hao, J., Yang, R., Han, J. (2025). Deep Quantile transformer for probabilistic net load forecasting with extreme solar penetration. In 2025 IEEE Power & Energy Society General Meeting (PESGM), Austin, TX, USA, pp. 1-5. <https://doi.org/10.1109/PESGM52009.2025.11225337>
- [11] Ke, G., Meng, Q., Finley, T., Wang, T., et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, pp. 1-9.
- [12] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8): 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [13] Nguyen, M.N., Pham, M.H., Wu, Y.K., Nguyen, N.T. (2021). An optimizing method based on combining edge computer and long short-term memory networks applied to solar power forecasting. In 2021 IEEE International Future Energy Electronics Conference (IFEEEC), Taipei, Taiwan, pp. 1-6. <https://doi.org/10.1109/IFEEEC53238.2021.9661669>
- [14] Boucetta, L.N., Amrane, Y., Arezki, S. (2023). Comparative analysis of LSTM, GRU, and MLP neural networks for short-term solar power forecasting. In 2023 International Conference on Electrical Engineering and Advanced Technology (ICEEAT), Batna, Algeria, pp. 1-

6.
<https://doi.org/10.1109/ICEEAT60471.2023.10426346>

[15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., et al. (2017) Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

[16] Liu, Y.J., Yang, M. (2025). Ultra-short-term power forecasting for photovoltaic cluster based on dynamic clustering of photovoltaic stations and optimized Ns-transformer combined model. *Energy Reports*, 14: 5656-5684. <https://doi.org/10.1016/j.egy.2025.12.001>

[17] Piantadosi, G., Dutto, S., Galli, A., De Vito, S., Sansone, C., Di Francia, G. (2024). Photovoltaic power forecasting: A transformer based framework. *Energy and AI*, 18: 100444. <https://doi.org/10.1016/j.egyai.2024.100444>

[18] Haque, A.U., Nehrir, M.H., Mandal, P. (2013). Solar PV power generation forecast using a hybrid intelligent approach. In *2013 IEEE Power & Energy Society General Meeting, Vancouver, BC, Canada*, pp. 1-5. <https://doi.org/10.1109/PESMG.2013.6672634>

[19] Phan, Q.T., Wu, Y.K., Phan, Q.D., Lo, H.Y. (2022). A novel forecasting model for solar power generation by a deep learning framework with data preprocessing and postprocessing. *IEEE Transactions on Industry Applications*, 59(1): 220-231. <https://doi.org/10.1109/TIA.2022.3212999>

[20] Sarp, S., Kuzlu, M., Cali, U., Elma, O., Guler, O. (2021). An interpretable solar photovoltaic power generation forecasting approach using an explainable artificial intelligence tool. In *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA*, pp. 1-5. <https://doi.org/10.1109/ISGT49243.2021.9372263>

NOMENCLATURE

Symbol	Description
PV	Photovoltaic
EMS	Energy Management System
FDIA	False Data Injection Attack
BDIA	Biased Data Integrity Attack
ML	Machine Learning
DL	Deep Learning
LSTM	Long Short-Term Memory
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
nRMSE	Normalized Root Mean Square Error
nMAPE	Normalized Mean Absolute Percentage Error

Greek symbols

P_t^f	the forecasted power value (kW)
P_t^r	the real power value (kW)
x_t	the input vector
h_t	the hidden state
C_t	the cell state
W	the weight matrices
b	bias vectors for the respective gates
σ	the sigmoid and hyperbolic tangent activation functions
\odot	denotes the element wise multiplication (Hadamard product)
$P_{capacity}$	Rated capacity of the PV plant (kW)
n	the number of forecast points
ε	Noise intensity level (%)
\tilde{C}_t	candidate cell state (candidate memory content)