

RESEARCH ARTICLE



Advanced RNN-Based Model Variants for Short-Term Solar Power Forecasting

Ngoc-Trung Nguyen,¹ Thanh-Trung Nguyen,¹ Manh-Hai Pham,¹ Minh-Phap Vu,^{1,2} * Tuan-Anh Nguyen¹

¹ Electric Power University, Ministry of Industry and Trade, Hanoi, Vietnam

² Institute of Science and Technology for Energy and Environment, Vietnam Academy of Science and Technology, Hanoi, Vietnam

ABSTRACT

Many existing solar power forecasting studies still focus on a single, fixed forecasting resolution, with limited investigation into how model performance changes when temporal data are aggregated to coarser time scales and formulated as single-step predictions, or how related Recurrent Neural Network (RNN) architectures compare across multiple short-term forecasting intervals. Therefore, this study is conducted to evaluate the performance of four advanced RNN architectures including GRU, BiGRU, LSTM, and BiLSTM, for short-term solar power forecasting using a single-step forecasting formulation across multiple temporally aggregated time resolutions: 5, 15, 30, and 60 minutes. The results demonstrated that all RNN-based models achieved reasonable accuracy in short-term prediction, with BiGRU consistently outperforming the others at most intervals. Notably, for the 5-minute resolution, BiGRU achieved an R^2 of approximately 0.94 and a NMAPE of just 3.41%, confirming its adaptability and precision. Its performance remained strong for 15- and 30-minute forecasts; although a gradual decline was observed as the temporal resolution became coarser. At the 1-hour mark, BiLSTM outperformed BiGRU slightly, achieving $R^2 = 0.85$ and NMAPE = 7.01%, indicating its better ability to model long-term dependencies in aggregated solar power data. These findings not only extend the practical applications of bidirectional RNNs in renewable energy forecasting but also provide a solid foundation for developing more intelligent and efficient power scheduling systems.

Keywords: Solar power forecasting, GRU, BiGRU, LSTM, BiLSTM, RNN-Based Models

1. INTRODUCTION

1.1. Research context and motivation

The growing global demand for clean, sustainable energy has accelerated the adoption of renewable energy sources, particularly solar power, due to its abundance, minimal environmental impact, and decreasing costs [1], [2]. Beyond utility-scale generation, solar energy is also increasingly harnessed directly at the building level; for instance, a recent parametric simulation study showed that an attached solar greenhouse can reduce the annual heating energy demand of a residential building by up to 20.6% [3], underscoring

the versatility of solar resources across both active and passive applications. According to the International Energy Agency (IEA), photovoltaic (PV) capacity is expected to surpass other electricity sources by 2027, reaching over 2,350 GW [4]. By 2050, solar power could supply nearly half of the world's energy demand [5]. Developing countries, such as Vietnam has actively embraced this transition, with significant solar projects underway. As reported by the International Renewable Energy Agency (IRENA), solar energy systems in Vietnam had a combined installed capacity of 18.854 GW in 2023, representing over one-third of the country's total renewable energy capacity [6]. This figure reflects the maximum

*Corresponding author.

trungnn@epu.edu.vn (N.-T. Nguyen); trungnt@epu.edu.vn (T.-T. Nguyen); haipm@epu.edu.vn (M.-H. Pham); vuminhphap@istee.vast.vn (M.-P. Vu); anhnt88@epu.edu.vn (T.-A. Nguyen)

Received 30 June 2026; Revised 11 August 2026; Accepted 28 September 2026; Published online 30 October 2026

Copyright: © 2026 by the Author(s). Licensee Solarlits Limited (Hong Kong). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Citation: Ngoc-Trung Nguyen, Thanh-Trung Nguyen, Manh-Hai Pham, Minh-Phap Vu, Tuan-Anh Nguyen, Advanced RNN-Based Model Variants for Short-Term Solar Power Forecasting, Journal of Daylighting, 13:2 (2026) 698–721. doi: [10.15627/jd.2026.38](https://doi.org/10.15627/jd.2026.38)

NOMENCLATURE

Abbreviations

AI	Artificial intelligence
BiGRU	Bidirectional gated recurrent unit
BiLSTM	Bidirectional long short-term memory
DL	Deep learning
GHI	Global horizontal irradiance
GRU	Gated recurrent unit
LSTM	Long short-term memory
MAE	Mean absolute error (kW)
MSE	Mean squared error (kW ²)
MAPE	Normalized mean absolute percentage error (%)
PV	Photovoltaic
RMSE	Root mean squared error (kW)
RNN	Recurrent neural network

Symbols

<i>i</i>	Index of the data samples
<i>n</i>	Total number of data samples
<i>P_{test}</i>	Measured PV power output in the test set, used to define the evaluation scenarios (kW)
<i>r</i>	Pearson correlation coefficient between an input feature and the PV power output
<i>R²</i>	Coefficient of determination
<i>y₁₁</i>	Actual (measured) PV power (kW)
<i>y₁₂</i>	Forecasted PV power (kW)
<i>y₁₂[¯]</i>	Mean of the actual (measured) values used in the <i>R²</i> denominator (kW)
<i>y_{rated}</i>	Installed capacity of the PV system (kW)

Input variables

<i>radiation</i>	Solar radiation (W/m ²)
<i>temperature</i>	Ambient temperature (°C)
<i>temperature_cell</i>	PV module (cell) temperature (°C)
<i>sin_hour</i> , <i>cos_hour</i>	Cyclic encodings of the hour of the day
<i>day_night</i>	Indicator separating generation and non-generation periods

output these systems are built to deliver. Under the national Power Development Plan VIII, solar capacity is expected to reach 20 GW by 2030. As part of this strategy, Vietnam plans to expand solar energy to account for 25% of its total renewable energy mix within the next eight years, aiming to reduce dependence on imported coal and foster the growth of its domestic solar industry [7].

However, integrating solar energy into power grids introduces challenges due to the variable and weather-dependent nature of solar irradiance [8], [9], [10]. The erratic fluctuation of PV output can destabilize power systems, leading to voltage issues, congestion, curtailment, or even infrastructure failure [11], [12]. Consequently, accurate short-term PV power forecasting plays a crucial role in grid stability, economic dispatching, and energy market operations [13], [14]. In particular, forecasting performance is highly sensitive to the temporal resolution at which the data is

modeled, as different aggregation levels can significantly alter the statistical properties and predictability of solar power time series.

A wide range of forecasting methodologies has been proposed in the literature, which can be broadly classified into four categories: physical models, persistence models, classical statistical models, and artificial intelligence (AI)-based methods [2], [10], [15]. Physical models rely on meteorological and system configuration data to simulate PV performance [10], [16], [17], [18], while persistence models use past observations to predict future values [19], [20]. Despite their simplicity, these models often fail under extreme or fast-changing weather conditions. Moreover, these approaches generally assume a fixed temporal resolution and do not explicitly account for how forecasting difficulty varies across different time scales.

AI-based methods, particularly deep learning (DL), have gained prominence due to their ability to capture non-linear, complex relationships between input variables and PV output. Recent studies have demonstrated the advantages of DL architectures, especially Recurrent Neural Networks (RNNs), in forecasting tasks where temporal dependencies are critical [21], [22], [23], [24]. Among RNN-based architectures, advanced variants such as Gated Recurrent Unit (GRU), Bidirectional GRU (BiGRU), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM) have shown significant potential for modeling time-series data with higher accuracy and robustness [25]. Nevertheless, it remains unclear how differences in their internal memory structures and directional learning mechanisms influence forecasting performance under varying temporal resolutions.

1.2. Research trends and gaps

When applied to photovoltaic (PV) power forecasting, these RNN-based models have demonstrated superior performance in capturing nonlinear relationships between weather conditions and power output. A number of studies have explored and compared these advanced architectures for solar energy prediction, showing that they provide increased accuracy, efficiency, and robustness over conventional approaches. For instance, hybrids of LSTM, BiLSTM, and GRU were employed for multi-step-ahead solar prediction in [26], and other studies confirmed the effectiveness of these models under different environmental conditions, including cloud coverage effects [27], or by leveraging hybrid configurations such as CNN-LSTM or CNN-GRU [28], [29]. In short, RNN-based models, especially their bidirectional and hybrid extensions, have shown remarkable potential for short-term and long-term PV power forecasting [30], [31]. However, most of these studies emphasize performance improvement at a single temporal resolution rather than investigating the underlying mechanisms that govern model behavior across multiple time scales.

Table 1 summarizes representative studies reviewed in this section and compares them with this study in terms of data, temporal resolution, models, evaluation metrics and error analysis. As the table shows, most of these studies use a single temporal

Table 1. Comparison of representative solar power forecasting studies with this study.

Study	Data	Temporal resolution	Models compared	Metrics	Multi-resolution	Analysis by power level
Zameer et al. [26]	Weather-station data, Amherst (USA), 2006–2013	5 min	LSTM, BiLSTM, GRU and their hybrids; linear and SVR baselines	RMSE, MSE, MAE, MAPE	No	No
Kuo et al. [27]	8 kW PV system with a sky imager, Feb–Jun 2021	1 min	ANN, LSTM, GRU	MAE, RMSE, MAPE	No	No
Mellit et al. [28]	4 kWp PV array, University of Trieste microgrid (Italy), Jan–Aug 2020	1 min (5, 30 and 60 min for LSTM only)	LSTM, BiLSTM, GRU, BiGRU, CNN1D, CNN1D–LSTM, CNN1D–GRU	r, RMSE, MAE, MAPE	Partly (LSTM only)	No
Elsaraiti and Merabet [31]	PV data, Halifax (Canada), 2017	30 min (aggregated from 5 min)	LSTM, MLP	MAE, MAPE, RMSE, R ²	No	No
Xu et al. [32]	Competition data and GEFCom2014 (Australia), 300 days each	1 h	GRU–XGBoost fusion; LSTM, ANN, CNN–LSTM, Transformer-based and statistical models	MSE, MAE	No	No
Kim et al. [33]	Two PV plants of 1 MW and 0.5 MW (South Korea), 2018–2021	1 h (24 steps ahead)	PVTransNet (Transformer), LSTM, AT–LSTM	MAE, NMAE, R ²	No	No
Ren et al. [34]	Three PV plants of 20–50 MW (China), 2019	15 min	QR–CNN–TCN, QR–LSTM, QR–CNN, QR–CNN–LSTM	MAE, RMSE, R ² , PICP, WS, CRPS	No	No
Bui et al. [35]	50 MW PV plant, central Vietnam, 2019–2020	5 min	LSTM	MAE, MAPE, MSE, RMSE	No	No
Quang et al. [36]	Phong Dien PV plant, 48 MWp (Vietnam), Dec 2018–Jan 2020	5 min	LSTM	MSE, RMSE, MAE, MAPE	No	No
Nguyen et al. [37]	40 MW PV plant, central Vietnam, 2021–2022	5 min	GRU	MAPE, MAPE*, RMSE	No	Partly (low-power points)
This study	49.5 MW PV plant, Quang Tri (Vietnam), one year	5, 15, 30 and 60 min	GRU, BiGRU, LSTM, BiLSTM under one protocol; Transformer benchmark at 5 min	MAE, MSE, RMSE, NMAPE, R ²	Yes	Yes ($P_{\text{test}} > 0, 100$ and 1,000 kW)

resolution, and apart from [28], none of them evaluates GRU, BiGRU, LSTM and BiLSTM together. Errors are also rarely analyzed by power level. The study in [28] compared the four variants on a 4 kWp system at 1-minute resolution, but its tests at 5, 30 and 60 minutes used LSTM alone. In contrast, this study compares the four variants under one protocol at four temporal resolutions for a 49.5 MW plant, using five metrics and three power-output thresholds.

Globally, there have been numerous case studies employing RNN-based models to forecast solar power generation. These studies often utilize high-quality datasets from experimental PV installations and evaluate multiple deep learning configurations under different forecasting horizons. For instance, researchers in [28] trained LSTM, BiLSTM, GRU, and BiGRU models using power data from the University of Trieste microgrid, while others proposed combinations of GRU and gradient boosting (XGBoost) to improve long-term solar forecasting [32]. A comprehensive study explored the state of PV power forecasting and highlighted a growing trend in using advanced RNN models. It was found that case studies published in 2022 and 2023 commonly utilized LSTM, BiLSTM, GRU, and BiGRU models, underlining the increasing relevance of these architectures in this domain [38]. Despite this progress, the majority of existing works adopt heterogeneous datasets, preprocessing pipelines, and evaluation metrics, which limits the comparability and generalizability of their findings.

More recently, attention-based deep learning architectures have attracted considerable interest in PV power forecasting. Transformer models and their efficient variants, such as Informer [39], dispense with recurrence and rely on self-attention to capture long-range dependencies, and several recent studies have applied them to multi-step and ultra-short-term PV forecasting with encouraging results [33], [40]. In parallel, explainable artificial intelligence techniques have been coupled with solar forecasting models to interpret the contribution of individual meteorological drivers [41], while probabilistic formulations based on quantile regression provide uncertainty intervals in addition to deterministic point forecasts [34]. Nevertheless, Transformer-based models are generally data- and computation-intensive and tend to excel at longer horizons, whereas gated RNN variants remain highly competitive for single-plant, short-horizon applications with moderate data volumes. A systematic understanding of how these established RNN variants behave across temporal resolutions therefore remains a necessary reference baseline, both for practitioners and for future comparisons against attention-based and probabilistic approaches.

In Vietnam, several studies have also investigated the use of RNN-based models for solar power forecasting. However, the majority of these efforts have focused primarily on LSTM and GRU. For example, LSTM has been applied to forecast power generation of large-scale solar plants considering seasonal weather variability

[35], [36], and a GRU model has been used for short-term load forecasting in the Vietnamese power system [42]. LSTM has also been applied to a solar plant operating under curtailment [43], while GRU has been utilized for forecasting using historical datasets from central Vietnam [37]. Additionally, there have been efforts to integrate LSTM or GRU with other models, such as ANN, to enhance prediction performance [44]. Despite these advancements, the studies in Vietnam tend to rely on unidirectional and standalone RNN models, with limited consideration of the superior predictive capacity of bidirectional configurations like BiGRU and BiLSTM. Existing research suggests that these bidirectional models generally outperform their unidirectional counterparts due to their ability to process input sequences in both forward and backward directions, thereby capturing more contextual information [45].

Another limitation observed in the literature is the lack of comparative studies evaluating the performance of different RNN-based models under a unified experimental framework. Given the heterogeneity among solar power plants, including differences in geographic location, weather patterns, and system configurations, it is crucial to conduct research that not only compares model performance but also develops tailored models specific to each power plant. Moreover, there is a need to establish a standardized methodology that guides the entire forecasting pipeline, from data acquisition and preprocessing to model training and performance evaluation. Such a framework would enable the identification of optimal models suited to the specific characteristics of each solar installation. Importantly, without a controlled experimental setup, it is difficult to isolate whether observed performance differences arise from model architecture or from external factors such as data quality and preprocessing choices.

An additional area that merits attention is the inconsistency in the use of performance metrics across different studies. While some research evaluates models using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) [26], [27], [29] or only RMSE and MAPE [30], others employ different combinations, including Mean Squared Error (MSE), coefficient of determination (R^2), Symmetric Mean Absolute Percentage Error (SMAPE), and Mean Relative Error (MRE) [44], [46], [47], [48]. This inconsistency makes it difficult to directly compare results across studies. In contrast, some researchers argue that the use of a comprehensive set of metrics provides a more reliable and well-rounded evaluation of model performance [49]. Therefore, future research should aim to incorporate a broader suite of evaluation indicators to facilitate meaningful comparisons and benchmarking across studies.

Another important research gap lies in the evaluation of forecasting models across different temporal resolutions. Existing studies often focus on fixed forecasting intervals, such as 5, 30 or 60 minutes, without analyzing how model performance varies across these time scales. For instance, [50] evaluated solar energy prediction at 10-minute resolutions, while [48] assessed

performance at 5-minute intervals. However, few studies have explicitly investigated which models are best suited for different temporal resolutions. More importantly, the fundamental reasons why forecasting difficulty changes across temporal resolutions, such as information loss due to aggregation, smoothing effects, and changes in temporal dependency structures, remain insufficiently explored. This is a critical issue, as forecasting needs differ depending on the application context. Short-term forecasts (e.g., minutes to one hour ahead) are essential for real-time control, energy pricing, and optimal operation scheduling. Mid-term forecasts (e.g., one day to several days ahead) are more suitable for generating unit control, maintenance planning, and risk management. Long-term forecasts (e.g., months to one year ahead) are primarily used for system planning and site selection of solar plants. Therefore, selecting models that align with the specific forecasting interval and application context is crucial. There is a need for comprehensive research that compares model performance across multiple temporal resolutions and identifies the optimal forecasting strategy for each use case.

Based on the above gaps, this study aims to address the key following research question:

- How does temporal resolution interact with the architectural characteristics of RNN-based models to influence forecasting performance in short-term solar power prediction?
- How does forecasting performance vary across different temporal resolutions and what underlying data characteristics contribute to these variations?
- How do different RNN architectures respond to changes in temporal resolution, and what explains the observed shift in model superiority across forecasting intervals?
- How do data-related factors, including input feature selection, outlier handling, and power distribution characteristics, influence model performance across different temporal resolutions?
- What practical insights can be derived regarding the selection of appropriate RNN architectures for different operational forecasting scenarios?

1.3. Research contribution

This study offers several novel contributions to the field of solar power forecasting, particularly in the context of deep learning-based time-series prediction. First, this research presents a comprehensive comparative analysis of advanced RNN-based models, including GRU, BiGRU, LSTM, and BiLSTM, some of which, particularly BiGRU and BiLSTM, have rarely been applied or evaluated in the context of solar power forecasting. By integrating bidirectional architectures, this study demonstrates their potential in capturing complex temporal dependencies in solar generation data. Unlike conventional benchmarking studies, all models are implemented under a unified and controlled experimental framework, ensuring that performance differences can be

attributed primarily to architectural characteristics rather than external factors.

Second, it provides a systematic and comparative analysis of single-step photovoltaic (PV) power forecasting conducted across multiple temporal resolutions, including 5 minutes, 15 minutes, 30 minutes, and 60 minutes. The proposed framework adopts temporal aggregation to formulate forecasting as a single-step prediction problem at each time scale. This design enables a structured investigation of how temporal resolution affects forecasting difficulty and model behavior, rather than treating each resolution independently. This design could reduce forecasting complexity and enable efficient and consistent performance across multiple very short-term intervals, which are particularly relevant for real-time monitoring, operational control, and grid integration of solar power.

Third, the research introduces a structured evaluation framework using consistent and widely accepted performance metrics, including MAE, MSE, RMSE, NMAPE, and R^2 . This not only ensures a fair comparison across models and forecasting intervals but also addresses inconsistencies found in prior works regarding evaluation standards. Furthermore, the analysis is extended to different output power thresholds, providing insights into model behavior under varying generation conditions, a factor that is rarely considered in prior research. In addition, the study provides empirical evidence that model superiority is not universal but depends on temporal resolution, offering practical guidance for selecting appropriate models in different operational scenarios.

Taken together, the fundamental novelty of this study lies not in proposing a new network architecture, but in the controlled empirical evidence that the three contributions above jointly provide: (i) a unified, single-site experimental design that isolates the effect of architectural characteristics (gating complexity and directionality) from confounding factors that hinder cross-study comparisons; (ii) the systematic documentation of a resolution-dependent crossover in model superiority – BiGRU dominates at the 5–30-minute resolutions while BiLSTM becomes superior at the one-hour resolution – which, to the best of the authors' knowledge, has not been explicitly quantified for a utility-scale PV plant in a tropical monsoon climate; and (iii) a mechanistic explanation linking this crossover to measurable data characteristics (autocorrelation decay, sample-size reduction, and the smoothing of cloud-induced ramps under temporal aggregation), which yields direct model-selection guidance for grid operators forecasting at different resolutions for different operational tasks.

To achieve the study's contribution, the paper is organized as follows: Section 2 provides an overview of the study workflow and the forecasting models used; Section 3 describes the dataset and preprocessing steps; Section 4 presents empirical results; Section 5 discusses key insights and model performance; and Section 6 concludes with key findings and suggests future directions.

2. METHODOLOGY

2.1. Study workflow

This study proposes four advanced RNN variants, namely GRU, BiGRU, LSTM, and BiLSTM, for short-term PV power forecasting under dynamic environmental conditions. The overall workflow is shown in [Figure 1](#). First, data from the selected PV plant, including environmental variables, generated power, and timestamps, are collected and pre-processed to extract relevant features. After normalization, the dataset is divided into training and testing sets. Outlier removal is applied only to the training data to improve model learning. The cleaned data are then used to train the proposed models, which are subsequently saved and evaluated on the testing set using standard performance metrics.

2.2. Time series deep learning models

2.2.1. Recurrent neural networks (RNNs)

RNNs are deep learning models designed to capture sequential dependencies through internal memory mechanisms. Unlike traditional neural networks that treat inputs independently, RNNs account for temporal order, making them suitable for sequential tasks [51]. By applying recurrent loops, RNNs generate outputs based on both current inputs and prior states [52]. This temporal awareness is advantageous in applications such as natural language processing, video analysis, and speech recognition, where context is essential [53], [54]. However, standard RNNs suffer from short-term memory limitations, hindering long-sequence learning [55]. To address this, advanced variants have been developed, including GRU [56], LSTM [57], bidirectional GRU [58], bidirectional LSTM [59]. [Figure 2](#) illustrates the architecture of a basic RNN.

2.2.2. Gated recurrent unit (GRU)

GRU is a refined variant of the RNN designed to mitigate the limitations of short-term memory in conventional RNNs, while maintaining a more compact architecture than LSTM networks [51]. Introduced by Cho et al. [60], GRU simplifies the gating mechanism by merging the input and forget gates found in LSTM into a single update gate, thereby reducing computational complexity. The internal configuration of a GRU unit is illustrated in [Figure 3](#) [61].

Compared to LSTM, GRUs are computationally more efficient due to their simpler architecture, yet still capable of capturing long-term dependencies and mitigating the vanishing gradient problem. Their performance has been shown to be comparable to LSTM across various sequence modeling tasks [62], [63].

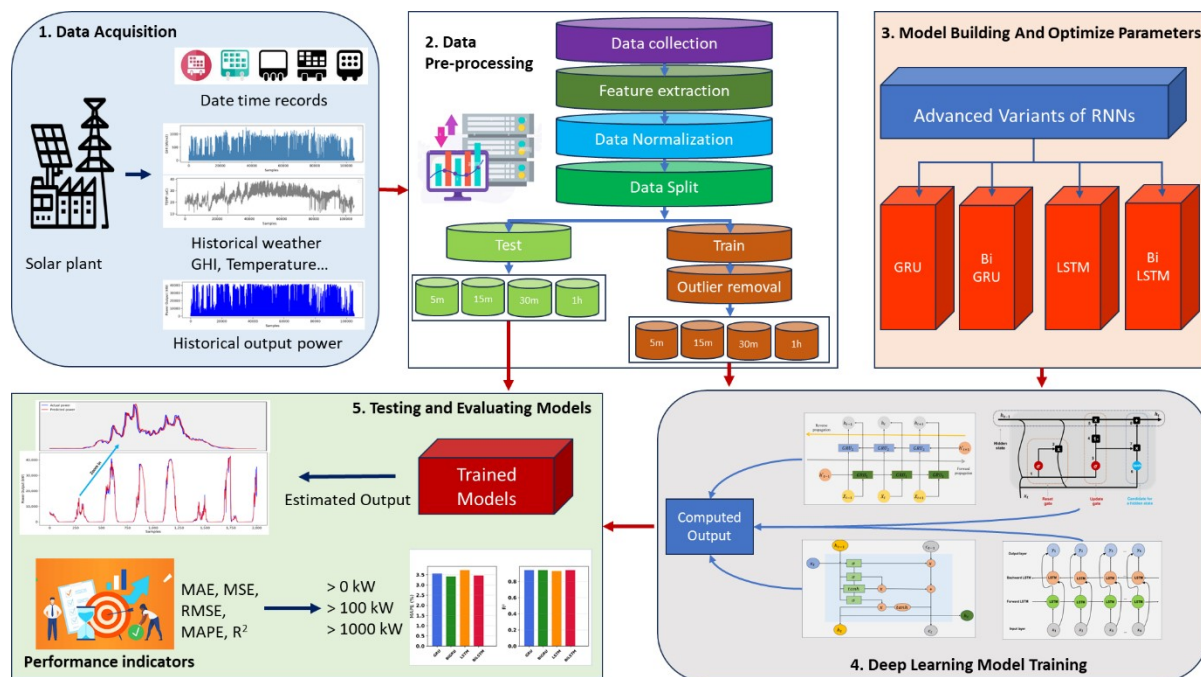


Figure 1. Proposed workflow for advanced RNN-based forecasting models.

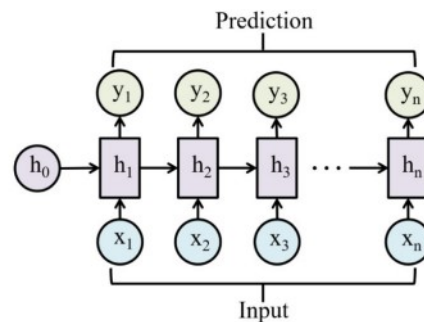


Figure 2. Basic RNN architecture.

2.2.3. Bidirectional gated recurrent unit (BiGRU)

As an evolution of the GRU network, the BiGRU integrates bidirectional processing for improved sequence modeling [64]. In other words, BiGRU enhances the standard GRU model by incorporating information from both preceding and succeeding elements within a sequence [65]. Unlike the traditional GRU, which operates solely in a forward temporal direction, the BiGRU captures dependencies from both directions by employing two GRU layers simultaneously, one that reads the input sequence from start to end and another that reads it in reverse [66]. A diagram of the BiGRU model is illustrated in Figure 4.

2.2.4. Long Short-Term Memory (LSTM)

LSTM is an advanced architecture within the RNN family, designed to address the challenge of learning long-term dependencies in sequential data. It was initially proposed in 1997 [57] and later improved with practical enhancements that increased its efficiency

and stability [67]. Compared to traditional RNNs, LSTM networks are more capable of preserving relevant information across extended sequences, which has led to their widespread adoption in deep learning applications [68], as shown in Figure 5 below.

At each time step, an LSTM unit processes both the current input and the previous hidden state, producing an output that is passed on to the next time step. This mechanism enables the model to maintain a memory of prior events while updating it dynamically. The output of the final hidden layer, or in some applications, the collection of all hidden layers is commonly used for classification tasks [69].

2.2.5. Bidirectional long short-term memory (BiLSTM)

Bi-LSTM networks represent a significant advancement over traditional LSTM models, particularly in tasks that require a deeper understanding of sequential data. While standard LSTM

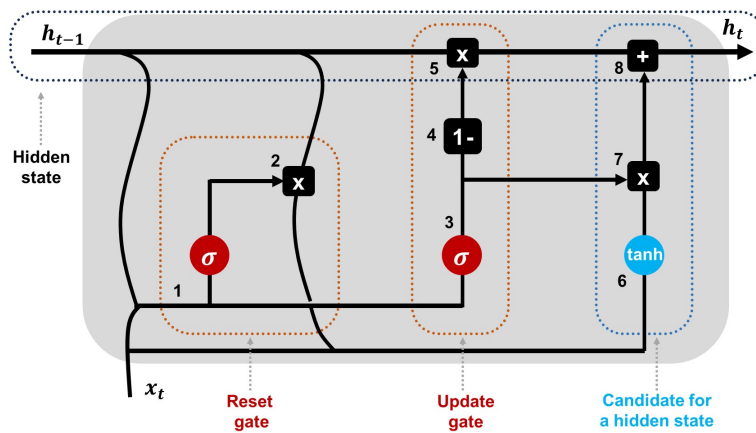


Figure 3. Architecture of the GRU network.

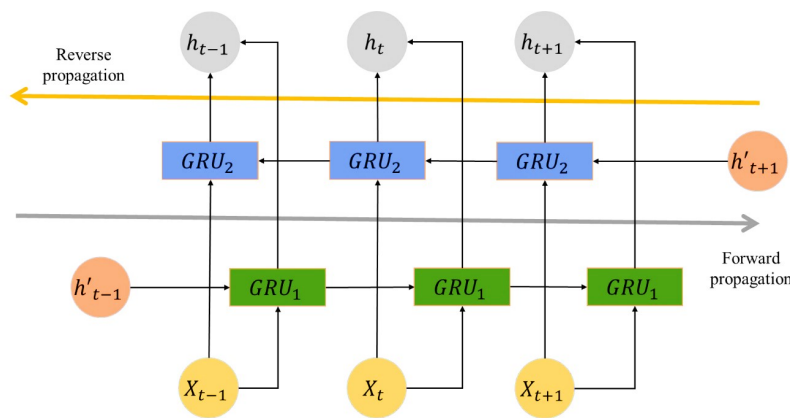


Figure 4. Schematic diagram of the BiGRU model.

architectures process information in a unidirectional manner, typically from past to future, Bi-LSTM enhances this by learning temporal dependencies in both directions [70]. The BiLSTM structure is shown in Figure 6 below.

The Bi-LSTM structure integrates two separate LSTM layers operating in parallel: one traverses the input sequence from left to right (forward direction), and the other processes it in the opposite direction, from right to left (backward direction) [71]. Through this bidirectional flow, the model gains access not only to historical context but also to future context within the input sequence.

2.3. Performance indicators

To evaluate forecasting performance, five metrics were used: MAE, MSE, RMSE, Normalized Mean Absolute Percentage Error (NMAPE), and R^2 . These metrics are determined as Equations (1), (2), (3), (4), and (5), respectively. MAE measures the average absolute prediction error, while MSE and RMSE quantify squared and root-squared errors, giving greater weight to large deviations. NMAPE expresses the prediction error as a percentage of actual values and is commonly used in forecasting studies [48], [72]. However, because NMAPE is sensitive to very small or zero power values, zero-output samples were excluded and progressive filtering was

applied to reduce distortion from low-power periods. R^2 was also used to assess the proportion of variance explained by the model, but it was interpreted together with error-based metrics because it does not directly reflect the absolute magnitude of forecasting errors.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{i1} - y_{i2}| \quad (1)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{i1} - y_{i2})^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i1} - y_{i2})^2} \quad (3)$$

$$NMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_{i1} - y_{i2}|}{y_{i1}} \times 100 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i1} - y_{i2})^2}{\sum_{i=1}^n (y_{i1} - \bar{y}_{i2})^2} \quad (5)$$

In above equations, y_{i1} and y_{i2} denote the actual and forecasted values, respectively. The parameter y_{rated} represents the installed capacity of the photovoltaic (PV) system under study (in this case, $y_{rated} = 49,500$ kW, corresponding to a solar PV plant located in the central region of Vietnam), n denotes the total number of data samples, and $\bar{y}_{i2}$ represents the mean of the actual (measured) values used in the R^2 denominator. All error metrics are computed on de-normalized (original-scale) data, and the same definitions apply at every temporal resolution.

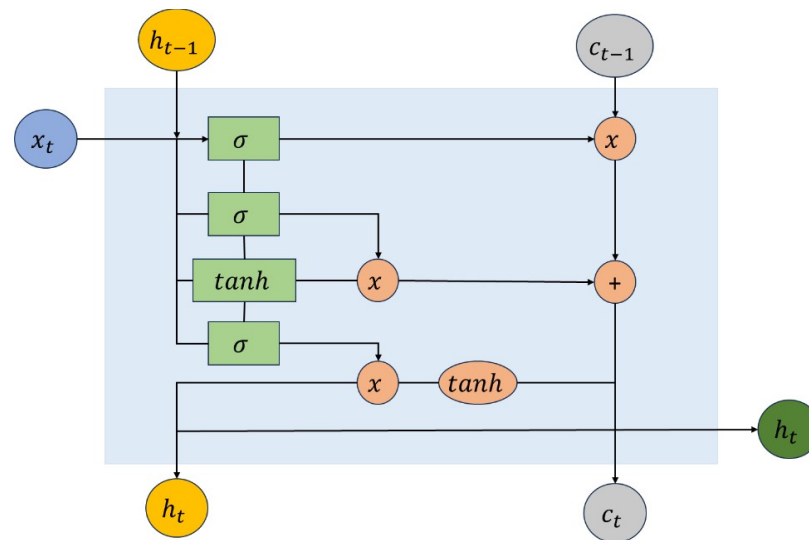


Figure 5. Architecture of the LSTM network.

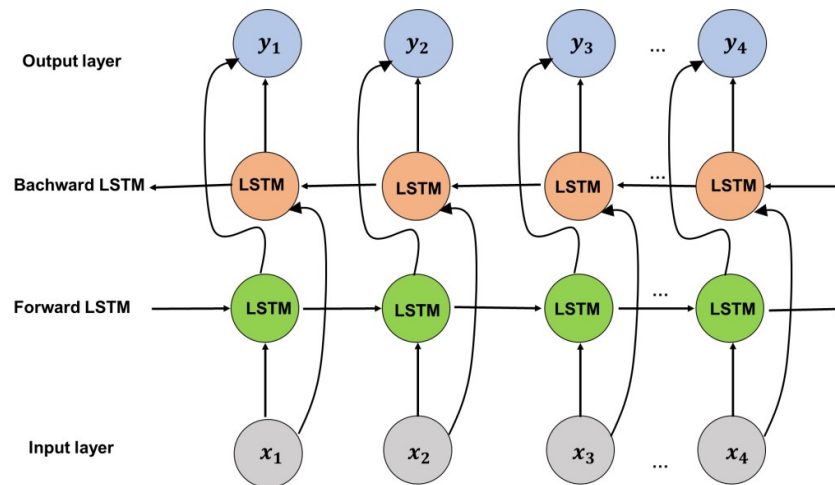


Figure 6. Schematic diagram of the BiLSTM model.

2.4. Model configuration and hyperparameter settings

All four models were implemented in TensorFlow/Keras under an identical training protocol, so that performance differences can be attributed to architectural characteristics rather than tuning effort. The final configurations (Table 2) were selected through preliminary trial-and-error experiments, in which the number of recurrent layers, hidden units, and dropout rates were varied and the configuration with the most stable validation loss was retained; no automated hyperparameter optimization (e.g., grid or Bayesian search) was applied, which is acknowledged as a limitation and left for future work.

All models shared the same settings: a 12-step input window with a single-step-ahead target; Min-Max normalization of inputs and target to [0, 1], fitted on the training set only; the Adam optimizer (initial learning rate 0.001); the MSE loss; a batch size of 64; and a fixed training budget of approximately 4,600 gradient

updates per run on the 5-minute dataset, within which the training and validation losses plateaued without overfitting. To account for stochastic initialization, each configuration was trained five times independently and the run with the lowest NMAPE was retained.

Computationally, all four models are lightweight by modern deep-learning standards: even the largest network (BiGRU) stores fewer than 2.4×10^5 weights (below 1 MB in single precision), and a single forecast requires only a few million multiply-accumulate operations, far below the timing constraints of 5-minute operational forecasting. Because the training protocol was identical, relative training cost scales directly with the parameter counts in Table 2: the GRU cell requires roughly 25% fewer parameters than the LSTM cell owing to its two-gate structure, and bidirectional variants approximately double the cost of their unidirectional counterparts. Wall-clock times, by contrast, depend on the hardware and software environment and transfer poorly across platforms, so these hardware-independent indicators are reported instead. In this light, BiGRU is the most accurate model at a

Table 2. Architecture and hyperparameter settings of the four RNN-based models.

Hyperparameter	GRU	BiGRU	LSTM	BiLSTM
Recurrent layers	2	2 (bidirectional)	2	2 (bidirectional)
Hidden units per layer	64 / 32	128 / 64 (per direction)	64 / 32	64 / 32 (per direction)
Recurrent activation	ReLU	ReLU	ReLU	ReLU
Dropout rates	0.2 / 0.2	0.3 / 0.3	0.3 / 0.2	0.2 / 0.2
Dense layers	1 (output: 1)	2 (hidden: 32, ReLU; output: 1)	1 (output: 1)	1 (output: 1)
Trainable parameters	23,649	233,793	31,137	78,657

Table 3. Monthly summary of key solar power generation parameters at the solar power plant.

Month	Total generated power (kW)	Average output power (kW)	Peak Power (kW)	Average temperature (°C)	Average solar radiation (W/m ² /day)	Average generation hours (hour/day)	Generated power standard deviation
Jan	44,188,075	4,949	41,414	21.1	108.1	11.2	9,408
Feb	30,314,954	3,759	41,517	18.6	80.3	11.6	7,342
Mar	70,644,607	7,913	41,864	24.2	172.5	12.2	12,283
Apr	68,892,096	7,974	41,518	24.6	178.6	12.5	12,317
May	73,557,020	8,239	41,711	27.1	185.8	13.0	12,107
Jun	91,513,665	10,592	41,318	30.8	254.4	13.3	13,485
Jul	90,334,275	10,118	41,482	29.5	238.1	13.2	13,528
Aug	81,118,197	9,086	41,606	29.0	206.8	12.7	12,810
Sep	69,447,330	8,038	41,576	28.0	183.3	12.3	12,193
Oct	38,010,655	4,257	41,349	25.3	114.8	11.7	7,697
Nov	62,621,249	7,248	41,603	25.8	153.3	11.5	11,267
Dec	26,294,145	2,945	40,367	20.0	68.5	11.2	6,139

computational cost that remains trivial for operational deployment, while the unidirectional GRU offers the best accuracy per parameter where the computational budget is the binding constraint.

3. DATA COLLECTION

3.1. Collecting data

The dataset consists of one year of 5-minute interval data collected from a 49.5 MW solar power plant in Quang Tri Province, Vietnam. The site is located in a climatic transition zone affected by tropical storms and the Foehn wind effect, leading to strong short-term variations in solar irradiance and temperature. The dataset includes solar radiation, ambient temperature, module surface temperature, time-related variables, and generated power. These variables allow the forecasting models to capture both meteorological variability and temperature-induced PV performance changes. An overview of the dataset is shown in **Figure 7**.

Figure 8 illustrates the variation in daily power output of the solar power plant throughout the month of April. As observed, the generated power fluctuates significantly from day to day throughout the month which can likely be attributed to changes in weather conditions. These fluctuations highlight the sensitivity of solar generation to meteorological factors.

Figure 9 provides a typical daily power output profile, specifically for April 8. It is evident that the plant operates for approximately 11 to 12 hours per day, typically starting around 6 a.m. and ceasing generation by 5 p.m. This corresponds to the expected solar window during this time of year.

The monthly characteristics of the dataset are summarized in **Table 3**, including generated power, average output, peak power, temperature, solar radiation, generation hours, and output variability.

Table 3 shows a clear seasonal pattern in solar power generation. The highest total generation occurred in June, reaching approximately 91.5 million kWh, followed by July and August, mainly due to higher solar radiation and longer generation hours. In contrast, December and February recorded the lowest generation levels, reflecting reduced irradiance and shorter daylight duration. Although monthly total and average generation varied substantially, peak power remained relatively stable at around 41 MW, indicating that the installed capacity was unchanged and that energy yield was mainly driven by seasonal and meteorological conditions.

Figure 10 further illustrates the monthly variation in total generated power, average output, and peak power. The figure confirms the strong increase in generation from March to September, with the highest output in June. The relatively stable peak power across months suggests consistent plant operation during optimal sunlight periods, while variations in total

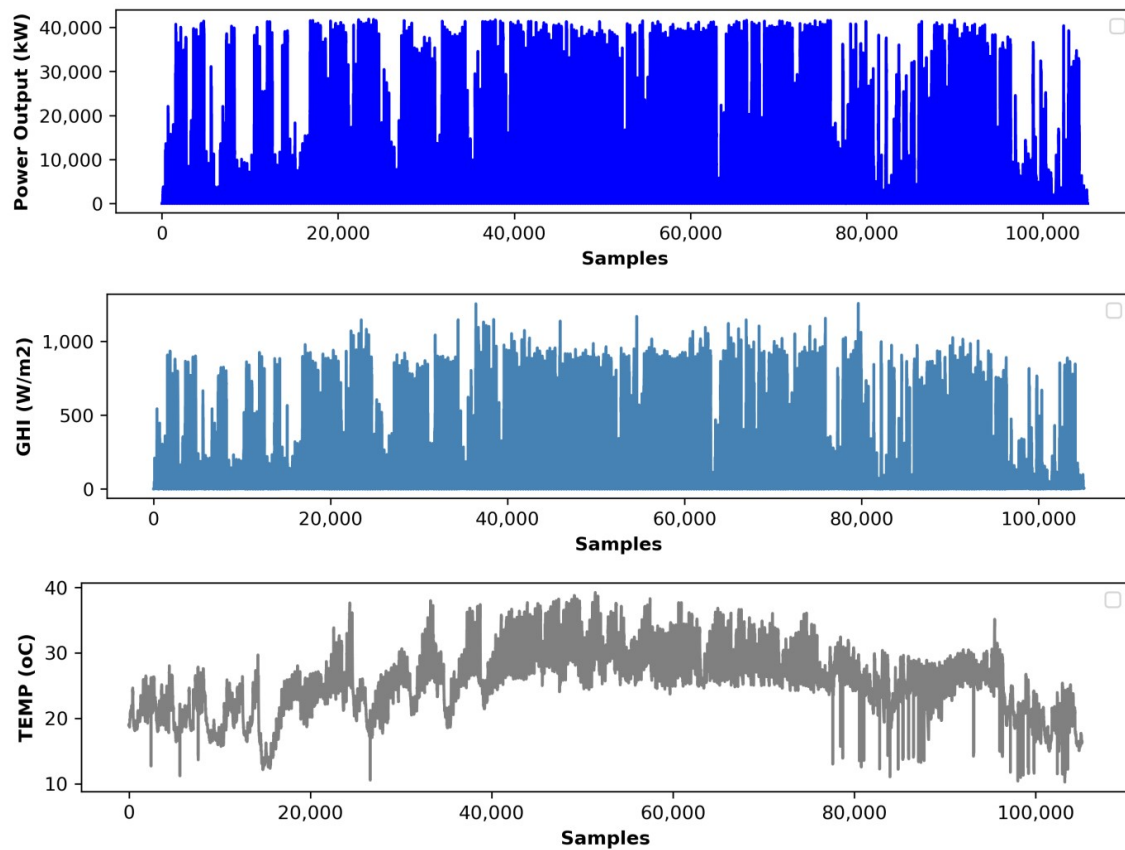


Figure 7. Power output, GHI, and temperature of the plant during the observation period.

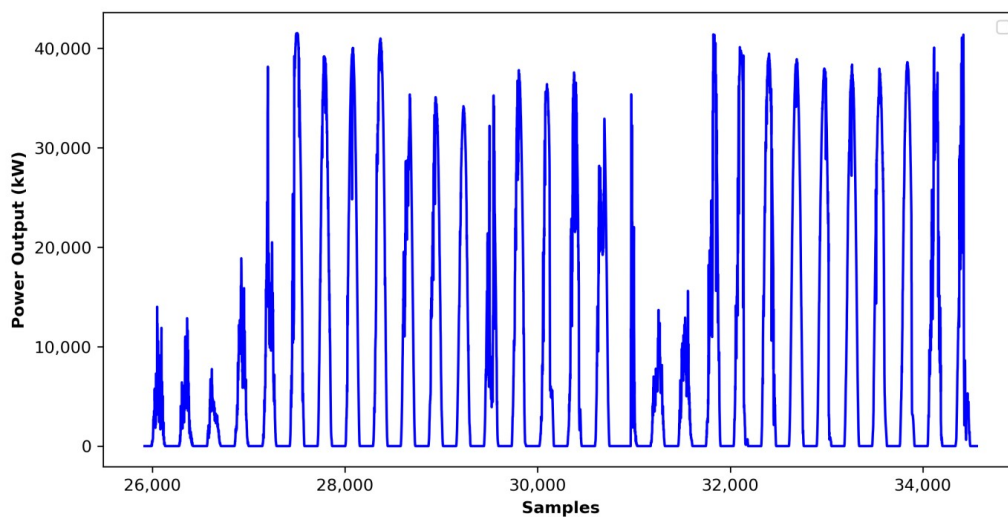


Figure 8. Example of monthly power output in the dataset (April).

generation are mainly associated with changes in solar radiation and daylight hours.

3.2. Data processing

From the raw dataset, key environmental and temporal features were prepared for solar power forecasting. The hour variable was converted into cyclic components ($\sin_{\text{hour}}$ and $\cos_{\text{hour}}$), and a

day_night feature was added to distinguish generation and non-generation periods.

The complete preprocessing pipeline can be summarized as follows. First, all sensor channels (power, irradiance, and temperatures) were aligned on a common 5-minute timestamp grid, and the coarser 15-, 30-, and 60-minute datasets were generated from this base series by averaging the records falling

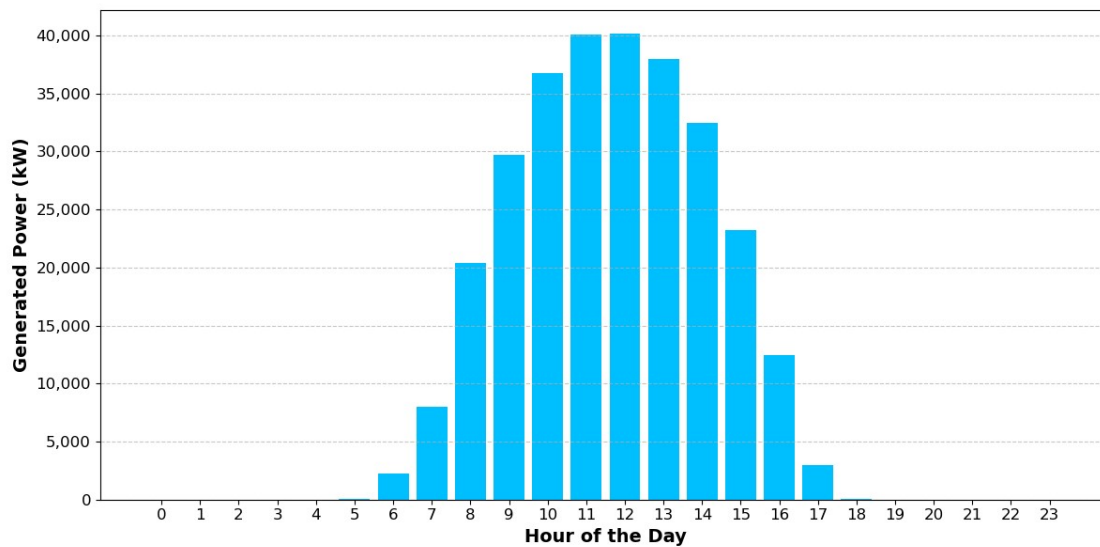


Figure 9. Example of daily power output in the dataset (April 8).

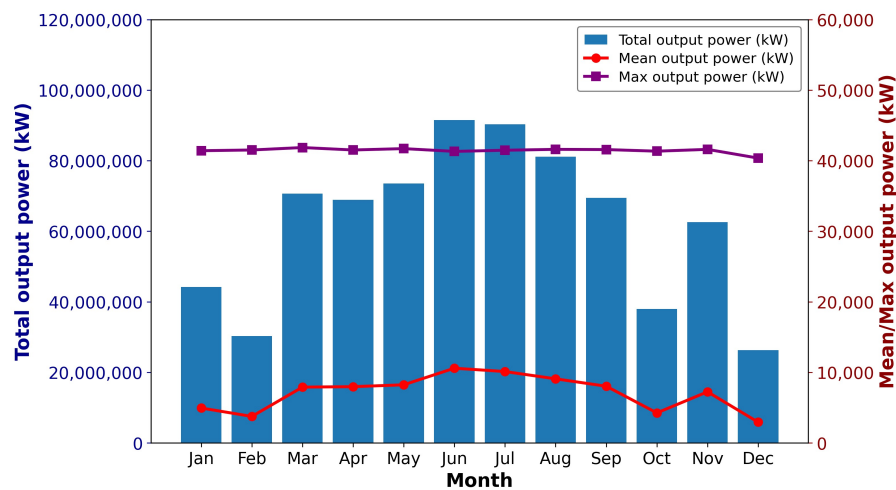


Figure 10. Monthly solar power generation statistics.

within each aggregation window, so that the aggregated series preserve the physical units and the diurnal envelope of the original data. Second, records containing missing or physically implausible sensor values were screened out during quality control, together with the radiation-power outlier filtering described below. Third, nighttime records were retained and explicitly flagged through the `day_night` variable so that the models learn the full diurnal cycle, whereas zero-power (non-generation) samples were excluded from the NMAPE-based evaluation, and results are additionally reported for progressively higher power thresholds ($P > 0$ kW, $P > 100$ kW, and $P > 1,000$ kW) to limit distortion from low-power periods. Fourth, anomalous records were removed from the training set only, while the test set was left untouched to reflect realistic operating conditions. Finally, all input features and the target variable were normalized to the range $[0, 1]$ using Min-Max scaling fitted exclusively on the training data, and all forecasts were

transformed back to physical units (kW) before any error metric was computed.

Figure 11 presents the correlation between the input features and power output. Solar radiation shows the strongest correlation with power generation ($r = 0.987$), followed by cell temperature ($r = 0.893$), `day_night` ($r = 0.612$), and ambient temperature ($r = 0.560$). In contrast, `cos_hour` shows a negative correlation ($r = -0.699$), reflecting the daily solar generation cycle. Feature selection was guided jointly by the Pearson correlation analysis and domain knowledge. Variables exhibiting a strong association with the power output (radiation, `temperature_cell`, `temperature`, and `day_night`) were retained, and the paired cyclic encodings `sin_hour` and `cos_hour` were kept together to preserve a unique, continuous representation of the time of day, while the calendar month was included to convey seasonal context. In addition, the historical generated power itself was included in the input vector,

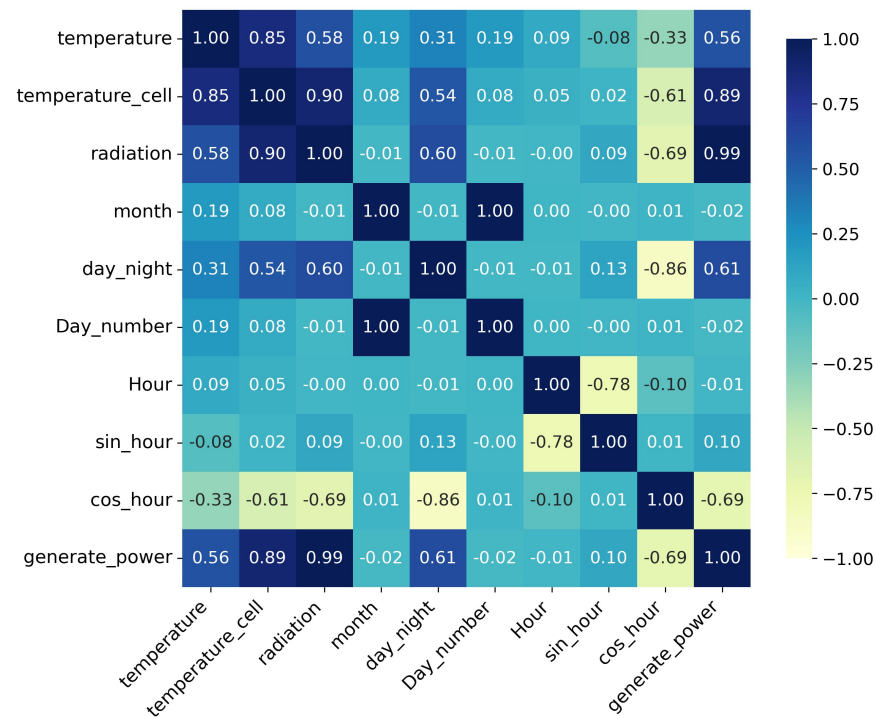


Figure 11. Correlation between input features and power output.

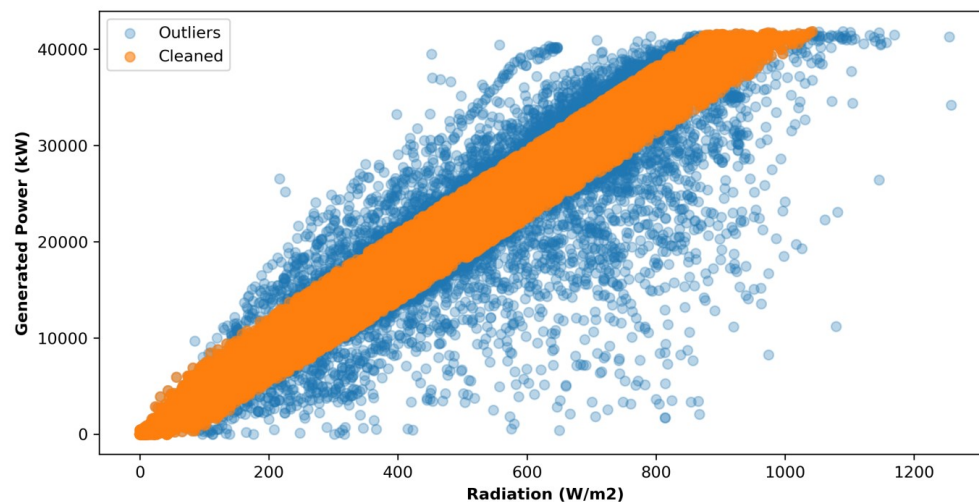


Figure 12. Result of outlier removal based on the radiation–power relationship.

so that each model receives autoregressive information alongside the exogenous meteorological drivers. The resulting input vector therefore comprises eight features: radiation, temperature, temperature_cell, day_night, sin_hour, cos_hour, month, and the lagged generated power.

To preserve seasonal representation, the 3rd, 6th, 13th, 16th, 23rd, and 26th days of each month were assigned to the test set, while the remaining data were used for training. This resulted in 19.7% testing data and 80.3% training data.

Outliers were removed from the training set using the radiation–power relationship and a 95% prediction interval. As shown in

Figure 12, 3,790 anomalous records were excluded, leaving 97,540 data points for the 5-minute forecasting task.

4. RESULTS

This section summarizes the model results for different forecasting time resolutions. Additionally, the results are also aggregated across different power output ranges of the solar power plant, corresponding to electricity generation during various time periods throughout the day.

4.1. Forecasting results for the 5-minute interval

The performance of four deep learning models, including GRU, BiGRU, LSTM, and BiLSTM, was evaluated for 5-minute solar power forecasting across three power output scenarios: $P_{\text{test}} > 0$ kW, $P_{\text{test}} > 100$ kW, and $P_{\text{test}} > 1,000$ kW. The results are summarized in **Table 4** below:

Referring to **Table 4** above, among the models, BiGRU consistently achieved the best performance, showing the lowest MAE, RMSE, and NMAPE values across all scenarios while maintaining high R^2 scores (~ 0.94). This indicates that BiGRU is highly effective for very short-term forecasting tasks. This trend is also clearly illustrated, as shown in **Figure 13**, which compares actual and predicted power values for the different models. From the graph, we can observe that all models perform relatively well, with predicted values closely aligning with actual measurements; however, BiGRU shows the highest degree of similarity, further confirming its high accuracy.

For 5-minute forecasting resolution, GRU ranked second in performance, showing slightly higher error metrics than BiGRU but still yielding reliable accuracy. BiLSTM performed comparably to GRU in the lower and middle output ranges but slightly outperformed it at higher thresholds. LSTM, in contrast, produced the highest errors across all metrics, indicating its relative limitations in this context. As the power output threshold increased, all models exhibited higher errors, reflecting greater variability in power generation during peak sunlight hours. Nevertheless, BiGRU retained strong predictive accuracy, especially in the $P_{\text{test}} > 1,000$ kW scenario, with the lowest RMSE (3,350 kW) and NMAPE (4.06%). Notably, the bidirectional models (BiGRU and BiLSTM) generally outperformed their unidirectional counterparts, highlighting the benefit of incorporating both past and future temporal contexts in short-term forecasting models.

Refer to **Figure 14**, examined models tend to perform better at lower power output levels. As the actual power increases, the predicted values tend to deviate downward from the reference line, indicating a consistent underprediction at higher power levels.

4.2. Forecasting results for the 15-minute interval

Table 5 below summarizes the forecasting performance of four deep learning models, including GRU, BiGRU, LSTM, and BiLSTM, over a 15-minute prediction interval.

For 15-minute resolution and under the full dataset scenario, the BiGRU model still demonstrates the best overall performance, achieving the lowest MAE (1,966 kW), RMSE (3,178 kW), and NMAPE (3.97%), while also maintaining a high R^2 score of 0.94. This indicates a strong fit between the predicted and actual values. BiLSTM followed closely with MAE and RMSE values of 1,979 kW and 3,232 kW, respectively, and an R^2 of 0.93. In contrast, the LSTM model showed the weakest performance in this scenario, with the highest MAE (2,290 kW), RMSE (3,494 kW), and the lowest R^2

(0.92), suggesting limitations in capturing short-term dependencies in the time series data.

As the minimum power output threshold increased, all models experienced a decline in performance. This trend is evidenced by rising MAE, RMSE, and NMAPE values, along with decreasing R^2 scores across all scenarios. For example, the GRU model's MAE increased from 2,129 kW to 2,362 kW and R^2 dropped from 0.93 to 0.92 as the threshold changed from $P_{\text{test}} > 0$ kW to $P_{\text{test}} > 1,000$ kW. This could suggest that forecasting becomes more challenging as the power output rises, likely due to increased variability and complexity in high-power scenarios.

Among the four models, BiGRU consistently delivered the most robust and stable performance across all power thresholds. Even under the high-power scenario ($P_{\text{test}} > 1,000$ kW), BiGRU maintained relatively low errors (MAE = 2,204 kW, RMSE = 3,496 kW) and a strong R^2 value of 0.93. This implies that the bidirectional structure of BiGRU enhances its ability to capture both past and future context in sequential data, contributing to improved accuracy and generalization.

In contrast, the LSTM model continued to show relatively poor performance in high-power scenarios, with the highest observed MAE (2,599 kW) and the lowest R^2 (0.90), highlighting its limitations in effectively managing more complex patterns in the data. While BiLSTM showed improved results over LSTM, its performance still fell slightly behind BiGRU, particularly in scenarios involving higher power thresholds.

As seen in **Table 5**, the BiGRU model still emerged as the most suitable architecture for short-term solar power forecasting at a 15-minute resolution. Its superior accuracy and resilience across varying power levels make it a strong candidate for real-world applications where prediction reliability is critical. BiLSTM also provides a viable alternative with relatively strong results.

4.3. Forecasting results for the 30-minute interval

The forecasting performance of four proposed models was evaluated over a 30-minute interval, with the results outlined in **Table 6**.

As shown in the performance table, all models exhibit a general decline in accuracy compared to the 15-minute interval, which is expected due to the coarser temporal resolution. Among them, BiGRU consistently delivers the best overall performance, with the lowest MAE (2,317 kW), RMSE (3,562 kW), and NMAPE (4.68%) under the general scenario ($P_{\text{test}} > 0$ kW), along with the highest R^2 value of 0.92. This suggests that BiGRU maintains robust predictive accuracy even over extended intervals.

GRU also performs relatively well, especially in lower-load scenarios, but shows a more noticeable drop in performance as the power level increases—its R^2 value falls from 0.91 to 0.86 when $P_{\text{test}} > 1,000$ kW. Meanwhile, LSTM records the weakest performance across all metrics, particularly for higher power outputs, indicating its limitations in capturing longer temporal dependencies compared to gated recurrent variants. BiLSTM performs

Table 4. 5-minute forecasting performance results.

Model	Scenario	MAE (kW)	MSE (kW ²)	RMSE (kW)	NMAPE (%)	R ²
GRU	P _{test} > 0 kW	1,763	9,612,588	3,100	3.56	0.94
	P _{test} > 100 kW	2,021	10,788,168	3,285	4.08	0.93
	P _{test} > 1,000 kW	2,176	11,858,173	3,444	4.40	0.92
BiGRU	P _{test} > 0 kW	1,687	9,894,949	3,146	3.41	0.94
	P _{test} > 100 kW	1,843	10,178,627	3,190	3.72	0.94
	P _{test} > 1,000 kW	2,008	11,221,858	3,350	4.06	0.93
LSTM	P _{test} > 0 kW	1,841	10,740,810	3,277	3.72	0.93
	P _{test} > 100 kW	2,016	11,122,690	3,335	4.07	0.93
	P _{test} > 1,000 kW	2,167	12,182,287	3,490	4.38	0.92
BiLSTM	P _{test} > 0 kW	1,713	9,693,621	3,113	3.46	0.94
	P _{test} > 100 kW	1,754	9,968,901	3,157	3.54	0.94
	P _{test} > 1,000 kW	1,893	10,946,173	3,309	3.82	0.93

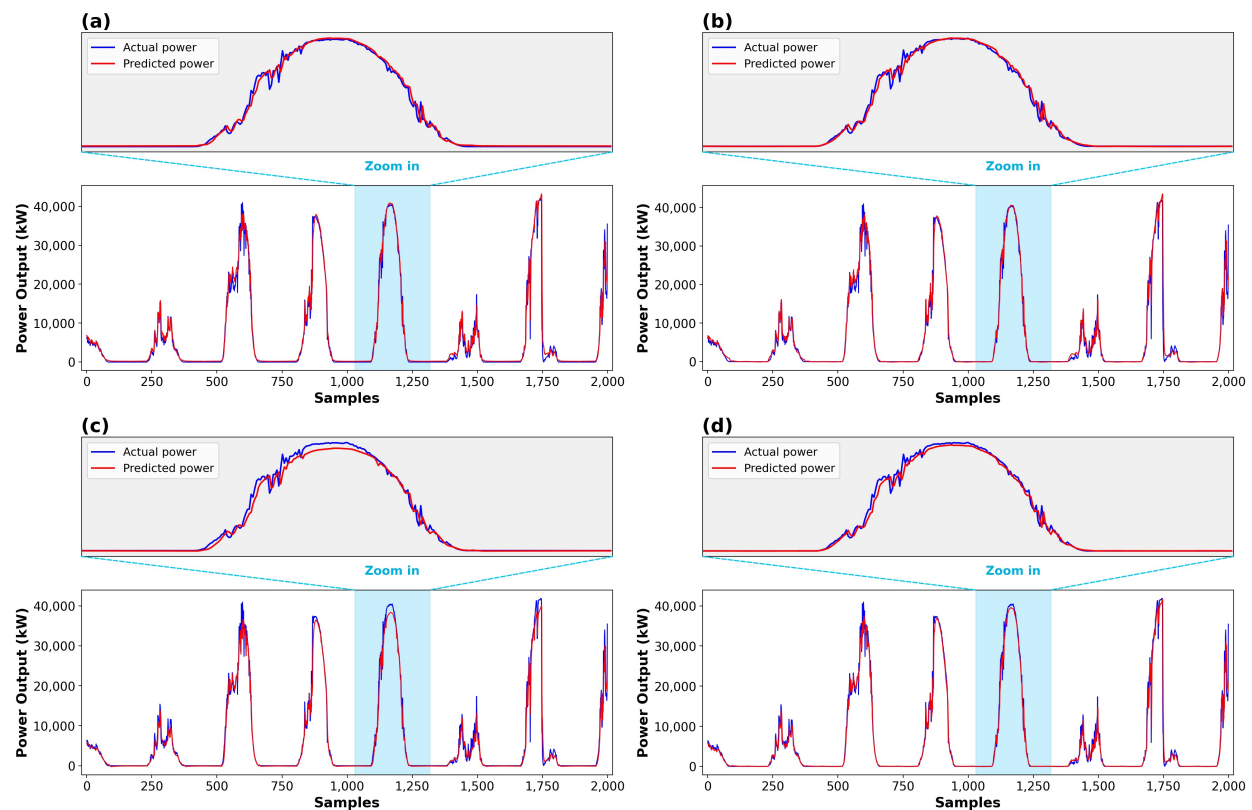


Figure 13. 5-Minute interval forecasting — (a) BiGRU, (b) GRU, (c) LSTM, (d) BiLSTM.

Table 5. 15-minute forecasting performance results.

Model	Scenario	MAE (kW)	MSE (kW ²)	RMSE (kW)	NMAPE (%)	R ²
GRU	$P_{\text{test}} > 0$ kW	2,129	11,202,948	3,347	4.30	0.93
	$P_{\text{test}} > 100$ kW	2,237	10,999,693	3,317	4.52	0.93
	$P_{\text{test}} > 1,000$ kW	2,362	11,886,779	3,448	4.77	0.92
BiGRU	$P_{\text{test}} > 0$ kW	1,966	10,101,846	3,178	3.97	0.94
	$P_{\text{test}} > 100$ kW	2,047	11,216,483	3,349	4.14	0.93
	$P_{\text{test}} > 1,000$ kW	2,204	12,218,956	3,496	4.45	0.93
LSTM	$P_{\text{test}} > 0$ kW	2,290	12,208,136	3,494	4.63	0.92
	$P_{\text{test}} > 100$ kW	2,441	13,151,289	3,626	4.93	0.92
	$P_{\text{test}} > 1,000$ kW	2,599	14,234,745	3,773	5.25	0.90
BiLSTM	$P_{\text{test}} > 0$ kW	1,979	10,443,060	3,232	4.00	0.93
	$P_{\text{test}} > 100$ kW	2,266	12,304,329	3,508	4.58	0.92
	$P_{\text{test}} > 1,000$ kW	2,404	13,397,038	3,660	4.86	0.91

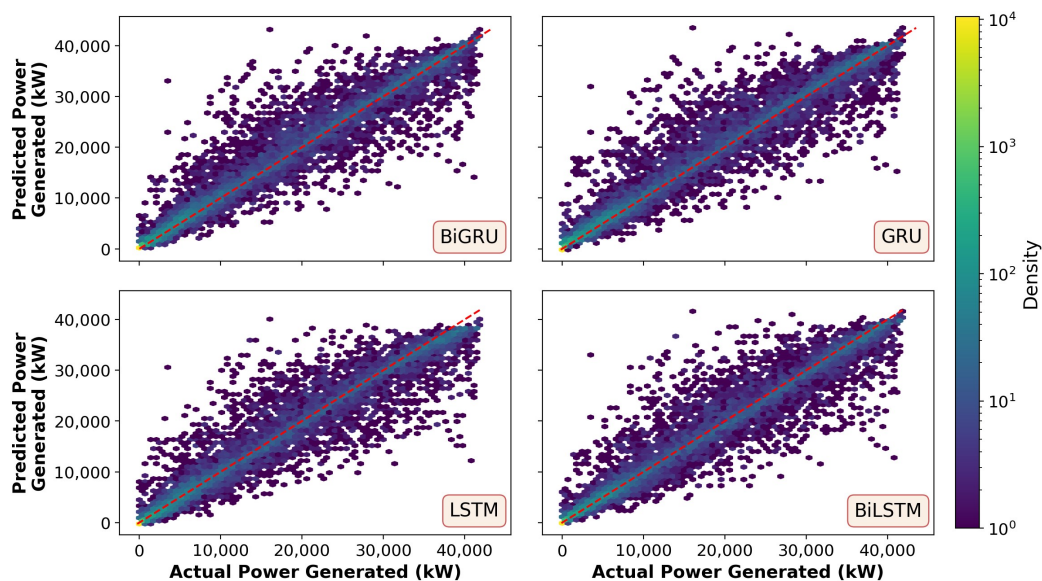


Figure 14. Model accuracy evaluation via density plot (5-minute forecasting).

However, it shows a substantial decline in performance under high-demand scenarios, with MAE rising to 3,325 kW and R² dropping to 0.86.

4.4. Forecasting results for the one-hour interval

The forecasting performance of four selected models over a one-hour prediction interval is gathered in Table 7 below.

When extending the forecasting resolution to a one-hour interval, the predictive performance of the examined models

further declines, reflecting the inherent difficulty of long-term solar power prediction due to insufficient historical data. Among the models, BiLSTM emerges as the most effective, particularly in low- and mid-range power scenarios, achieving the lowest RMSE (4,785 kW), MAE (3,471 kW), and NMAPE (7.01%) when $P_{\text{test}} > 0$ kW, alongside an R² value of 0.85. These results suggest that BiLSTM can effectively retain and utilize longer-term sequential patterns, enhancing its stability over extended intervals.

While GRU continues to perform reasonably well in lower power output ranges, its prediction accuracy deteriorates significantly in

high-power scenarios. Specifically, its R^2 value drops from 0.88 ($P_{\text{test}} > 0$ kW) to 0.77 ($P_{\text{test}} > 1,000$ kW), indicating increased error dispersion under more volatile conditions. BiGRU, although previously leading at shorter intervals, appears to struggle with one-hour forecasting, posting a higher MAE and RMSE than BiLSTM, especially for $P_{\text{test}} > 0$ kW. However, its performance remains more consistent than LSTM's.

LSTM demonstrates the weakest performance across all evaluated conditions, particularly at higher power levels. For instance, with $P_{\text{test}} > 1,000$ kW, it yields the highest error values—MAE of 5,601 kW, RMSE of 7,140 kW, and NMAPE of 11.32%, with a relatively low R^2 of 0.64, signaling that its ability to generalize deteriorates significantly over extended time frames.

As observed from Figure 15, all four forecasting models perform well when actual power outputs vary smoothly. However, they face significant challenges in accurately predicting sudden fluctuations in power output.

Regarding Figure 16, although the number of data points significantly decreases in the one-hour forecasting interval, it is difficult to visually distinguish accuracy differences among the models based solely on the density plots. However, when combined with the summary in Table 7, it becomes evident that the prediction error widens noticeably over the one-hour interval, with LSTM exhibiting the highest variance. In contrast, BiLSTM demonstrates a tighter alignment with actual outputs, particularly in the mid-range power bands. These findings underscore BiLSTM's relative advantage in long-term solar power forecasting while also highlighting the challenge of maintaining model accuracy as the forecast resolution extends.

4.5. Benchmark comparison with a representative attention-based model

To situate the RNN family within the broader landscape of deep-learning forecasting approaches, an additional benchmark was conducted at the 5-minute resolution against a vanilla Transformer encoder, the archetypal member of the attention-based family emphasized in recent literature [33], [39], [40], whose popular variants (e.g., the Temporal Fusion Transformer) share the same core self-attention mechanism. The 5-minute resolution was chosen as the representative case because it is the headline operational resolution of this study and provides the largest training set, which is the most favorable condition for data-intensive attention models. The Transformer was evaluated under the identical experimental protocol used for the RNN models: the same eight input features, Min-Max scaling fitted on the training set only, a 12-step input window with a single-step target, the same optimizer, loss function, batch size, and epoch budget, and five independent training runs with the best run selected by NMAPE. The encoder uses a model dimension of 64, four attention heads, two encoder blocks, a feed-forward dimension of 128, a dropout rate of 0.1, and sinusoidal positional encoding, totaling 69,633

trainable parameters, which places it between the LSTM and BiLSTM in size. The results are summarized in Table 8, in which the best-performing RNN at this resolution (BiGRU) is reproduced from Table 4 for direct comparison.

As shown in Table 8, the vanilla Transformer and the bidirectional RNNs occupy the same accuracy tier under the identical, deliberately modest training budget. BiGRU attains the lowest errors in the full-data scenario (NMAPE of 3.41% versus 3.54%, and MAE of 1,687 kW versus 1,754 kW), while the margins at the higher power thresholds are small in either direction, well within the modest-margin range discussed in Section 5.4. This parity indicates that, for single-plant, short-horizon forecasting with moderate data volumes, gated recurrent architectures remain fully competitive with attention-based models, which is consistent with recent comparative studies reporting that Transformers tend to show their clearest advantages at longer horizons and larger data scales [33], [39]. At the same time, attention architectures offer natural extensions toward multi-step decoding and probabilistic forecasting [34].

5. DISCUSSION

5.1. Overall forecasting performance across forecast intervals

Figure 17 below illustrates the performance of the four proposed models in the task of solar power generation forecasting across four different forecast intervals within the scope of the study.

In relation to Figure 17, the forecasting performance tends to decline as the forecast interval increases. Specifically, at shorter intervals such as 5 and 15 minutes, the models achieve high accuracy with low MAE, RMSE, and NMAPE values, and high R^2 scores. However, performance significantly deteriorates at the 1-hour interval. This phenomenon can be explained by the fact that temporal aggregation reduces high-frequency variability while simultaneously introducing uncertainty due to information loss, making it more difficult for models to reconstruct short-term dynamics.

Among the compared models, BiGRU demonstrates superior performance across most time frames, particularly at the 5, 15, and 30-minute intervals. This highlights the advantage of the bidirectional architecture in capturing short-term temporal dependencies in the data. More specifically, the GRU structure, with its simpler gating mechanism compared to LSTM, is more responsive to rapid fluctuations in high-frequency data, allowing it to adapt quickly to short-term changes in solar irradiance. When combined with bidirectional learning, BiGRU effectively captures both immediate past and near-future context, which is particularly beneficial in short-interval forecasting where temporal continuity is strong. Conversely, LSTM tends to deliver the weakest results, with higher error metrics and lower R^2 values, especially at the 30-minute and 1-hour intervals. This may be attributed to its more

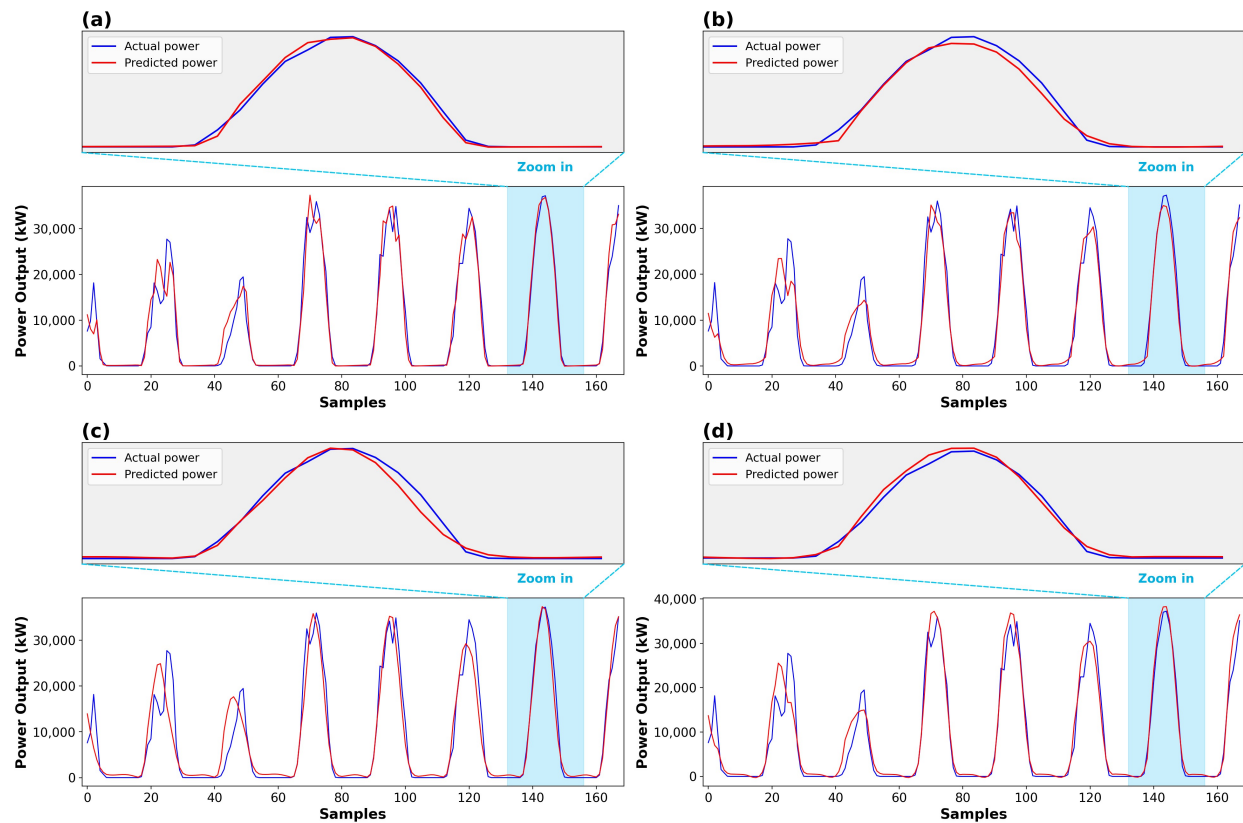


Figure 15. One-hour interval forecasting — (a) BiGRU, (b) GRU, (c) LSTM, (d) BiLSTM.

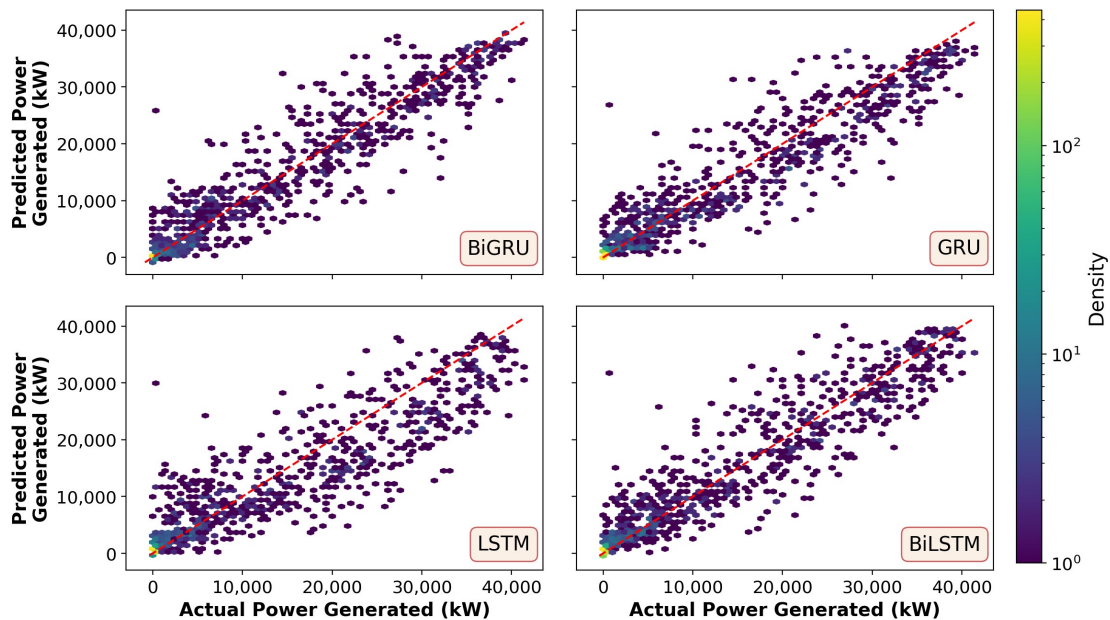


Figure 16. Model accuracy evaluation via density plot (one-hour forecasting).

complex gating structure, which can lead to slower adaptation to rapid changes and potential over-smoothing in highly dynamic conditions.

A more mechanistic explanation can be offered by considering how temporal aggregation alters the statistical structure of the

series. Averaging twelve 5-minute records into an hourly value act as a low-pass filter: it attenuates the high-frequency, cloud-induced ramps that dominate sub-15-minute variability, but it also weakens the strong short-lag autocorrelation that short-horizon

Table 6. 30-minute forecasting performance results.

Model	Scenario	MAE (kW)	MSE (kW ²)	RMSE (kW)	NMAPE (%)	R ²
GRU	P _{test} > 0 kW	2,430	14,328,650	3,785	4.91	0.91
	P _{test} > 100 kW	3,148	18,871,183	4,344	6.36	0.88
	P _{test} > 1,000 kW	3,373	20,423,496	4,519	6.81	0.86
BiGRU	P _{test} > 0 kW	2,317	12,686,418	3,562	4.68	0.92
	P _{test} > 100 kW	2,435	13,428,170	3,664	4.92	0.91
	P _{test} > 1,000 kW	2,587	14,351,030	3,788	5.23	0.90
LSTM	P _{test} > 0 kW	2,835	17,212,567	4,149	5.73	0.89
	P _{test} > 100 kW	2,955	18,180,101	4,264	5.97	0.88
	P _{test} > 1,000 kW	3,114	19,063,289	4,366	6.29	0.87
BiLSTM	P _{test} > 0 kW	2,378	13,481,783	3,672	4.80	0.91
	P _{test} > 100 kW	3,088	18,471,773	4,298	6.24	0.88
	P _{test} > 1000 kW	3,325	19,971,904	4,469	6.72	0.86

Table 7. One-hour forecasting performance results.

Model	Scenario	MAE (kW)	MSE (kW ²)	RMSE (kW)	NMAPE (%)	R ²
GRU	P _{test} > 0 kW	3,112	18,764,421	4,332	6.29	0.88
	P _{test} > 100 kW	4,068	30,025,874	5,480	8.22	0.80
	P _{test} > 1,000 kW	4,494	33,418,365	5,781	9.08	0.77
BiGRU	P _{test} > 0 kW	3,745	25,705,124	5,070	7.57	0.83
	P _{test} > 100 kW	3,961	27,233,136	5,219	8.00	0.82
	P _{test} > 1,000 kW	4,418	30,417,057	5,515	8.92	0.79
LSTM	P _{test} > 0 kW	3,918	28,669,291	5,354	7.91	0.81
	P _{test} > 100 kW	5,049	45,820,542	6,769	10.20	0.7
	P _{test} > 1,000 kW	5,601	50,973,014	7,140	11.32	0.64
BiLSTM	P _{test} > 0 kW	3,471	22,899,539	4,785	7.01	0.85
	P _{test} > 100 kW	3,624	24,200,197	4,919	7.32	0.84
	P _{test} > 1,000 kW	3,909	26,246,063	5,123	7.90	0.82

Table 8. Benchmark results of the vanilla Transformer at the 5-minute resolution, with BiGRU reproduced from Table 4 for comparison.

Model	Scenario	MAE (kW)	MSE (kW ²)	RMSE (kW)	NMAPE (%)	R ²
Transformer	P > 0 kW	1,754	10,097,495	3,178	3.54	0.94
	P > 100 kW	1,797	10,385,303	3,223	3.63	0.94
	P > 1,000 kW	1,946	11,414,147	3,378	3.93	0.93
BiGRU	P > 0 kW	1,687	9,894,949	3,146	3.41	0.94
	P > 100 kW	1,843	10,178,627	3,190	3.72	0.94
	P > 1,000 kW	2,008	11,221,858	3,350	4.06	0.93

highly correlated, so the forecasting problem is dominated by local pattern continuation; the compact two-gate GRU cell, with fewer parameters to fit (Table 2), adapts quickly to these local dynamics and is less prone to overfitting, while the backward pass of the bidirectional wrapper supplies additional context about the shape of the current ramp within the input window. As the resolution coarsens, each 12-step input window spans a progressively longer physical duration (one hour at the 5-minute scale versus half a day at the hourly scale), the number of training samples shrinks by a factor of twelve, and the residual predictable structure shifts from local continuation toward slower, weather-regime-scale dependencies. Under these conditions, the explicit cell state of the LSTM, with its separate input, forget, and output gates, becomes advantageous for retaining information across the longer effective horizon, which explains why BiLSTM overtakes BiGRU at the one-hour resolution. Physically, the plant's location in a climatic transition zone affected by tropical storms and the Foehn wind effect produces frequent abrupt irradiance transients; these transients are the dominant error source at fine resolutions, whereas at the hourly scale forecast errors are governed mainly by broader cloud-cover and weather-regime evolution.

5.2. Key input variables that significantly affect model performance

One of the key factors affecting the performance of solar power forecasting models is the distribution characteristics of the input power data. The experimental results in this study clearly demonstrate that deep learning models, particularly RNN architectures such as GRU, BiGRU, LSTM, and BiLSTM, tend to deliver more accurate forecasts at lower power levels, where fluctuations are less abrupt. This indicates that model performance is closely linked to the degree of nonlinearity and volatility present in the data, which varies not only with power levels but also across temporal resolutions. Conversely, as power levels increase, especially near peak capacity, the forecasting error also increases significantly. This is evident through the underestimation of predicted points relative to the reference line in the density plots. Such underestimation can be explained by the difficulty of capturing sudden irradiance drops (e.g., due to cloud movement), which become more pronounced when data is aggregated over longer intervals. These findings are consistent with other studies in literature. For example, a study by Nguyen Tuan Anh et al. using a GRU model showed that the NMAPE increased from 3.14% (for power values > 0 kW) to 3.52% (for power values > 1,000 kW) [37]. Another study at a power plant in central Vietnam reported a full-day NMAPE error (all-hour average) ranging from about 4.0% to 4.7%, while the NMAPE error from 6:00 to 18:00 rose to about 8.0% to 9.4% [35].

Outlier removal in the training dataset is also critical, as it significantly contributes to the quality of forecasting models. In this study, outliers were removed using the linear relationship

between solar radiation and generated power. This preprocessing step ensures that the models learn the true physical relationship rather than noise, which is particularly important when comparing architectures under different temporal resolutions. This highlights the importance of selecting and processing input features, such as indicators related to solar irradiance, temperature, and on-site weather conditions. In a study using the LSTM model [73], the authors proposed replacing time-based variables with solar irradiance data under clear-sky conditions, which led to a significant performance gain, reducing NMAPE from 3.491% to 3.08%, an improvement of up to 12%. This helps reduce noise in model training. Moreover, developing additional features that are closely correlated with power generation, such as day/night classification or solar generation phase angles, could further enhance forecasting quality. These improvements suggest that feature engineering can partially compensate for information loss at coarser temporal resolutions.

Another important consideration is the method used to divide the training and test datasets. The test data should be evenly distributed across different time intervals instead of being concentrated at the end of the dataset. This helps avoid model bias and ensures that the model learns general trends and periodic patterns in both input and output variables. This is especially important when analyzing model behavior across temporal resolutions, as uneven sampling could distort the observed relationship between resolution and performance.

5.3. Developing and optimizing advanced RNN-based forecasting architectures

In this study, the performance of four deep learning models based on RNN architectures was thoroughly evaluated across different forecast intervals. Unlike conventional studies that focus only on accuracy, this analysis aims to explain how architectural differences influence learning behavior under varying temporal conditions. BiGRU consistently outperformed other models in short-term forecasts, particularly at the 5- and 15-minute intervals, achieving average R^2 values between 0.93–0.94 and NMAPE from 3.41% to 3.97%, respectively. This highlights the clear advantage of bidirectional models in leveraging both past and future context to improve accuracy. At short temporal resolutions, the strong autocorrelation in the data allows BiGRU to efficiently exploit local temporal patterns using its relatively lightweight gating mechanism.

GRU and BiLSTM models also performed well, while LSTM generally showed the weakest performance, especially in scenarios with high power output or coarser forecast intervals. Interestingly, at the 1-hour interval, BiLSTM outperforms BiGRU, which can be explained by its more complex memory cell structure. The input, forget, and output gates in LSTM allow it to retain long-term dependencies more effectively, which becomes critical when short-term information is lost due to temporal aggregation. In this

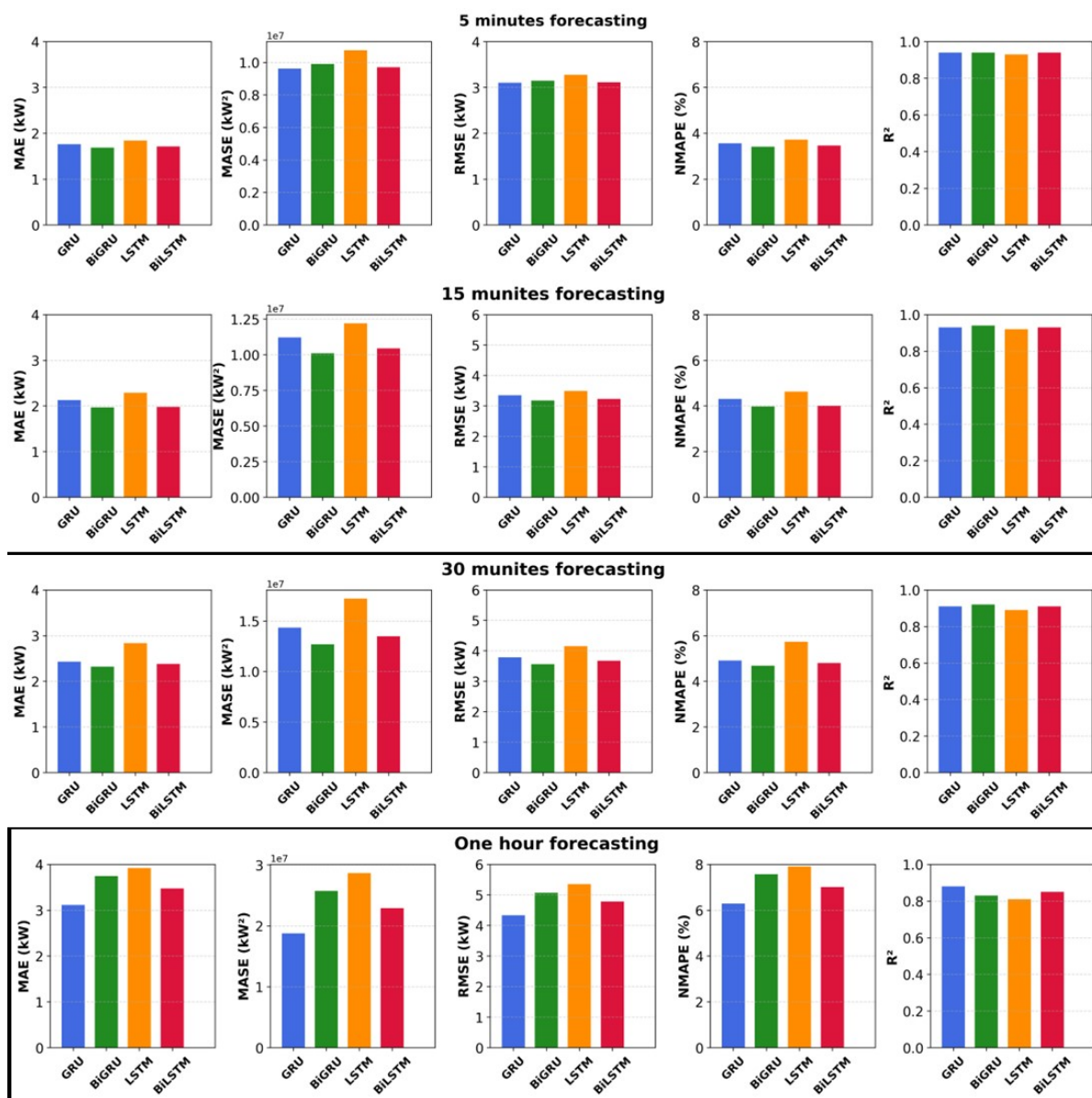


Figure 17. Model performance across forecast intervals.

context, the bidirectional nature of BiLSTM further enhances its ability to reconstruct broader temporal patterns, making it more suitable for coarser resolutions. This is aligned with the study that BiLSTM and BiGRU demonstrated superior predictive accuracy compared to their unidirectional counterparts [45].

Compared to previous studies conducted in the literature, the performance metrics in this study show promising potential for applying advanced variants of RNNs to solar power generation forecasting. In another study applying models such as CatBoost, LightGBM, and XGBoost, the R^2 value reached a maximum of only about 0.608 in the training set and 0.46 in the test set [74], which is significantly lower than the results achieved by BiGRU in this study. A separate study using the LSTM model [48], applied to the

Thanh Cong 1 solar power plant, used forecasted weather data as input. The use of this data caused a significantly higher NMAPE compared to using historical weather data (10.857% versus 3.491%). However, after improving the model by incorporating new features and addressing forecast-related errors, the NMAPE was reduced to 9.881%. This indicates that even when changing the input data source (i.e., forecasted weather), the predictive model's NMAPE remained relatively high. This further suggests that model architecture alone cannot guarantee performance; its effectiveness depends on how well it aligns with the temporal characteristics of the data.

5.4. Evaluating forecasting performance using multiple metrics

The simultaneous use of multiple evaluation metrics such as MAE, MSE, RMSE, NMAPE, MASE, and R^2 allows for a comprehensive assessment of model performance. In particular, R^2 evaluates how well the model fits the actual data, while NMAPE and MAE reflect the absolute and relative errors of the forecasts. In this study, BiGRU demonstrated stability by maintaining NMAPE at low values among other proposed models which is below 4.7% and R^2 above 0.92 in most scenarios (except one-hour forecasting interval), even under high power output conditions. This stability reflects its ability to generalize well under high-frequency conditions where local temporal dependencies dominate.

It is worth noting that, at the finest resolution, the margins separating the leading models are modest (e.g., an NMAPE of 3.41% for BiGRU versus 3.46% for BiLSTM at 5 minutes). The conclusions of this study are therefore not drawn from any single pairwise difference, but from the consistency of the observed ranking: BiGRU attains the lowest errors across all three power thresholds and at all three sub-hourly resolutions, while the reversal in favor of BiLSTM at the one-hour resolution is comparatively large (NMAPE of 7.01% versus 7.57%, and RMSE of 4,785 kW versus 5,070 kW). Moreover, every metric is computed over a large and chronologically representative test set (72 days evenly distributed across all twelve months of the year), and each architecture was trained five times under an identical protocol with the best run retained for every model, so that no architecture benefits from a more favorable training draw. Because the present work adopts a deterministic (point-forecast) formulation for a single plant, in which forecast errors of consecutive test samples are strongly autocorrelated, formal significance procedures designed for comparisons across many independent datasets, as well as probabilistic uncertainty intervals, fall outside its scope; extending the framework toward probabilistic forecasting, in which prediction intervals can be quantified explicitly [34], is identified as a direction for future work.

Compared to other studies [48], when LSTM was applied under different forecasted weather data conditions, where four performance metrics were employed, RMSE increased significantly from 996 kW to 3,216 kW, and NMAPE rose from 1.085% to 4.432%. Another study, which evaluated performance using only two metrics, NMAPE and RMSE, employed a white-box and combination model on two real-world datasets from Central Australia and Vietnam. The model achieved NMAPE and RMSE values of around 4% for real-time or intraday PV power prediction [50]. These comparisons highlight that a multi-metric evaluation framework is necessary to fully capture model behavior, especially when performance varies across temporal resolutions.

5.5. Forecasting performance across multiple temporal resolutions and model scalability

The ability of forecasting models to scale to coarser prediction intervals (such as one hour and one day) is a crucial requirement for practical applications, such as system planning, maintenance schedule [5], and it lays the foundation for the development of medium- and long-term prediction models. Based on the one-year dataset used in this study, the results show that while the BiGRU model outperforms others at shorter intervals, the BiLSTM model stands out at the 1-hour forecast window, achieving an MAE of 3,471 kW and a NMAPE of 7.01%. This transition in model superiority provides direct empirical evidence for the research question, confirming that temporal resolution and model architecture interact in a non-trivial way to determine forecasting performance. This indicates the flexibility in selecting models depending on operational requirements. However, all models exhibit a decline in performance as the forecast interval increases, especially under conditions of high-power output and significant fluctuations. This decline can be attributed to the combined effects of information loss, increased uncertainty, and reduced temporal resolution of key features. This presents a direction for future research to investigate the subjective and objective causes of these power output sudden fluctuations throughout the day, and to develop models capable of accurately predicting such volatile periods.

Moreover, it is important to emphasize that the number of published scientific studies on solar power forecasting in Vietnam remains limited. Most current solutions originate from commercial software, which does not publicly share source code or data [74], [75], creating challenges for system expansion and upgrades. Therefore, developing and comparing advanced predictive models in academic research is essential. On one hand, this helps evaluate the practical applicability of research models; on the other, it enables identification of strengths and weaknesses among models of the same family across multiple temporal resolutions, thereby paving the way for future work such as the development of hybrid models. More importantly, this study contributes to a deeper understanding of the mechanisms behind model performance variation, rather than merely reporting comparative results.

6. CONCLUSION

The growing penetration of solar photovoltaic energy presents significant challenges for grid operators, particularly in ensuring reliable, real-time power forecasting. This study evaluates and compares the performance of four advanced RNN architectures including GRU, BiGRU, LSTM, and BiLSTM, for short-term solar power forecasting at multiple temporally aggregated time resolutions: 5, 15, 30, and 60 minutes. Results demonstrate that all models perform reasonably well for short-term predictions, with BiGRU consistently outperforming the others across most forecasting intervals from 5 to 30 minutes. BiGRU achieved an R^2 of

approximately 0.94 and a NMAPE as low as 3.41% at the 5-minute interval.

While BiGRU maintained strong performance up to 30 minutes (NMAPE = 4.68%, $R^2 = 0.92$), its accuracy declined at the 60-minute resolution ($R^2 = 0.83$). In contrast, BiLSTM demonstrated better performance at this coarser resolution, achieving $R^2 = 0.85$ and NMAPE = 7.01%, indicating its ability to capture long-range temporal dependencies. The analysis also revealed that forecasting accuracy tended to deteriorate as power output increased, particularly near peak output or during periods of rapid fluctuation. In addition, a benchmark against a representative attention-based baseline at the 5-minute resolution showed that a vanilla Transformer performs on par with the bidirectional RNNs under the same training budget, with BiGRU retaining the lowest overall errors, which confirms that gated recurrent architectures remain competitive with attention-based models for short-horizon, single-plant solar power forecasting.

Despite these promising results, the study is limited by its reliance on data from a single location, which may affect the generalizability of results. Additionally, while the temporal resolution aggregation employed in this study improves forecasting complexity and reduces high-frequency noise, it may limit the ability of the forecasting models to represent short-lived power ramps within each aggregated interval. Moreover, model accuracy declined at peak output levels and over longer time intervals, indicating difficulty in capturing extreme variations. Future work should incorporate multi-location datasets to improve robustness and explore hybrid models, such as BiGRU with Transformers or reinforcement learning, to better handle volatility. Enhancing input preprocessing and integrating real-time or external meteorological data may also help reduce forecast errors.

FUNDING

This research received no external funding.

AUTHOR CONTRIBUTIONS

Ngoc-Trung Nguyen: Writing – original draft, Visualization, Software, Investigation. Thanh-Trung Nguyen: Writing – review & editing, Methodology, Formal analysis. Manh-Hai Pham: Writing – review & editing, Validation, Resources, Data curation. Nguyen Tuan Anh: Writing – review & editing, Validation, Data curation. Minh-Phap Vu: Writing – review & editing, Validation, Supervision, Methodology, Conceptualization..

DECLARATION OF COMPETING INTEREST

The author declares no conflicts of interest.

Declaration Of Generative AI And AI-Assisted Technologies

During the preparation of this work, the authors used ChatGPT and Claude in order to check spelling and improve the quality of the writing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

REFERENCES

- [1] M. A. Elhadidy, S. M. Shaahid, Parametric study of hybrid (wind + solar + diesel) power generating systems, *Renewable Energy*, 21:2 (2000) 129–139.
- [2] K. Wang, X. Qi, H. Liu, A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network, *Applied Energy*, 251 (2019) 113315.
- [3] K. Wang, X. Qi, H. Liu, A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network, *Applied Energy*, 251 (2019) 113315.
- [4] I. Citaristi, International energy agency-iea, in: *The Europa directory of international organizations 2022*, Routledge, 2022, pp. 701–702.
- [5] R. Asghar, F. R. Fulginei, M. Quercio, A. Mahrouch, Artificial Neural Networks for Photovoltaic Power Forecasting: A Review of Five Promising Models, *IEEE Access*, 12 (2024) 90461–90485.
- [6] N. T. P. Thanh, Vietnam's Solar Energy Market: A Comprehensive Outlook for Investors, 2024.
- [7] M. Ha-Duong, Vietnam's Power Development Plan 8 (PDP8): A bold step towards a net-zero future, *The Mekong Eye*, 2023.
- [8] C. Philibert, Solar energy perspectives, (No Title), 2011.
- [9] H. Wang, Y. Liu, B. Zhou, C. Li, G. Cao, N. Voropai, E. Barakhtenko, Taxonomy research of artificial intelligence for deterministic solar power forecasting, *Energy Conversion and Management*, 214 (2020) 112909.
- [10] R. Ahmed, V. Sreeram, Y. Mishra, M. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, *Renewable and sustainable energy reviews*, 124 (2020) 109792.
- [11] <https://doi.org/10.1016/j.rser.2020.109792>
- [12] B. Uzum, A. Onen, H. M. Hasanien, S. M. Mueen, Rooftop Solar PV Penetration Impacts on Distribution Network and Further Growth Factors–A Comprehensive Review, *Electronics*, 10:1 (2021) 55.
- [13] B. Kraas, M. Schroedter-Homscheidt, R. Madlener, Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the Spanish electricity market, *Solar Energy*, 93 (2013) 244–255.
- [14] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, *Solar Energy*, 136 (2016) 78–111.
- [15] B. Chen, P. Lin, Y. Lai, S. Cheng, Z. Chen, L. Wu, Very-short-term power prediction for PV power plants using a simple and effective RCC-LSTM model based on short term multivariate historical datasets, *Electronics*, 9:2 (2020) 289.
- [16] T. Huld, G. Friesen, A. Skoczek, R. P. Kenny, T. Sample, M. Field, E. D. Dunlop, A power-rating model for crystalline silicon PV modules, *Solar Energy Materials and Solar Cells*, 95:12 (2011) 3359–3369.
- [17] L. Visser, T. AlSkaif, W. van Sark, Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution, *Renewable Energy*, 183 (2022) 267–282.
- [18] Z. Zhen et al., Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image, *IEEE Transactions on Industry Applications*, 56:4 (2020) 3385–3396.
- [19] A. Nespoli, E. Ogliari, S. Leva, A. Massi Pavan, A. Mellit, V. Lughi, A. Dolara, Day-ahead photovoltaic forecasting: A comparison of the most effective techniques, *Energies*, 12:9 (2019) 1621.
- [20] S. Dutta et al., Load and renewable energy forecasting for a microgrid using persistence technique, *Energy Procedia*, 143 (2017) 617–622.
- [21] A. Agga, A. Abbou, M. Labbadi, Y. El Houm, Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models, *Renewable Energy*, 177 (2021) 101–112.

- [22] N. D. Tuyen, N. T. Thanh, V. X. S. Huu, G. Fujita, A combination of novel hybrid deep learning model and quantile regression for short-term deterministic and probabilistic PV maximum power forecasting, *IET Renewable Power Generation*, 17:4 (2023) 794–813.
- [23] M. O. Moreira, P. P. Balestrassi, A. P. Paiva, P. F. Ribeiro, B. D. Bonatto, Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting, *Renewable and Sustainable Energy Reviews*, 135 (2021) 110450.
- [24] S. Aslam, H. Herodotou, S. M. Mohsin, N. Javaid, N. Ashraf, S. Aslam, A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids, *Renewable and Sustainable Energy Reviews*, 144 (2021) 110992.
- [25] N. Bacanin, L. Jovanovic, M. Zivkovic, V. Kandasamy, M. Antonijevic, M. Deveci, I. Strumberger, Multivariate energy forecasting via metaheuristic tuned long-short term memory and gated recurrent unit neural networks, *Information Sciences*, 642 (2023) 119122.
- [26] A. Zameer, F. Jaffar, F. Shahid, M. Muneeb, R. Khan, R. Nasir, Short-term solar energy forecasting: Integrated computational intelligence of LSTMs and GRU, *Plos one*, 18:10 (2023) e0285410.
- [27] W.-C. Kuo, C.-H. Chen, S.-Y. Chen, C.-C. Wang, Deep learning neural networks for short-term PV Power Forecasting via Sky Image method, *Energies*, 15:13 (2022) 4779.
- [28] A. Mellit, A. M. Pavan, V. Lughi, Deep learning neural networks for short-term photovoltaic power forecasting, *Renewable Energy*, 172 (2021) 276–288.
- [29] S. Khan, T. Mazhar, T. Shahzad, W. Waheed, A. Waheed, M. M. Saeed, H. Hamam, Optimizing deep neural network architectures for renewable energy forecasting, *Discover Sustainability*, 5:1 (2024) 394.
- [30] V. Gundu, S. P. Simon, Short term solar power and temperature forecast using recurrent neural networks, *Neural processing letters*, 53:6 (2021) 4407–4418.
- [31] M. Elsaraiti, A. Merabet, Solar power forecasting using deep learning techniques, *IEEE access*, 10 (2022) 31692–31698.
- [32] Y. Xu, S. Zheng, Q. Zhu, K.-c. Wong, X. Wang, Q. Lin, A complementary fused method using GRU and XGBoost models for long-term solar energy hourly forecasting, *Expert Systems with Applications*, 254 (2024) 124286.
- [33] J. Kim, J. Obregon, H. Park, J.-Y. Jung, Multi-step photovoltaic power forecasting using transformer and recurrent neural networks, *Renewable and Sustainable Energy Reviews*, 200 (2024) 114479.
- [34] X. Ren, Y. Liu, F. Zhang, L. Li, A deep learning quantile regression photovoltaic power-forecasting method under a priori knowledge injection, *Energies*, 17:16 (2024) 4026.
- [35] D. Bui, Q. Nguyen, V. Doan, T. Pham, D. Le, Evaluating an Effectiveness of a Solar Power Plant Output Forecasting Model Based on LSTM Method Using Validation in Different Seasons of a Year in Vietnam, *GMSARN International Journal*, 18:1 (2024) 114.
- [36] N. Quang, L. Duy, B. Van, Q. Dinh, Applying Artificial Intelligence in Forecasting the Output of Industrial Solar Power Plant in Vietnam, *EAI Endorsed Transactions on Energy Web* (2021) 169166.
- [37] T.-A. Nguyen, M.-H. Pham, V. M. Phap, Q.-H. Do, N.-T. Nguyen, D.-T. Nguyen, T. N. Nguyen, Forecasting of solar power generation in Vietnam deploying a simple GRU model, in: *Proceedings of the 2023 Asia Meeting on Environment and Electrical Engineering (EEE-AM)*, IEEE, 2023, pp. 1–6.
- [38] S. Husein, E. Gago, B. Hasan, M. Pegalajar, Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies, *Heliyon*, 2024.
- [39] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: Beyond efficient transformer for long sequence time-series forecasting, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 35:12 (2021) 11106–11115.
- [40] J. Wang, W. Hu, L. Xuan, F. He, C. Zhong, G. Guo, TransPVP: A transformer-based method for ultra-short-term photovoltaic power forecasting, *Energies*, 17:17 (2024) 4426.
- [41] R. M. Rizk-Allah, L. M. Abouelmagd, A. Darwish, V. Snasel, A. E. Hassanien, Explainable AI and optimized solar power generation forecasting model based on environmental conditions, *PLOS ONE*, 19:10 (2024) e0308002.
- [42] T. H. B. Huy, D. N. Vo, H. D. Nguyen, H. P. Truong, K. T. Dang, K. H. Truong, Short-term load forecasting in power system using recurrent neural network, in: *Proceedings of the 2023 International Conference on System Science and Engineering (ICSSE)*, IEEE, 2023, pp. 67–71.
- [43] L. D. Bui, N. Nguyen Quang, E. R. Sanseverino, B. Doan Van, Forecasting Energy Output Solar Power Plant in Curtailment Condition Based on Lstm Using P/Ghi Factor and Validation in Training Process, a Case Study in Vietnam.
- [44] T. A. Nguyen, M. H. Pham, T. K. Duong, M. P. Vu, A Recent Invasion Wave of Deep Learning In Solar Power Forecasting Techniques Using Ann, in: *Proceedings of the 2021 IEEE International Future Energy Electronics Conference (IFEEEC)*, 16–19 November 2021, pp. 1–6.
- [45] M. Jaihani et al., A novel recurrent neural network approach in forecasting short term solar irradiance, *ISA Transactions*, 121 (2022) 63–74.
- [46] A.-N. Buturache, S. Stancu, Solar energy production forecast using standard recurrent neural networks, long short-term memory, and gated recurrent unit, *Engineering Economics*, 32:4 (2021) 313–324.
- [47] Y. Zhou, J. Wang, Z. Li, H. Lu, Short-term photovoltaic power forecasting based on signal decomposition and machine learning optimization, *Energy Conversion and Management*, 267 (2022) 115944.
- [48] N. Q. Nguyen, L. D. Bui, B. V. Doan, E. R. Sanseverino, D. D. Cara, Q. D. Nguyen, A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam, *Electric Power Systems Research*, 199 (2021) 107427.
- [49] R. C. Deo, M. Şahin, J. F. Adamowski, J. Mi, Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: A new approach, *Renewable and Sustainable Energy Reviews*, 104 (2019) 235–261.
- [50] T. Nguyen-Duc, H. Vu-Xuan-Son, H. Do-Dinh, N. Nguyen-Vu-Nhat, G. Fujita, S. Tran-Thanh, An adaptive method for real-time photovoltaic power forecasting utilizing mathematics and statistics: Case studies in Australia and Vietnam, *IET Renewable Power Generation*, 18:14 (2024) 2589–2604.
- [51] S. Abbaspour, F. Fotouhi, A. Sedaghatbaf, H. Fotouhi, M. Vahabi, M. Linden, A Comparative Analysis of Hybrid Deep Learning Models for Human Activity Recognition, *Sensors*, 20:19 (2020) 5707.
- [52] W. Fang, Y. Chen, Q. Xue, Survey on Research of RNN-Based Spatio-Temporal Sequence Prediction Algorithms, *Journal on Big Data*, 3:3 (2021) 97–110.
- [53] J. Xiao, Z. Zhou, Research Progress of RNN Language Model, in: *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 27–29 June 2020, pp. 1285–1288.
- [54] A. Shewalkar, Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU, *Journal of Artificial Intelligence and Soft Computing Research*, 9 (2019).
- [55] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, K.-W. Chau, Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting, *Water*, 12:5 (2020) 1500.
- [56] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, *arXiv preprint arXiv:1412.3555*, 2014.
- [57] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural computation*, 9:8 (1997) 1735–1780.
- [58] J. X. Chen, D. M. Jiang, Y. N. Zhang, A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification, *IEEE Access*, 7 (2019) 118530–118540. <https://doi.org/10.1109/ACCESS.2019.2936817>
- [59] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, J. Schmidhuber, A Novel Connectionist System for Unconstrained Handwriting Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:5 (2009) 855–868.
- [60] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, *arXiv preprint arXiv:1406.1078*, 2014.
- [61] I. D. Mienye, T. G. Swart, G. Obaido, Recurrent neural networks: A comprehensive review of architectures, variants, and applications, *Information*, 15:9 (2024) 517.
- [62] S. M. Vatanchi, H. Etemadfar, M. F. Maghrebi, R. Shad, A comparative study on forecasting of long-term daily streamflow using ANN, ANFIS, BiLSTM and CNN-GRU-LSTM, *Water Resources Management*, 37:12 (2023) 4769–4785.
- [63] R. Cahuantzi, X. Chen, S. Güttel, A comparison of LSTM and GRU networks for learning symbolic sequences, in: *Proceedings of the Science and Information Conference*, Springer, 2023, pp. 771–785.
- [64] C. Chen et al., Reconstruction of long-term strain data for structural health monitoring with a hybrid deep-learning and autoregressive model considering thermal effects, *Engineering Structures*, 285 (2023) 116063.

- [65] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:1409.0473, 2014.
- [66] F. M. Shiri, T. Perumal, N. Mustapha, R. Mohamed, M. A. B. Ahmadon, S. Yamaguchi, Recognition of Student Engagement and Affective States Using ConvNeXtLarge and Ensemble GRU in E-Learning, in: *Proceedings of the 2024 12th International Conference on Information and Education Technology (ICIET)*, 18–20 March 2024, pp. 30–34.
- [67] A. Graves, Generating sequences with recurrent neural networks, arXiv preprint arXiv:1308.0850, 2013.
- [68] D. Gaur, S. Kumar Dubey, Development of activity recognition model using LSTM-RNN deep learning algorithm, *Journal of Information and Organizational Sciences*, 46:2 (2022) 277–291.
- [69] S. Minaee, E. Azimi, A. Abdolrashidi, Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models, arXiv preprint arXiv:1904.04206, 2019.
- [70] F. M. Shiri, E. Ahmadi, M. Rezaee, T. Perumal, Detection of Student Engagement in E-Learning Environments Using EfficientnetV2-L Together with RNN-Based Models, *Journal of Artificial Intelligence* (2579–0021), 6 (2024).
- [71] T. H. H. Aldhyani, H. Alkahtani, A Bidirectional Long Short-Term Memory Model Algorithm for Predicting COVID-19 in Gulf Countries, *Life*, 11:11 (2021) 1118.
- [72] Q.-T. Phan, Y.-K. Wu, Q.-D. Phan, Application of a new Transformer-based model and XGBoost to improve one-day-ahead solar power forecasts, in: *Proceedings of the 2023 IEEE/IAS 59th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, IEEE, 2023, pp. 1–7.
- [73] L. Bui Duy et al., Refining Long Short-Term Memory Neural Network Input Parameters for Enhanced Solar Power Forecasting, *Energies*, 17:16 (2024) 4174.
- [74] H. N. Nguyen, Q. T. Tran, C. T. Ngo, D. D. Nguyen, V. Q. Tran, Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms, *PloS one*, 20:1 (2025) e0315955.
- [75] L. D. Bui, N. Q. Nguyen, B. Van Doan, E. R. Sanseverino, Forecasting energy output of a solar power plant in curtailment condition based on LSTM using P/GHI coefficient and validation in training process, a case study in Vietnam, *Electric Power Systems Research*, 213 (2022) 108706.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Solarlits and/or the editor(s). Solarlits and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.