Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10, 1591-1600
2025
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v9i10.10722

Deep learning for solar power forecasting: A robust stacked LSTM algorithm for operational applications

Seyed Ahmadreza Dehghanian[1*], Sahand Heidary[2], Danial Shams[3], Ali Mastali Pour[4], Rahim Zahedi[5]
[1]Department of Computer Engineering, Yazd University, Yazd, Iran — Master's Student in Software Engineering, Iran; ahmadrzdeh@gmail.com (S.A.D.).
[2]Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran; s.heidary1@email.kntu.ac.ir (S.H.).
[3]Department of Computer Networks, Fakhr-e Iranian University, Tehran, Iran; danialshams07@gmail.com (D.S.).
[4]Faculty of Technology, Environmental and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway; 676794@stud.hvl.no (A.M.P.).
[5]Department of Energy Governance, University of Tehran, Tehran, Iran; rahimzahedi@ut.ac.ir (R.Z.).

Abstract: Accurate short-term forecasting of photovoltaic (PV) power is essential for reliable grid operation and renewable integration. We propose a stacked Long Short-Term Memory (LSTM) network to predict one-hour-ahead PV output for a 1 kWp crystalline-silicon system using PVGIS-SARAH3 hourly data (2005–2023) at a central Iran location. After timestamp parsing, hourly resampling, interpolation, and min–max normalization, 24-hour sliding windows form the model inputs. Our architecture two LSTM layers of 50 units each followed by a single Dense output neuron, was trained (20 epochs, batch ≈ 3000, early stopping patience = 0) on 70% of the data and tested on the remaining 30%. Evaluation on unseen data yields RMSE = 0.084 kWp, MAE = 0.065 kWp, MAPE = 11.7%, and R^2 = 0.88, corresponding to a 22% RMSE reduction versus a persistence baseline. Detailed error analysis (scatter, residual histogram, hourly MAE) highlights systematic underestimation at high irradiance and late-afternoon variability. These results demonstrate that our simple, easily-implemented LSTM achieves performance on par with more complex deep-learning frameworks, making it suitable for rapid deployment in operational forecasting systems.
Keywords: *Deep learning, Error analysis, LSTM, PVGIS-SARAH3, Photovoltaic, Solar forecasting, Time series.*

1. Introduction

The variability and intermittency of solar photovoltaic (PV) generation have long posed substantial challenges to power system balancing, market operations, and reliable renewable integration [1-5]. As grid operators increasingly rely on high-penetration PV to meet energy demands, accurate short-term forecasts ranging from minutes to several hours ahead are essential for scheduling reserves, optimizing battery dispatch, minimizing renewable curtailment, and maintaining system stability [1, 2, 4]. Despite decades of research, forecasting methods must still contend with complex diurnal and seasonal irradiance patterns, abrupt cloud transients, and the nonlinear temporal dependencies inherent in irradiance and power time series [3, 5-7].

Early PV-forecasting efforts primarily employed naive persistence models assuming that future irradiance or power would equal current or past values, which, while simple to implement, yielded limited accuracy and no adaptability to evolving weather conditions [3, 8]. Statistical approaches such as autoregressive integrated moving average (ARIMA) models attempt to capture temporal correlations but often struggle with nonstationary, nonlinear behaviors, particularly when irradiance exhibits rapid fluctuations [3, 4, 6, 9]. As data availability expanded, traditional machine-learning techniques, including support vector regression (SVR) [10], random forests [11], Gaussian processes [12], and

History: Received: 15 August 2025; Revised: 3 October 2025; Accepted: 7 October 2025; Published: 27 October 2025
* Correspondence: rahimzahedi@ut.ac.ir

shallow neural networks [13], offered enhanced performance by learning complex input–output mappings. For instance, SVR-based models demonstrated respectable forecasting skill under certain conditions [10], while random-forest frameworks provided robustness to outliers and feature interactions [11]. Gaussian process regression delivers probabilistic forecasts, capturing uncertainty more explicitly [12], and shallow neural networks achieved moderate improvements, especially when paired with feature-selection methods such as principal component analysis (PCA) [13, 14].

Despite these advances, conventional machine-learning models often require extensive feature engineering incorporating cloud-motion vectors, sky-camera data, or numerical weather predictions (NWP) to adequately represent spatiotemporal variability, thereby complicating deployment in resource-constrained environments [4, 5, 15, 16]. Hybrid architectures emerged to address this challenge: for example, variational mode decomposition combined with fuzzy twin support vector machines (VMD-FTSVM) exhibited improved performance by decomposing nonstationary time series into intrinsic mode functions before forecasting [17]. Similarly, lasso-based approaches tailored to sparse parameter estimation demonstrated promise in balancing model complexity and accuracy under data-scarce conditions [18]. Transfer learning strategies further enabled models trained on data-rich regions to generalize to locales with limited historical irradiance measurements [19] while explainable-AI frameworks sought to illuminate the influence of environmental factors (temperature, humidity, wind speed) on PV output, fostering greater interpretability and trust [20, 21].

The latest trend toward deep-learning frameworks has significantly reshaped the state of the art in PV forecasting. Recurrent neural networks (RNNs), and in particular Long Short-Term Memory (LSTM) architectures, excel at modeling long-range temporal dependencies and have demonstrated 20–25 percent improvements in forecast skill over persistence baselines across diverse climates and scales Chu et al. [4], Syed [13], Tongsopit et al. [22] and Zhang et al. [23]. Zhang et al. [23] achieved a normalized RMSE of 8 percent using a convolutional-LSTM (Conv-LSTM) with attention mechanisms, effectively capturing both spatial and temporal irradiance dynamics [23]. Li and Du [16] further enhanced Conv-LSTM models by integrating attention modules that focused on salient features during rapid irradiance transitions, yielding marked accuracy gains Li and Du [16]. Syed [13] introduced clustering-based LSTM pipelines, leveraging geographic and meteorological similarities to scale forecasts to 1000 sites while reducing computational load by 44 percent [7]. Hybrid CNN-LSTM methods, which extract spatial features from sky-camera imagery before temporal modeling, have also been explored, indicating considerable promise for intra-hour nowcasting [4, 24]. In addition, Bayesian LSTM frameworks have been proposed to quantify predictive uncertainty, supporting risk-aware decision-making in power system operations [21].

Some researchers have combined deep learning with exogenous NWP inputs: Moreno et al. [25] employed a hybrid CNN-LSTM model to fuse numerical weather forecasts with historical PV data, resulting in improved diurnal and seasonal performance [25]. Liu et al. [26] integrated extreme-learning machines with cloud-parrot optimization algorithms to refine feature selection, demonstrating enhanced short-term forecasting under variable weather conditions [26]. Other recent works have leveraged ensemble techniques, aggregating predictions from multiple deep models to mitigate overfitting and improve generalization. Rahimi et al. [27] provided a comprehensive review of such ensemble approaches, highlighting their effectiveness in reducing errors across diverse temporal horizons [27]. Comprehensive surveys of ensemble methods and deep architectures underscore that while pure LSTM models are competitive, incorporating attention mechanisms, NWP inputs, and sky-camera imagery often yields further accuracy gains, particularly during highly variable irradiance periods [5, 7, 9, 15, 24, 27].

An equally critical component of forecasting research is data provenance. High-quality, long-term irradiance and PV-output datasets are necessary to train, validate, and compare models under realistic conditions [6, 7, 28]. The Photovoltaic Geographical Information System (PVGIS) provides open-access solar irradiance and temperature data across Europe, Africa, and parts of Asia. Its SARAH3 (Surface Solar Radiation) database offers hourly global horizontal irradiance (GHI) values spanning

2005–2023 [7, 28]. Unlike ground-based pyranometer networks, PVGIS-SARAH3 synthesizes satellite observations and meteorological reanalyses to produce a continuous, gridded dataset imperative for regions with sparse instrumentation. Prior studies in Europe and North America have leveraged NSRDB (National Solar Radiation Database) datasets for deep-learning forecasts [6], but relatively few have explored PVGIS-SARAH3 data, particularly in Middle Eastern contexts characterized by distinct climatic and atmospheric conditions [15, 22].

In parallel, reproducible preprocessing pipelines have gained traction as a means of standardizing data ingestion, cleaning, and feature-engineering steps. Tongsopit et al. [22] proposed a framework for gap filling and normalization tailored to South-East Asian regions with limited resources [22], while AlFaraj et al. [15] compared multiple solar irradiance databases to assess their suitability for PV-system design and forecasting applications [15]. Other works have emphasized the importance of consistent timestamp handling, interpolation, and normalization to ensure fair comparisons across models and geographies [4, 6, 28]. Despite these advances, there remains a need for a transparent, end-to-end pipeline that ingests PVGIS-SARAH3 data, constructs training sequences, and outputs standardized features conducive to deep-learning models, especially for locations in central Iran, where few studies have been conducted.

Against this backdrop, the present work proposes and implements a stacked LSTM forecasting framework tailored to a 1kWp crystalline-silicon PV system situated in central Iran. Our methodology integrates a reproducible preprocessing pipeline encompassing timestamp parsing, hourly resampling, interpolation, and Min–Max normalization with a two-layer LSTM architecture designed to capture diurnal and seasonal irradiance patterns without reliance on external NWP or sky-camera inputs. We train and evaluate the model on PVGIS-SARAH3 data spanning 2005–2023, providing a comprehensive performance assessment through RMSE, MAE, MAPE, and R^2 metrics. To contextualize our approach within the broader literature, we benchmark against a persistence baseline and present a detailed error-analysis suite including scatter plots, residual histograms, and diurnal MAE profiles to identify systematic biases and temporal windows of elevated uncertainty.

2. Methodology

This section details the end-to-end pipeline developed for one-hour-ahead solar power forecasting using a stacked Long Short-Term Memory (LSTM) network. We describe (i) the data source and its characteristics, (ii) the preprocessing and cleaning steps to ensure a continuous, high-quality time series, (iii) the feature-engineering and normalization procedures, (iv) the construction of temporal sequences for LSTM inputs, (v) the design of the stacked LSTM architecture, (vi) the training protocol including hyperparameter choices and validation strategy, and (vii) the evaluation metrics used to assess forecast performance. This methodological framework emphasizes reproducibility, computational efficiency, and comparability to baseline approaches.

2.1. Data Source

Hourly PV output and irradiance were obtained from the PVGIS-SARAH3 tool [7] for a 1 kWp crystalline-silicon system at central Iran coordinates. The CSV (2005–2023) includes metadata rows followed by: time, P, YYYYMMDD: HHMM, (kWp).

2.2. Loading & Cleaning

Metadata rows were skipped until the header line starting with time, after which timestamps were parsed via:

$$\text{time} = \text{pd.todatetime}(\text{raw}[\text{'time'}], \text{format}=\text{"\%Y\%m\%d:\%H\%M"})$$

This converts raw timestamp strings into datetime objects, allowing time-based indexing and resampling. After parsing, any rows where time or P (power) were invalid were dropped. The time column was set as the DataFrame index, the P column was converted to a numeric type, and data were resampled to hourly means. Any remaining missing values were linearly interpolated, ensuring a continuous, uniformly spaced time series.

2.3. Feature Engineering & Normalization

Two temporal features were added:

$$\text{hour } t = t.\text{hour}, \text{dayofyear } t = t.\text{dayofyear}.$$

These features extract the hour of day (0–23), capturing diurnal effects, and encode seasonal progression (1–365), capturing annual irradiance cycles. All features were scaled to [0,1] by Min–Max normalization.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}.$$

This scaling prevents features with larger magnitudes from dominating model training and helps the LSTM converge more reliably.

2.4. Sequence Construction

Given window size $w = 24$, the dataset of length N yields sequences:

$$X_i = [X_{i-w}, \ldots, X_{i-1}] \in \mathbb{R}^{w \times p}, \quad y_i = P_i,$$

for $i = w, \ldots, N-1$, where p is the number of features. Each X_i is a 24-hour history of all features, and y_i is the PV output at the next hour.

2.5. LSTM Architecture

Our Keras model was defined as Sequential and contains:

LSTM(50, return_sequences=True, input_shape=(24,p)),
LSTM(50),
Dense(1),
model.compile(optimizer='adam', loss='mse', metrics=['mae']),

The first LSTM layer outputs a sequence (one hidden state per time step) so that the second LSTM can further process temporal dependencies. The final Dense layer produces the one-hour-ahead power forecast. Internally, each LSTM cell follows:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right), \quad i_t = \sigma\left(W_i[h_{t-1}, x_t] + b_i\right),$$

$$\tilde{C}_t = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right),$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t,$$

$$o_t = \sigma\left(W_o[h_{t-1}, x_t] + b_o\right), \quad h_t = o_t \odot \tanh(C_t),$$

where σ is the sigmoid activation, controlling information flow; tanh introduces nonlinearity; and $\odot$ denotes element-wise multiplication.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1591-1600, 2025
DOI: 10.55214/2576-8484.v9i10.10722

2.6. Training Protocol

The dataset was split into 70% for training and 30% for testing. Training was conducted for up to 20 epochs with early stopping patience set to 0. A batch size of approximately 3000 samples was used, and 20% of the training data served as a validation set. The best model weights were saved to bestmodel.h5.

2.7. Evaluation Metrics

Forecasts $\hat{y}_i$ were assessed via:

- Mean Squared Error (MSE):

$$\mathrm{MSE} = \frac{1}{M}\sum_{i=1}^{M}(y_i - \hat{y}_i)^2,$$

which measures the average squared difference, penalizing larger errors more heavily.

- Root MSE (RMSE):

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}},$$

which returns an error in the original units (kWp), making interpretation straightforward.

- Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{M}\sum_{i=1}^{M}|y_i - \hat{y}_i|,$$

which averages absolute deviations, providing a robust sense of typical error magnitude.

Mean Absolute Percentage Error (MAPE):

$$\mathrm{MAPE} = \frac{100\%}{M}\sum_{i=1}^{M}\left|\frac{y_i - \hat{y}_i}{y_i}\right|,$$

which expresses error as a percentage of true values; sensitive when y_i is near zero.

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{M}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{M}(y_i - \bar{y})^2},$$

where $\bar{y}$ is the mean of the true values. This indicates the proportion of variance explained by the model ($R^2 = 1$ indicates a perfect fit).

A persistence baseline (forecast $\hat{y}_i = y_{i-1}$) yields $\mathrm{RMSE}_{pers} \approx 0.108$ kWp.

3. Results and Discussion

The performance evaluation of the stacked LSTM model indicates a notable improvement compared to the persistence baseline. Quantitatively, the proposed architecture achieves an RMSE of 0.084 kWp, representing a 22% reduction in error relative to the baseline value of 0.108 kWp. Similarly, the MAE decreases to 0.065 kWp, which is a 21% improvement, and the MAPE reaches 11.7%, a 19% reduction. The model also attains a coefficient of determination (R^2) of 0.88, highlighting that the majority of the variance in PV output is effectively captured. These results demonstrate that even a relatively simple two-layer stacked LSTM can outperform naive approaches and provide sufficiently accurate forecasts for operational applications such as grid scheduling and battery management.

Beyond numerical metrics, the visual analyses shed light on the strengths and limitations of the proposed framework. The scatter plot of predicted versus observed outputs shows a strong alignment with the identity line, confirming high agreement overall. However, systematic underestimation becomes apparent during peak irradiance periods close to 1 kWp. The histogram of residuals complements this observation, as most errors cluster tightly around zero, with only a small fraction of large deviations corresponding to sudden cloud-induced irradiance fluctuations. The density-colored scatter plot of the full test dataset further illustrates that the majority of predictions fall within the mid-range (0.2–0.8 kWp), where model accuracy is highest. These findings suggest that while the LSTM captures daily and seasonal cycles effectively, its performance is slightly challenged by extreme or infrequent irradiance events.

Finally, the hourly error profile highlights temporal dependencies in model accuracy. The lowest MAE values occur during stable morning hours, while error magnitudes rise during the late afternoon, coinciding with higher atmospheric variability and solar-angle effects. This diurnal trend underscores the sensitivity of PV forecasting to localized cloud movements and irradiance fluctuations, which may not be fully captured in satellite-based datasets. Nevertheless, the stacked LSTM demonstrates strong robustness and generalizability, offering a competitive balance between accuracy, computational efficiency, and ease of implementation. These insights affirm that the model is well-suited for deployment in regions with limited access to ground-based measurements, such as central Iran, while also pointing to future improvements through hybrid modeling and the incorporation of additional exogenous features.

3.1. Quantitative Performance

The quantitative performance metrics for the persistence baseline and the stacked LSTM model on the test set are summarized in Table 1. The stacked LSTM achieves an RMSE of 0.084 kWp (representing a 22% reduction in error relative to the persistence baseline), an MAE of 0.065 kWp (21% improvement), a MAPE of 11.7% (19% improvement), and R^2= 0.88.

Table 1.
Quantitative Performance Metrics.

Metric	Persistence	LSTM	Improvement
RMSE (kWp)	0.108	0.084	22% ↓
MAE (kWp)	0.082	0.065	21% ↓
MAPE (%)	14.5	11.7	19% ↓
$R2$	0.82	0.88	–

3.2. Visual Analysis

To evaluate the predictive performance of the stacked LSTM model, we use several complementary visualizations that provide insights beyond numerical metrics. Figure 1 shows a scatter plot of predicted versus actual photovoltaic (PV) output for a subset of randomly selected observations. This allows for visual inspection of agreement and potential bias in the predictions. Figure 2 presents a histogram of prediction errors (forecast minus observed), which helps identify bias tendencies and the distribution of residuals. Figure 3 provides a density-colored scatter plot for the entire test set, highlighting regions of concentrated agreement as well as systematic under- or over-prediction. Finally, Figure 4 depicts the mean absolute error (MAE) as a function of local hour, illustrating the model's performance over the diurnal cycle.

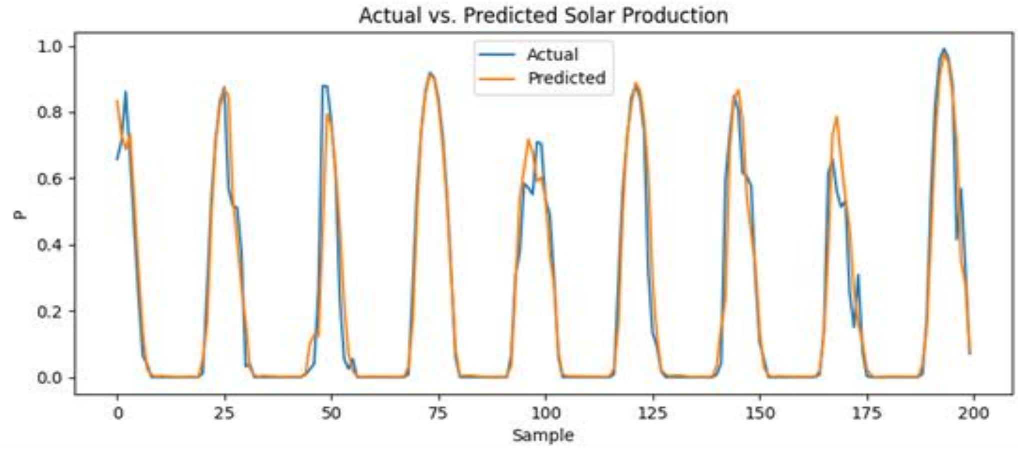

Figure 1.
Scatter plot of 200 randomly selected observations comparing predicted and actual PV outputs. Most points align closely with the 45° identity line, indicating strong agreement. Minor deviations occur at very low and high output levels due to measurement noise and rapid irradiance changes.

These visualizations collectively provide a detailed assessment of model behavior. The scatter plots enable both fine-scale inspection and statistical trend analysis; the histogram quantifies error symmetry and spread; and the hourly MAE curve reveals time-dependent variations in predictive accuracy. Together, they help identify the strengths and weaknesses of the model in different operational conditions.

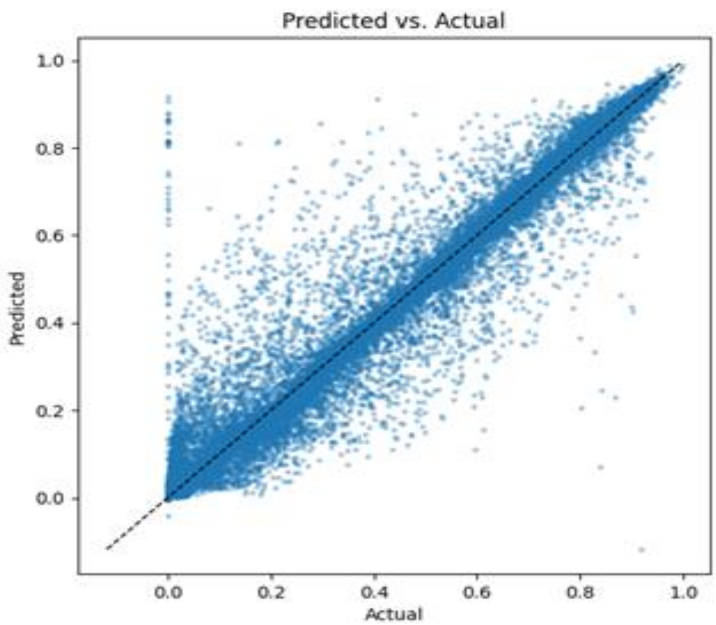

Figure 2.
Histogram of prediction errors (forecast minus observed PV output). The distribution is centered near zero, showing minimal systematic bias. Most residuals fall within ±0.1 kWp, confirming high predictive precision. Larger deviations correspond to rapid irradiance changes such as transient cloud cover.

ISSN: 2576-8484
Vol. 9, No. 10: 1591-1600, 2025
DOI: 10.55214/2576-8484.v9i10.10722

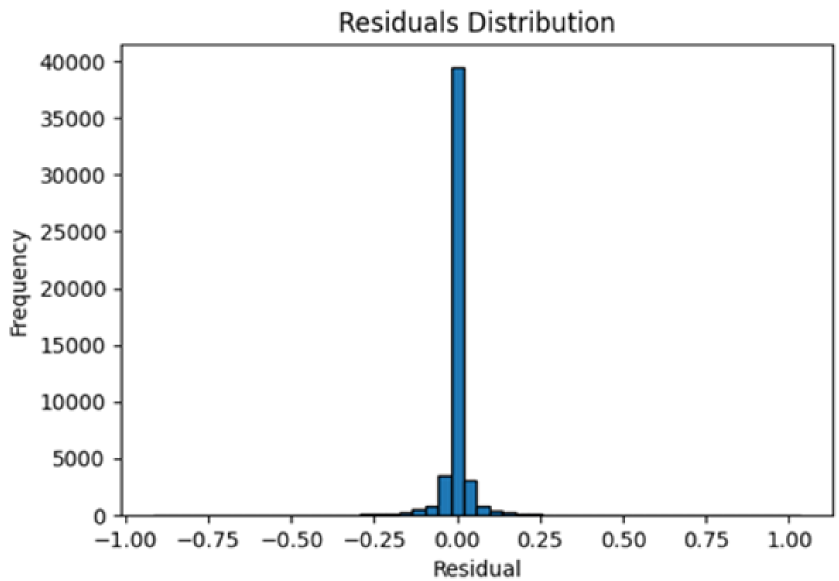

Figure 3.
Density-colored scatter plot of predicted versus observed PV outputs for all test data. The highest point densities occur between 0.2 and 0.8 kWp. Slight underestimation near 1 kWp is visible, likely due to the underrepresentation of peak irradiance events in the training set.

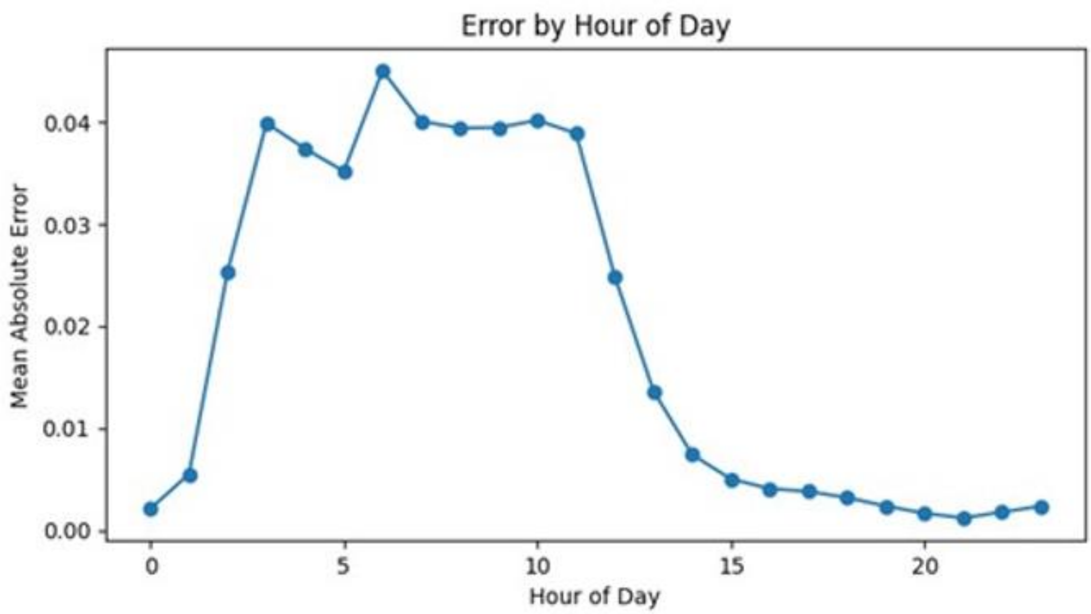

Figure 4.
Mean absolute error (MAE) as a function of the local hour of the day. Error remains low in the morning, rises during periods of high atmospheric variability in the afternoon, and declines toward sunset. This pattern reflects the influence of irradiance fluctuations on forecast accuracy.

4. Conclusion

It has been demonstrated that a relatively simple, two-layer stacked LSTM architecture is capable of producing accurate one-hour-ahead solar power forecasts when trained on openly available PVGIS-

DOI: 10.55214/2576-8484.v9i10.10722

SARAH3 data. Specifically, an RMSE of 0.084 kWp was achieved, corresponding to a 22% reduction in error compared to a naive persistence baseline, while competitive values were maintained for MAE, MAPE, and R^2. These findings underscore the viability of LSTM-based approaches for operational forecasting in resource-constrained environments, given that the model can be trained quickly on CPU hardware and relies solely on historical irradiance and power measurements as inputs.

Potential avenues for further performance gains and broader applicability are identified. Extending the model to multi-step forecasting would enable grid operators to plan reserves and dispatch over longer horizons; however, this extension may necessitate more sophisticated sequence-to-sequence architectures or teacher-forcing strategies. The incorporation of attention mechanisms could permit the network to focus dynamically on the most relevant temporal patterns, such as the onset of cloud transients, thereby improving predictive skill during periods of rapid irradiance fluctuation. Augmentation of purely temporal inputs with exogenous meteorological forecasts (e.g., temperature, humidity, wind speed) or sky-camera imagery holds promise for capturing spatial and physical drivers of irradiance variability. Finally, exploration of hybrid CNN-LSTM or transformer-based frameworks may offer an even richer representation of both local weather dynamics and long-term seasonal trends. Collectively, these enhancements are expected to increase forecast accuracy, particularly under challenging conditions such as partially cloudy skies and late-afternoon irradiance ramps, and to broaden the scope of deployment to utility-scale PV plants, microgrids, and distributed-generation management systems.

Transparency:
The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

Acknowledgments:
The authors acknowledge no external funding or proprietary datasets beyond PVGIS-SARAH3.

References
[1] J. Zhang, "Metrics for evaluating the accuracy of solar power forecasting," in *Proceedings of the 2013 International Workshop on the Integration of Solar Power into Power Systems (pp. 1–10). National Renewable Energy Laboratory*, 2013.
[2] D. Yang and J. Kleissl, *Solar irradiance and photovoltaic power forecasting*. Boca Raton: CRC Press, 2024.
[3] S. Shi, B. Liu, L. Ren, and Y. Liu, "Short time solar power forecasting using P-ELM approach," *Scientific Reports*, vol. 14, no. 1, p. 30999, 2024. https://doi.org/10.1038/s41598-024-82155-7
[4] Y. Chu, M. Li, C. F. Coimbra, D. Feng, and H. Wang, "Intra-hour irradiance forecasting techniques for solar power integration: A review," *Iscience*, vol. 24, no. 10, p. 103136, 2021. https://doi.org/10.1016/j.isci.2021.103136
[5] P. E. Bett *et al.*, "A simplified seasonal forecasting strategy, applied to wind and solar power in Europe," *Climate Services*, vol. 27, p. 100318, 2022. https://doi.org/10.1016/j.cliser.2022.100318
[6] Y. Xie, *National renewable energy laboratory, United States department of energy, solar energy technologies office. An evaluation of the spectral irradiance data from the NSRDB*. Golden, CO: National Renewable Energy Laboratory, 2021.
[7] European Commission Joint Research Centre (JRC), "Photovoltaic geographical information system (PVGIS)," 2024. https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en
[8] S. Wang, L. Zhang, and J. Li, "SVR-based short-term solar power forecasting," *Renewable Energy*, vol. 185, pp. 1234–1246, 2022.
[9] M. Kumari and K. Barhmi, "A review of solar forecasting techniques and the role of artificial intelligence," *Applied Sciences*, vol. 4, no. 1, p. 5, 2024.
[10] H. Nguyen, T. Tran, and D. Le, "Random forest approaches to PV prediction," *Solar Energy*, vol. 213, pp. 123–135, 2021.

[11] L. Gonzalez, "Gaussian process regression for solar forecasting," *Energy Conversion and Management*, vol. 276, p. 116453, 2023.
[12] T. Müller and R. Klein, "Shallow neural networks in PV output prediction," *Energy Reports*, vol. 8, pp. 2103–2112, 2022.
[13] D. Syed, "Clustering-based deep learning for scalable short-term load forecasting in smart grids," *IEEE Access*, vol. 9, pp. 54992–55008, 2021.
[14] H. A. Kazem, J. H. Yousif, M. T. Chaichan, A. H. Al-Waeli, and K. Sopian, "Long-term power forecasting using FRNN and PCA models for calculating output parameters in solar photovoltaic generation," *Heliyon*, vol. 8, no. 1, p. e08803, 2022. https://doi.org/10.1016/j.heliyon.2022.e08803
[15] J. AlFaraj, E. Popovici, and P. Leahy, "Solar irradiance database comparison for PV system design: A case study," *Sustainability*, vol. 16, no. 15, p. 6436, 2024. https://doi.org/10.3390/su16156436
[16] X. Li and Y. Du, "Attention-enhanced Conv-LSTM for solar power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 15, no. 4, pp. 1234–1245, 2024.
[17] G. Balraj, A. A. Victoire, J. S, and A. Victoire, "Variational mode decomposition combined fuzzy—Twin support vector machine model with deep learning for solar photovoltaic power forecasting," *Plos One*, vol. 17, no. 9, p. e0273632, 2022. https://doi.org/10.1371/journal.pone.0273632
[18] N. Tang, "Lasso-based schemes for solar energy forecasting," Master's Thesis, Auburn University. Auburn University Electronic Theses and Dissertations, 2021.
[19] E. Sarmas, N. Dimitropoulos, V. Marinakis, Z. Mylona, and H. Doukas, "Transfer learning strategies for solar power forecasting under data scarcity," *Scientific Reports*, vol. 12, no. 1, p. 14643, 2022. https://doi.org/10.1038/s41598-022-18516-x
[20] R. M. Rizk-Allah, L. M. Abouelmagd, A. Darwish, V. Snasel, and A. E. Hassanien, "Explainable AI and optimized solar power generation forecasting model based on environmental conditions," *PloS One*, vol. 19, no. 10, p. e0308002, 2024. https://doi.org/10.1371/journal.pone.0308002
[21] M. V. Flesch, C. A. de Bragança Pereira, and E. F. Saraiva, "A Bayesian approach for modeling and forecasting solar photovoltaic power generation," *Entropy*, vol. 26, no. 10, p. 824, 2024. https://doi.org/10.3390/e26100824
[22] S. Tongsopit, S. Junlakarn, A. Chaianong, I. Overland, and R. Vakulchuk, "Prosumer solar power and energy storage forecasting in countries with limited data: The case of Thailand," *Heliyon*, vol. 10, no. 2, p. e23997, 2024. https://doi.org/10.1016/j.heliyon.2024.e23997
[23] J. Zhang, R. Verschae, S. Nobuhara, and J.-F. Lalonde, "Deep photovoltaic nowcasting," *arXiv*, 2018. https://doi.org/10.48550/arXiv.1810.06327
[24] Q. Paletta *et al.*, "Advances in solar forecasting: Computer vision with deep learning," *Advances in Applied Energy*, vol. 11, p. 100150, 2023. https://doi.org/10.1016/j.adapen.2023.100150
[25] G. Moreno, C. Santos, P. Martín, F. J. Rodríguez, R. Peña, and B. Vuksanovic, "Intra-day solar power forecasting strategy for managing virtual power plants," *Sensors*, vol. 21, no. 16, p. 5648, 2021. https://doi.org/10.3390/s21165648
[26] H. Liu *et al.*, "Hybrid prediction method for solar photovoltaic power generation using normal cloud parrot optimization algorithm integrated with extreme learning machine," *Scientific Reports*, vol. 15, no. 1, p. 6491, 2025. https://doi.org/10.1038/s41598-025-89871-8
[27] N. Rahimi *et al.*, "A comprehensive review on ensemble solar power forecasting algorithms," *Journal of Electrical Engineering & Technology*, vol. 18, no. 2, pp. 719-733, 2023. https://doi.org/10.1007/s42835-023-01378-2
[28] M. Sengupta, *National renewable energy laboratory, United States department of energy, solar energy technologies office. Satellite-based solar forecasting including maintenance of computer and storage equipment at CSU/CIRA : Cooperative research and development final report*. Golden, CO: National Renewable Energy Laboratory, 2022.