



Attention-Based BiLSTM with Mutual Information Feature Selection for Improved Solar Radiation Prediction

Gautam Kumar¹, Yash Hooda^{1*}, Ajay Kumar¹, Aayush Kumar¹,
Sayan Khan¹, Shovon Dey¹

¹ University Institute of Engineering & Department of Computer Science, Chandigarh
University, Mohali, Punjab, India
yashhooda2017@gmail.com

Abstract. Solar radiation prediction is one of the important factors to be considered for the efficient usage of solar energy. Solar radiation is highly unpredictable because of the complex weather conditions and it is also highly variable. Therefore, the prediction of solar radiation is considered to be a challenging task. In this paper, an effective framework is presented for prediction of solar radiation. It is based on the concept of temporal feature engineering, feature selection using mutual information, and the usage of the Attention-based Bidirectional Long Short-Term Memory model combined with Mutual Information-based feature selection and lag-sequence temporal modelling. Lag based temporal features are used to efficiently predict the time varying data. Moreover, the mutual information feature selection method is used to select the most pertinent meteorological attributes. Unlike conventional LSTM-based forecasting methods, the proposed framework integrates nonlinear feature relevance analysis with temporal attention learning to improve forecasting robustness. In the proposed model, Performance is assessed using the publicly available solar radiation dataset and evaluated using the standard performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). From the experimental results, it is evident that the proposed model is able to achieve high accuracy with an RMSE of 0.0477 and R^2 of 0.9171. The results demonstrate that the proposed model is effective in modelling complex relationships in solar radiation data and shows capability for real world solar energy applications.

Keywords: Solar radiation prediction, Mutual information, Feature selection, BiLSTM, Attention mechanism, Deep learning

1 Introduction

The prediction of solar radiation plays an important role in efficient planning of solar energy systems and large-scale utilization of renewable energy resources. Solar radiation data, however, is highly non-stationary and non-linear in nature and depends

heavily on complex atmospheric and meteorological conditions, which make the prediction of solar radiation a complex and challenging task.

Latest innovations in deep learning (DL) have notably improved the performance of solar radiation forecasting models, which can automatically learn temporal features from historical data. Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) have shown promising results in foreseeing solar radiation compared to other statistical and machine learning (ML) models. The performance of these models can, however, be affected by redundant and less informative meteorological features. Therefore, efficient feature selection techniques must be used to select the most relevant input features. It has been observed that mutual information-based feature selection techniques perform better compared to other techniques, like correlation-based techniques, which consider only linear relationships between features and the target variable.

Moreover, solar radiation data exhibits strong temporal dependencies, which can be captured effectively by incorporating lag-based features. This will allow the model to learn patterns from past observations. Attention mechanisms have also shown promising results in deep learning models, which can be used to improve the performance of solar radiation forecasting models [22], [23], [24].

This paper proposes an improved solar radiation prediction framework, which incorporates efficient feature selection techniques, lag-based temporal feature engineering, and an Attention-based BiLSTM model. The performance of the proposed model has been estimated using standard performance metrics, like MAE, RMSE, and R^2 , which confirm the efficiency of the proposed model in capturing complex patterns in solar radiation data.

2 Related Work

Various techniques based on statistical, ML, and DL approaches have been extensively looked into for predicting solar radiation, as presented in [5]. Initially, statistical techniques were used, but these techniques were found to be less accurate in handling the extremely non-linear and nonstationary nature of solar radiation data. Recently, with the advent of high computational capabilities and readily available data, DL techniques have proved to be capable for handling solar radiation forecasting problems, as presented in [3], [8], and [9]. In recent times, Recurrent Neural Network (RNN) architectures, such as LSTM and BiLSTM, have proved their potential in handling long-term temporal dependencies in solar radiation data, as presented in [2] and [6].

Feature selection has also been identified as another key element that needs to be considered for the improvement of the performance of the models. The traditional methods of feature selection, such as Pearson and Spearman correlation, have been commonly used for selecting relevant meteorological features. However, these methods

are only able to handle linear relationships. The use of more sophisticated methods of feature selection, such as mutual information-based feature selection methods, is also able to handle non-linear relationships and is thus considered to be a more informative method of feature selection.

Moreover, temporal feature engineering methods have also been considered for the improvement of the ability of the models to handle time-dependent relationships in solar radiation data. The recent advances have also identified the significance of attention mechanisms for the improvement of the performance of deep learning models [11], [15], [19] – [27].

While considerable progress has been made in the forecasting of solar radiation with the help of deep learning techniques, existing studies are mostly concentrated on certain aspects such as feature selection methods or temporal modelling techniques. There is little work that effectively combines the best of feature selection techniques, temporal feature engineering techniques, and attention-based deep learning techniques.

In this paper, an effective approach is proposed that combines the best of mutual information-based feature selection techniques, temporal feature engineering techniques, and attention-based deep learning techniques. Recent advancements in solar radiation prediction using deep learning have shown significant improvements [3], [5], [11], [19–27].

3 Dataset Description

In order to maintain transparency, reproducibility, and reliability, this research utilizes a publicly accessible dataset for solar radiation [1]. The dataset is collected from the Kaggle Solar Radiation Prediction (HI-SEAS) repository. The repository includes actual meteorological conditions and corresponding solar radiation data.

The dataset includes several weather-related variables collected over a period of time, such as temperature, humidity, air pressure, and wind speed. The target variable is the solar radiation level, which is generally indicated by Global Horizontal Irradiance (GHI) levels. The variables are diverse and sufficient for an efficient feature selection and deep learning modelling process. The dataset has meteorological observations collected at regular temporal intervals, including temperature, humidity, pressure and wind speed.

Before training the model, several preprocessing activities are carried out to refine the data. The date and time variables are merged and normalized into a unified datetime format to maintain temporal order. Meanwhile, other time-based variables such as hour, day, and month are also extracted to include periodic patterns of solar radiation. To include temporal dependencies in the data, lag-based variables are created based on past radiation data. This will allow the model to learn from past trends. In addition, all

numerical variables are normalized using Min-Max normalization to ensure efficient training by preventing dominance of variables with greater scales.

The dataset is divided into the training and testing set in 80:20 data distribution. The training set is utilized to help and train the model, whereas the testing set is used to evaluate the execution of the model. This is to ensure the model is evaluated fairly. Lag features such as lag-2, lag-3, and lag-4 were generated using previous solar radiation observations to capture short-term and long-term temporal patterns.

The selected data set and the data preprocessing steps offer a strong platform for implementing the mutual information-based feature selection method and the proposed Attention- based BiLSTM model to achieve the accurate prediction of the solar radiation.



Fig. 1. Workflow of the proposed solar radiation prediction framework.

4 Mutual Information-Based Feature Selection

Feature selection is a crucial move in enhancing the performance of DL models. Feature selection helps remove irrelevant information and reduces redundancy in the data. The meteorological variables considered for predicting solar radiation have a complex and non-linear relationship with the target variable. The inclusion of all variables without

proper feature selection may cause a noisy problem and affect the generalization of the model negatively. In this research work, the feature selection method based on mutual information (MI) is considered for measuring the dependency of each feature and the target variable, Global Horizontal Irradiance (GHI). The mutual information method is considered for measuring the dependency of variables because it is proficient in measuring both linear and non-linear relationships between variables.

The mutual information between two variables X and Y is defined as:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) \left(\frac{p(x, y)}{p(x)p(y)} \right)$$

where $p(x, y)$ represents the joint probability distribution of X and Y , and $p(x)$, $p(y)$ denote their marginal probability distributions and $\log$ denotes logarithm base 2.

The parameters with high mutual information score values are considered more fitting as they provide more information about the target variable. Features with low mutual information score values are discarded to reduce the level of noise and enhance the efficiency of the model. Features that received higher MI scores were used for model training.

Table 1. Mutual Information Scores of Meteorological Features

Feature	MI Score	Selected
Temperature	0.81	Yes
Humidity	0.74	Yes
Pressure	0.42	Yes
Wind Speed	0.31	No
Wind Direction	0.28	No

In addition, the correlation heatmap is used for visualization purposes to understand the correlations between the meteorological variables and the solar radiation, as depicted in the final group of selected parameters are taken as input to the suggested Attention-based BiLSTM model, enabling the efficient learning of complex patterns for the prediction process.

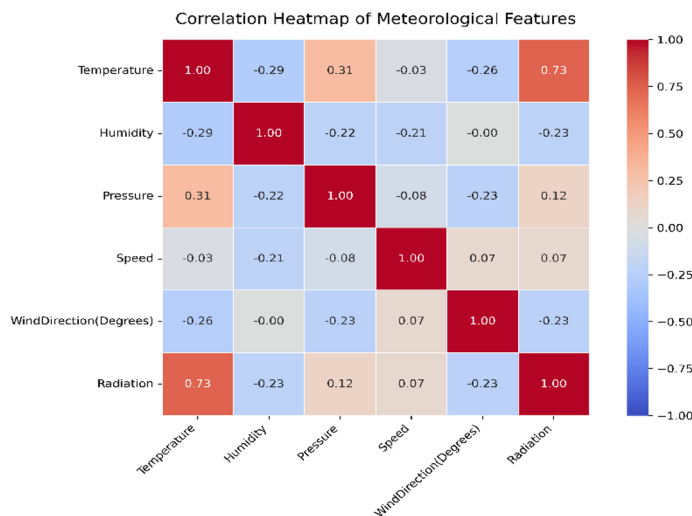


Fig. 2. Correlation heatmap of meteorological features and solar radiation (used for visualization only)

5 Temporal Feature Engineering

It is observed that the time series data related to solar radiation is associated with complex behaviours due to the changing atmospheric conditions. In order to overcome the complexities involved in modelling the temporal dependencies associated with the raw data, temporal feature engineering is used. In this study, instead of applying signal decomposition techniques, the lag-based feature engineering approach is used, where the past observations related to solar radiation are used as input features.

This is done specifically by creating multiple lag features by shifting the radiation values over previous time steps. Lag-2, lag-3, lag-4, and lag-7 temporal windows were taken in account for the purpose of extracting temporal dependencies. The lag features allow the model to learn short-term changes as well as long-term temporal dependencies in the data. The addition of context helps the model to comprehend data more effectively, improving prediction accuracy.

The features are then used in combination with some of the meteorological variables to create the final input representation. The method enhances the temporal learning capabilities of the DL model without affecting the computational complexity of the model.

On the other hand, the use of lag-based temporal feature engineering minimizes the effects of non-stationarity, hence improving the model’s capacity to learn significant patterns. Consequently, more precise solar radiation prediction is achieved.

6 Hybrid Model

Deep learning models are highly dependent on appropriate architecture design and hyperparameter tuning, such as the batch size, and number of epochs [2], [6]. Otherwise, they will be unable to achieve a stable training state or optimal performance. In this paper, hyperparameters will be selected empirically after multiple experimental trials, and the best configuration will be determined based on minimum validation error. Fig. 3 shows the proposed Attention-based BiLSTM architecture used for solar radiation prediction.



Fig. 3. Proposed Architecture for solar radiation prediction

Table 2. Hyperparameter Configuration of the Proposed Model

Parameter	Value
BiLSTM Units	64
Batch Size	32
Epochs	100
Optimizer	Adam
Learning Rate	0.001
Dropout	0.2
Loss Function	Mean Squared Error

Early stopping along with validation watch was applied to avoid overfitting. The model architecture is based on an "Attention-enhanced Bidirectional Long Short-Term Memory" architecture, which is effective in handling temporal dependencies and key patterns in solar radiation sequences [12], [22], [24], [27].

6.1 BiLSTM

BiLSTM networks are a further extension of traditional LSTM networks, where input sequences are processed in both forward and backward directions [2]. Such a two-directional processing of sequences helps in a more comprehensive understanding of time series.

For a given input sequence, a BiLSTM layer is composed of two parallel LSTM layers, where one-layer processes the sequence in sequential order, while the other layer processes it in reverse order. The results are then merged to give a unified representation, which helps in a better learning of complex temporal relationships in solar radiation data.

6.2 Attention Mechanism

Furthermore, to upgrade the capabilities of the model, an attention mechanism is also considered after the BiLSTM layer. This mechanism helps the model focus on different time steps of the input order by providing different weights to different temporal features.

Unlike the traditional approach of treating all time steps equally for the purpose of making predictions, the attention mechanism helps focus on important patterns that are significant for the predictions. This helps improve the accuracy of the predictions and also provides better interpretability of the complex and non-linear data such as solar radiation.

6.3 Proposed Attention-Based BiLSTM Architecture

In this proposed framework, the pre-processed input data, which is comprised of a number of selected meteorological features, is first processed using a Bidirectional LSTM layer, which is designed to effectively capture temporal dependencies in both directions.

The result of this BiLSTM layer is then sent through an attention mechanism, which is designed to effectively identify and weight the most informative time steps in the sequence.

The result of this attention-based weighting layer is then sent through another BiLSTM layer, which is designed to further enhance temporal feature extraction. Finally, a fully connected layer, which is designed to output a predicted value of solar radiation, is used.

The proposed framework is capable of effectively capturing both short-term and long-term dependencies in solar radiation statistics, which enables it to significantly improve its prediction capabilities.

7 Hyperparameter Optimization

Deep learning models are highly sensitive to hyperparameters such as the batch size, and the number of training epochs. Therefore, if these parameters are not set appropriately, it can affect the performance of the model.

In this study, the hyperparameters are tuned using empirical methods. A split of 20% was used during training for hyperparameter tuning. The early stopping patience was selected empirically from the validation loss which stabilizes after around 100 epochs. The effectiveness of the proposed model is tested with different configurations. The parameters such as the batch size, sequence length, and the quantity of epochs are adjusted to obtain the best performance. In this study, the best model is selected based on the accomplishment of the model on the validation set. Therefore, Model is trained to obtain minimum loss on the validation set. Early stopping is also applied to ensure the model is trained efficiently. This model is both effective and efficient in the task of forecasting the solar radiation as it achieved stable convergence and forecasting performance.

8 Experimental Setup

This section provides an overview of the experimental setup for testing the proposed solar radiation prediction framework. Chronological splitting was used instead of random splitting to protect the sequential nature of time-series data. The experiments are carried out on a pre-processed dataset that includes mutual information-based feature selection and lag-based temporal feature engineering.

8.1 Data Division

The dataset is divided into 2 sets based on an 80:20 ratio. The First set is utilized for training and validating the model, and the remaining data is held for testing the performance of the trained model on new data points. This is a typical method for ensuring a fair and accurate assessment of the model's generalization capacity.

8.2 Model Setup

The proposed Attention-based BiLSTM model is developed and trained using the TensorFlow/Keras platform. The input to the model is the selected meteorological features and lag-based temporal features. The selected meteorological variables and lag-based temporal features were given as sequential inputs into the network.

A Bidirectional LSTM layer is integrated into the model for extracting temporal dependencies in the data in both directions. The extracted feature representation is processed by an attention mechanism for focusing on the paramount time points in the sequence. The final predicted value of solar radiation is obtained by processing the

feature representation through a dense layer. The hyperparameters of the model are empirically set based on the validation performance.

8.3 Evaluation Measures

The quality of prediction of the suggested framework is evaluated by the following measures of a regression:

- RMSE
- R^2
- MAE

Suggested model has an R^2 of 0.9171, which depicts a high accuracy and the ability to capture complex patterns in solar radiation.

9 Result and Discussion

This section discusses the performance of the proposed solar radiation prediction framework, along with an evaluation of the impact of feature selection, temporal feature engineering, and the Attention-based BiLSTM model on the prediction accuracy.

9.1 Model Performance

The performance of the proposed model is quite strong, as indicated by low error values and high accuracy. The low RMSE and high R^2 score show the capability of the proposed framework to apprehend nonlinear temporal patterns effectively. The results obtained using the model are as follows:

- MAE: 0.02597
- RMSE: 0.04776
- R^2 : 0.9171

These results indicate that the model has performed well in understanding complex and nonlinear relationships in solar radiation data. The R^2 value indicates an excellent level of agreement between forecasted and real values, thereby validating the effectiveness of the model.

Table 3. Performance Metrics of the Proposed Attention-Based BiLSTM Model

Model	MAE	RMSE	R ²
Proposed Attention-BiLSTM	0.02597	0.04776	0.9171

9.2 Contrast with Existing Models

To further assess the effectiveness of the suggested model, a comparison is made with other traditional models such as LSTM and BiLSTM. The comparison is based on the R² score, which indicates the accuracy of the predictions made by the models [11], [15], [20], [21]. The comparison is provided in Table 4.

Table 4. Comparison of Prediction Accuracy with Existing Models

Model	R ²
Linear Regression	0.71
SVR	0.79
Random Forest	0.84
LSTM	0.85
BiLSTM	0.88
Proposed Attention-BiLSTM	0.9171

The results clearly show that the proposed Attention-based BiLSTM outperforms the traditional models. This is due to the use of mutual information-based feature selection and lag-based temporal feature engineering.

9.3 Ablation Study

Table 5. Ablation Study of the Proposed Framework

Configuration	R ²
BiLSTM Only	0.88
BiLSTM + Attention	0.90
BiLSTM + MI Feature Selection	0.89
Proposed Full Model	0.9171

Ablation study results verify that all components have positive impact on forecasting performance improvement.

9.4 Impact of Feature Selection and Temporal Modelling

The mutual information-based feature selection method has a significant impact on the increasing the effectiveness of the model, where the most important meteorological features can be selected, and redundant information can be removed. Unlike other feature selection methods, mutual information can handle both linear and nonlinear relationships, making it a more informative feature selection method.

Furthermore, lag-based temporal feature engineering enables the model to learn time dependent patterns using historical values of radiation. This improves the model's ability to apprehend short as well as long-term trends in solar radiation data. These methods improve data representation, thereby contributing to an increase in model accuracy.

9.5 Prediction Analysis

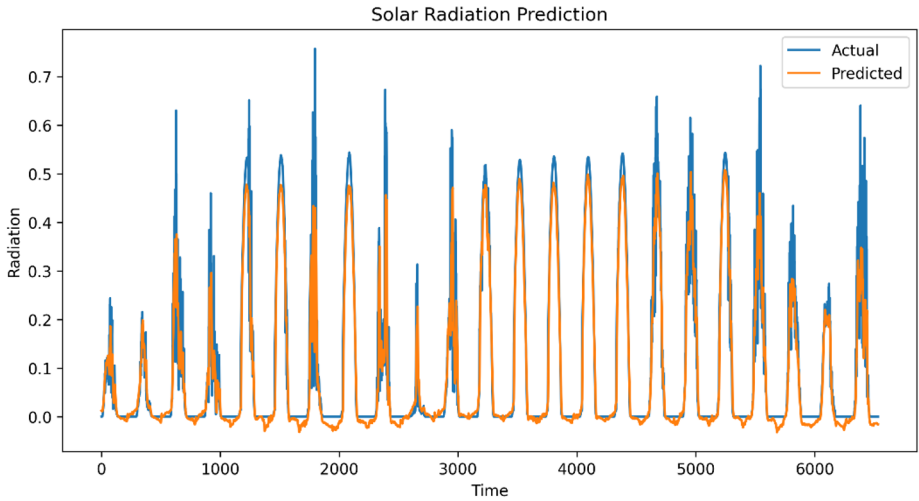


Fig. 4. Comparison between actual and predicted solar radiation values

The comparison between the forecasted and real solar radiation values is depicted in the Fig. 4. The predicted curve follows the actual observations very closely, thus showing that the model has captured the short-term changes as well as the long-term time dependence. The prediction deviations were greater during sudden changes in radiation and cloudy weather because of the sudden changes of the meteorological variables.

9.6 Discussion

The results verify that the combination of mutual information-based feature selection, lag-based temporal feature engineering, and the Attention-based BiLSTM model can greatly improve the performance of solar radiation prediction. All these techniques play a role in improving the effectiveness of the solar radiation prediction. First, feature selection plays a role in reducing redundancy in the features. Then, there is the use of temporal features in improving performance. Lastly, there is the use of the attention mechanism in improving performance.

The use of the framework in solar radiation prediction has better accuracy and stability compared to the use of other deep learning techniques. The evaluation of the dataset has shown that the use of the framework in solar radiation prediction is promising and effective in the conditions of the obtained data set.

10 Conclusions and Future Work

This paper proposes an effective framework for solar radiation prediction by utilizing mutual information-based feature selection, lag-based feature engineering, and the Attention based BiLSTM model. New framework addresses the challenges associated with solar radiation prediction.

The use of mutual information-based feature selection allows for the selection of the most relevant meteorological features. The addition of lag-based features allows for the consideration of the time dependence of solar radiation. The use of the Attention-based BiLSTM model allows for consideration of the most significant time steps in the data. The experiment results prove that the new framework can achieve high accuracy in solar radiation prediction. The RMSE of the model is 0.04776, and the R^2 score is 0.9171. The experiment results prove the efficiency of the new framework in modelling the complex solar radiation data. The new framework can be further enhanced in the future by incorporating more efficient optimization algorithms such as Particle Swarm Optimization or Genetic Algorithm for hyperparameter tuning. The new framework can also be further enhanced by considering more efficient deep learning models such as the Transformer model [22]–[27]. The new framework can also be further enhanced by considering more diverse data [19]–[21].

References

1. Kaggle: Solar Radiation Prediction Dataset (HI-SEAS). <https://www.kaggle.com/datasets/dronio/SolarEnergy> (2020)
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Computation* **9**(8), 1735–1780 (1997)
3. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
4. Voyant, C., et al.: Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* **105**, 569–582 (2017)
5. Yang, Z., et al.: Deep learning approaches for solar radiation prediction. *Energy Conversion and Management* **208**, 112125 (2020)
6. Rajagukguk, R., Ramadhan, R., Lee, H.: Deep learning models for solar irradiance forecasting: A review. *Energies* **13**(24), 6623 (2020)
7. Kumari, P., Toshniwal, R.: Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production* **318**, 128566 (2021)
8. Mellit, A., Kalogirou, S.: Artificial intelligence techniques for photovoltaic applications. *Progress in Energy and Combustion Science* **34**(5), 574–632 (2008)
9. Salcedo-Sanz, S., et al.: Prediction of daily global solar radiation using feature selection and machine learning. *Energy Conversion and Management* **50**(12), 2935–2945 (2009)
10. Abdulrahman, M., et al.: Solar irradiance forecasting using ensemble machine learning methods. *Renewable and Sustainable Energy Reviews* **135**, 110220 (2021)
11. Liu, J., et al.: Short-term solar radiation forecasting using attention-based deep learning models. *Applied Soft Computing* **103**, 107135 (2021)
12. Vaswani, A., et al.: Attention Is All You Need. In: *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008 (2017)

13. Huang, L., et al.: Solar radiation prediction using CNN–LSTM deep learning model. *Applied Energy* **246**, 66–77 (2019)
14. Zhang, J., et al.: Short-term solar irradiance forecasting based on CNN and LSTM. *IEEE Access* **7**, 110145–110154 (2019)
15. Wang, Y., et al.: CEEMDAN-based hybrid deep learning model for solar radiation forecasting. *Energy* **215**, 119120 (2021)
16. Li, C., et al.: Hour-ahead solar irradiance prediction using BiLSTM and CNN–BiLSTM. *International Journal of Information Systems Science* **30**(7), 711–724.
17. Brownlee, J.: Feature selection techniques for machine learning. *Machine Learning Algorithms*, 1–15 (2018)
18. Zhang, X., et al.: Multi-scale solar radiation prediction using signal decomposition and LSTM networks. *Energy* **174**, 649–659 (2019)
19. Anwar, M.T., et al.: Forecasting meteorological solar irradiation using machine learning. *Proceedings of the ACM Conference* (2023)
20. Allal, Z., et al.: Machine learning algorithms for solar irradiance prediction. *Results in Engineering* (2024)
21. Tandon, A., et al.: Machine learning-driven solar irradiance prediction. Springer Nature (2025)
22. Naveed, M.S., et al.: Leveraging advanced AI algorithms with transformer-infused recurrent neural networks to optimize solar irradiance forecasting. *Frontiers in Energy Research* (2024)
23. Alorf, A., et al.: Solar Irradiance Forecasting Using Temporal Fusion Transformer Models. *Energy Reports* (2025)
24. Chiranjeevi, M., et al.: Solar Irradiance Forecasting Performance Enhancement Using Hybrid Fuzzy Based CNN–BiLSTM–Transformer Model. *IEEE Access* (2025)
25. Niu, Y., et al.: Solar Multimodal Transformer: Intraday Solar Irradiance Predictor Using Public Cameras and Time Series. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2025)
26. Kibet, D., et al.: Minimalist Deep Learning for Solar Power Forecasting. *Energies* (2025)
27. Basu, A., et al.: SolarTformer: A Transformer Based Deep Learning Approach for Short-Term Solar Power Forecasting. *arXiv preprint* (2026)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

