Mathematical Modelling and Control

http://www.aimspress.com/journal/mmc

MMC, 6(2): 205–215.
DOI: 10.3934/mmc.2026016
Received: 15 July 2024
Revised: 06 September 2024
Accepted: 27 November 2024
Published: 12 June 2026

*Research article*

# BiLSTM networks for solar radiation forecasting in greenhouse environments

**Divyadharshini Venkateswaran[1], Yongyun Cho[1,*], Jinyoung Kim[1], Myeongbae Lee[1], Wonjun Hwang[1] and Sooho Jung[2]**

[1] Department of Information and Communication Engineering, Sunchon National University, Suncheon, Republic of Korea

[2] Jeollanamdo Agricultural Research and Extension Services, Naju-si, Republic of Korea

* **Correspondence:** Email: yycho@scnu.ac.kr.

**Abstract:** This paper investigated an innovative prediction approach using the bidirectional long short-term memory (BiLSTM) model for the forecasting of solar radiation in the smart farm in Naju, South Jeolla Province. Accurate solar radiation prediction holds substantial significance in facilitating efficient solar power generation, significantly impacting agricultural sustainability and resource utilization. This study rigorously assessed the effectiveness of various predictive models for forecasting solar radiation, a crucial factor in optimizing solar energy utilization in greenhouse settings. The variety of statistical and machine learning models aims to improve the accuracy of solar radiation prediction, which is crucial for precise energy management and cultivation practices. An hour-ahead mean absolute error (MAE) of 13.53 demonstrated the proposed BiLSTM model's impressive forecasting capabilities. The comparative analysis among individual models then revealed that the LSTM model achieves an MAE of 15.23, extreme gradient boosting (XGBoost) at 24.64, and autoregressive integrated moving average (ARIMA) at 28.52. This exploration underscored the important role of accurate solar radiation forecasting, specifically highlighting the effectiveness of the BiLSTM model in optimizing solar power generation within greenhouse environments. Its importance extended to the promotion of sustainable agricultural practices and the improvement of resource-efficient energy control strategies.

**Keywords:** solar radiation; bidirectional LSTM; long-short-term memory; energy optimization; solar photovoltaic data; greenhouse

## 1. Introduction

The World Energy Outlook (WEO) 2017 highlights a projected surge in global energy demand due to the expanding world population. By 2040, it is expected to increase by an additional 30% from 2016 levels, with a significant portion attributed to the anticipated rise in consumer electricity usage. Furthermore, the WEO predicts that around two-thirds of future global investments in new power plants will be directed toward renewable energy sources by 2040, with solar energy emerging as the primary contender (see reference [1]). Many countries are currently investigating the integration of renewable energy sources into their power systems, which presents additional problems due to the fluctuation of resources such as solar radiation and wind. Therefore, effective prediction of solar radiation, especially during periods of high-energy integration (see [2]), is crucial. One of the greatest challenges for the near future of global energy supply will be the extensive integration of renewable energy sources, particularly unpredictable ones such as wind and solar, into existing or future energy supply structures. Electrical operators must ensure a precise balance between electricity production and consumption at all times to ensure seamless electrical network operation (see Figure 1). However, the agricultural industry has experienced a notable shift toward controlled environment agriculture (CEA) practices, with greenhouse cultivation emerging as a prominent method to enhance crop growth and yield. Greenhouses offer growers the ability to regulate environmental parameters such as temperature, humidity, and light exposure, creating optimal conditions for plant growth. Among these environmental factors, solar radiation plays a pivotal role in influencing plant photosynthesis, transpiration, and general physiological processes. These compelling reasons underscore the critical role of forecasting in both the energy production and consumption domains. Consequently, the necessity for effective forecasting models becomes apparent. Solar photovoltaic projects, which use multiple solar panels to convert sunlight directly into electricity, highlight the critical need for precise forecasting approaches to optimize energy use and ensure grid stability (see [3]).

An energy system is being investigated in this smart farm environment for its potential to combine a backup wood pellet boiler with solar photovoltaic (PV)-powered heat pumps, therefore guaranteeing the total system's stability. In both winter and summer, this system maintains the air quality of the farm to its ideal. The main power source, a grid-connected solar PV system, allows heat pumps, water circulation circuits, and fan coil units (FCUs) to run continuously throughout the year. Without relying on heat pumps, PV panels run continuously, with the grid balancing any excess or shortage of solar power. During the winter months, if solar energy is insufficient to meet the heat demand, the wood pellet boiler supplements the heating load. In contrast, during periods of no heating or cooling demand, surplus solar energy is exported to the grid, contributing to overall energy sustainability. The authors recently conducted research on the efficacy of an integrated renewable energy framework, consisting of PV solar power plants, air source heat pumps (ASHPs), and a wood pellet boiler, in meeting the power requirements of smart farms located in South Korea. [4–6]. This research serves as foundational motivation for the current study, with the objective of further optimizing energy efficiency and sustainability in agricultural operations. Recently, Figures 2 and 3 illustrated the design and diversity of the farm in South Korea. This greenhouse grows a diverse variety of crops, including bell peppers, tomatoes, apples, and mangoes, all of which thrive under carefully controlled conditions. Researchers have focused on the comprehensive technological and ecological assessment of hybrid renewable energy systems for smart agriculture in the Republic of Korea [4].

Time series analysis and forecasting have traditionally relied on autoregressive integrated moving average (ARIMA) and its variants, such as seasonal ARIMA (SARIMA) and ARIMA with Explanatory Variables (ARIMAX) [7]. Several time series issues have been modeled using these techniques [8–10], but new ways of evaluating time series data have emerged, based on deep learning (DL) and machine learning (ML). Various forecasting models, such as random forests, gradient-boosted trees, and deep learning, were used to model the Standard & Poor's 500 (S&P 500) Index in [11–13]. The authors acknowledged the significant difficulty of training DL algorithms, including neural networks [11, 14]. According to Lee [15], a method for forecasting stock returns was presented using recurrent neural networks (RNNs). The method involved building portfolios by modifying the RNN's internal layers to vary the threshold levels of returns.

The most appropriate research that contrasts the capabilities of bidirectional long-short-term memory (BiLSTM) is [16, 17]. Kim and Moon found that the unidirectional LSTM performs worse than the BiLSTM model, which is based on multivariate time series data [16]. Bidirectional and unidirectional LSTM networks were combined into a stacked architecture in [17], which showed better performance than unidirectional and BiLSTM networks in predicting network traffic speed. This article expands on earlier research [18,19] by comparing three-time series modeling standards, ARIMA, LSTM, and BiLSTM. The comparison was made to predict economic and financial time series and parameter tuning. Although bidirectional training has shown promise for classic prediction tasks such as building schedulers [20] and anticipating software vulnerabilities [21], its effectiveness in learning time series data, particularly financial and economic data, is still up for debate. In order to fill this research gap, the following questions will be investigated in this paper: (i) Do learning time-series data in both directions—from the past to the

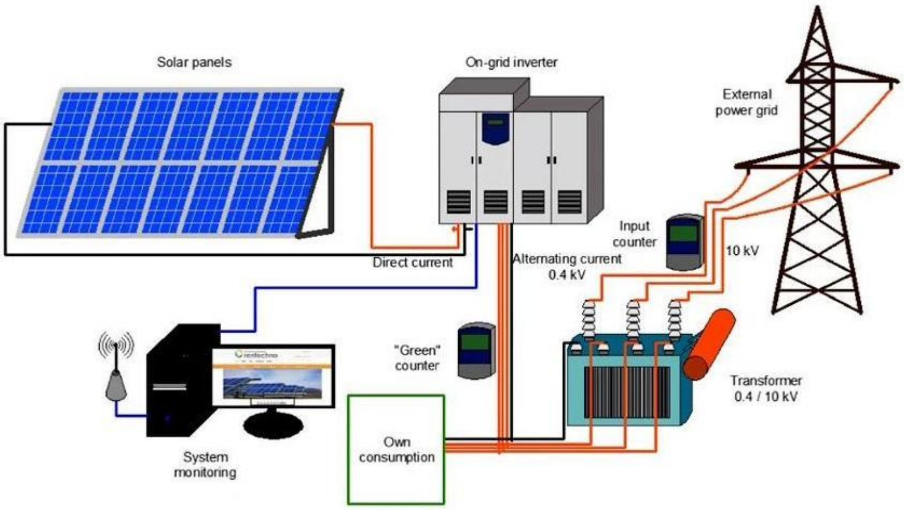
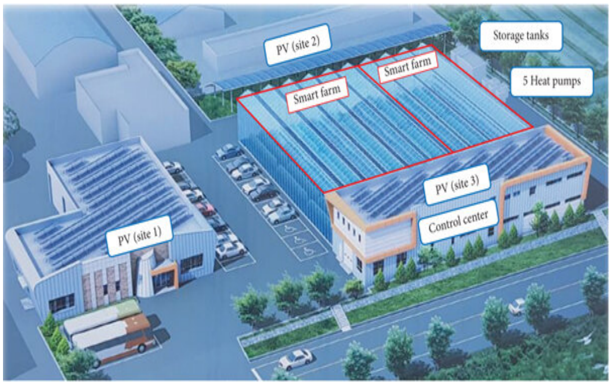



**Figure 1.** Schematic diagram of a solar power plant [22].

future and from the future to the past—improve prediction accuracy? (ii) How is input data treated differently in LSTM and BiLSTM architectures? Therefore, the main goals of our current work are to (1) find out if adding more training layers to the LSTM design enhances prediction performance; and (2) research how the LSTM and BiLSTM architectures handle input data differently.

Based on the preceding ideas, we conduct an analysis of solar radiation forecasting in a smart farm. Our research is conducted in the Jeollanamdo Agricultural Research and Extension Services, 1508 Senam-ro, Sanpo-myeon, Naju-si, Jeollanam-do, Republic of Korea. In addition, this study rigorously assesses the effectiveness of various predictive models in forecasting solar radiation, a crucial factor in optimizing solar energy utilization in greenhouse settings. The range of statistical and machine learning (ML) models aims to improve the accuracy of solar radiation prediction, which is vital for precise energy management and cultivation practices. In particular, the proposed BiLSTM model achieves an hour-ahead mean absolute error (MAE) of 13.53, showcasing its impressive forecasting capabilities. Further comparative analysis among individual models highlights the superiority of the BiLSTM model. Finally, simulation results are visualized to offer insights into the effectiveness of each model.




**(a)**

**(b)**

**Figure 2.** Visualization of the smart farm in Naju, South Jeolla province: (a) schematic representation, (b) photograph of the operational farm (see reference [4]).


**(a)**


**(b)**

**Figure 3.** Solar-powered energy system for a smart farm in Naju, South Jeolla province: (a) paprika and tomato, and (b) apple and mango (see reference [4]).

## 2. Methodology

### 2.1. ARIMA

ARIMA models are used to examine temporal patterns in time series data by using a linear regression-based forecasting technique. Moving average (MA) and autoregressive (AR) components are combined in these models. The general term ARIMA$(p, d, q)$ characterizes this class of ARIMA models. It consists of three terms: MA, integrated (or differencing), and AR terms. The corresponding orders of these terms are $p, d$, and $q$. The differencing term converts a nonstationary series into a stationary one, whereas the AR and MA terms depend on how strongly the time series correlates over time [23]. According to [24], the ARIMA $(p, d, q)$ model for a time series is represented as follows:

$$y'_t = + \Upsilon_1 y'_{t-1} + \cdots + \Upsilon_p y'_{t-p} + \widehat{\theta}_1 \widehat{\varepsilon}_{t-1} + \cdots + \widehat{\theta}_q \varepsilon_{t-q} + \widehat{\varepsilon}_t,$$

where $\Upsilon_1 y'_{t-1} + \cdots + \Upsilon_p y'_{t-p}$ is the AR term with $\Upsilon_1$ to $\Upsilon_p$ as coefficients at $p$ order, $\theta_1 \varepsilon_{t-1} + \cdots + \widehat{\theta}_q \varepsilon_{t-q}$ is the MA term with $\widehat{\theta}_1$ to $\widehat{\theta}_q$ as coefficients at $q$ order, $\widehat{\varepsilon}_t$ is an error term for random background noise at time $t$, and $y'_t$ is the differentiating series. For first-order differentiation:

$$y'_t = y_t - y_{t-1} \quad \text{(first-order differencing)},$$

where $y_t$ is the observation at time $t$.

### 2.2. XGBoost

A popular machine learning technique for problems such as regression and classification, extreme gradient boosting (XGBoost) is based on the gradient boosting decision tree (GBDT) method [25, 26]. It incorporates an updated version of the GBDT approach that is made to prevent overfitting and strengthen resilience. Using weak learners (usually classification and regression tree (CART)) and applying gradient descent to enhance model performance, XGBoost, which operates as an ensemble tree technique, takes advantage of the boost principle. Based on a gradient-boosted tree architecture, XGBoost uses an additive approach.

Consider a dataset represented as

$$D = \{(\boldsymbol{x}_i, y_i)\}, (|D| = n, \boldsymbol{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}),$$

where $|D| = n$ denotes the number of instances, $\boldsymbol{x}_i \in \mathbb{R}^m$ represents the feature vector of length $m$, and $y_i \in \mathbb{R}$ is the corresponding target value. XGBoost employs $K$ regression trees to make predictions. Each regression tree is constructed iteratively to minimize a loss function. Given an input feature vector $\boldsymbol{x}_i$, XGBoost computes the prediction by aggregating the predictions of all $K$ trees. The final prediction is obtained as follows:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i), f_k \in F. \tag{2.1}$$

According to the XGBoost framework, $f$ is a tree in the $F$-space, where $K$ is the number of regression trees and $\boldsymbol{x}_i$ is the $i$th occurrence. Every node in the regression tree has a score associated with it. A regularized model is part of XGBoost to prevent overfitting. The loss function is defined as follows:

$$L(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k). \tag{2.2}$$

In the XGBoost framework, $l$ represents the loss function quantifying the disparity between $\hat{y}_i$ and $y_i$. The term $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ serves to penalize the complexity of regression trees. To illustrate, $\Omega(f_k)$ is defined as $T$, where $T$ is the number of tree leaves, and $\gamma$ and $\lambda$ are factors that control the complexity of the tree. The score vector of the leaves within $T$ is also indicated by $w$. The main goal of the regularized objective function is to make it easier to get a simplified regression model that can still make predictions.

### 2.3. *Long short-term memory*

In the year 1997, Sepp Hochreiter and Jurgen Schmidhuber laid the groundwork for the development of LSTM networks. One of the most prevalent challenges encountered when training classic RNNs is the vanishing gradient problem. The LSTM was developed with the intention of resolving this issue. When gradients become smaller as they propagate back through time during training, a phenomenon known as the vanishing gradient problem develops. This phenomenon makes it difficult for the network to understand long-term relationships. The mathematical expressions that govern the behavior of LSTMs are referred to as Eq (2.3) to Eq (2.7), are as follows:

$$\text{Input gate: } \Gamma_i = \sigma(\mathscr{W}_{yi}y_t + \mathscr{W}_{hi}h_{t-1} + \mathscr{W}_{ci}\mathscr{C}_{t-1} + b_i), \tag{2.3}$$
$$\text{Forget gate: } \Gamma_f = \sigma(\mathscr{W}_{yf}y_t + \mathscr{W}_{hf}h_{t-1} + \mathscr{W}_{cf}\mathscr{C}_{t-1} + b_f), \tag{2.4}$$
$$\text{CSU: } \mathscr{C}_t = \Gamma_f \cdot \mathscr{C}_{t-1} + \Gamma_i \cdot \tanh(\mathscr{W}_{yc}y_t + \mathscr{W}_{hc}h_{t-1} + b_c), \tag{2.5}$$
$$\text{Output gate: } \Gamma_o = \sigma(\mathscr{W}_{yo}y_t + \mathscr{W}_{ho}h_{t-1} + \mathscr{W}_{co}\mathscr{C}_t + b_o), \tag{2.6}$$
$$\text{HSU: } h_t = \Gamma_o \cdot \tanh(\mathscr{C}_t), \tag{2.7}$$

where $y_t$ indicates the input at time step $t$, $h_t$ refers to the hidden state at time step $t$, and $\mathscr{C}_t$ corresponds to the cell state at time step $t$. The symbols $\sigma$ and $\cdot$ represent the activation function of the sigmoid and the multiplication by elements, respectively. $W$ stands for the weight matrices, and $b$ represents the bias vectors (refer to Figure 4).

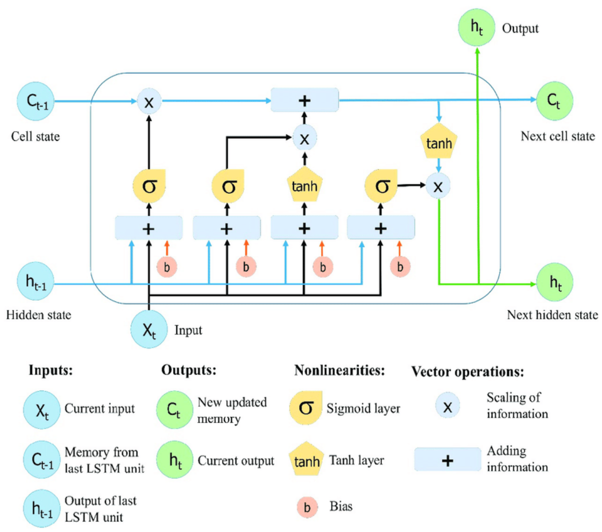



**Figure 4.** The structure of the LSTM neural network. Reproduced from Yan [27].

### 2.4. *Bidirectional long short-term memory*

The BiLSTM architecture is an expansion of the conventional LSTM architecture that integrates information from past and future time steps. Alex Graves and Jurgen Schmidhuber were the ones that initially presented it to the public in the year 2005. Through the integration of information from both previous and subsequent contexts, the BiLSTM model enhances the LSTM skills. It has two LSTM layers; one looks at the input sequence forward, and one looks at it backward. By recording bidirectional dependencies, more complete input sequence representations are produced. When processing a sequence, forward LSTM layers go from the first time step all the way to the last, while backward LSTM layers do the opposite. In the LSTM layer forward, $\vec{\hbar_t}$ represents the hidden states and $\vec{\mathscr{C}_t}$ represents the cell states, where $t$ is the current time step. In the same way, $\hbar_t^{-}$ and $\mathscr{C}_t^{-}$ denote LSTM layers backward.

Using the following Eq (2.8) to Eq (2.10), the forward LSTM layer updates its $\vec{\hbar_t}$ and $\vec{\mathscr{C}_t}$.

$$I_t^{\rightarrow} = \sigma\left(\mathscr{W}_{xi}X_t + \mathscr{W}_{hi}\hbar_{t-1}^{\rightarrow} + \mathscr{W}_{ci}\mathscr{C}_{t-1}^{\rightarrow} + b_i\right), \tag{2.8}$$

$$F_t^{\rightarrow} = \sigma\left(\mathscr{W}_{xf}X_t + \mathscr{W}_{hf}\hbar_{t-1}^{\rightarrow} + \mathscr{W}_{cf}\mathscr{C}_{t-1}^{\rightarrow} + b_f\right), \tag{2.9}$$

$$\mathscr{C}_t^{\rightarrow} = F_t^{\rightarrow} \odot \mathscr{C}_{t-1}^{\rightarrow} + I_t^{\rightarrow} \odot \tanh\left(\mathscr{W}_{xc}X_t + \mathscr{W}_{hc}\hbar_{t-1}^{\rightarrow} + b_c\right), \tag{2.10}$$

$$O_t^{\rightarrow} = \sigma\left(\mathscr{W}_{xo}X_t + \mathscr{W}_{ho}\hbar_{t-1}^{\rightarrow} + \mathscr{W}_{co}\mathscr{C}_t^{\rightarrow} + b_o\right), \tag{2.11}$$

$$\hbar_t^{\rightarrow} = O_t^{\rightarrow} \odot \tanh\left(\mathscr{C}_t^{\rightarrow}\right). \tag{2.12}$$

Equations (2.10) to (2.17) are used by the LSTM layer backward to update both its hidden state and cell state.

$$I_t^{\leftarrow} = \sigma\left(\mathscr{W}_{xi}X_t + \mathscr{W}_{hi}\hbar_{t+1}^{\leftarrow} + \mathscr{W}_{ci}\mathscr{C}_{t+1}^{\leftarrow} + b_i\right), \tag{2.13}$$

$$F_t^{\leftarrow} = \sigma\left(\mathscr{W}_{xf}X_t + \mathscr{W}_{hf}\hbar_{t+1}^{\leftarrow} + \mathscr{W}_{cf}\mathscr{C}_{t+1}^{\leftarrow} + b_f\right), \tag{2.14}$$

$$\mathscr{C}_t^{\leftarrow} = F_t^{\leftarrow} \odot \mathscr{C}_{t+1}^{\leftarrow} + I_t^{\leftarrow} \odot \tanh\left(\mathscr{W}_{xc}X_t + \mathscr{W}_{hc}\hbar_{t+1}^{\leftarrow} + b_c\right), \tag{2.15}$$

$$O_t^{\leftarrow} = \sigma\left(\mathscr{W}_{xo}X_t + \mathscr{W}_{ho}\hbar_{t+1}^{\leftarrow} + \mathscr{W}_{co}\mathscr{C}_t^{\leftarrow} + b_o\right), \tag{2.16}$$

$$\hbar_t^{\leftarrow} = O_t^{\leftarrow} \odot \tanh\left(\mathscr{C}_t^{\leftarrow}\right). \tag{2.17}$$

By integrating all the hidden states from both the forward and backward directions, $\hbar_t$, as shown in Eq (2.18), we can derive the final hidden state of the BiLSTM network at time step $t$.

$$\hbar_t = [\hbar_t^{\rightarrow}; \hbar_t^{\leftarrow}]. \tag{2.18}$$

A concatenation operation is represented by the symbol $[\cdot]$. Take a look at Figure 5.

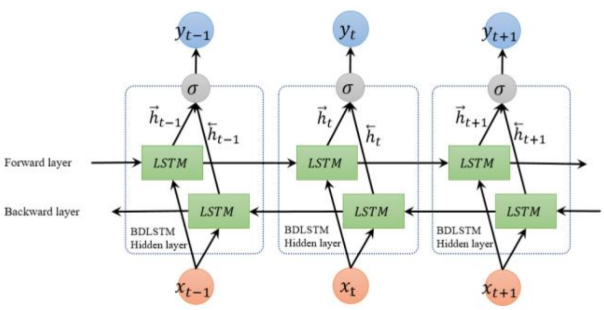



**Figure 5.** Unfolded architecture of BiLSTM with three consecutive steps [28].

## 3. Experimenis

### 3.1. Data collection

The dataset was acquired using a comprehensive surveillance system at the study site, which integrates various sensors to ensure accurate data collection. It includes real-time greenhouse power consumption and solar radiation data within the smart farm (as shown in Figure 6). The data was sourced from the Republic of Korea's JEONNAM AGRICULTURAL RESEARCH & EXTENSION SERVICES (https://www.jares.go.kr/main/business) and covers the period from June 28, 2022, to December 13, 2023. The use of modern communication technologies enables seamless integration of data from multiple sensors, ensuring a robust dataset for thorough research.

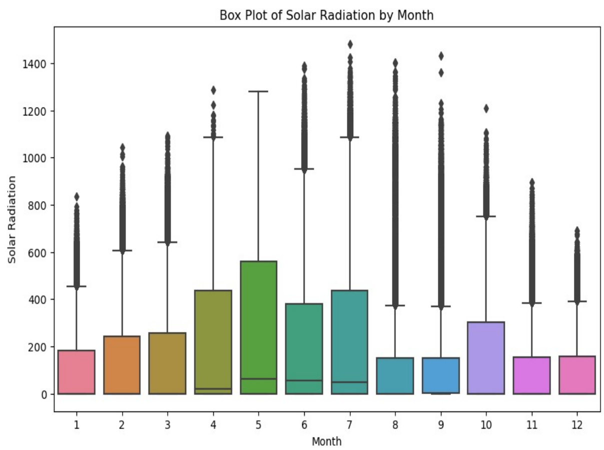



**Figure 6.** Solar radiation analysis by month.

### 3.2. Data preprocessing

Solar radiation data collected in real-time from PV plants in greenhouses is the dataset used in this research. A number of preprocessing processes are carried out to improve the quality of the data and maximize the performance of the model before training the hybrid model to forecast solar radiation. Table 1 summarizes the features of the dataset and uses a forward fill method to fill missing values within the dataset to ensure the continuity of the dataset. Additionally, monthly variations in solar radiation are analyzed using box plots, providing insights into seasonal trends (Figure 6). The multivariate time series data includes attributes alongside solar radiation measurements, enabling a comprehensive understanding of environmental factors influencing solar radiation patterns. BiLSTM is employed to forecast solar radiation, leveraging its ability to capture complex temporal patterns. BiLSTM decomposes the time series, enhancing the extraction of the patterns. The data is then segmented into equal-length sequences for the BiLSTM input. To evaluate performance, the dataset is split into training (80%) and testing (20%) sets for each year, enabling comparison with established techniques like XGBoost, ARIMA, and LSTM. This approach ensures a comprehensive assessment of the forecast's performance.

**Table 1.** Dataset features.

| No | Attribute | Description |
|---|---|---|
| 1 | Solar-radiation(kWh/m2) | Energy emitted by the Sun (kWh/m2) |
| 2 | outside-windspeed(WS) | Average outside-(WS) (Km/h) |
| 3 | outside-temperature | The Noon temperature (C) |
| 4 | Dew-point | Dew point temperature (C) |
| 5 | Absolute-humidity | Air humidity (%) |
| 6 | Outside-rain | Precipitation |

## 4. Experimental results and discussion

The effectiveness of the designed deep learning models is evaluated through precision metrics and a range of error indicators, which are thoroughly discussed in Sections 2 and 3. Real-world data samples are utilized to guarantee consistent evaluation across all experiments. The Adam optimizer and mean squared error (MSE) loss function are used, with early stopping configured after 500 iterations. Running on a Windows 10 system with 32.00 GB RAM and an Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz, the LSTM model predicts one-hour-ahead power consumption in greenhouses. The evaluation metrics include MAE, MSE, coefficient of determination ($R^2$), and root mean squared error (RMSE). These metrics highlight the model's accuracy and reliability, such as MAE $= \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$, MSE $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)2$, $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$, and $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$, In this context, where $y_i$ denotes the actual solar radiation value, $\hat{y}_i$ represents the predicted solar radiation value, $\bar{y}$ is the average of the observed solar radiation values, and $n$ stands for the total count of samples. These metrics provide insight into the accuracy and effectiveness of our model in predicting solar radiation.

### 4.1. *Solar radiation forecast: one hour ahead*

One of the most important parts of energy management is solar radiation forecasting, especially when it comes to using solar power and other renewable resources. Several meteorological factors, including cloud cover and atmospheric conditions, are taken into account by hour-ahead forecasts to estimate the amount of solar radiation that will reach solar panels in the next hour. In order to maximize energy efficiency, we provide a BiLSTM model for real-time prediction of solar radiation in this research. Organizing data on past solar radiation is the first step in making predictions. The results are shown in Table 2, which contrasts the performance of BiLSTM with that of the XGBoost, LSTM, and ARIMA models. By comparing the R2 values of each model at respective time steps, we can see that BiLSTM performs better. Table 1 shows that compared to the other models, BiLSTM has significantly lower MAE, MSE, and RMSE, indicating that it is the most predictively powerful. Furthermore, BiLSTM's constantly high R2 values highlight the great quality of fit it achieves. Figure 7 shows side-by-side bar graphs that compare MAE and MSE on several forecast horizons and models to make visual comparisons easier. We will describe the most effective experimental outcomes after running the BiLSTM model independently many times. With the help of Figure 8, we can see how well BiLSTM predicts solar radiation. To further illustrate their prediction accuracy across the given time range, Figure 7 provides a detailed comparison of the hour-ahead performance of the BiLSTM, XGBoost, LSTM, and ARIMA models.

**Table 2.** Comparison of different models based on various error indicators.

| Method | Resoloution | Hidden Layer | Epoch | MAE (↓) | MSE (↓) | RMSE (↓) | R2 (↑) |
|---|---|---|---|---|---|---|---|
| ARIMA | 1Hr | 32 | 500 | 28.0930 | 5458.3886 | 73.8809 | 0.92086 |
| XGBoost | 1Hr | 32 | 500 | 24.8190 | 5001.7666 | 70.7231 | 0.9274 |
| LSTM | 1Hr | 32 | 500 | 15.2394 | 2571.2052 | 50.7070 | 0.9624 |
| **BiLSTM** | 1Hr | 32 | 500 | **13.5327** | **2564.4634** | **50.6405** | **0.9625** |

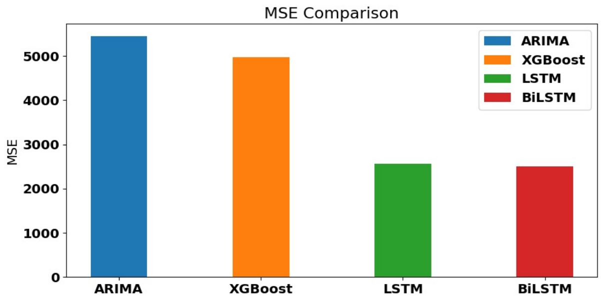



**Figure 7.** Comparison graph of MAE and MSE for all models.

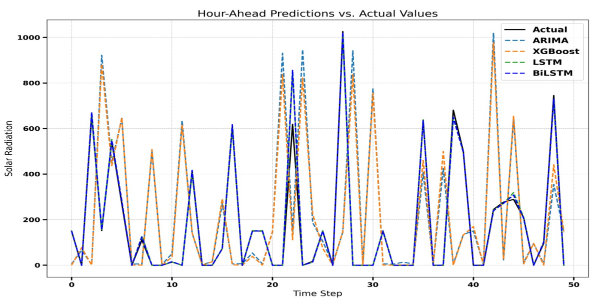



**Figure 8.** Comparison analysis of ARIMA, XGBoost, LSTM, and BiLSTM models

*4.2. Results and discussion*

Accurate forecasting of solar radiation, particularly for short time intervals, is essential for effective energy consumption management. This study presents a BiLSTM model designed to predict solar radiation one hour into the future, utilizing preprocessing techniques on historical data. Table 2 summarizes the performance of various forecasting models based on several error metrics.

- **Hour-ahead predictions:** Our BiLSTM model demonstrates superior performance compared to other forecasting methods. It achieves the lowest MAE of 13.53 and MSE of 2564.46 while maintaining a high R-squared ($R^2$) value of 0.96. The RMSE of 50.64 further confirms the model's accuracy in predicting solar radiation one hour ahead.

The consistent ability of the BiLSTM model to forecast solar radiation accurately, even over short-term intervals, is evident from these results. Using its bidirectional nature, the BiLSTM model effectively captures patterns and temporal relationships in solar radiation data, leading to improved prediction accuracy. Compared to ARIMA, XGBoost, and standard LSTM models, the BiLSTM shows reduced error metrics (RMSE, MAE, and MSE) and a higher $R^2$ value, making it the most accurate for short-term predictions. Although the BiLSTM model performs exceptionally well for one-hour forecasts, the ARIMA model exhibits slightly higher errors for longer forecast horizons. This suggests that forecasting solar radiation beyond one hour remains a challenge. Future research could focus on enhancing the BiLSTM model by incorporating additional meteorological variables or exploring alternative modeling approaches to address the complexities of long-term solar radiation forecasting. Such advances could significantly improve the accuracy of renewable energy predictions and improve long-term solar radiation forecasts.

## 5. Conclusions

The study rigorously assessed the effectiveness of various predictive models for forecasting solar radiation, a critical factor in optimizing solar energy utilization in greenhouse settings. By employing a range of statistical and machine learning models, the goal was to improve the accuracy of solar radiation predictions, thereby supporting precise energy management and cultivation practices within greenhouses in Naju, South Jeolla Province. The BiLSTM model proposed in this study demonstrated strong forecasting abilities, with an MAE of 1 hour. A comparative analysis revealed that the LSTM model had an MAE of 15.23, while XGBoost and ARIMA recorded MAEs of 24.64 and 28.52, respectively. These results highlight the crucial importance of accurate solar radiation forecasting, particularly emphasizing the effectiveness of the BiLSTM model in enhancing solar power generation within greenhouse settings. In future research efforts, there is an opportunity to explore the potential of hybrid deep learning models, such as combining LSTM with other neural network architectures or statistical methods, to enhance prediction accuracy further. Investigating the adaptability of the BiLSTM model to larger greenhouse setups, along with expanding the forecast horizon beyond short-term predictions, could provide crucial insights for developing more effective long-term energy management approaches in greenhouse agriculture.

## Use of Generative-AI tools declaration

The authors declare that they have not used artificial intelligence (AI) tools in the creation of this article.

## Availability of data and materials

Data will be made available on request.

## Acknowledgments

This work was supported by Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government(MOTIE) (20202020900060, The Development and Application of Operational Technology in Smart Farm Utilizing Waste Heat from Particulates Reduced Smokestack).

## Conflict of interest

The authors of this paper declare no conflicts of interest.

## References

1. R. G. Newell, D. Raimi, S. Villanueva, B. Prest, Global energy outlook 2020: energy transition or energy addition, *Resources for the Future*, 2020. Available from: `https://media.rff.org/documents/GEO_2020_Report.pdf`.

2. B. Espinar, J. L. Aznarte, R. Girard, A. M. Moussa, G. Kariniotakis, Photovoltaic forecasting: a state of the art, *5th European PV-hybrid and mini-grid conference*, Tarragona, Spain, 2010. Available from: `https://minesparis-psl.hal.science/hal-00771465`.

3. A. Ghosh, Nexus between agriculture and photovoltaics (agrivoltaics, agriphotovoltaics) for sustainable development goal: a review, *Sol. Energy*, **266** (2023), 112146. https://doi.org/10.1016/j.solener.2023.112146

4. K. Rabea, S. Michailos, G. T. Udeh, J. Park, Y. W. Lee, S. Kim, et al., A comprehensive technoeconomic and environmental evaluation of a hybrid renewable energy system for a smart farm in South Korea, *Int. J. Energy Res.*, **2023** (2023), 1–18. https://doi.org/10.1155/2023/4951589

5. D. Venkateswaran, Y. Cho, Efficient solar power generation forecasting for greenhouses: a hybrid deep learning approach, *Alex. Eng. J.*, **91** (2024), 222–236. https://doi.org/10.1016/j.aej.2024.02.004

6. D. Venkateswaran, Y. Cho, C. Shin, Hybrid LSTM-Markovian model for greenhouse power consumption prediction: a dynamical approach, *Eur. Phys. J.-Spec. Top.*, **234** (2025), 2673–2683. https://doi.org/10.1140/epjs/s11734-024-01244-w

7. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015. Available from: `https://books.google.co.kr/books?id=rNt5CgAAQBAJ`.

8. M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and arima models for time series forecasting, *Appl. Soft Comput.*, **11** (2011), 2664–2675. https://doi.org/10.1016/j.asoc.2010.10.015

9. A. A. Adebiyi, A. O. Adewumi, C. K. Ayo, Stock price prediction using the arima model, *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, IEEE, 2014, 106–112. https://doi.org/10.1109/UKSim.2014.67

10. A. M. Alonso, C. G. Martos, Time series analysis-forecasting with arima models, *Universidad Carlos III de Madrid, Universidad Politecnica de Madrid*, 2012.

11. C. Krauss, X. A. Do, N. Huck, Deep neural networks, gradient-boosted trees, random forests: statistical arbitrage on the S&P 500, *Eur. J. Oper. Res.*, **259** (2017), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031

12. V. E. Sathishkumar, J. Park, Y. Cho, Using data mining techniques for bike sharing demand prediction in metropolitan city, *Comput. Commun.*, **153** (2020), 353–366. https://doi.org/10.1016/j.comcom.2020.02.007

13. V. E. Sathishkumar, Y. Cho, Season wise bike sharing demand analysis using random forest algorithm, *Comput. Intell.*, **40** (2024), e12287. https://doi.org/10.1111/coin.12287

14. V. E. Sathishkumar, S. Venkatesan, J. Park, C. Shin, Y. Kim, Y. Cho, Nutrient water supply prediction for fruit production in greenhouse environment using artificial neural networks, *Basic Clin. Physiol. Pharmacol.*, **126** (2020), 257–258.

15. S. I. Lee, S. J. Yoo, A deep efficient frontier method for optimal investments, *arXiv preprint*, 2017. https://doi.org/10.48550/arXiv.1709.09822

16. J. Kim, N. Moon, Bilstm model based on multivariate time series data in multiple field for forecasting trading area, *J. Ambient Intell. Human Comput.*, 2019. https://doi.org/10.1007/s12652-019-01398-9

17. Z. Cui, R. Ke, Z. Pu, Y. Wang, Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction, *arXiv Preprint*, 2018. https://doi.org/10.48550/arXiv.1801.02143

18. S. S. Namini, N. Tavakoli, A. S. Namin, A comparison of arima and lstm in forecasting time series, *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227

19. N. Tavakoli, Modeling genome data using bidirectional lstm, *IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, IEEE, 2019, 183–188. https://doi.org/10.1109/COMPSAC.2019.10204

20. N. Tavakoli, D. Dai, Y. Chen, Client-side straggler-aware I/O scheduler for object-based parallel file systems, *Parallel Comput.*, **82** (2019), 3–18. https://doi.org/10.1016/j.parco.2018.07.001

21. Y. Pang, X. Xue, A. S. Namin, Predicting vulnerable software components through n-gram analysis and statistical feature selection, *14th International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2015, 543–548. https://doi.org/10.1109/ICMLA.2015.99

22. M. K. Anser, M. Mohsin, Q. Abbas, I. S. Chaudhry, Assessing the integration of solar power projects: SWOT-based AHP–F–TOPSIS case study of Turkey, *Environ. Sci. Pollut. Res.*, **27** (2020), 31737–31749. https://doi.org/10.1007/s11356-020-09092-6

23. D. C. Montgomery, C. L. Jennings, M. Kulahci, *Introduction to time series analysis and forecasting*, John Wiley & Sons, 2015. Available from: `https://books.google.com/books?id=JCFiBwAAQBAJ`.

24. R. J. Hyndman, G. Athanasopoulos, *Forecasting: principles and practice*, OTexts, 2018. Available from: `https://otexts.com/fpp2`.

25. T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, 785–794. https://doi.org/10.1145/2939672.2939785

26. H. Abbasimehr, R. Paki, A. Bahrini, A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models, *Sustain. Comput.-Inf. Syst.*, **38** (2023), 100863. https://doi.org/10.1016/j.suscom.2023.100863

27. S. Yan, Understanding LSTM and its diagrams, *ML Review*, 2017.

28. Z. Cui, R. Ke, Z. Pu, Y. Wang, Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values. *Transport Res. C-Emer.*, **118** (2020), 102674. https://doi.org/10.1016/j.trc.2020.102674