



OPEN

Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: a study on solar potential in Ghardaia, Algeria

Boumediene Ladjal¹, Mohamed Nadour², Mohcene Bechouat¹, Nadji Hadroug²,
Moussa Sedraoui³, Abdelaziz Rabehi⁴, Mawloud Guermoui^{4,5} & Takele Ferede Agajie⁶✉

This paper provides an in-depth analysis and performance evaluation of four Solar Radiance (SR) prediction models. The prediction is ensured for a period ranging from a few hours to several days of the year. These models are derived from four machine learning methods, namely the Feed-forward Back Propagation (FFBP) method, Convolutional Feed-forward Back Propagation (CFBP) method, Support Vector Regression (SVR), and the hybrid deep learning (DL) method, which combines Convolutional Neural Networks and Long Short-Term Memory networks. This combination results in the CNN-LSTM model. Additionally, statistical indicators use Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Normalized Root Mean Squared Error (nRMSE). Each indicator compares the predicted output by each model above and the actual output, pre-recorded in the experimental trial. The experimental results consistently show the power of the CNN-LSTM model compared to the remaining models in terms of accuracy and reliability. This is due to its lower error rate and higher detection coefficient ($R^2 = 0.99925$).

Keywords Artificial neural networks, Convolutional neural network, Convolutional feed-forward back propagation, Deep learning, Feed-forward back propagation, Long short-term memory, Solar radiance forecasting

In recent years, advancements in renewable energy technologies have gained global importance due to the increasing demand for sustainable and environmentally friendly energy solutions. Solar energy, as an abundant and clean resource, has attracted significant research attention, particularly in optimizing its production and integration into power grids. However, the efficiency of solar energy systems depends on the accurate forecasting of Direct Normal Irradiance (DNI), which is essential for planning energy generation, optimizing photovoltaic (PV) and concentrated solar power (CSP) systems, and ensuring grid stability. Accurate DNI prediction enables better energy management, reduces operational costs, and supports large-scale adoption of solar power^{1–3}. Solar radiation forecasting is inherently complex due to its dependence on various meteorological factors such as cloud cover, atmospheric aerosols, temperature variations, and humidity levels. These factors exhibit nonlinear and dynamic behaviors, making it difficult for traditional forecasting methods to provide precise and reliable predictions. Conventional approaches, including Numerical Weather Prediction (NWP) models, statistical regression techniques, and empirical models, often fail to capture the intricate relationships between these meteorological variables. NWP models rely on physics-based simulations, which require high computational power, while statistical models assume linear dependencies, limiting their accuracy in highly variable weather conditions. In most cases, conventional statistical approaches, such as Linear Regression (LR), are inadequate for predicting solar radiation, necessitating more advanced methodologies^{4–6}. This paper emphasizes the importance of accurate solar radiation (SR) forecasting, as it is essential for the effective development of

¹Department of Automation and Electromechanics, Faculty of Science and Technology, University of Ghardaia, Ghardaia, Algeria. ²Applied Automation and Industrial Diagnostics Laboratory (LAADI), Faculty of Sciences and Technology, University of Djelfa, 17000 Djelfa, Algeria. ³Laboratory of Inverse Problems, Modeling, Information and Systems (PI-MIS), University 8 Mai 1945, Guelma, Algeria. ⁴Telecommunications and Smart Systems Laboratory, University of Djelfa, PO Box 3117, 17000 Djelfa, Algeria. ⁵Centre de Développement des Energies Renouvelables, Unité de Recherche Appliquée en Energies Renouvelables, URAER, CDER, 47133 Ghardaia, Algeria. ⁶Department of Electrical and Computer Engineering, Faculty of Technology, Debre Markos University, 269 Debre Markos, Ethiopia. ✉email: takele_ferede@dmu.edu.et

renewable energy technologies. The city of Ghardaïa, Algeria, has been chosen as the research focus due to its high solar potential and suitability for future solar energy projects. Through studies on SR forecasting, this research aims to extend predictive capabilities to other variables related to renewable energy generation. Several forecasting methods exist for solar radiation, each suited to different timescales. Long-term forecasts often rely on statistical and physical models, while short-term and very short-term forecasts require more advanced computational techniques^{7,8}. Ghardaïa's unique climate, characterized by high variability in cloud conditions, necessitates specialized forecasting models, which are a key focus of this study. Additionally, the high demand for solar energy in Algeria's desert regions requires precise predictions, making microclimatic-level modeling using numerical meteorological data essential⁹.

To address these limitations, Artificial Intelligence (AI) and Deep Learning (DL) techniques have emerged as powerful tools for solar irradiance forecasting. Unlike traditional models, deep learning approaches can automatically learn complex patterns from large datasets, enabling improved predictive accuracy. Among the various deep learning architectures, hybrid models combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have shown superior performance in time-series forecasting tasks. CNNs specialize in spatial feature extraction, identifying patterns in meteorological inputs such as cloud movement, temperature fluctuations, and pressure variations. Meanwhile, LSTMs are designed to process sequential data, capturing long-term dependencies and trends in solar irradiance fluctuations. By integrating CNNs and LSTMs into a single hybrid framework, forecasting models can effectively analyze both spatial and temporal dependencies, resulting in more accurate DNI predictions^{10–13}. Recent studies in renewable energy systems, such as microbial electrosynthesis¹⁴ and biohydrogen production¹⁵, have demonstrated the transformative potential of hybrid architectures and adaptive optimization. These advancements underscore the need for similarly innovative approaches in solar forecasting, particularly in regions like Ghardaïa with high climatic variability.

We focus on the integration of advanced neural networks and hybrid deep learning (DL) techniques. Below are the key highlights of our study, showcasing significant advancements in AI-based approaches while addressing the limitations of conventional forecasting models:

1. **Advanced AI Techniques Integration:** The research utilizes a combination of conventional Artificial Neural Networks (ANNs) and DL methods, specifically CNN-LSTM models, to improve solar radiation forecasting by effectively handling complex meteorological data.
2. **Focus on Ghardaïa, Algeria:** Centered in Ghardaïa, a region with significant solar potential, the study tailors its findings to local climatic conditions marked by variable cloud cover and atmospheric factors.
3. **Hybrid CNN-LSTM Model:** By combining CNNs for spatial feature extraction and LSTMs for capturing temporal dependencies, the hybrid model effectively addresses the nonlinear nature of SR data.
4. **Evaluation of ANN Architectures:** The study compares various ANN architectures, including Feed-Forward Back Propagation (FFBP), Convolutional FFBP (CFBP), Support Vector Regression (SVR), and the hybrid CNN-LSTM approach, using advanced training algorithms to identify the most effective techniques for solar radiation forecasting.
5. **Acknowledgement of Limitations and Contributions:** While demonstrating the advantages of AI-driven approaches, the research acknowledges challenges such as data scarcity and model complexity, emphasizing the importance of optimizing input parameters for accurate solar predictions and contributing to renewable energy management.

This study focuses on developing a hybrid CNN-LSTM-based DNI forecasting framework and applying it to the Ghardaïa region in Algeria, known for its high solar potential and climate variability. The region experiences significant variations in temperature, humidity, wind speed, and dust storms, all of which impact solar irradiance levels. A reliable forecasting model tailored to these conditions is crucial for optimizing solar energy production and supporting Algeria's renewable energy initiatives.

The dataset used in this study is sourced from NASA's Prediction of Worldwide Energy Resources (POWER) project, covering the period from 2017 to 2023, and consists of 2575 daily observations. The key meteorological parameters include DNI, maximum and minimum temperature, wind speed, specific humidity, and surface pressure. These variables play a critical role in training deep learning models to capture variations in solar radiation and improve forecasting accuracy.

To evaluate the performance of the proposed CNN-LSTM model, its predictions are compared against conventional forecasting techniques, including Feed-Forward Back Propagation (FFBP), Convolutional FFBP (CFBP), and Support Vector Regression (SVR). The models are assessed using multiple statistical performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Normalized Root Mean Squared Error (nRMSE), and the coefficient of determination (R^2).

The results demonstrate that the CNN-LSTM model significantly outperforms traditional approaches, achieving lower forecasting errors and higher accuracy, making it a promising solution for real-world solar energy management. Beyond its technical contributions, this study highlights practical applications in optimizing solar power plant operations, improving energy grid stability, and supporting long-term renewable energy policies.

The structure of the paper is section “[Related works](#)” reviews related work on AI-based solar irradiance forecasting. Section “[Materials and methods](#)” describes the methodology, including dataset preparation, preprocessing, and model architecture. Section “[Results and discussions](#)” presents experimental results and comparative performance analysis. Finally, Section “[Conclusion](#)” concludes the study and discusses potential future research directions.

Related works

In recent years, researchers have shown considerable interest in SR forecasting, dedicating numerous studies to this field. SR is influenced by geographic characteristics of the Earth's surface and surrounding weather conditions. Unfortunately, it is subjected to uncontrollable changes, such as seasonal and diurnal variations, city locations, and cloud abundance. Consequently, accurate estimation of SR on Earth is crucial for maximizing solar operations. SR forecasting is crucial in renewable energy systems. Accurate predictions can minimize energy costs and optimize energy production, enabling plants to contribute effectively to energy supply chains, and supporting commerce, industry, and transportation services consistently. Various types of forecasting problems, such as electric load, wind, and solar power requests, have been a focus for decades. Today, solar power is essential and increasingly contributes to the electricity mix, becoming competitive with conventional sources. Regardless of the specific applications, solar power forecasting involves system-referenced interfaces encompassing inputs as controllable factors and the proposed output results. Paper¹⁶ provides an overview of advancements in forecasting methods for wind and solar energy, highlighting improvements in forecast skill, the impact of blockchain technology on data transactions, evolving business models in the renewable sector, and challenges posed by integrating high levels of renewable energy into power systems. The authors in the paper¹⁷ explore various methodologies for energy prediction and management in the context of renewable energy and electricity consumption forecasting. Article¹⁸ presents a comprehensive review of solar irradiance resources and forecasting techniques, focusing on using sensor networks for accurate solar power generation predictions. The authors in the paper¹⁹ investigate various methods of wind power forecasting, focusing on their analysis, prediction time scales, error measurements, and accuracy improvements to enhance the efficiency of wind energy systems. ANNs have been predominantly used in SR estimation. SR patterns undergo dynamic spatial variation in any region where geographic and weather conditions are not constant over time. Consequently, ANNs emerge as a flexible transfer function, making excellent forecasts of solar data patterns. Numerous studies have used ANNs for SR forecasting in various areas, showcasing the potential of this approach. The generalization capabilities of ANNs were highlighted by authors in²⁰, demonstrating their strength in handling noisy and incomplete real-world data. This capability is crucial in modeling solar radiation, as it involves intricate and variable datasets. The review of 54 papers by Eisa underscores the importance of selecting performance criteria and network architecture as foundational steps toward successful forecasting in diverse contexts. Based on this foundation, Voyant et al.²¹ investigate the predictability of solar irradiation, particularly for insular locations without weather stations. This paper presents innovative work using physical phenomena to improve ANN predictions by addressing challenges posed by non-stationarity in solar data. By incorporating extraterrestrial irradiation coefficients, they successfully improve prediction quality, showcasing ANNs' adaptability in unique environmental conditions. In the paper²², the authors advance the discourse by introducing radial basis function networks for estimating SR in areas without direct measurements. Their approach includes univariate and multivariate forecasting techniques, highlighting the versatility of ANNs in different climatic contexts. The authors demonstrate that ANNs can effectively model solar radiation, emphasizing the need for selecting appropriate input parameters to optimize performance. Hoyos-Gómez et al., in a paper²³, address the increasing demand for accurate solar irradiance forecasting due to the proliferation of solar power generation systems. They highlight the robustness of ANNs, fuzzy logic, and hybrid models in capturing systematic errors, reinforcing ANNs' status as a leading technique in solar irradiance modeling. Their review of existing studies indicates a significant gap in research focused on tropical environments, signaling an area for future exploration. Bou Nassif et al. review various methodologies combining ANN with statistical techniques for short-term load forecasting. They note that integrating additional input parameters can enhance ANN accuracy, further supporting the notion that ANNs can be effectively tailored to specific forecasting challenges²⁴. Recent advancements are reflected in²⁵, which focuses on the critical role of short-term PV forecasts in energy dispatch decisions. Their study illustrates the evolution of prediction methods, emphasizing the necessity of incorporating a broader range of meteorological features alongside traditional solar irradiance data. This approach aligns with the trend of enhancing ANN models to improve forecasting precision. Nawaeb et al., in²⁶, provide a comparative review of empirical and AI-based approaches to solar irradiation forecasting, stressing that ANNs are widely used across different geographical settings. Region-specific studies have also gained prominence. For instance, Benali et al.²⁷ applied a hybrid CNN-LSTM model to North African solar datasets, emphasizing the importance of localized feature engineering for arid climates. Their findings validate our methodology of integrating NASA POWER data with lagged SR measurements for Ghardaïa. Additionally, Chen et al. in²⁸ explored the fusion of satellite-derived cloud dynamics with ground-based sensors using CNN-LSTM and achieved sub-hourly irradiance predictions, complementing our use of NASA's CERES/MERRA2 dataset. Their results indicate that although empirical methods remain relevant, the trend shows a tendency toward AI models more accurately, ANNs when accuracy in predictions is required. Finally, innovative ML techniques in²⁹ are applied to solar energy estimation, notably the source-free domain adaptation approach. Their work highlights the potential of advanced ANN configurations, such as LSTM networks, to adapt dynamically to changing environmental conditions, thereby improving predictive performance across different regions. Recent comparative studies, such as Wang & Li³⁰, have shown that hybrid CNN-LSTM models outperform transformer-based architectures for short-term forecasts due to their localized spatial feature extraction capabilities. This aligns with our model's design, which prioritizes short-term accuracy for energy dispatch planning. Furthermore, ethical considerations in renewable energy AI are increasingly addressed through explainability frameworks. For example, Ibrahim et al. in³¹ integrated SHAP (SHapley Additive exPlanations) into solar forecasting models to interpret feature contributions, highlighting a direction we recognize as crucial for future work. Recent advancements in hybrid models have further improved solar irradiance forecasting. Zhang et al. in³² proposed a spatio-temporal transformer combined with LSTM, achieving state-of-the-art accuracy in arid regions like Ghardaïa. Similarly, Kumar et al. in³³ demonstrated that attention-augmented CNN-LSTM architectures significantly improve forecasting robustness for non-stationary

Category	Details
Source	NASA POWER (CERES/MERRA2)
Location	Latitude: 32.4466° N, Longitude: 3.6992° E
Elevation	522.19 m
Time Period	January 1, 2017–December 31, 2023
Resolution	0.5°×0.625° latitude/longitude

Table 1. Dataset overview and metadata.

Variable	Description
ALLSKY_SFC_SW_DWN	All sky surface shortwave downward irradiance (kW-hr/m ² /day)
WS2M	WS at 2 m above ground level (m/s)
T2M_MAX	Maximum temperature at 2 m above ground level (°C)
T2M_MIN	Minimum temperature at 2 m above ground level (°C)
QV2M	Specific humidity at 2 m above ground level (g/kg)
PS	Surface pressure (kPa)

Table 2. Parameters and descriptions.

solar data. These studies highlight the growing emphasis on hybrid architectures for capturing both spatial and temporal dependencies, a critical requirement for regions with high climatic variability.

In this study, we will present a comparative analysis of a CNN-LSTM model against several traditional forecasting methods, including SVR and multiple configurations of Feed Forward and Convolutional Feed-forward ANNs. The performance of each model will evaluate using key metrics: MSE, RMSE, MAE, MAPE, nRMSE, and the coefficient of determination (R²).

Materials and methods

This section outlines the methodology employed in this study, including dataset acquisition, preprocessing techniques, model architecture, and performance evaluation. We focus on the integration of advanced deep learning techniques, particularly CNN-LSTM models, to enhance solar radiation forecasting by leveraging both spatial feature extraction and temporal sequence learning.

Dataset collection

The dataset used in this study is sourced from NASA’s Prediction of Worldwide Energy Resources (POWER) project, covering the period from 2017 to 2023. The study focuses on Ghardaia, Algeria (Table 1), an area with high solar potential and significant climatic variability, making it ideal for evaluating deep learning-based solar forecasting models. The dataset consists of 2575 daily observations of meteorological parameters critical for solar irradiance forecasting (Table 2).

Data preprocessing

Data preprocessing is essential to ensure the quality and consistency of inputs for deep learning models. In this study, missing values are handled using linear interpolation to maintain data continuity, while all features are normalized to the [0,1] range using Min–Max normalization to enhance convergence during training. Feature engineering involves introducing temporal lag variables to help LSTMs capture time-dependent patterns and conducting correlation analysis to select the most relevant meteorological features. The dataset is then split into 85% training and 15% testing data to evaluate model performance on unseen observations. Finally, the data is formatted into 3D input sequences, making it suitable for CNN-LSTM models by incorporating multiple timesteps to capture sequential dependencies effectively. The different ML and DL models used for solar irradiance forecasting are represented in Fig. 1.

The CNN design model

The CNN is employed to extract spatial features from meteorological data, such as temperature, humidity, and cloud cover, which are critical for understanding local weather patterns. For example, convolutional layers can identify regions of high cloud coverage that may reduce solar irradiance. The dataset, which includes climatic variables measured at 10-min intervals, is well-suited for CNN processing due to its multidimensional structure and large scale^{29,34–38}. This allows the model to capture complex spatial patterns that influence solar irradiance. The architecture of CNNs typically consists of multiple layers, including convolutional layers with local receptive fields. These layers allow the model to apply convolution matrices to specific segments of the input data, mathematically represented as follows:

$$(f * w)(i, j) = \sum_m \sum_n f(m, n) \cdot w(i - m, j - n) \tag{1}$$

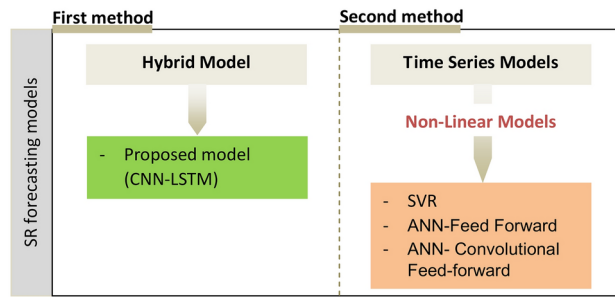


Fig. 1. Different models for solar irradiance forecasting.

This feature extraction process not only reduces the size but also enhances the model's resilience by maintaining invariant properties. The activation function, usually of the Rectified Linear Unit (ReLU) type, is applied post-convolution and is expressed as follows:

$$a(x) = \max(0, x) \quad (2)$$

Additionally, max pooling is employed to down sample the feature maps, which can be represented as:

$$p(i, j) = \max_{m, n \in \text{pooling region}} f(i + m, j + n) \quad (3)$$

Ultimately, the output of the fully connected layer is computed using the equation:

$$y = W.a + b \quad (4)$$

Such robustness is critical in this study, especially when considering the minimal spatial invariance of cloud patterns that influence solar irradiance. In designing the solar radiance model, the first two layers function as convolutional layers, where they capture essential features from the input data. Following these are the flatten layer, which converts the multidimensional outputs of the convolutional layers into a one-dimensional vector, allowing for seamless integration with the LSTM layers.

The LSTM networks

The LSTM networks were first proposed in 1997 as a specific type of Recurrent Neural Networks (RNNs) designed explicitly to overcome the long-term dependency problem that architectures more general than RNNs usually suffer from. We utilize, in this study, LSTM networks in effectively forecasting direct normal irradiance by harnessing the strengths that come with LSTM in dealing with time-series data. The architecture of LSTM networks involves several types of gates, which control the flow of input data while training. This gating mechanism is relevant in our context because it allows the model to forget or remember information selectively; this is important to prevent the vanishing gradient often found in long sequences. This capability can empower LSTMs in transferring relevant information through the forecasting process, ensuring that important temporal patterns in solar irradiance data are recognized and utilized. The LSTM model utilizes three main gates to control information flow: the input gate, forget gate, and output gate. These gates are represented mathematically to decide what information should be added to or removed from the cell state. In Fig. 2 illustrate the internal structure of the LSTM cell is shown, which is used as a basic component for processing sequences in time-series data.

The input gate controls how much new information is added to the cell state and defined as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

where i_t represents the input gate, W_i the weight matrix, h_{t-1} the previous hidden state, x_t the current input, and b_i the bias. The forget gate determines how much of the past information is retained and is given by:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

where f_t is the forget gate, W_f the weight matrix, and b_f the bias term. The cell state update equation, which integrates both the forget and input gates, is expressed as:

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (7)$$

where C_t is the cell state at time t , C_{t-1} is the previous cell state, and W_c the weight matrix associated with the cell state update. Finally, the output gate determines the next hidden state that is passed forward and is represented as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

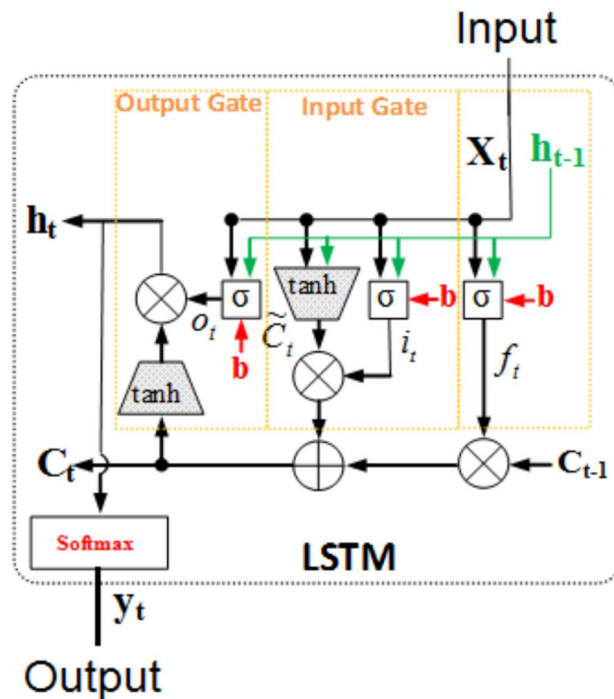


Fig. 2. The internal structure of the LSTM cell.

$$h_t = o_t * \tanh C_t \quad (9)$$

where h_t is the new hidden state, influenced by the output of the cell state C_t , and o_t is the output gate. By integrating LSTMs into our hybrid CNN-LSTM model, we aim to enhance the model's ability to capture intricate temporal patterns in solar irradiance, thereby contributing to more accurate solar energy forecasting in the region of Ghardaïa, Algeria.

The framework of the hybrid CNN-LSTM

The proposed CNN-LSTM hybrid model is designed for short-term forecasting of Direct Normal Irradiance (DNI) using meteorological data from Ghardaïa, Algeria. This framework combines the spatial feature extraction capability of CNNs with the temporal sequence learning of LSTMs, ensuring improved forecasting accuracy. CNNs identify spatial dependencies by applying convolutional filters to meteorological variables such as temperature variations, cloud coverage, and humidity distributions, which significantly influence DNI. By utilizing localized feature extraction, CNNs detect patterns such as temperature gradients preceding an increase in DNI or cloud formations reducing solar radiation, while also reducing data dimensionality and minimizing noise. The extracted spatial features are then passed to the LSTM component, which is optimized for learning long-term dependencies in solar irradiance fluctuations. LSTMs utilize memory cells and gating mechanisms to selectively retain relevant past information while discarding insignificant variations, addressing the vanishing gradient problem often encountered in deep sequence models. This allows the LSTM to recognize key temporal trends, including seasonal changes in solar radiation, daily DNI fluctuations due to sunrise and sunset, and short-term variations caused by atmospheric disturbances. The CNN-LSTM framework is trained on historical datasets, with hyperparameters optimized to minimize forecasting errors. The final model outputs precise DNI predictions, enhancing solar resource management and energy planning in arid and semi-arid regions. The design of this hybrid framework aligns with recent advancements in system optimization for renewable energy. For example,¹⁵ demonstrated how self-regulation and flow rate optimization enhance biohydrogen production, underscoring the value of adaptive architectures like CNN-LSTM for handling dynamic solar data. Similarly,^{39,40} emphasized the role of energy transfer mechanisms in microbial systems, which inspired our approach to capturing temporal dependencies in irradiance data. The algorithm 1 describes a CNN-LSTM model designed to forecast DNI for very short-term solar irradiance prediction.

The SVR model designing

The SVR model is an effective and powerful ML technique based on the statistical learning theory introduced as a nonlinear development of support vector machines. SVM is one of the most used and effective classification and regression algorithms that perform the classification and/or regression tasks based on the principles of dimensions. It separates the two classes, or predicts numerical values, by finding the widest margin between the observations, reducing the penalty of violating the margin, and ensuring that the forecast error does not exceed an error tolerance. SVR is a non-probabilistic model, and it is included in the so-called "kernel method," which operates directly with the input vectors in a high-dimensional implicit feature space, as a function of the distance

Inputs:

. Put $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ as training dataset, where X_i are meteorological feature vectors and Y_i are the corresponding direct normal irradiance (DNI) values.

- . **Model Parameters:** Initialize CNN and LSTM model parameters, including:
 - Convolutional layer parameters: kernel size, number of filters, and activation functions.
 - LSTM layer parameters: number of units and activation functions.

Initialization:

- . Assign initial values to the CNN layer weights, bias terms, and LSTM weights.
- . Initialize the learning rate and the set early stopping criteria based on the validation loss.

Processing:**1. CNN Feature Extraction:**

- **Convolution Operation:** For each input X_i , perform the convolution operator using filters W_c to extract spatial features.

$$F_{conv} = \text{ReLU}(W_c * X + b_c)$$

- **Pooling Layer:** Apply max or average pooling to reduce dimensionality and retain important features.

- **Flattening:** Collapse the output of the final convolutional layer into the one-dimensional vector F_{flat} to be fed into the LSTM.

2. LSTM Temporal Sequence Modeling:

- **Sequence Processing:** Pass the flattened features F_{flat} through the LSTM layer to capture temporal dependencies.

- **LSTM Gate Mechanisms:**

. **Forget Gate:**

$$f_t = \sigma(W_f[h_{t-1}, F_{flat}] + b_f)$$

. **Input Gate:**

$$i_t = \sigma(W_i[h_{t-1}, F_{flat}] + b_i), \quad \tilde{C}_t = \tanh(W_c[h_{t-1}, F_{flat}] + b_c)$$

. **Output Gate:**

$$o_t = \sigma(W_o[h_{t-1}, F_{flat}] + b_o)$$

. **Cell State Update:**

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

. **Hidden State Update:**

$$h_t = o_t \cdot \tanh(C_t)$$

3. Output Layer:

. Use the final LSTM hidden state h_t to make the DNI prediction according to the following expression:

$$\hat{Y} = W_y \cdot h_t + b_y$$

End for**Outputs:**

- . **Predicted DNI Values:** $\hat{Y}$, the forecasted solar irradiance values.

Algorithm 1. Algorithm corresponding the hybrid CNN-LSTM model.

from a prescribed center located through a kernel function by building a map of features. This space is designed to facilitate the prediction task of regression, whereas in the original feature space, the number of samples is very large. This method has shown remarkable forecasting performance, proving to be a strong generalization tool that does not need to know in advance the distribution of the error, as well as being able to easily manage small datasets with high computational efficiency. The aim is to find the hypothesis that best approximates the functional relationship between input vectors (independent variable) and dependent scalar values (response) to propose a refined forecast^{34–36}.

Design of ANNs

In general, the ANNs are computational models inspired by biological neural systems, mimicking the synaptic connections found in the human brain. These systems learn tasks or solve problems through adaptation, optimization, or experiential learning rather than following predefined rules. Similar to the human brain, ANNs represent knowledge through the weighted connections between nodes, analogous to neuronal connections. ANNs are particularly well-suited for modeling complex, non-linear relationships and are typically trained using supervised learning methods, wherein the model makes statistical inferences or decisions based on input data. By incorporating new observations, ANNs can iteratively improve their performance on specific tasks. These networks are designed to perform specific functions and can generalize across diverse types and varieties of features. Over the past decade, ANN models have demonstrated significant promise in various applications. Notably, FFBP and CFBP models have achieved high levels of success in data classification and recognition tasks. In FFBP, data flows unidirectionally from input to output through fully connected layers, with each neuron computing activations based on the weighted sum of its inputs:

$$a^{(l)} = f(W^{(l)}a^{(l-1)} + b^{(l)}) \quad (10)$$

and training is performed using backpropagation, updating weights via the following expression

$$W^{(l)} \leftarrow W^{(l-1)} - \eta \frac{\partial E}{\partial W^{(l)}} \quad (11)$$

Here, the CFPB enhances the FFBP by adding convolutional layers for automated feature extraction, where feature maps are computed using the following expression.

$$z_{i,j,k} = \sum_{m,n} X_{i+m,j+n} \cdot K_{m,n,k} + b_k \quad (12)$$

followed by pooling layers to reduce dimensions. Both architectures minimize a loss

$$E = \frac{1}{n} \sum_{i=1}^N L(y_i, \hat{y}_i) \quad (13)$$

through gradient descent, with FFBP excelling in structured data tasks and CFPB excelling in spatially hierarchical data like images.

Results and discussions

In this section, we present the performance of the proposed hybrid CNN-LSTM model for time series prediction and compare it with other models, including SVR, CFPB, and FFBP models. The performance evaluation starts with analyzing the correlation between SR and the input meteorological variables, followed by a discussion on the predictive performance of all models. Figure 3 illustrates the workflow for SR forecasting using machine learning models.

Correlation analysis

Figure 4 depicts the relationship between SR and the key meteorological variables, such as maximum temperature (T_{max}), minimum temperature (T_{min}), wind speed (WS), time, humidity (Hum), and pressure (Pr). According to Fig. 4, there are strong positive correlations between SR and both T_{max} and T_{min} , indicating that higher SR values are associated with higher temperatures. Additionally, SR shows a positive correlation with time, reflecting its diurnal pattern. Conversely, SR has negative correlations with WS , Hum , and Pr , suggesting that higher SR values are linked to lower wind speeds, lower humidity, and decreased atmospheric pressure. These relationships highlight the complex interactions between SR and meteorological factors, which are essential for feature selection in predictive modeling.

Based on the observed positive linear relationship between the current SR measurement $y(i)$ and its lagged value $y(i-1)$ as shown in the scatter plots, the proposed architecture (illustrated in Fig. 5) can be expressed mathematically by

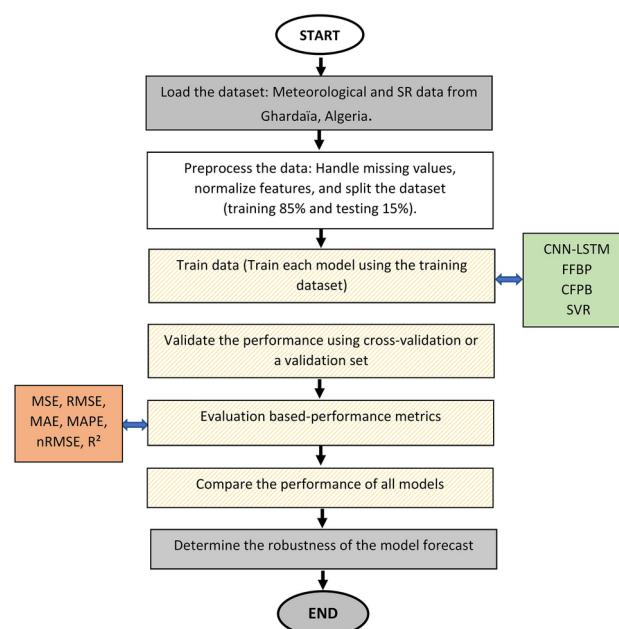


Fig. 3. Flowchart for SR forecasting using machine learning models.

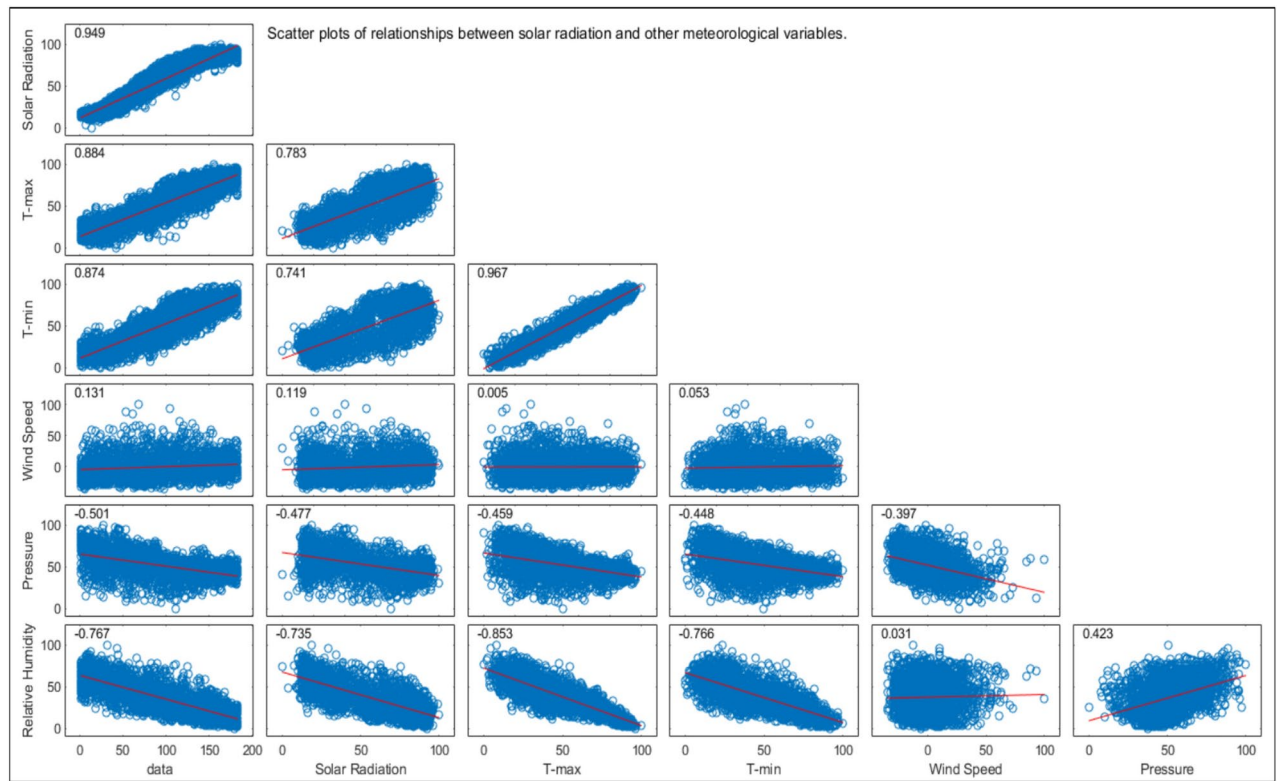


Fig. 4. Correlation Matrix Between Solar Radiation (SR) and Meteorological Variables (Temperature, Humidity, Wind Speed) in Ghardaia (2017–2023).

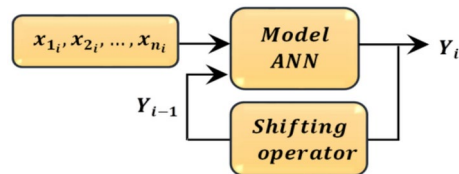


Fig. 5. Proposed ANN architecture incorporating lagged SR and meteorological features for prediction step.

$$y(i) = f(y(i-1), x_1, x_2, \dots, x_n) \quad (14)$$

where $y(i)$ and $y(i-1)$ are, respectively, the current SR measurement and the lagged SR measurement, which presents the one-time step before occurring $y(i)$. Also, $x_1, x_2, \dots, x_n$ are additional input features, including T_{max} , T_{min} , WS , Hum and pressure Pr . Finally, $f(\cdot)$ is the nonlinear mapping function learned by the proposed Model ANN architecture. Moreover, Eq. (14) indicates that the proposed model utilizes the temporal dependency that is captured through the lagged value $y(i-1)$. It simultaneously utilizes the influence of meteorological features ($x_1, x_2, \dots, x_n$) to predict $y(i)$. Also, the nonlinear function $f(\cdot)$, implemented via the ANN model with a shifting operator, effectively captures the observed positive linear relationship and complex interactions between the inputs and the target SR values. This approach highlights the importance of incorporating both temporal dynamics and environmental variables for accurate SR forecasting. The proposed ANN Architecture incorporating lagged SR and meteorological features for prediction is shown in Fig. 5.

Figure 6 gives a scatter plot collection illustrating the relationship of time-varying SR measures $y(i)$ and their lagged values $y(i-1)$, $y(i-2)$, $y(i-3)$, $y(i-4)$. More generally, this is known as a scatter plot matrix of the correlation coefficients. Thus, the plots of current SR measure and its past measures illustrate the temporal dependencies that may be present. The scatter plots show that there is a positive linear relationship between $y(i)$ and its lagged versions; hence, high values of current SR are usually associated with high values of past measurements. Such consistency in trends reflects the strong temporal relationship within the SR data. In addition, the coefficient of determination, R^2 seen in the upper-right corner of each plot represents the percentage of variance in $y(i)$ explained by each measurement. For example, the R^2 value of 0.986 for the relationship of $y(i)$ to $y(i-1)$ indicates that 98.6% of the variability in $y(i)$ can be explained by its most recent past value. These temporal correlations and R^2 values underline the predictive relationships between current

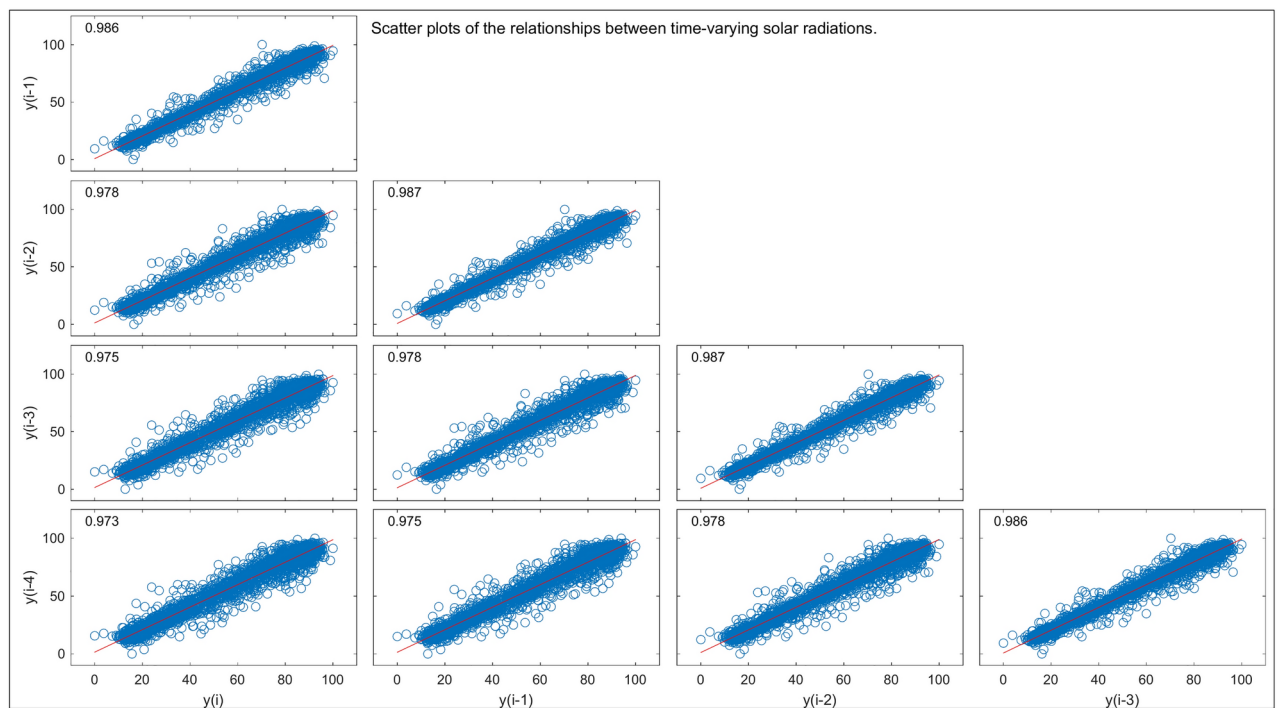


Fig. 6. Scatter plots depicting temporal correlations between SR measurements and their lagged values.

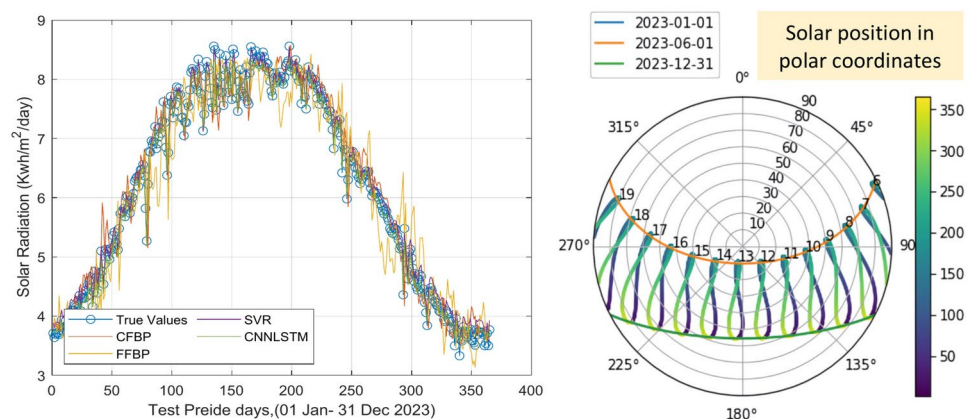


Fig. 7. Actual vs. Predicted Solar Irradiance (SR) Values for CNN-LSTM, SVR, CFBP, and FFBP Models (2020–2023).

and lagged solar radiation, which imply that it is important to include lagged variables in time series forecasting models for an effective capture of the dynamics of SR over time.

Model performance comparison

In this section, we compare the performance of the proposed hybrid CNN-LSTM model with other prediction models, including SVR, CFBP, and FFBP, in terms of their accuracy in forecasting SR values over a one-year period (See Fig. 7).

In Fig. 7, the performance comparison of the proposed hybrid CNN-LSTM model with the SVR, CFBP, and FFBP models is presented. The SR predictions are plotted alongside the actual observed SR measurements. This figure demonstrates that the CNN-LSTM model closely follows the observed solar irradiance trends, particularly during periods of high variability. This highlights the model's ability to capture complex temporal dependencies, which is critical for accurate short-term forecasting. In contrast, the SVR and FFBP models show larger deviations, especially during peak solar irradiance periods.

Figure 8 presents a Taylor diagram comparing the predictive performance of five models: CNN-LSTM, SVR, CFBP, FFBP, and the actual SR values. The diagram evaluates the models based on three key metrics: standard deviation, correlation coefficient, and centered root mean square difference (CRMSD). The 'True Values' point

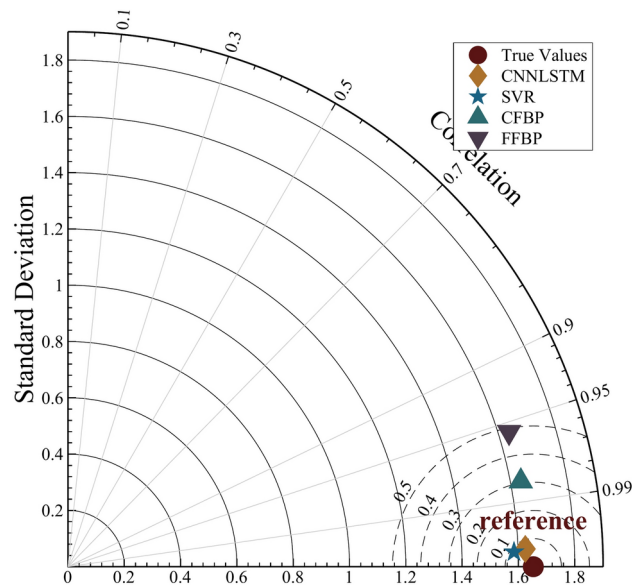


Fig. 8. Taylor diagram comparing model performances of CNN-LSTM, SVR, CFBP and FFBP with observed SR.

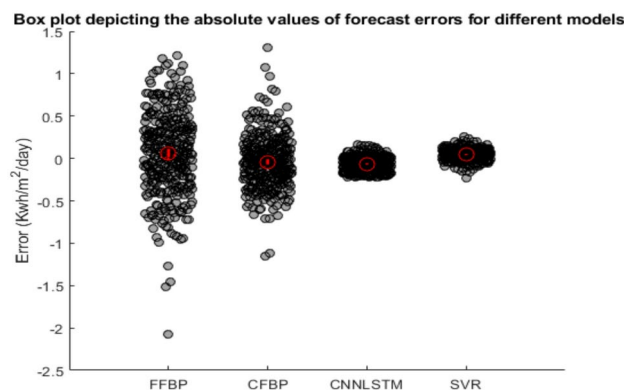


Fig. 9. Box plot and error diagrams for FFBP, CFBP, CNN-LSTM, and SVR approaches.

serves as the reference, with a standard deviation and correlation coefficient of 1.0, indicating perfect agreement. The CNN-LSTM model is closest to the true values, exhibiting high correlation coefficients and low CRMSD, signifying strong agreement. In contrast, the SVR, CFBP, and FFBP models show greater discrepancies from the true values, with lower correlation coefficients and higher CRMSD. This indicates that the CNN-LSTM model significantly outperforms the others in SR prediction.

Figure 9 presents a box plot comparing the absolute forecast errors of four models: FFBP, CFBP, CNN-LSTM, and SVR. The box plot displays the distribution of forecast errors, highlighting the median error, interquartile range, and minimum and maximum values for each model. The CNN-LSTM model shows the smallest median error and the narrowest interquartile range, indicating the most accurate and consistent forecasts. In comparison, the SVR model has a slightly higher median error and a wider interquartile range. The FFBP model exhibits the largest median error and the widest interquartile range, indicating higher and more variable forecast errors. The CFBP model performs better than FFBP but not as well as CNN-LSTM. Although outliers are present, indicating occasional large forecast errors for all models, the CNN-LSTM model remains the most reliable and accurate overall.

Comparative performance evaluation using error metrics

In this section, the comparative performance of different SR prediction models is evaluated using multiple error metrics, including MSE, RMSE, MAE, MAPE, and nRMSE. These metrics are commonly used to assess the accuracy of predictive models by quantifying the discrepancies between predicted and observed values⁴¹. Here, the MSE calculates the average squared difference between the predicted and actual values. It is defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (15)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations. Also, the RMSE presents the square root of the MSE. It provides an error value in the same units as the predicted data, making it easier to interpret. It is defined as^{42–44}:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

Moreover, the MAE presents the average of the absolute differences between the actual and predicted values. It is particularly useful when all errors are treated equally. It is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

Similarly, the MAPE calculates the percentage error between predicted and actual values. It is often used to express the accuracy of a model in terms of percentage. It is expressed by^{45,46}:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (18)$$

Finally, the nRMSE normalizes the RMSE by dividing it by the range or mean of the observed values, making it scale-independent. It is defined as^{47,48}:

$$\text{nRMSE} = \frac{\text{RMSE}}{\text{Range or Mean of } y_i} \quad (19)$$

These metrics provide different perspectives on model performance. MSE and RMSE penalize larger errors more heavily, whereas MAE treats all errors equally. MAPE is useful for understanding the error in percentage terms, and nRMSE offers a normalized error value that accounts for the scale of the data^{48,49–52}. The use of these metrics helps in selecting the most appropriate model for SR prediction based on the specific characteristics of the data and the desired application. The table presents a comparative performance analysis of four SR prediction models: FFBP, CFBP, SVR, and CNN-LSTM across five error metrics: MSE, RMSE, MAPE, MAE, and nRMSE⁵³.

The CNN-LSTM model consistently outperforms the other models, achieving the lowest values in all metrics, indicating its superior predictive accuracy. Specifically, it has the smallest MSE (0.0069), RMSE (0.0833), MAPE (1.1807), MAE (0.0679), and nRMSE (0.0135). These results are significantly lower than values reported in recent studies. For example, Ahmed et al.⁶ reported an MSE of 0.012 for hybrid models, and Kumari and Toshniwal⁵ achieved an RMSE of 0.15 for standalone LSTM architectures. The superior performance of the CNN-LSTM model can be attributed to its ability to capture both spatial and temporal dependencies in solar irradiance data, a feature not fully exploited by standalone or less integrated architectures.

The MSE calculates the average squared difference between the predicted and actual values, penalizing larger errors more heavily. The RMSE, derived as the square root of MSE, provides an error value in the same units as the predicted data, facilitating interpretation. The MAE, which averages absolute differences, treats all errors equally, while the MAPE expresses errors as percentages, offering intuitive insights into model accuracy. Finally, the nRMSE normalizes RMSE by the data range, enabling scale-independent comparisons.

As shown in Table 3, the CNN-LSTM model's MSE (0.0069) is 42.5% lower than the hybrid models reported by Ahmed et al.⁶, and its RMSE (0.0833) is 44.5% lower than the standalone LSTM results from Kumari and Toshniwal⁵. These improvements highlight the effectiveness of combining CNNs for spatial feature extraction (e.g., cloud cover patterns) with LSTMs for temporal modeling (e.g., diurnal cycles). In contrast, the FFBP model exhibits the highest error values across all metrics (MSE = 0.2403, RMSE = 0.4902), reflecting its inability to handle the non-linear and dynamic nature of solar irradiance data. The SVR and CFBP models show intermediate performance, with SVR achieving an RMSE of 0.1073, which, while competitive, still lags behind the CNN-LSTM.

The radar chart in Fig. 10 and Taylor diagram in Fig. 8 further corroborate these findings. The CNN-LSTM model's proximity to the "True Values" reference in the Taylor diagram (high correlation, low CRMSD) underscores its alignment with observed data. Additionally, the narrow interquartile range in the box plot (Fig. 9) confirms the model's consistency, with minimal outliers compared to FFBP and CFBP. These results demonstrate that the CNN-LSTM's hybrid architecture is uniquely suited for solar irradiance forecasting in regions like Ghardaia, where both spatial variability (e.g., sudden cloud cover) and temporal trends (e.g., seasonal changes) significantly impact solar potential.

Figure 11 presents a radar chart comparing the performance of four models: CNN-LSTM, SVR, CFBP, and FFBP across five evaluation metrics: MSE, RMSE, MAPE, MAE, and nRMSE. The CNN-LSTM model consistently outperforms the others, achieving the lowest values across all metrics, indicating superior accuracy and reliability. The SVR model shows comparable but slightly inferior performance, while the CFBP model

	FFBP	CFBP	SVR	CCN-LSTM
MSE	0.2403	0.0957	0.0115	0.0069
RMSE	0.4902	0.3093	0.1073	0.0833
MAPE	6.7293	4.0022	1.7752	1.1807
MAE	0.3865	0.2355	0.0862	0.0679
nRMSE	0.0793	0.0501	0.0174	0.0135

Table 3. Error Metrics for CNN-LSTM, SVR, CFBP, and FFBP Models.

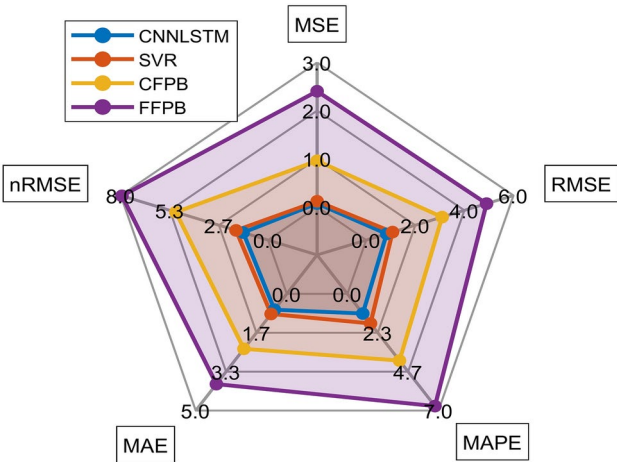


Fig. 10. Comparison of prediction errors for different SR models based on various metrics.

exhibits moderate results. In contrast, the FFPB model performs the worst, with the highest error values in all metrics. This comparison highlights the robustness of the CNN-LSTM model for the given task

Table 4 presents the coefficient of determination (R^2) values for four models: FFBP, CFPB, SVR, and CNN-LSTM. The R^2 values highlight the predictive accuracy of each model, with higher values indicating better performance. The FFBP model shows a reasonable fit with $R^2=0.95643$, while the CFPB model improves to $R^2=0.98267$, reflecting stronger accuracy. The SVR model achieves near-perfect performance with $R^2=0.99845$, and the CNN-LSTM model outperforms all others with $R^2=0.99925$, indicating an almost perfect correlation between predicted and actual values. These results confirm the superiority of the CNN-LSTM and SVR models for this predictive task.

Conclusion

This study evaluated the performance of FFBP, CFBP, SVR, and the hybrid CNN-LSTM model for solar radiation (SR) prediction using statistical metrics and visual analytics. The CNN-LSTM model outperformed all other methods, achieving the lowest errors (MSE=0.0069, RMSE=0.0833, MAE=0.0679, MAPE=1.18%, nRMSE=0.0135) and the highest coefficient of determination ($R^2=0.99925$), demonstrating near-perfect alignment between predicted and actual values. These results surpass recent benchmarks in hybrid models and standalone architectures, validating the effectiveness of combining spatial feature extraction (CNNs) and temporal sequencing (LSTMs). Visual tools, including radar charts, Taylor diagrams, and scatter plots, confirmed the model’s robustness in capturing complex spatial–temporal dependencies inherent in SR data. While the SVR model showed strong performance, it lagged behind the CNN-LSTM, and the FFBP/CFBP models exhibited limitations in modeling temporal dynamics.

The CNN-LSTM’s hybrid architecture positions it as a reliable tool for solar energy forecasting, particularly in regions like Ghardaia, Algeria, where variable weather conditions challenge conventional methods. Practical applications include optimizing solar farm operations through short term forecasts for energy storage management, reducing reliance on fossil fuels during low-irradiance periods, and guiding policymakers in renewable energy planning for grid stability.

Future research should integrate additional meteorological variables such as humidity and wind speed to refine predictions and explore hybrid architectures with attention mechanisms to enhance long-term forecasting accuracy. By advancing these models, we can better support global efforts to transition toward sustainable energy systems, particularly in sun-rich regions where precise solar forecasting is critical for climate resilience and energy security.

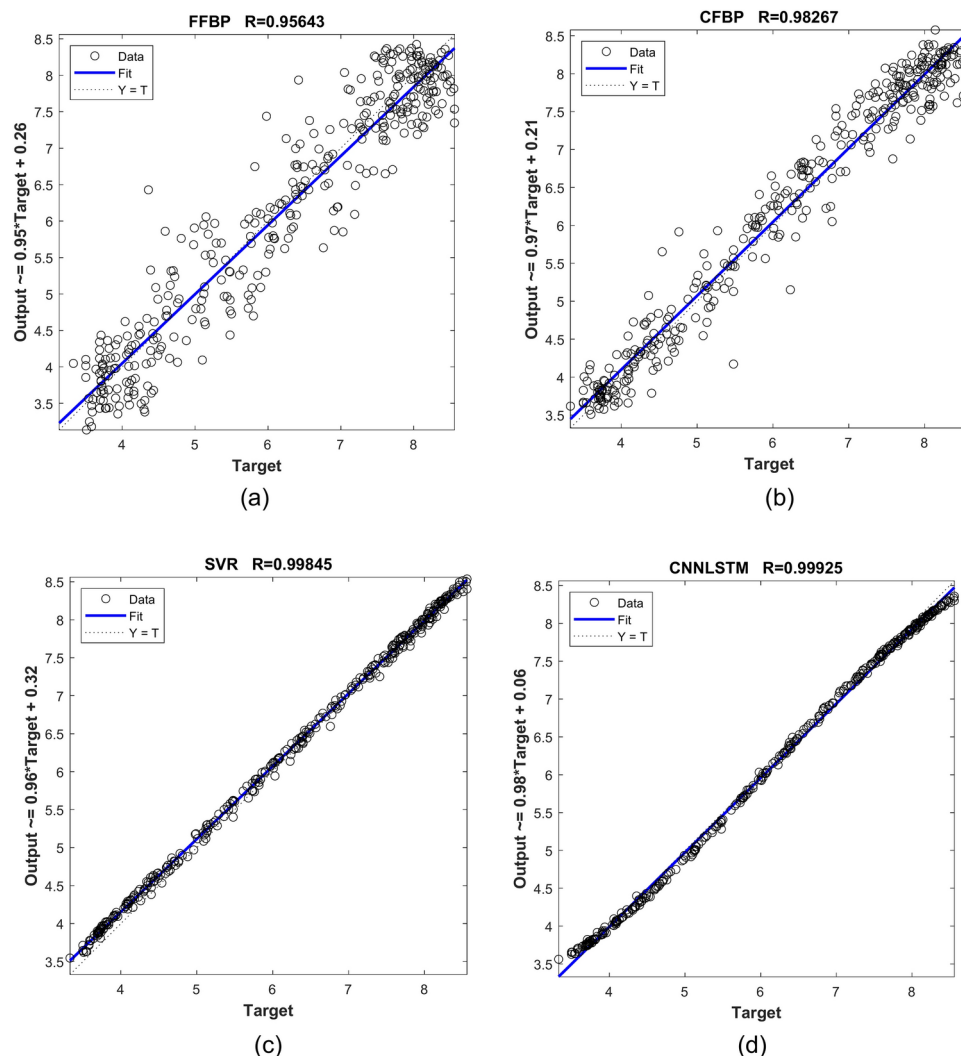


Fig. 11. Radar chart comparison of model performance metrics. (a) FFBP Model; (b) CFBP Model; (c) SVR Model and (d) CNN-LSTM Model.

Model	FFBP	CFBP	SVR	CNNLSTM
coefficient of determination R^2	0.95643	0.98267	0.99845	0.99925

Table 4. Coefficient of determination (R^2) for model performance comparison.

Data availability

The data used and/or analyzed during the current study are available from co-author Dr. Abdelaziz Rabehi (rab_ghi@hotmail.fr) on reasonable request.

Received: 19 January 2025; Accepted: 12 March 2025

Published online: 02 May 2025

References

- Hassan, Q. et al. The renewable energy role in the global energy transformations. *Renew. Energy Focus* **48**, 100545 (2024).
- Sovacool, B. K., Baum, C. M. & Low, S. Climate protection or privilege? A whole systems justice milieu of twenty negative emissions and solar geoengineering technologies. *Polit. Geogr.* **97**, 102702 (2022).
- Qadir, Z. et al. Predicting the energy output of hybrid PV-wind renewable energy system using feature selection technique for smart grids. *Energy Rep.* **7**, 8465–8475 (2021).
- Irfan, M. et al. Radiative forcing on climate change: assessing the effect of greenhouse gases on energy balance of Earth. *Adv. Technol. Dev. Greenhouse Gases: Emission, Capture Convers.* **2024**, 137–167 (2024).
- Kumari, P. & Toshniwal, D. Deep learning models for solar irradiance forecasting: a comprehensive review. *J. Clean. Prod.* **318**, 128566 (2021).

6. Ahmed, R. et al. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. *Renew. Sustain. Energy Rev.* **124**, 109792 (2020).
7. Zhou, Y. et al. A review on global solar radiance prediction with machine learning models in a comprehensive perspective. *Energy Convers. Manage.* **235**, 113960 (2021).
8. Krishnan, N. et al. How solar radiance forecasting impacts the utilization of solar energy: a critical review. *J. Clean. Prod.* **388**, 135860 (2023).
9. Guermoui, M. et al. An analysis of case studies for advancing photovoltaic power forecasting through multi-scale fusion techniques. *Sci. Rep.* **14**(1), 6653 (2024).
10. AlKandari, M. & Ahmad, I. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Appl. Comput. Inf.* **20**(3/4), 231–250 (2024).
11. Messeri, L. & Crockett, M. J. Artificial intelligence and illusions of understanding in scientific research. *Nature* **627**, 49–58 (2024).
12. Tehrani, A. A. et al. Predicting SRin the urban area: a data-driven analysis for sustainable city planning using artificial neural networking. *Sustain. Cities Soc.* **100**, 105042 (2024).
13. Shouman, E. R. M. Solar power prediction with artificial intelligence. *Adv. Solar Photovoltaic Energy Syst.* **2024**, 456 (2024).
14. Malakouti, S. M. Estimating the output power and wind speed with ML methods: a case study in Texas. *Case Stud. Chem. Environ. Eng.* **7**, 100324 (2023).
15. Malakouti, S. M. et al. Advanced techniques for wind energy production forecasting: Leveraging multi-layer perceptron, Bayesian optimization, ensemble learning, and CNN-LSTM models. *Case Stud. Chem. Environ. Eng.* **10**, 100881 (2024).
16. Sweeney, C. et al. The future of forecasting for renewable energy. *Wiley Interdiscipl. Rev.: Energy Environ.* **9**, 2 (2020).
17. Ahmad, T., Zhang, H. & Yan, B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustain. Cities Soc.* **14**, 10 (2020).
18. Kumar, D. S. et al. Solar irradiance resource and forecasting: a comprehensive review. *IET Renew. Power Gener.* **14**(10), 1641–1656 (2020).
19. Hanifi, S., Liu, X., Lin, Z. & Lotfian, S. A critical review of wind power forecasting methods—past, present and future. *Energies* **13**, 15 (2020).
20. Mohammed Eisa, H. *Streamflow Forecasting Using Artificial Neural Networks (Case Study)* 203 (2024). <https://core.ac.uk/download/pdf/71673066.pdf>.
21. Voyant, C. et al. *Predictability of PV Power Grid Performance on Insular Sites Without Weather Stations: Use of Artificial Neural Networks* (2010).
22. Carles-Martí-Pérez, P. & Gasque-Albalade, M. Improvement of temperature-based ANN models for solar radiance estimation through exogenous data assistance. *Energy Convers. Manage.* **52**, 990 (2011).
23. Hoyos-Gómez, L., Ruiz-Muñoz, J. & Ruiz-Mendoza, J. Short-term forecasting of global solar irradiance with incomplete data. *Appl. Energy* **307**, 118192 (2021).
24. Bou-Nassif, A., Soudan, B., Azzeh, M., Attilli, I. & Almulla, O. Artificial intelligence and statistical techniques in short-term load forecasting: a review. *Appl. Energy* **307**, 408 (2021).
25. Jiang, X., Jiang, M. & Zhou, Q. Day-ahead PV power forecasting based on MSTL-TFT. *Int. Rev. Model. Simul.* **14**, 408 (2023).
26. Nawab, F. et al. Solar irradiation prediction using empirical and artificial intelligence methods: a comparative review. *Heliyon* **9**, 17038 (2023).
27. Benali, L. et al. Localized feature engineering in hybrid models for North African solar datasets. *Solar Energy* **267**, 112–125 (2023).
28. Chen, H. et al. Satellite-cloud dynamics fusion for sub-hourly irradiance prediction using deep learning. *Appl. Energy* **357**, 122543 (2024).
29. Shazid-Islam, M. et al. *Location Agnostic Source-Free Domain Adaptive Learning to Predict Solar Power Generation* (Springer, 2024).
30. Wang, J. & Li, X. CNN-LSTM vs. transformers: a comparative study in short-term solar forecasting. *IEEE Trans. Sustain. Energy* **15**(2), 876–889 (2024).
31. Ibrahim, S. et al. Explainable AI for renewable energy: interpreting feature contributions in solar forecasting models. *Energy AI* **16**, 100345 (2024).
32. Zhang, Y. et al. Spatio-temporal transformer networks for solar irradiance forecasting in arid regions. *Renew. Energy* **215**, 109002 (2023).
33. Kumar, R. et al. Attention-augmented CNN-LSTM for high-resolution solar forecasting. *Energy Convers. Manage.* **301**, 118045 (2024).
34. Santos, O. et al. Solar irradiance forecasting using dynamic ensemble selection. *Appl. Sci.* **12**(7), 3510 (2022).
35. Khelifi, R. et al. Short-term PV power forecasting using a hybrid TVF-EMD-ELM strategy. *Int. Trans. Electr. Energy Syst.* **2023**(1), 6413716 (2023).
36. Beniwal, M., Singh, A. & Kumar, N. Forecasting long-term stock prices of global indices: a forward-validating genetic algorithm optimization approach for support vector regression. *Appl. Soft Comput.* **145**, 110566 (2023).
37. Wang, Y. et al. Cost prediction of building projects using the novel hybrid RA-ANN model. *Eng. Constr. Archit. Manage.* **31**(6), 2563–2582 (2024).
38. Ferkous, K. et al. A novel learning approach for short-term photovoltaic power forecasting—a review and case studies. *Eng. Appl. Artif. Intell.* **133**, 108502 (2024).
39. Malakouti, S. M. Babysitting hyperparameter optimization and 10-fold cross-validation to enhance the performance of ML methods in predicting wind speed and energy generation. *Intell. Syst. Appl.* **19**, 200248 (2023).
40. Malakouti, S. M. Improving the prediction of wind speed and power production of SCADA system with ensemble method and 10-fold cross-validation. *Case Stud. Chem. Environ. Eng.* **8**, 100351 (2023).
41. Al-Kahtani, N. et al. Digital health transformation in Saudi Arabia: a cross-sectional analysis using Healthcare Information and Management Systems Society’ digital health indicators. *Digital Health* **8**, 205520762211177 (2022).
42. Malakouti, S. M., Menhaj, M. B. & Suratgar, A. A. The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Clean. Eng. Technol.* **15**, 100664. <https://doi.org/10.1016/j.clet.2023.100664> (2023).
43. Guermoui, M. & Rabehi, A. Soft computing for solar radiation potential assessment in Algeria. *Int. J. Ambient Energy* **41**(13), 1524–1533 (2020).
44. Khelifi, R., Guermoui, M., Rabehi, A. & Lalmi, D. Multi-step-ahead forecasting of daily solar radiation components in the Saharan climate. *Int. J. Ambient Energy* **41**(6), 707–715 (2020).
45. Guermoui, M., Abdelaziz, R., Gairaa, K., Djemoui, L. & Benkacali, S. New temperature-based predicting model for global solar radiation using support vector regression. *Int. J. Ambient Energy* **43**(1), 1397–1407 (2022).
46. Guermoui, M., Rabehi, A., Gairaa, K. & Benkacali, S. Support vector regression methodology for estimating global solar radiation in Algeria. *Eur. Phys. J. Plus* **133**, 1–9 (2018).
47. Rabehi, A., Guermoui, M. & Lalmi, D. Hybrid models for global solar radiation prediction: a case study. *Int. J. Ambient Energy* **41**(1), 31–40 (2020).
48. Rabehi, A., Guermoui, M., Khelifi, R. & Mekhalif, M. L. Decomposing global solar radiation into its diffuse and direct normal radiation. *Int. J. Ambient Energy* **41**(7), 738–743 (2020).
49. Guermoui, M., Boland, J. & Rabehi, A. On the use of BRL model for daily and hourly solar radiation components assessment in a semiarid climate. *Eur. Phys. J. Plus* **135**(2), 1–16 (2020).

50. Rabehi, A., Guermoui, M., Djafer, D. & Zaiani, M. Radial basis function neural networks model to estimate global solar radiation in semi-arid area. *Leonardo Electron. J. Pract. Technol.* **27**, 177–184 (2015).
51. Rabehi, A., Rabehi, A. & Guermoui, M. Evaluation of different models for global solar radiation components assessment. *Appl. Solar Energy* **57**, 81–92 (2021).
52. Rabehi, A., Helal, H., Zappa, D. & Comini, E. Advancements and prospects of electronic nose in various applications: a comprehensive review. *Appl. Sci.* **14**(11), 4506 (2024).
53. Guermoui, M., Arrif, T., Belaid, A., Hassani, S. & Bailek, N. Enhancing direct normal solar Irradiation forecasting for heliostat field applications through a novel hybrid model. *Energy Convers. Manage.* **304**, 118189 (2024).

Author contributions

Boumediene Ladjal, Mohamed Nadour, Mohcene Bechouat, and Nadji Hadroug: conceptualization, methodology, software, visualization, investigation, writing- original draft preparation. Moussa Sedraoui, Abdelaziz Rabehi, Mawloud Guermoui, and Takele Ferede AGAJIE: Data curation, validation, supervision, resources, writing -review & Editing, Project administration, Funding Acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-94239-z>.

Correspondence and requests for materials should be addressed to T.F.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025