Energy Conversion and Management: X 27 (2025) 101072

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

Ultra-short term PV power forecasting under diverse environmental conditions: A case study of Norway

Opy Das [a], Dounia Dahlioui [a], Muhammad Hamza Zafar [a], Naureen Akhtar [a]
Syed Kumayl Raza Moosavi [a], Filippo Sanfilippo [a,b,*]

[a] Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway
[b] Department of Software Engineering, Kaunas University of Technology, Kaunas, 51368, Lithuania

ARTICLE INFO

Keywords:
Ultra short term power forecasting
Temporal data
PV power forecasting
Temporal attention

ABSTRACT

Accurate short-term solar forecasting is critical for power plant operations, grid balancing, real-time dispatching, automatic generation control, and energy trading. In Norway, where solar radiation is limited in winter and highly variable in summer, accurate predictions are essential. This study focuses on ultra-short-term forecasting of solar radiation and power output from a 37.8 kWp solar photovoltaic (PV) power plant at the University of Agder (UiA), Grimstad (58.335322° N, 8.577718° E), in southern Norway. We propose a novel forecasting model that integrates Spatial Attention, Temporal Attention, Self-Attention, CNN, and BiLSTM architectures to enhance prediction accuracy. Using a custom dataset collected from the UiA PV plant, the model's effectiveness was validated through comprehensive ablation studies and comparative analysis with state-of-the-art methods. The proposed model achieved a low RMSE of 0.162 kW using seven days of data, demonstrating its superiority in predicting short-term PV power outputs and associated uncertainties, outperforming conventional forecasting techniques.

1. Introduction

The global energy landscape is undergoing a rapid transformation as a result of the significant shift towards sustainable and renewable energy sources. The necessity of utilising green energy has become increasingly apparent because of the limited fossil resource reserves and the adverse environmental impacts of fossil fuels. Photovoltaic (PV) power has experienced rapid development as a result of the abundant source of solar energy, which has provided a variety of applications. As a result of its broad distribution and low power generation costs, the production of renewable energy has become increasingly appealing. Norway has installed a total of 299 MW of solar electricity [1]. However, the installation of solar power presents both significant opportunities and obstacles. The intermittent and unpredictable volatility of PV power output have resulted in significant issues that could potentially impact the power grid's ability to operate safely and consistently. In regions like Norway, where solar radiation is highly variable due to seasonal and weather-related factors, this intermittency may lead to sudden drops or spikes in power generation. Such fluctuations can cause issues including voltage instability, frequency deviations, difficulties in load balancing, and increased dependency on backup generation. Consequently, it is imperative to conduct comprehensive research, evaluate, and forecast the generation of solar power with accuracy [2].

There are two established approaches to solar power estimation. The first is direct measurement, which involves monitoring the actual output power of the solar system in real time. While this provides accurate, immediate data on system performance, it does not involve predictive modelling and thus is not considered a forecasting approach. The second is indirect prediction, which involves analysing solar radiation data to estimate future photovoltaic (PV) output power. This method falls within the domain of forecasting, as it relies on environmental variables to anticipate future energy production [3].

Photovoltaic (PV) generation is inherently intermittent and uncertain due to its dependence on variable solar irradiance. These fluctuations can occur across multiple time scales — from weeks and days to minutes and seconds — introducing significant challenges for system stability and reliability. This variability, especially in grid-connected PV systems, highlights the critical need for accurate forecasting to support grid management and planning [4].

Forecasting plays a crucial role in addressing these challenges. Medium- and long-term forecasting forms the basis for ensuring reliable grid operation and maintaining long-term stability, supporting strategic planning and large-scale decision-making processes [5]. In contrast, short-term forecasting helps in the daily management of power

* Corresponding author at: Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway.
E-mail address: filippo.sanfilippo@uia.no (F. Sanfilippo).

https://doi.org/10.1016/j.ecmx.2025.101072
Received 18 February 2025; Received in revised form 15 May 2025; Accepted 15 May 2025
Available online 11 June 2025

grid systems, assisting in operational planning and informed decision-making [6]. Ultra-short-term forecasting is particularly vital for immediate control, enhancing system efficiency, and preventing imbalances between production and demand caused by sudden fluctuations [7].

There are several modelling techniques available for forecasting PV (photovoltaic) power. The first approach can be classified as physical models which relies on atmospheric and meteorological data to predict solar irradiance. Two primary approaches in this category are Numerical Weather Prediction (NWP) models and radiative transfer models [8,9]. Although these models are based on the physical principles they often suffer from limitations such as low spatial and temporal resolution, leading to reduced accuracy, particularly for localised PV systems and short-term forecasts. Moreover, these models may not accurately capture the abrupt weather fluctuations, which are essential for precise photovoltaic power predictions. Statistical and machine learning approaches, including artificial neural networks (ANNs), have been utilised to tackle several of these obstacles. In [10], the authors compared different deterministic models for forecasting PV output power, where the forecasts were based on weather data provided by the forecasting service. The comparison was conducted using three- and five-parameter electric equivalent circuits and statistical models based on artificial neural networks (ANN). The study found that ANN, when combined with clear sky solar radiation, provided superior forecasting performance. However, these models can be prone to overfitting, particularly when trained on limited datasets, and may exhibit poor generalisation to unseen conditions. These limitations highlight the necessity for robust and adaptive forecasting methodologies, especially those that can effectively model intricate non-linear relationships and incorporate uncertainty across varying weather scenarios.

Statistical models have gained significant attention due to their strong theoretical foundation and strict statistical assumptions [11–13]. These models are typically effective for short- and medium-term forecasting [14,15]. Commonly used statistical prediction methods include regression analysis and time series analysis [16,17]. However, the high volatility, non-linear relationships, and complex temporal patterns inherent in PV systems often fail to meet the strong assumptions of statistical models, which can ultimately result in abnormal prediction errors [18]

The significant advancements in Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) have created immense opportunities in the fields of forecasting and prediction, particularly for handling high-dimensional, unstructured, or noisy datasets effectively. In the field of machine learning, algorithms such as classifiers, regressors, support vector machines, and random forests are commonly employed for modelling and forecasting. These methods often require extensive manual feature engineering and access to large datasets [19]. The authors in [20] discuss various machine learning methods, including Random Forests, Support Vector Machines (SVMs), and Gradient Boosting, for solar forecasting. Shallow machine learning techniques often suffer from limitations such as inadequate robustness and a tendency towards local optimisation [21]. Other machine learning and deep learning models are also presented for the PV forecasting i.e., XGBoost [22,23], vanilla LSTM [24], Seq2Seq with attention [25], adaptive neuro fuzzy logic! [26] and temporal convolutional network [27] . There are also some recent advancements in photovoltaic (PV) power forecasting that have demonstrated the effectiveness of integrating multi-task learning frameworks with nonlinear weather correction modules [28,29].

Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), address these issues by effectively capturing spatial and temporal relationships in datasets [30]. Advanced architectures like Long Short-Term Memory (LSTM) networks further enhance temporal prediction capabilities, providing a robust solution to these challenges [31–33]. Deep learning approaches, particularly neural networks, are increasingly utilised for historical data modelling and prediction. Due to its robust generalisation skills and superior non-linear mapping proficiency, Artificial Neural Networks (ANNs) are being increasingly employed for solar energy forecasting [34]. Diverse models utilising the ANN framework, including the Generalised Regression Neural Network (GRNN), Extreme Learning Machine Neural Network (ELMNN), Support Vector Machine (SVM), and additional methodologies with distinct attributes and particular benefits, have been established [35–37]. The authors provide a complete classification of solar PV output power forecasting approaches, focussing on model selection methodologies and performance metrics [38]. The difficulties connected with data collection, model creation, accuracy evaluation, and the integration of context change detection and incremental learning to improve prediction precision is discussed in [39].

Integrating scientific principles with deep learning has helped in the development of hybrid models that considerably increase the accuracy of solar energy predictions. Convolutional Neural Networks (CNNs) play an important role in these models by extracting non-linear characteristics and recognising invariant patterns in historical power data that are frequently observed over multiple time periods [40]. The capacity to record complicated linkages and temporal interactions improves PV systems' predictive performance. The attention mechanism is first introduced by Bahdanau et al. in [41], is a neural network component that enables models to dynamically focus on the most relevant parts of the input data when doing predictions. It has emerged as a valuable tool, allowing models to focus on certain deviations in the data, improving generalisation and prediction accuracy across a variety of applications [42]. The authors proposed an architecture that integrates dynamic inner partial least squares (DiPLS) and bidirectional long short-term memory (BiLSTM) networks, as well as attention mechanisms, to improve the interpretability and accuracy of solar power projections [43]. In the authors present a hybrid model that uses Temporal Convolutional Networks (TCN) and Gated Recurrent Units (GRU) with an attention mechanism to enhance global horizontal irradiance prediction significantly [44]. In [45]the authors have further highlighted the role of attention-based models, particularly Temporal Fusion Transformers (TFT), in energy forecasting tasks. Spectral clustering was utilised to extract interpretable temporal patterns which, when fed into a TFT model, provided high-accuracy forecasts along with feature attribution capabilities. Similarly in [46] a TFT-based model enhanced by interpretable feature selection (TIME) has been explored, enabling the identification of key variables influencing PV output which reinforce the importance of self attention in not only improving forecasting but also improving the model interpretability (see Table 1).

1.1. Contributions and paper organisation

The integration of photovoltaic (PV) systems into power grids presents significant technical challenges primarily due to the inherent variability and intermittency of solar energy production. This unpredictability, caused by constantly changing weather conditions, cloud movements, and diurnal cycles, can lead to voltage fluctuations, frequency deviations, and overall grid instability when PV penetration reaches high levels. Consequently, accurate ultra-short-term PV power forecasting — particularly at the one-minute interval resolution — has emerged as a critical requirement for effective grid operation, energy dispatch planning, and real-time balancing of supply and demand. Despite considerable research advances in PV forecasting, many current methodologies demonstrate substantial limitations in their ability to simultaneously capture both the complex spatial dependencies (such as cloud movement patterns across multiple array installations) and temporal dependencies (rapid changes in irradiance) that characterise solar power generation. Furthermore, while substantial research efforts have been directed towards short-term (hours ahead), medium-term (days ahead), and long-term (weeks to months ahead) PV power prediction, the ultra-short-term forecasting domain (seconds to minutes ahead) remains comparatively underexplored despite its critical importance

Table 1
Comparison of different PV forecasting techniques for PV power.

Ref	Year	Techniques used	Summary	Results
[28]	2025	Two-stage deep learning: Multi-Task Learning (MTL) + neural network weather correction (attention-based)	Presents a novel PV forecasting architecture that explicitly accounts for meteorological impacts. Stage 1: a custom MTL framework predicts PV power alongside multiple weather variables simultaneously. Stage 2: uses a neural network to perform a nonlinear Weather Correction, adjusting the initial PV forecast using the Stage 1 predicted weather details	RMSE: 1.1199, MAE: 0.4173
[47]	2025	Adaptive Masked Network (ASMNet) combining masking mechanism + Transformer for PV forecasting	Proposed ASMNet selectively covers less significant historical PV time steps to decrease redundancy and enhance ultra-short-term PV forecast precision. It learns adaptively which time steps are more crucial and concentrates attention on them.	RMSE: 0.2824, MAE: 0.0786
[48]	2024	EEMD-DARNN (Ensemble Empirical Mode Decomposition + Dual-stage Attention-based Recurrent Neural Network)	A hybrid model was developed that employs EEMD to decompose the univariate agricultural load sequence and utilises a Dual Attention-based RNN (DARNN) to enhance the accuracy of multi-step ultra-short-term forecasting. The decomposition effectively captures intricate features, while the attention mechanisms dynamically select essential temporal and spatial information.	RMSE: 12.298, MAE: 7.335, (15 min horizon)
[22]	2022	XGBoost-based probabilistic forecasting	Created a probabilistic model for estimating solar irradiance utilising XGBoost. The model sought to elucidate non-linear correlations between meteorological variables and solar irradiation, enhancing precision compared to traditional models. It additionally integrated uncertainty quantification to produce predicted intervals.	
[27]	2023	Hybrid forecasting framework combining TCN (Temporal Convolutional Networks) + feature engineering	Proposed a hybrid architecture based on TCN for projecting utility-scale photovoltaic output hours in advance. It amalgamates meteorological forecast data and historical photovoltaic data by feature selection and engineering to enhance multi-hour-ahead predictions. TCN adeptly captures sequential dependencies with extensive receptive fields and superior training speed compared to RNNs.	RMSE 83.84 W/m^2, MAE 54.38 W/m^2, (for 1-hour ahead)
[46]	2024	TIME-TFT model: Temporal Importance Model Explanation for feature selection + Temporal Fusion Transformer (deep learning)	Introduces a PV forecasting framework that is interpretable and consists of two stages. Initially, the most impactful input features are selected using the Temporal Importance (TIME) method, which results in interpretability-driven feature rankings. Subsequently, a Temporal Fusion Transformer (TFT) model is implemented to forecast PV power, with built-in attention and limiting to identify the features that are actually utilised.	

for instantaneous grid stability, especially in microgrids and islanded systems where rapid power fluctuations have more pronounced effects on system reliability.

To address these limitations, there is a pressing need for advanced models that leverage state-of-the-art deep learning techniques, including attention mechanisms and hybrid architectures, to improve forecasting accuracy and robustness in dynamic and complex scenarios. The contributions of this work are illustrated in graphical abstract and described as:

- Developed a novel forecasting model integrating Spatial Attention, Temporal Attention, Self-Attention, CNN, and BiLSTM architectures to improve PV power prediction accuracy.
- Conducted a comprehensive comparative analysis of the proposed model against state-of-the-art approaches, demonstrating its effectiveness in forecasting performance.
- Highlighted the role of attention mechanisms and hybrid deep learning architectures in enhancing the accuracy and robustness of PV power forecasting models.

2. Dataset description

The dataset utilised in this study comprises comprehensive meteorological and PV power generation data, collected over an extended period with 10,080 observations. The temporal resolution of the data is 1 min. The input features encompass eight crucial environmental parameters: Global Horizontal Irradiance (GHI) and Direct Normal Irradiance (DNI) measured in W/m^2, which represent the total solar radiation received on a horizontal surface and the direct solar radiation perpendicular to the sun's rays, respectively. Additional parameters include Wind Speed (WS) measured in m/s, Wind Direction (WDIR) in degrees from true north, Precipitation (Preci) in mm, Ambient Temperature (Temp) in °C, Atmospheric Pressure (Press) in hPa, and Relative Humidity (Hum) as a percentage. These parameters were systematically measured and averaged to capture dynamic environmental conditions that influence PV power generation. The target variable in the dataset is the actual power output from the PV system, measured in kilowatts (kW), representing the dependent variable for the forecasting model. This extensive dataset provides a robust foundation for developing and validating predictive models, as it captures the complex interactions between various meteorological parameters and their collective impact on PV power generation. The temporal resolution of the data collection ensures capture of both diurnal and seasonal variations in environmental conditions. The Global Horizontal Irradiance (GHI) and Direct Normal Irradiance (DNI) serve as primary indicators of solar resource availability, while the other meteorological parameters account for atmospheric attenuation and module efficiency factors. For instance, ambient temperature directly affects the PV module efficiency through

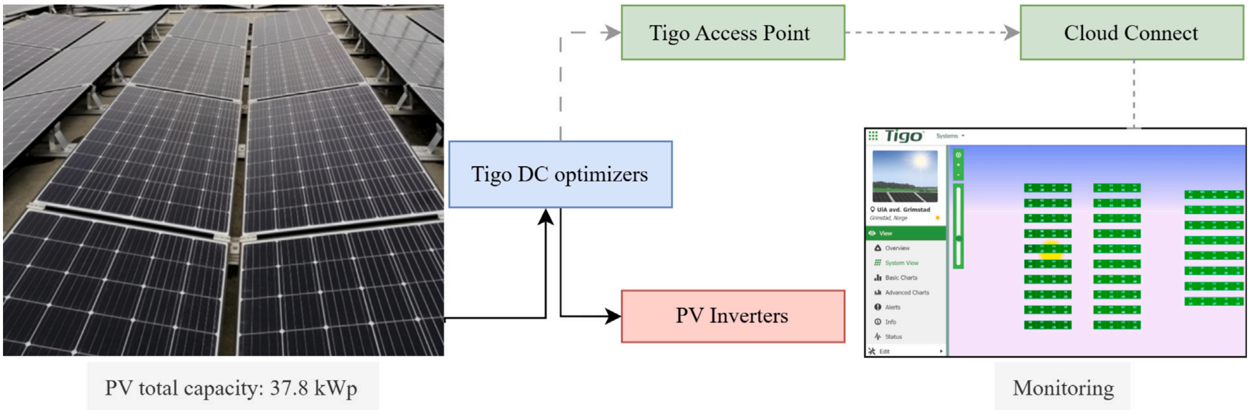

Fig. 1. Description of experimental setup and hardware used to collect data.

temperature coefficients, while wind speed influences the module operating temperature through convective cooling [49,50]. The inclusion of atmospheric pressure and relative humidity enables the modelling of aerosol and water vapour effects on solar radiation transmission.

2.1. Data acquisition

The system comprises three subsystems with different module technologies, all mounted at a 10° tilt in an East/West orientation. The first subsystem consists of 40 polycrystalline silicon PV modules, each with a capacity of 270 Wp. The second subsystem uses monocrystalline modules, each rated at 315 Wp, while the third subsystem also features monocrystalline modules with a capacity of 360 Wp each. Hence a total of 37.8 kWp for the entire system. The modules are connected in series within strings and linked to a total of 11 DC/AC inverters from Solax. Additionally, Tigo optimisers are installed on the backside of each module, enabling individual performance monitoring. An illustration of the complete system is provided in Fig. 1. Here we can elaborate on the experimental setup, description of the PV systems, The Hardware connected with the PV system to collect data.

2.2. Dataset analysis

Based on a comprehensive analysis of the PV power generation dataset, several significant patterns and relationships have emerged. The data reveals that GHI demonstrates the strongest positive correlation (0.748) with power output, reinforcing its fundamental role in PV generation. Direct Normal Irradiance and ambient temperature also show notable positive correlations of 0.455 and 0.383, respectively, indicating the positive impact on power generation efficiency. Interestingly, the relative humidity exhibits a negative correlation (−0.344), suggesting that increased atmospheric moisture might adversely affect power output. Wind direction and precipitation display weak negative correlations, while atmospheric pressure shows negligible correlation with power generation. The statistical analysis of power output characteristics reveals considerable variability in the system's performance. The power generation ranges from 0 to 5125 kW, with a mean output of 2211.63 kW and a substantial standard deviation of 2273.18 kW, indicating significant fluctuations in power production. The Global Horizontal Irradiance measurements span from −2.39 to 736.04 W/m^2, with an average of 109.85 W/m^2 and a standard deviation of 164.02 W/m^2. These statistics underscore the dynamic nature of solar power generation and its strong dependence on environmental conditions. The relatively high standard deviations in both power output and GHI measurements reflect the inherent variability in solar resource availability and the subsequent PV system response. This analysis not only confirms the paramount importance of solar radiation components in PV power generation but also highlights the complex interplay between various meteorological parameters affecting system performance.

3. Proposed technique

The proposed model architecture, as shown in Fig. 2, is designed to efficiently process time series data with multiple features such as past PV power, solar irradiance, temperature, and other meteorological data. The architecture incorporates attention mechanisms at multiple stages to capture spatial, temporal, and inter-temporal relationships in the data, followed by a Bi-LSTM layer to capture sequential dependencies.

3.1. Spatial attention mechanism

The spatial attention mechanism is responsible for learning which features are most important at each time step. In the context of PV power forecasting, different input features (such as solar irradiance, temperature, wind speed, etc.) contribute differently to the prediction. The spatial attention mechanism assigns a weight to each feature dynamically, based on its relevance to the current time step. Mathematically, let the input data $\mathbf{X} \in \mathbb{R}^{T\times F}$ represent the time series data, where T is the number of time steps and F is the number of features. The spatial attention mechanism computes the attention scores $\mathbf{A}_s \in \mathbb{R}^{T\times F}$, which reflect the importance of each feature at each time step. The attention scores are computed using a dense layer followed by a softmax activation function, which ensures that the attention scores are normalised across the feature dimension. The equation for calculating the attention weights is:

$$\mathbf{A}_s = \mathrm{softmax}(\mathbf{W}_s\mathbf{X} + \mathbf{b}_s) \tag{1}$$

where, $\mathbf{W}_s \in \mathbb{R}^{F\times F}$ is a learnable weight matrix, and $\mathbf{b}_s$ is a bias term. The softmax function normalises the attention weights such that:

$$\mathbf{A}_{s,i} = \frac{\exp(\mathbf{W}_s\mathbf{X}_i + \mathbf{b}_s)}{\sum_{j=1}^{F}\exp(\mathbf{W}_s\mathbf{X}_j + \mathbf{b}_s)} \tag{2}$$

The input data is then weighted by the attention scores, yielding the spatially attended input:

$$\mathbf{X}_s = \mathbf{X} \odot \mathbf{A}_s \tag{3}$$

Here, $\odot$ denotes element-wise multiplication, ensuring that the most important features are emphasised, while less relevant features are suppressed. This step helps the model focus on the most critical factors influencing PV power output, allowing it to learn the patterns in the data more effectively.

3.2. Temporal attention mechanism

While the spatial attention mechanism focuses on selecting the most important features, the temporal attention mechanism is designed to highlight the most relevant time steps in the sequence. In time series

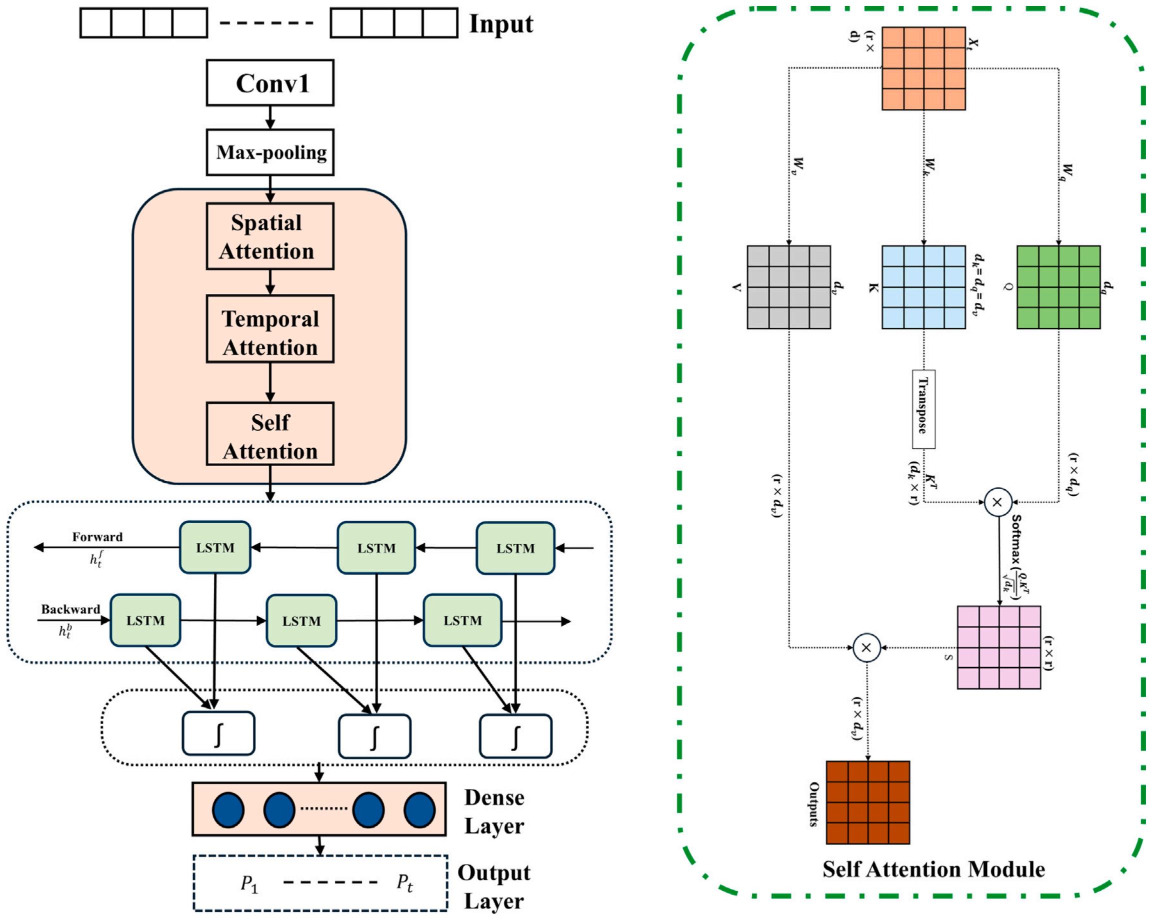

Fig. 2. Detailed architecture of proposed model.

data, not all time steps contribute equally to the forecasting task. Some time steps may have more predictive power based on factors such as diurnal patterns, cloud coverage variations, or sudden weather changes. Given the spatially attended data $\mathbf{X}_s \in \mathbb{R}^{T\times F}$, the temporal attention mechanism computes attention scores across the time dimension. This allows the model to assign higher weights to time steps that are more relevant for predicting future PV power generation.

The temporal attention weights are computed as:

$$\mathbf{A}_t = \text{softmax}(\mathbf{W}_t\mathbf{X}_s + \mathbf{b}_t) \tag{4}$$

where, $\mathbf{W}_t \in \mathbb{R}^{F\times 1}$ is a learnable weight matrix, and $\mathbf{b}_t$ is a bias term. The attention scores are applied across the time dimension, producing a weighted version of the input data that emphasises the most critical time steps:

$$\mathbf{X}_t = \mathbf{X}_s \odot \mathbf{A}_t \tag{5}$$

By allowing the model to focus on important time intervals, the temporal attention mechanism ensures that transient patterns, such as sudden cloud cover or temperature changes, are given more significance in the prediction process. This is particularly important in PV forecasting, where short-term weather changes can significantly impact power output.

3.3. Self-attention mechanism

Self-attention, also known as scaled dot-product attention, is crucial for capturing the relationships between different time steps and features in the input sequence. This mechanism allows the model to weigh the importance of each time step in relation to others, capturing long-range dependencies and interactions between time steps that might not be captured by traditional recurrent architectures. The self-attention mechanism operates by generating three vectors for each input: the query vector $\mathbf{Q}$, the key vector $\mathbf{K}$, and the value vector $\mathbf{V}$. These vectors are computed using linear transformations of the input data:

$$\mathbf{Q} = \mathbf{X}_t\mathbf{W}_q, \quad \mathbf{K} = \mathbf{X}_t\mathbf{W}_k, \quad \mathbf{V} = \mathbf{X}_t\mathbf{W}_v \tag{6}$$

where, $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{F\times d_k}$ are learnable weight matrices, and d_k is the dimensionality of the query, key, and value vectors. The attention scores are computed as the scaled dot-product of the query and key vectors, normalised by the square root of the dimensionality d_k:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{7}$$

This produces an attention-weighted representation of the input sequence, which captures the relationships between time steps and allows the model to focus on the most relevant parts of the sequence for making predictions. The output of the self-attention mechanism is given by:

$$\mathbf{X}_{sa} = \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \tag{8}$$

By incorporating self-attention, the model can capture complex temporal dependencies in the PV power data, such as recurring patterns or correlations between different time intervals, which are essential for accurate forecasting.

3.4. Convolutional layers

The convolutional layers in this work are integral for extracting localised temporal features from the input time series data, enabling the model to identify patterns such as variations in solar irradiance

Table 2
Hyperparameters used for the CNN-Bi-LSTM with spatial, temporal, and self-attention model.

Hyperparameter	Value
Filters	16
Kernel size	5
Activation function (Conv1D)	ReLU
Padding (Conv1D)	Same
Pooling size (MaxPooling1D)	1
LSTM units	75
Dense units	20
Learning rate	0.01
Loss function	Mean squared error
Optimiser	Adam
Epochs	150
Batch size	32
Validation split	20% (Chronological)

or temperature over short time intervals. These layers apply multiple convolutional filters to the input data, capturing critical dependencies between adjacent time steps and features. Mathematically, the output of the jth filter at time step t is expressed as:

$$H_{t,j} = \sigma\left(\sum_{i=1}^{k} W_{j,i} \cdot X_{t+i-1} + b_j\right) \tag{9}$$

where, $W_{j,i}$ is the weight vector for the ith element of the jth filter, b_j is the bias term, and $\sigma(\cdot)$ denotes the activation function, such as ReLU ($\max(0, x)$). The convolution operation slides these filters over the input data, producing feature maps that highlight relevant patterns.

Following the convolutional layers, max pooling is applied to reduce the temporal dimension and emphasise the most prominent features. The pooling operation for a window size p is defined as:

$$P_{t,j} = \max_{i=1,\ldots,p} H_{t+i-1,j} \tag{10}$$

where, $P_{t,j}$ represents the pooled output at time step t for the jth filter. This step reduces computational complexity and ensures robustness to temporal variations in the data. In the context of PV forecasting, the convolutional layers capture meaningful short-term patterns, such as sudden changes in weather conditions or periodic variations, providing a rich representation for subsequent attention and recurrent layers. This localised feature extraction enhances the model's ability to learn and predict PV power generation effectively.

3.5. Bi-LSTM layer

The self-attention output $\mathbf{X}_{sa}$ is then passed into a Bidirectional Long Short-Term Memory (Bi-LSTM) layer. The Bi-LSTM network is well-suited for time series prediction tasks as it captures both forward and backward dependencies in the data. This is particularly important for ultra-short-term PV forecasting, where both recent and past time steps influence future power generation. The Bi-LSTM processes the sequence in two directions: the forward LSTM captures dependencies from the past to the present, while the backward LSTM captures dependencies from the future to the past. The hidden states for the forward and backward passes at each time step t are given by:

$$\mathbf{h}_t^{\rightarrow} = \mathrm{LSTM}(\mathbf{X}_{sa}), \quad \mathbf{h}_t^{\leftarrow} = \mathrm{LSTM}(\mathbf{X}_{sa}) \tag{11}$$

The final hidden representation at each time step is the concatenation of the forward and backward hidden states:

$$\mathbf{h}_t = [\mathbf{h}_t^{\rightarrow}; \mathbf{h}_t^{\leftarrow}] \tag{12}$$

The Bi-LSTM enables the model to learn complex temporal patterns and capture long-range dependencies in both directions, providing a rich representation of the input sequence for forecasting.

Table 3
Hyperparameter combinations and their corresponding best validation losses.

Filters	LSTM Units	Dense Units	Learning Rate	Best validation loss
32	100	30	0.001	0.005718
64	150	50	0.0005	0.004426
16	50	20	0.005	0.001727
32	75	25	0.001	0.002843
64	100	30	0.0001	0.004888
32	150	40	0.001	0.001903
16	75	20	0.01	0.001354
64	50	25	0.0005	0.003903
32	200	50	0.0005	0.002898
16	125	30	0.005	0.001934

3.6. Output layer

The final hidden representation from the Bi-LSTM layer is passed through a dense output layer with a single neuron to predict the next time step's PV power generation. The dense layer is defined as:

$$\hat{y} = \mathbf{W}_o \mathbf{h}_t + b_o \tag{13}$$

where, $\mathbf{W}_o$ is a weight matrix, and b_o is the bias term. The output $\hat{y}$ represents the predicted PV power at the next time step.

3.7. Loss function and optimisation

The model is trained using the Mean Squared Error (MSE) loss function, which is suitable for regression tasks such as PV power forecasting. The MSE loss is given by:

$$\mathcal{L}(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{14}$$

where, $\hat{y}_i$ is the predicted PV power, y_i is the true value, and N is the number of samples. The model is optimised using the Adam optimiser, which is an adaptive learning rate optimisation algorithm that has proven effective for training deep neural networks. The details of the network and their values are shown in Table 2.

4. Results and discussion

4.1. Hyperparameters evaluation

Hyperparameters play a crucial role in the performance of deep learning models. Common hyperparameters such as `filters`, `learning rate`, `LSTM units`, and `epochs` are typically tuned to optimise model performance, as discussed in previous studies [51,52]. In this study, we explored the impact of various hyperparameters on the performance of our model, focusing on `filters`, `LSTM units`, `dense units`, and `learning rate`.

The hyperparameter tuning results for the proposed model, as shown in Table 3, demonstrate that hyperparameter combinations significantly influence model performance, particularly with respect to validation loss. The optimal configuration, `filters: 16`, `lstm_units: 75`, `dense_units: 20`, `learning_rate: 0.01`, achieved the lowest validation loss of 0.001354, suggesting that fewer filters, moderate LSTM units, and a higher learning rate work best for this forecasting task.

Configurations with `filters: 16`, `lstm_units: 50`, `dense_units: 20`, `learning_rate: 0.005` and `filters: 32`, `lstm_units: 150`, `dense_units: 40`, `learning_rate: 0.001` also performed well. However, setups with more filters or lower learning rates resulted in poorer outcomes. This finding highlights that increased complexity does not always guarantee improved performance and emphasises the importance of hyperparameter optimisation in ultra-short-term PV power forecasting. The optimised hyperparameters are summarised in Table 2.

Algorithm 1 Hierarchical Attention Fusion for PV Power Prediction

function ModelArchitecture(X) ▷ X: Input data with shape [samples, time_steps, features]
  $F_{conv} \leftarrow$ Conv1D(X) ▷ Extract features using 1D convolution
  $F_{pool} \leftarrow$ MaxPooling1D(F_{conv}) ▷ Apply max pooling
  $F_{spatial} \leftarrow$ SpatialAttention(F_{pool}) ▷ Apply spatial attention
  $F_{temporal} \leftarrow$ TemporalAttention($F_{spatial}$) ▷ Apply temporal attention
  $F_{self} \leftarrow$ SelfAttention($F_{temporal}$) ▷ Apply self-attention
  $F_{lstm} \leftarrow$ BidirectionalLSTM(F_{self}) ▷ Process with Bi-LSTM
  $y_{pred} \leftarrow$ Dense(F_{lstm}) ▷ Final prediction layer
  return y_{pred}
end function

function SpatialAttention(F) ▷ F: Input feature maps
  $A \leftarrow$ Permute(F, (2, 1)) ▷ Transpose to operate on feature dimension
  $A \leftarrow$ Dense(A, activation=softmax) ▷ Generate attention weights for features
  $A \leftarrow$ Permute(A, (2, 1)) ▷ Transpose back to original shape
  $F_{attended} \leftarrow F \odot A$ ▷ Apply attention via element-wise multiplication
  return $F_{attended}$
end function

function TemporalAttention(F) ▷ F: Spatially attended features
  $A \leftarrow$ Dense(F, activation=softmax) ▷ Generate attention weights for time steps
  $A \leftarrow$ Mean(A, axis=-1) ▷ Average weights across features
  $A \leftarrow$ RepeatVector(A, features) ▷ Repeat for each feature
  $A \leftarrow$ Permute(A, (2, 1)) ▷ Reshape to match input
  $F_{attended} \leftarrow F \odot A$ ▷ Apply attention via element-wise multiplication
  return $F_{attended}$
end function

function SelfAttention(F) ▷ F: Temporally attended features
  $Q \leftarrow$ Dense(F) ▷ Query transformation
  $K \leftarrow$ Dense(F) ▷ Key transformation
  $V \leftarrow$ Dense(F) ▷ Value transformation
  $A \leftarrow$ BatchDot(Q, K, axes=[2,2])$/\sqrt{d_k}$ ▷ Scaled dot-product attention
  $A \leftarrow$ Softmax(A) ▷ Attention weights
  $F_{attended} \leftarrow$ BatchDot(A, V, axes=[2,1]) ▷ Apply attention to values
  return $F_{attended}$
end function

4.2. Evaluation matrices

To comprehensively assess the performance of the proposed ultra-short-term PV power prediction model, five different evaluation metrics were employed: Root Mean Square Error (RMSE), Normalised Root Mean Square Error (NRMSE), Mean Absolute Error (MAE), coefficient of determination (R^2), and Performance Parameter (PP). These metrics provide complementary insights into the model's accuracy and reliability [53].

4.2.1. Root Mean Square Error (RMSE)

RMSE is a widely used metric that measures the standard deviation of prediction errors, giving higher weights to larger errors due to its quadratic nature [54]. It is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

where n is the number of samples, y_i represents the actual PV power output, and $\hat{y}_i$ denotes the predicted value.

4.2.2. Normalised Root Mean Square Error (NRMSE)

NRMSE normalises the RMSE by the range of observed values, enabling comparison across different datasets and scales:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \times 100 \tag{16}$$

where y_{max} and y_{min} are the maximum and minimum observed values in the dataset, respectively.

4.2.3. Mean Absolute Error (MAE)

MAE provides the average magnitude of errors without considering their direction, making it less sensitive to outliers compared to RMSE [55]:

$$\text{MAE (\%)} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{17}$$

4.2.4. Coefficient of determination (R^2)

R^2 indicates the proportion of variance in the dependent variable that is predictable from the independent variable(s):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{18}$$

where $\bar{y}$ is the mean of observed values. R^2 ranges from 0 to 1, with 1 indicating perfect prediction.

4.2.5. Performance Parameter (PP)

PP is a specialised metric that considers the dataset's inherent variability through its standard deviation:

$$PP = 1 - \frac{RMSE}{\alpha} \tag{19}$$

where α is the standard deviation of the dataset. PP approaches 1 for high-quality predictions and can be negative for poor predictions. This metric is particularly useful for PV power prediction as it accounts for the natural variability in solar power generation. These metrics collectively provide a comprehensive evaluation framework, with:

- RMSE and MAE quantifying prediction errors in the original units
- NRMSE enabling cross-dataset comparison
- R^2 assessing the model's explanatory power
- PP accounting for dataset-specific variability

4.3. Ablation study

To systematically evaluate the contribution of each component in our proposed architecture, we conducted a comprehensive ablation study with the time stamp prediction of 1 min. The study progressively removes key components from our full model (STSA-CNN-BiLSTM) to quantify their individual impacts on prediction performance. Table 4 presents the comparative results across five different model configurations. Also, we run the proposed model for 30 times and the training and validation losses are shown in Fig. 3.

4.3.1. Model configurations

The ablation study examined five distinct configurations in decreasing order of complexity. The complete proposed architecture (STSA-CNN-BiLSTM) incorporates Spatial Attention, Temporal Attention, and

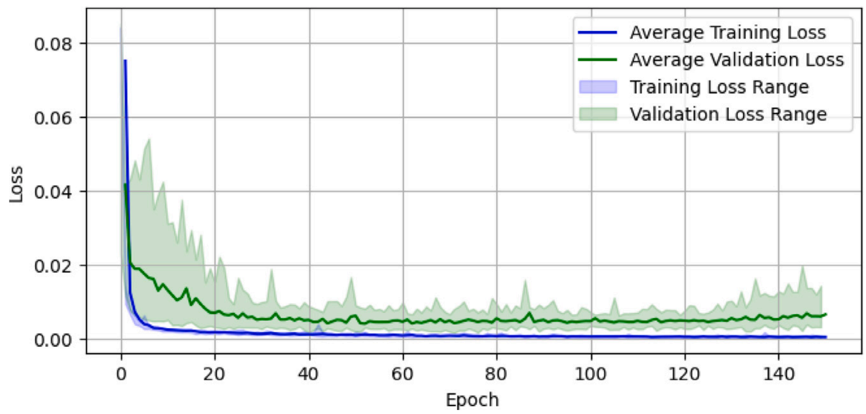

Fig. 3. Training and validation losses over the epochs with 30 runs.

Self-Attention mechanisms with CNN-BiLSTM. The second configuration (STA-CNN-BiLSTM) removes Self-Attention while retaining Spatial and Temporal Attention. The third variant (TA-CNN-BiLSTM) further removes Spatial Attention, keeping only Temporal Attention. The fourth configuration (CNN-BiLSTM) represents the base architecture without any attention mechanisms, while the final variant (CNN-LSTM) replaces BiLSTM with unidirectional LSTM to evaluate the impact of bidirectional processing.

Table 4
Performance comparison on test data for ablation study of the proposed model.

Model	RMSE (kW)	MAE	NRMSE	R-squared	PP
STSA-CNN-BiLSTM	0.162	2.91%	0.0325	0.9923	0.8893
STA-CNN-Bi-LSTM	0.206	3.32%	0.0668	0.9848	0.8592
TA-CNN-Bi-LSTM	0.389	4.99%	0.1210	0.9675	0.7341
CNN-Bi-LSTM	0.404	5.21%	0.1633	0.9502	0.7239
CNN-LSTM	0.5692	7.99%	0.2198	0.9358	0.6109

4.3.2. Impact of self-attention

The integration of Self-Attention demonstrates significant performance improvements, as evidenced by comparing STSA-CNN-BiLSTM against STA-CNN-BiLSTM. The addition of Self-Attention reduces RMSE by 21.4% from 0.206 kW to 0.162 kW, while MAE improves by 12.3% from 3.32% to 2.91%. The NRMSE shows a substantial decrease of 51.3% from 0.0668 to 0.0325, accompanied by an R^2 enhancement from 0.9848 to 0.9923 and a PP increase from 0.8404 to 0.8893. These improvements highlight the crucial role of Self-Attention in capturing complex dependencies within the input data.

4.3.3. Contribution of spatial attention

The removal of Spatial Attention, observed by comparing STA-CNN-BiLSTM with TA-CNN-BiLSTM, reveals substantial performance degradation across all metrics. The RMSE increases significantly by 88.8% from 0.206 kW to 0.389 kW, while MAE rises by 50.3% from 3.32% to 4.99%. The NRMSE shows a considerable deterioration of 81.1% from 0.0668 to 0.1210, with R^2 decreasing from 0.9848 to 0.9675 and PP dropping from 0.8404 to 0.8102. These results underscore the importance of Spatial Attention in capturing spatial dependencies within the PV power prediction task.

4.3.4. Effect of temporal attention

The impact of Temporal Attention becomes apparent when comparing TA-CNN-BiLSTM with CNN-BiLSTM. While the improvements are more modest compared to other components, they remain significant. The inclusion of Temporal Attention yields a 3.9% improvement in RMSE from 0.404 kW to 0.389 kW and a 4.2% reduction in MAE from 5.21% to 4.99%. Notable improvements are also observed in NRMSE with a 25.9% enhancement from 0.1633 to 0.1210, along with R^2 improvement from 0.9502 to 0.9675 and PP increase from 0.7913 to 0.8102. These results demonstrate the value of Temporal Attention in capturing time-dependent patterns in the data.

4.3.5. Bidirectional vs. Unidirectional LSTM

The comparison between CNN-BiLSTM and CNN-LSTM reveals the substantial advantages of bidirectional processing in the LSTM layer. The bidirectional approach achieves 29.0% lower RMSE (0.404 kW vs. 0.5692 kW) and a 34.8% reduction in MAE (5.21% vs. 7.99%). The NRMSE shows a 25.7% improvement (0.1633 vs. 0.2198), accompanied by superior R^2 (0.9502 vs. 0.9358) and PP (0.7913 vs. 0.7633) values. These results demonstrate that bidirectional processing significantly enhances the model's ability to capture temporal dependencies in both forward and backward directions.

The comprehensive ablation study demonstrates that each component of our proposed architecture contributes meaningfully to the overall performance. The Self-Attention mechanism provides the most substantial individual improvement in prediction accuracy, followed by the Spatial Attention mechanism. The Temporal Attention component, while showing more modest gains, still contributes meaningfully to the model's performance. The combination of all three attention mechanisms yields the best performance across all metrics, validating our architectural design choices. Furthermore, the bidirectional nature of the LSTM layer proves crucial, offering significant advantages over its unidirectional counterpart. These findings collectively support the design decisions in our proposed STSA-CNN-BiLSTM architecture and demonstrate its effectiveness for ultra-short-term PV power prediction.

4.4. Comparative analysis

The results in Table 5 highlight the superior performance of the proposed Spatial Attention Temporal Attention Self-Attention-based CNN BiLSTM model for ultra-short-term photovoltaic (PV) power forecasting. The model achieves the lowest RMSE (0.162 kW) and MAE (2.91%), indicating higher accuracy in forecasting compared to other models,as shown in Fig. 5. Furthermore, its NRMSE (0.0325%) is substantially lower than the alternatives, showcasing its robustness in minimising relative errors. The high R-squared value (0.9923) further affirms the model's ability to explain the variability in PV power output with excellent reliability. Notably, the proposed model's Performance Parameter (PP) value of 0.8893 significantly outperforms other models, underlining its superior forecasting quality and alignment with actual PV power trends.

In comparison, traditional models like XGBoost and Vanilla LSTM show weaker performance, with RMSE values of 0.292 kW and 0.301 kW, respectively, and higher MAE and NRMSE values, reflecting their limited ability to capture the complexities of PV power forecasting. Seq2Seq with Attention and TCN models, despite utilising advanced

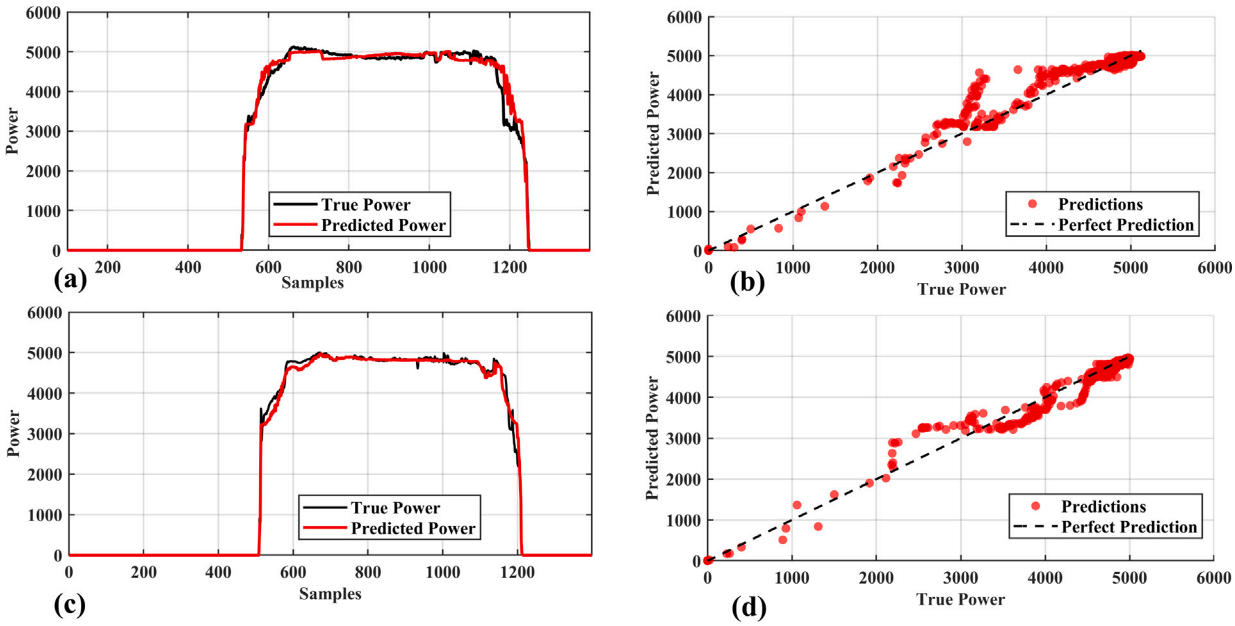

Fig. 4. True vs. Predicted power (KW) of proposed model (a) Line plot of day 1 (b) Scatter plot of day 1 (c) Line plot of day 2 (d) scatter plot of day 2.

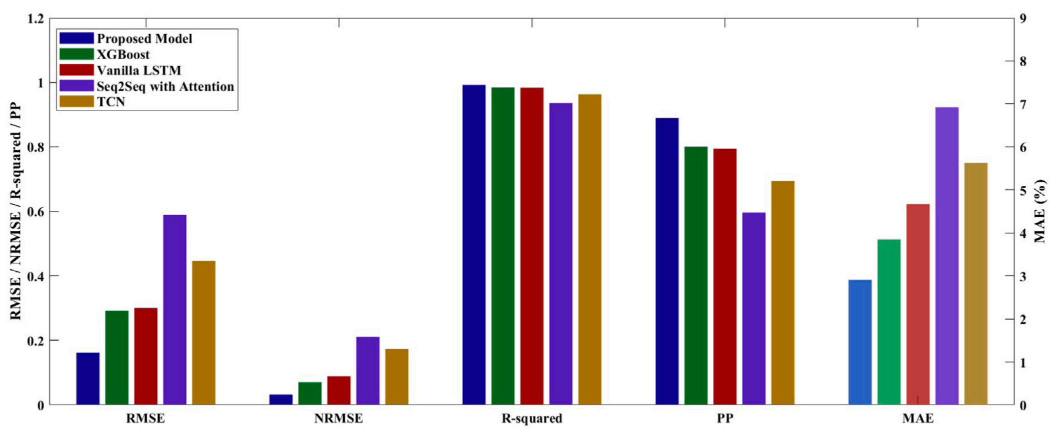

Fig. 5. Bar graph of the evaluation matrices of the state of the art algorithms.

Table 5
Performance metrics for different models.

Model	RMSE (kW)	MAE	NRMSE (%)	R-squared	PP
Proposed model	0.162	2.91%	0.0325	0.9923	0.8893
XGBoost	0.292	3.85%	0.0705	0.9842	0.8004
Vanilla LSTM	0.301	4.67%	0.0885	0.9832	0.7942
Seq2Seq with attention	0.590	6.92%	0.2115	0.9358	0.5967
TCN	0.447	5.63%	0.1733	0.9631	0.6944

architectures, demonstrate even higher errors and lower PP values (0.5967 and 0.6944, respectively). These results emphasise the impact of incorporating attention mechanisms and hybrid architectures like CNN BiLSTM in the proposed model, which effectively leverage spatial and temporal dependencies in the data. The attention mechanisms enhance the model's capacity to focus on critical patterns and reduce noise, making it highly suitable for ultra-short-term PV forecasting tasks (see Fig. 4).

5. Conclusions

This paper tackles the challenge of predicting ultra-short-term solar power generation with minimal data, a common issue in Norway, where sunlight can change rapidly over short time frames. Forecasting the next minute's power output under such conditions is particularly difficult for traditional models. Ultra-short-term predictions in solar power provide precise, localised insights into rapid power output fluctuations, essential for real-time energy system management. To tackle this issue, we propose a model that combines Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory networks (BiLSTM) with three distinct attention mechanisms. These mechanisms enhance the model's ability to capture spatial and temporal characteristics while focusing on the importance of specific time steps. An ablation study reveals that using all three attention mechanisms significantly improves predictive performance. The proposed model is tailored to Norway's weather conditions and predicts solar radiation on tilted surfaces with high accuracy. It achieves a root mean square error (RMSE) of 0.162 kW, outperforming models that use one or two attention mechanisms. This model is not only effective for solar photovoltaic (PV) power forecasting but also applicable to other renewable energy technologies, addressing uncertainties arising from variable weather conditions. Additionally, it can be integrated into energy systems to mitigate challenges associated with renewable energy intermittency, allowing for more efficient operation by anticipating output fluctuations

CRediT authorship contribution statement

Opy Das: Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Dounia Dahlioui:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Muhammad Hamza Zafar:** Writing – review & editing, Writing – original draft, Validation, Software, Investigation. **Naureen Akhtar:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources. **Syed Kumayl Raza Moosavi:** Writing – review & editing, Writing – original draft, Resources, Formal analysis. **Filippo Sanfilippo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration.

Declaration of competing interest

NONE

All authors claim that there is not any conflict of interest regarding the above submission. The work of this submission has not been published previously. It is not under consideration for publication elsewhere. Its publication is approved by all authors and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

Acknowledgements

This research is supported by the Artificial Intelligence, Biomechatronics, and Collaborative Robotics research group at the Top Research Center Mechatronics (TRCM), University of Agder (UiA), Norway. The authors gratefully acknowledge the contribution of Anne Gerd Imenes.

Data availability

Data will be made available on request.

References

[1] Klyve ØS, Nygård MM, Riise HN, Fagerström J, Marstein ES. The value of forecasts for PV power plants operating in the past, present and future Scandinavian energy markets. Sol Energy 2023;255:208–21, URL https://www.sciencedirect.com/science/article/pii/S0038092X23002049.

[2] Das O, Zafar MH, Sanfilippo F, Rudra S, Kolhe ML. Advancements in digital twin technology and machine learning for energy systems: A comprehensive review of applications in smart grids, renewable energy, and electric vehicle optimisation. Energy Convers Manag: X 2024;24:100715, URL https://www.sciencedirect.com/science/article/pii/S2590174524001934.

[3] Iheanetu KJ. Solar photovoltaic power forecasting: A review. Sustainability 2022;14(24). URL https://www.mdpi.com/2071-1050/14/24/17005.

[4] Hussain A, Chen Z, Ma T, Ahmad I. A review on solar power forecasting methods for smart grid and energy market requirements. Optim Eng 2021;22:1267–83, URL https://link.springer.com/content/pdf/10.1007/s11081-020-09583-y.pdf.

[5] Kumar NM, Kumar P, Balamurugan P. A review on the forecasting of solar irradiance and solar power. Energy 2020;190:116467, URL https://www.sciencedirect.com/science/article/pii/S0196890420311717?via%3Dihub.

[6] Wang Y, Sun Y, Zhang C, Gu C, Zhang J. A comprehensive review of solar photovoltaic forecasting: The role of deep learning techniques. Appl Energy 2021;285:116452, URL https://www.sciencedirect.com/science/article/pii/S0306261921007054?via%3Dihub.

[7] Liu Y, Zhang J, Zhang Q, Wang Y, Ma J. A review on artificial intelligence in solar photovoltaic energy systems. IEEE Access 2021;9:32774–93, URL https://ieeexplore.ieee.org/document/9405472.

[8] Magnaldo M-A, et al. Evaluation of surface shortwave downward radiation forecasts by the numerical weather prediction model AROME. Geosci Model Dev 2024;17:1091–109, URL https://gmd.copernicus.org/articles/17/1091/2024/.

[9] Du J, Zhang Y, Zheng S, Wang C, Zhang X. Short-term solar irradiance forecasts using sky images and radiative transfer model. Energies 2018;11(5):1107, URL https://www.mdpi.com/1996-1073/11/5/1107.

[10] Voyant C, Notton G, Kalogirou S, Nivet M-L, Paoli C, Motte F, Fouilloy A. Machine learning methods for solar radiation forecasting: A review. Renew Energy 2017;105:569–82, URL https://www.sciencedirect.com/science/article/pii/S096014811730455X.

[11] Wang C, Dong Z, Zhang X, Zhou Y, Zhao B. A comprehensive review of solar photovoltaic power forecasting methods: Models and applications. Sensors 2021;21(16):5648, URL https://www.mdpi.com/1424-8220/21/16/5648.

[12] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. Sol Energy 2013;85(7):1189–209, URL https://www.sciencedirect.com/science/article/pii/S0038092X08002107?via%3Dihub.

[13] Mahata S, Chatterjee R, Saha H. A novel GAN architecture reconstructed using bi-LSTM and style transfer for PV temporal dynamics simulation. IEEE Trans Sustain Energy 2024. URL https://ieeexplore.ieee.org/document/10601515.

[14] Mellit A, Pavan AM. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Sol Energy 2010;84(5):807–21, URL https://www.sciencedirect.com/science/article/pii/S0038092X08002107?via%3Dihub.

[15] Nespoli L, Medici V, Caputo B, Ferracuti F, Fonti A, Priori M, Verdolini E. Machine learning for solar energy prediction and forecasting: A systematic review. Appl Energy 2021;293:116983, URL https://www.sciencedirect.com/science/article/pii/S0306261921008394?via%3Dihub.

[16] Zhao J, Li M, Zhang X, Wang Y. A comprehensive review of photovoltaic power forecasting methods: Models and applications. Autom Electr Power Syst 2022;46(10):14–25, URL http://www.aeps-info.com/aeps/article/abstract/20220324006.

[17] Liu J, Zhang L, Sun H, Liu Y, Song Q. Statistical machine learning for power flow analysis considering the influence of weather factors on photovoltaic power generation. IEEE Trans Power Syst 2024. URL https://ieeexplore.ieee.org/document/10494312.

[18] Chen Z, Wang Z, Li X, Liu P, Tang Y. A review on solar power forecasting: Models, methods, and challenges. IEEE Access 2021;9:118176–99, URL https://ieeexplore.ieee.org/document/9380330.

[19] Li M, Wang F, Zhang X, Liu J. Research on solar power prediction models and techniques: A review. Power Syst Technol 2016;40(2):430–8.

[20] Fathima AS, Palanisamy K. Optimization in microgrids with renewable energy and storage: A review. Renew Sustain Energy Rev 2018;45:431–46, URL https://www.sciencedirect.com/science/article/pii/S096014811730455X.

[21] Pan H, Qin H, Zheng J, Li L, Liu H. Review on solar photovoltaic power forecasting: Application, development and challenges. IEEE Access 2021;9:122611–28, URL https://doi.org/10.1109/ACCESS.2021.3099169.

[22] Li X, Ma L, Chen P, Xu H, Xing Q, Yan J, Lu S, Fan H, Yang L, Cheng Y. Probabilistic solar irradiance forecasting based on XGBoost. Energy Rep 2022;8:1087–95.

[23] Ray B, Lasantha D, Beeravalli V, Anwar A, Nabi MN, Sheng H, Rashid F, Muyeen S. A comprehensive framework for effective long-short term solar yield forecasting. Energy Convers Manag: X 2024;22:100535.

[24] Srinivasan H, Lekhashree V, Manohar S. Time-series forecasting in retail industry using bidirectional, stacked, and vanilla LSTMs. In: Proceedings of eighth international conference on information system design and intelligent applications. Springer; 2024, p. 503–13.

[25] Kharlova E, May D, Musilek P. Forecasting photovoltaic power production using a deep learning sequence to sequence model with attention. In: 2020 international joint conference on neural networks. IJCNN, IEEE; 2020, p. 1–7.

[26] Salameh T, Farag MM, Hamid A-K, Hussein M. Adaptive neuro-fuzzy inference system for accurate power forecasting for on-grid photovoltaic systems: A case study in Sharjah, UAE. Energy Convers Manag: X 2025;26:100958.

[27] Li Y, Song L, Zhang S, Kraus L, Adcox T, Willardson R, Komandur A, Lu N. A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting. IEEE Trans Smart Grid 2023;14(5):4073–85.

[28] Tian Z, Chen Y, Wang G. Enhancing PV power forecasting accuracy through nonlinear weather correction based on multi-task learning. Appl Energy 2025;386:125525, URL https://www.sciencedirect.com/science/article/pii/S0306261925002557.

[29] Mbungu NT, Bashir SB, Michael NE, Farag MM, Hamid A-K, Ismail AAA, Bansal RC, Abo-Khalil AG, Elnady A, Hussein M. Predictive control technique for solar photovoltaic power forecasting. Energy Convers Manag: X 2024;24:100768.

[30] Yang D, Kleissl J, Gueymard CA, Pedro HT, Coimbra CF. History and trends in solar irradiance and photovoltaic forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101, URL https://www.sciencedirect.com/science/article/pii/S0038092X17309052?via%3Dihub.

[31] Ahmad S, Zhang D, Huang Q, Zhang H, Chen G, Yan B, Cao G. Review of various solar PV power forecasting techniques. Electr Power Syst Res 2021;191:107427, URL https://doi.org/10.1016/j.epsr.2021.107427.

[32] Diagne A, Chen D, Wang Z, Gu W, Bu S. Solar power forecasting: A review of the state-of-the-art and future directions. IEEE Syst J 2021;15(1):207–19, URL https://doi.org/10.1109/JSYST.2020.3007184.

[33] Lodhi E, Dahmani N, Bukhari SMS, Gyawali S, Thapa S, Qiu L, Zafar MH, Akhtar N. Enhancing microgrid forecasting accuracy with SAQ-MTCLSTM: A self-adjusting quantized multi-task ConvLSTM for optimized solar power and load demand predictions. Energy Convers Manag: X 2024;24:100767.

[34] Pedro HT, Coimbra CF. Assessment of forecasting techniques for solar power production with no exogenous inputs. Energy Procedia 2017;135:22–31, URL https://www.sciencedirect.com/science/article/pii/S187661021735854X.

[35] Antonanzas J, Osorio N, Escobar R, Urraca R, de Pison FM, Antonanzas-Torres F. Review of photovoltaic power forecasting. Appl Energy 2016;228:383–402, URL https://www.sciencedirect.com/science/article/pii/S0306261916317032.

[36] Hossain M, Oo A, Ali A. Application of extreme learning machine for short-term output power forecasting of three grid-connected PV systems. J Clean Prod 2017;167:395–405, URL https://scholar.google.com/scholar_lookup?title=Application%20of%20extreme%20learning%20machine%20for%20short%20term%20output%20power%20forecasting%20of%20three%20grid-connected%20PV%20systems&author=M.%20Hossain&publication_year=2017.

[37] Massidda L, Marrocu M, Riccio A. Daily solar irradiation forecasting using a hybrid neural network. Appl Energy 2016;170:333–41, URL https://www.sciencedirect.com/science/article/pii/S0306261916301180.

[38] Iheanetu KJ. Solar photovoltaic power forecasting: A review. Sustainability 2022;14(24):17005, URL https://www.mdpi.com/2071-1050/14/24/17005.

[39] Al-Dahidi S, Madhiarasan M, Al-Ghussain L, et al. Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework. Energies 2024;17(16):4145, URL https://www.mdpi.com/1996-1073/17/16/4145?utm_source=chatgpt.com.

[40] Li G, Xie S, Wang B, Xin J, Li Y, Du S. Photovoltaic power forecasting with a hybrid deep learning approach. IEEE Access 2020;8:175871–80.

[41] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. In: Proceedings of the 3rd international conference on learning representations. ICLR, 2015, URL https://arxiv.org/abs/1409.0473.

[42] Yao T, Wang J, Cao H, Liu F, Wang X, Wang Y, Chi X. A multi-level attention-based LSTM network for ultra-short-term solar power forecast using meteorological knowledge. In: Memmi G, Yang B, Kong L, Zhang T, Qiu M, editors. Knowledge science, engineering and management. Cham: Springer International Publishing; 2022, p. 16–28.

[43] Zhong Y, He T, Mao Z. Enhanced solar power prediction using attention-based DiPLS-BiLSTM model. Electronics 2024;13(23). URL https://www.mdpi.com/2079-9292/13/23/4815.

[44] Rao Z, Yang Z, Yang X, Li J, Meng W, Wei Z. TCN-GRU based on attention mechanism for solar irradiance prediction. Energies 2024;17(22). URL https://www.mdpi.com/1996-1073/17/22/5767.

[45] Zheng Y, Chen W, Zhang L, Chen Z, Wang J. Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture. Appl Energy 2023;343:121192.

[46] Zhou X, Zhang L, Wang J, Sun H. Interpretable feature selection and deep learning for short-term probabilistic PV power forecasting in buildings using local monitoring data. Appl Energy 2024;355:121654.

[47] Ma Q, Fu X, Yang Q, Qiu D. Adaptive masked network for ultra-short-term photovoltaic forecast. Eng Appl Artif Intell 2025;139:109555.

[48] Yang F, Fu X, Yang Q, Chu Z. Decomposition strategy and attention-based long short-term memory network for multi-step ultra-short-term agricultural power load forecasting. Expert Syst Appl 2024;238:122226.

[49] Skoplaki E, Palyvos JA. On the temperature dependence of photovoltaic module electrical performance: A review. Sol Energy 2009;83(5):614–24.

[50] Aste N, Del Pero C, Leonforte F. Thermal behaviour and efficiency of photovoltaic panels under different wind conditions. Energy Procedia 2015;82:309–14.

[51] Aszemi NM, Dominic PDD. Hyperparameter optimization in convolutional neural network using genetic algorithms. Int J Adv Comput Sci Appl 2019;10(6).

[52] Gorgolis N, Hatzilygeroudis I, Istenes Z, Gyenne L-G. Hyperparameter optimization of LSTM network models through genetic algorithm. In: 2019 10th international conference on information, intelligence, systems and applications. IISA, IEEE; 2019, p. 1–4.

[53] Zhang J, Florita A, Hodge B-M, Lu S, Hamann HF, Banunarayanan V, Brockway AM. A suite of metrics for assessing the performance of solar power forecasting. Sol Energy 2015;111:157–75.

[54] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast 2006;22(4):679–88.

[55] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30(1):79–82.