Review article

Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies

Husein Mauladdawilah [a],*, E.J. Gago [a], Hasan Balfaqih [b], M.C. Pegalajar [c]

[a] University of Granada, School of Civil Engineering, 18001, Granada, Spain
[b] Operations and Supply Chain Management Department, College of Business, Effat University, Jeddah, 21478, Saudi Arabia
[c] University of Granada, Department of Computer Science and Artificial Intelligence, 18014, Granada, Spain

A R T I C L E   I N F O

A B S T R A C T

Time series forecasting still awaits a transformative breakthrough like that happened in computer vision and natural language processing. The absence of extensive, domain-independent benchmark datasets and standardized performance measurement units poses a significant challenge for it, especially for photovoltaic forecasting applications. Additionally, since it is often time domain-driven, a plethora of highly unique and domain-specific datasets were produced. The lack of uniformity among published models, developed under diverse settings for varying forecasting horizons, and assessed using non-standardized metrics, remains a significant obstacle to the progress of the field as a whole. To address these issues, a systematic review of the state-of-the-art literature on prediction tasks is presented, collected from the Web of Science and Scopus databases, published in 2022 and 2023, and filtered using keywords such as "photovoltaic," "deep learning," "forecasting," and "time series." Finally, 36 case studies were selected. Before comparing, a state-of-the-art demonstration of key elements in the topic was presented, such as model type, hyperparameters, and evaluation metrics. Then, the 36 articles were compared in terms of statistical analysis, including top publishing countries, data sources, variables, input, and output horizon, followed by an overall model comparison demonstrating every proposed model categorized into model type (artificial neural network units, recurrent units, convolutional units, and transformer units). Due to the mostly utilization of specific private datasets measured at the targeted location, having universal error metrics is crucial for clear global benchmarking. Root Mean Squared Error and Mean Absolute Error were the most utilized metrics, although they specifically demonstrate the accuracy relative to their respective sites. However, 33% utilized universal metrics, such as Mean Absolute Percentage Error, Normalized Root Mean Squared Error, and the Coefficient of Determination. Finally, trends, challenges, and future research were highlighted for the relevant topic to spotlight and bypass the current challenges.

1. Introduction

One of the primary strategies to reduce the carbon footprint left by fossil fuels is the adoption of renewable energy sources. Additionally, renewable energy sources are used globally to supplement an energy infrastructure that is already reliant on highly

---

* Corresponding author.
E-mail addresses: mauladdawilah@correo.ugr.es (S.M. Husein), ejadraque@ugr.es (E.J. Gago), hbalfaqih@effatuniversity.edu.sa (B. Hasan), mcarmen@decsai.ugr.es (M.C. Pegalajar).

Nomenclature

Abbreviations

AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criterion
BiGRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long-Short-Term Memory
BO	Bayesian Optimization
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CNN	Convolutional Neural Network
CO2	Carbon Dioxide
COCO	Common Object in Context
1D-CNN	1 dimensional CNN
EEG	Electroencephalogram
EEMD	Ensemble Empirical Mode Decomposition
FFNN	Feedforward Neural Network
GHI	Global Horizontal Irradiance
GPT	Generative Pretrained Transformer
GRU	Gated Recurrent Unit
GWO	Grey Wolf Optimizer
ICEEMDAN	Improved CEEMDAN
IEA	The International Energy Agency
IMF	Intrinsic Mode Function
IoT	The Internet of Things
KNN	K-nearest Neighbors
KPCA	Kernel Principal Component Analysis
LGBM	Light Extreme Boost
LSTM	Long-Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multi Layer Perceptron
MLR	Multiple Linear Regression
MSE	Mean Squared Error
NASA	The National Aeronautics and Space Administration
NLP	Natural Language Processing
NREL	National Renewable Energy Laboratory
NWP	Numerical Weather Prediction
PCC	Pearson Correlation Coefficients
POA	Plane of Array
PSO	Swarm Optimization
PV	Photovoltaic
PVGIS	PV Global Information Systems
PVPS	Photovoltaic Power Systems

R^2	Coefficient of Determination
ReLU	Rectified Linear Unit
RF	Random Forest
RGB	Red, Green, and Blue
RMSE	Root mean squared error
RNN	Recurrent neural network
SARIMA	Seasonal ARIMA
SCC	Spearman Correlation Coefficients
SGD	Stochastic Gradient Descent
SSA	Salp Swarm Algorithm
SVM	Support Vector Machines
SVR	Support Vector Regression
SWT	Stationary Wavelet Transform
tanh	Hyperbolic Tangent Function
VMD	Variational Mode Decomposition
XGB	Extreme Gradient Boost

Notations/Symbols

W	Weight
b	Bias
n	Number of inputs
h_t	Hidden state
c_t	Cell state
σ	Sigmoid activation function
$\overrightarrow{h_t}$	Hidden vectors of forward network
$\overleftarrow{h_t}$	Hidden vectors of backward network
M	Sub-matrix
$\otimes$	Convolutional operation
f	CNN filter
s	CNN stride
d_k	Dimension of k
Q	Queries of the Attention
K	Keys of the Attention
V	Values of the Attention
pos	Transformer encoding position
y_i	Labeled output
$\hat{y}_i$	Predicted output
θ	Weight/Bias
η	Learning rate
J	Cost function
Z	Perceptron input function

Units

GW	Giga Watt
GWp	Giga Watt peak
kWh	Kilo Watt hour
MW	Mega Watt

polluting sources. This becomes crucial when the demand for energy in a certain location increases or access to conventional energy sources is limited. For instance, the European Union's 2021–2022 gas crisis and the threat to its 2050 decarbonization ambitions caused considerable alarm [1]. Therefore, energy from renewable sources is highly valued by both public and private institutions as a way of addressing this issue. This study specifically focuses on solar energy as one type of renewable energy.

Solar cells are electronic devices that convert solar energy, or irradiance, into electrical energy using the photovoltaic effect, a phenomenon that enables the generation of voltage and current when a substance is exposed to sunlight. The fundamental components of conventional solar panels are solar cells, and these panels are often grouped together to form a solar string. Solar energy has grown in popularity due to the enormous demand for electricity and the negative environmental effects of non-renewable energy sources [2]. As the number of PV systems being installed worldwide continues to rise, there is a growing demand for optimizing the
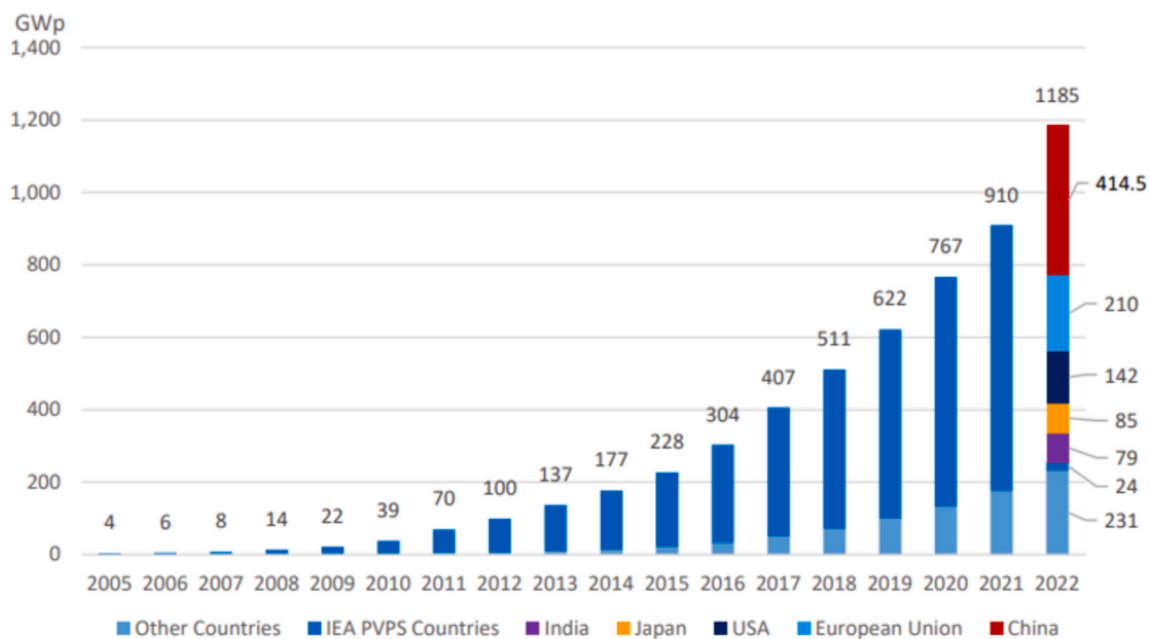
Fig. 1. Trend of Worldwide total installed PV panels capacity. (source: IEA PVPS 2023).

performance and reducing the cost of these systems. Forecasting the output of these systems is essential for PV optimization and cost savings.

In the IEA PVPS 2023 report [3], it is revealed that worldwide installations of PV panels have grown significantly in the past few years. The new installations of PV arrays increased by 30%, 20.6%, and 37% in 2020, 2021, and 2022, respectively, marking a new benchmark growth since the inception of this trend [4]. Fig. 1 illustrates the global installation of photovoltaic panels over the past two decades. Major countries leading in PV systems installations include China, EU countries, the USA, Japan, and India. Globally, a total of more than 1185 GWp is operating as of 2022 [3]. As the need for sustainable energy sources continues to rise, solar energy technology is expected to grow at an unprecedented rate in the coming years.

China has experienced the most significant growth in solar energy generation, adding 104 GW in 2022 alone, constituting over 34% growth in a single year. This pushed their total capacity to 414.5 GW, which is double that of Europe's capacity. European countries come second in capacity, installing 39 GW in 2022, accounting for 23% of the new capacity. Leading in Europe are Spain (20.8%), Germany (19.2%), Poland (12.6%), and Holland (10%), bringing the total capacity to 210 GW. The USA market achieved a 15.1% increase in new installed capacity, and Japan demonstrated steady growth with an addition of 6.4 GW, similar to the previous year. Surprisingly, India is exhibiting ambitious growth, ranking second highest in solar energy growth with a 30% increase in new installations. This highlights the country's vision to expand in the market. Photovoltaic panels are playing a major role in reducing CO_2 emissions and providing an independent source of energy, unlike fossil fuels. On the other hand, its stability is considered a very challenging topic.

PV performance depends on many factors [5], primarily solar irradiance and then meteorological variables (i.e., cloud thickness, precipitation, and temperature), PV module type, the structure of the installation, and electrical components that are connected within the system. The degree of effect of the aforementioned variables is not constant over time; it varies with the year and season. [6]. In the literature, many methods have been presented to predict PV panel power generation. It can be categorized into four methods: physical, persistent, classical (statistical), and artificial intelligence models.

Physical models predict PV panel output by derived equations, such as that are used in PV Global Information Systems (PVGIS) [7]. It takes inputs that consist of PV's information (i.e., unit type and mounting angle), meteorological measurements, and the latitude and longitude of the panels. The main input for a physical model is irradiance data [8]. It can be collected from various sources, including satellite data, weather station databases, and Numerical Weather Prediction (NWP) [9]. In [10], day-ahead forecasting physical model was implemented. The most notable advantage is that it outperforms statistical methods with lower errors. On the other hand, the physical model is fixed on a specific weather condition where sudden extreme climate fluctuation is not considered [11].

The persistence (Naive) model is predicted by assuming that next time t will be equal to a similar historical time t (i.e., tomorrow is equal to today). PV power outputs were forecasted in [12]. It is noted that this model is mostly used as a benchmark for comparison purposes with other models [13].

Traditional statistical techniques require prior collected data. On this specific topic, measured PV-generated power is used in most studies in the literature. It fits a curve over historical data that will be extended into the future time frame [14]. Models that have been applied to the forecast are the Multiple Linear Regression Model (MLR), Autoregressive Integrated Moving Average (ARIMA),
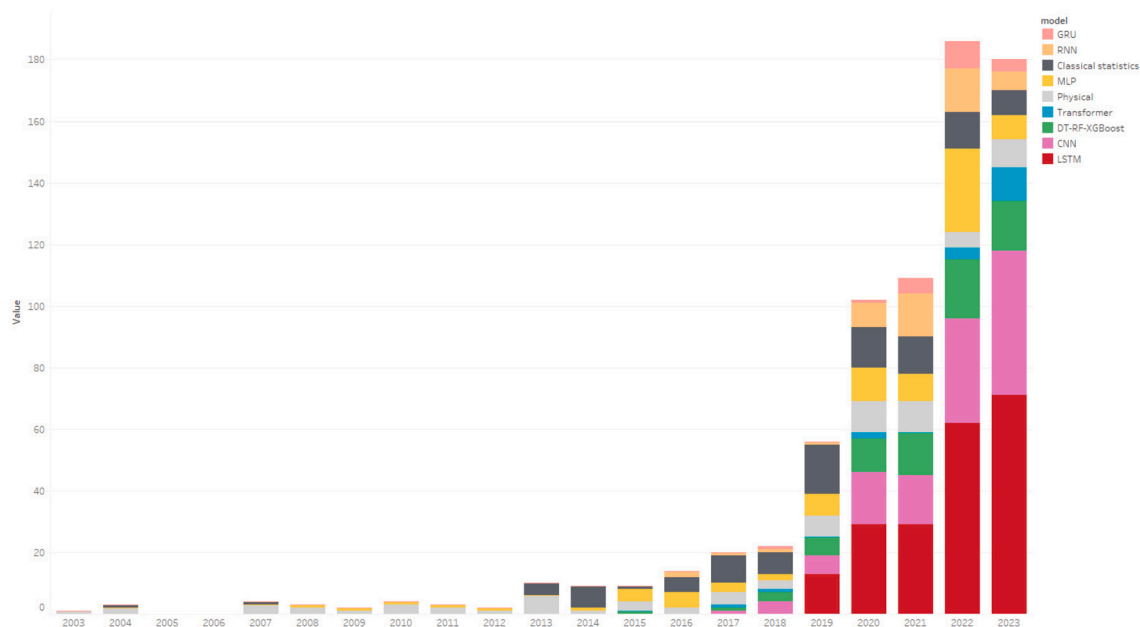
Fig. 2. Techniques adapted in the publications of PV power forecasting over time.

and Seasonal ARIMA (SARIMA) [15–17]. In contrast, machine learning (ML) algorithms do not require a long learning process. ML for regression is a technique that uses a computer's capability to perform a large number of repetitive computations. Models such as K-Nearest Neighbors (KNN) [18], Decision Tree techniques (such as random forest (RF) [19], Extreme Gradient Boost (XGB) [19], [20], Light Extreme Boost (LGBM) [20]) and Deep Learning Neural Networks. Many variations of neural networks with gradient descent and back-propagation algorithms were developed. These models include the Forward Neural Network (FFNN), Multilayer Perceptron (MLP) in [13], Recurrent Neural Network (RNN), Long-Short-Term Memory (LSTM-RNN) [21], [22], Gated Recurrent Unit (GRU-RNN) [23].

PV time series deep learning is an area that has experienced remarkable growth and innovation over the past decade. Fig. 2 sheds light on the evolution of PV power forecasting techniques as depicted in published works collected from Scopus database. A clear trend emerges: deep learning is the dominant force in this field. Among deep learning techniques, LSTM stands out as the most widely employed, followed by CNN, Transformer, MLP, RNN, and GRU. Notably, classical statistical methods and physical modeling, alongside other machine learning techniques, have also experienced a consistent increase in the number of publications. This evidence underlines the vibrant and accelerating nature of PV power forecasting as an expanding field.

Fig. 3 presents a network graph illustrating the key keywords from PV forecasting papers published in 2022 and 2023, as indexed in the Web of Science and Scopus databases. These keywords are color-coded and grouped based on co-occurrence in the papers (e.g., forecasting, machine learning, and deep learning). The graph reveals five distinct regions, offering insights into the topics of most interest to researchers.

Deep learning techniques are central in the network (the blue region), given our focus on them. The red region highlights the machine learning connections in solar forecasting related to smart grid applications, including models such as ANN, CNN, LSTM, Random Forest, and SVM. Irradiance forecasting with PCA is connected to recurrent model types, as depicted in the green region. Moreover, the Transformer unit and Attention Mechanism are linked with irradiance and PV power forecasting in the purple region. Other techniques, such as Bayesian Optimization, NWP, and Transfer Learning, are also present. Lastly, these studies commonly employ time series data and sky images as types of data.

For the systematic review methodology in Fig. 4, a comprehensive literature search was conducted across multiple electronic databases, including Web of Science and Scopus. The search strategy involved a combination of relevant keywords for "deep learning," "time series," "photovoltaic (PV) power," and "forecasting." The search was filtered to include articles published between 2022 and 2023, resulting in a total of 89 publications. After removing duplicate records, the remaining number of publications was 65. The titles and abstracts of the identified studies were screened against predefined inclusion and exclusion criteria. The full texts of potentially eligible studies were then thoroughly assessed for final inclusion in the review. The final set of included studies, comprising 36 publications, formed the basis for synthesizing the current state of research on deep learning techniques for time series PV power forecasting. This review article has its objectives as follows:

1. Demonstrate the trends in PV power forecasting techniques, with a primary focus on those utilizing deep learning models.
2. Compare different deep learning models with varying hyperparameter settings to elucidate the most promising configurations.
3. Showcase the variables utilized and their sources in published article to create a clear visual representation of their contributions and effects on the topic.
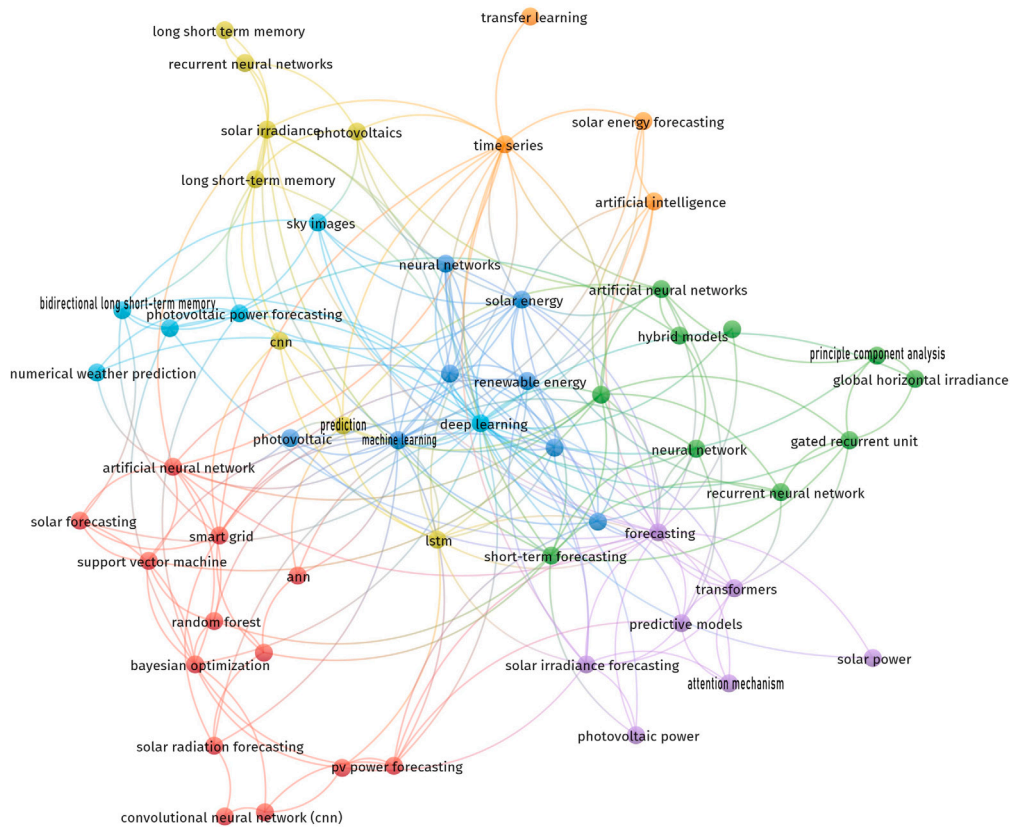
Fig. 3. Network connection of related publication in 2023.

4. Review evaluation metrics utilized in the latest publications to highlight current trends, with a focus on metrics that allow for global model comparison.
5. Identify gaps in this line of investigation to pave the way for future studies to be more robust and efficient.

The structure of this review article is organized as follows: In section 2, a detailed explanation is provided regarding the time series concept of data, deep learning techniques, and evaluation matrices. Section 3 compares recent case studies on photovoltaic (PV) forecasting using deep learning models in two subsections. Firstly, an evaluation of characteristic selection is conducted, encompassing input and output variables. Secondly, an overall comparison of published articles categorizes each proposed model setting to predict PV power output into four primary categories based on their architectural components: ANN-unit, RNN-unit, CNN-unit, and attention-unit. This comprehensive analysis showcases each model unit type, optimization techniques, and evaluation metrics to discern the most efficient, accurate, and commonly used models for PV power output time series data forecasting. In section 4, a comprehensive discussion highlighting challenges, trends, and future research directions is presented. Section 5 presents recommendations and conclusions.

2. Problem definition and deep learning models

The investigation into deep learning models for PV power forecasting consists of several key topics that will be thoroughly discussed in this section. Firstly, we delve into the study of time series data, focusing on the historical records of PV power production and its correlated variables over time. Secondly, the deep learning algorithm employed for the task can be further subdivided into the robustness of the deep learning core unit, the optimization technique implemented to enhance model accuracy, and the type of cost function steering the learning process of the model.

2.1. Time series data in PV power forecasting

The data structure for deep learning models can take various forms, including text (tokenized and converted into arrays of individual words for Natural Language Processing (NLP) 5.c tasks such as mapping human understanding of words and conversation [24]), images 5.b (digitized and represented numerically, e.g., through RGB values of pixels for computer vision tasks like recognition and manipulation [25]), or time series, each form determining the necessary methodology required to build the appropriate algorithm. Time series data, comprising chronologically measured observations 5.a, is central to PV power forecasting. It exhibits
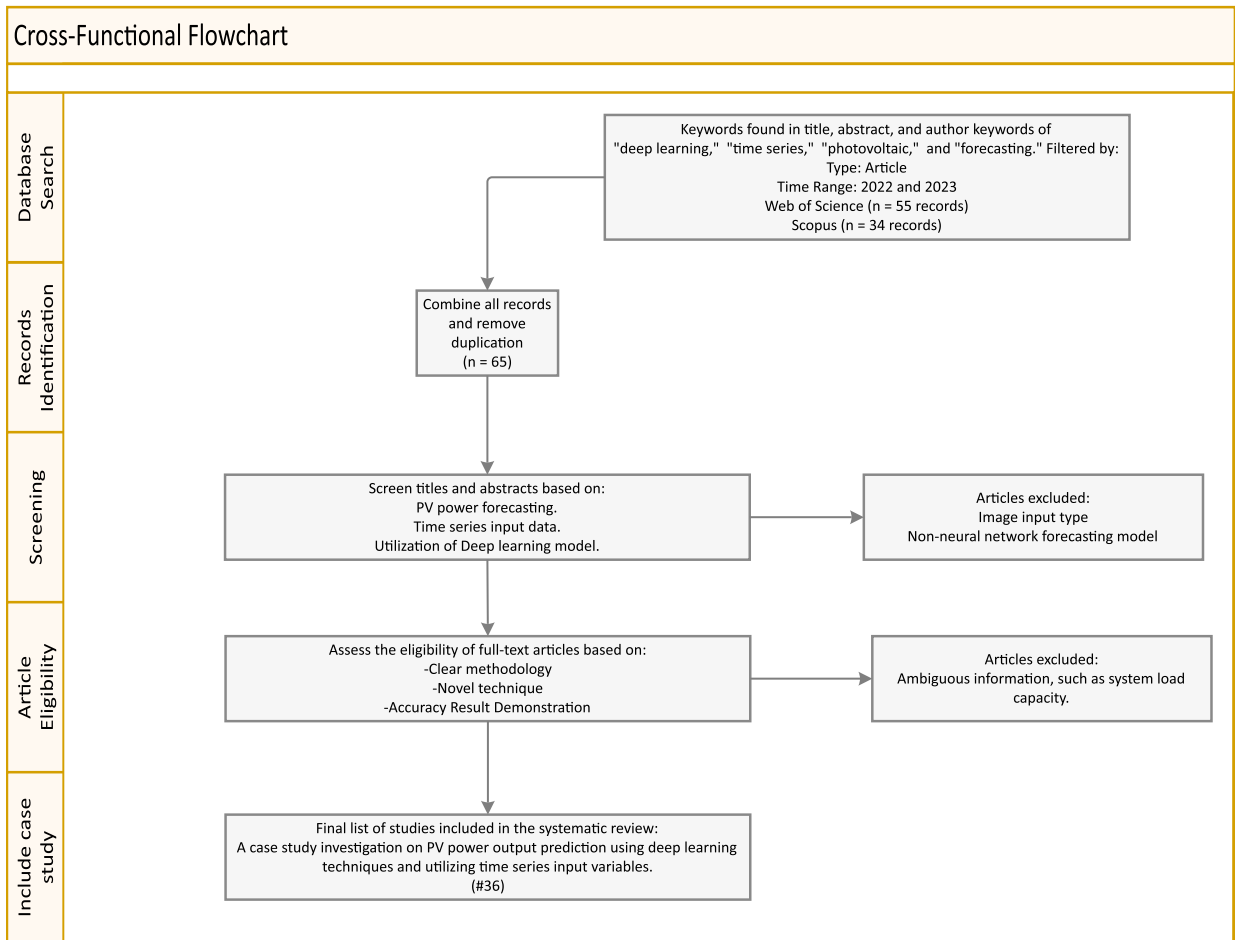
Cross-Functional Flowchart


Fig. 4. Systematic review methodology.

characteristics such as large size, high dimensionality, and temporal patterns like trends, seasonality, and random variations [26,27]. The forecasting objective is to predict specific features while gaining insights into influential variables. Time series data used for PV forecasting can be classified as univariate, containing historical power generation or irradiance, or multivariate, including meteorological variables like temperature, humidity, wind, and pressure. Additionally, cell parameters (temperature, voltage, current) and time indices may be incorporated. The temporal resolution of data, ranging from minutes to months, affects model complexity and accuracy. For recurrent models utilizing historical information, the sequence length is crucial. Instead of single inputs, data is organized into sequences, often using recurrent neural networks adept at memorizing sequences. Furthermore, the forecasting horizon ranges from ultra-short-term (seconds to hours) for network quality and demand response [28], to short-term (intra-day) for maintenance planning [29], medium-term (day-ahead), and long-term (beyond one day) for decision-making and investment planning [30]. Long-term forecasting is challenging due to the stochastic nature of cloud distribution, leading to increased errors.

2.2. Deep learning

This section provides an overview of a deep learning forecasting techniques. The review encompasses various aspects, including the model's architectural components, such as ANN-unit, RNN-unit, CNN-unit, and attention-unit; optimization techniques, including stochastic gradient descent, among others; evaluation metrics such as MAE, MSE, RMSE, MAPE, and R^2; that are published in the most recent research papers of this domain.

#### 2.2.1. Unit structure of deep learning models

The most basic form of the ANN is shown in Fig. 6. It was introduced by [31] in 1957, inspired by [32]'s works on modeling the nervous activity. It simulates the work of neurons in our nervous system. It consists of three parts: A single perceptron gets the input signal and processes it linearly through an activation function, resulting in an output in the range [0 1], as expressed mathematically in Equation (1) where W is for weight, b for bias, and n for number of inputs, it has proven to be able to capture sharp fluctuations in data behavior.

X_{t-7}	X_{t-6}	X_{t-5}	X_{t-4}	X_{t-3}	X_{t-2}	X_{t-1}	X_t
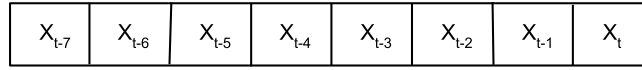
(a)　　　Time series data representation
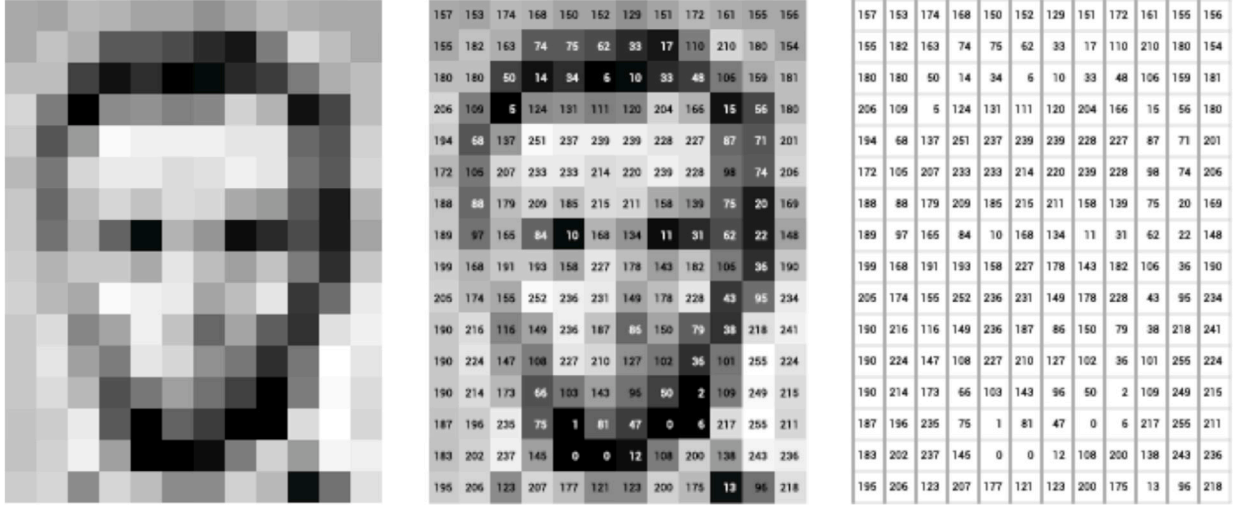
(b)　　　Image data representation
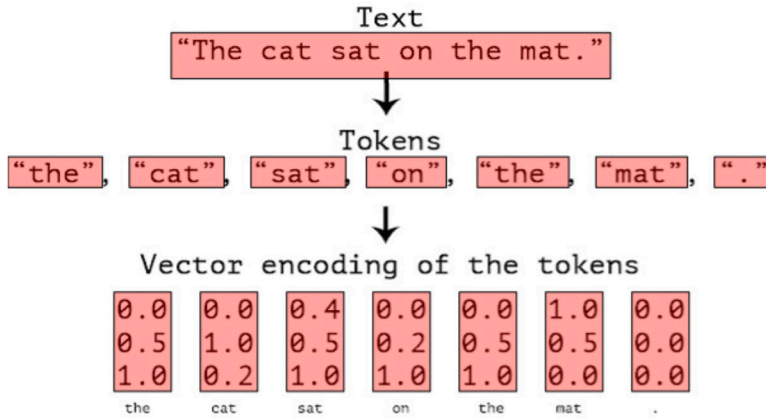
(c)　　　Word data representation

Fig. 5. Data representation.

$$y = f\left(b + \sum_{i=1}^{n} x_i W_i\right) \tag{1}$$

Multi-layer perceptron is one of the approach to predict using ANN. [33] published a paper demonstrating time series weather prediction using multiple ANN-units organized in layers. Many areas of application of MLP time series forecasting have been covered (i.e., economic trend prediction, natural language processing, voice recognition, and computer vision). One of the great implementations of PV power forecasting using multilayer neural networks was done by [34], mean daily data were used to predict the next day's generated power. It has achieved an accuracy of 98% coefficient of determination (R^2). However, the disadvantage of basic MLP is that it needs a large amount of data for the learning process, big computational resources, and complex network architecture [35].

RNN-unit is an enhanced version of the ANN unit that makes use of the strong correlations between historical records and the future target in time-series data. Temporal information between historical data is utilized in the recurrent unit. It was introduced by [36] in 1986. [37] demonstrated multiple ANN perceptron in series that are used to take a sequence input in which previous results are included to compute the current output, as shown in Fig. 7. It can be seen that every cell produces an output and a hidden state.
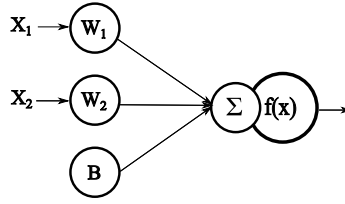
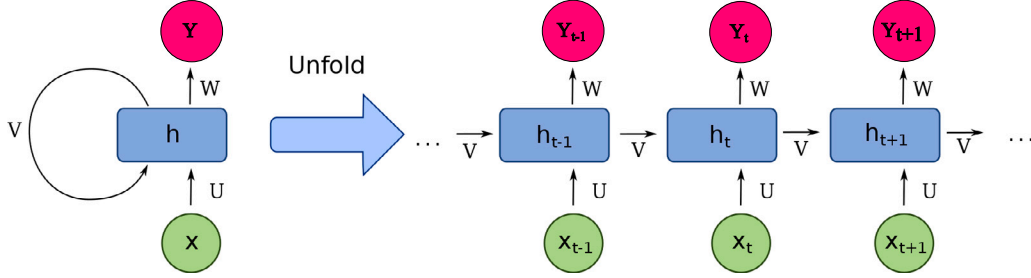**Fig. 6.** Connected neural network with two inputs and one bias.

Fig. 7. Schematic diagram of Recurrent cells.

The outputs of RNN-unit are calculated with Eq. (2) where h_{t-1} is the previous hidden state and V, U, and W are weights for the previous hidden state, the current input, and the current hidden state, respectively.

In PV power production, [38] Designed a cascade architecture consisting of RNN models that utilized hierarchical clustered data to tackle PV power fluctuations and designed an accurate model. [39] experimented with 3 RNN layers to predict 5 min, 15 min, 1 h, and 3 h ahead, with an input sampling of 12 time-steps. The input was imported from local sensors connected to the internet. The experiment aimed to focus on the weather effects on PV power forecasting accuracy, utilizing different sets of weather inputs and highlighted the impact of each type. [40] conducted several experiments on a forecasting model consisting of a Stacked ensemble RNN combined with another deep learning model like (ANN, SVR, LSTM, and CNN) for 1 and 3 days ahead power prediction, utilizing measured weather input data and their statistical features. However, normal RNN units could result in gradient vanishing and explosion problems while dealing with long sequences. Thus, many variations of recurrent models have been introduced, like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU).

$$h_t = f(V h_{t-1} + U x_t + b_h)$$
$$y_t = f(W_y h_t + b_y)$$

(2)

LSTM was designed by [41] in 1997, to alleviate the weaknesses of RNN. By the following couple of decades, it had won many prizes for achieving benchmark in speech and handwriting recognition [42]. One of the earliest published articles for predicting PV power generation using LSTM is by [43] that was used to predict a day-ahead using hourly weather data, resulting in a significant improvement compared to the previous mentioned models by more than 46%. The architecture of LSTM consists of four neural networks that are used to control three main tasks (Forget, input, and output gates). The LSTM is calculated as Eq. (3) where f_t represents forget gate output, i_t and g_t are neural networks that correspond with input data, and o_t is the result of the output gate. W is for weights, and t for the time step. σ and tanh are sigmoid and hyperbolic tangent activation functions, respectively. The Forget gate is to determine how much previous data should be carried out. The input gate has inputs x_t, h_{t-1}, and c_{t-1} which are model sequence input, previous hidden state, and previous cell state, respectively. Both x_t and h_{t-1} are being concatenated together to form the main input for all neural networks in the algorithm. The unit outputs are hidden state h_t and current cell state c_t.

$$Forget NN : f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$Input NN : i) \, i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$ii) \, g_t = tanh(W_g \cdot [h_{t-1}, x_t] + b_g)$$
$$Output NN : o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$Output : i) \, c_t = f_t * c_{t-1} + i_t * g_t$$
$$ii) \, h_t = o_t * tanh(c_t)$$

(3)

GRU is a lighter version of the LSTM developed by [44]. It was developed with 3 neural networks to control two gates (update and reset). The update gate is to select how much of the previous state should be preserved. The reset gate decides how to combine current input data with previous memory. u is for hidden states' weights. [45] has compared GRU unit with the previously mentioned units. The paper developed a model to predict long-term irradiance by using hourly and daily historical solar data. It found that error-

wise, GRU is slightly better than LSTM, performing faster and using significantly fewer computational resources. Eq. (4) expresses GRU mathematically where h_t is the output of GRU unit. [46] Experimented with a vanilla GRU to predict using pre-processed public weather and power production data, with a 1 h data resolution. The authors experimented with hyperparameter tuning and observed the impact on the prediction accuracy.

$$Reset\,NN : r_t = \sigma(W_r x_t + u_r h_{t-1} + b_r)$$

$$Pass\,NN : z_t = \sigma(W_z x_t + u_z h_{t-1} + b_z)$$

$$Memory\,State\,NN : \tilde{h}_t = tanh(W_h x_t + u_h r_t h t - 1 + b_h F)$$

$$Output : h_t = (1 - z_t) \cdot \tilde{h}_t + z_t \cdot h_{t-1}$$

(4)

Furthermore, [47] proposed a Bidirectional Recurrent unit to increase the model's accuracy, which can be trained using all available input data in the sequence (past to future samples and vice versa), between $t = 0$ and $t = t_{now}$. Output is the combination of two directions of processing. For instance, bidirectional LSTM is expressed mathematically in Eq. (5). $W_{\overrightarrow{h_y}}$ and $W_{\overleftarrow{h_y}}$ are weights for forward and backward LSTM units, $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are hidden vectors of forward and backward LSTMs, respectively, and b_y is output layer bias. [48] Conducted a series of experiments in prediction using a bidirectional LSTM for various forecasting horizons of 5, 15, and 30 minutes ahead. The experiments utilized actual PV power output data solely, in the absence of meteorological data. [49] Experimented with BiLSTM, Bi-GRU, and a hybrid CNN-BiLSTM model for 1-day ahead prediction of global horizontal irradiance. The authors utilized historical irradiance measurement data from the public NASA database without considering other meteorological variables. The authors took into consideration the models' parameters to fit and tune their proposed forecasting models.

$$y_t = W_{\overrightarrow{h_y}}\overrightarrow{h_t} + W_{\overleftarrow{h_y}}\overleftarrow{h_t} + b_y$$

(5)

CNN)-layers was introduced by [50]. It was inspired by the pattern of biological cortical neurons processing in the animal visual cortex, which focuses on a target area named the receptive field while ignoring the rest of the surrounding area. Although the CNN was created to deal with visual processing (images), [51] applied deep CNN to forecast timeseries PV power production for two farms in Belgium. The model forecasted five steps into the future with acceptable accuracy and robustness. In [52], the authors designed a CNN-BiLSTM prediction model, utilizing 1 year of meteorological input data from the target location and neighboring sites. The authors stated that the CNN was used to extract spatial features from the input variables, and the BiLSTM was added to extract the temporal characteristics. The process of the CNN algorithm is shown in Fig. 8. CNN extracts features (pixel trends and edges) from the input and groups them into small packets (convolutional operation). Two hyperparameters are related: kernel and filter. The kernel decides the number of elements in the input to be looked at once and slides over the input sequence, performing element-wise multiplication. Filter is the number of matrix features in the map (filters are randomly initialized). The CNN unit takes an input of size $N \times N$ and filters it using a matrix of size of $m \times m$, resulting in an output that has a size of $(N - m + 1) \times (N - m + 1)$ for convolution in the i-th layer. The procedure can be described mathematically in Eq. (6) where w, b, x, M, and $\otimes$ are the weight, bias, input, sub-matrix, and convolutional operation, respectively, "i" and "j" represent the spatial coordination, σ represents the activation function for the unit output. Pooling layer is often added next to the convolutional layer to reduce the size of the newly created future map, which will increase the speed of processing. It is tuned with three main hyperparameters: Filter size to decide the new matrix dimension, stride that is similar to masking, and average pooling. It is described in Eq. (7) where "f" denotes the filter and "s" refers to the stride.

$$h_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} \otimes x_{(i+a)(i+b)}$$

$$y_{ij} = \sigma(h_{ij})$$

(6)

$$Pooling\,layer\,Output : ((n_h - f)/s + 1) \times ((n_w - f)/s + 1)$$

(7)

Transformer Model was introduced by Vaswani et al. in 2017 [53], achieved state-of-the-art results in various natural language processing tasks, especially in language translation. The model utilizes The Attention mechanism, as proposed by Parikh et al. [54], to create a matrix representation of an input sequence by establishing connections between its elements. Unlike the sequential recurrence process in models like RNN, LSTM, and GRU, where the sequence is processed linearly, the Transformer processes the sequence in parallel using recurrent attention mechanisms, which has been proven to be a more efficient method for dealing with sequences. In [55], the authors demonstrated a prediction model utilizing an attention mechanism alongside CNN and LSTM layers. The model was designed to capture short-term and long-term temporal patterns to forecast day-ahead PV power output. A public dataset from the DKASC website was utilized, with the variables serving as input to this model. However, the authors stated the challenge of requiring huge computational resources when utilizing the attention mechanism. Saoud et al. [56] implemented a transformer to forecast energy consumption for households, using Time2vec positional encodings with weights for the input, and achieved state-of-the-art results. [57] developed a novel multi-model forecasting approach that utilizes sky-image-derived information, coupling the correlations between sky images and historical data to improve ultra-short-term forecasting performance (10 minutes ahead). Specifically, it employs the Informer model to encode historical and empirically estimated clear-sky global horizontal irradiance (GHI) data. The ground-based sky images are transformed into optical flow maps, which can be processed by a vision transformer. An attention mechanism is proposed to explore the coupling correlations between these two modalities. Finally, a generation decoder
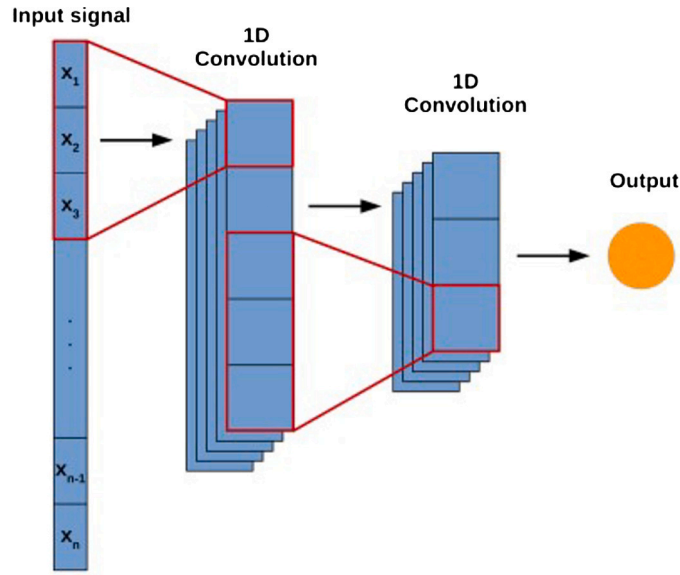
Input signal

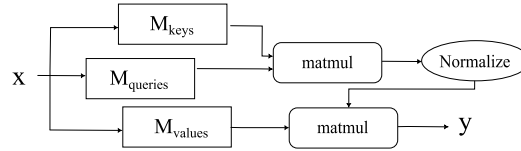**Fig. 8.** Convolutional schematic.

Fig. 9. Attention mechanism schematic.

is used to implement multi-step ultra-short-term forecasting. The architecture of Transformers is constructed of stacked Attention blocks (Multi-Head Attention). Fig. 9 demonstrates the attention mechanism. This mechanism involves scaling the dot product of three copies of the input (Queries, Keys, and Values) and computing the output matrix as shown in Eq. (8). In this equation, each element in the Query matrix is multiplied with all elements in the Keys matrix and divided by the square root of the dimension of k denoted by d_k. To obtain weights for the Values matrix, a Softmax function is applied to the resulting matrix. Finally, the subspaces W_i^Q, W_i^K, and W_i^V are concatenated and linearly passed with the weight matrix W^o to obtain the final output.

$$Attention(Q, K, V) = softmax((\frac{QK^T}{\sqrt{d_k}})V)$$

$$Multihead(Q, K, V) = Concat(head_1, ..., head_t)W^o \qquad (8)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

In addition to the attention mechanisms, the Transformer model incorporates Encoder-Decoder layers, as presented by [44], which have proven to be a powerful approach in sequence-to-sequence modeling. The encoder is responsible for processing the input sequence, which typically consists of time-based serial data, and converting it into a compressed representation known as the latent space. The latent space is a lower-dimensional representation that captures the essential information from the input sequence while removing unnecessary details. This compression enables the model to focus on the most important aspects of the input. On the other hand, the decoder takes the latent space representation produced by the encoder and generates the output sequence. In the context of sequence prediction, the decoder is given a one-time-shifted target sequence as an additional input. This means that at each time step, the decoder receives the elements of the target sequence up to the present time, without having access to future elements. By incorporating the target sequence into the decoder's input, the model can utilize the ground truth information during training, which helps improve the quality of the generated sequence. This approach allows the model to capture dependencies and patterns in the data, making it effective in various sequence-to-sequence tasks. As shown in Fig. 10, the encoder is structured with three main components: an input, a positional encoding, and a stack of identical encoder layers. Positional encoding incorporates sequential information for the time series data and is applied using sine and cosine functions. This involves adding the positional encoding vector element-wise to the input vector. The resulting matrix is then passed through the encoder layers. Each encoder layer consists of two sub-layers: a Multi-Head Attention and a fully connected feed-forward layer. Following each sub-layer, a normalization layer is applied. The output matrix of the encoder is then passed into the decoder. The decoder block is structured similarly to the Encoder block with an input, decoder, and output layer, containing two sub-layers: Multi-Head Attention and a fully connected FFNN layer.
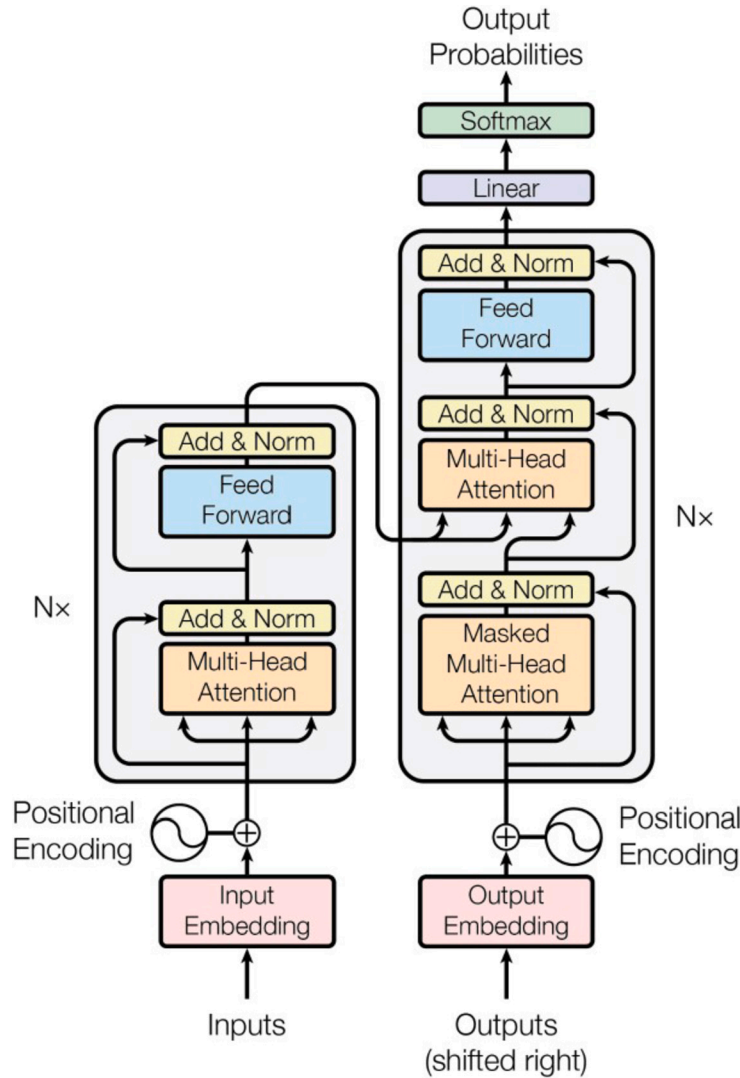
Fig. 10. Transformer architecture.

Additionally, the decoder has a third sub-layer that applies self-attention mechanisms to the encoder output matrix. To prevent any future information leakage, the decoder implements look-ahead masking, ensuring that the prediction of a time series data point is solely based on preceding data points. Positional encoding is used to retain the temporal information of the sequence while enabling parallel computation. It is added at the beginning, before the encoder and decoder layers. The positional encodings used in this model are sine and cosine functions with varying frequencies. They have the same dimension as d_{model}, allowing them to be added element-wise to the input sequence, as shown in Eq. (9). Where "pos" denotes the position and "i" refers to the dimension. The sine-cosine function was employed because it enables the model to associate each element with its corresponding time position, thereby facilitating the identification of correlations.
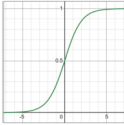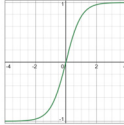
$$PE_{pos,2i} = \sin(pos/10000^{2i/d_{model}})$$
$$PE_{pos,2i+1} = \cos(pos/10000^{2i/d_{model}})$$
(9)

2.2.2. Hyperparameters

Determining the appropriate hyperparameters for deep learning models is crucial for achieving high accuracy in time series forecasting of photovoltaic (PV) power generation. Hyperparameters can be categorized into two types: those related to the model's architectural design and those related to the learning process [58]. Hyperparameters pertaining to the model's architectural design include the number of hidden layers, the number of units (or neurons) in each layer, and the type of activation function employed in the neurons. These hyperparameters significantly influence the model's capacity and complexity, affecting its ability to capture intricate patterns in the data. On the other hand, hyperparameters related to the learning process such as the number of epochs

Table 1
Most popular activation functions.

Activation function	Function	First derivative	Advantage	Plot
Sigmoid(Z)	$\frac{1}{1+e^{-Z}}$	$sigmoid(Z) * (1 - sigmoid(Z))$	Smooth gradient	
Tanh(Z)	$\frac{e^Z - e^{-Z}}{e^Z + e^{-Z}}$	$tanh^2(Z)$	Zero centric output	
ReLU(Z)	$max(0, Z)$	1	Computationally efficient	

(iterations over the entire dataset), learning rate, and batch size, govern the optimization and training of the deep learning model. Additionally, selecting a suitable optimization technique is a key hyperparameter that can enhance the model's performance and convergence. This section discusses the hyperparameters commonly utilized in deep learning models for time series forecasting of PV power and their impact on the model's accuracy.

2.2.2.1. Design-related hyperparameters   Number of hidden layers, the number of units (or neurons) in each layer, and the type of activation function are hyperparameters related to the model's design. Furthermore, the internal flow connections are important. Techniques such as Residual Networks (ResNets) [59] are utilized to overcome the issue of vanishing gradients by skipping layers. Another popular network connection is the Densely Connected Convolutional Network (DenseNet). The innovation in DenseNet is that it connects each layer to every other layer in the network instead of only feed-forwarding to the next layer.

Activation functions play a crucial role in introducing non-linearity into deep learning models, enabling them to learn complex relationships in the data. The most commonly used activation functions are shown in Table 1, sigmoid, tanh, and ReLU, while Z represents the neural input signal. Sigmoid is the most frequently used function; it outputs a signal between [0, 1] and provides a smooth gradient through its first derivative. However, its output is not zero-centered, which can make the model unstable. Tanh is an improved version that centers the signal around zero. However, this comes at the cost of increasing the steepness of the gradient, often leading to the problem of vanishing gradients. ReLU is a computationally efficient activation function that outputs a signal in the range of [0, Z] and has a derivative of 1. However, this efficiency reduces the model's capability to fit the data more accurately. Other activation functions, such as the softmax for output layers in classification tasks or the leaky ReLU, which addresses the dying ReLU problem, can also be considered depending on the specific requirements of the time series forecasting task for PV power.

2.2.2.2. Learning-related hyperparameters   Selecting the appropriate optimization technique and tuning hyperparameters related to the learning process are crucial for enhancing the performance and efficiency of PV power generation forecasting models. Batch size, which refers to the number of training samples propagated through the model in each iteration, is a key hyperparameter for training deep learning models using mini-batch gradient descent, a variant of the Stochastic Gradient Descent (SGD) optimization algorithm. A smaller batch size generally leads to more frequent updates of the model parameters, which can sometimes result in better convergence but also increased computational overhead. Conversely, larger batch sizes can lead to smoother convergence but may get stuck in sub-optimal solutions. The learning rate is another critical hyperparameter that determines the step size at which the optimizer updates the model's weights and biases during the training process. SGD, which stands as the most popular optimization technique for training deep learning models, including those used for time series forecasting of PV power, utilizes backpropagation to navigate the learning process and reduce the error. Other common optimization techniques include RMSprop, particle swarm optimization (PSO), gray wolf optimization, Bayesian optimization (BO), and the Marquardt algorithm.

Stochastic Gradient Descent (SGD) stands as the most popular optimization technique in contemporary deep learning. It finds its roots in the mathematical gradient descent iterative algorithm pioneered by [60] and [61], which successfully located local minima of differentiable functions. However, this algorithm systematically processes every individual data point, a practice that, when dealing with machine learning big data, demands substantial computational resources. To address this challenge, the stochastic variant of the algorithm was introduced by [62] and further elucidated in [63]. This approach significantly reduces the computational burden by selecting random data points for each iteration.

In this algorithm, the first derivative of the cost function is computed to determine the direction in which weights and biases should be updated. Referring to Eq. (10), where θ represents the weights or biases, they are updated by adding or subtracting the gradient. The learning rate (η) and the derivative of the cost function (J) jointly determine the magnitude of the gradient descent.
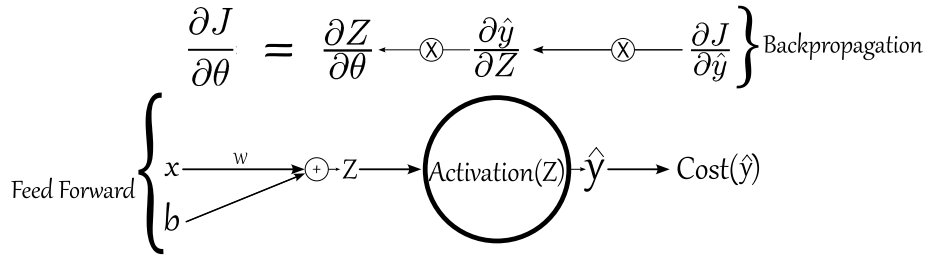
Fig. 11. Feed forward neural network and gradient backpropagation schematics.

$$\theta_{new} = \theta_{old} - \eta \cdot \frac{\partial J}{\partial \theta} \tag{10}$$

Fig. 11 demonstrates the algorithmic flows of feedforward and backpropagation for a single perceptron network. For explanation purposes, the MSE cost function and the ReLU activation function were selected as examples of how the gradient of a single perceptron is calculated. The derivatives of the MSE cost function are computed as follows: $\frac{\partial J}{\partial \hat{y}} = \frac{\partial(\frac{1}{2}(\hat{y}-y)^2)}{\partial \hat{y}} = \hat{y} - y$. For the ReLU activation function, the derivative is $\frac{\partial \hat{y}}{\partial Z} = 1$. The weight gradient is calculated using Eq. (11), and similarly, the bias gradient is computed using Eq. (12).

Gradient for weight: $\frac{\partial Z}{\partial W} = \frac{\partial(b+xW)}{\partial W} = x$

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial Z} \frac{\partial Z}{\partial W} = (\hat{y} - y) \bullet (1) \bullet (x) \tag{11}$$

Gradient for bias: $\frac{\partial Z}{\partial b} = \frac{\partial(b+xW)}{\partial b} = 1$

$$\frac{\partial J}{\partial b} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial Z} \frac{\partial Z}{\partial b} = (\hat{y} - y) \bullet (1) \bullet (1) \tag{12}$$

The cost function, also known as the loss function, is a crucial component in training deep learning models as it evaluates the model's performance by measuring the difference between the actual samples in the dataset and the model's predictions. It quantifies the error between the labeled and predicted values, presenting it as a single scalar value. For regression tasks, such as time series forecasting of PV power generation, the Mean Squared Error (MSE) is more commonly employed as the cost function compared to the Mean Absolute Error (MAE). MSE calculates the average squared difference between the predicted and actual values, as shown in Equation (13), where $\hat{y}i$ and y_i represent the predicted and true values, respectively, for the ith sample. MSE is particularly sensitive to outliers due to the squaring operation. In contrast, MAE measures the average absolute deviation between the predicted and actual values, as shown in Equation (14), without squaring the errors. While MAE is less sensitive to outliers, it does not provide as much weight to larger errors compared to MSE.

$$MSE : \frac{1}{2N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{13}$$

$$MAE : \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \tag{14}$$

The choice of the cost function can significantly impact the model's training and performance, especially in the presence of noise or outliers in the PV power time series data. Therefore, understanding the properties and implications of different cost functions is essential for developing accurate and robust forecasting models in this domain. Similarly, other optimization algorithms are also utilized and investigated, such as PSO [64], GWO [65], BO [66], and the Marquardt algorithm [67]. Weights and biases are updated in every iteration, thus minimizing the error of the model. The learning rate, number of perceptrons, epochs, types of cost function, and activation function are the hyperparameters for a deep learning model. Tuning these hyperparameters is the main focus in achieving robust and accurate models.

2.2.3. Performance evaluation methods

For showcasing the performance of the deep learning model, it is essential to use evaluation metrics. Publications in this topic mostly use two primary groups of metrics: accuracy metrics, and error metrics. Error metrics quantify the difference between the predicted values and the true target values, while accuracy metrics assess how closely the predicted values match the ground truth. The most widely used error measures, mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and root mean square error (RMSE), each have distinct interpretations. MAE has a key advantage of its intuitive interpretability. The MAE directly measures the average magnitude of errors in the same units as the original data, making it straightforward to comprehend the implications of the error value. Moreover, the MAE treats all errors equally, without disproportionately penalizing larger errors. [68] However, it does not inherently facilitate direct comparisons of model performance across different datasets or

Table 2
Evaluation Metrics.

Metric	Abbreviation	Formula
Mean Absolute Error	MAE	$\frac{1}{N} \sum_{i=1}^{N} \lvert y_i - \hat{y}_i \rvert$
Mean Squared Error	MSE	$\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$
Mean Absolute Percentage Error	MAPE	$\frac{1}{N} \sum_{i=1}^{N} \frac{\lvert y_i - \hat{y}_i \rvert}{y_i} \times 100$
Root Mean Squared Error	RMSE	$\sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$
Normalized Root Mean Squared Error	NRMSE	$\frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}}{y_{max} - y_{min}}$
Coefficient of determination	R^2	$1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$

target variables. The magnitude of the MAE is inherently tied to the scale and units of the specific dataset used for evaluation. MAPE offers a valuable advantage in its ability to provide a scale-independent and unitless measure of error while still preserving interpretability. By expressing the error as a percentage of the actual values, MAPE eliminates the influence of the target variable's unit of measurement, making standard metrics useful when comparing models across different datasets [69]. On the other hand, MAPE should be interpreted cautiously due to its tendency to exhibit a bias towards smaller values in the dataset. When the actual or observed values are relatively low, even minor absolute errors can result in disproportionately high percentage errors, effectively inflating the MAPE metric [70]. This can be problematic when dealing with datasets that contain seasonality with both low and high values, such as in the case of generated power forecasting.

Furthermore, MSE is known to be sensitive to outliers due to the squaring of errors in its calculation. Its sensitivity stems from the fact that MSE amplifies the impact of larger errors by squaring them, effectively giving more weight to extreme deviations from the true values [71]. Consequently, the presence of outliers in the dataset can significantly inflate the MSE, potentially distorting the overall assessment of the model's performance. This characteristic makes MSE more suitable for applications where outliers are expected to be infrequent or when the primary concern is minimizing the impact of large errors. On the other hand, RMSE offers a distinct advantage in its ability to provide an error metric that is directly interpretable in the same units as the target variable. By taking the MSE, RMSE effectively transforms the squared errors back to the original scale, allowing for a more intuitive understanding of the average magnitude of errors [72]. This property makes RMSE particularly valuable when evaluating model performance on datasets with high volatility or variability, such as weather time series data. Though, its interpretability is related to the dependent variable, making it challenging to compare across different datasets and models. Furthermore, The Normalized Root Mean Squared Error (NRMSE) is intended to facilitate comparisons of RMSE across multiple datasets and different forecasting models. RMSE is normalized by either the mean or the range of the true values. However, while NRMSE is often used as a standard metric for universal comparison, its interpretability depends on the chosen normalization method.

Regarding accuracy metrics, the coefficient of determination (R^2) offers a key advantage in its intuitive interpretability as a measure of goodness-of-fit between the predicted and actual values. Specifically, R^2 quantifies the proportion of variance in the dependent variable that can be explained by the independent variables in the model [73]. Expressed as a value between 0 and 1, or alternatively as a percentage, R^2 provides a readily comprehensible indication of how well the model's predictions align with the ground truth data. Although R^2 gives an instant measure of model performance, it can be misleading due to potential overfitting issues that may be present in the results. Selecting appropriate evaluation metrics is crucial, as it allows for a fair and objective comparison of different forecasting approaches.

The most commonly utilized evaluation metrics among researchers for time series forecasting models are MAE (Eq. (14)), MSE (Eq. (13)), MAPE, RMSE, NRMSE, and R^2, as presented in Table 2. To comprehensively assess the performance of their proposed prediction models, investigators typically employ a combination of these metrics to provide an interpretable and multifaceted evaluation of the models' accuracy and efficacy. By leveraging the strengths of each metric, researchers can gain insights into different aspects of the models' performance, such as the average magnitude of errors, the sensitivity to outliers, and the goodness of fit, ultimately facilitating a more robust and comprehensive analysis.

$\hat{y}_i$ denotes predicted PV power output, y_i denotes actual value, $\bar{y}$ denotes mean of the observed data, and N denotes the total number of samples.

3. State of the art

This section will explore the current state of PV power forecasting and the existing gaps in the field. Fig. 12 displays the types of models utilized in case studies published in 2022 and 2023. LSTM, CNN, and MLP were featured in PV power forecasting publications either as proposed models or as benchmarks for evaluations, with 31, 21, and 7 publications, respectively. Additionally, both BiLSTM and GRU were used in nine publications each. Notably, compared to previous years, the state-of-the-art attention-based Transformer has garnered more publications in the field, totaling 7 articles. Moreover, new models are emerging, such as the Autoformer, specifically in our domain. In contrast, basic RNN models have seen a significant decline, and many classical machine learning models have vanished.
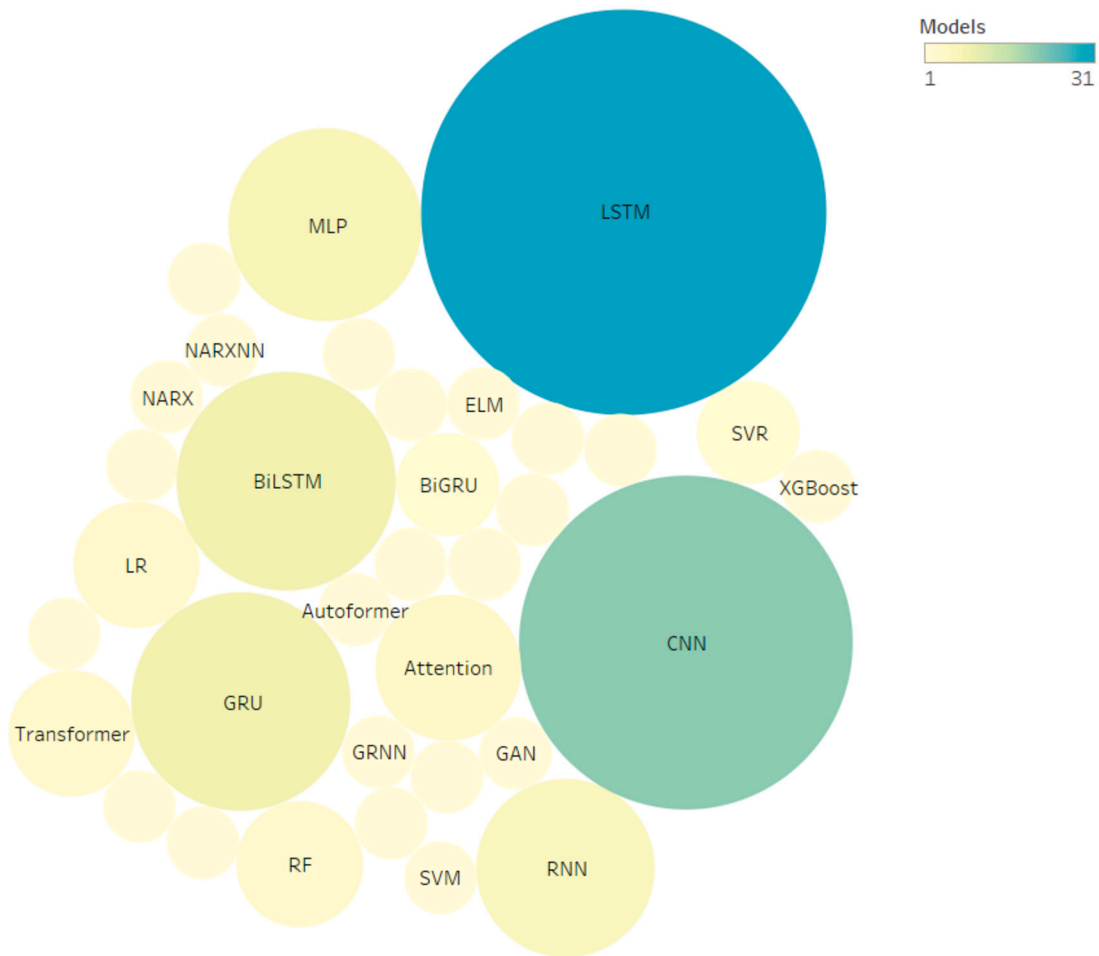
Fig. 12. PV forecasting models recent published articles (2022 and 2023).

From these publications, a selection of 36 case studies implementing unique algorithms to accomplish this task will be reviewed in two sections: an analysis of the system characteristics employed and an overall comparison of the proposed models.

3.1. Model characteristics analysis

This section demonstrates the literature statistics including top productive countries, datasets source, input variables, input length, and output forecasting horizon. Fig. 13 illustrates the distribution of implemented case studies worldwide. The United States leads in the number of case studies, followed by Australia and South Korea in second place, and China and the UK in third place. For this section, six main system parameters - input variable types and quantities, data granularity, lookback length, output types, and forecasting horizon - will be analyzed.

3.1.1. Data variables characteristics

Datasets play a crucial role in conducting investigations on this topic. Therefore, it is essential to review the sources of the datasets used in the selected case studies. Fig. 14 illustrates the sources of the datasets in these studies. Twenty-three articles have developed forecasting models using self-collected datasets, obtained from installed PV systems, as seen in [74]. On the other hand, public datasets are an alternative for developing forecasting models, even though they might not be specific to precise locations. Several online databases were utilized, including NASA like in [75], NREL in [76], and weather station datasets in [77]. Additionally, obtaining datasets through physical calculations was implemented in a few articles, as demonstrated in [78].

Fig. 15 illustrates the various types of variables utilized for PV power forecasting in recent publications. The selection of these variables depends on their correlations with main PV power production, availability, and ease of access to such data. The most common variable used is measured power, employed in 67% of the recently published articles, such as [79–83]. This direct method captures the energy production behavior at specific times and locations. Although PV power generation exhibits daily and yearly cycles, it often fluctuates unpredictably due to various factors. Hence, authors are incorporating additional variables to fill gaps and enhance prediction accuracy. Solar irradiance variables, including global horizontal, direct, and diffused irradiances, are widely used
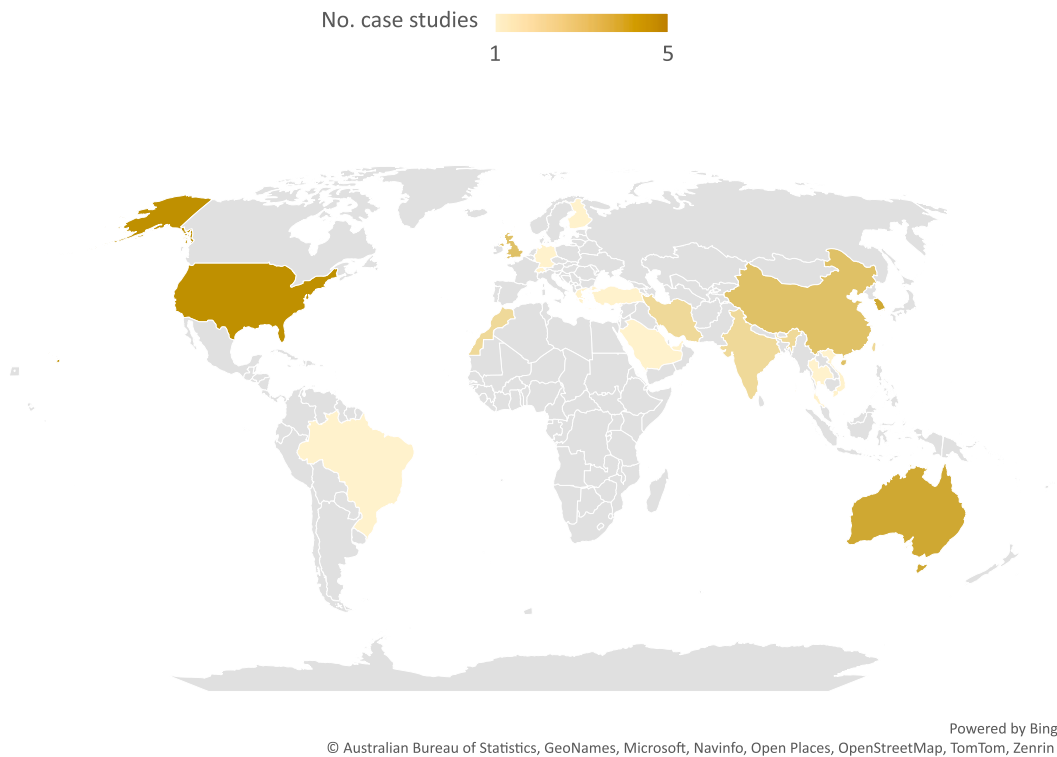
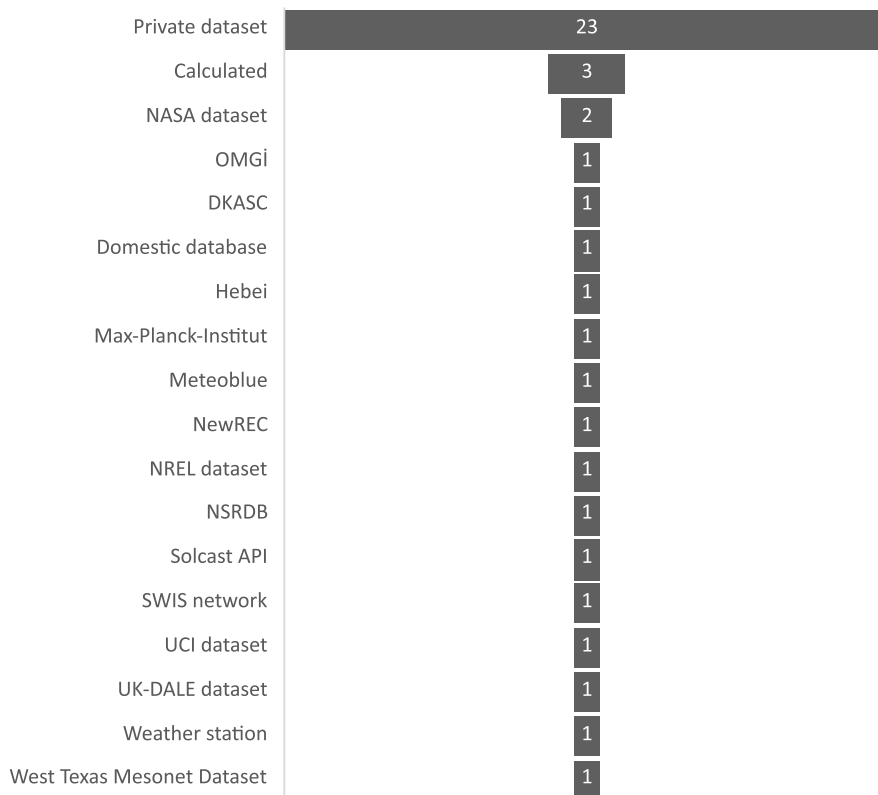**Fig. 13.** The distribution of the selected 36 case studies around the world.

Fig. 14. Sources of datasets utilized in selected case studies.
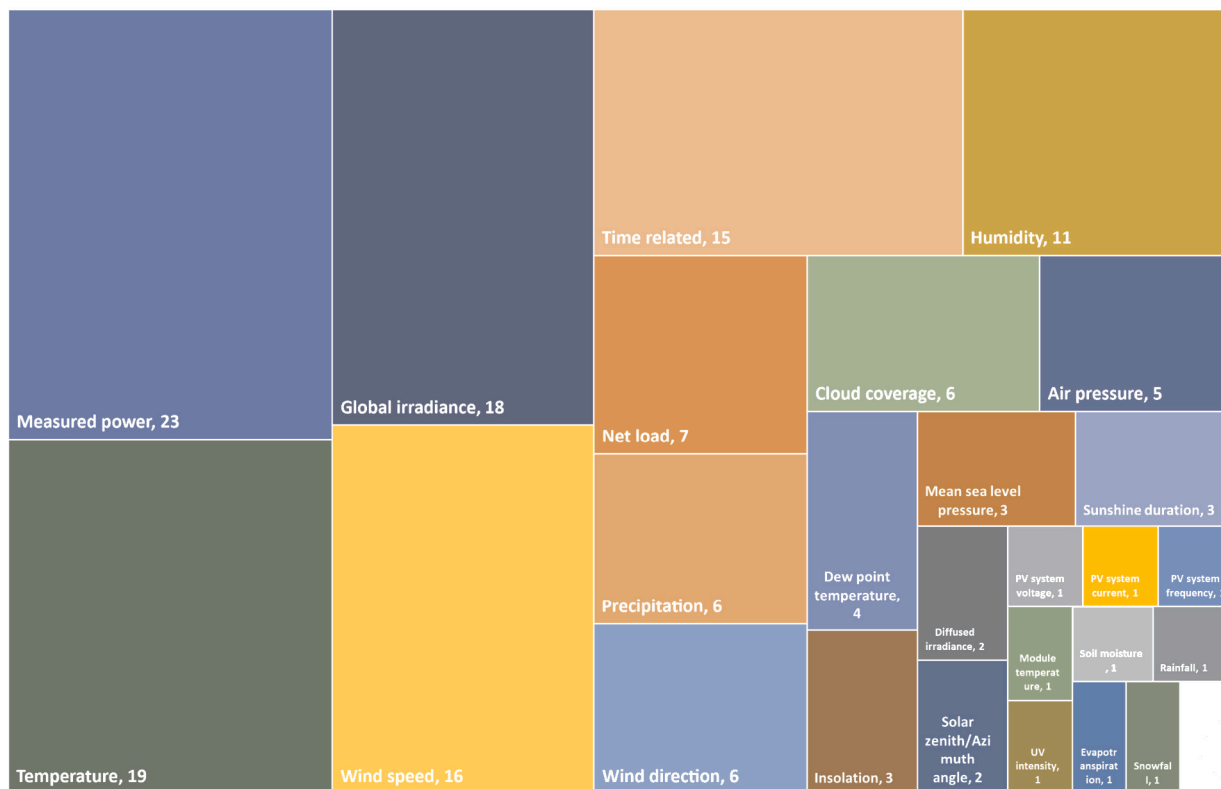
MODEL VARIABLES

Fig. 15. Variables types of recent published papers, (count).

in articles such as [78,84–87]. These variables enable the prediction model to capture more insights or compensate for the lack of historical measured PV power production records. They exhibit very high Pearson correlation coefficients (> 0.9). Furthermore, other variables are often implemented. Temperature variables are among the most popular, featured in recent publications like [88–91,75], due to their easy accessibility and high Pearson correlation coefficients, which can reach 0.9. Similarly, wind speed variables, as implemented in [92,81,77], and humidity [81,93], which have moderate Pearson correlation coefficients, are also utilized. Time-related variables, presented in [94,95], such as day of the year, datetime, and trigonometric representations of time, are frequently used to extract insights from the cyclic behavior of solar radiation produced by the sun. Other variables capturing insights from human activity, such as net load data demonstrated in articles like [96,56], or PV module parameter trends including the array's voltage, current, temperature, etc., showcased in [97], are also employed.

The number of variables varies from author to author, as illustrated in Fig. 16. Recent papers on this topic have utilized anywhere from single input variables to as many as 17 variables. Univariate forecasting models that use a single variable are the most commonly implemented, as observed in publications like [98,85,99]. This method is convenient due to the low dimensionality of the input and the high correlation between historical data and future results. In contrast, multivariate systems can capture additional trends from historical data, which can include variables like weather, time features, and other indirectly related factors that might influence the system. However, increasing the dimensionality of the input data also escalates the processing load. Such systems, utilizing five variables, were frequently employed, as evidenced in [92,100,101]. It's noteworthy that the highest number of variables used in a study was 17, as demonstrated in [87].

3.1.2. Characteristics of model inputs

Data granularity is an important parameter that affects model performance and computational requirements. In Fig. 17, the prevalent data granularities used in the compared articles have been shown as a function of number of published articles in which they have been utilized. In this set of articles, Hourly data is the most commonly employed, featuring in 19 different papers such as [88, 80,85,86,83]. This granularity strikes a balance between including substantial information in the model and managing computational resources effectively. The shortest data resolution used is 5 minutes, found in [92,102,56]. While this fine granularity provides highly accurate information for modeling, it significantly increases computational demands due to the large volume of data. On the other end of the spectrum, monthly data granularity is the longest resolution, noted in two publications [96,93]. This granularity is well-suited for long-term predictions but necessitates extensive datasets for effective forecasting. Fifteen-minute data granularity ranks second in popularity, presented in articles such as [79,94,77,99]. Daily granularity follows closely behind demonstrated in [98,91,75].
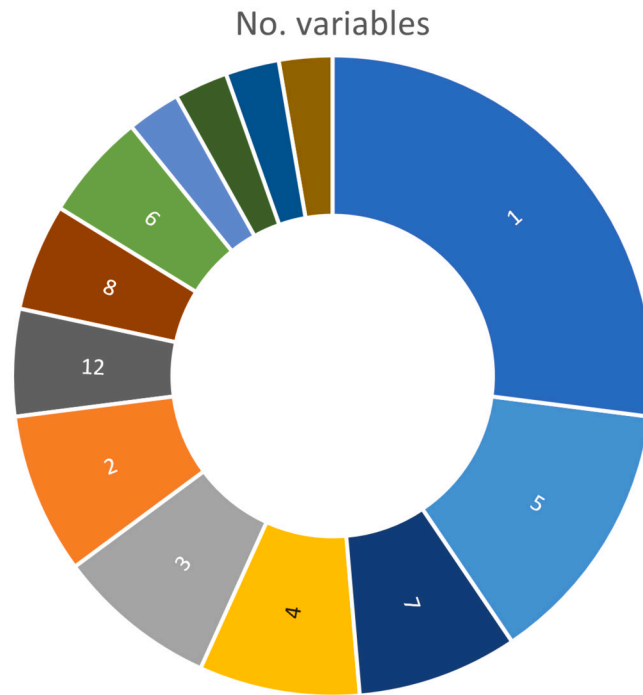
Fig. 16. Number of variables utilized in recent published papers.
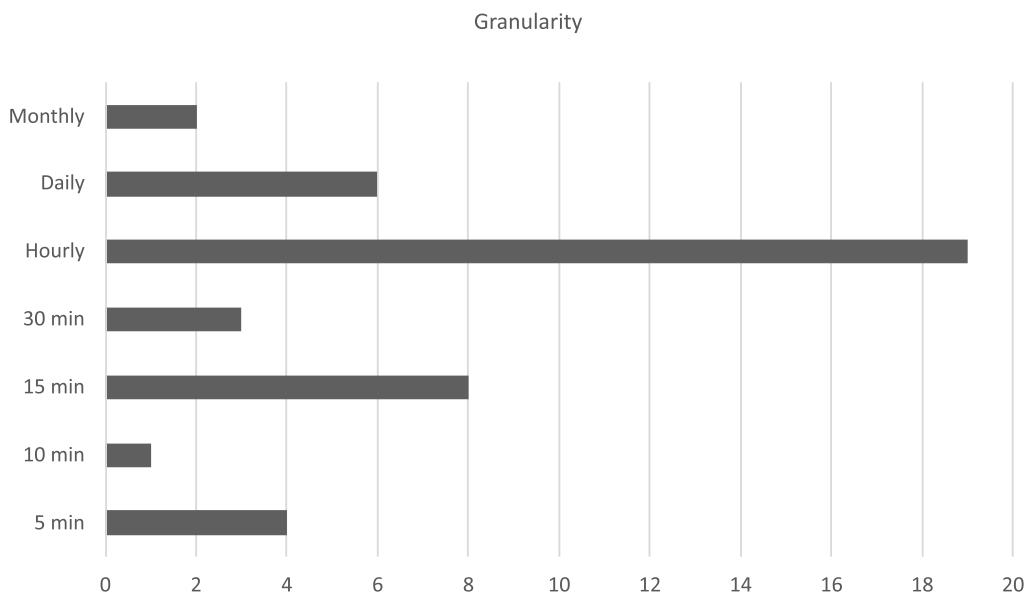
Fig. 17. Data resolutions of recent published papers.

Interestingly, 10-minute granularities were the least favored option, appearing in only one publication [97]. This variation in data granularity highlights the trade-off between data accuracy and computational resources in PV power forecasting models.

The input sequence length is a crucial parameter in time-series data forecasting, containing historical information essential for predicting the future. This length can vary widely, from as short as a single previous historical element to as long as the entire dataset. Fig. 18 illustrates the different input sequence lengths used in recent papers. The longest sequence utilized comprised 26,298 past samples by [97], incorporating half a year of data with a resolution of 10 minutes, encompassing conventional daily, weekly, and monthly trends. It is noteworthy that the most commonly proposed lookback steps are 96 previous time steps, as demonstrated in [94,103]. This represents a one-day time frame using a 15-minute sampling rate for input data. It enables a more accurate daily behavior of the input variable in the predictive model, making it particularly suitable for short- to medium-term
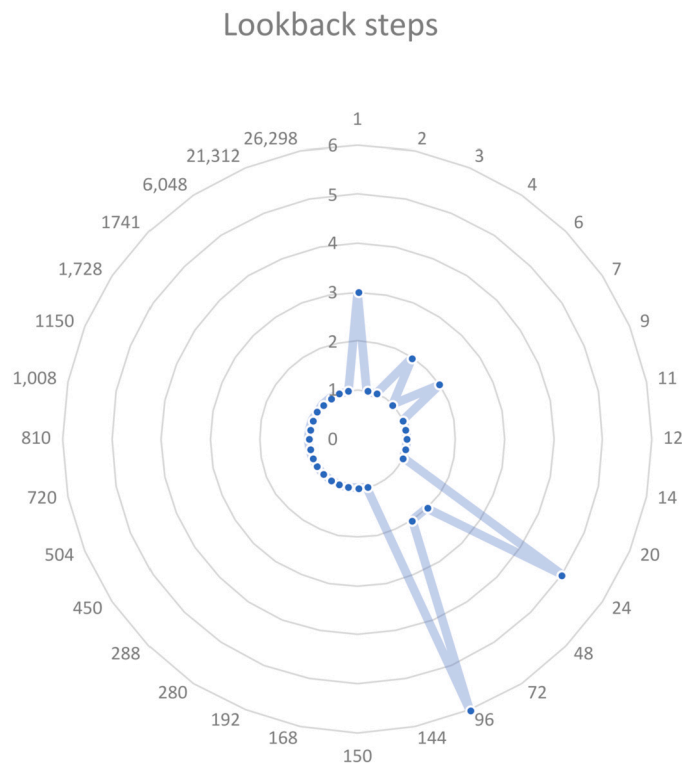
Fig. 18. Input sequences length of recent published papers.

forecasting applications, where predicting daily trends is crucial. Similarly, 24-step lookbacks were implemented in five articles, such as [93,85], corresponding to one day of hourly data. However, this approach uses less data, potentially affecting the accuracy of forecasting results, in exchange for a reduced computational burden. Other input sequence lengths were utilized, such as 1 by [78,94], followed by 4 and 7 input sequence lengths by [104,75], and 48 and 72 lookback steps in [105,101], where the authors aimed to capture specific trends within certain time frames for various operational and management applications.

3.1.3. Characteristics of model outputs

The outputs generated by models in the compared articles on PV power forecasting using deep neural networks can be categorized into direct PV forecasting or indirect PV forecasting, which predicts highly correlated variables such as solar irradiance, temperature, or wind speed, as illustrated in Fig. 19. A majority of works (61%) adopted a direct approach, accomplishing the task by setting the model's label as historical PV generation records, as demonstrated in studies like [79,80,90,82,83]. In contrary, while setting the model's label based on historical PV generation offers the highest accuracy, obtaining real PV historical data from specific locations proves to be extremely challenging. Therefore, indirect methods have been explored. Solar irradiance stands out as the most correlated variable to PV power production. Articles such as [89,85–87] illustrate PV power forecasting methods by setting the model's label to solar irradiance. Similar approaches have been used with different variables such as temperature [93] and weather [84]. This underscores the advantage of time series data, allowing tasks to be achieved using various datasets with specific levels of correlation, especially in situations where data scarcity is a challenge.

Fig. 20 illustrates the PV model forecasting horizons implemented in recent publications. Here, the numbers indicate the steps into the future, irrespective of the granularity of the data. As anticipated, the majority (15 articles) predicted a one-time step ahead since achieving high accuracy is less complicated in this scenario. This approach has been demonstrated in articles such as [79,89,90,86,106]. Forecasting 24 steps ahead can provide more valuable insights for power applications while considering moderate use of computer resources. This method, exemplified in papers like [78,107,100,74], generally corresponds to predicting one day ahead with hourly sampled data. Similarly, forecasting one day ahead using 15-minute data granularity resulted in the utilization of 96 samples, as seen in works like [88,94,82]. Although implementing higher resolution increases the computational load, researchers believed it was worth it for gaining more accurate insights into future events. Researchers have explored various forecasting horizons, including 7, 12, and 2016 next samples ahead in studies like [75,87,102], and 4, 6, 288, and 336 forecasting horizons in papers such as [108,97,102,109]. This exploration showcases the effort to understand the effect of different forecasting horizons on enhancing the model's accuracy and gaining valuable insights into future events. Notably, the longest forecasting horizon utilized was 16,128 next samples by [92], corresponding to two months ahead with a data granularity of 5 minutes.

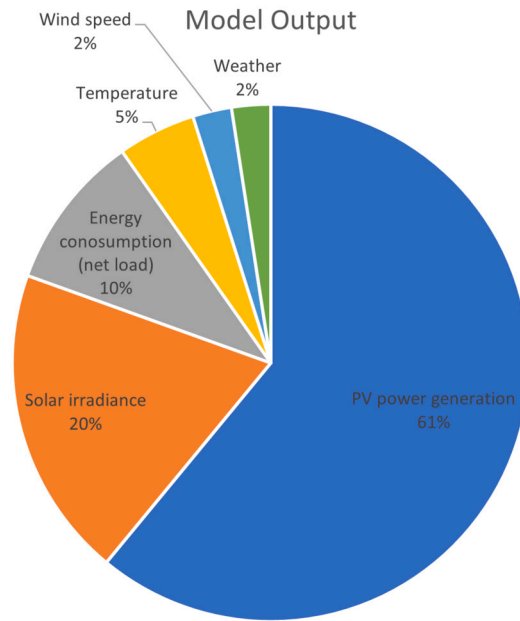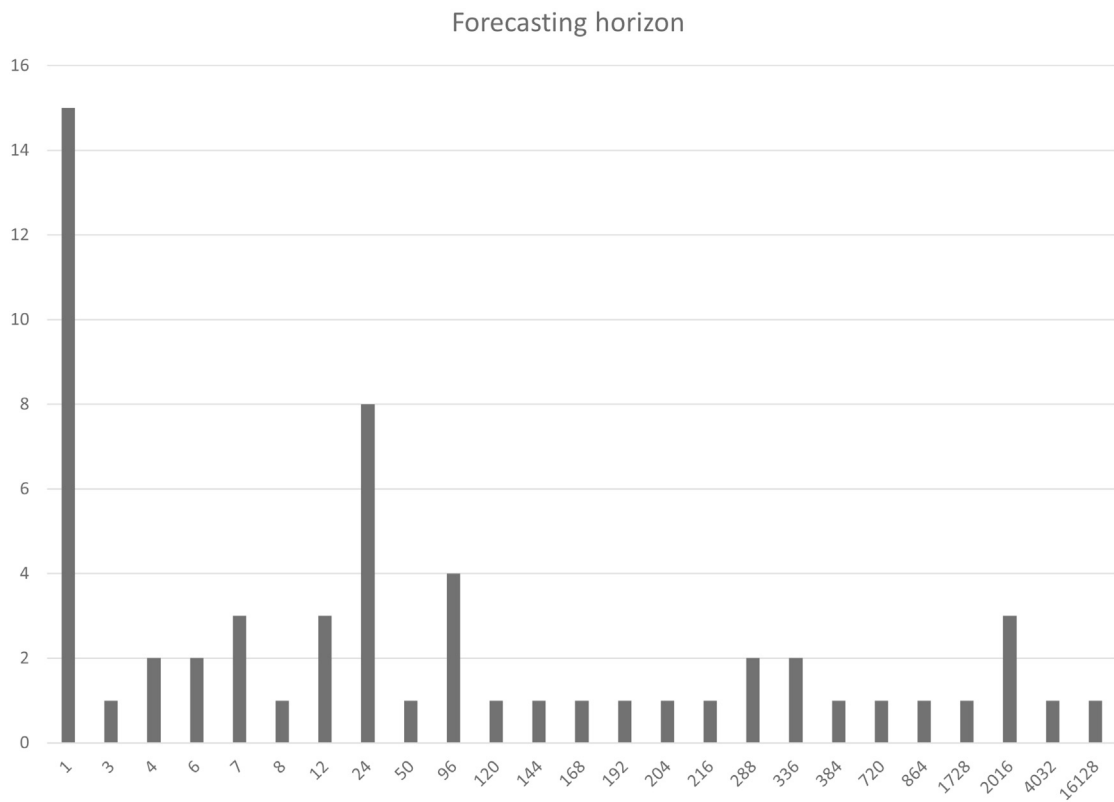**Fig. 19.** Output types of recent published papers.

Fig. 20. Forecasting horizons of recent published papers.

3.2. Models comparison

Unique algorithms that utilized deep learning model to forecast PV power generation will be compared in this section. Future PV power production can be determined through either direct or indirect forecasting methods. In direct forecasting, the model directly

Table 3
Summary of ANN Models and Parameters.

Ref.	Year	Model	No. Input Variables	Granularity	Lookback Length (steps)	Pre-Forecasting	Forecasting Horizon (steps)	Data Coverage	Source Country
Alaraj et al. [79]	2023	SSA-ANN	3	15 min	-	-	1	1 year	Saudi Arabia
Scott et al. [88]	2023	ANN	8	15-min, hourly, daily	7, 24, 96	-	7, 24, 96	10 months	UK
Brester et al. [78]	2023	ANN	8	Hourly	1, 2, 3, 6, 9	-	24	6 months	Finland

Table 4
Recurrent based models papers.

Ref.	Year	Model	No. Input Variables	Granularity	Lookback Length (steps)	Pre-Forecasting	Forecasting Horizon (steps)	Data Coverage	Source Country
Al-Jaafreh et al. [89]	2022	LSTM	16	Hourly	-	PCC	1	6 years	Switzerland
Singh et al. [80]	2022	LSTM, GRU	3	Hourly	-	-	1	6 months	India
Piotrowski and Farret [98]	2022	LSTM	1	Daily	450; 810; 1,150	-	50	-	Brazil
Agga et al. [107]	2022	LSTM, MLP	1	Hourly	288	-	24	1 year	Morocco
Admasie et al. [84]	2023	GWO-LSTM	3	Hourly	192	-	24	-	South Korea
Sadeghi et al. [92]	2023	GSA-LSTM	5	5 min	150; 280	SCC, PCC	12; 144; 204; 216; 864; 1,728; 2,016; 4,032; 16,128	1 month	Iran
Wang et al. [76]	2023	SVR-BO-LSTM	1	5, 30, 60 (min)	504; 1,008; 1,728; 6,048	EEMD	168; 288; 336; 2,016	-	Arizona
Phan et al. [104]	2023	GRU	6	Hourly	4	KPCA	1	-	Taiwan
Rubasinghe et al. [94]	2023	LSTM	5	15 min	1; 96	ICEEMDAN	96	4 years	Australia

predicts PV power generation as its output. On the other hand, PV power forecasting can also be done indirectly by predicting related factors such as irradiance, temperature, and so on. The models are analyzed based on their deep learning architecture, algorithms, optimization techniques, and input variables. To facilitate model comparison, we categorize them into four types of building units: ANN-unit (MLP models), RNN-unit (RNN, LSTM, GRU, and bidirectional models), CNN-unit, and Attention-unit (Transformer and Autoformer models).

3.2.1. Artificial neural network-based model

Table 3 displays the latest paper that employs a basic neural network for PV power generation forecasting. In the study by [79], solar PV power forecasting was performed using an ANN model, optimized with the Salp Swarm Algorithm (SSA), and trained with one year of historical PV output records from Qasim, Saudi Arabia [110]. The model was fed with three variables, comprising two meteorological variables and measured PV output data. SSA optimization played a pivotal role as a propagation technique, facilitating the update of system weights and the fine-tuning of hyperparameters. [88] conducted research to evaluate the effectiveness of a basic MLP model in predicting power generation from a 30 kWh roof-mounted PV system located in the UK. The study explored three distinct data granularities: 15-minute, hourly, and daily, with the aim of forecasting the subsequent 96, 24, and 7 time steps ahead, respectively. The model was provided with a total of eight variables, including time-related factors like the time of day. [78] presented a study focused on day-ahead PV power forecasting, examining two different approaches. One approach incorporated input data derived from measured meteorological variables, while the other relied on data from Numerical Weather Prediction (NWP) output. The study utilized a dataset comprising six months of hourly data from three buildings in Finland, with a combined capacity of 33.2 kWh. The model was supplied with eight variables and considered various lookback periods, including 1, 2, 3, 6, and 9 past hours for analysis. The latest models employing MLPs have shown a trend of utilizing diverse input variables (meteorological, time-related, and numerical weather prediction data), optimizing parameters through techniques like Salp Swarm Algorithm and Stochastic Gradient Descent, and forecasting PV power generation at multiple time scales ranging from 15 minutes to 1 week ahead.

3.2.2. Recurrent neural network-based model

RNN represents an evolved version of the basic neural network, specially designed for time-series learning, encompassing basic RNN, LSTM, and GRU architectures. Table 4 showcases the latest papers employing recurrent models for PV power output forecasting. In their work, [89] demonstrated short-term solar energy forecasting for PV output in Basel, Switzerland, using a vanilla LSTM unit. They investigated the impact of various meteorological variables on system accuracy and employed three groups of meteorological sets, each containing 6, 14, and 16 variables. The Pearson correlation coefficient was used to assess and rank the variables. Similarly, [80] explored vanilla GRU and LSTM units to predict power generation for campus rooftop PV arrays in Allahabad, India, with three meteorological variables as inputs. In contrast, [98] delved into the influence of historical data input quantity on PV power output forecasting accuracy. They conducted tests with three distinct lookback sizes: 450, 810, and 1150 past days, using a vanilla LSTM model. Additionally, [107] demonstrated short-term PV power prediction for a 15 kW grid-connected PV plant in Rabat, Morocco,

Table 5
Stacked and Two directional recurrent based models papers.

Ref.	Year	Model	No. Input Variables	Granularity	Lookback Length (steps)	Pre-Forecasting	Forecasting Horizon (steps)	Data Coverage	Source Country
Parida et al. [90]	2022	Stacked LSTM	2	Hourly	720	KELM	1	2 years	Alabama
Hong et al. [85]	2022	Stacked PSO-LSTM	1	Hourly	24	-	1	-	Taiwan
Garip et al. [81]	2023	stacked LSTM and BiLSTM	15	Daily	1,741	-	1	-	Turkey
Xiao et al. [95]	2023	BiLSTM	7	Hourly	24	-	6	-	China
Zhang et al. [103]	2023	BiLSTM	1	15 min	96	AP-CEEMDAN	1	-	Australia

using the LSTM model. Their input consisted of one year of historical data sequenced for the 12 previous days. The study also involved investigating and selecting the model's 52 hyperparameters using Gridsearch algorithms. In a different approach, [84] presented day-ahead PV power output forecasts based on an empirical formula derived from forecasted weather data. They employed a vanilla LSTM model with an input sequence spanning 192 previous hours to predict weather variables. Additionally, the study explored the use of the Grey Wolf Optimizer (GWO) to optimize the model's hyperparameters. Furthermore, [92] investigated PV power plant output forecasting across various time horizons using LSTM units in Iran. Their analysis considered different forecasting periods: 12 hours, 3 days, and 2 weeks. They employed two types of correlation coefficients for variable selection: Pearson Correlation Coefficients (PCC) to assess linear relationships and Spearman Correlation Coefficients (SCC) for rank correlation. The study also emphasized the presence of autocorrelation at various lag intervals and employed Gridsearch to fine-tune hyperparameters for optimizing model performance. [76] conducted an experiment involving an LSTM model for PV power prediction. The model received two sub-signal inputs derived from decomposed PV historical data collected at two sites in Arizona, USA. Ensemble Empirical Mode Decomposition (EEMD) was applied to reduce randomness and volatility from the data. The model's hyperparameters were selected using Bayesian optimization, and the output was filtered using the SVR model. Various tests were conducted with different forecasting horizons and input sequence lengths. Furthermore, [104] forecasted the next 1 hour of PV power generation using a model based on 3 GRU units tailored for different times of the day. They incorporated data from 10 PV sites in Taiwan, using four past PV generation records, solar irradiance data, and four NWP meteorological results as input. They also implemented a new error correction technique, involving the removal of samples with a standard deviation exceeding 1, and used Kernel Principal Component Analysis (KPCA) to reduce input variable dimensions. Lastly, [94] examined the improved complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) to decompose PV historical data for 15-minute and 1-day (96 samples) ahead of PV net load prediction. They employed vanilla LSTM units, which were fed with historical, weather, and decomposed IMF subsignals data.

3.2.2.1. Stacked recurrent units   Table 5 presents various stacked recurrent and dual-directional model architectures used for PV power generation forecasting. In their work, [90] demonstrated one-hour-ahead solar power forecasting using stacked LSTM units in the context of Alabama, US. The model incorporated two meteorological variables, and the input dataset was categorized into three groups: summer, rainy, and winter. Grey wolf optimization was employed to update the model's parameters, and a Karnel extreme learning machine was utilized to mitigate the weight's randomness, thereby reducing the impact of noise and outliers on the data. Additionally, [85] showcased one-hour-ahead solar Global Horizontal Irradiance (GHI) forecasting for PV power prediction, employing three cascaded LSTMs with a sequence input spanning the previous 24 hours. This research was conducted across five cities in Taiwan, and the input data was categorized into four seasons. The author used particle swarm optimization (PSO) to select the model's hyperparameters. Furthermore, [81] exhibited one-day-ahead PV power forecasting by comparing two models, stacked LSTM and BiLSTM, using data from 26 solar panels in Istanbul, Turkey. The model was fed with two historical datasets and 13 weather variables, with a sequence length covering the entire record. This study introduced two unique evaluation metrics: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

3.2.2.2. Bidirectional recurrent units   [95] presented a model for next-six-hour PV power prediction using the BiLSTM model, based on data from Gansu province, China. In this study, NeuralProphet was incorporated to provide additional variables to the models, including trend and seasonal decompositions of the original data. The model inputs included historical PV output, meteorological data, and time index variables, with input sequence lengths spanning 24 hours. Moreover, [103] explored the benefits of utilizing decomposed PV power output to enhance prediction accuracy. The study focused on ultra-short-term (15-minute) PV power forecasting horizons using BiLSTM units, which received an input sequence comprising the previous 96 values (equivalent to 1 day of data). The dataset was categorized into three different weather conditions: sunny, changeable, and cloudy days. To decompose the main signal into its constituent components, the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) was implemented, resulting in 12 intrinsic mode function (IMF) components.

Researchers explored different approaches to forecasting using recurrent models (vanilla, stacked, and bidirectional models). For vanilla models, authors utilized different input variables, lookback periods, and hyperparameter tuning techniques for PV power prediction using vanilla recurrent models (RNNs, LSTMs, and GRUs). Authors investigated the impact of varying numbers of mete-

orological variables, ranging from 3 to 16. They also examined diverse lookback periods, from as short as 12 previous days to as long as 1150 past days. Hyperparameter tuning techniques employed included grid search, GWO, and Bayesian optimization. Additionally, some studies incorporated techniques such as error correlation analysis and input data decomposition to enhance model performance. For stacked and bidirectional LSTM architectures, the models were used for short-term (one-hour) and day-ahead PV power forecasting, incorporating meteorological and historical data. They employed optimization techniques like GWO and PSO for parameter tuning and introduced new evaluation metrics like AIC and BIC. The studies highlight the use of bidirectional architectures along with data preprocessing techniques like decomposition and additional input features to improve short-term PV forecasting performance.

3.2.3. Convolution neural network-based model

Table 6 presents the most recent papers demonstrating the prediction of power generation from PV panels, which incorporate CNN layers in their architecture, either alone or combined with other models. [86] conducted experiments involving an input feature extractor employing two layers of CNN for multistep short-term PV power forecasting. The study utilized four stacked LSTM layers and was based on data collected from Abu Dhabi, UAE. The investigation focused on two types of input variables: Global horizontal irradiance (GHI) and plane of array (POA) irradiance. Similarly, [91] conducted an experiment employing a CNN-LSTM model to forecast multiple upcoming days of PV power generation in Rabat, Morocco. The study incorporated seven variables, encompassing historical power generation, weather data, and energy consumption information, as inputs to the model. The approach involved utilizing three 1D CNN layers to extract spatial features and generate a novel time series representation for the sequence input. Subsequently, the transformed data was fed into three LSTM layers and a fully connected layer to predict the subsequent power generation. The hybrid model demonstrated the capability to offer more accurate predictions for both general and specific locations when compared to stand-alone models. Similarly, [96] delved into the CNN-LSTM model for predicting targeted domestic household PV systems by harnessing historical solar generation data from households. The study employed historical records of PV power generation at both regional (aggregated) and individual levels in Sydney, Australia, as inputs for the system. Diverse granularities, encompassing half-hourly, daily, and monthly intervals, were explored to enhance the capacity for predicting individual household solar power production using a broader dataset. [77] employed a comparable approach for predicting the next-step PV power output using the CNN-BiGRU model. The study involved the utilization of decomposed historical PV production achieved through variational mode decomposition, in conjunction with four meteorological data variables in Korea. The system incorporated an input look-back length of 20 time steps. To enhance the predictive capacity, CNN layers were employed to establish a novel matrix representing the extracted input-output relationship. This matrix was then fed into the BiGRU model for subsequent power output prediction. [82] demonstrated PV power prediction for the next 96 steps using a combined input of decomposed historical data and predicted numerical weather inputs. The approach involved utilizing a CNN-LSTM model with decomposed historical power generation data as the input. The measured data was decomposed using a discrete wavelet transform. Simultaneously, numerical weather input served as the input for the MLP model, which was subsequently combined with the CNN-LSTM model's output. Both models were fed sequences of 96 elements. [102] investigated PV power forecasting using measured data from 1.5 MWp floating PV power plants across 111 universities in Thailand. The study employed two models: Bi-LSTM for one-week ahead forecasting and CNN-BiGRU for day-ahead forecasting, with the input sequence length set to the 74 previous time steps. Three months' worth of historical power and meteorological data were incorporated into the forecasting model, including the grouping of regular and cloudy weather conditions. [93] computed multi-step, long-term PV power output using recurrent (LSTM and GRU) and CNN forecasting models based on forecasted global solar irradiance and temperature data in the region of Zahedan City, Iran. The models were provided with four input variables from monthly data, spanning the previous 24 months. The forecasted results were then compared with the actual performance of a 20 MW PV plant. In contrast, [100] undertook the prediction of the following day's PV power generation for a specific location, employing the BiLSTM-CNN model with generalized input variables. The study utilized historical PV and meteorological data with a granularity of 1-hour from Yunnan, China. To enhance spatial recognition in areas with limited historical data, PV plants were grouped into subregions using the k-means algorithm, and one robust CNN-BiLSTM model in the region was prepared. In conclusion, this section showcase forecasting models incorporated CNN layers, either alone or combined with other models like LSTMs, BiL-STMs, and GRUs. It was demonstrated to extract spatial and temporal features from diverse input data sources (PV historical record, meteorological variables, NWE, and decomposed signal components). CNN-based hybrid models demonstrate enhanced forecasting capabilities across various time horizons, ranging from multi-step short-term to long-term predictions spanning multiple days or months ahead. Moreover, the studies highlight the importance of data preprocessing techniques, such as decomposition methods, and the strategic utilization of input variables tailored to specific locations or granularities.

3.2.4. Transformer-based model

Table 7 presents published articles on the cutting-edge deep learning transformer model used for time series PV power generation forecasting.

[56] conducted an experiment to test the benefits of decomposing historical energy data to forecast the next time step using a vanilla transformer model. This was done to enhance the model's robustness and enable it to handle high noise, disturbances, and signal dips. The experiment utilized a dataset from 5 houses in London, UK, decomposed into 3 levels of sub-signals using the Stationary Wavelet Transform (SWT) and sequenced into 12 past readings input for the transformer model. Nevertheless, the proposed model still needs improvement for multi-step forecasting. [108] performed large-scale forecasting of time-series energy consumption using a vanilla Transformer model. The model predicted the power consumption of Birmingham Main Library for the next time step using only historical consumption records sequenced with an input length of 4 past readings. The author incorporated

Table 6
Ensemble with CNN models papers.

Ref.	Year	Model	No. Input Variables	Granularity	Lookback Length (steps)	Pre-Forecasting	Forecasting Horizon (steps)	Data Coverage	Source Country
Li et al. [82]	2022	CNN-LSTM-ANN	7	15 min	96	DWT	96	-	China
Michael et al. [86]	2022	CNN-Stacked LSTM	2	Hourly	-	-	1	-	UAE
Agga et al. [91]	2022	CNN-LSTM	7	Daily	14	-	1; 3; 7	-	Morocco
de C. Costa [96]	2022	CNN, LSTM, CNN-LSTM	2	30 min, daily, monthly	-	-	1	-	Australia
Zhang et al. [77]	2022	BiGRU, CNN	5	15 min	20	VMD	1	-	South Korea
Li et al. [100]	2023	CNN-BiLSTM, BiLSTM-CNN	5	Hourly	11	k-means clusting	24	-	China
Azizi et al. [93]	2023	MLP, LSTM, GRU, CNN, CNN-LSTM	4	Monthly	24	-	120	37 Y	Iran
Khortsriwong et al. [102]	2023	CNN-BiGRU, BiLSTM	4	5 min	21,312	-	288; 2,016	3 M	Thailand

Table 7
Transformer based models papers.

Ref.	Year	Model	No. Input Variables	Granularity	Lookback Length (steps)	Pre-Forecasting	Forecasting Horizon (steps)	Data Coverage	Source Country
Saoud et al. [56]	2023	Transformer	1	5 min	12	SWT	1	-	UK
Tian et al. [83]	2022	Transformer	13	Hourly	168	-	1	-	-
Huang and Kaewunruen [108]	2023	Transformer	1	30 min	4	-	4	-	UK
Jeong [99]	2023	Transformer	1	15 min	48; 96; 144	-	4, 8, 12, 24	-	South Korea
Kothona et al. [101]	2022	Transformer	5	Hourly	72	-	24	-	Greece
Trong et al. [97]	2023	CNN-Transformer	4	10 min	26,298	VMD	6	-	Vietnam
Ziyabari et al. [87]	2022	Transformer-CNN Time embedded	17	Hourly	24	-	12	-	Philadelphia
Sherozbek et al. [106]	2023	Vision Transformer-CNN	6	Hourly	1	-	1	-	South Korea
Demir et al. [105]	2022	Transformer	1	Hourly	48		384	10 Y	Texas
Pospíchal et al. [75]	2022	Vision Transformer (AdamW)	12	Daily	7	-	7	-	India
Santos et al. [74]	2022	Temporal Fusion Transformer	12	Hourly	72	-	24	-	Germany and Australia

skip layers by [111] to enhance computational efficiency by bypassing minor errors in deep layers that do not affect the learning process.

In the area of PV power forecasting, [83] investigated the vanilla transformer for ultra-short-term (1 step ahead) PV power forecasting. The model used up to 12 historical time series variables categorized into time indices, meteorological data, and PV data, collected from a real household microgrid dataset. To evaluate the model's stability, the dataset was divided into three subsets: Summer, Autumn, and Spring. The robustness of the trained model was validated by applying it to a Public DC competition dataset. However, their proposed transformer setup required 571,409 parameters to perform the task. Similarly, [105] forecasted the next 384 time steps of irradiance using vanilla transformers for PV power generation in Lubbock, Texas. Both supervised and unsupervised machine learning approaches were employed, using 48 past readings as input. For unsupervised forecasting, the rolling concept proposed by [112] in 2018 was implemented. The author found that, for unsupervised machine learning, the transformer is only capable of short-term forecasting (up to 48 steps ahead). In different approach, [99] studied how the forecasting horizon length and input sequence length affect the Transformer model's ability to predict PV power generation. The study also demonstrated that it's possible to make predictions using only a historical PV power output dataset from neighboring sites. For the experiment, data was collected from 50 neighboring sites in Suncheon, South Korea. The Transformer proved effective for long forecasting horizons. However, it's worth noting that a long forecasting horizon requires a similarly long or longer input sequence length. On the other hand, for short-term forecasting, the Transformer provides similar results to recurrent-based models but demands more computational resources. Moreover [101] investigated different meteorological variables to forecast the next 24 time steps using a transformer model. The historical dataset, comprising PV power output, irradiation, and cloudiness index data, was collected from 5 PV systems in Kozani, Greece, and aggregated. Additionally, forecasted irradiance and cloud coverage datasets were imported to support the model's predictions. The model inputs were sequences of 72 and 24 elements for historical and forecasted data,

respectively. Although adding cloud coverage data significantly improved accuracy, solar irradiance data did not reduce the error. Furthermore, [97] experimented with Variational Mode Decomposition, as proposed by [113], to simplify a historical dataset and remove nonlinearity and nonstationarity from PV power output. This approach aimed to forecast the next 6 time steps ahead using a Transformer model. To enhance model performance, Trong replaced the input positional encoding in the original architecture with two 1-dimensional convolutional layers, allowing for the extraction of spatial information. Trong utilized the sub-signals from VMD and PV system parameters (current, voltage, and frequency) as input data. However, replacing positional encoding with convolutional layers increased the number of computational parameters.

In contrast, [87] conducted an investigation involving a CNN-transformer to predict solar irradiance over the next 12 hours, utilizing 17 variables. In this study, the original transformer model's Feedforward Neural Network (FFNN) layers were replaced with a 1D-CNN network to enhance the model's ability to capture local data trends. To facilitate this process, a sliding mechanism incorporating a 24-element input sequence was employed. Notably, achieving favorable results required implementing over one million parameters. Similarly, [106] adopted a similar approach to forecast PV power generation, integrating a 1D-CNN layer in place of the fully connected feed-forward network. The objective was to predict PV power production for the ensuing 1 hour in Buang-gun, Republic of Korea. The system used a retrospective input consisting of historical PV data, solar radiance, insolation, and temperature variables. Before entering the multihead attention blocks, the input was normalized. Furthermore, [75] proposed a 1-step ahead spatiotemporal solar irradiance forecasting using a variant of the transformer model. The Vision Transformer architecture [114] was implemented, with the normalization layer placed before the attention layer in the encoder-decoder blocks. The model was fed with historical irradiance and weather data from the NASA Power Bot application, time variables encoded using sine-cosine encoding, and regional coordination vector encoding, with a sequence length of 7. To address the challenges posed by long input sequences in transformer models for time series problems, [115] proposed a solution that tackles two specific issues: improving input representation scores and selecting relevant time steps. To enhance the input representation scores, a local convolution block was introduced within the attention mechanism to further improve the representation matrix scores. To address the memory challenge associated with long input sequences, the LogSparse attention mechanism was introduced to dynamically select a subset of time steps that are most relevant for the prediction task. This mechanism employs a logarithmic sparsity pattern, where attention scores decrease logarithmically as the temporal distance between elements in the current time step increases.

In different approach, [116] introduced the Temporal Fusion Transformer (TFT), a variant of the Transformer model designed for multi-horizon forecasting. This innovative approach employs multiple networks and transforms static insights into time series inputs. The process begins with input variable selection networks, which identify variables that significantly impact the model's accuracy. To incorporate static features within the time series dataset, LSTM units are integrated into the encoder-decoder layers, enabling their computation alongside other input variables. Gate Residual Networks are strategically deployed to disregard parameters with minimal impact on the model's outcomes. Finally, Quantile forecasting is employed to determine the target range for multi-horizon forecasting. [74] conducted an investigation using the Temporal Fusion Transformer to forecast the PV power output for the upcoming 24 hours across six PV facilities situated in Germany and Australia. The study reveals that the key inputs encompass three meteorological datasets: solar horizontal irradiation, temperature, and humidity; zenith and azimuth angles; along with sine and cosine representations of the monthly cycles. However, it is noteworthy that the model entails 361,000 parameters to accomplish the specified tasks. Summarizing published Attention Mechanisms and Transformer architecture, researchers have shown a promising approach for PV power forecasting. While vanilla Transformer models have demonstrated their capability for next-step and short-term forecasting, researchers have proposed various modifications and hybrid architectures to enhance their performance and address inherent limitations. These improvements include incorporating convolutional layers for spatial feature extraction, employing data decomposition techniques to simplify input representations, and introducing mechanisms like LogSparse attention and local convolutions to handle long input sequences effectively. Furthermore, innovative variants like the Temporal Fusion Transformer have been developed to tackle multi-horizon forecasting challenges by fusing static and temporal data through variable selection networks and dedicated components. Despite the computational demands of these models, their ability to capture intricate temporal dependencies and leverage auxiliary data sources, such as meteorological variables and NWE.

4. Discussion

In delving into the discussion of trends, challenges, and future directions in deep learning-based time series forecasting for PV systems, it becomes evident that the field is witnessing a surge of innovation and exploration. The remarkable success of the transformer model in NLP has spurred researchers to adapt it for time series applications, resulting in various variants of the original architecture. In relation to timeseries forecasting, [117] aimed to predict future time series data related to influenza-like illnesses using the transformer model. Leveraging the attention mechanism inherent in transformers, the model effectively captured temporal data patterns, achieving state-of-the-art performance in the field. However, the model faces limitations when dealing with long input sequences. Transformers usually operate with fixed-size input windows, and as the input sequence length grows, so do the computational and memory requirements. The dot-product nature of the attention mechanism demands significant computational resources to build a robust model. A notable example of this resource-intensive nature is the Generative Pretraining Transformer (GPT) by [118], which required 1.5 billion parameters to achieve state-of-the-art results. Lastly, to confront the challenges faced by the transformer model in long-term series forecasting, such as handling complex temporal patterns and improving computation efficiency in data utilization, [109] introduced a novel approach called Autoformer. The Autoformer aims to enhance the natural periodicity of time series data and establish connections between input matrix representations in a series-wise manner rather than a point-wise approach. Several techniques were incorporated into the original transformer model to achieve these goals. Firstly, an Auto-Correlation mechanism

was introduced to capture period-based dependencies. This mechanism focuses on the similarity between sub-series by aggregating the auto-correlation values across different time delays. By emphasizing period-based dependencies, the Autoformer model can effectively capture the relationships and patterns that occur at specific intervals within the time series. Secondly, a series decomposition block was implemented to aggregate the input long-term trend from medium prediction. In the encoder layers, the trend component was eliminated, leaving only the seasonal part intact. This decomposition helps isolate the repeating patterns and inherent periodicity of the time series data, facilitating their integration into the model's architecture. In the encoder, only the seasonal part of the input signal is retained, while the trend component is discarded. On the other hand, the decoder incorporates the trend-cyclical and seasonal parts of the original input signal, rather than relying solely on the shifted target sequence as in traditional transformer models.

One notable trend is the proliferation of hybrid architectures that leverage the strengths of various deep learning models to enhance forecasting accuracy and robustness. Researchers are increasingly integrating recurrent hybrid with models, such as LSTM, GRU [80,107], BiLSTM-NP [95] and CNN [82,86,91,96,77,100,93,102], to capture temporal dependencies and spatial features effectively. Moreover, the emergence of novel architectures like CNN-transformers signals a future trend [97,87,106], with studies indicating their potential to revolutionize PV forecasting by accommodating long input sequences and improving computational efficiency.

Additionally, researchers have investigated the utilization of new input variables, such as meteorological conditions [79,90, 85,103], cloud coverage data [101], historical data input quantities and spatiotemporal properties [98]. Further, techniques like numerical weather prediction [78,104] and empirical formulas derived from forecasted weather data [84] have been explored to enhance the forecasting accuracy.

Input decomposition methods have also gained significant attention, with researchers employing techniques like EMD, CEEMDAN, and ICEEMDAN [76,103,94], wavelet transforms, and stationary wavelet transform (SWT) [82,56], and variational mode decomposition (VMD) [77,97] to extract relevant features and handle non-stationarity in the data.

Furthermore, optimization techniques for model hyperparameter tuning, such as Grid Search Algorithms [107,92], Grey Wolf Optimizer (GWO) [84,90], Bayesian Optimization [76], and Particle Swarm Optimization (PSO) [85], have been widely explored to improve model performance. These trends underscore the evolving landscape of deep learning methodologies, paving the way for more accurate and comprehensive PV forecasting models.

Table 8 shows the error metrics employed by the reviewed papers. The most commonly used error metrics were the RMSE and MAE. Most studies utilized their own data and reported error metrics based on the statistical details of their data (maximum, minimum, variation, and standard deviation). These metrics were favored due to their straightforward representation of the error magnitude. However, for cross-comparison with other PV forecasting models, it is challenging to make meaningful comparisons without additional information such as the site's rated capacity or the data range. For universal benchmarking purposes, nearly 30% of the papers presented the MAPE, R^2, and a few utilized the NRMSE, which allows readers to evaluate the overall accuracy strength instantly. Finally, very few studies employed pre-trained scaling or normalization techniques on their datasets, as this could potentially compromise the location-specific characteristics of the data.

Despite the advancements in deep learning for PV power forecasting, several challenges persist. One of the most significant challenges highlighted in the reviewed studies is the limited availability or insufficient amount of high-quality, comprehensive data required to effectively train deep learning models [79,88,78,98,85,75]. Additionally, the computational demands and resource requirements for training and deploying complex deep learning models, particularly when utilizing the transformers model, have been identified as a substantial limitation [83,97,87,74]. Handling non-stationary input signals, detecting and removing outliers [74], and data preprocessing steps such as decomposition techniques to transform non-stationary data to stationary form and reconstruct them back [82], have also been recognized as challenges.

In terms of future research directions, a huge potential is with the exploration of hybrid and ensemble model architectures that combine different deep learning units or techniques to leverage their respective advantages [100,102,108]. Additionally, researchers have emphasized the need for robust hyperparameter tuning optimization techniques to find the best set of hyperparameter values and architectures for specific forecasting tasks [76,85,86,105,106]. The development of universal or generalized models that can be applied to diverse scenarios and locations, by utilizing public datasets such as NASA POWER, has been pushed forward as a future research direction [75,96,100,108]. Furthermore, incorporating additional input variables, such as aerosol impact, variability indices, cloud data, and longer historical data sequences, have been suggested as potential avenues for improving forecasting accuracy [79,97,101].

5. Conclusion

The reliability of PV integrated systems is significantly influenced by the inherent variability of solar irradiance. Fluctuations in solar radiation levels lead to intermittent power generation, resulting in voltage and power fluctuations, posing challenges for utility companies, energy markets, and power distribution networks. Reliable forecasting techniques, spanning a wide range of forecast horizons from ultra-short-term to long-term predictions, are pivotal in addressing these challenges. Current research and industry practices heavily rely on data-driven models for short-term, day-ahead forecasting due to their higher accuracy and ability to capture cloud dynamics effectively. Conversely, long-term forecasts have gained increasing importance for strategic planning and resource allocation in power systems.

Achieving highly accurate predictions remains a formidable challenge; consequently, various methodologies encompassing physical, statistical, and artificial intelligence approaches have been extensively investigated. Among these methods, deep learning neural

Table 8
Errors table.

Ref.	Peak Power	RMSE	MAE	MAPE	R^2	MSE	NRMSE
Brester et al. [78]	20 kW	436.507 W	172.032 W	-	-	-	2.15%
Alaraj et al. [79]	6.3 kW	1.45%	-	18%	-	2.12%	-
Scott et al. [88]	30 kWh	1.76 kW	-	-	-	-	-
Singh et al. [80]	1 MW	1.34 kW	1.0163 kW	-	-	2.1816 kW	-
Agga et al. [107]	15 kW	6.76	4.5	23.667%	-	-	-
Wang et al. [76]	100 MW	4.12 MW	2.054 MW	-	-	-	-
Rubasinghe et al. [94]	3500 MW	-	42.85 MW	-	-	-	-
Al-Jaafreh et al. [89]	Scaled	0.035	-	-	-	0.0013	-
Piotrowski and Farret [98]	550.68 kWh	151.89	-	-	-	-	-
Admasie et al. [84]	900 W/m^2	0.0298	-	-	-	-	-
Sadeghi et al. [92]	10 MW	-	-	-	97.95%	0.15	-
Phan et al. [104]	907.68 kW	-	-	-	-	-	8.11%
Parida et al. [90]	Scaled	0.012	0.01	1.57%	-	-	-
Xiao et al. [95]	2.4 kW	0.29 kW	0.16 kW	3.09%	-	-	-
Zhang et al. [103]	6 MW	0.055 error/MW	0.029 error/MW	2.77%	-	-	-
Hong et al. [85]	1957.7 W/m^2	0.04035	0.2871	-	95.79%	-	-
Garip et al. [81]	157.1 kWh	19.17 kW	13.15 kW	-	96%	-	-
Michael et al. [86]	8.12 kWh/m^2	0.36	0.18	3.11%	98%	0.13	11%
Agga et al. [91]	46 kWh	6.65 kWh	4.97 kWh	19.85%	-	-	-
Zhang et al. [77]	20 MW	0.63 MW	0.49 MW	5.77%	98.8%	-	7.47%
Khortsriwong et al. [102]	1.5 MW	67.53 kW	28.34 kW	8.32%	-	-	-
Li et al. [100]	18 MW	0.0409	0.1019	-	-	0.0018	-
Azizi et al. [93]	-	12.87 w/m^2	10.42 w/m^2	-	95.77%	-	5.23%
Li et al. [82]	700 kW	69.07 kW	39.05 kW	-	-	-	-
de C. Costa [96]	3 kWh	540.8 W	426 W	-	-	-	19%
Saoud et al. [56]	300 Wh	4 W	1.8 W	4.88%	-	-	-
Trong et al. [97]	42 kW	0.43 kW	0.36 kW	-	-	-	-
Ziyabari et al. [87]	-	0.06 $W/m2$	0.04 W/m2	-	87%	-	-
Sherozbek et al. [106]	3.5 kWh	210 Wh	170 Wh	-	-	40 Wh	-
Santos et al. [74]	6 kW	0.064	0.033	-	99.8%	-	-
Tian et al. [83]	-	-	0.092	-	0.879	0.036	-
Pospíchal et al. [75]	8 $kWh/m^2/day$	-	131.5 W	3.45%	-	-	-
Jeong [99]	3	-	0.1722	-	-	0.0971	-
Kothona et al. [101]	60 kWh	-	1388.7 Wh	-	92%	-	-
Huang and Kaewunruen [108]	27000 kWh	76.9611 kWh	-	-	82.38%	-	-
Demir et al. [105]	1,124.90 W/m^2	84 W/m^2	-	-	-	-	-

networks, particularly transformer architectures or their hybrid variants employed in ensemble setups with intricate hyperparameter tuning, have demonstrated the most promising potential in terms of predictive accuracy. The optimization technique plays a crucial role in the model's learning process, making it an increasingly attractive line of investigation to explore for developing more robust, accurate, and computationally efficient forecasting models. In addition, unifying accuracy measurement and presentation will increase the cross-comparison between available models to accelerate the development of accurate models.

Looking ahead, a promising avenue for future research lies in exploring hybrid and ensemble architectures that combine various deep learning modules or techniques to capitalize on their respective strengths synergistically. Moreover, researchers have underscored the importance of robust hyperparameter optimization strategies to identify the optimal set of hyperparameter values and architectures tailored to specific forecasting tasks. The development of universal or generalized models, facilitated by leveraging publicly available datasets like NASA POWER, has garnered attention as a potential future research direction, enabling the application of these models across diverse scenarios and locations. Additionally, the creation of a shared database dedicated to PV farm data from various locations will accelerate the development of PV power forecasting. Furthermore, incorporating additional input variables, such as aerosol effects, variability indices, cloud data, and extended historical data sequences, has been proposed as a potential pathway to further enhance forecasting accuracy.

CRediT authorship contribution statement

Husein Mauladdawilah: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis. **E.J. Gago:** Validation, Supervision, Project administration. **Hasan Balfaqih:** Writing – review & editing, Validation, Supervision, Methodology. **M.C. Pegalajar:** Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors declare that there are no data deposited in a public repository, as no data were utilized for the research described in the article.

References

[1] M. Mišík, The EU needs to improve its external energy security, Energy Policy 165 (2022) 112930, https://doi.org/10.1016/j.enpol.2022.112930.

[2] J. Heng, J. Wang, L. Xiao, H. Lu, Research and application of a combined model based on frequent pattern growth algorithm and multi-objective optimization for solar radiation forecasting, Appl. Energy 208 (2017) 845–866, https://doi.org/10.1016/j.apenergy.2017.09.063.

[3] G. Masson, M. de l'Epine, I. Kaizuka, Trends in PV applications 2023, Tech. Rep., IEA PVPS TCP, 2023.

[4] G. Masson, E. Bosch, A.V. Rechem, M. de l'Epine, Snapshot of global PV markets 2023, Tech. Rep., IEA PVPS TCP, 2023.

[5] Solar Energy Perspectives, OECD, https://doi.org/10.1787/9789264124585-en, 2011.

[6] B. Chen, P. Lin, Y. Lai, S. Cheng, Z. Chen, L. Wu, Very-short-term power prediction for PV power plants using a simple and effective RCC-LSTM model based on short term multivariate historical datasets, Electronics 9 (2) (2020) 289, https://doi.org/10.3390/electronics9020289.

[7] T. Huld, G. Friesen, A. Skoczek, R.P. Kenny, T. Sample, M. Field, et al., A power-rating model for crystalline silicon PV modules, Sol. Energy Mater. Sol. Cells 95 (12) (2011) 3359–3369, https://doi.org/10.1016/j.solmat.2011.07.026.

[8] T. Muneer, S. Etxebarria, E. Gago, Monthly averaged-hourly solar diffuse radiation model for the UK, Build. Serv. Eng. Res. Technol. 35 (6) (2014) 573–584, https://doi.org/10.1177/0143624414522639.

[9] Z. Zhen, J. Liu, Z. Zhang, F. Wang, H. Chai, Y. Yu, et al., Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image, IEEE Trans. Ind. Appl. 56 (4) (2020) 3385–3396, https://doi.org/10.1109/tia.2020.2984617.

[10] L. Visser, T. AlSkaif, W. van Sark, Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution, Renew. Energy 183 (2022) 267–282, https://doi.org/10.1016/j.renene.2021.10.102.

[11] R. Ahmed, V. Sreeram, Y. Mishra, M. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization, Renew. Sustain. Energy Rev. 124 (2020) 109792, https://doi.org/10.1016/j.rser.2020.109792.

[12] S. Dutta, Y. Li, A. Venkataraman, L.M. Costa, T. Jiang, R. Plana, et al., Load and renewable energy forecasting for a microgrid using persistence technique, Energy Proc. 143 (2017) 617–622, https://doi.org/10.1016/j.egypro.2017.12.736.

[13] A. Nespoli, E. Ogliari, S. Leva, A.M. Pavan, A. Mellit, V. Lughi, et al., Day-ahead photovoltaic forecasting: a comparison of the most effective techniques, Energies 12 (9) (2019) 1621, https://doi.org/10.3390/en12091621.

[14] M.Q. Raza, M. Nadarajah, C. Ekanayake, On recent advances in PV output power forecast, Sol. Energy 136 (2016) 125–144, https://doi.org/10.1016/j.solener.2016.06.073.

[15] E. Kim, M.S. Akhtar, O.B. Yang, Designing solar power generation output forecasting methods using time series algorithms, Electr. Power Syst. Res. 216 (2023) 109073, https://doi.org/10.1016/j.epsr.2022.109073.

[16] J. Zhang, Z. Liu, T. Chen, Interval prediction of ultra-short-term photovoltaic power based on a hybrid model, Electr. Power Syst. Res. 216 (2023) 109035, https://doi.org/10.1016/j.epsr.2022.109035.

[17] M. Bouzerdoum, A. Mellit, A.M. Pavan, A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant, Sol. Energy 98 (2013) 226–235, https://doi.org/10.1016/j.solener.2013.10.002.

[18] F. Wang, Z. Zhen, B. Wang, Z. Mi, Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting, Appl. Sci. 8 (1) (2017) 28, https://doi.org/10.3390/app8010028.

[19] D.P. Mishra, S. Jena, R. Senapati, A. Panigrahi, S.R. Salkuti, Global solar radiation forecast using an ensemble learning approach, Int. J. Power Electr. Drive Syst. 14 (1) (2023) 496, https://doi.org/10.11591/ijpeds.v14.i1.pp496-505.

[20] Z. Wang, L. Jia, Short-term photovoltaic power generation prediction based on LightGBM-LSTM model, in: 2020 5th International Conference on Power and Renewable Energy (ICPRE), IEEE, 2020.

[21] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, M. Shi, A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework, Energy Convers. Manag. 212 (2020) 112766, https://doi.org/10.1016/j.enconman.2020.112766.

[22] Y. Jung, J. Jung, B. Kim, S. Han, Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: case study of South Korea, J. Clean. Prod. 250 (2020) 119476, https://doi.org/10.1016/j.jclepro.2019.119476.

[23] A. Mellit, A.M. Pavan, V. Lughi, Deep learning neural networks for short-term photovoltaic power forecasting, Renew. Energy 172 (2021) 276–288, https://doi.org/10.1016/j.renene.2021.02.166.

[24] K.R. Chowdhary, Natural language processing, in: Fundamentals of Artificial Intelligence, Springer, India, 2020, pp. 603–649.

[25] M. van Heel, G. Harauz, E.V. Orlova, R. Schmidt, M. Schatz, A new generation of the IMAGIC image processing system, J. Struct. Biol. 116 (1) (1996) 17–24, https://doi.org/10.1006/jsbi.1996.0004.

[26] T. chung Fu, A review on time series data mining, Eng. Appl. Artif. Intell. 24 (1) (2011) 164–181, https://doi.org/10.1016/j.engappai.2010.09.007.

[27] H. Balfaqih, Artificial Intelligence in Logistics and Supply Chain Management: A Perspective on Research Trends and Challenges, Springer International Publishing, ISBN 9783031089541, 2022, pp. 1241–1247.

[28] J. Yan, L. Hu, Z. Zhen, F. Wang, G. Qiu, Y. Li, et al., Frequency-domain decomposition and deep learning based solar pv power ultra-short-term forecasting model, IEEE Trans. Ind. Appl. 57 (4) (2021) 3282–3295, https://doi.org/10.1109/tia.2021.3073652.

[29] L.A. Fernandez-Jimenez, A. Muñoz-Jimenez, A. Falces, M. Mendoza-Villena, E. Garcia-Garrido, P.M. Lara-Santillan, et al., Short-term power forecasting system for photovoltaic plants, Renew. Energy 44 (2012) 311–317, https://doi.org/10.1016/j.renene.2012.01.108.

[30] G. Reikard, Predicting solar radiation at high resolutions: a comparison of time series forecasts, Sol. Energy 83 (3) (2009) 342–349, https://doi.org/10.1016/j.solener.2008.08.007.

[31] F. Rosenblatt, The Perceptron, a Perceiving and Recognizing Automaton Project Para, Cornell Aeronautical Laboratory, 1957.

[32] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115–133, https://doi.org/10.1007/bf02459570.

[33] A.G. Ivakhnenko, V.G. Lapa, Cybernetic Predicting Devices, Joint Publications Research Service, 1966, available from the Clearinghouse for . . . .

[34] A. Mellit, A.M. Pavan, 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 84 (5) (2010) 807–821, https://doi.org/10.1016/j.solener.2010.02.006, https://ui.adsabs.harvard.edu/abs/2010SoEn...84..807M.

[35] S. Sobri, S. Koohi-Kamali, N.A. Rahim, Solar photovoltaic generation forecasting methods: a review, Energy Convers. Manag. 156 (2018) 459–497, https://doi.org/10.1016/j.enconman.2017.11.019.

[36] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536, https://doi.org/10.1038/323533a0.

[37] J.L. Elman, Finding structure in time, Cogn. Sci. 14 (2) (1990) 179–211, https://doi.org/10.1207/s15516709cog1402_1.

[38] W. Ouyang, K.M. Yu, N. Sodsong, K.H. Chuang, Short-term solar PV forecasting based on recurrent neural network and clustering, in: R. Su (Ed.), International Conference on Image and Video Processing, and Artificial Intelligence, SPIE, 2019.

[39] H.K. Ahn, N. Park, Deep RNN-based photovoltaic power short-term forecast using power IoT sensors, Energies 14 (2) (2021) 436, https://doi.org/10.3390/en14020436.

[40] A.A.H. Lateko, H.T. Yang, C.M. Huang, H. Aprillia, C.Y. Hsu, J.L. Zhong, et al., Stacking ensemble method with the RNN meta-learner for short-term PV power forecasting, Energies 14 (16) (2021) 4733, https://doi.org/10.3390/en14164733.

[41] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[42] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, Adv. Neural Inf. Process. Syst. 27 (2014).

[43] X. Qing, Y. Niu, Hourly day-ahead solar irradiance prediction using weather forecasts by lstm, Energy 148 (2018) 461–468, https://doi.org/10.1016/j.energy.2018.01.177, https://www.sciencedirect.com/science/article/pii/S0360544218302056.

[44] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, et al., Learning phrase representations using RNN encoder-decoder for statistical machine translation 2014, https://doi.org/10.48550/ARXIV.1406.1078, arXiv:1406.1078.

[45] M. Aslam, J.M. Lee, H.S. Kim, S.J. Lee, S. Hong, Deep learning models for long-term solar radiation forecasting considering microgrid installation: a comparative study, Energies 13 (1) (2019) 147, https://doi.org/10.3390/en13010147.

[46] A.N. Buturache, S. Stancu, Solar energy production forecast using standard recurrent neural networks, long short-term memory, and gated recurrent unit, Eng. Econ. 32 (4) (2021) 313–324, https://doi.org/10.5755/j01.ee.32.4.28459.

[47] M. Schuster, K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681, https://doi.org/10.1109/78.650093.

[48] H. Zhen, D. Niu, K. Wang, Y. Shi, Z. Ji, X. Xu, Photovoltaic power forecasting based on ga improved bi-lstm in microgrid without meteorological information, Energy 231 (2021) 120908, https://doi.org/10.1016/j.energy.2021.120908.

[49] S. Boubaker, M. Benghanem, A. Mellit, A. Lefza, O. Kahouli, L. Kolsi, Deep neural networks for predicting solar radiation at hail region, Saudi Arabia, IEEE Access 9 (2021) 36719–36729, https://doi.org/10.1109/access.2021.3062205.

[50] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, Handb. Brain Theory Neural Netw. 3361 (10) (1995) 1995.

[51] H. Wang, H. Yi, J. Peng, G. Wang, Y. Liu, H. Jiang, et al., Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network, Energy Convers. Manag. 153 (2017) 409–422, https://doi.org/10.1016/j.enconman.2017.10.008, https://www.sciencedirect.com/science/article/pii/S019689041730910X.

[52] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, G. Sun, Short-term global horizontal irradiance forecasting based on a hybrid cnn-lstm model with spatiotemporal correlations, Renew. Energy 160 (2020) 26–41, https://doi.org/10.1016/j.renene.2020.05.150.

[53] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al., Attention is all you need 2017, https://doi.org/10.48550/ARXIV.1706.03762, arXiv:1706.03762.

[54] A.P. Parikh, O. Täckström, D. Das, J. Uszkoreit, A decomposable attention model for natural language inference 2016, https://doi.org/10.48550/ARXIV.1606.01933, arXiv:1606.01933.

[55] J. Qu, Z. Qian, Y. Pei, Day-ahead hourly photovoltaic power forecasting using attention-based cnn-lstm neural network embedded with multiple relevant and target variables prediction pattern, Energy 232 (2021) 120996, https://doi.org/10.1016/j.energy.2021.120996.

[56] L.S. Saoud, H. Al-Marzouqi, R. Hussein, Household energy consumption prediction using the stationary wavelet transform and transformers, IEEE Access 10 (2022) 5171–5183, https://doi.org/10.1109/access.2022.3140818.

[57] J. Liu, H. Zang, L. Cheng, T. Ding, Z. Wei, G. Sun, A transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting, Appl. Energy 342 (2023) 121160, https://doi.org/10.1016/j.apenergy.2023.121160.

[58] T. Yu, H. Zhu, Hyper-parameter optimization: a review of algorithms and applications 2020, https://doi.org/10.48550/ARXIV.2003.05689, arXiv:2003.05689.

[59] A. Feroz Mirza, M. Mansoor, M. Usman, Q. Ling, Hybrid inception-embedded deep neural network ResNet for short and medium-term PV-wind forecasting, Energy Convers. Manag. 294 (2023) 117574, https://doi.org/10.1016/j.enconman.2023.117574.

[60] H.B. Curry, The method of steepest descent for non-linear minimization problems, Q. Appl. Math. 2 (3) (1944) 258–261, http://www.jstor.org/stable/43633461.

[61] R. Courant, Variational methods for the solution of problems of equilibrium and vibrations, Bull. Am. Math. Soc. 49 (1) (1943) 1–23, https://doi.org/10.1090/s0002-9904-1943-07818-4.

[62] S. Amari, A theory of adaptive pattern classifiers, IEEE Trans. Electron. Comput. EC-16 (3) (1967) 299–307, https://doi.org/10.1109/pgec.1967.264666.

[63] S. ichi Amari, Backpropagation and stochastic gradient descent method, Neurocomputing 5 (4–5) (1993) 185–196, https://doi.org/10.1016/0925-2312(93)90006-o.

[64] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks, IEEE, 1995.

[65] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61, https://doi.org/10.1016/j.advengsoft.2013.12.007.

[66] J. Snoek, H. Larochelle, R.P. Adams, Practical Bayesian optimization of machine learning algorithms 25, https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf, 2012.

[67] M. Hagan, M. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Netw. 5 (6) (1994) 989–993, https://doi.org/10.1109/72.329697.

[68] C. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. 30 (2005) 79–82, https://doi.org/10.3354/cr030079.

[69] S. Kim, H. Kim, A new metric of absolute percentage error for intermittent demand forecasts, Int. J. Forecast. 32 (3) (2016) 669–679, https://doi.org/10.1016/j.ijforecast.2015.12.003.

[70] R.J. Hyndman, A.B. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. 22 (4) (2006) 679–688, https://doi.org/10.1016/j.ijforecast.2006.03.001.

[71] T.O. Hodson, Root-mean-square error (rmse) or mean absolute error (mae): when to use them or not, Geosci. Model Dev. 15 (14) (2022) 5481–5487, https://doi.org/10.5194/gmd-15-5481-2022.

[72] T. Chai, R.R. Draxler, Root mean square error (rmse) or mean absolute error (mae)?, https://doi.org/10.5194/gmdd-7-1525-2014, 2014.

[73] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021) e623, https://doi.org/10.7717/peerj-cs.623.

[74] M.L. Santos, X. García-Santiago, F.E. Camarero, G.B. Gil, P.C. Ortega, Application of temporal fusion transformer for day-ahead PV power forecasting, Energies 15 (14) (2022) 5232, https://doi.org/10.3390/en15145232.

[75] J. Pospíchal, M. Kubovčík, I.D. Luptáková, Solar irradiance forecasting with transformer model, Appl. Sci. 12 (17) (2022) 8852, https://doi.org/10.3390/app12178852.

[76] L. Wang, M. Mao, J. Xie, Z. Liao, H. Zhang, H. Li, Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model, Energy 262 (2023) 125592, https://doi.org/10.1016/j.energy.2022.125592.

[77] C. Zhang, T. Peng, M.S. Nazir, A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables, Electr. Power Syst. Res. 213 (2022) 108796, https://doi.org/10.1016/j.epsr.2022.108796.

[78] C. Brester, V. Kallio-Myers, A.V. Lindfors, M. Kolehmainen, H. Niska, Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations, Renew. Energy 207 (2023) 266–274, https://doi.org/10.1016/j.renene.2023.02.130.

[79] M. Alaraj, I. Alsaidan, A. Kumar, M. Rizwan, M. Jamil, Advanced intelligent approach for solar PV power forecasting using meteorological parameters for Qassim region, Saudi Arabia, Sustainability 15 (12) (2023) 9234, https://doi.org/10.3390/su15129234.

[80] P. Singh, N.K. Singh, A.K. Singh, Solar photovoltaic energy forecasting using machine learning and deep learning technique, in: 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), IEEE, 2022.

[81] Z. Garip, E. Ekinci, A. Alan, Day-ahead solar photovoltaic energy forecasting based on weather data using LSTM networks: a comparative study for photovoltaic (PV) panels in Turkey, Electr. Eng. (2023), https://doi.org/10.1007/s00202-023-01883-7.

[82] L. Li, J. Cao, T. Hong, M. Lu, W. Zhao, L. Fang, Photovoltaic power prediction based on wavelet analysis, in: Lecture Notes in Electrical Engineering, Springer Nature, Singapore, 2022, pp. 216–222.

[83] F. Tian, X. Fan, R. Wang, H. Qin, Y. Fan, A power forecasting method for ultra-short-term photovoltaic power generation using transformer model, Math. Probl. Eng. 2022 (2022) 1–15, https://doi.org/10.1155/2022/9421400.

[84] S. Admasie, J.S. Song, C.H. Kim, Optimal coordinated generation scheduling considering day-ahead PV and wind power forecast uncertainty, IET Gener. Transm. Distrib. 17 (11) (2023) 2545–2562, https://doi.org/10.1049/gtd2.12868.

[85] Y.Y. Hong, Y.T. Pan, C.C. Hsu, C.L.P.P. Rioflorido, J.B.D. Santos, One-hour ahead spatio-temporal solar GHI forecasting using long short-term memory, in: 2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA), IEEE, 2022.

[86] N.E. Michael, M. Mishra, S. Hasan, A. Al-Durra, Short-term solar power predicting model based on multi-step CNN stacked LSTM technique, Energies 15 (6) (2022) 2150, https://doi.org/10.3390/en15062150.

[87] S. Ziyabari, L. Du, S.K. Biswas, Short-term solar irradiance forecasting based on self-attentive transformers, in: 2022 IEEE Power & Energy Society General Meeting (PESGM, IEEE, 2022.

[88] C. Scott, M. Ahsan, A. Albarbar, Machine learning for forecasting a photovoltaic (PV) generation system, Energy 278 (2023) 127807, https://doi.org/10.1016/j.energy.2023.127807.

[89] T.M. Al-Jaafreh, A. Al-Odienat, Y.A. Altaharwah, The solar energy forecasting using LSTM deep learning technique, in: 2022 International Conference on Emerging Trends in Computing and Engineering Applications (ETCEA), IEEE, 2022.

[90] A.K. Parida, D. Kumar, R.R. Sahoo, B.K. Balabantaray, Medium term solar power prediction using stacked LSTM based deep learning technique, in: 2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), IEEE, 2022.

[91] A. Agga, A. Abbou, M. Labbadi, Y.E. Houm, I.H.O. Ali, CNN-LSTM: an efficient hybrid deep learning architecture for predicting short-term photovoltaic power production, Electr. Power Syst. Res. 208 (2022) 107908, https://doi.org/10.1016/j.epsr.2022.107908.

[92] D. Sadeghi, A. Golshanfard, S. Eslami, K. Rahbar, R. Kari, Improving PV power plant forecast accuracy: a hybrid deep learning approach compared across short, medium, and long-term horizons, Renew. Energy Focus 45 (2023) 242–258, https://doi.org/10.1016/j.ref.2023.04.010.

[93] N. Azizi, M. Yaghoubirad, M. Farajollahi, A. Ahmadi, Deep learning based long-term global solar irradiance and temperature forecasting using time series with multi-step multivariate output, Renew. Energy 206 (2023) 135–147, https://doi.org/10.1016/j.renene.2023.01.102.

[94] O. Rubasinghe, T. Zhang, X. Zhang, S.S. Choi, T.K. Chau, Y. Chow, et al., Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high PV penetration, Appl. Energy 333 (2023) 120641, https://doi.org/10.1016/j.apenergy.2023.120641.

[95] J. Xiao, F. Li, F. Wang, G. Liu, X. Wang, Q. Liu, et al., A forecasting method of photovoltaic power generation based on NeuralProphet and BiLSTM, in: 2022 IEEE 10th International Conference on Information, Communication and Networks (ICICN), IEEE, 2022.

[96] R.L. de C. Costa, Convolutional-LSTM networks and generalization in forecasting of household photovoltaic generation, Eng. Appl. Artif. Intell. 116 (2022) 105458, https://doi.org/10.1016/j.engappai.2022.105458.

[97] T.N. Trong, H.V.X. Son, H.D. Dinh, H. Takano, T.N. Duc, Short-term PV power forecast using hybrid deep learning model and variational mode decomposition, Energy Rep. 9 (2023) 712–717, https://doi.org/10.1016/j.egyr.2023.05.154.

[98] L.J. Piotrowski, F.A. Farret, Forecasting of photovoltaic power generation using deep learning AI, in: 2022 14th Seminar on Power Electronics and Control (SEPOC), IEEE, 2022.

[99] H. Jeong, Predicting the output of solar photovoltaic panels in the absence of weather data using only the power output of the neighbouring sites, Sensors 23 (7) (2023) 3399, https://doi.org/10.3390/s23073399.

[100] G. Li, S. Guo, X. Li, C. Cheng, Short-term forecasting approach based on bidirectional long short-term memory and convolutional neural network for regional photovoltaic power plants, Sustain. Energy, Grids Netw. 34 (2023) 101019, https://doi.org/10.1016/j.segan.2023.101019.

[101] D. Kothona, K. Spyropoulos, C. Valelis, C. Sarigiannidis, K.C. Chatzisavvas, G.C. Christoforidis, Efficient 24-hour ahead PV energy production forecasting employing a transformer-based model, in: 2022 2nd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED), IEEE, 2022.

[102] N. Khortsriwong, P. Boonraksa, T. Boonraksa, T. Fangsuwannarak, A. Boonsrirat, W. Pinthurat, et al., Performance of deep learning techniques for forecasting PV power generation: a case study on a 1.5 MWp floating PV power plant, Energies 16 (5) (2023) 2119, https://doi.org/10.3390/en16052119.

[103] J. Zhang, Y. Hao, R. Fan, Z. Wang, An ultra-short-term PV power forecasting method for changeable weather based on clustering and signal decomposition, Energies 16 (7) (2023) 3092, https://doi.org/10.3390/en16073092.

[104] Q.T. Phan, Y.K. Wu, Q.D. Phan, H.Y. Lo, A novel forecasting model for solar power generation by a deep learning framework with data preprocessing and postprocessing, IEEE Trans. Ind. Appl. 59 (1) (2023) 220–231, https://doi.org/10.1109/tia.2022.3212999.

[105] A. Demir, L.F. Gutierrez, A.S. Namin, S. Bayne, Solar irradiance prediction using transformer-based machine learning models, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022.

[106] J. Sherozbek, J. Park, M.S. Akhtar, O.B. Yang, Transformers-based encoder model for forecasting hourly power output of transparent photovoltaic module systems, Energies 16 (3) (2023) 1353, https://doi.org/10.3390/en16031353.

[107] A. Agga, A. Abbou, M. Labbadi, R. Touileb, Short-term PV plant power production forecasting and hyperparameters grid search for LSTM and MLP models, in: Digital Technologies and Applications, Springer International Publishing, 2022, pp. 181–189.

[108] J. Huang, S. Kaewunruen, Forecasting energy consumption of a public building using transformer and support vector regression, Energies 16 (2) (2023) 966, https://doi.org/10.3390/en16020966.

[109] H. Wu, J. Xu, J. Wang, M. Long, Autoformer: decomposition transformers with auto-correlation for long-term series forecasting 2021, https://doi.org/10.48550/ARXIV.2106.13008, arXiv:2106.13008.

[110] S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, S.M. Mirjalili, Salp swarm algorithm: a bio-inspired optimizer for engineering design problems, Adv. Eng. Softw. 114 (2017) 163–191, https://doi.org/10.1016/j.advengsoft.2017.07.002.

[111] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, pp. 770–778.

[112] S. Siami-Namini, N. Tavakoli, A.S. Namin, A comparison of ARIMA and LSTM in forecasting time series, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018.

[113] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544, https://doi.org/10.1109/tsp.2013.2288675.

[114] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., An image is worth 16x16 words: transformers for image recognition at scale, https://doi.org/10.48550/ARXIV.2010.11929, 2020.

[115] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.X. Wang, et al., Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting 2019, https://doi.org/10.48550/ARXIV.1907.00235, arXiv:1907.00235.

[116] B. Lim, S.O. Arık, N. Loeff, T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, Int. J. Forecast. 37 (4) (2021) 1748–1764, https://doi.org/10.1016/j.ijforecast.2021.03.012.

[117] N. Wu, B. Green, X. Ben, S. O'Banion, Deep transformer models for time series forecasting: the influenza prevalence case 2020, https://doi.org/10.48550/ARXIV.2001.08317, arXiv:2001.08317.

[118] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI Blog 1 (8) (2019) 9.