Development and trending of deep learning methods for wind power predictions

Hong Liu[1,2] · Zijun Zhang[1,2]

Abstract

With the increasing data availability in wind power production processes due to advanced sensing technologies, data-driven models have become prevalent in studying wind power prediction (WPP) methods. Deep learning models have gained popularity in recent years due to their ability of handling high-dimensional input, automating data feature engineering, and providing high flexibility in modeling. However, with a large volume of deep learning based WPP studies developed in recent literature, it is important to survey the existing developments and their contributions in solving the issue of wind power uncertainty. This paper revisits deep learning-based wind power prediction studies from two perspectives, deep learning-enabled WPP formulations and developed deep learning methods. The advancement of WPP formulations is summarized from the following perspectives, the considered input and output designs as well as the performance evaluation metrics. The technical aspect review of deep learning leveraged in WPPs focuses on its advancement in feature processing and prediction model development. To derive a more insightful conclusion on the so-far development, over 140 recent deep learning-based WPP studies have been covered. Meanwhile, we have also conducted a comparative study on a set of deep models widely used in WPP studies and recently developed in the machine learning community. Results show that DLinear obtains more than 2% improvements by benchmarking a set of strong deep learning models. Potential research directions for WPPs, which can bring profound impacts, are also highlighted.

Abbreviations

ACE	Average coverage error
AE	Autoencoder
AM	Attention-based model
ARIMA	Auto-regressive integrated moving average

✉ Zijun Zhang
zijzhang@cityu.edu.hk

1 School of Data Science, City University of Hong Kong, Hong Kong SAR, China

2 Centre for Systems Informatics Engineering, City University of Hong Kong Shenzhen Research Institute, Guangdong, Shenzhen 518057, China

AS   Atomic search
Bi-LSTM   Bi-directional LSTM
CNN   Convolutional neural network
CRPS   Continuous ranked probability score
CS   Cuckoo search
CSA   Clonal selection algorithm
CSO   Crisscross optimization
CWC   Coverage width-based criterion
DA   Dragonfly algorithm
DBN   Deep belief network
DL   Deep learning
DL-WPP   Deep learning based WPP
DNN   Deep neural network
DT   Decision tree
ELM   Extreme learning machine
EMD   Empirical mode decomposition
EN   Elastic net
EWT   Empirical wavelet transform
GA   Genetic algorithm
GCN   Graph convolutional neural network
GMM   Gaussian mixture model
GNN   Graph neural network
GRU   Gated recurrent unit
GS   Grid search
GWO   Grey wolf optimization
KDE   Kernel density estimation
KF   Kalman filters
LASSO   Least absolute shrinkage and selection operator
LR   Linear regression
LSTM   Long short term memory
LUBE   Lower upper bound estimation
MAE   Mean absolute error
ACE   Average coverage error
AE   Autoencoder
AM   Attention-based model
ARIMA   Auto-regressive integrated moving average
AS   Atomic search
Bi-LSTM   Bi-directional LSTM
CNN   Convolutional neural network
CRPS   Continuous ranked probability score
CS   Cuckoo search
CSA   Clonal selection algorithm
CSO   Crisscross optimization
CWC   Coverage width-based criterion
DA   Dragonfly algorithm
DBN   Deep belief network
DL   Deep learning
DL-WPP   Deep learning based WPP
DNN   Deep neural network

DT	Decision tree
ELM	Extreme learning machine
EMD	Empirical mode decomposition
EN	Elastic net
EWT	Empirical wavelet transform
GA	Genetic algorithm
GCN	Graph convolutional neural network
GMM	Gaussian mixture model
GNN	Graph neural network
GRU	Gated recurrent unit
GS	Grid search
GWO	Grey wolf optimization
KDE	Kernel density estimation
KF	Kalman filters
LASSO	Least absolute shrinkage and selection operator
LR	Linear regression
LSTM	Long short term memory
LUBE	Lower upper bound estimation
MAE	Mean absolute error

1 Introduction

Wind power is a critical pillar in the pursuit of global carbon neutrality, and its installation capacity has steadily increased in recent decades as reported by Global Wind Energy Council (GWEC 2022). This upward trend provides a solid foundation for powering our society with clean and renewable energy, while also mitigating the environmental pollution caused by fossil fuels. However, the volatility of wind power and its increasing penetration pose new challenges to the safety and stability of power grid operations. Studying wind power prediction (WPP) is critical and valuable because accurate results can facilitate power grids and wind farms to better manage the wind power generation uncertainty. Accurate wind power predictions can benefit many downstream applications, such as more efficient wind power integration (Wang et al. 2019a), intelligent market operations (Usaola et al. 2004), and monitoring wind turbine performance (Ma et al. 2014). WPP has become a classical problem in the renewable energy field and has attracted a large volume of studies (Landberg 1999; Liu et al. 2010; Liu et al. 2021b, c, d, e, a; Woo et al. 2019). These studies persistently aim to develop new technologies that achieve state-of-the-art performance in terms of accuracy and reliability.

In the literature, various approaches have been proposed for improving the accuracy of wind power predictions. The advancements in this field can be analyzed from two aspects, the problem formulation and methodological advancement, as displayed in Fig. 1. The formulation of the WPP problem can be further categorized based on input and output design considerations. Early studies attempted to estimate wind power using environmental and physical attributes (Landberg 1999), such as the topological information and meteorological information, with integrating wind power conversion dynamics or historical wind power generation records (Liu et al. 2010). With the wide deployment and continuous advancement of supervisory control and data acquisition (SCADA) systems in commercial wind farms, recent studies (Liu et al. 2021d; Woo et al. 2019; Khodayar and Wang 2018)
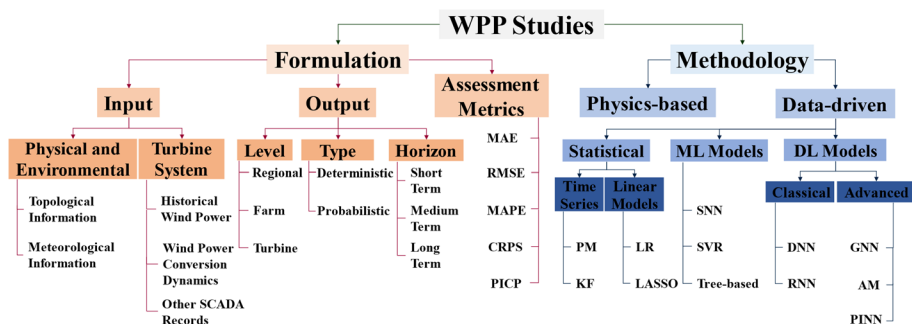
Fig. 1 The categorization of WPP studies in literature

have attempted to learn spatial–temporal correlations from the SCADA data to future wind power. To analyze temporal correlations, multivariate time series data were used to reflect the system dynamics, while to learn spatial correlations due to environmental factors, data collected from different sites were integrated into a tensor (Liu et al. 2021d). Another study (Woo et al. 2019) projected wind turbines into a 2-dimensional grid to reflect their geo-spatial relationship while such strategy might lose its effectiveness in cases that turbines were sparsely distributed. An alternative solution (Khodayar et al. 2018) involved organizing wind turbines into a graph constructed using locations and mutual information. Existing WPP studies can also be categorized based on various output considerations, which include the power output level, prediction horizon, and prediction type. Prediction of wind power outputs has been studied at three levels, the region, wind farm, and wind turbine. Depending on the prediction horizon, WPP tasks can be classified as short-term (0–6h ahead), medium-term (6–24h ahead), or long-term prediction (more than 24h ahead) (Khodayar et al. 2018). According to the considered output type, WPP tasks are classified into deterministic and probabilistic predictions. Deterministic predictions provide estimated spot values of future wind power while probabilistic predictions quantify the uncertainty of future wind power by inferring confidence intervals, quantiles, or even distributions.

Metrics for assessing the performance of WPP constitute another important consideration in problem formulation. The choice of evaluation metric depends on the type of WPP tasks. For deterministic WPP tasks, RMSE is the most widely used evaluation metric. Other metrics, such as mean absolute error (MAE), mean absolute percentage error (MAPE), and their standard deviations, are also meaningful for measuring WPP performance from different perspectives. In probabilistic WPP tasks, well-known and commonly used metrics include continuous ranked probability score (CRPS) and prediction interval coverage probability (PICP).

In addition to the problem formulation, WPP studies can also be categorized based on methodological development as shown in the right part of Fig. 1. Existing efforts in studying WPP methods have been devoted into two directions, physics-based and data-driven. Physics-based methods (Landberg 1999) focused on developing WPP models based on physical attributes and principles. On the other hand, data-driven methods (Sideratos and Hatziargyriou 2007; Brown et al. 1984; Treiber et al. 2016; Madhiarasan and Deepa 2017) focused on developing models for estimating future wind power outputs based on SCADA data and even a joint consideration of SCADA data as well as historical measurements and predictions of physical and environmental attributes. Three series of

data-driven models (Sideratos and Hatziargyriou 2007) have been developed in literature, statistical models (Brown et al. 1984), classical machine learning (ML) models (Treiber et al. 2016), and recent deep learning (DL)-based models (Madhiarasan and Deepa 2017). Statistical models include time series models, such as persistent method (PM) (Bludszuweit et al. 2008), auto-regressive integrated moving average (ARIMA) (Chen et al. 2009), and Kalman filters (KF) (Bossanyi 1985), which model historical patterns of wind power to estimate its future, as well as linear regression (LR) models, such as least absolute shrinkage and selection (LASSO) (Cavalcante et al. 2017), which incorporates data of exogenous factors in addition to the consideration of historical wind power records. Classical machine learning algorithms, such as support vector regressor (SVR) (Zendehboudi et al. 2018), shallow neural networks (SNN) (Wang et al. 2021b), tree-based models like decision tree (DT) for regression (Heinermann and Kramer 2016) and random forest (RF) for regression (Lahouar and Slama 2017), etc., have been applied to model nonlinearities in wind power generation using SCADA data. With the recent development of deep learning, there has been increasing interest in applying DL-based models for better WPP performance. Recent DL-based models include classical ones, such as deep neural network (DNN) (Methaprayoon et al. 2007), convolutional neural network (CNN) (Liu et al. 2021d), recurrent neural network (RNN) (Cali and Sharma 2019), as well as latest ones, such as graph neural network (GNN) (Wu et al. 2022), attention-based model (AM) (Tian et al. 2022) and Physics-informed graph network (PINN) (Pombo et al. 2022).

DL-based methods have brought unique advantages on data processing and feature engineering into WPP modeling. Convolutional and recurrent mechanisms automate the process of embedding a vast input space from spatial and temporal aspects to derive a low-dimensional representation. Researchers have hypothesized that this new data processing paradigm carries a larger scope of information in data to benefit WPPs. Two well-known variants of RNN models, long short term memory (LSTM) (Liu et al. 2019b, a) and gated recurrent unit (GRU) (Liu et al. 2022b, a), have been frequently applied to extract temporal patterns of wind power data. Signal processing techniques, such as the empirical mode decomposition (EMD) (Abedinia et al. 2020), variational mode decomposition (VMD) (Abdoos 2016), and SSA (Dong et al. 2017), have been used to filter out noises and extract data fluctuation modes to facilitate recurrent neural networks to better learn latent patterns in time series for WPPs.

DL techniques have been developed to solve both deterministic and probabilistic WPP tasks. Various DL-based structures (Lahouar and Slama 2017; Methaprayoon et al. 2007; Liu et al. 2021d; Cali and Sharma 2019) have been developed for direct inference of deterministic values of future wind power outputs. To perform probabilistic WPP tasks, DL-based models have been extended for the quantile regression (QR) (Nielsen et al. 2006), lower upper bound estimation (LUBE) (Khosravi and Nahavandi 2013), kernel density estimation (KDE) (Bessa et al. 2012), and mixture density network (MDN) (Zhang et al. 2020a, b, c). In a few studies, clustering algorithms, such as K-means (Wang et al. 2018), C-means (Yang et al. 2021a), EM (Liu et al. 2018a, b, c), etc., have been employed to group wind power time series into different clusters representing different wind power fluctuation patterns, which help specify deep learning models into prediction scenarios.

Although deep learning has proven effectiveness in processing large volumes of data, its robustness and generalization ability across different WPP tasks still require further improvement. One possible reason for this is the fact that the hyperparameters and configurations of deep models, such as the number of layers and neurons in each layer, can greatly affect their performance. To address this challenge, a number of optimization algorithms have been developed, such as the grid search (GS) (Zhang et al.

2014), cuckoo search (CS) (Li et al. 2021), genetic algorithm (GA) (Huang et al. 2015), particle swarm optimization (PSO) (Amjady et al. 2011), and grey wolf optimization (GWO) (Lu et al. 2020), which aim to attain the optimal architecture for a given deep learning model. Another attempt observed is incorporating attention mechanism for dynamically processing features after the input or latent layers to adaptively respond WPP scenarios (Yang and Zhang 2021). An emerging trend is developing PINN based WPP models (Huang and Wang 2022). Physics and domain knowledge were leveraged to better govern the network design and prediction performance to enable a better generalization in WPPs (Lagomarsino-Oneto et al. 2023).

This paper provides a survey of recent WPP studies powered by deep learning from two dimensions, the WPP formulations and WPP methods. We discuss recent innovations in WPP formulations in terms of the input designs, output types, and evaluation metrics. In terms of methodology, this paper offers a systematic summary of recent advancements in four main components of many deep learning-based WPP (DL-WPP) modeling pipelines, the input signal processing, data pattern clustering, latent feature engineering, and model architecture optimization. In addition, we review three types of learning schemes based on these four components. The review focuses particularly on the latent feature engineering component, which is a significant benefit of deep learning due to its ability to accommodate high-dimensional input space and automate the engineering of latent features representing such input space. Based on the characteristics of the deep learning models, 28 deep learning network configurations covered in this review are grouped into 8 types. Finally, future research directions are discussed by analyzing the limitations of current WPP studies and identifying the hot spots of great potential for benefiting the technical development in WPP studies. The contributions of this paper can be summarized as follows.

- Compared with existing WPP reviews (Vargas et al 2019; Qian et al. 2019; Khalid and Javaid 2020; Wang et al. 2021c, 2016, 2020; Jung and Broadwater 2014; Marugán et al. 2018; Liu et al. 2020a, b), which commonly listed a large set of reported WPP methods, this paper surveyed the literature with jointly landing discussions in the WPP problem formulation evolution as well as the mechanism advancement in data-driven WPP from classical models to carefully generate a clearer image on the WPP advancement driven by the recent rapid DL development.
- This paper sheds light on surveying over 140 recent studies investigating DL-WPP methods. Based on this comprehensive analysis of recent state-of-the-art studies, this paper provides a high-level overview of the current frontiers of data-driven WPPs.
- This paper serves as an interface for accessing emerging WPP developments, indexing them from multiple perspectives. Readers can quickly refer to their interested WPP studies according to the data organization, feature engineering, evaluation metrics, deep learning-based modeling frameworks, etc.
- This paper provides details of current state-of-the-art DL-based modeling methods, guiding the audience to replicate these WPP models. Meanwhile, a comparative analysis of the performance of latest DL-based WPP models is conducted based on datasets of three commercial wind farms to facilitate the audience to further understand the effectiveness of these models on WPP tasks.
- Discussions of the up-to-date promising future research directions are provided via analyzing the limitations of existing WPP studies and identifying new DL hot spots for possibly advancing WPP performance. These future trends aim to serve as

guidance for researchers to study more advanced WPP methods leading to a more accurate and reliable WPP performance.

The remaining parts of this paper are organized as follows. Section 2 summarizes recent DL-WPP studies in terms of the problem formulations. Section 3 revisits and elaborates details of emerging deep learning-based methods for WPP tasks. Section 4 provides a discussion of promising future research trends and Sect. 5 concludes the insights of this survey study.

2 Data-driven wind power prediction problem formulation

Let x represent the input data related to wind power generations and y represent the actual wind power output. The objective of data-driven WPP studies is to develop a model $f(\cdot)$ that predicts the wind power output, where the predicted output is denoted by $\hat{y} = f(x)$. Typically, the model $f(\cdot)$ is trained by minimizing the error measures between the prediction $\hat{y}$ and the true value y.

Based on this formulation, new developments have occurred on the design of input x and the output y considered in WPP studies. Along with the innovation in problem formulation, we also observe new developments in metrics for more effectively evaluating WPP performance. Thus, a survey from this perspective is also needed. In this section, we first review on the scope of the considered input x and its organization forms. Then, we summarize the different settings of output y considered in WPP studies. Finally, we survey the evaluation metrics used in WPP tasks.

2.1 WPP input and its organization

As shown in Table 1, recent WPP studies have mainly focused on numerical weather predictions (NWP) and SCADA data as inputs. The NWP input data, $x \in \mathbb{R}^{T \times N_{NWP}}$, where T represents the length of the sequence and N_{NWP} represents the number of NWP attributes (e.g., the predicted temperature and wind speed), have been utilized in WPP model development (Bessa et al. 2012). These NWP data are usually extracted from the data provided by other weather forecasting sources. For example, in studies (Hong et al. 2016; Klinges et al. 2022; Kokkos et al. 2021), the NWP data extracted from the global gridded weather data were utilized to enable the consideration of gridded macroclimatic variables as the input.

Wind turbine SCADA data is another frequently employed data source for developing inputs $x \in \mathbb{R}^{T \times N_{SCADA}}$, where N_{SCADA} describes the number of SCADA attributes, e.g., the historical wind power, historical wind power ramp, wind speed, wind direction, generator torque, blade pitch angle, etc., in WPP studies (Methaprayoon et al. 2007; Liu et al. 2021b, c, d, e, a). With the advancement of the SCADA system, more related attributes become available for enhancing WPP. Several recent studies (Valsaraj et al. 2020, 2022; Külüm et al. 2023; Weide et al. 2022) targeted at enriching the information supply for WPP tasks by including anemometric data collected at higher heights, which aimed to overcome the deficit of anemometers mounted on the back end of the turbine nacelle on measuring wind speeds.

Table 1 Inputs of WPP studies

Data source	Spatial influence	References
NWP	Not considered	Bessa et al. (2012); Dong et al. (2016); Ghadi et al. (2014); Ahmadpour and Farkoush (2020); Yang et al. (2021a, b); Wen et al. (2019); Li et al. (2015a); Yang et al. (2013); Men et al. (2016)
SCADA data	Not considered	Treiber et al. (2016); Methaprayoon et al. (2007); Liu et al. (2021c); Liu et al. (2019b, a); Liu et al. (2022b, a); Abedinia et al. (2020); Abdoos (2016); Li et al. (2022); Liu et al. (2018a, b, c); Huang et al. (2015); Lu et al. (2020); Chitsaz et al. (2015); Osório et al. (2015); Catalão et al. (2010); Qureshi et al. (2017); Azimi et al. (2016); Wu et al. (2020a, b); Yin et al. (2017); Sun (2021); Wang et al. (2021a); Peng et al. (2016); Prósper et al. (2019); Yuan et al. (2019); Banik et al. (2020); Song et al. (2020); Shi et al. (2012); Li et al. (2020a, b, c); Zu and Song (2018); Jahangir et al. (2020); Han et al. (2019a, b); Dong et al. (2017); Ouyang et al. (2020); Amjady (2011); Li et al. (2015a, b); Jiao et al. (2018); Zhang et al. (2019a, b); Han et al. (2019a); Kisvari et al. (2021); Wang et al. (2019a, b); Hu et al. (2020); Yan and Wu (2020); Mahmoud et al. (2018); Zhang et al. (2016); Wang et al. (2017a, b, c); Wan et al. (2016a, b); Yan et al. (2015); Wan et al. (2013a, b); Wan et al. (2016a); Hu et al. (2018); Farah et al. (2022)
SCADA data	Considered	Liu et al. (2021d); Woo et al. (2019, 2018); Khodayar and Wang (2018); Severiano et al. (2021); Niu et al. (2020); Ju et al. (2019); Wang et al. (2021a); Hong and Rioflorido (2019); Yu et al. (2019a); Wang et al. (2017a, b, c); Sun and Zhao. (2020)
NWP and SCADA data	Not considered	Cali et al. (2019); Wang et al. (2018); Zhang et al. (2020b, 2015); Bessa et al. (2009); Zhang et al. (2019a, b); Yan et al. (2018); Ding et al. (2019); Gallego-Castillo et al. (2016); Kou et al. (2014); Wan et al. (2013a, b)
NWP and SCADA data	Considered	Yu et al. (2022a, 2020); Higashiyama et al. (2018); Zhang and Wang (2018); Kim and Hur (2018)

A joint consideration of both weather and historical data as the input for WPP tasks has also been reported (Ghadi et al. 2014; Cali et al. 2019). Both NWP and SCADA attributes were simply integrated into one input described as $x = [x_{NWP}, x_{SCADA}]$, where $x_{NWP} \in \mathbb{R}^{T_{NWP} \times N_{NWP}}$ represents the part containing NWP data, $x_{SCADA} \in \mathbb{R}^{T_{SCADA} \times N_{SCADA}}$ represents the input SCADA data, T_{NWP} is the time length of NWP data considered, and T_{SCADA} is the time length of SCADA data considered.

To incorporate more relevant attributes and enrich WPP input, recent studies (Pombo et al. 2022; Huang and Wang 2022; Li and Zhang 2022; Zhang and Zhao 2021) have also considered applying the physics-informed methods to augment data and improve accuracy. Such efforts jointly consider knowledge from astronomy, fluid dynamics, power curve modelling, and the spatiotemporal correlation among sensors and actuators. Astronomy knowledge usually comprises of the solar position, which can be extracted by means of the azimuth and elevation angles, as well as the weather conditions which is usually captured by installed cameras. Fluid dynamics consider the flux which measures a quantity's flow rate carried by a moving fluid per unit of normalized area, as well as the turbulence intensity which evaluates the level of velocity fluctuation of a fluid. Power curve presents the relationship between wind speed and output power.

To further improve WPP performance, studies (Liu et al. 2021d; Woo et al. 2019; Khodayar and Wang 2018) have explored incorporating higher dimensional information from data sources to build more meaningful WPP inputs. Data considered in the input was expanded from a single wind turbine or a single wind farm site to a group of its neighbors with a size P which enables a consideration of the spatial influences, such as the wake effects, geostrophic wind, ground roughness, etc., as shown in Fig. 2. These studies (Liu et al. 2021d; Woo et al. 2019; Khodayar and Wang 2018) organized high dimensional input data using three strategies, the stacking, projection, and graph-based model. Meanwhile, although the following description is offered with using the SCADA data as an illustrative example, the joint consideration of meteorological measurements and predictions as well as SCADA data can also be expanded to a group of neighbors of the targeted turbine for learning the spatial pattern among those attributes.
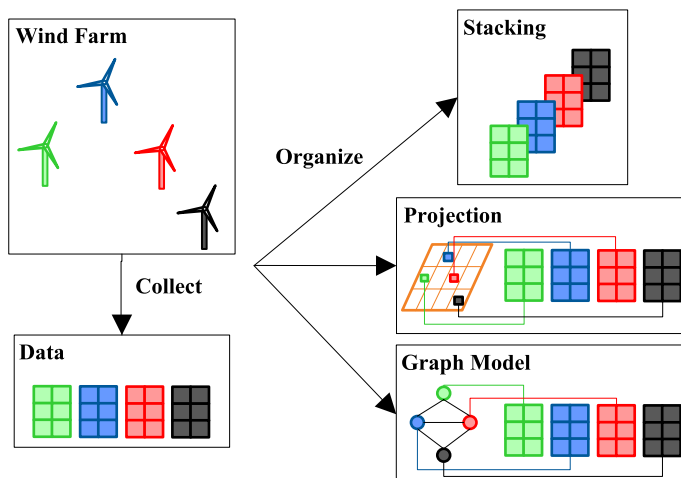

Fig. 2 Three types of data organization considering spatial influence

2.1.1 Stacking strategy

In (Liu et al. 2021d), SCADA data collected from P turbines were directly stacked into a 3-dimensional tensor, $x \in \mathbb{R}^{T \times P \times N_{SCADA}}$ to consider richer information. However, the organized input x does not reflect the actual geographical distribution of the turbines.

2.1.2 Projection strategy

In (Woo et al. 2019), data from multiple sources were projected onto a 2-dimensional grid to construct a 4-dimensional tensor $x \in \mathbb{R}^{T \times W \times H \times N_{SCADA}}$ using the SCADA data of P turbines, where W and H are the width and height of grids. This strategy considers the geographical distribution of wind turbines. However, once the distribution of the turbines is sparse, the majority values in x are blank, leading to inefficiency and impaired WPP performance.

2.1.3 Graph-based modeling strategy

In (Khodayar and Wang 2018; Liu et al. 2023), a graph is developed to model the relationship between wind turbines via incorporating geographical information, such as the longitude, latitude, and altitude as well as the mutual information between the wind turbines. This strategy expands the input x to a tuple $x = [G, x_{SCADA}]$, where $x_{SCADA} \in \mathbb{R}^{T \times P \times N_{SCADA}}$ is the stacked SCADA data and G is the modeled graph given by the correlation matrix. Efficient correlation matrices considered include following candidates:

- Mutual information matrix: $G_{ij} = MI(i,j)$.
- Exponential negative distance matrix: $G_{ij} = e^{-Dis(i,j)}$.
- Multiplication of mutual information and exponential Negative Distance: $G_{ij} = MI(i,j) \times e^{-Dis(i,j)}$
- Mutual information controlled exponential Negative Distance:
  $G_{ij} = \begin{cases} 0, if MI(i,j) < \tau \\ e^{-Dis(i,j)}, if MI(i,j) \geq \tau \end{cases}$

where $MI(i,j)$ is the mutual information of turbines i and j, $Dis(i,j)$ is the distance between turbines i and j, as well as τ is a predefined threshold.

2.2 WPP output settings

As shown in Table 2, recent WPP studies considered different output settings. Regarding the targeted prediction level, most studies targeted on either the wind turbine level power prediction (Liu et al. 2021d; Woo et al. 2019; Khodayar and Wang 2018; Zhang et al. 2019a, b) or the wind farm level power prediction (Dong et al. 2016; Ghadi et al; 2014). A few studies (Osório et al. 2015; Catalão et al. 2010) uniquely discussed predicting a regional level wind power output, which is the total power output of multiple wind farms.

Depending on the prediction output type, two WPP tasks, the deterministic WPP and probabilistic WPP, are studied. Deterministic WPPs (Treiber et al. 2016; Madhiarasan and Deepa 2017; Bludszuweit et al. 2008) predict the spot value of wind power. However, point

Table 2 Output settings of WPP studies

Level	Type	Horizon	References
Regional	Deterministic	Short-term	Osório et al. (2015), Catalão et al. (2010)
		Median-term	Jahangir et al. (2020)
		Long-term	Ahmadpour et al. (2020), Hu et al. (2020)
	Probabilistic	Short-term	Kou et al. (2014)
		Long-term	Ahmadpour et al. (2020), Yu et al. (2020)
Wind Farm	Deterministic	Short-term	Liu et al. (2021b, c, d, e, a), Tian et al. (2022), Liu et al. (2019b, a), Abdoos (2016), Huang et al. (2015); Lu et al. (2020); Chitsaz et al. (2015); Wu et al. (2020a, b); Yin et al. (2017); Sun (2021); Wang et al. (2021a, b, c, d); Peng et al. (2016); Zu and Song (2018); Han et al. (2019a, 2019b); Dong et al. (Ouyang et al. (2020; Amjady et al. (2011; Jiao et al. (2018; Zhang et al. (2019a, b; Yan and Wu (2020, 2017), Zhang et al. (2016), Wang et al. (2017a, b, c), Niu et al. (2020), Wang et al. (2021a), An et al. (2011), Han and Tong (2020), Tan et al. (2021)
		Median-term	Cali et al. (2019), Abedinia et al. (2020), Wang et al. (2018), Dong et al. (2016), Amjady et al. (2011), Hong and Rioflorido (2019), Zhang et al. (2019a), Amjady and Abedinia (2017)
		Long-term	Ghadi et al. (2014), Ahmadpour and Farkoush (2020), Azimi et al. (2016), Li et al. (2015a, b), Bessa et al. (2009), Yan et al. (2018), Higashiyama et al. (2018), Zhang et al. (2020a, b, c)
	Probabilistic	Short-term	Liu et al. (2022b, a), Khosravi and Nahavandi (2013), Wen et al. (2019), Li et al. (2015a), Wang et al. (2019a, b), Mahmoud et al. (2018), Zhang et al. (2016), Wang et al. (2017a, b, c), Wan et al. (2016a, b, 2013a), Yan et al. (2015), Hu et al (2018), Severiano et al. (2021), Shi et al. (2017a), Zhang and Wang (2016), Dowell and Pinson (2015)
		Median-term	Gallego-Castillo et al. (2016), Zhang et al. (2015), Zhang and Wang (2018), Hong et al. (2016), Li et al. (2020a, b, c)
		Long-term	Bessa et al. (2012), Ahmadpour and Farkoush (2020), Yang et al. (2013)
	Deterministic	Short-term	Liu et al. (2021d, 2018a, 2021c), Woo et al. (2019, 2018), Treiber et al. (2016), Li et al. (2022), Song et al. (2020), Shi et al. (2012), Kisvari et al. (2021), Ju et al. (2019), Yu et al. (2019a, b), Sun and Zhao (2020), Yu et al. (2022a, 2019a), Heydari et al. (2021)
		Median-term	Yang et al. (2021a), Ding et al. (2020), Viet et al. (2020)
		Long-term	Li et al. (2020a, b, c)

Table 2 (continued)

Level	Type	Horizon	References
Wind Turbine		Short-term	Ma et al. (2014), Yuan et al. (2019), Wang et al. (2017a, b, c), Liu et al. (2022b, a)
	Probabilistic	Median-term	Kim and Hur (2018)
		Long-term	Men et al. (2016)

Monitoring wind farm performances (Ma et al. 2014)

predictions are prone to errors and lack of the capability to quantify the future wind power uncertainty. To address this issue, probabilistic WPPs (Khosravi and Nahavandi 2013; Bessa et al. 2012) investigate the provision of the confidence interval, quantile, or distribution of the future wind power, which enable the uncertainty quantification.

Meanwhile, WPP prediction tasks are typically classified into three types based on the prediction horizons which aim to serve different downstream tasks:

1. Short-term WPP (0–6h ahead): Short-term WPPs are the most commonly discussed. The results can be applied to enhance the efficiency of wind power utilization in grids, scheduling (Wang et al. 2019a), reducing regulation costs in electricity market operations (Usaola et al. 2004), as well as optimizing and monitoring wind farm performances (Ma et al. 2014).
2. Median-term WPP (6–24h ahead): Median-term WPPs mainly aim to support dynamic operations in power systems, such as balancing between the wind generation and load (Menemenlis et al. 2012), energy scheduling (Shi et al. 2012), and load following (Paterakis et al. 2014).
3. Long-term WPP (more than 24 h ahead): Long-term WPPs contribute into a variety of downstream tasks, such as electricity pricing (Wang et al. 2017a, b, c), unit commitment (Wang et al. 2008), turbine maintenance (Ren et al. 2021), storage management (Blonbou et al. 2011), and power trading (Pircalabu et al. 2017).

    Another important extension of WPP involves wind power ramp prediction (Cui et al. 2023; Hu et al. 2023), which uniquely sheds the research light on better supporting WPP tasks against the future sudden large wind speed changes. Both the point prediction of wind power ramps (Gallego et al. 2014) and the probabilistic prediction of wind power ramps (He et al. 2023) have been studied.

2.3 WPP performance evaluation metrics

To evaluate the WPP performance, a variety of assessment metrics have been designed and applied.

Table 3 summarizes the metrics applied in recent WPP studies (Chitsaz et al. 2015; Osório et al. 2015; Catalão et al. 2010; Qureshi et al. 2017). These metrics can be divided into two groups, metrics for evaluating deterministic and probabilistic WPPs.

In deterministic WPPs, absolute error-based errors including MAE, NMAE, MAPE and symmetric MAPE (sMAPE) as well as squared error-based metrics including mean square error (MSE), mean square error (RMSE), and normalized mean square error (NRMSE) are most utilized. MAE is a basic but widely considered metric that takes the average absolute error of each pair of observed and predicted wind power outputs. NMAE presents the normalized version of MAE according to the maximal wind power generation. MAPE and sMAPE present the proportion of the absolute error comparing with the actual wind power generation and the average of prediction and actual wind power, respectively. Squared error metrics penalize the large prediction error via taking the square of the error. In the case that two models show the same performances in terms of MAE on a dataset, the one with larger errors in certain data points is more likely to obtain a larger RMSE.

Evaluating probabilistic WPP performance is more complex than evaluating deterministic WPP. Three types of metrics, prediction interval-based, distribution-based, and quantile-based metrics, are considered. Regarding prediction intervals, reliability and sharpness are critical performance measures. PICP is the most widely applied reliability metric,

Table 3 Metrics considered in evaluating WPP performance

Type	Metric	Formulation	References
	MSE	$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2$	Huang et al. (2015), (Wang et al. (2020), Kisvari et al. (2021), Ju et al. (2019), Yu et al. (2019a, 2022a), Viet et al. (2020)
	RMSE	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2}$	Liu et al. (2021d, 2018a), Woo et al. (2019), Tian et al. (2022), Liu et al. (2019b, a), Li et al. (2022), Huang et al. (2015), Lu et al. (2020), Yang et al. (2021a, b), Qureshi et al. (2017), Azimi et al. (2016), Yin et al. (2017), Sun (2021), Peng et al. (2016), Prósper et al. (2019), Song et al. (2020), Shi et al. (2012), Hong and Rioflorido (2019), Zhang et al. (2020a), Ding et al. (2019), Higashiyama et al. (2018)
	NRMSE	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \widehat{y}_i}{y_{max}}\right)^2}$	Abedinia et al. (2020), Abdoos (2016), Li et al. (2022), Dong et al. (2016), Chitsaz et al. (2015), Wang et al. (2021a, b, c, d), Li et al. (2020a, b, c), Li et al. (2015a, 2015b), Zhang et al. (2019a, b), Zhang et al. (2016), Wang et al. (2017a, b, c), Yan et al. (2015), Severiano et al. (2021), Niu et al. (2020), Wang et al. (2021a), Zhang et al.(2019a), Yan et al. (2018), An et al. (2011)
Deterministic	MAE	$\frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y}_i|$	Liu et al. (2021d), Cali et al. (2019), Tian et al. (2022), Liu et al. (2019b, a), Li et al. (2022), Liu et al. (2018a), Huang et al. (2015), Lu et al. (2020), Qureshi et al. (2017), Wu et al. (2020a, b), Yin et al. (2017), Peng et al. (2016), Zhang et al. (2020a), Higashiyama et al. (2018)
	MAPE	$\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \widehat{y}_i|}{y_i}$	Liu et al. (2021d), Woo et al. (2019), Cali et al. (2019), Liu et al. (2019b, a), Li et al. (2022), Wang et al. (2018), Huang et al. (2015), Lu et al. (2020), Osório et al. (2015), Catalão et al. (2010), Li et al. (2020a, b, c), Zu et al. (2018), Jahangir et al. (2020)
	sMAPE	$\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \widehat{y}_i|}{(y_i + \widehat{y}_i)/2}$	Hu et al. (2020), Severiano et al. (2021)
	NMAE	$\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \widehat{y}_i|}{y_{max}}$	Abdoos (2016), Dong et al. (2016, 2017), Chitsaz et al. (2015), Osório et al. (2015), Wang et al. (2021a, b, c, d), Prósper et al. (2019), Li et al. (2020a, b, c), Bessa et al. (2009), An et al. (2011)
	PICP	$\frac{1}{N}\sum_{i=1}^{N}I(L_i \leq y_i \leq U_i)$	Ma et al. (2014), Liu et al. (2022b, a), Liu et al. (2020a, b), Wen et al. (2019), Yuan et al. (2019), Banik et al. (2020), Wang et al. (2019a, b), Mahmoud et al. (2018), Zhang et al. (2016), Wang et al. (2017a, b, c), Wan et al. (2016a, b), Yu et al. (2020), Shi et al. (2017a)
	PINAW	$\frac{1}{N}\sum_{i=1}^{N}\frac{U_i - L_i}{y_{max}}$	Ma et al. (2014), Liu et al. (2022b, a), Zhang et al. (2020a, b, c), Liu et al. (2020a, b), Banik et al. (2020), Mahmoud et al. (2018), Zhang et al. (2016), Wang et al. (2017a, b, c), Yu et al. (2020), Shi et al. (2017a), Liu et al. (2022b, a)
Probabilistic	ACE	PICP-PINC	Zhang et al. (2020a, b, c), Ahmadpour and Farkoush (2020); Mahmoud et al. (2018); Wang et al. (2017a, b, c); Wan et al (2016a, b)
	CWC	$PINAW(1 + \gamma(\mu - PICP)e^{-\eta(PICP-\mu)})$	Liu et al. (2022b, a); Khosravi et al. (2013); Banik et al. (2020); Wang et al. (2019a, b); Wang et al. (2017a, b, c)

Table 3 (continued)

Type	Metric	Formulation	References
	CRPS	$\frac{1}{N}\sum_{i=1}^{N}\int_0^1 F_i(y) - I(y_i^* \le y)dy$	Ma et al. (2014); Zhang et al. (2020a, b, c); Men et al. (2016); Wang et al. (2017a, b, c); Dowell and Pinson (2015)
	Skill score	$\frac{1}{Nr}\sum_{i=1}^{N}\sum_{j=1}^{r}\left(I\left(y_i \le \widehat{q}_i^{\alpha_j}\right) - \alpha_j\right)(y_i - \widehat{q}_i^{\alpha_j})$	Bessa et al. (2012); Wan et al. (2016a, b); Zhang and Wang (2018, 2016)

N is the number of samples in the dataset. y_i and $\widehat{y}_i$ are the actual and predicted wind power generation of the ith sample in the dataset. y_{max} is the maximum possible wind power output (rated wind power for a wind turbine or installation capacity for a wind farm or a region). η and μ are hyperparameters which depend on PINC. $\gamma(x) = \begin{cases} 1(if x > 0) \\ 0(if x \le 0) \end{cases}$; $I(x) = \begin{cases} 1(if x is true) \\ 0(if x is false) \end{cases}$; L_i and U_i are the estimated lower and upper bound of the ith sample in the dataset. $F_i(y)$ is the estimated cumulative distribution function of the ith sample in the dataset. y_i^* is the normalized actual wind power output of the ith sample in the dataset. r is the number of estimated quantile in a sample. α_j is the jth nominal quantile in a sample. $\widehat{q}_i^{\alpha_j}$ is the jth estimated quantile of the ith sample

reflecting the proportion of prediction intervals containing the actual wind power generation. Average coverage error (ACE) metric is another reliability metric, which measures the difference between PICP and the nominal confidence of the prediction interval. prediction interval normalized average width (PINAW) is a typical sharpness metric, measuring the average width of prediction intervals. To comprehensively consider both reliability and sharpness, the coverage width-based criterion (CWC) metric is defined as a multivariate function of PINAW and PICP. CRPS and skill score are typical distribution-based and quantile-based metrics, respectively, which measure the fitness of the predicted distribution by comparing it with the actual wind power distribution.

3 Deep learning based wind power prediction methods

This section presents a comprehensive review of the DL-based methods in recent WPP studies. As shown in Fig. 3, most of reported DL-based WPP (DL-WPP) methods can be categorized into four groups, Scheme 1–Scheme 4, based on the adopted learning scheme, which consists of four information processing components, the signal processing component $f_{sp}(\cdot)$, clustering component $f_c(\cdot)$, feature engineering component $f_{fe}(\cdot)$, and optimization component $f_o(\cdot)$. First, we provide a review of Scheme 1–Scheme 4.

3.1 Learning schemes

Scheme 1 (Ma et al. 2014; Woo et al. 2019; Treiber et al. 2016) presents the simplest end-to-end modeling pipeline based on deep learning. The $f_{fe}(\cdot)$ is developed using deep learning algorithms as models to take the inputs and generate the wind power predictions. Optionally, a meta-learning process $f_o(\cdot)$ can be conducted to optimize hyper-parameters of $f_{fe}(\cdot)$. The power prediction $\widehat{y}$ is obtained using Eqs. (1) and (2).

$$\widehat{y} = f_{fe}^*(x) \tag{1}$$

$$f_{fe}^* = f_o(f_{fe}, x', y') \tag{2}$$

where f_{fe}^* is regarded as the optimal feature engineering component for generating predictions, x', y' are inputs and outputs of the training set.

Scheme 2 extends Scheme 1 via incorporating wind power sequence mode decomposition to generate prediction modes sharing similar patterns (Zu and Song 2018; Han et al. 2019a, b; Dong et al. 2017). The formulation of Scheme 2 is illustrated via Eqs. (3)–(6). The $f_{sp}(\cdot)$ is used to decompose the input into m subseries $s_1, s_2, \ldots, s_m$. The ith subseries is processed using the corresponding feature engineering component $f_{fe_i}(\cdot)$. As in Scheme 1, optionally, all of the feature engineering components can also be optimized via $f_o(\cdot)$ in Scheme 2.

$$\widehat{y} = \sum_{i=1}^m \widehat{y}_i \tag{3}$$

$$\widehat{y}_i = f_{fe_i}^*\left(s_i\right) \forall i = 1, 2, \ldots, m \tag{4}$$
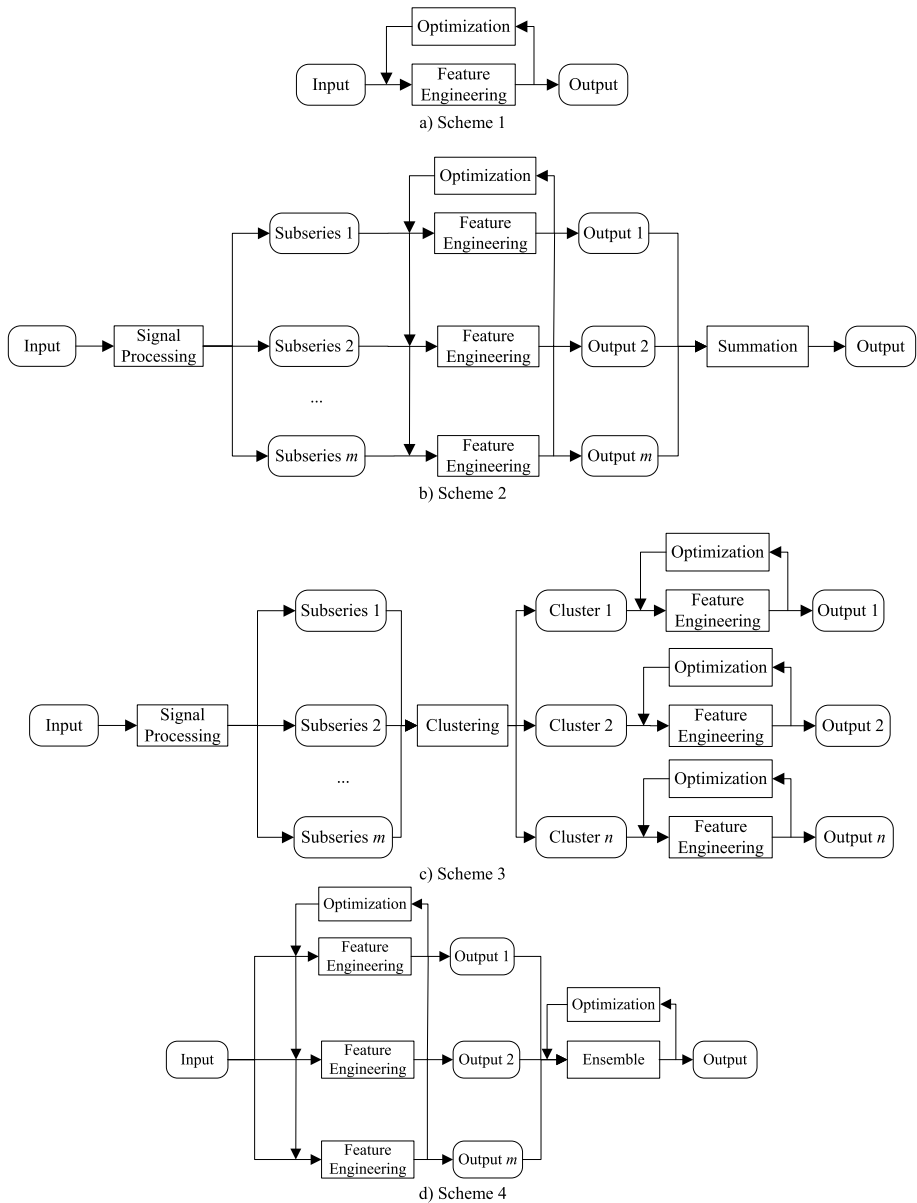
Fig. 3 Four learning schemes for developing DL-WPP models

$$s_1, s_2, \ldots, s_m = f_{sp}(x) \tag{5}$$

$$f_{fe_1}^*, f_{fe_2}^*, \ldots, f_{fe_m}^* = f_o(f_{fe_1}, \ldots, f_{fe_m}, f_{sp}, x', y') \tag{6}$$

Scheme 3 presents a further advancement on top of Scheme 2 by clustering data based on wind power sequence modes to create prediction modeling scenarios (Liu et al. 2021c;

Abedinia et al. 2020; Azimi et al. 2016). The formulation of Scheme 3 is described in Eqs. (7)–(10). The $f_{sp}(\cdot)$ is applied to decompose the input into m subseries $s_1, s_2, \ldots, s_m$. The generated sub-sequences are then used to group data into n clusters $cl_1, cl_2, \ldots, cl_n$ using a clustering component $f_c(\cdot)$. Data of the ith cluster are processed using the corresponding feature engineering component $f_{fe_i}(\cdot)$, which can be optionally optimized by the ith optimization component $f_{o_i}(\cdot)$.

$$\widehat{y} = f^*_{fe_j}(x) \tag{7}$$

$$j = f_c(s_1, s_2, \ldots, s_m) \tag{8}$$

$$s_1, s_2, \ldots, s_m = f_{sp}(x) \tag{9}$$

$$f^*_{fe_i} = f_{o_i}\left(f_{fe_i}, f_{sp}, x', y'\right) \forall i = 1, 2, \ldots, n \tag{10}$$

Scheme 4 develops hybrid models to attain a greater flexibility in the specification. In Scheme 4, multiple models are used to generate predictions. The final prediction is calculated using an optimized ensemble process based on the performance of each model. The formulation of Scheme 4 is described by Eqs. (11)–(14).

$$\widehat{y} = \sum_{i=1}^{m} c_i \widehat{y}_i \tag{11}$$

$$c_1, c_2, \ldots, c_m = f_e(f^*_{fe_1}, f^*_{fe_2}, \ldots, f^*_{fe_m}, x', y') \tag{12}$$

$$\widehat{y}_i = f^*_{fe_i}(x) \forall i = 1, 2, \ldots, m \tag{13}$$

$$f^*_{fe_i} = f_{o_i}\left(f_{fe_i}, x', y'\right) \forall i = 1, 2, \ldots, m \tag{14}$$

Table 4 summarizes recent WPP studies and their adopted learning schemes. Next, we will conduct a comprehensive review on four components, $f_{sp}(\cdot), f_c(\cdot),$ $f_{fe}(\cdot),$ and $f_o(\cdot),$ of WPP methods.

3.2 Signal processing, clustering, feature engineering and optimization components

3.2.1 Signal processing component

In WPP studies, time series inputs are typically treated as a signal. Therefore, advanced signal processing methods may better capture the patterns of the input time series from different perspectives. As shown in Table 5, there are four types of signal processing methods, the frequency-based methods, mode decomposition-based methods, singular spectrum analysis-based methods, and combined methods.

3.2.1.1 Frequency-based methods The frequency-based methods aim to study the raw signal in different frequency domains for improving WPP performance. The Fourier

Table 4 Learning schemes in DL-WPP model development

Schemes	References
Scheme 1	Ma et al. (2014), Woo et al. (2019), Treiber et al. (2016), Liu et al. (2021b, c, d, e, a), Cali et al. (2019), Khosravi et al. (2013), Bessa et al. (2012), Zhang et al. (2020a, b, c), Liu et al. (2018a, b, c), Huang et al. (2015), Ghadi et al. (2014), Li et al. (2015a, b), Men et al. (2016), Qureshi et al. (2017), Prósper et al. (2019), Yuan et al. (2019), Banik et al. (2020), Shi et al. (2012), Jahangir et al. (2020), Amjady et al. (2011), Li et al. (2015a), Jiao et al. (2018), Kisvari et al. (2021), Wang et al. (2019a, b), Hu et al. (2020), Mahmoud et al. (2018), Wang et al. (2017a, b, c), Wang et al. (2017a, b, c), Wan et al. (2016a, b), Yan et al. (2015), Wan et al. (2013a, b), Hu et al. (2018), Severiano et al. (2021), Ju et al. (2019), Wang et al. (2021a, b, c, d), Hong et al. (2019), Yu et al. (2019a, b), Woo et al (2018), Bessa et al. (2009), Zhang et al. (2019a), Yan et al. (2018), Ding et al. (2019), Gallego-Castillo et al. (2016), Zhang et al. (2015), Kou et al. (2014), Wan et al. (2013a), Yu et al. (2022a), Yu et al. (2020), Zhang and Wang (2018), Shi et al. (2017a, b), Dowell and Pinson (2015), Yu et al. (2019a), Liu et al. (2021b, c, d, e, a), Liu et al. (2022b, a)
Scheme 2	Yang et al. (2013), Zu and Song (2018), Han et al. (2019a, b), Dong et al. (2017), Wang et al. (2019a, b), Yan and Wu (2020), Zhang et al. (2016), Wang et al. (2017a, b, c), Sun and Zhao (2020), Han and Tong (2020), Heydari et al. (2021), Ding et al. (2020)
Scheme 3	Liu et al. (2021d), Abedinia et al. (2020), Azimi et al. (2016), Sun (2021), Zhang et al. (2019a, b)
Scheme 4	Bokde et al. (2019), Hossain et al. (2021), Hong and Rioflorido. (2019), Guo et al. (2022), Khazaei et al. (2022), Shahid et al. (2020a), Liu et al. (2020a)

transform (Zhou et al. 2022a, b) and wavelet transform (Catalão et al. 2010; Ahn and Hur 2023; Nascimento et al. 2023; Zhang et al. 2022a, b; Chi and Yang 2023; Aly 2022) are two most widely applied frequency-based methods. The Fourier transform decomposes the input signal into frequency components, which are represented as the sum of sine and cosine of different frequencies. In comparison, the wavelet transform decomposes the input signal into wavelets, which are obtained via shifting and scaling a continuously differentiable wavelet function. To enhance the performance of wavelet transform methods, some variants, such as wavelet packet decomposition (WPD) (Zu and Song 2018; Meng et al. 2016) and empirical wavelet transform (EWT) (Yan et al. 2020; Liu et al. 2018a, b, c), have been proposed.

- Mode decomposition-based methods: The mode decomposition-based methods including the VMD (Abedinia et al 2020) and EMD (Abdoos 2016) aim to decompose the input signal into several intrinsic mode functions and the residual.
- Singular spectrum analysis-based methods: singular spectrum analysis (SSA) (Zhang et al. 2019a) aims to obtain spectrum information on the input signal via singular value decomposition on trajectory matrix of the input time series.
- Secondary methods: Secondary methods (Wu et al. 2020a) combine multiple series decomposition methods to improve the efficiency of WPP.

The signal processing component can also serve filtering out the noises (Saffari et al. 2021; Zhang et al. 2022a; Wang et al. 2023). In (Peng et al. 2020), the wavelet transform is utilized to effectively denoise the original signals, while avoiding distortion and information loss to some extent, The mother wavelet in the wavelet transform can be scaled and time-shifted by the scale factor and the time-shifting factor, producing a series of sub-wavelets to extract the target features under different resolutions.

Table 5 Signal processing methods

Type	Model name	References
Frequency-based	WT	Liu et al. (2019), Yang et al. (2013, 2023), Chitsaz et al. (2015), Osório et al. (2015), Catalão et al. (2010), Azimi et al. (2016), Li et al. (2020a, b, c), Ouyang et al. (2020), Wang et al. (2017a, b, c), An et al. (2011), Liu et al. (2018b), Ahn and Hur (2023), Nascimento et al. (2023), Zhang et al. (2022a, b), Chi and Yang (2023), Shahid et al. (2020a, b), Yu et al. (2022a, b), Dhiman et al. (2022)
	WPD	Zu and Song (2018), Meng et al. (2016)
	EWT	Yan et al. (2020), Liu et al. (2018a, b, c)
	FT	Zhou et al. (2022a, b)
Mode decomposition-based	EMD	Liu et al. (Tian et al. (2022), Abedinia et al. (2020), Li et al. (2022), Lu et al. ((2020), Amjady and Abedinia (2017), Heydari et al. (2021), Meng et al. (2022a, 2022b)
	VMD	Liu et al. (Abdoos (2016), Han et al. (2019a, b), Wang et al. (2019a, b), Zhang et al. (2016), Sun and Zhao (2020), Ding et al. (2020), Meng et al. (2022a, 2022b)
Singular spectrum analysis-based	SSA	Dong et al. (2017), Zhang et al. (2019a)
	VMD+EMD	Wu et al. (2020a)
	VMD+EWT	Wu et al. (2020a)
Secondary	EMD+WPD	Yin et al. (2017)
	WT+VMD	Sun (2021)
	WPD+VMD+SSA	Han and Tong (2020)

3.2.2 Clustering component

Clustering methods aim to extract intrinsic information from wind data by dividing the input data into groups based on their similarity. Each group is then processed by a different feature engineering module. In the WPP literature, K-means (Wang et al. 2018), C-means (Yang et al. 2021a), and expectation maximization (Liu et al. 2018a) are the most widely used clustering algorithms. Time series clustering methods, such as the K-shape algorithm (Liu et al. 2021d), may also improve WPP performance by directly analyzing the similarity among the time series.

3.2.3 Optimization component

Optimization algorithms are applied to attain the best architecture of the deep learning model, which includes determining the number of neurons and layers. The grid search (GS) algorithm (Liu et al. 2021d) is the simplest and most widely used algorithm. In the GS algorithm, a set of candidate architectures are defined based on domain knowledge or preliminary trials, and the best architecture is selected based on its performance on the validation set.

However, the performance of the GS algorithm is highly dependent on expert knowledge, as the number of candidates is usually limited to reduce computational cost. To explore a larger solution space, meta-heuristic algorithms, such as the GA (Liu et al. 2021a), PSO (Ma et al. 2014), shark smell optimization (SSO) (Abedinia et al. 2020), atomic search (AS) (Li et al. 2020a, b, c), CS (Li et al. 2021), clonal selection algorithm (CSA) (Chitsaz et al. 2015), crisscross optimization (CSO) (Yin et al. 2017), dragonfly algorithm (DA) (Shi et al. 2017a, b), sparrow search (SS) (Abdoos 2016), and GWO (Lu et al. 2020), have been proposed. Table 6 summarizes the optimization algorithms used in WPP studies.

3.2.4 Feature engineering component

One significant advantage of deep learning is the ability of automating and adaptively learning a low-dimensional embedding, which is a set of latent features representing the raw high-dimensional input. Many deep learning methods can simultaneously serve the feature extraction and the prediction in WPP modeling. The feature extraction module $f_{fe}(\cdot)$ aims to extract informative latent features from the input. The prediction module $f_p(\cdot)$ aims to model the mapping from extracted features to the wind power output. Therefore, the entire process including the feature engineering to the prediction can be expressed by Eqs. (15) and (16),

$$\hat{y} = f_{fe}(z) = f_p(z_f) \tag{15}$$

$$z_f = f_f(z) \tag{16}$$

where z is the input of the feature engineering module. It can be either the original input x or the subseries s.

Next, we will review a total of 28 deep learning models grouped into 8 types as listed in Table 7. Among these models, fully connected neural networks can be used for both the

Table 6 Optimization algorithms

| Algorithm	References
GS	Liu et al. (2021d), Zhang et al. (2019a, b), Kisvari et al. (2021), Wang et al. (2021a, b, c, d)
GA	Liu et al. (2021a), Huang et al. (2015), Viet et al. (2020), Saroha et al. (2014)
PSO	Ma et al. (2014), Liu et al. (2022b, a), Osório et al. (2015), Catalão et al. (2010), Yuan et al. (2019), Amjady et al. (2011), Jiao et al. (2018)
SSO	Abedinia et al. (2020), Wang et al. (2021b), Amjady et al. (2017), Li et al. (2020a, b, c), Abedinia and Amjady (2015)
CS	Li et al. (2021), Wu et al. (2020a, b)
CSA	Chitsaz et al. (2015)
CSO	Yin et al. (2017, 2021), Meng et al. (2022a, b)
DA	Shi et al. (2017a, b), Li et al. (2020b), Zhou et al. (2022a), Liu et al. (2021b, c, d, e, a)
AS	Li et al. (2020a, b, c)
SS	Li et al. (2022), Song et al. (2022), An et al. (2021), Wu and Wang (2021), Yu and Luo (2022)
GWO	Lu et al. (2020), Han and Tong (2020), Heydari et al. (2021), Ding et al. (2020)

Table 7 Deep learning models

Type	Model name	Feature extraction	Prediction	References
Fully connected	DNN	✓	✓	Methaprayoon et al. (2007), Abedinia et al. (2020), Huang et al. (2015), Ghadi et al. (2014), Ronay et al. (2013)
	RBM	✓	✓	Peng et al. (2016)
	DBN	✓	✓	Wang et al. (2018), Zhang et al. (2019a)
	Fuzzy NN	✓	✓	Khodayar et al. (2022), Bilal et al. (2023)
	ELM	✓	✓	Abdoos (2016), Wu et al. (2020a, b), Yin et al. (2017), Mahmoud et al. (2018), Ding et al. (2020)
	AE	✓		Yu et al. (2022a), Liu et al. (2020a, b), Liu and Zhang (2021)
Autoencoders	SAE	✓		Wang et al. (2021a, b, c, d), Jiao et al. (2018)
	SSAE	✓		Liu and Chen (2019), Deng et al. (2020)
	RAE	✓		Khodayar et al. (2017)
	QR		✓	Wen et al. (2019), Yu et al. (2020), Zhang and Wang (2018), He et al. (2021)
Probabilistic output	LUBE		✓	Khosravi and Nahavandi (2013), Banik et al. (2020), Wang et al. (2019a, b), Yan et al. (2015), Shi et al. (2017a)
	KDE		✓	Bessa et al. (2012), Yuan et al. (2019), Zhang et al. (2015), Zhang and Wang (2016)
	MDN		✓	Zhang et al. (2020a, b, c), Men et al. (2016), Wang et al. (2017a, b, c)
	1DCNN	✓		Zhang et al. (2020a), Harbola and Coors (2019)
Convolutional	2DCNN	✓		Yu et al. (2019a), Yin et al. (2019)
	3DCNN	✓		Mehrkanoon (2019), Higashiyama et al. (2018)
	GRU	✓		Tian et al. (2022), Liu et al. (2022b, a), Kisvari et al. (2021), Wang et al. (2019a, b), Niu et al. (2020)
Recurrent	LSTM	✓		Liu et al. (2019b, a), Zhang et al. (2019a, b), Yu et al. (2020), Yu et al. (2019a, b), Shahid et al. (2021)
	Bi-LSTM	✓		Jahangir et al. (2020), Ko et al. (2020), Zhen et al. (2020), Jahangir et al. (2020)
Convolutional and recurrent	ConvLSTM	✓		Woo et al. (2019, 2018), Sun and Zhao (2020)
Graph-based	GNN	✓		Park and Park (2019), Bentsen et al. (2022)
	GCN	✓		Xu et al. (2022a, b)
	Transformer	✓		Tian et al. (2022), Niu et al. (2020), Nascimento et al. (2023)
	Informer	✓		Zhou et al. (2021), Zheng and Zhang (2023), Cao et al. (2023)

Table 7 (continued)

Type	Model name	Feature extraction	Prediction	References
Self attention-based	Autoformer	√		Wu et al. (2021)
	Pyformer	√		Liu et al. (2021b)
	Fedformer	√		Zhou et al. (2022a, b), Deng et al. (2022)
	DLinear	√	√	Zeng et al. (2023)
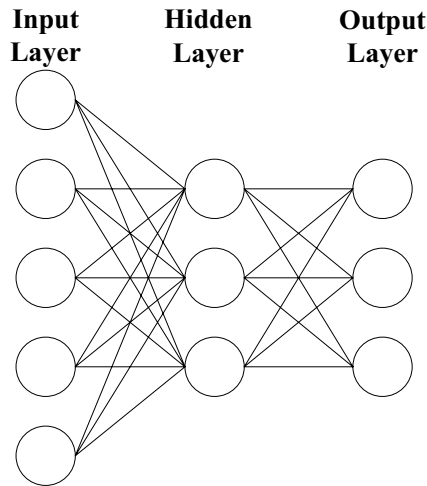
Fig.4 Fully connected neural networks

feature extraction and prediction modules. Probabilistic output models can only be used for the prediction module. Deep learning models including autoencoders, convolutional networks, recurrent networks, etc., are usually regarded as the feature extraction module, which learn a low-dimensional representation from the input to be fed into an additional regression layer for generating predictions.

3.2.5 Fully connected neural networks

Figure 4 illustrates the structure of fully connected neural networks, which consist of three components: an input layer, one or more hidden layers, and an output layer. The input z or z_f is initially flattened to a one-dimensional vector z_1, which is fed into the neural network. It is processed by n neural layers including $n-1$ hidden layers and one output layer. Finally, results are generated from the output layer. The last hidden layer provides the extracted features.

Let z_i denote the output of the ith layer, which is also the input of $(i + 1)^{th}$ layer, the formulation of fully connected neural network is described in Eqs. (17)–(19),

$$z_{i+1} = f_a\left(W_i z_i + b_i\right) \forall i = 1, 2, \ldots, n - 1 \tag{17}$$

$$z_f = f_f(z) = z_{n-1} \tag{18}$$

$$f_{fe}(z) = f_p\left(z_f\right) = z_n \tag{19}$$

where W_i is a learnable matrix, b_i is a learnable vector, and f_a is an activation function.

In WPP studies, the output layer of FNN is designed to provide an accurate prediction of the wind power $\hat{y}$. Therefore, the parameters W_i and b_i are obtained by minimizing the error between $\hat{y}$ and actual output y. Let $z\prime$ and $y\prime$ be the input and output of the training set respectively, and $\hat{y}\prime$ as the prediction. The parameters W_i and b_i of FNN can be formulated as Eqs. (20) and (21),

$$\hat{y}\prime = f_{FNN}(W_1, b_1, \ldots, W_n, b_n; z\prime) \tag{20}$$

$$W_1, b_1, \ldots, W_n, b_n = \text{argmin} L(\hat{y}\prime, y\prime) \tag{21}$$

where f_{FNN} is the FNN model serving as either the feature extraction or prediction module, $L(\hat{y}\prime, y\prime)$ is a pre-defined loss function. Let N_t be the number of the instances in the training set. Common loss functions used are provided as follows, where MSE is the most frequently applied one.

- MSE: $L(\hat{y}\prime, y\prime) = \frac{1}{N_t} \sum_{i=1}^{N_t} (\hat{y}_i\prime - y_i\prime)^2$

- MAE: $L(\hat{y}\prime, y\prime) = \frac{1}{N_t} \sum_{i=1}^{N_t} |\hat{y}_i\prime - y_i\prime|$

- Huber Loss Function: $L(\hat{y}\prime, y\prime) = \frac{1}{N_t} \sum_{i=1}^{N_t} \begin{cases} \frac{1}{2}(\hat{y}_i\prime - y_i\prime)^2 (if |\hat{y}_i\prime - y_i\prime| > 1) \\ |\hat{y}_i\prime - y_i\prime| - \frac{1}{2}(Otherwise) \end{cases}$

Apart from the generic DNN (Methaprayoon et al. 2007; Abedinia et al. 2020), some variants of the FNN, such as restricted Boltzmann machine (RBM) (Peng et al. 2016), deep belief network (DBN) (Wang et al. 2018; Zhang et al. 2019a) and extreme learning machine (ELM) (Yin et al. 2017; Ding et al. 2020), are proposed to enhance the prediction accuracy. The fuzzy neural network (Khodayar et al. 2022; Bilal et al. 2023; Qiao et al. 2022; Xu et al. 2022a; Li et al. 2020a), which utilizes fuzzy influence techniques to determine the values of the neurons, has also received discussions. The Fuzzy NN is well-known for its effectiveness in tackling the uncertainty in the SCADA data, making them well-suited for WPP with incomplete information or ambiguous data.

3.2.5.1 Autoencoders The generic autoencoder (AE) has the same structure as the FNN shown in Fig. 4. AEs aim to encode the input z into a latent representation z_f using the information of z itself. Hence, different from FNN, the output layer of AE targets to output the reconstruction of the input z, and the values of the hidden nodes are regarded as the feature z_f. The parameters θ_{AE} of AE are inferred via minimizing the reconstruction loss between the input z and its reconstruction $\hat{z}$. The formulation of AE is provided in Eqs. (22) and (23),

$$\hat{z}\prime = W f_{AE}(\theta_{AE}; z\prime) + b \tag{22}$$

$$W, b, \theta_{AE} = \text{argmin} L(\hat{z}\prime, z\prime) \tag{23}$$

where f_{AE} is the AE model, which serves as the feature extraction module.

Recently, a few variants of the AE have been proposed to improve the performance of WPPs:

- Stacked AE (SAE): In study (Wang et al. 2021a, b, c, d), the AE model is improved by stacking multiple AEs together. The first AE takes the original input z and outputs the latent feature z_1. The subsequent $i^{th}(i > 1)$ AE takes the output z_{i-1} of $(i-1)^{th}$ AE as the input and produces the latent feature z_i.
- Sparse SAE: In study (Yin et al. 2021), a Sparse SAE is proposed to learn more concise features by introducing a sparse penalty term into the loss function of AE, is proposed. Let $\bar{z}_i$ denote the average value of z_i, and ρ denote a sparse parameter,

which is set to a small number near 0. The sparse penalty term is defined as the Kullback–Leibler (KL) divergence of ρ and z_i, $KL\left(\rho||\bar{z}_i\right) = \rho\log\left(\frac{\rho}{\bar{z}_i}\right) + (1-\rho)\log(\frac{1-\rho}{1-\bar{z}_i})$.

- In study (Li et al. 2020a, b, c), rough neurons in the SAE are introduced to address the uncertainty of the wind. Different from generic AEs, the output z_i is determined using the rough set theory.

3.2.5.2 Probabilistic output models In the literature, four probabilistic output models for generating probabilistic outputs are observed: QR, LUBE, KDE, and MDN. The descriptions of these models are provided as follows.

- QR: The QR aims to directly estimate the quantile with neural networks. In QR, the best parameters of NNs θ_{QR} can be obtained by minimizing the negative skill score.
- LUBE: The LUBE model aims to directly estimate the quantile with neural networks, which is usually trained by minimizing the CWC metric.
- KDE: KDE methods attempt to estimate the distribution of wind power, which is modeled by a probability density function.
- MDN: The MDN method mixes multiple PDFs to allow sufficient flexibility in modeling the wind power distribution. Usually, the Gaussian mixture model (GMM) is adopted to model the PDF because of its simplicity and convenience for sampling and computing the distribution. However, the GMM may lead to density leakage problems in the mixture model. Recent studies (Zhang et al. 2020a, b, c) addressed this issue by replacing the GMM with the beta kernel. The KDE and MDN models are usually optimized by maximizing the likelihood of the distribution. One recent study (Yang et al. 2021a, b) further improved the training process by using a Wasserstein distance-based adversarial learning algorithm.

3.2.5.3 Convolutional neural networks CNNs (Liu et al. 2019b, a; He et al. 2020) are known for their shift invariance, meaning they can detect objects equally well regardless of their locations in the input. CNNs consist of convolution layers and pooling layers. In each convolution layer, a learnable kernel g is used to learn local features of the data. Typically, a pooling layer is used immediately after each convolution layer to aggregate the local information of the output of the convolution layer and select the most concise and efficient features. The general formulation of CNN is provided in Eqs. (24)–(26),

$$f_f(z) = z_{n+1} \tag{24}$$

$$z_{i+1} = Pool\left(g_i * z_i\right) \forall i = 1, 2, \ldots, n \tag{25}$$

$$z_1 = z \tag{26}$$

where n is the number of convolution and pooling layers, g_i is the kernel of the ith convolution layer, $*$ denotes the convolution operator, and $Pool(\cdot)$ is a pooling function. Depending on the type of CNN used (1DCNN, 2DCNN, or 3DCNN), different types of convolution operators and pooling functions are utilized in WPP studies.
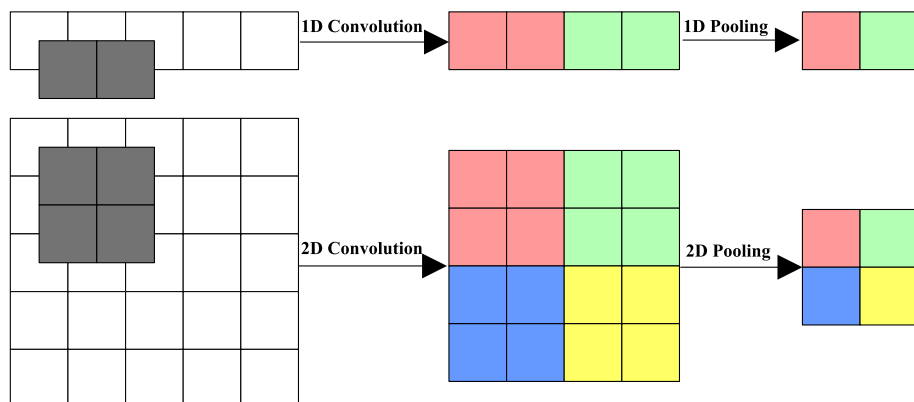
Fig. 5 1D and 2D convolutional neural networks

3.2.5.4 1-dimensional CNN (1DCNN) As shown in the top part of Fig. 5, 1DCNN takes a one-dimensional vector as input. If the original input is multidimensional, it should be flattened to a vector before being fed to the 1DCNN. In 1DCNN, 1-dimensional convolution operator and pooling functions are utilized. The formulation of 1D convolution operator is provided in Eq. (27), where k is the size of the kernel. Common pooling functions including the max pooling and average pooling are described in Eqs. (28)–(29) and Eqs. (30)–(31), respectively, where l describes the size of pooling kernel.

$$(g * z)_p = \sum_{j=p-\frac{k}{2}}^{p+\frac{k}{2}} g_{j-p-\frac{k}{2}} z_j \tag{27}$$

$$Maxpool(z) = \widetilde{z} \tag{28}$$

$$\widetilde{z}_p = \max_{j=p\times l, p\times l+1, \ldots, p\times l+l-1} z_j \tag{29}$$

$$Avgpool(z) = \bar{z} \tag{30}$$

$$\bar{z}_p = \frac{1}{l} \sum_{j=p\times l}^{p\times l+l-1} z_j \tag{31}$$

3.2.5.5 2-dimensional CNN (2DCNN) As shown in the bottom part of Fig. 5, the 2DCNN expands 1DCNN to a 2-dimensional grid with considering the input of a 2-dimensional matrix form. If the original input is three dimensional, $z \in \mathbb{R}^{T \times P \times N_{feat}}$, where T is the length of time series, P is the number of wind turbines, and N_{feat} is the number of features in each time step and wind turbine of input z, a common organization is to regard the time steps as different input channels, and the features of different wind turbines are organized as a matrix.

3.2.5.6 3-dimensional CNN (3DCNN) The 3DCNN is typically used to extract spatial–temporal features from the 3-dimensional input $z \in \mathbb{R}^{T \times P \times N_{feat}}$. The definition of 3DCNN is similar to that of 1DCNN and that of 2DCNN.

Fig. 6 Recurrent neural networks
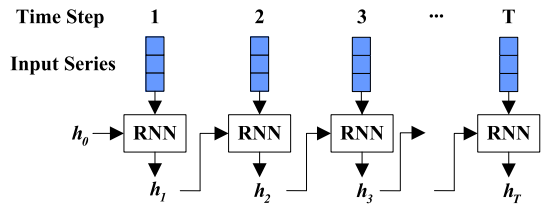
To obtain the best parameter estimation of the kernels g, the extracted features $f_{CNN}(z) = z_{n+1}$ are usually transformed into the prediction via an FNN. The loss $L(\cdot, \cdot)$ between the prediction and actual wind power in the training set is applied to optimize g, as shown in Eqs. (32) and (33).

$$\hat{y}' = f_{FNN}(f_{CNN}(z)) \tag{32}$$

$$g_1, g_2, \ldots, g_n = \mathrm{argmax} L(\hat{y}', y) \tag{33}$$

3.2.5.7 Recurrent neural networks As shown in Fig. 6, RNNs are efficient models for processing 2-dimensional multi-variate time series data $z \in \mathbb{R}^{T \times N_{feat}}$ across T time steps. In each step, the RNN takes z_t, the tth time step of z, and the last hidden state h_{t-1} as inputs and outputs the current hidden state h_t. The general formulation of RNN is provided in Eqs. (34) and (35).

$$f_f(z) = [h_1, h_2, \ldots, h_T] \tag{34}$$

$$h_t = RNN(h_{t-1}, z_t) \forall i = 1, 2, \ldots, T \tag{35}$$

where h_0 is pre-defined vector usually set to a vector of zeros. As the vanilla RNN encounters the gradient vanishing and explosion issues, variants of RNN, such as the GRU and LSTM, as well as an advanced development on top of RNN, such as the attention mechanism, BiLSTM, and ConvLSTM, more frequently appear in WPP studies.

- GRU: In each time step t, the GRU (Tian et al. 2022) utilizes a reset gate r_t and an update gate u_t to control the hidden state h_t. The reset gate r_t determines whether the last state h_{t-1} is considered or reset to a new state, and the update gate u_t determines whether h_t is updated by the new input z_t or remains the old value h_{t-1}
- LSTM: Similar to GRU, the LSTM (Neshat et al. 2021) uses an input gate i_t, a forget gate f_t and an output gate o_t to control the input, forget and output process of the hidden state h_t.
- BiLSTM: The BiLSTM (Jahangir et al. 2020; Huang et al. 2022) is an efficient variant of LSTM. Different from the generic LSTM, which only considers the past information from h_{t-1}, the BiLSTM takes into consideration of both past and future information in each time step.
- Attention Mechanism: The gradient of the RNN models accumulates across all time steps, potentially leading to gradient vanishing and gradient exploding issues when the length of the input time series is long. To alleviate such issues, the attention

Fig. 7 Illustration of ConvLSTM
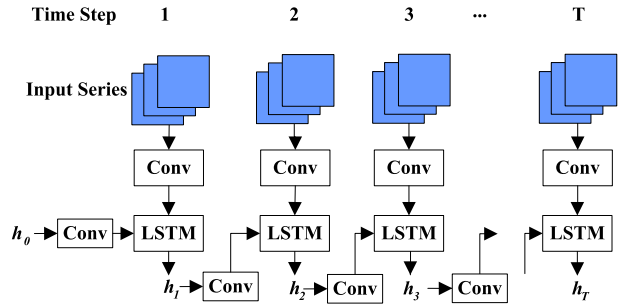
mechanism (Yang and Zhang 2021; Ren et al. 2022; Zhang et al. 2023) is introduced with considering all of the hidden states $h_1, h_2, \ldots, h_{t-1}$ in the past time steps.

- ConvLSTM: As shown in Fig. 7, the ConvLSTM (Wilms et al. 2021) is a model to extract spatial–temporal patterns by leveraging convolution operations to modeling spatial correlations and the LSTM for learning temporal patterns. The ConvLSTM takes a four-dimensional input $z \in \mathbb{R}^{W \times H \times T \times N_{feat}}$, which is placed on a $W \times H$ grid according to the distribution of wind turbines. The formulation of ConvLSTM is provided in Eqs. (36)−(42).

$$f_{ConvLSTM}(z) = [h_1, h_2, \ldots, h_T] \tag{36}$$

$$i_t = sigmoid\big(W_{zi} * z_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i\big) \forall t = 1, 2, \ldots, T \tag{37}$$

$$f_t = sigmoid\big(W_{zf} * z_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f\big) \forall t = 1, 2, \ldots, T \tag{38}$$

$$g_t = tanh(W_{zg} * z_t + W_{hg} * h_{t-1} + b_g) \forall t = 1, 2, \ldots, T \tag{39}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t \forall t = 1, 2, \ldots, T \tag{40}$$

$$o_t = sigmoid\big(W_{zo} * z_t + W_{ho} * h_{t-1} + W_{co} \circ c_{t-1} + b_o\big) \forall t = 1, 2, \ldots, T \tag{41}$$

$$h_t = o_t tanh(c_t) \forall t = 1, 2, \ldots, T \tag{42}$$

where $f_{ConvLSTM}(\cdot)$ is the ConvLSTM model, $W_{zi}, W_{hi}, W_{zf}, W_{hf}, W_{zg}, W_{hg}, W_{zo}, W_{ho}$ are learnable kernels, and $W_{ci}, W_{cf}, W_{co}, b_i, b_f, b_g, b_o$ are learnable vectors.

However, in WPP tasks which faces the sparse distribution of wind turbines, the large portion of blank values may degrade the efficiency and performance of the model. In such cases, graph-based models are better choices for extracting spatial features.

3.2.5.8 Graph-based neural networks Graph based neural networks leverage the graph that represents the geographical correlation of the wind turbines to extract the spatial features. As shown in Fig. 8, GNN is a generic graph-based neural network. In GNN, the input z_i in each wind turbine i is transformed to the feature space via a DNN. The feature is then concatenated

Fig. 8 Illustration of GNN
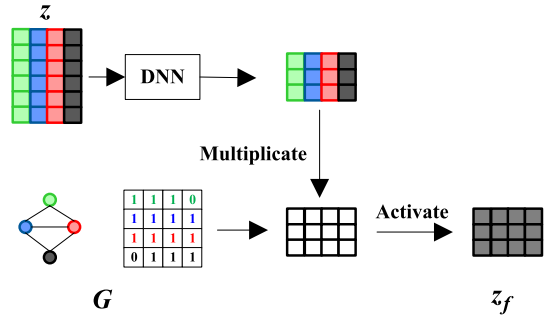
to form a matrix M_z, which is multiplied by the graph matrix G, and activated by an activation function f_a. The general formulation of GNN is provided in Eqs. (43)–(45).

$$f_{GNN}(z) = z_f \tag{43}$$

$$z_f = M_z G \tag{44}$$

$$M_z = [f_{DNN}(z_1), f_{DNN}(z_2), \dots, f_{DNN}(z_P)] \tag{45}$$

where $f_{GNN}(\cdot)$ represents the GNN model.

The structure of GNN can be expanded to extract spatial–temporal features of the input z. There are two types of improvements to achieve this goal.

The first improvement is replacing the DNN with RNN. By this mean, the temporal features of each wind turbine are first extracted by the RNN, and the spatial–temporal features are finally provided by the GNN.

The second improvement is to reorganize the input z as a time series. In each time step, the GNN is used to extract the spatial feature. These features are finally concatenated together and processed by an RNN to obtain the spatial–temporal features.

Graph convolutional neural network (GCN) is an efficient variant of GNN. Unlike the original GNN, GCN utilizes a graph Laplacian L instead of G, which is formulated as Eq. (46).

$$L = I_P - D^{-\frac{1}{2}} G D^{-\frac{1}{2}} \tag{46}$$

where I_P is an identity matrix with order P, the D is a diagonal matrix formulated as Eq. (47).

$$D_{ij} = \begin{cases} 0 (i \neq j) \\ \sum_{j=1}^{P} G_{ij} (i = j) \end{cases} \tag{47}$$

3.2.5.9 Self-attention-based neural networks Self-attention-based neural networks are constructed based on the attention mechanism, which is shown in Fig. 9. Transformer (Tian et al. 2022) is a generic self-attention-based neural network that has been designed to extract temporal features of the input. In Transformer, the input $z \in \mathbb{R}^{T \times N\prime}$ is regarded as a time series with T time steps. Here, $N\prime = N_{feat}$, if only data of the target turbine are considered, and having

Fig. 9 Illustration of self attention mechanism
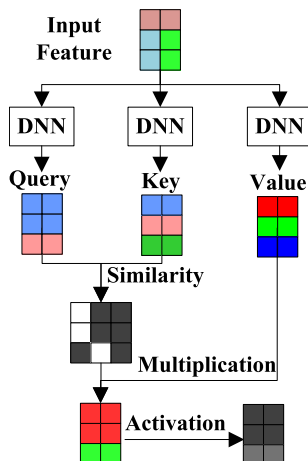
$N' = P \times N_{feat}$, if data from neighboring P wind turbines are considered. To learn the relationship between time steps, the input z_t of each time step t is first transformed to the query, key and value vectors $Q_t, K_t, V_t \in \mathbb{R}^{T \times d_h}$, where d_h is the number of hidden dimensions, via three different DNNs respectively. The vectors $Q_t, K_t, V_t (t = 1, 2, \dots, T)$ are concatenated to form multivariate time series Q, K, V. The attention A of the input is evaluated as the similarity between Q and K. Finally, the features are produced by the activation value of multiplication of the A and V. The formulation of Transformer is provided in Eqs. (48)–(51).

$$f_{Trans}(z) = z_f = f_a(AV) \tag{48}$$

$$A = Sim(Q, K) \tag{49}$$

$$Q, K, V = \left[Q_1, Q_2, \dots, Q_T\right], \left[K_1, K_2, \dots, K_T\right], \left[V_1, V_2, \dots, V_T\right] \tag{50}$$

$$Q_t, K_t, V_t = f_{DNN,Q}(z_t), f_{DNN,K}(z_t), f_{DNN,V}(z_t) \forall t = 1, 2, \dots, T \tag{51}$$

where $f_{Trans}(z)$ is the Transformer model, $f_{DNN,Q}$, $f_{DNN,K}$, $f_{DNN,V}$ are three DNN models. $Sim(\cdot, \cdot)$ is a similarity function.

Latest WPP studies have explored the performance of Transformer-based models, such as the Informer (Zhou et al. 2021; Nascimento et al. 2023; Huang et al. 2022), Autoformer (Wu et al. 2021), Pyraformer (Liu et al. 2021b), and Fedformer (Zhou et al. 2022a, b; Deng et al. 2022). Studies observed that the transformer-based models could improve WPP efficiency by sampling a set of informative queries and keys with incorporating signal processing methods. Additionally, the DLinear model (Zeng et al. 2023) reported the performance improvement of WPPs through a simpler attention scheme. In DLinear, the attention mechanism was defined as a linear transformation from the historical time steps to the future time steps.

3.3 Prediction performance analyses

In this section, we first conduct an analysis based on results of existing studies to consolidate views of the WPP improvement brought by DL-WPP methods comparing with

Table 8 RMSE improvements of deep learning methods

References	Prediction level	Installed capacity	Prediction horizon	Deep learning method		Best performed conventional model		RMSE improvement (%)
				Models used	RMSE	Models	RMSE	
Huang et al. (2022)	Farm	2.5 MW	10 m	CNN, Bi-LSTM	0.34 MW	SVR	2.76 MW	87.68
Zhang et al. (2020a, b, c)	Farm	0.73 MW	24 h	LSTM	129 KW	RF	134 KW	3.7
Zhao et al. (2023)	Farm	2 MW	1–24 h	CNN, GRU	1.58 MW	NARX	3.10 MW	49.0
Yu et al. (2022b)	Farm	0.8 MW	30 m	GNN, AE, ConvLSTM	1.21 MW	SVR	1.61 MW	24.8
Tian et al. (2022)	Farm	2 MW	10 m	Transformer, GRU	112 KW	EN	208 KW	46.2
Liu et al. (2021d)	Turbine	2 MW	10–60 m	CNN, GRU	308 KW	LR	319 KW	3.4
Meng et al. (2022a, b)	Farm	2 MW	10 m	ELM, AE	1.70 MW	PM	13.73 MW	87.61
Liu and Zhang (2022a, b)	Turbine	2 MW	10–60 m	Informer	193 KW	MARS	198 KW	2.5
Liu et al. (2023)	Turbine	2 MW	10–60 m	GCN, LSTM	336 KW	LASSO	345 KW	2.6
Huang and Jiang (2022)	Farm	0.73 MW	10 m	Informer	307 KW	DNN	871 KW	64.8

Table 9 Dataset Descriptions

Dataset	Collected period	Rated wind power (MW)	No. of SCADA attributes	No. of WTs	Sampling interval (min)
Dataset 1	Sep. 2019–Mar. 2020	2	10	33	10
Dataset 2	Sep. 2019–Mar. 2020	2	10	25	10
Dataset 3	Jan. 2016–Jun. 2017	2	10	11	10

traditional ones as reported in Table 8. Next, to horizontally compare the effectiveness of developed DL models in WPP tasks, we conduct a computational experiment replicating famous and recent DL models based on our collected wind farm SCADA datasets.

The performance improvement of the developed DL-WPP method over the best-performed traditional machine learning WPP method is analyzed article-wise based on results reported of considered articles. Results of such analytics are reported in Table 8. In each article, the DL configuration of the reported DL-WPP method is analyzed and the best performed classical machine learning based WPP method is identified. As the Root Mean Square Error (RMSE) is the only metric that simultaneously utilized in all considered articles, RMSE values of the developed DL-WPP method and the best-performed traditional WPP method are retrieved and the RMSE improvement percentage is computed for each article. Analytical results in Table 8 revealed that existing studies unanimously observed the improvement generated by DL models and RMSE improvement could range from 2.5% to 87.68% across studies. Such significant variation can be caused by multiple reasons. First, the WPP task setups across studies differ in terms of considered prediction horizons and targets. Secondly, results of reported WPP methods were examined based on datasets collected by different research groups, which might possess completely different wind patterns. Moreover, the coverage of benchmarking models applied is different in studies reported in Table 8. Although we can conclude the performance improvement based on each work in Table 8, it is difficult to derive a fair conclusion via a horizontal comparison of DL-WPP methods developed in existing studies due to previously mentioned three reasons.

To discover more meaningful insights, in this work, we would like to further verify the effectiveness of recent DL-WPP method development via reproducing and comparing latest DL-WPP models based on our SCADA datasets collected from three commercial wind farms, which cover a larger population of wind turbines and more recent samples. Specific descriptions of three datasets are offered in Table 9.

Since deterministic short-term WPP is most frequently studied, we consider the wind turbine power output prediction with horizons ranging from 10 to 60 min in this computational experiment. Meanwhile, processed data are divided into the training set, validation set, and test set respectively with a 0.6:0.2:0.2 split ratio. The commonly considered RMSE metric is employed to evaluate the WPP performance of developed models. An extended set of promising DL models including DLinear, Informer, Transformer, CNN, LSTM, GRU and DNN are considered in this further computational experiment based on recent studies (Zhou et al. 2022a, b; Zhou et al. 2021; Zeng et al.2023).

Results of our computational experiments are reported in Table 10. It is observable that the DLinear model significantly outperforms other candidates on most datasets. Meanwhile, we also identify that, in more than 69% test cases, DLinear obtains 2.0% improvements compared with the best performed one from other considered DL models. The performance of Informer, Transformer, and GRU models are also promising, which

Table 10 RMSE (KW) of deep learning models for short-term WPP

Dataset	Model	Prediction Horizon (min)					
		10	20	30	40	50	60
Dataset 1	DLinear	**122.10**	**172.55**	**203.50**	**226.68**	**245.22**	**260.08**
	Informer	131.08	178.63	205.31	230.15	249.21	263.84
	Transformer	135.52	182.53	206.10	230.08	247.35	262.75
	GRU	141.69	181.74	208.65	231.07	245.31	261.07
	LSTM	164.79	196.61	221.54	241.15	257.22	270.86
	CNN	194.06	215.48	233.48	248.45	260.69	271.80
	DNN	187.72	213.04	233.05	249.60	263.61	275.99
Dataset 2	DLinear	**171.22**	**256.02**	**308.63**	**346.94**	**376.83**	**401.07**
	Informer	191.29	264.18	313.56	353.21	379.57	405.82
	Transformer	212.17	277.27	321.45	355.12	385.15	407.16
	GRU	215.20	276.00	319.56	354.77	382.94	406.91
	LSTM	242.82	292.77	333.99	367.92	397.43	418.97
	CNN	304.10	333.88	372.06	395.97	425.88	443.58
	DNN	273.22	308.84	359.45	388.66	410.67	429.44
Dataset 3	DLinear	**155.69**	**204.78**	**233.55**	**255.65**	**274.68**	292.10
	Informer	158.63	209.43	236.78	258.31	275.78	**291.57**
	Transformer	161.35	212.58	237.51	256.27	277.37	292.88
	GRU	157.23	210.89	240.93	263.49	283.29	300.56
	LSTM	165.42	211.28	239.23	260.56	280.23	296.62
	CNN	186.17	222.78	247.03	266.16	283.46	299.29
	DNN	177.10	214.38	239.51	259.62	277.65	293.79

are only 2.0%, 3.1% and 3.9% worse in average than that of the DLinear, respectively. Comparing the performance on different datasets, the DLinear model is 7.6%, 9.4% and 2.8% better than the candidates on average. Thus, the DLinear model may be a better option for achieving the state-of-the-art performance in the considered short-term WPP task by comparing with other DL models.

4 Promising trends for future deep learning-based WPP studies

In future WPP studies, DL techniques will take an increasingly important role with the rapid development and advancement. Although existing WPP studies have achieved promising performance in WPP, the WPP can be further improved by addressing the following limitations and issues.

- Inputs: Although spatial–temporal correlations are already considered in some studies, most of them are based on the self-correlation in the data and the locations of the wind farm sites. Some important factors, such as the wake effect and topological information, can be considered to improve the performance. In (Park and Park 2019), a Physics-informed graph network (PIGN) has been developed to attack such issues and promising results in the wind power estimation have been reported. More advanced mechanisms can be developed to bring more values into the WPP tasks.

- Features: Existing deep learning-based WPP studies engineer the WPP features using well-designed models, which highly depend on the training data. More robust and reliable WPP features are required to improve the WPP performance.
- Models: With rapid development of the deep learning techniques, especially in the natural language processing and computer vision, the deep models evolve at a fast pace. Apart from the advanced deep models in other research communities, we can also design advanced models specific to WPP tasks based on our domain knowledge.
- Evaluation: RMSE and MAE are the most widely used metrics in recent studies. However, such metrics only present the error in average, which are inadequate in real grid applications. More specific metrics are required to evaluate the WPP performance specific to certain downstream grid operations.

Next, we present four promising research trends for applying DL techniques in WPP.

4.1 More advanced design of WPP input organization

In the WPP input organization, we observe three main trends, the incorporation of geographical information, the privacy-preserving data sharing paradigm, and the usage of multiple resolution data.

Geographical information is a crucial factor in physics-based methods (Landberg 1999). Current WPP studies (Khodayar and Wang 2018) have utilized the location of the wind turbines to depict their distribution and improve WPP performances. Other geographical information, such as topography and roughness, could also be utilized to better represent the actual geographical situations of wind turbines and wind farms.

Privacy-preserving data sharing is another important aspect in WPP studies. When data are distributed across different wind farms, sharing them may raise safety and privacy concerns in situations where WPP methods and service providers involve external entities. Moreover, wind power units from different power plants located in different regions or countries may not always be accessible due to various imposed regulations. To address these issues, (Liu and Zhang 2022a, b) proposed a bi-party data-driven modeling framework to learn the spatial–temporal features in different wind farms while preserving privacy. More methods with advanced privacy-preserving schemes for various WPP modeling tasks could be developed to further enhance the performance of WPP.

Most existing WPP studies have considered the sampling interval of SCADA data to be the same as the desirable WPP resolution. However, a recent study (Liu and Zhang 2022a) has discovered that the usage of multiple sampling resolution data may significantly improve WPP performance. It is also interesting to investigate whether multiple resolution data in spatial dimension can enhance WPP. In other words, it may be possible to utilize turbine-level SCADA data to predict farm-level wind power output.

In summary, the WPP input organization plays a crucial role in determining the amount of information conveyed to WPP tasks. Therefore, further studies should be conducted to investigate advanced WPP input organizations, providing better references for predictions.

4.2 Identifying WPP features Benefiting Domain Generalization

As stated in Sect. 3, the input space considering spatial temporal correlation is $\mathbb{R}^{T \times P \times N_{SCADA}}$, which is extremely large. In applying DL methods into WPPs, learning latent representations from the overwhelming input space may be too specific towards a particular

WPP task and dataset, resulting in the lack of generalizability. To alleviate this issue, the ML field has presented a study trend on identifying a subset of meaningful features which possesses the causal relationship towards the concerned output. These features are considered as ones helping the model domain generalizability.

To identify a subset of latent features beneficial to the domain generalization, several deep learning methods (Arjovsky et al. 2019) utilize the invariant risk minimization (IRM), which is a learning scheme from the causal learning paradigm that optimizes the loss function under different environments. This approach can identify causal features that are robust to shifts in the environment. Therefore, it is promising to apply such a technique to WPP studies due to the high volatility of wind and the high-dimensional attributes after considering the spatial–temporal-system dynamics correlations. It is foreseeable that more WPP studies will consider the domain generalization issue into the WPP feature engineering, which seem to be scarce in the current literature.

4.3 Efficient feature engineering models

Currently, various complex models, including CNN, RNN, and transformer-based models, are commonly used in WPP studies. However, these models cannot be considered superior to others in terms of the best performance in all situations. For instance, one study (Zeng et al. 2023) showed that a simple linear model outperformed most of the complicated transformer-based models when predicting a long sequence. However, the reason behind this observation is still unclear. More studies are required to scientifically identify the optimal WPP methods for different WPP tasks.

Physics-informed WPP is also an interesting direction. Currently, physicals-informed techniques (Gijón et al. 2023; Wu et al. 2023; Tartakovsky et al. 2023; Pombo et al. 2022) are commonly utilized in expanding the dataset to obtain richer information. Because of the intrinsic relationship among the system dynamics in SCADA data, it is also promising to utilize physical principles for guiding the model design. Physics and domain knowledge can be leveraged to better govern the network design and prediction performance to enable a better generalization in WPPs (Lagomarsino-Oneto et al. 2023). With such development, it is possible to obtain a more reliable and robust prediction. In (Park and Park 2019), a PIGN has been developed to attack such issues and promising results in the wind power estimation have been reported. More advanced mechanisms can be developed to bring more values into the WPP tasks based on PINN.

4.4 Evaluation of the WPP performance in different scenarios

As stated in Sect. 2, deterministic WPP studies commonly utilized RMSE and MAE based metrics to evaluate the WPP performance. These metrics merely evaluate the difference between the prediction and actual power output. All instances are equally treated in these metrics. However, in some power grid operations, there are three additional requirements for wind power predictions.

- Concentration on the high wind power output instances: It is crucial to concentrate on high wind power output instances as it is more challenging for power systems to operate during periods of high wind power. Although predicting peaks is difficult, predicting high wind power output instances in WPPs should receive more attention.

- Prevention of adverse effect: It is important to prevent adverse effects which occur when the WPP prediction decreases while the actual WPP output increases, causing undesired operations in the power system. Therefore, more effective metrics need to be developed and utilized to evaluate the WPP performance, taking into account these requirements and other factors in power system operations.
- Wind power ramp consideration: The wind power ramp events possess a great potential of facilitating WPP models to address impacts brought by sudden large wind changes. Hence, it is of great practical value to study WPP methods together with better utilizing the historical wind power ramp pattern and with integrating more effective wind power ramp predictions.

Overall, these requirements highlight the need for more research in developing effective WPP methods and metrics that can meet the diverse needs of power grid operations.

5 Conclusions

This paper provided a comprehensive review of recent deep learning development in WPPs. It covered more than 140 recent WPP studies with advanced deep learning from two perspectives: WPP formulations and WPP methods.

First, different developments in WPP formulations including the new designs of inputs, new settings of outputs, and new application of evaluation metrics reported in recent WPP studies were summarized. The evolution of input designs was propelled by the broader availability of input sources and an increased interest in leveraging high-dimensional data for WPPs. Early data-driven studies mostly considered one source of input, either historical SCADA or NWP data. Subsequently, the input design evolved to jointly consider data sources and model the spatial influence to convey richer information in modeling. Moreover, the scope of input data considered expanded from a targeted wind turbine or wind farm site to include a group of neighboring sites. This change enabled a more comprehensive analysis of the influence of spatial–temporal data patterns on WPP modeling. To efficiently utilize data information under such setting, new data organization strategies including stacking, projection, and graph modelling were presented. To cope with incorporating more relevant attributes and enriching the WPP input, physics-informed methods were also employed to augment data and subsequently improve the accuracy based on a joint consideration of relevant knowledge. Meanwhile, in the WPP model output setting, the distribution of the power output at different levels was considered to provide more value prediction outcomes and to enable the quantification of future power output uncertainty. Recent studies also explored the link between the wind power ramp and WPP tasks, which aimed at coping with the impact of sudden large wind changes on WPP performances.

Next, a comprehensive review of the deep learning-based modeling methods for WPPs was conducted. Based on the process of converting inputs to wind power prediction, most of presented deep learning based modeling frameworks could be decomposed into four components, the signal processing component, clustering component, feature engineering component, and optimization component. It was observed that many advances in WPP modeling facilitated by deep learning were seen in the feature engineering component, mainly leveraging higher dimensional information and engineering low-dimensional but representative latent features for attaining better WPP performances. To conduct a more

in-depth review of the development of feature engineering techniques, a total of 8 groups of 28 state-of-the-art deep learning models were compared. The FNN model was one basic but well-known NN-based option. Via stacking multiple neural layers, the FNNs were able to transform the input to a set of latent features representing the useful information beneficial to the WPP task. Unlike FNNs, the AE models focused on extracting the latent representation based on the information of input itself. To address the uncertainty of the future wind power, probabilistic models were studied to provide the quartiles, confidence intervals, or distributions instead of a spot value of future wind power in WPP. Convolutional models and recurrent models were typically designed to extract the local spatial features and temporal features of the high-dimensional input data, respectively. To jointly analyze the spatial and temporal patterns, the ConvLSTM was proposed via integrating the advantages of convolution and recurrent models. However, the input data of ConvLSTM needed to be projected into a two-dimensional grid, which could be inefficient with a sparse distribution of the wind turbines or wind farms. To efficiently analyze the spatial patterns of wind data, graph-based models were also investigated. The attention-based models were also explored in WPP studies to adaptively analyze the spatial or temporal patterns of the input via attention mechanisms. Recent results reported that the attention-based models offered higher efficiency and better performance than convolutional and recurrent models in WPP tasks.

To verify the performance advancement brought by deep learning methods, we surveyed the existing literature and reported the improvement generated by deep learning methods against traditional models. We further verify the effectiveness of recent WPP development via reproducing and comparing latest deep learning models based on SCADA datasets collected from three commercial wind farms. The results demonstrated that DLinear achieved better performance on all of datasets considered in this work. The Informer, Transformer, and GRU model could also obtain promising results.

The future trends in WPP studies including the advanced input organization design, interpretable WPP features, more emerging modeling mechanisms, and more effective evaluation metrics were introduced and discussed. It is expected that these research areas will receive more attention in future WPP studies, as improvements in these aspects are likely to lead to further improvements in WPP accuracy and reliability.

In summary, this review serves as a guide for researchers and software developers dedicated to WPP studies and its downstream tasks.

Author contributions H. Liu and Z. Zhang both contributed into the conceptualization of this work and wrote the main manuscript. H. Liu conducted the literature review and technical analysis in this review work.

Declarations

Conflict of 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.

References

Abdoos AA (2016) A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. Neurocomputing 203:111–120

Abedinia O, Amjady N (2015) Short-term wind power prediction based on hybrid neural network and chaotic shark smell optimization. Int J Precis Eng Manuf-Green Technol 2(3):245–254

Abedinia O, Lotfi M, Bagheri M, Sobhani B, Shafie-Khah M, Catalao J (2020) Improved EMD-based complex prediction model for wind power forecasting. IEEE Trans Sustain Energy 11(4):2790–2802

Ahmadpour A, Farkoush SG (2020) Gaussian models for probabilistic and deterministic Wind Power Prediction: Wind farm and regional. Int J Hydrogen Energy 45(51):27779–27791

Ahn EJ, Hur J (2023) A short-term forecasting of wind power outputs using the enhanced wavelet transform and arimax techniques. Renew Energy 212:394–402

Aly HHH (2022) A hybrid optimized model of adaptive neuro-fuzzy inference system, recurrent Kalman filter and neuro-wavelet for wind power forecasting driven by DFIG. Energy 239:122367

Amjady N, Abedinia O (2017) (2017) Short term wind power prediction based on improved Kriging interpolation, empirical mode decomposition, and closed-loop forecasting engine. Sustainability 9(11):2104

Amjady N, Keynia F, Zareipour H (2011) Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization. IEEE Trans Sustain Energy 2(3):265–276

An X, Jiang D, Liu C, Zhao M (2011) Wind farm power prediction based on wavelet decomposition and chaotic time series. Expert Syst Appl 38(9):11280–11285

An G, Jiang Z, Chen L, Cao X, Li Z, Zhao Y, Sun H (2021) Ultra short-term wind power forecasting based on sparrow search algorithm optimization deep extreme learning machine. Sustainability 13(18):10453

Arjovsky M., Bottou L., Gulrajani I., Lopez-Paz D. (2019) Invariant risk minimization. arXiv preprint arXiv1907.02893.

Azimi R, Ghofrani M, Ghayekhloo M (2016) A hybrid wind power forecasting model based on data mining and wavelets analysis. Energy Convers Manage 127:208–225

Banik A, Behera C, Sarathkumar TV, Goswami AK (2020) Uncertain wind power forecasting using LSTM-based prediction interval. IET Renew Power Gener 14(14):2657–2667

Bentsen LØ, Warakagoda ND, Stenbro R, Engelstad P (2022) Wind park power prediction: attention-based Graph networks and deep learning to capture wake losses. J Phys: Conf Series. 2265(2):022035

Bessa RJ, Miranda V, Gama J (2009) Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting. IEEE Trans Power Syst 24(4):1657–1666

Bessa RJ, Miranda V, Botterud A, Wang J, Constantinescu EM (2012) Time adaptive conditional kernel density estimation for wind power forecasting. IEEE Trans Sustain Energy 3(4):660–669

Bilal B, Adjallah KH, Sava A et al (2023) Wind turbine output power prediction and optimization based on a novel adaptive neuro-fuzzy inference system with the moving window. Energy 263:126159

Blonbou R, Monjoly S, Dorville JF (2011) An adaptive short-term prediction scheme for wind energy storage management[J]. Energy Convers Manage 52(6):2412–2416

Bludszuweit H, Domínguez-Navarro JA, Llombart A (2008) Statistical analysis of wind power forecast error. IEEE Trans Power Syst 23(3):983–991

Bokde N, Feijóo A, Villanueva D et al (2019) A review on hybrid empirical mode decomposition models for wind speed and wind power prediction. Energies 12(2):254

Bossanyi E (1985) Short-term wind prediction using kalman filters. Wind Eng 9(1):1–8

Brown BG, Katz RW, Murphy AH (1984) Time series models to simulate and forecast wind speed and wind power. J Appl Meteorol Climatol 23(8):1184–1195

Cali U, Sharma V (2019) Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection. Int J Smart Grid and Clean Energy 8(2):103–110

Cao Y, Liu G, Luo D, Bavirisetti DP, Xiao G (2023) Multi-timescale photovoltaic power forecasting using an improved Stacking ensemble algorithm based LSTM-Informer model. Energy 2783:128669

Catalão JPS, Pousinho HMI, Mendes VMF (2010) Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal. IEEE Trans Sustain Energy 2(1):50–59

Cavalcante L, Bessa RJ, Reis M (2017) LASSO vector autoregression structures for very short-term wind power forecasting. Wind Energy 20(4):657–675

Chen P, Pedersen T, Bak-Jensen B, Chen Z (2009) ARIMA-based time series model of stochastic wind power generation. IEEE Trans Power Syst 25(2):667–676

Chi D, Yang C (2023) Wind power prediction based on WT-BiGRU-attention-TCN model. Front Energy Res 11:1156007

Chitsaz H, Amjady N, Zareipour H (2015) Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm. Energy Convers Manage 89:588–598

Cui Y, Chen Z, He Y, Xiong X, Li F (2023) An algorithm for forecasting day-ahead wind power via novel long short-term memory and wind power ramp events. Energy 263:125888

Deng Z, Li Y, Zhu H, Huang K, Tang Z, Wang Z (2020) Sparse stacked autoencoder network for complex system monitoring with industrial applications. Chaos, Solitons Fractals 137:109838

Deng B, Wu Y, Liu S, Xu Z. (2022) Wind Speed Forecasting for Wind Power Production Based on Frequency-Enhanced Transformer. 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), pp. 151–155.

Dhiman HS, Deb D, Guerrero JM (2022) On wavelet transform based convolutional neural network and twin support vector regression for wind power ramp event prediction. Sustain Comput: Inform Syst 36:100795

Ding M, Zhou H, Xie H, Wu M, Nakanishi Y, Yokoyama R (2019) A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. Neurocomputing 365:54–61

Ding J, Chen G, Yuan K (2020) Short-term wind power prediction based on improved grey wolf optimization algorithm for extreme learning machine. Processes 8(1):109

Dong L, Wang L, Khahro S, F, Gao S., Liao X. (2016) Wind power day-ahead prediction with cluster analysis of NWP. Renew Sustain Energy Rev 60:1206–1212

Dong Q, Sun Y, Li P (2017) (2017) A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: a case study of wind farms in China. Renew Energy 102:241–257

Dowell J, Pinson P (2015) Very-short-term probabilistic wind power forecasts by sparse vector autoregression. IEEE Trans Smart Grid 7(2):763–770

Farah S, Humaira N, Aneela Z, Steffen E (2022) Short-term multi-hour ahead country-wide wind power prediction for Germany using gated recurrent unit deep learning. Renew Sustain Energy Rev 167:112700

Gallego C, Cuerva A, Costa A (2014) Detecting and characterising ramp events in wind power time series. J Phys: Conf Ser 555(1):012040

Gallego-Castillo C, Bessa R, Cavalcante L, Lopez-Garcia O (2016) On-line quantile regression in the RKHS (Reproducing Kernel Hilbert Space) for operational probabilistic forecasting of wind power. Energy 113:355–365

Ghadi MJ, Gilani SH, Afrakhte H, Baghramian A (2014) A novel heuristic method for wind farm power prediction: a case study. Int J Electr Power Energy Syst 63:962–970

Gijón A, Pujana-Goitia A, Perea E, et al. Prediction of wind turbines power with physics-informed neural networks and evidential uncertainty quantification. arXiv preprint arXiv:2307.14675, 2023.

Global Wind Energy Council (2022) Global wind report 2022.

Guo H, Wang J, Li Z et al (2022) A multivariable hybrid prediction system of wind power based on outlier test and innovative multi-objective optimization. Energy 239:122333

Han Y, Tong X (2020) Multi-step short-term wind power prediction based on three-level decomposition and improved grey wolf optimization. IEEE Access 8:67124–67136

Han L, Jing H, Zhang R, Gao Z (2019a) Wind power forecast based on improved long short term memory network. Energy 189:116300

Han L, Zhang R, Wang X, Bao A, Jing H (2019b) Multi-step wind power forecast based on VMD-LSTM. IET Renew Power Gener 13(10):1690–1700

Harbola S, Coors V (2019) One dimensional convolutional neural network architectures for wind prediction. Energy Convers Manage 195:70–75

He J, Yu C, Li Y, Xiang H (2020) Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning. Energy Convers Manage 205:112418

He Y, Li H, Wang S, Yao X (2021) Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression. Neurocomputing 430:121–137

He Y, Zhu C, An X (2023) A trend-based method for the prediction of offshore wind power ramp. Renew Energy 209:248–261

Heinermann J, Kramer O (2016) Machine learning ensembles for wind power prediction. Renew Energy 89:671–679

Heydari A, Majidi Nezhad M, Neshat M, Garcia DA, Keynia F, De Santoli L, Bertling Tjernberg L (2021) A combined fuzzy GMDH neural network and grey wolf optimization application for wind turbine power production forecasting considering SCADA data. Energies 14(12):3459

Higashiyama K, Fujimoto Y, Hayashi Y (2018) Feature extraction of NWP data for wind power forecasting using 3D-convolutional neural networks. Energy Procedia 155:350–358

Hong YY, Rioflorido CLPP (2019) A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. Appl Energy 250:530–539

Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ (2016) Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond. Int J Forecast 32(3):896–913

Hossain MA, Chakrabortty RK, Elsawah S et al (2021) Very short-term forecasting of wind power generation using hybrid deep learning model. J Clean Prod 296:126564

Hu J, Heng J, Tang J, Guo M (2018) Research and application of a hybrid model based on Meta learning strategy for wind power deterministic and probabilistic forecasting. Energy Convers Manage 173:197–209

Hu H, Wang L, Lv S (2020) Forecasting energy consumption and wind power generation using deep echo state network. Renew Energy 154:598–613

Hu J, Zhang L, Tang J, Liu Z (2023) A novel transformer ordinal regression network with label diversity for wind power ramp events forecasting. Energy 280:128075

Huang X, Jiang A (2022) Wind power generation forecast based on multi-step informer network. Energies 15(18):6642

Huang B, Wang J (2022) Applications of physics-informed neural networks in power systems-a review. IEEE Trans Power Syst 38(1):572–588

Huang D, Gong R, Gong S (2015) Prediction of wind power by chaos and BP artificial neural networks approach based on genetic algorithm. J Elect Eng Technol 10(1):41–46

Huang L, Li L, Wei X, Zhang D (2022) Short-term prediction of wind power based on BiLSTM–CNN–WGAN-GP. Soft Comput 26(20):10607–10621

Jahangir H, Tayarani H, Gougheri SS, Golkar MA, Ahmadian A, Elkamel A (2020) Deep learning-based forecasting approach in smart grids with micro-clustering and bi-directional LSTM network. IEEE Trans Ind Electron 68(9):8298–8309

Jiao R, Huang X, Ma X, Han L, Tian W (2018) A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. IEEE Access 6:17851–17858

Ju Y, Sun G, Chen Q, Zhang M, Zhu H, Rehman MU (2019) A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting. IEEE Access 7:28309–28318

Jung J, Broadwater RP (2014) Current status and future advances for wind speed and power forecasting. Renew Sustain Energy Reviews 31:762–777

Khalid R, Javaid N (2020) A survey on hyperparameters optimization algorithms of forecasting models in smart grid. Sustain Cities Soc 61:102275

Khazaei S, Ehsan M, Soleymani S et al (2022) A high-accuracy hybrid method for short-term wind power forecasting. Energy 238:122020

Khodayar M, Wang J (2018) Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans Sustain Energy 10(2):670–681

Khodayar M, Kaynak O, Khodayar ME (2017) Rough deep neural architecture for short-term wind speed forecasting. IEEE Trans Industr Inf 13(6):2770–2779

Khodayar M, Saffari M, Williams M et al (2022) Interval deep learning architecture with rough pattern recognition and fuzzy inference for short-term wind speed forecasting. Energy 254:124143

Khosravi A, Nahavandi S (2013) Combined nonparametric prediction intervals for wind power generation. IEEE Trans Sustain Energy 4(4):849–856

Kim D, Hur J (2018) Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method. Energy 157:211–226

Kisvari A, Lin Z, Liu X (2021) Wind power forecasting–a data-driven method along with gated recurrent neural network. Renew Energy 163:1895–1909

Klinges DH, Duffy JP, Kearney MR, Maclean IM (2022) mcera5: Driving microclimate models with ERA5 global gridded climate data. Methods Ecol Evol 13(7):1402–1411

Ko MS, Lee K, Kim JK, Hong C, Dong Z, Hur K (2020) Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting. IEEE Trans Sustain Energy 12(2):1321–1335

Kokkos N, Zoidou M, Zachopoulos K et al (2021) Wind climate and wind power resource assessment based on gridded scatterometer data: A Thracian Sea case study. Energies 14(12):3448

Külüm E, Genç MS, Karagöz F (2023) Evaluation of wind measurement methods for determination of realistic wind shear: A case study in Aksaray. Turkey. Flow Meas Instrum 93:102408

Kou P, Liang D, Gao F, Gao L (2014) Probabilistic wind power forecasting with online model selection and warped gaussian process. Energy Convers Manage 84:649–663

Lagomarsino-Oneto D, Meanti G, Pagliana N et al (2023) Physics informed machine learning for wind speed prediction. Energy 268:126628

Lahouar A, Slama JBH (2017) Hour-ahead wind power forecast based on random forests. Renew Energy 109:529–541

Landberg L (1999) Short-term prediction of the power production from wind farms. Wind Eng Ind Aerodyn 80(1–2):207–220

Li X, Zhang W (2022) Physics-informed deep learning model in wind turbine response prediction. Renew Energy 185:932–944

Li P, Guan X, Wu J (2015a) Aggregated wind power generation probabilistic forecasting based on particle filter. Energy Convers Manage 96:579–587

Li S, Wang P, Goel L (2015b) Wind power forecasting using neural network ensembles with feature selection. IEEE Trans Sustain Energy 6(4):1447–1456

Li L, Chang Y, Tseng M, Liu J, Lim M (2020a) Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm. J Clean Prod 270:121817

Li L, Yin X, Jia X, Sobhani B (2020b) Day ahead powerful probabilistic wind power forecast using combined intelligent structure and fuzzy clustering algorithm. Energy 192:116498

Li L, Zhao X, Tseng M, Tan R (2020c) Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. J Clean Prod 242:118447

Li L, Cen Z, Tseng M, Shen Q, Ali MH (2021) Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic-Support vector regression machine. J Clean Prod 279:123739

Li Z, Luo X, Liu M, Cao X, Du S, Sun H (2022) Wind power prediction based on EEMD-Tent-SSA-LS-SVM. Energy Rep 8:3234–3243

Liu H, Chen C (2019) Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction. Appl Energy 254:113686

Liu X, Zhang Z (2021) A two-stage deep autoencoder-based missing data imputation method for wind farm SCADA data. IEEE Sens J 21(9):10933–10945

Liu H, Zhang Z (2022b) A bilateral branch learning paradigm for short term wind power prediction with data of multiple sampling resolutions. J Clean Prod 380:134977

Liu H, Tian H, Chen C, Li Y (2010) A hybrid statistical method to predict wind speed and wind power. Renew Energy 35(8):1857–1861

Liu H, Mi X, Li Y (2018a) Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Convers Manage 156:498–514

Liu T, Wei H, Zhang K (2018b) Wind power prediction with missing data using Gaussian process regression and multiple imputation. Appl Soft Comput 71:905–916

Liu Z, Hajiali M, Torabi A, Ahmadi B, Simoes R (2018c) Novel forecasting model based on improved wavelet transform, informative feature selection, and hybrid support vector machine on wind power forecasting. J Ambient Intell Humaniz Comput 9(6):1919–1931

Liu H, Chen C, Lv X, Wu X, Liu M (2019a) Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. Energy Convers Manage 195:328–345

Liu Y, Guan L, Hou C, Han H, Liu Z, Sun Y, Zheng M (2019b) Wind power short-term prediction based on LSTM and discrete wavelet transform. Appl Sci 9(6):1108

Liu H, Duan Z, Chen C (2020a) Wind speed big data forecasting using time-variant multi-resolution ensemble model with clustering auto-encoder. Appl Energy 280:115975

Liu H, Li Y, Duan Z, Chen C (2020b) A review on multi-objective optimization framework in wind energy forecasting techniques and applications. Energy Convers Manage 224:113324

Liu H, Chen D, Lin F, Wan Z (2021a) Wind power short-term forecasting based on LSTM neural network with dragonfly algorithm. J Phys: Conf Series. 1748(3):032015

Liu J, Shi Q, Han R, Yang J (2021b) A hybrid GA–PSO–CNN model for ultra-short-term wind power forecasting. Energies 14(20):6500

Liu X, Cao Z, Zhang Z (2021d) Short-term predictions of multiple wind turbine power outputs based on deep neural networks with transfer learning. Energy 217:119356

Liu X, Yang L, Zhang Z (2021e) Short-term multi-step ahead wind power predictions based on a novel deep convolutional recurrent network method. IEEE Trans Sustain Energy 12(3):1820–1833

Liu H, Han H, Sun Y, Shi G, Su M, Liu Z, Wang H, Deng X (2022a) Short-term wind power interval prediction method using VMD-RFG and Att-GRU. Energy 251:123807

Liu X, Yang L, Zhang Z (2022b) The attention-assisted ordinary differential equation networks for short-term probabilistic wind power predictions. Appl Energy 324:119794

Liu H, Yang L, Zhang B, Zhang Z (2023) A two-channel deep network based model for improving ultra-short-term prediction of wind power via utilizing multi-source data. Energy 283:128510

Liu H., Zhang Z. (2022) A Bi-party Engaged Modeling Framework for Renewable Power Predictions with Privacy-preserving. IEEE Transactions on Power Systems, in press.

Liu S., Yu H., Liao C., Li J., Lin W., Liu A. (2021b) Dustdar S. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. International Conference on Learning Representations.

Lu H, Ma X, Huang K, Azimi M (2020) Prediction of offshore wind farm power using a novel two-stage model combining kernel-based nonlinear extension of the Arps decline model with a multi-objective grey wolf optimizer. Renew Sustain Energy Rev 27:109856

Ma J, Yang M, Lin Y (2014) Ultra-short-term wind turbine power forecast based on empirical dynamic modeling. IEEE Trans Sustain Energy 11(2):906–915

Madhiarasan M, Deepa SN (2017) Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting. Artif Intell Rev 48:449–471

Mahmoud T, Dong Z, Ma J (2018) An advanced approach for optimal wind power generation prediction intervals by using self-adaptive evolutionary extreme learning machine. Renew Energy 126:254–269

Marugán AP, Márquez FPG, Perez JMP, Ruiz-Hernández D (2018) A survey of artificial neural network in wind energy systems. Appl Energy 228:1822–1836

Mehrkanoon S (2019) Deep shared representation learning for weather elements forecasting. Knowl-Based Syst 179:120–128

Men Z, Yee E, Lien FS, Wen D, Chen Y (2016) Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. Renew Energy 87:203–211

Menemenlis N, Huneault M, Robitaille A (2012) Computation of dynamic operating balancing reserve for wind power integration for the time-horizon 1–48 hours. IEEE Trans Sustain Energy 3(4):692–702

Meng A, Ge J, Yin H, Chen S (2016) Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. Energy Convers Manage 114:75–88

Meng A, Chen S, Ou Z, Ding W, Zhou H, Fan J, Yin H (2022a) A hybrid deep learning architecture for wind power prediction based on bi-attention mechanism and crisscross optimization. Energy 238:121795

Meng A, Zhu Z, Deng W, Ou Z, Lin S, Wang C, Xu X, Wang X, Yin H, Luo J (2022b) A novel wind power prediction approach using multivariate variational mode decomposition and multi-objective crisscross optimization based deep extreme learning machine. Energy 260:124957

Methaprayoon K, Yingvivatanapong C, Lee WJ, Liao JR (2007) An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty. IEEE Trans Ind Appl 43(6):1441–1448

Nascimento EGS, de Melo TAC, Moreira DM (2023) A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy. Energy 278:127678

Neshat M, Nezhad MM, Abbasnejad E, Mirjalili S, Groppi D, Heydarib A, Tjernberg BL, Garcia DA, Alexander B, Shi Q, Wagner M (2021) Wind turbine power output prediction using a new hybrid neuro-evolutionary method. Energy 229:120617

Nielsen HA, Madsen H, Nielsen TS (2006) Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy: Int J Prog Appl Wind Power Convers Technol 9(1–2):95–108

Niu Z, Yu Z, Tang W, Wu Q (2020) Reformat M. Wind power forecasting using attention-based gated recurrent unit network. Energy 196:117081

Osório GJ, Matias JCO, Catalão JPS (2015) Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. Renew Energy 75:301–307

Ouyang T, Huang H, He Y, Tang Z (2020) Chaotic wind power time series prediction via switching data-driven modes. Renew Energy 145:270–281

Park J, Park J (2019) Physics-induced graph neural network: an application to wind-farm power estimation. Energy 187:115883

Paterakis NG, Erdinc O, Bakirtzis AG et al (2014) Load-following reserves procurement considering flexible demand-side resources under high wind power penetration[J]. IEEE Trans Power Syst 30(3):1337–1350

Peng X, Xiong L, Wen J, Xu Y, Fan W, Feng S, Wang B (2016) A very short term wind power prediction approach based on multilayer restricted Boltzmann machine. IEEE PES Asia-Pacific Power and Energy Eng Conf (APPEEC) 2016:2409–2413

Peng Z, Peng S, Fu L et al (2020) A novel deep learning ensemble model with data denoising for short-term wind speed forecasting. Energy Convers Manage 207:112524

Pircalabu A, Hvolby T, Jung J, Høg E (2017) Joint price and volumetric risk in wind power trading: a copula approach. Energy Econ 62:139–154

Pombo DV, Rincón MJ, Bacher P, Bindner HW, Spataru SV, Sørensen PE (2022) Assessing stacked physics-informed machine learning models for co-located wind–solar power forecasting. Sustain Energy, Grids and Netw 32:100943

Prósper MA, Otero-Casal C, Fernández FC, Miguez-Macho G (2019) Wind power forecasting for a real onshore wind farm on complex terrain using WRF high resolution simulations. Renew Energy 135:674–686

Qian Z, Pei Y, Zareipour H, Chen N (2019) A review and discussion of decomposition based hybrid models for wind energy forecasting applications. Appl Energy 235:939–953

Qiao B, Liu J, Wu P et al (2022) Wind power forecasting based on variational mode decomposition and high-order fuzzy cognitive maps. Appl Soft Comput 129:109586

Qureshi AS, Khan A, Zameer A, Usman A (2017) Wind power prediction using deep neural network based meta regression and transfer learning. Appl Soft Comput 58:742–755

Ren Z, Verma AS, Li Y, Teuwen JJ, Jiang Z (2021) Offshore wind turbine operations and maintenance: A state-of-the-art review. Renew Sustain Energy Rev 144:110886

Ren J, Yu Z, Gao G et al (2022) A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism. Energy Rep 8:437–443

Ronay Aka, YFL, Vitellia, V, Zio E (2013) A genetic algorithm and neural network technique for predicting wind power under uncertainty. Chemical Engineering, 33.

Saffari M, Williams M, Khodayar M, et al (2021). Robust wind speed forecasting: A deep spatio-temporal approach. 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe): 1–6.

Saroha S., Aggarwal S. K. (2014) Multi step ahead forecasting of wind power by genetic algorithm based neural networks. 2014 6th IEEE Power India International Conference (PIICON), 1–6.

Severiano CA, e Silva, P.C.D.L., Cohen, M.W. and Guimarães, F.G., (2021) Evolving fuzzy time series for spatio-temporal forecasting in renewable energy systems. Renew Energy 171:764–783

Shahid F, Zameer A, Mehmood A et al (2020a) A novel wavenets long short term memory paradigm for wind power prediction. Appl Energy 269:115098

Shahid F, Khan A, Zameer A et al (2020b) Wind power prediction using a three stage genetic ensemble and auxiliary predictor. Appl Soft Comput 90:106151

Shahid F, Zameer A, Muneeb M (2021) A novel genetic LSTM model for wind power forecast. Energy 223:120069

Shi J, Guo J, Zheng S (2012) Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renew Sustain Energy Rev 16(5):3471–3480

Shi J, Lee WJ, Liu X (2017a) Generation scheduling optimization of wind-energy storage system based on wind power output fluctuation features. IEEE Trans Ind Appl 54(1):10–17

Shi Z, Liang H, Dinavahi V (2017b) Direct interval forecast of uncertain wind power based on recurrent neural networks. IEEE Trans Sustain Energy 9(3):1177–1187

Sideratos G, Hatziargyriou ND (2007) An advanced statistical method for wind power forecasting. IEEE Trans Power Syst 22(1):258–265

Song L, Xie Q, He Y, Dang P. (2020) Ultra-short-term wind power combination forecasting model based on MEEMD-SAE-Elman. 2020 IEEE 4th information technology, networking, electronic and automation control conference (itnec) 1:1844–1850.

Song J., Peng X., Yang Z., Wei P., Wang B., Wang Z., (2022) A Novel Wind Power Prediction Approach for Extreme Wind Conditions Based on TCN-LSTM and Transfer Learning. IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), 1410–1415.

Sun H (2021) Hybrid model with secondary decomposition, randomforest algorithm, clustering analysis and long short memory network principal computing for short-term wind power forecasting on multiple scales. Energy 221:119848

Sun Z, Zhao M (2020) Short-term wind power forecasting based on VMD decomposition, convlstm networks and error analysis. IEEE Access 8:134422–134434

Tan B., Ma X., Shi Q., Guo M., Zhao H., Shen X. (2021) Ultra-short-term Wind Power Forecasting Based on Improved LSTM. 6th International Conference on Power and Renewable Energy (ICPRE), 1029–1033.

Tartakovsky AM, Ma T, Barajas-Solano DA et al (2023) Physics-informed Gaussian process regression for states estimation and forecasting in power grids. Int J Forecast 39(2):967–980

Tian C, Niu T, Wei W (2022) Developing a wind power forecasting system based on deep learning with attention mechanism. Energy 257:124750

Treiber NA, Heinermann J, Kramer O (2016) Wind power prediction with machine learning. In: Lässig J, Kersting K, Morik K (eds) Computational sustainability. Springer, Cham, pp 13–29

Usaola J, Ravelo O, Gonzalez G, Soto F, Davila MC, Diaz-Guerra B (2004) Benefits for wind energy in electricity markets from using short term wind power prediction tools; a simulation study. Wind Eng 28(1):119–127

Valsaraj P, Thumba DA, Asokan K et al (2020) Symbolic regression-based improved method for wind speed extrapolation from lower to higher altitudes for wind energy applications. Appl Energy 260:114270

Valsaraj P, Thumba DA, Kumar S (2022) Machine learning-based simplified methods using shorter wind measuring masts for the time ahead wind forecasting at higher altitude for wind energy applications. Renew Energy Environ Sustain 7:24

Vargas SA, Esteves GRT, Maçaira PM, Bastos BQ, Oliveira FLC, Souza RC (2019) Wind power generation: a review and a research agenda. J Clean Prod 218:850–870

Viet DT, Phuong VV, Duong MQ, Tran QT (2020) Models for short-term wind power forecasting based on improved artificial neural network using particle swarm optimization and genetic algorithms. Energies 13(11):2873

Wan C, Xu Z, Pinson P, Dong Z, Wong K (2013a) Optimal prediction intervals of wind power generation. IEEE Trans Power Syst 29(3):1166–1174

Wan C, Xu Z, Pinson P, Dong Z, Wong K (2013b) Probabilistic forecasting of wind power generation using extreme learning machine. IEEE Trans Power Syst 29(3):1033–1044

Wan C, Lin J, Wang J, Song Y, Dong Z (2016a) Direct quantile regression for nonparametric probabilistic forecasting of wind power generation. IEEE Trans Power Syst 32(4):2767–2778

Wan C, Song Y, Xu Z, Yang G, Nielsen AH (2016b) Probabilistic wind power forecasting with hybrid artificial neural networks. Electric Power Compon Syst 44(15):1656–1668

Wang J, Shahidehpour M, Li Z (2008) Security-constrained unit commitment with volatile wind power generation. IEEE Trans Power Syst 23(3):1319–1327

Wang J, Song Y, Liu F, Hou R (2016) Analysis and application of forecasting models in wind power integration: a review of multi-step-ahead wind speed forecasting models. Renew Sustain Energy Rev 60:960–981

Wang C, Zhou K, Yang S (2017a) A review of residential tiered electricity pricing in China. Renew Sustain Energy Rev 79:533–543

Wang H, Li G, Wang G, Peng J, Jiang H, Liu Y (2017b) (2017a) Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 188:56–70

Wang Y, Hu Q, Meng D, Zhu P (2017c) Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. Appl Energy 208:1097–1112

Wang K, Qi X, Liu H, Song J (2018) Deep belief network based k-means cluster approach for short-term wind power forecasting. Energy 165:840–852

Wang R, Li C, Fu W, Tang G (2019a) Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction. IEEE Trans Neural Netw Learn Syst 99:1–14

Wang Y, Hu Q, Srinivasan D, Wang Z (2019b) Short-term wind speed or power forecasting with heteroscedastic support vector regression. IEEE Trans Sustain Energy 10(1):16–25

Wang Y, Yu Y, Cao S, Zhang X, Gao S (2020) A review of applications of artificial intelligent algorithms in wind farms. Artif Intell Rev 53:3447–3500

Wang L, Tao R, Hu H, Zeng Y (2021a) Effective wind power prediction using novel deep learning network: stacked independently recurrent autoencoder. Renew Energy 164:642–655

Wang S, Li B, Li G, Yao B, Wu J (2021b) Short-term wind power prediction based on multidimensional data cleaning and feature reconfiguration. Appl Energy 292:116851

Wang S, Wang J, Lu H, Zhao W (2021c) A novel combined model for wind speed prediction–Combination of linear model, shallow neural networks, and deep learning approaches. Energy 234:121275

Wang Y, Zou R, Liu F, Zhang L, Liu Q (2021d) A review of wind speed and wind power forecasting with deep neural networks. Appl Energy 304:117766

Wang Y, Xu H, Song M et al (2023) A convolutional Transformer-based truncated Gaussian density network with data denoising for wind speed forecasting. Appl Energy 333:120601

Weide Luiz E, Fiedler S (2022) Spatiotemporal observations of nocturnal low-level jets and impacts on wind power production. Wind Energy Sci 7(4):1575–1591

Wen H., Gu J., Ma J., Jin Z. (2019) Probabilistic wind power forecasting via Bayesian deep learning based prediction intervals. In: 2019 ieee 17th international conference on industrial informatics (indin) 1:1091–1096.

Wilms H., Cupelli M., Monti A., Gross T. (2021) Exploiting spatio-temporal dependencies for RNN-based wind power forecasts. IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), 921–926.

Woo S, Park J, Park J, Manuel L (2019) Wind field-based short-term turbine response forecasting by stacked dilated convolutional lstms. IEEE Trans Sustain Energy 11(4):2294–2304

Woo S., Park J., Park J. (2018) Predicting wind turbine power and load outputs by multi-task convolutional LSTM model. 2018 ieee power & energy society general meeting (pesgm), 1–5.

Wu Z, Wang B (2021) An Ensemble neural network based on variational mode decomposition and an improved sparrow search algorithm for wind and solar power forecasting. IEEE Access 9:166709–166719

Wu F, Cattani C, Song W, Zio E (2020a) Fractional ARIMA with an improved cuckoo search optimization for the efficient Short-term power load forecasting. Alex Eng J 59(5):3111–3118

Wu Z, Xia X, Xiao L, Liu Y (2020b) Combined model with secondary decomposition-model selection and sample selection for multi-step wind power forecasting. Appl Energy 261:114345

Wu H, Xu J, Wang J, Long M (2021) Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Adv Neural Inf Process Syst 34:22419–22430

Wu Q, Zheng H, Guo X, Liu G (2022) Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks. Renew Energy 199:977–992

Wu Z, Sun B, Feng Q et al (2023) Physics-informed AI surrogates for day-ahead wind power probabilistic forecasting with incomplete data for smart grid in smart cities[J]. CMES-Comput Model Eng Sci 137(1):527–554

Xiao L, Wang J, Dong Y, Wu J (2015) Combined forecasting models for wind energy forecasting: a case study in China. Renew Sustain Energy Rev 44:271–288

Xu Y, Jia L, Yang W (2022b) Correlation based neuro-fuzzy Wiener type wind power forecasting model by using special separate signals. Energy Convers Manage 253:115173

Xu H., Zhen Z., Wang F. (2022) NWP Feature Selection and GCN-based Ultra-short-term Wind Farm Cluster Power Forecasting Method. 2022 IEEE Industry Applications Society Annual Meeting (IAS) ,1–22.

Yan H, Wu Z (2020) A hybrid short-term wind power prediction model combining data processing, multiple parameters optimization and multi-intelligent models apportion strategy. IEEE Access 8:227126–227140

Yan J, Li K, Bai E, Deng J, Foley AM (2015) Hybrid probabilistic wind power forecasting using temporally local Gaussian process. IEEE Trans Sustain Energy 7(1):87–95

Yan J, Zhang H, Liu Y, Han S, Li L, Lu Z (2018) Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping. IEEE Trans Power Syst 33(3):3276–3284

Yang L, Zhang Z (2021) A deep attention convolutional recurrent network assisted by k-shape clustering and enhanced memory for short term wind speed predictions. IEEE Trans Sustain Energy 13(2):856–867

Yang M, Fan S, Lee WJ (2013) Probabilistic short-term wind power forecast using componential sparse Bayesian learning. IEEE Trans Ind Appl 49(6):2783–2792

Yang L, Zheng Z, Zhang Z (2021a) An improved mixture density network via wasserstein distance based adversarial learning for probabilistic wind speed predictions. IEEE Trans Sustain Energy 13(2):755–766

Yang M, Shi C, Liu H (2021b) Day-ahead wind power forecasting based on the clustering of equivalent power curves. Energy 218:119515

Yang Q, Huang G, Li T et al (2023) A novel short-term wind speed prediction method based on hybrid statistical-artificial intelligence model with empirical wavelet transform and hyperparameter optimization. J Wind Eng Ind Aerodyn 240:105499

Yin H, Dong Z, Chen Y, Ge J, Lai L, Vaccaro A, Meng A (2017) An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. Energy Convers Manage 150:108–121

Yin H, Ou Z, Huang S, Meng A (2019) A cascaded deep learning wind power prediction approach based on a two-layer of mode decomposition. Energy 189:116316

Yin H, Ou Z, Fu J, Cai Y, Chen S, Meng A (2021) A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture. Energy 234:121271

Yu R, Gao J, Yu M, Lu W, Xu T, Zhao M, Zhang J, Zhang R, Zhang Z (2019a) Lstm-efg for wind power forecasting based on sequential correlation features. Futur Gener Comput Syst 93:33–42

Yu R, Liu Z, Li X, Lu W, Yu M, Wang J, Li B (2019b) Scene learning: Deep convolutional networks for wind power prediction by embedding turbines into grid space. Appl Energy 238:249–257

Yu Y, Han X, Yang M, Yang J (2020) Probabilistic prediction of regional wind power based on spatiotemporal quantile regression. IEEE Trans Ind Appl 56(6):6117–6127

Yu C, Li Y, Chen Q et al (2022a) Matrix-based wavelet transformation embedded in recurrent neural networks for wind speed prediction. Appl Energy 324:119692

Yu R, Sun Y, He D, Gao J, Liu Z, Yu M (2022b) Spatio-temporal graph cross-correlation auto-encoding network for wind power prediction. Int J Mach Learn Cybern 15(1):51–63

Yu X., Luo L., (2022b) Day-Ahead Wind Power Prediction Based on BP Neural Network Optimized by Improved Sparrow Search Algorithm. 4th Asia Energy and Electrical Engineering Symposium (AEEES). 2022;230–235.

Yuan X, Chen C, Jiang M, Yuan Y (2019) Prediction interval of wind power using parameter optimized Beta distribution based LSTM model. Appl Soft Comput 82:105550

Zendehboudi A, Baseer MA, Saidur R (2018) Application of support vector machine models for forecasting solar and wind energy resources: a review. J Clean Prod 199:272–285

Zeng A, Chen M, Zhang L, Xu Q (2023) Are transformers effective for time series forecasting? Proceed of the AAAI Conf Artif Intell 37:1–15

Zhang Y, Wang J (2016) K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting. Int J Forecast 32(3):1074–1080

Zhang Y, Wang J (2018) A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without direct access to off-site information. IEEE Trans Power Syst 33(5):5714–5726

Zhang J, Zhao X (2021) Three-dimensional spatiotemporal wind field reconstruction based on physics-informed deep learning. Appl Energy 300:117390

Zhang H, Chen L, Qu Y, Zhao G, Guo Z (2014) Support vector regression based on grid-search method for short-term wind power forecasting. J Appl Math. https://doi.org/10.1155/2014/835791

Zhang Y, Wang J, Luo X (2015) Probabilistic wind power forecasting based on logarithmic transformation and boundary kernel. Energy Convers Manage 96:440–451

Zhang Y, Liu K, Qin L, An X (2016) Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods. Energy Convers Manage 12:208–219

Zhang J, Yan J, Infield D, Liu Y, Lien F-s (2019a) Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and gaussian mixture model. Appl Energy 241:229–244

Zhang Y, Le J, Liao X, Zheng F, Li Y (2019b) A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum analysis and locality-sensitive hashing. Energy 168:558–572

Zhang H, Liu Y, Yan J, Han S, Li L, Long Q (2020a) Improved deep mixture density network for regional wind power probabilistic forecasting. IEEE Trans Power Syst 35(4):2549–2560

Zhang X, Han P, Xu L, Zhang F, Wang Y, Gao L (2020b) Research on bearing fault diagnosis of wind turbine gearbox based on 1DCNN-PSO-SVM. IEEE Access 8:192248–192258

Zhang Y, Li Y, Zhang G (2020c) Short-term wind power forecasting approach based on Seq2Seq model using NWP data. Energy 213:118371

Zhang W, Lin Z, Liu X (2022a) Short-term offshore wind power forecasting-A hybrid model based on Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and deep-learning-based Long Short-Term Memory (LSTM). Renew Energy 185:611–628

Zhang Y, Zhang J, Yu L et al (2022b) A short-term wind energy hybrid optimal prediction system with denoising and novel error correction technique. Energy 254:124378

Zhang Z, Wang J, Wei D et al (2023) A novel ensemble system for short-term wind speed forecasting based on two-stage attention-based recurrent neural network. Renew Energy 204:11–23

Zhao Z, Yun S, Jia L et al (2023) Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. Eng Appl Artif Intell 121:105982

Zhen H, Niu D, Yu M, Wang K, Liang Y, Xu X (2020) A hybrid deep learning model and comparison for wind power forecasting considering temporal-spatial feature extraction. Sustainability 12(22):9490

Zheng Z, Zhang Z. (2023) A Stochastic Recurrent Encoder Decoder Network for Multistep Probabilistic Wind Power Predictions. IEEE Transactions on Neural Networks and Learning Systems, 2023.

Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W (2021) Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceed AAAI Conf Artif Intell 35(12):11106–11115

Zhou Y, Wang J, Lu H, Zhao W (2022b) (2022a) Short-term wind power prediction optimized by multi-objective dragonfly algorithm based on variational mode decomposition. Chaos, Solitons Fractals 157:111982

Zhou T., Ma Z., Wen Q., Wang X., Sun L., Jin R. (2022) FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. Proceedings of the 39th International Conference on Machine Learning 162: 27268–27286.

Zu X, Song R (2018) Short-term wind power prediction method based on wavelet packet decomposition and improved GRU. J Phys: Conf Ser 1087(2):022034

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