Transactions on Soft Computing

www.tsc.reapress.com

Trans. Soft. Comput. Vol. 1, No. 2 (2025) 149–167.

Paper Type: Original Article

Application of Deep Learning Models in Wind Farm Power Output Forecasting

Mojtaba Heydari Vahed[1], Mahdieh Ghazvini[2,*], Fahimeh Ghasemian[2]

[1] Department of Electrical Engineering, Shahid Bahonar University, Kerman, Iran; mojtaba@eng.uk.ac.ir.

[2] Department of Computer, Shahid Bahonar University, Kerman, Iran; mghazvini@uk.ac.ir; f_ghasemian@uk.ac.ir.

Citation:

Received: 11 February 2024 *Revised: 24 March 2024* *Accepted: 08 August 2024*	Heydari Vahed, M., Ghazvini, M., & Ghasemian, F. (2025). Application of deep learning models in wind farm power output forecasting. *Transactions on soft computing, 1*(2), 149-167.

Abstract

Wind energy stands out as one of the fastest-growing clean energy sources worldwide, driven by global commitments to decarbonization, enhanced energy security, and sustainable development. Accurate forecasting of wind farm power output thus becomes a key prerequisite for better harnessing renewables and optimizing grid management. The non-stationary, stochastic, and multi-scale nature of wind speed and direction poses serious limitations for traditional forecasting models. In recent years, the rise of deep learning has sparked a major revolution in time series prediction, particularly outperforming classical statistical and machine learning methods in wind energy forecasting. This paper analyzes recent research on wind power prediction, focusing on deep learning architectures, spatial-temporal models, physics-informed approaches, and hybrid designs. It also reviews data sources such as Supervisory Control And Data Acquisition (SCADA), Meteorological Data (Meteo), Numerical Weather Prediction (NWP), remote sensing, and multi-source structures. Beyond offering a conceptual classification, quantitative and qualitative comparisons of methods, analysis of research challenges, computational complexity reviews, and identification of knowledge gaps, the study outlines future research directions. Findings reveal that attention-based models excel at capturing inter-turbine relationships, while decomposition-driven hybrids like Wavelet–LSTM, VMD–BiLSTM, and EMD–GRU deliver the highest accuracy and stability. Persistent challenges include data heterogeneity, spatial-temporal wind variations, the need for probabilistic forecasts, lack of interpretability, and demand for lightweight architectures.

Keywords: Wind power forecasting, Deep learning, Decomposition methods, Numerical weather prediction.

1 | Introduction

The prediction of wind power generation is one of the most important pillars of intelligent energy network management in today's world. Energy production from wind is inherently unstable, and this characteristic makes power generation planning and network control highly dependent on very accurate forecasting models. In recent years, attention to wind energy as one of the key renewable energy sources has increased significantly

Corresponding Author: mghazvini@uk.ac.ir

https://doi.org/10.48314/tsc.v1i2.49

in many countries due to the growing environmental crises, the depletion of fossil fuel resources, and the ever-increasing demand for energy. Wind energy, resulting from temperature differences at the Earth's surface and the Coriolis effect, is considered a clean and renewable resource; however, its variable and unpredictable nature poses serious challenges to ensuring the stability, reliability, and efficiency of power generation systems. Moreover, issues such as high initial investment costs, the fixed characteristics of some power plants, and difficulties in identifying suitable areas for wind exploitation remain major obstacles to the development of this technology.

Over the past few decades, installed wind power generation capacity has increased rapidly; for example, according to the global wind energy Council, there was a 59% growth in 2020 compared to the previous year alone. This rapid expansion, due to the random, nonlinear, and unstable nature of wind, has brought challenges such as instability in the balance and control of power grids. As the share of wind energy in the electricity grid increases, the need for accurate prediction of wind power generation becomes more pressing to ensure operational efficiency and system security.

Selecting appropriate methods for predicting and modeling power generation is one of the main challenges in this field. Traditional statistical and physical methods, due to limitations in modeling complexities and sudden fluctuations in weather conditions, often lack sufficient accuracy, leading to suboptimal operation and reduced economic efficiency of power plants. In recent years, deep learning has gained a significant position in this context, as it can identify complex and sometimes hidden patterns in wind data and use them for more accurate predictions.

The aim of this article is to provide a comprehensive review of the latest deep learning models in the field of wind power forecasting a review that not only introduces the methods but also analytically examines their strengths, limitations, and future research directions. In the following sections, Section 1 presents the fundamentals and nature of wind and the challenges of wind power prediction. Section 2 introduces traditional methods for forecasting wind power generation. Section 3 examines the forecasting time horizons for wind power plants. Sections 4 and 5 review various data types and deep learning architectures. Section 6 provides a review of deep learning-based research for predicting wind power generation, and Section 7 discusses the strengths and challenges of deep learning models in wind power prediction. Section 8 identifies research gaps, and finally, Section 9 concludes the article and outlines future research directions.

1.1 | Nature of Wind and Wind Power Forecasting Challenges

Wind energy ranks as a vital renewable resource, harnessed from air currents in Earth's atmosphere. It arises from uneven surface heating and the Coriolis force. Due to its abundance and cleanliness, wind energy has gained prominence as a primary option for global energy supply. Harnessing wind power brings numerous benefits, including:

I. Reduced reliance on fossil fuels: wind serves as a clean, renewable source that cuts fossil fuel use and greenhouse gas emissions.

II. Lower air pollution: unlike fossil fuels, wind turbines emit no pollutants, thereby improving air quality.

III. Sustainability: wind offers an inexhaustible and steady resource, in contrast to finite fossil energy sources.

IV. Job creation: the wind energy sector generates new employment in turbine manufacturing, installation, and maintenance.

V. Despite these advantages, accurately forecasting wind farm output proves challenging due to wind speed's random fluctuations and other factors. Such precision plays a pivotal role in grid operations planning, management, and integrating wind farms with other energy sources.

Wind power forecasting may seem straightforward at first glance, but it involves complex, multifaceted issues. Wind's inherently unstable and chaotic nature leads to changes over very short timescales, with spatial-temporal patterns hard to capture using linear or classical statistical models. Moreover, factors like local topography, turbine height, ambient temperature, microscale airflow traits, and turbine wake interactions all influence output, amplifying the problem's difficulty. A key aspect of these challenges is that even within one wind farm, turbines just tens of meters apart can experience distinct wind speed and direction patterns. Overlooking such local-scale differences can yield substantial errors in power estimates. Thus, forecasting models must adeptly handle intricate nonlinear relationships alongside temporal and spatial dependencies. Yet system complexity and data constraints persist as major hurdles. These can be categorized as follows:

Inherent instability and variability of wind: wind speed and direction vary sharply under meteorological, climatic, and seasonal influences. Instantaneous shifts from pressure fronts, convection currents, or jet streams render wind behavior modeling and output prediction inherently uncertain and tough.

Accuracy dependence on forecasting horizon: short-term forecasts typically achieve higher precision; however, extending the horizon amplifies uncertainty from unpredictable weather shifts. Selecting the right timeframe—whether for operations, grid planning, or electricity markets—carries special weight.

Complexity in modeling environmental and local factors: topography, surface roughness, natural obstacles, and microclimatic structures profoundly shape wind dynamics. Accurately modeling them demands high-resolution local data and airflow dynamic simulations, requiring vast datasets and intensive computations.

Data limitations and meteorological access issues: in many areas, wind data features sparse spatial coverage and long intervals. This scarcity forces models to rely on interpolated or simulated data, introducing significant error sources.

Challenges in grid integration of wind power: wind farm output naturally fluctuates, potentially disrupting grid stability through frequency variations, heightened storage needs, or backup plant reliance. Managing these demands strategies like energy storage systems, complementary generation, and smart load control. Given wind's behavioral complexity and the broad influences on turbine output, traditional statistical methods alone fall short for modern energy systems. Lately, machine learning and deep learning approaches have emerged as powerful tools, excelling at extracting hidden patterns, nonlinear ties, and spatio-temporal interactions with greater fidelity. These strengths enable smarter models to better cope with rapid wind shifts and weather uncertainties, yielding more precise and reliable predictions. Examples include artificial neural networks, fuzzy systems, and hybrids like particle swarm optimization with adaptive time-series modeling—advances that have markedly boosted wind power forecast accuracy. Wider adoption of deep learning, especially leveraging multi-source data and multilayer structures, promises to cut errors and enhance forecasting efficiency.

In essence, achieving accurate wind farm power forecasts goes beyond technical necessity; it underpins operations planning, electricity market management, grid stability, and reserve cost reductions. Thus, collaboration among climatologists, energy engineers, data miners, and AI experts in crafting hybrid, interdisciplinary models pave the way for robust, efficient forecasting systems.

2 | Traditional Methods for Wind Farm Power Output Forecasting

Forecasting wind speed and estimating wind farm power output has long stood as a fundamental and daunting topic in renewable energy. Given wind's chaotic, nonlinear, and multi-scale behavior, researchers over recent decades have developed a wide array of modeling and prediction approaches. Broadly, these fall into four main categories: physical models, time-series statistical models, classical machine learning methods, and deep learning models. Across all these, wind speed emerges as the primary and most influential input parameter for turbine output estimation, owing to the highly nonlinear and sensitive relationship between wind speed and generated power. Alongside speed, variables like wind direction, air density, pressure, temperature, and measurement or hub height also play key roles in boosting forecast accuracy and typically feature in more advanced models [1].

Before data-driven modern techniques took hold, wind speed prediction and wind farm power estimation relied mainly on a set of traditional models. Encompassing physical, statistical, hybrid, and certain classical signal processing techniques, these served for years as the core tools for analyzing wind behavior and laid crucial groundwork for today's advanced approaches. Modeling here either drew on physical laws governing airflow or leveraged historical patterns and time-series trends. While they delivered acceptable performance in their era, limitations such as assuming linear wind behavior, struggles with sharp fluctuations, and demands for precise environmental data gradually gave way to more sophisticated data-centric methods [1].

Among traditional methods, physical models held a special place. Primarily including Numerical Weather Prediction (NWP) models and microscale ones like WAsP and CFD-based approaches, these seek to simulate local or regional wind patterns by solving governing equations for fluid dynamics and atmospheric conditions. They suit medium- and long-term forecasts and aid wind farm development planning projects. That said, they demand high computational power, highly accurate input data, and intricate parameter tuning; plus, their precision at very local scales—like turbine hub height—often remains constrained [1–3].

Another traditional group comprises time-series statistical models, long used for short-term wind speed prediction. Models like AR, MA, ARMA, ARIMA, and their seasonal variant SARIMA rest on the assumption that short-term wind behavior follows linear patterns and correlations from past data. Their simplicity, fast execution, and high interpretability made them hugely popular; yet they falter against wind's nonlinear, chaotic, and multi-scale traits. Moreover, for noisy or non-stationary data—common in Supervisory Control And Data Acquisition (SCADA) systems—they offer limited performance and usually require heavy preprocessing and differencing to stationaries the series [1], [4–6].

Besides statistical and physical models, hybrid or combined methods emerged, blending physical model outputs with statistical algorithms or filters like Kalman to cut prediction errors. These somewhat offset atmospheric model uncertainties but still hinged on input data quality and weather model accuracy. Meanwhile, classical signal processing techniques such as wavelet transforms or Kalman filters helped denoise, separate frequency levels, and refine data quality in forecasting pipelines. Useful for data prep though they were, their capacity to learn deep, complex patterns stayed limited [7–9].

Table 1 outlines the categorization and traits of traditional wind and wind farm power forecasting methods. Overall, despite their foundational role in wind prediction literature, these traditional approaches—tied to simplifying assumptions, inability to model nonlinear structures, and needs for precise environmental data—failed to deliver the accuracy and flexibility required for operational wind farm forecasts. These very shortcomings paved the way for classical machine learning and, especially, deep learning methods, which offer researchers far superior modeling of complex, nonlinear, and multi-scale relationships.

Table 1. Categorization and characteristics of traditional wind and wind farm power forecasting methods.

Category	Example Models	Features and Advantages	Limitations	Common Applications
Physical models	NWP models (GFS, ECMWF, WRF), microscale models like WAsP, CFD models	Based on physical airflow equations; suited for medium- and long-term forecasts; able to simulate weather conditions and topography	High computational demands; limited accuracy at local scales (turbine hub height); reliant on input data quality	Wind farm development planning, long-term analysis, wind potential assessment
Time-series statistical models	AR, MA, ARMA, ARIMA, SARIMA	Simple, interpretable, fast to run; good for short-term forecasts in fairly stable conditions	Struggles with nonlinear and chaotic behavior; noise-sensitive; requires series stationarities; poor on SCADA data	Short-term wind speed prediction, past trend analysis

Table 1. Continued.

Category	Example Models	Features and Advantages	Limitations	Common Applications
Time-series statistical models	AR, MA, ARMA, ARIMA, SARIMA	Simple, interpretable, fast to run; good for short-term forecasts in fairly stable conditions	Struggles with nonlinear and chaotic behavior; noise-sensitive; requires series stationarities; poor on SCADA data	Short-term wind speed prediction, past trend analysis
Hybrid methods	NWP combined with Kalman filter, statistical models + adaptive filters	Somewhat reduces errors by blending models; offsets certain uncertainties	Still tied to base model accuracy; added complexity; affected by weather model errors	Short- and medium-term forecasting; uncertainty reduction in physical models
Classical signal processing methods	Wavelet transform, Kalman filter, frequency filters	Improves data quality; noise removal; frequency component extraction	Lacks ability to learn complex patterns; useful only for data prep	SCADA data preprocessing, noise reduction, wind signal analysis

3 | Forecasting Horizons for Wind Farm Power Output

Wind turbine power forecasting occurs across various time horizons, depending on power grid operational needs, electricity market planning, and energy resource management. Generally, the forecasting horizon plays a decisive role in selecting the model type, input data, and expected accuracy level. Based on common classifications in scientific studies and power industry applications, wind power prediction falls into four main ranges.

In the ultra-short-term horizon, the goal is to predict instantaneous power changes over 1 to 30 minutes. These forecasts prove vital for controlling rapid power fluctuations, grid protection, reactive power regulation, and real-time turbine commands. The main challenge here lies in wind's chaotic behavior and high noise in SCADA data, demanding models with quick response and high sensitivity to moment-to-moment shifts.

The short-term horizon spans roughly 30 minutes to 6 hours and finds widest use in generation planning, energy storage management, and electricity market decisions. Here, forecast accuracy directly affects grid balancing costs and plant performance optimization. Models for this range must adeptly extract short-term patterns and temporal dependencies in wind data.

The medium-term horizon, typically covering 6 to 24 hours, supports next-day plant planning, grid resource management, and power system stability analysis. It relies heavily on NWP models, as large-scale atmospheric changes dominate tomorrow's output. Blending physical and data-driven models often yields better results in this window.

Finally, the long-term horizon stretches from one day to a week or more, mainly serving overall energy management, economic planning, annual production estimates, and operational strategy design. Its key challenge stems from wind behavior's strong ties to large-scale weather conditions and macro-atmospheric phenomena, heightening prediction uncertainty. Physical and bulk statistical models take center stage here, with accuracy typically lower than shorter horizons.

Overall, picking the right model for each horizon hinges on data nature, grid operational demands, and tolerable uncertainty levels—each brings its own unique challenges, applications, and computational requirements. Deep learning models outperform all others mainly in ultra-short- and short-term forecasting, while in longer horizons they usually integrate with NWP data.

4 | Types of Data Used in Wind Farm Power Output Forecasting

Forecasting wind turbine power output hinges heavily on the quality, variety, and precision of input data, given the complex, multi-scale nature of wind behavior and turbine performance under environmental influences. In research and industrial applications, data for wind power prediction typically sorts into four main groups: SCADA data, Meteorological Data (Meteo), NWP model data, and remote sensing data. Each source carries its own traits, constraints, and uses, and smartly combining them proves decisive in boosting accuracy for deep learning and data-driven models.

4.1 | Supervisory Control And Data Acquisition Data

SCADA data stands as the most common and valuable for short-term and ultra-short-term wind power forecasting, as it mirrors turbine behavior at the local level with high temporal resolution. This includes details like wind speed and direction, real-time output power, yaw and pitch status, generator and equipment temperatures, structural vibrations, and control system states. Sampling occurs at 1- to 10-minute intervals, making it ideal for deep learning models that thrive on frequent, continuous inputs. That said, SCADA data often grapples with heavy noise, outliers, missing values, and strong ties to each turbine's specifics. These issues double down on the need for advanced preprocessing like noise detection and removal, wavelet-based filters, outlier spotting, and missing data imputation techniques.

4.2 | Meteorological Data

Meteo complements SCADA by aiding short- and medium-term wind variation modeling. It covers variables such as temperature, pressure, humidity, cloud cover, precipitation, and solar radiation. Pairing it with SCADA typically lifts forecast accuracy markedly, offering broader atmospheric context. Its value shines in two areas: smart grid predictions and models linking wind swings to environmental patterns.

4.3 | Numerical Weather Prediction Models Data

Numerical weather models rank among the most effective data sources for medium- to long-term forecasts. Tools like WRF, ECMWF, and GFS solve physical equations and blend observational data to generate atmospheric forecast maps at set spatial and temporal resolutions. NWP outputs come every three to six hours, featuring large-scale info like wind speed at various heights, surface pressure, temperature, and humidity. A main drawback is their coarser local resolution (e.g., at turbine hub height), so many studies tap deep learning to refine and downscale NWP results.

4.4 | Remote Sensing Data

Remote sensing sources—including LIDAR, SODAR, weather radars, and satellite data like Sentinel and MODIS—enable wind pattern measurements at diverse heights and distances from turbines. They prove especially handy for offshore wind farms, where uniform sea conditions and lack of surface obstacles allow highly precise data capture. High accuracy, vertical flow profiling, and wide-scale coverage mark their strengths, though high costs and environmental sensitivity pose challenges.

A comparison of data types for wind farm power forecasting appears in *Table 2*.

Table 2. Types of data used in wind farm power output forecasting.

Data Type	Features and Content	Advantages	Challenges/ Limitations	Common Modeling Applications
SCADA data	Wind speed and direction, generated power, operational parameters (pitch, yaw), equipment temperature and vibrations, turbine status	High temporal resolution (1–10 minutes), direct access, deep learning friendly, reflects real turbine behavior	High noise and outliers, turbine-specific dependency, missing data, heavy preprocessing needs	Short- and ultra-short-term forecasting, fault detection, turbine dynamics modeling
Meteorological data (meteo)	Temperature, pressure, humidity, cloudiness, precipitation, solar radiation, atmospheric parameters	Complements SCADA, better environmental representation, boosts model accuracy	Lower resolution than SCADA, station-turbine location mismatches	Short- and medium-term forecasting, SCADA+Meteo hybrids
NWP Data	Physical model outputs (WRF, ECMWF, GFS), wind speed/direction at heights, pressure, temperature	Suits medium- and long-term forecasts, wide access, broad spatial coverage	Limited spatial resolution, 3–6 hour updates, needs downscaling	Medium-term forecasting, deep learning error correction inputs, grid management
Remote sensing data (LIDAR, SODAR, radar, satellite)	Wind at various heights, atmospheric flows, surface roughness, cloud and terrain conditions	High hub-height accuracy, offshore suited, wide coverage	High costs, environmental sensitivity, advanced equipment needs	Wind profile estimation, NWP enhancement, site selection, short-term offshore forecasting

5 | Deep Learning Approaches for Wind Farm Power Output Forecasting

Deep learning is this advanced offshoot of machine learning that relies on building up data through layered, hierarchical models. The beauty of it is that these models pull out features automatically from basic ones right up to the more sophisticated stuff without you having to hand-craft them yourself. The whole concept goes back to the 1940s with the McCulloch-Pitts artificial neuron, but it's really taken off lately thanks to beefier computing power like GPUs, massive datasets, and smarter optimization techniques. Nowadays, you'll find deep learning everywhere: computer vision, natural language processing, speech recognition, robotics, bioinformatics, recommender systems you name it.

At its heart, deep learning runs on multilayer artificial neural networks. Think of them as bunches of processing units called neurons, stacked into input layers, hidden ones, and output layers. Each neuron throws in some nonlinearity via activation functions ReLU, Sigmoid, Tanh and that's what lets the model approximate really tricky functions. You train them using backpropagation, usually with optimizers like SGD, Adam, or RMSProp that tweak things via stochastic gradient descent.

When it comes to forecasting wind farm power, deep learning shines because it nails those intricate nonlinear patterns, tracks time and space dependencies, and blends data from all sorts of sources. Stuff like wind speed, direction, air density, pressure, temperature, and turbine height are the key inputs that help predict those short- and medium-term swings in output. Here's a rundown of the main deep learning families making waves in wind power prediction:

I. Feedforward Neural Networks (FNN): the straightforward ones where data just flows input to output, no loops—great for basic regression jobs.

II. Convolutional Neural Networks (CNN): they use convolution and pooling layers to grab spatial features. For wind, they capture how nearby turbines influence each other, perfect for short-term power guesses.

III. Recurrent Neural Networks (RNN, LSTM, GRU): built for sequences and time series. LSTM and GRU are champs at remembering long-term wind wobbles, making them solid for short- and medium-term forecasts.

IV. Temporal Convolutional Networks (TCN): with their dilated convolutions that widen the receptive field, they chew through long time series fast and often beat LSTMs on accuracy in studies.

V. Attention and transformer models: these handle far-flung dependencies in time series and level up predictions by mixing SCADA with NWP data. Transformers are stars for medium- and long-term stuff, but they guzzle compute resources.

VI. Hybrid models: mashups like CNN-LSTM, Wavelet-LSTM, or Transformer-CNN crank up accuracy and tame noise, ideal for messy multi-source environmental data.

VII. Graph Neural Networks (GNN): wind farm turbines don't act alone they're interconnected. GNNs map out that spatial layout and turbine interactions for whole-farm power predictions.

Table 3 gives a quick snapshot of each model's features, pros, cons, and best uses handy for picking the right one in research or real-world setups. Pulling all this together lets deep learning deliver spot-on wind power forecasts, smooth out fluctuations, and make renewable grids way more dependable.

Table 3. Types of deep learning models in wind power forecasting.

DL Model Category	Features	Advantages	Limitations	Common Applications
CNN	Grabs local ties in time-space data; convolution setup	Handles multi-scale and grid-like data well; spots local patterns	Struggles with super-long time dependencies	Forecasting clustered turbines, local wind flow breakdowns
RNN / LSTM / GRU	Time-series pros with long memory; tracks time links	Manages swings and extended dependencies; nails short-term accuracy	Training drags on; compute-heavy; classic RNNs hit vanishing gradients	Short/medium-term wind speed/power calls, SCADA series digs
TCN	Dilated convolutions with growing reach; long inputs no sweat	Trains quicker than RNNs; edges out on accuracy sometimes	Fussy about architecture and kernel tweaks	Long-haul time series; medium-term predictions
Attention/ Transformer	Attention for distant links; blends SCADA-NWP seamlessly	Masters far-off series dependencies; top-tier accuracy	Data hog and compute beast; pretty complex	Sharp wind power forecasts, medium/long-range, multi-data mixes
Hybrid models (CNN-LSTM, wavelet-LSTM, transformer-CNN)	Model combos for peak performance; Wavelet preprocesses	Better accuracy; cuts noise; mixes the best of breeds	More complex; parameter tuning marathon	Precise short/medium-term; uncertainty busters
GNN	Maps space structures and turbine interplays	Tailor-made for farms; nails spatial turbine bonds	Wants exact layout data; compute intensive	Turbine group power estimates, farm tweaks

Deep learning has carved out a sweet spot in wind power forecasting by auto-extracting features from gnarly, multi-source data. They master nonlinear ties, long-haul dependencies, and wind's space-time dances with real precision. CNNs pull local and spatial nuggets from structured stuff like SCADA grids or images. RNNs, LSTMs, and GRUs handle wind/power series thanks to their memory for long dependencies. Transformers with attention gobble distant series and multi-source links, amping accuracy alongside SCADA/NWP. Autoencoders and GANs focus on data reps, denoising, and fake data gen, while hybrids like CNN-LSTM or Wavelet-LSTM fuse space-time powers for better results. GNNs tackle turbine interlinks in farms. *Table 3* breaks down inputs, feature pulls, train speeds, data needs—each setup fits data type and goal like a glove.

5.1 | Data Structures in Deep Learning Models

One key factor in the success of deep learning models for wind power forecasting lies in how input data gets organized and structured. Deep networks can only pull-out meaningful features when data has been prepped in a way that fits the model's architecture. In wind farm applications, data typically comes from diverse sources, each with its own structure and nature. That's why understanding data types and picking the right input format plays such a crucial role in landing accurate, stable predictions. This section covers the main data structures used in deep learning models.

I. 1D time series data: the most common format for wind power forecasting is one-dimensional time series, usually sequences of wind speed, output power, or key SCADA variables. This setup meshes directly with architectures like LSTM, GRU, and TCN. RNN-based models and their evolved versions shine at spotting short- and long-term dependencies, with strong skills in picking up discontinuous, nonlinear patterns in wind data. This data type suits short-term and ultra-short-term forecasts.

II. Multivariate inputs: in many cases, a single variable won't cut it for wind power prediction you need a mix from SCADA, Meteo, and NWP outputs all at once. Here, model inputs turn multivariate. This lets the network learn ties between physical parameters like wind speed, direction, pressure, humidity, temperature, and turbine traits. Deep learning models then extract complex cross-variable features, markedly lifting prediction accuracy.

III. Spatial-temporal grids: for wind turbines, especially in large farms, spatial links between turbines matter as much as temporal ones. Spatial-temporal data often shows up as grids, enabling simultaneous modeling of spatial and temporal dynamics. These form the backbone for advanced models like CNNs, ConvLSTMs, spatial-temporal Transformers, and GNNs. GNNs in particular excel at analyzing turbine correlations, grasping the physical layout of a wind farm. This structure adds huge value for forecasting in tricky setups, like offshore farms.

IV. Multi-modal data: some cutting-edge models pull in data from varied modules—satellite images, SCADA time series, tabular data, or turbine connection graphs all combined. Known as multi-modal or multi-source data, this demands hybrid network setups: CNN layers for images, LSTM for time series, and fully connected or GNN layers for tabular or graph data. These architectures handle multilevel relationships and work great for complex, long-term predictions.

5.2 | Role of Preprocessing in Deep Learning Model Performance

Preprocessing data stands as one of the most critical steps when building deep learning models for wind power forecasting. Data quality directly shapes model accuracy, stability, and generalizability. SCADA system data often comes loaded with noise, outliers, missing values, and turbine-specific inconsistencies, so preprocessing isn't just about cleaning it's key to boosting the model's learning power.

A first step involves weeding out outliers, typically caused by sensor glitches, blade icing, or control disruptions. Next comes filling in missing values, often via simple averaging, interpolation, or more sophisticated techniques like autoencoders, KNN-imputation, or GANs. After that, normalization scales variables to a common range, speeding up and sharpening the optimization process.

Equally vital is pulling out physics-statistical features such as moving averages, standard deviations, speed gradients, or power-to-speed ratios. Picking the right time window matters a lot too, since input length affects how well the model picks up temporal patterns.

Finally, signal decomposition think Wavelet, VMD, EMD, or ICEEMDAN ranks among the most potent tools for enhancing feature quality. Pairing these with deep learning models, like Wavelet-LSTM or VMD-CNN-LSTM, consistently slashes Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) in recent studies. By breaking signals into frequency components, they help models better grasp behaviors across different timescales.

5.3 | Performance Metrics for Deep Learning Models in Wind Farm Power Forecasting

When it comes to gauging how well deep learning models predict wind farm power output, researchers turn to a standard set of statistical metrics. Each one sheds light on different sides of the error profile and prediction quality, and using them together gives a well-rounded view of a model's performance.

Root mean square error: RMSE tops the list as one of the most telling and sensitive metrics for forecasting tasks. It captures the gap between actual and predicted values by averaging the squared errors and then taking the square root. Because of that squaring step, it really flags big errors, making it perfect for models where major slip-ups could spell real operational trouble. A drop in RMSE signals sharper predictions and fewer wild error swings.

$$\mathrm{RMSE} = \mathrm{sqrt}\left(\left(\frac{1}{n}\right) * \Sigma\,(yi - \hat{y}i)^2\right). \tag{1}$$

Mean absolute error: MAE simply averages the absolute differences between predictions and actuals, without the squaring that amplifies outliers like RMSE does. That makes it more robust and steadier, especially when errors skew uneven or aren't normally distributed.

$$\mathrm{MAE} = (1/n) \times \Sigma\,|y_i - \hat{y}_i|. \tag{2}$$

It's straightforward to interpret and great for line-by-line comparisons across forecasting methods.

Mean absolute percentage error: Mean Absolute Percentage Error (MAPE) puts errors in percentage terms relative to actual values, which lets you compare models across datasets of different scales or units. The catch? It gets shaky when actual values dip near zero, potentially blowing up unrealistically.

$$\mathrm{MAPE} = (100/n) \times \Sigma\,|\frac{(y_i - \hat{y}_i)}{}y_i|. \tag{3}$$

Despite that, it remains one of the most intuitive and widely used yardsticks for deep learning performance.

Coefficient of determination (R^2): R^2 tells you what fraction of the real data's variance your model explains, by pitting its errors against a dead-simple baseline (just the data mean). Values range from 0 to 1 the closer to 1, the better the model mirrors the true data patterns.

$$R^2 = 1 - [\Sigma\,(y_i - \hat{y}_i)^2] / [\Sigma\,(y_i - \bar{y})^2]. \tag{4}$$

Here, $\bar{y}$ is the mean of actual values. High R^2 means the model does a solid job teasing out wind power's dynamic patterns.

Beyond these numbers, studies also weigh forecasting horizons (prediction timeframes) and model complexity (parameter counts, layers, compute demands) for a fuller take on real-world viability. Blending all these lets you stack up deep learning models head-to-head and pick the winner for actual wind farm use.

6 | Deep Learning Studies on Wind Farm Power Prediction

Over recent years, plenty of research has explored how deep learning models perform in forecasting wind farm power output. The table below sums up the key approaches.

Table 4. Summary of deep learning-based wind farm power forecasting methods.

Author/Year	Data & Inputs	Output	Architecture	Method Summary	Advantages	Drawbacks	Horizon	Metrics
[10]	SCADA (wind tower measurements) + NWP (12 variables), data from 3 farms in Hebei province (2018–2019)	Day-ahead wind power (MW)	DWT (denoise) → Autoencoder (feature compress) → BiLSTM (two-layer stacked)	Three-stage setup: DWT cleans noise, AE squeezes to 30D vector, bidirectional BiLSTM predicts; Adam training (lr=1e-3, batch=128)	Strong accuracy & stability across sites; PA ~90.7%; outperforms BP, LSTM	Private data; no public code; geographically limited	Day-ahead	NRMSE, MAE, MAPE, PA (e.g., PA boost 3.22–3.86%; MAE=8.27; MAPE=6.55)
[11]	SCADA; preprocessing: DWT / FSST / CNN; RF for feature ranking in some	Power (short-term)	Hybrid: DWT/FSST/CNN (feature) → BiLSTM / BiGRU (temporal)	Compared hybrid architectures; CNN+BiGRU leads ($R^2 \approx 0.9093$). 10-fold CV curbs overfitting	$R^2 \approx 0.91$; high accuracy; thorough feature engineering + sequence modeling	Incomplete dataset docs; sparse training/runtime details; reproducibility unclear	Short-term	$R^2 \approx 0.9093$; MSE/RMSE/MAE
[12]	SCADA two turbines (Xintianbao Wind Farm) — 10-min & 30-min intervals (1,000 samples/dataset)	Power (short-term)	Peephole LSTM + ISSA meta-optimizer (improved Sparrow Search: chaotic init, adaptive weights, Cauchy mutation, opposition-based learning)	ISSA tunes LSTM hyperparameters; Peephole-LSTM boosts stability/accuracy; vs ARIMA, SVM, PSO-LSTM; DM test for significance	Sharp RMSE/MAE/MAPE drops; Diebold–Mariano stats proof; detailed ISSA params	Small datasets (2×1000), private → limited repro; higher compute	Short-term (10–30 min)	RMSE/MAE/MAPE
[13]	Public 2018 wind-power dataset (high-dim features); preprocessing: interpolation, normalization; PLS for latent components	Power (day-ahead / short-term)	PLS (feature reduction) → LSTM	PLS extracts latent components from high-dim inputs then LSTM predicts; targets multicollinearity & overfitting	Greater stability, better training; RMSE=0.1524; R^2=0.558	Average R^2; incomplete runtime/hyperparams	Medium / day-ahead	RMSE=0.1524; R^2=0.558
[14]	SCADA (3 real-world datasets, varying volatility); multi-stage decomp: ICEEMDAN, SE, VMD	Power (short-term)	Decomposition → BiTCN (Bidirectional TCN) → Transformer (global dependency) hybrid	Breaks signal into subcomponents (ICEEMDAN, Sample Entropy, VMD), BiTCN per subcomponent, Transformer fuses/rebuilds	Top accuracy (MAPE 1.66–1.90%; R^2 up to 0.991); volatility-proof	High compute; private data/code → limited repro	Short-term	MAPE ≈1.66–1.90%; R^2 0.991
[15]	SCADA + weather vars (wind speed/dir, temp, humidity) — China data	Power (deterministic & probabilistic short-term)	Conditional Diffusion (generative) + Channel Attention (+ LSTM in pipeline)	Gradual generative diffusion for conditional prediction, channel attention weights features; outputs uncertainty distributions	Superior uncertainty handling; tops LSTM/CNN	High compute from diffusion sampling; long training	Short-medium	RMSE, R^2

Table 4. Continued.

Author/Year	Data & Inputs	Output	Architecture	Method Summary	Advantages	Drawbacks	Horizon	Metrics
[16]	SCADA historic (day-ahead); GAN for extreme/rare samples	Day-ahead power	cGAN / GAN augmentation → LSTM predictor	GAN creates extreme/weather-shift samples to train LSTM; aims for rare-condition generalization	RMSE/MAPE down ~10–15% vs baselines; stronger on non-stationary	GAN training unstable; extended training time	Day-ahead	RMSE & MAPE
[17]	10-min coastal wind (Inani Beach, Bangladesh; Jul 2014–Aug 2015)	Short-term wind speed/power proxy	CNN front-end → BiLSTM (bidirectional)	LSTM vs CNN-LSTM vs CNN-BiLSTM; latter best (RMSE=0.52 m/s; MAE=0.36; R^2=0.80)	Coastal-friendly; bidirectional context lifts accuracy	Single-site; limited hyperparams/runtime disclosure	Ultra-short/short	RMSE=0.52 m/s; MAE=0.36; R^2=0.80
[18]	SCADA + MERRA-2 meteo (Esenkoy, Turkey)	Turbine power/speed	CNN (feature) → GRU (temporal) hybrid (vs ANN, LSTM)	DL/ML comparison; CNN-GRU pulls spatial features + temporal modeling	Solid performance for complexity; competitive R^2/errors	Some high RMSEs (e.g., 75.24) flag data/processing sensitivity	Short-medium	RMSE, MMAPE
[19]	Hourly wind speed (Halifax) — March & July 2015	Wind speed (multi-step)	Modified LSTM with state-updating	Reuses internal states across steps for long-horizon stability; MATLAB training, scheduling, gradient clipping	Big RMSE cuts both seasons; better multi-step stability	Wind speed focus (not power); short dataset; no public code/data	Medium (multi-step wind speed)	RMSE: Spring=8.5128; Summer=4.7796
[12]	Xinjiang wind farm (35,040 samples ex) — meteo + SCADA; VMD decomp	Ultra-short power (minutes)	VMD → CNN/LSTM/Attention → GBDT fusion	Multi-stage: VMD frequency split, CNN features, LSTM/Attention temporal, GBDT final correction; MAE/RAE/MAPE & param details	Multi-scale denoising + ensemble ~10% accuracy gain; param docs	High compute; pipeline complexity & dense data needs	Ultra-short (1–60 min)	MAPE/MAE/MSE
[20]	SCADA (turbine-level) + weather features	Farm/single-turbine power (short-term)	Transformer-based (WPF-adapted)	Transformer model grabs long-term time-series dependencies via attention, SCADA-tuned; focuses on complex temporal patterns & point-error reduction	Long-term dependency modeling; beats LSTM samples	More data & compute vs simple RNNs	Short-term	RMSE/MAE
[21]	SCADA + sometimes NWP hybrids	Short-medium power	Powerformer (Transformer variant: sparse attention + LSTM embedding + gated residuals)	"Powerformer": Transformer special for power series—sparse self-attention cuts compute, LSTM time embedding, gated residual denoises	High accuracy; compute savings vs plain Transformer	Still complex; hyperparam-sensitive; needs impl tweaks	Short-medium	—

Table 4. Continued.

Author/Year	Data & Inputs	Output	Architecture	Method Summary	Advantages	Drawbacks	Horizon	Metrics
[22]	SCADA (wind farm; 1–10-min res)	Short-term power	Wavelet decomp + LSTM (or BiLSTM)	DWT splits/denoises signal first, selected features to LSTM for time-series prediction; SCADA noise removal & stability boost	Major noise reduction & RMSE/MAE gains on noisy data; simple, interpretable	DWT param picks (wavelet type, levels); multi-stage ops complexity	Short-term	RMSE
[23]	Multi-turbine SCADA + turbine layout/topography	Turbine/farm power (short-term spatial-temporal)	Spatial-temporal GNN (graph for spatial) + gated dilated inception (temporal)	Farm-as-graph model + gated dilated inception for local/long temporal; targets wake effects & turbine interactions	Precise spatial turbine relations; spatial output accuracy	Precise spatial data & graph engineering needs	Short-term spatial	—
[24]	SCADA (e.g., minute-level outputs)	Ultra-short power (1–15 min)	Conditional GAN (cGAN) for augmentation + prediction model (e.g., CNN–LSTM)	cGAN generates samples to fix shortages or imbalances; synthetic data trains the forecaster to boost ultra-short horizon accuracy	Greater stability & accuracy in low-sample or imbalanced cases; handy for ultra-short horizons	GAN training is tricky to optimize; poor training risks generating noisy samples	Ultra-short-term	—
[16]	SCADA and/or NWP (dataset-dependent)	Short-term power	Multiple Transformer versions (Time2Vec embedding, variants)	Reviews & compares Transformer architectures (with time embeddings) for WPF; offers implementation tips & hyperparam sensitivity analysis	Practical guide for picking/tuning variants; highlights attention-based superiority	Relies on ample data; comparisons vary by dataset	Short-term	RMSE/MAE

Recent studies make it clear that deep learning models, especially when blending multi-source data like SCADA, weather variables, NWP, and sometimes remote sensing, consistently outpace traditional methods think ARIMA, SVM, or linear regression in wind power forecasting accuracy. This edge shows up reliably when data quality and sample size are solid, or with solid preprocessing steps like decomposition or denoising [16], [27], [28].

LSTM and GRU pop up as strong baselines in tons of studies for short- and medium-term predictions, often beating traditional models, particularly with complex time series that carry long memory or when data's spotty/noisy. A few key caveats though: 1) the superiority really kicks in with big enough training sets, augmented samples, or proper preprocessing, 2) plain LSTM/GRU without tweaks like DWT, SG, or VMD can stumble on super-noisy data, and 3) in small single-site studies, gains might be marginal or just overfitting [12], [16], [29], [30].

CNNs, originally built for images, have found new life in recent work analyzing spatial and multi-sensor wind farm data. They pull out local wind shift patterns and turbine-to-turbine relationships. Pairing CNN with LSTM in CNN-LSTM setups lets you model spatial and temporal features at once, bumping accuracy noticeably. Plenty of examples with CNN-LSTM or CNN-BiLSTM prove that CNN's local feature grab plus LSTM's time memory sharpens ultra-short and short-term forecasts [11], [17]. It shines with multi-channel data (speed, direction, power, turbine params), but watch the architecture tuning and overfitting.

Transformer models with attention mechanisms have caught fire for longer-term and multi-source predictions. Research shows Transformers, especially fused with SCADA and NWP, lift accuracy on extended horizons by nailing distant dependencies. The evidence points to transformers (and variants like powerformer/informer/TFT) holding real advantages in medium- to long-term; they learn long-range links better than RNNs and pair with NWP to refine 1–24-hour forecasts. That said, the gains come with hefty compute costs and demands for big, labeled data [15], [21], [28].

Generative models like autoencoders and GANs have also stepped into wind power forecasting. Autoencoders compress data and strip noise to tease out useful prediction features, while GANs whip up synthetic wind samples to beef up training sets and dodge data shortages. Autoencoders often prep features for forecasters, proven to cut noise (e.g., DWT→AE→BiLSTM in [10]). GANs and diffusion models handle augmentation or rare/extreme cases, boosting generalization in tough spots but GAN training's finicky and unstable, and diffusion eats compute [16], [28].

Hybrid models have exploded in recent papers. Hybrids clearly show that mixing preprocessing wavelet, VMD, ICEEMDAN, feature extractors CNN, temporal blocks LSTM/GRU/TCN, and attention/Transformer layers beats any single piece. Wavelet/DWT + LSTM or ensembles slash RMSE/MAPE across studies [10], [14].

Looking at metrics like RMSE/MAE, horizons, data types, and model complexity reveals: 1) SCADA + NWP combos usually win, especially longer horizons, 2) hybrids often top single models, 3) CNN/TCN/WaveNet lead ultra-short, and 4) Transformers/attention variants rule medium/long. One caution: these are dataset-bound tied to reported metrics and setups so no blanket claims without study context [14], [17], [21].

Table 5's breakdown shows each deep learning family carving out its niche in wind power prediction. RNN-based models hold their foundational role for clean, single-purpose time series. CNNs and their LSTM hybrids excel at fast fluctuations and multi-scale patterns, offsetting classic RNN limits with spatial-temporal fusion. Decomposition+DL setups dominate accuracy by preprocessing signals to ditch SCADA noise. Transformers and attention architectures, surging lately, handle long-range complexities, perfect for medium-term SCADA-weather blends. GNNs push beyond single-turbine by modeling interactions. As farms grow, spatial and attention models will matter more. Finally, generative tools like GAN/diffusion tackle data scarcity and quality, though complexity and compute hold them back from industry scale.

Table 5. Deep learning model categories from reviewed papers.

Model Group	Sample Architectures	Strengths	Weaknesses	Suitable Time Horizon
RNN-based	LSTM, GRU, BiLSTM, Peephole-LSTM	Great for short time series, simple structure, works with limited data	Noise-sensitive, weak on long-range dependencies	5–60 minutes
CNN-based	CNN-LSTM, CNN-BiLSTM, MCNN	Pulls spatio-temporal features, handles messy data well	Poor long-term dependency handling	1–10 minutes
Decomposition +DL	DWT-LSTM, VMD-LSTM, ICEEMDAN-TCN-TR	Most accurate results, highly noise-resistant	High complexity, heavy compute	1–60 minutes
Transformer-based	Transformer, Attention, Diffusion+Attention	Long-term dependencies, high stability	Data-hungry	10–60 minutes
GNN-based	GNN, GAT-ForecastNet	Models turbine spatial dependencies	Limited for single turbines	10–60 minutes
Generative	GAN-LSTM, Diffusion	Boosts data quality, good for scarce samples	Tough training, expensive	—

Table 6 shows no single architecture nails every time horizon perfectly model choice has to match the specific timeframe. For ultra-short predictions (1–10 minutes), CNNs, TCNs, and WaveNets shine brightest since these windows deal with fine-grained, sudden swings, and convolutional networks excel at pulling local patterns from high-frequency data. As horizons stretch to 10–60 minutes, CNN-LSTM hybrids and Transformers take the lead they balance short-term patterns with longer temporal ties.

Table 6. Model performance across time horizons.

Time Horizon	Top Models	Explanation
1–10 minutes	CNN / MCNN / TCN / WaveNet / VMD-LSTM	Ideal for ultra-short forecasts, handles rapid fluctuations
10–60 minutes	CNN-LSTM / BiLSTM / transformer / decomposition models	Best accuracy in noisy conditions
1–6 hours	Transformer / Informer / GRU-Attention	Long-range dependencies and data fusion
12–48 hours	NWP + LSTM / NWP + Transformer / Diffusion models	Suited for day-ahead and longer-range

For medium-term (1–6 hours), Attention models, Informer, and attention-boosted RNNs stand out, as these spans hinge on weather patterns and long-range structures. Beyond that (12–48 hours), blending NWP with LSTM or Transformer becomes essential to properly model larger-scale environmental uncertainties. Overall, the table makes clear that longer horizons demand deeper learning capacity, less noise sensitivity, and multi-data integration skills.

Findings from *Table 7* highlight how wind power forecasting research has shifted toward more complex, multi-scale, spatial-temporal models. Transformers' rapid rise stems from their unmatched ability to handle long-range dependencies and tricky weather patterns. Meanwhile, widespread Wavelet and VMD preprocessing shows researchers grasp that input quality drives output—many studies use decomposition upfront to strip SCADA noise. Growing GNN focus reflects that turbines aren't solo acts; wind flow interactions, wake effects, and spatial ties matter hugely for output. Probabilistic models especially

Transformer and Diffusion variants directly tackle grid needs for 24–48 hour uncertainty estimates. Broad trends point from simple, one-dimensional setups to versatile, attention-driven, farm-layout-aware architectures.

Table 7. Key model trends from attached paper analysis.

Trend	Description
Transformer surge	Usage jumped from ~5% to over 30% in wind power forecasting over 3 years
Wavelet/VMD growth	In 30% of studies, noise reduction cut RMSE by 8–15%
GNN expansion	Strong gains in large farms (10+ turbines)
Probabilistic advances	Diffusion and Attention best manage uncertainty at 24–48 hours

State-of-the-art analysis confirms Transformers and hybrids set the gold standard, WaveNet/TCN excel in fast forecasts, and Wavelet with LSTM/Transformer outperforms on noisy data. GNNs are rapidly becoming a standard tool in large-scale wind farms.

7 | Strengths and Challenges of Deep Learning Models in Wind Power Forecasting

7.1 | Strengths of Deep Learning in Wind Power Forecasting

Deep learning models have swiftly taken over from traditional methods and even many classical machine learning ones in recent years, thanks to their knack for picking up on those intricate, nonlinear, chaotic wind-dependent patterns. One standout strength is automatic feature extraction—no need for hand-crafted engineering or picking physical indicators like in older models; the network just learns the variable relationships on its own.

These models also handle multi-source data beautifully SCADA, NWP, weather stats, even satellite images—uncovering shared patterns across them. That's why they've become the go-to for short- and ultra-short-term forecasts, where reacting to sudden wind shifts really counts. Another big plus is their ability to blend spatial and temporal dependencies; setups like ConvLSTM, spatio-temporal transformers, or graph networks (GCN/GNN) model entire turbine clusters in sync. This shines especially in offshore farms or multi-turbine setups.

On top of that, deep learning's flexibility in hybrid designs like Wavelet-LSTM, VMD-CNN-LSTM, or Transformer-GNN—delivers top accuracy in tons of studies, holding steady even in noisy, non-stationary conditions.

7.2 | Limitations and Challenges of Deep Learning

Despite their impressive wins in wind power prediction, deep learning models still face some hefty hurdles. First off, they crave massive, high-quality data; SCADA feeds often come riddled with noise, outliers, and gaps, tanking accuracy and demanding heavy preprocessing.

Training costs a fortune in compute power too especially transformers, ConvLSTMs, or graph nets that need beefy GPUs and heaps of memory. That's a real pain in time-sensitive industrial settings.

Generalization or domain shift is another headache. Models trained on one wind farm don't always play nice on another with different wind regimes, layouts, or controls, spotlighting the need for domain adaptation or transfer learning. Interpretability's a sore spot as well; many deep models act like black boxes, leaving grid operators and tech decision-makers in the dark on why a prediction landed where it did. Finally, while uncertainty's crucial in wind forecasting, probabilistic deep learning lags behind. Point forecasts aren't enough operators want uncertainty bands for scenario planning, but that's not fully matured yet.

7.3 | Research Gaps

Despite big strides in deep learning for wind power, a deep dive into the literature reveals some glaring holes. Top one: lack of interpretable models. Most like LSTM, Transformers, or hybrids behave like black boxes, unable to explain their decisions to grid operators or industry pros critical in power systems where trust and transparency rule. Second, too few physics-informed deep learning efforts. Current ones are purely data-driven, ignoring wind flow physics, atmospheric stability, turbulence, or turbine aerodynamics. That leaves them shaky outside training data ranges. Then there's the push for lightweight, industry-ready models. Fancy Transformers and graphs pack too many params, train forever, and guzzle hardware tough for real farms or control centers. Transferability's unsolved too. Models from one farm flop on another with unique geography, climate, or geometry, calling for more general, adaptable, domain-robust designs. Lastly, probabilistic approaches and uncertainty quantification are thin. Point predictions fall short for grid ops needing uncertainty ranges for planning; wind's not a single number but a probability distribution, and deep models haven't nailed that yet.

8 | Conclusion

In recent years, a wide range of experimental studies have shown that deep learning models especially when paired with solid preprocessing and multi-source data like SCADA, NWP, and sometimes remote sensing imagery sharply outperform traditional methods in wind power forecasting. In practice, various architectures have proven their worth: advanced RNNs (LSTM, GRU, BiLSTM) serve as reliable baselines for short- to medium-term predictions, while convolutional networks (TCN, WaveNet, CNN) shine particularly in ultra-short-term scenarios. Hybrid and multi-stage models (e.g., Wavelet/DWT → CNN → LSTM or VMD → CNN/LSTM → GBDT) tend to deliver the sharpest results, cutting noise while extracting spatio-temporal features. Meanwhile, attention-based and Transformer architectures suit medium- to long-term horizons and NWP-SCADA blending better, though they come with higher compute costs and data demands. GNNs and probabilistic models (GAN, diffusion) have also gained traction lately for mapping spatial relationships in large farms and handling uncertainty, despite practical hurdles like precise spatial data needs, compute intensity, and training stability [10], [14], [17], [28].

Looking at literature gaps reveals that despite deep learning's big leaps, wind power forecasting still holds major research opportunities to shape next-gen models. One key path is tailoring Transformer architectures for energy data. They've excelled in NLP and vision, but adapting them to noisy, non-stationary, multi-scale wind series calls for more work. Models like Informer, Autoformer, and FEDformer hint at sharper medium/long-term forecasts by trimming compute and extending memory. Another big avenue is ramping up graph networks GNNs for wind farm modeling. Turbines in a farm aren't isolated wind flow, turbulence, shadowing, and aerodynamics create spatial ties. GNNs, with their knack for network relations, naturally fit this structure and could elevate predictions from single turbines to whole-farm levels.

Multimodal learning is emerging as a cornerstone too. Merging SCADA, NWP outputs, satellite images, LiDAR/SODAR, and topography paints a fuller wind picture, letting models view the phenomenon from multiple angles for tougher, more accurate forecasts. On the applied side, industry's data privacy push is steering researchers toward federated learning. SCADA data stays confidential and can't move between farms, so federated setups train models distributedly without raw data shifts boosting security and generalization to new sites. Finally, one of the most exciting frontiers is physics-informed deep learning. These weave in physical knowledge from blade aerodynamics to boundary layer behavior and atmospheric stability into deep structures. The result? Models with top accuracy plus rock-solid stability in unknowns, crucial for industry.

Next-gen wind power forecasters will likely blend multi-source data, lightweight real-time architectures, better interpretability, and uncertainty analysis. Standardizing datasets and shared benchmarks will streamline comparisons and guide research. Overall, these paths promise an era where deep learning models aren't just sharper, but more reliable, adaptable, and ready for real-world use.

Author Contributions

Conceptualization, Heydari Vahed M., Ghazvini M., and Ghasemian F.; Methodology, Ghazvini M.; Software, Ghasemian F.; Validation, Heydari Vahed M. and Ghazvini M.; Formal analysis, Ghazvini M.; Investigation, Ghasemian F.; Resources, Heydari Vahed M.; Data maintenance, Ghasemian F.; Writing creating the initial draft, Ghazvini M.; Writing reviewing and editing, Heydari Vahed M. and Ghasemian F.; Visualization, Ghasemian F.; Supervision, Heydari Vahed M.; Project administration, Ghazvini M. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors gratefully acknowledge the reviewers and editorial members for their valuable observations and constructive recommendations, which greatly contributed to improving the scientific content, organization, and clarity of this article.

Funding

This study was carried out without any dedicated financial support from governmental, private, or non-profit funding organizations.

Data Availability

The data and materials analyzed in this study, including publicly available wind farm datasets and related computational outputs, are available from the corresponding author upon reasonable request.

References

[1] Schaffarczyk, A. P. (2024). *Introduction to wind turbine aerodynamics*. Springer Nature. https://doi.org/10.1007/978-3-031-56924-1

[2] Wang, X., Guo, P., & Huang, X. (2011). A review of wind power forecasting models. *Energy procedia*, *12*, 770–778. https://doi.org/10.1016/j.egypro.2011.10.103

[3] Maradin, D. (2021). Advantages and disadvantages of renewable energy sources utilization. *International journal of energy economics and policy*, *11*(3), 176–183. https://doi.org/10.32479/ijeep.11027

[4] Lange, M., & Focken, U. (2006). *Physical approach to short-term wind power prediction* (Vol. 208). Springer. https://doi.org/10.1007/3-540-31106-8%0A%0A

[5] Shadab, A., Said, S., & Ahmad, S. (2019). Box Jenkins multiplicative ARIMA modeling for prediction of solar radiation: A case study. *International journal of energy and water resources*, *3*(4), 305–318. https://doi.org/10.1007/s42108-019-00037-5

[6] Naderipour, M., Hosseini, H. A.-S., & Jamali, M. B. (2025). A novel model based on the ARIMA model in predicting stock prices of Tehran Stock Exchange companies. *System engineering and productivity*, *6*(3), e732324. https://doi.org/10.22034/sep.2025.2075221.1411

[7] Cassola, F., & Burlando, M. (2012). Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output. *Applied energy*, *99*, 154–166. https://doi.org/10.1016/j.apenergy.2012.03.054

[8] Grigonytė, E., & Butkevičiūtė, .E. (2016). Short-term wind speed forecasting using ARIMA model. *Energetika*, *62*(1–2), 45–55. https://doi.org/10.6001/energetika.v62i1-2.3313

[9] Jiao, J. (2018). A hybrid forecasting method for wind speed. *MATEC web of conferences* (Vol. 232, p. 3013). EDP Sciences. https://doi.org/10.1051/matecconf/201823203013

[10] Xu, P., Zhang, M., Chen, Z., Wang, B., Cheng, C., & Liu, R. (2023). A deep learning framework for day ahead wind power short-term prediction. *Applied sciences*, *13*(6), 4042. https://doi.org/10.3390/app13064042

[11] Vishnutheerth, E. P., Vijay, V., Satheesh, R., & Kolhe, M. L. (2024). A comprehensive approach to wind power forecasting using advanced hybrid neural networks. *IEEE access*, *12*, 124790–124800. https://doi.org/10.1109/ACCESS.2024.3450096

[12] Tang, J., Yue, G., Lv, J., & Yue, J. (2025). An ensemble learning model for ultra-short-term wind power prediction based on variational mode decomposition and adaptive weighting. *Natural resource modeling*, *38*(4), e70015. https://doi.org/10.1111/nrm.70015

[13] Daryabi, S., Srujana, N. D., & Rama Sudha, K. (2025). Partial least squares enhanced long short-term memory models for wind power forecasting. *Intelligent strategies for ICT* (pp. 325–338). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-96-5607-3_29

[14] Shi, J., Teh, J., & Lai, C. M. (2025). Wind power prediction based on improved self-attention mechanism combined with bi-directional temporal convolutional network. *Energy*, *322*, 135666. https://doi.org/10.1016/j.energy.2025.135666

[15] Peng, H., Sun, H., Luo, S., Zuo, Z., Zhang, S., Wang, Z., & Wang, Y. (2024). Diffusion-based conditional wind power forecasting via channel attention. *IET renewable power generation*, *18*(3), 306–320. 10.1049/rpg2.12825

[16] Liu, R., Song, Y., Yuan, C., Wang, D., Xu, P., & Li, Y. (2023). GAN-based abrupt weather data augmentation for wind turbine power day-ahead predictions. *Energies*, *16*(21). https://doi.org/10.3390/en16217250

[17] Dev, U., Ahmad, S., Uddin, M. R., Dhar, R., Mubarak, H., Hazari, M. R., … & Alshammari, O. (2025). Short-term wind speed forecasting using different lstm deep learning methods. *Innovations in electrical and electronics engineering* (pp. 199–211). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-9112-5_12

[18] Karakan, A. (2024). Predicting energy production in renewable energy power plants using deep learning. *Energies*, *17*(16), 4031. https://doi.org/10.3390/en17164031

[19] Elsaraiti, M., & Merabet, A. (2021). Application of long-short-term-memory recurrent neural networks to forecast wind speed. *Applied sciences*, *11*(5), 2387. https://doi.org/10.3390/app11052387

[20] Huang, S., Yan, C., & Qu, Y. (2023). Deep learning model-transformer based wind power forecasting approach. *Frontiers in energy research*, *10*, 1055683. https://doi.org/10.3389/fenrg.2022.1055683

[21] Wang, W., Yang, J., Li, Y., Ren, G., & Li, K. (2025). Data-driven deep learning model for short-term wind power prediction assisted with WGAN-GP data preprocessing. *Expert systems with applications*, *275*, 127068. https://doi.org/10.1016/j.eswa.2025.127068

[22] Liu, Z., Gao, W., Wan, Y. H., & Muljadi, E. (2012). Wind power plant prediction by using neural networks. *2012 IEEE energy conversion congress and exposition (ECCE)* (pp. 3154–3160). IEEE. https://doi.org/10.1109/ECCE.2012.6342351

[23] Wang, R., Qiu, H., Jiang, G., Liu, X., & Cheng, X. (2024). Class-imbalanced spatial--temporal feature learning for blade icing recognition of wind turbine. *IEEE transactions on industrial informatics*, *20*(8), 10249–10258. https://doi.org/10.1109/TII.2024.3393550

[24] Zhang, J., Zhao, Z., Yan, J., & Cheng, P. (2023). Ultra-short-term wind power forecasting Based on CGAN-CNN-LSTM model supported by Lidar. *Sensors*, *23*(9), 4369. https://doi.org/10.3390/s23094369

[25] Al-qaness, M. A. A., Ewees, A. A., Aseeri, A. O., & Abd Elaziz, M. (2024). Wind power forecasting using optimized LSTM by attraction–repulsion optimization algorithm. *Ain shams engineering journal*, *15*(12), 103150. https://doi.org/10.1016/j.asej.2024.103150

[26] Daenens, S., Verstraeten, T., Daems, P. J., Nowe, A., & & Helsen, J. (2024). Spatio-temporal graph neural networks for power prediction in offshore wind farms using scada data. *Wind energy science discussions*, *2024*, 1–19. https://doi.org/10.5194/wes-10-1137-2025

[27] Xue, L., Zhang, X., Jiang, W., Huo, K., & Shen, Q. (2023). A classification performance evaluation measure considering data separability. *Artificial neural networks and machine learning ICANN 2023* (pp. 1–13). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-44207-0_1

[28] Peng, S., Guo, L., Huang, H., Liu, X., & Peng, J. (2024). ForecastNet wind power prediction based on spatio-temporal distribution. *Applied sciences*, *14*(2), 937. https://doi.org/10.3390/app14020937

[29] Tang, F. (2025). Short-term wind power prediction based on improved sparrow search algorithm optimized long short-term memory with peephole connections. *Wind engineering*, *49*(1), 71–90. https://doi.org/10.1177/0309524X241257429

[30] Qureshi, S., Shaikh, F., Kumar, L., Ali, F., Awais, M., & Gürel, A. E. (2023). Short-term forecasting of wind power generation using artificial intelligence. *Environmental challenges*, *11*, 100722. https://doi.org/10.1016/j.envc.2023.100722