Proceedings of CONF-SPML 2027 Symposium: Intelligent Network Security and Machine Learning for Communication Systems
DOI: 10.54254/2755-2721/2027.MELB38462

A Review of the Accuracy and Computational Cost of AI Regression Models for Short-Term Wind Power Forecasting

Huazhang Liao

School of Electronic Information, Northwest University, Xi'an, China
501286645@qq.com

Abstract. Short-term wind power forecasting models must meet both accuracy requirements and computational limits. Published comparisons use different datasets, forecast horizons, and timing boundaries. In this focused narrative review, twelve publications on artificial intelligence (AI) regression models and their evaluation are examined. Support vector regression, random forests, XGBoost, and long short-term memory networks are the main model families discussed. Prediction errors and computational costs are compared within studies, and changes in input information are examined. Under the reported experimental conditions, lower prediction errors do not always require higher computational costs. Runtime rankings can also differ between training and inference. Meteorological inputs and data enrichment affect accuracy and the performance gaps between models. However, cost reporting remains incomplete. The evidence cannot establish whether these accuracy gains require longer computation times or more resources. A conditional screening procedure is therefore suggested. The forecasting task and input availability should be checked first. Parameter search, training, and inference costs should then be compared separately. Results should also be tested across several time windows. These recommendations can support model screening under operational constraints. Their applicability still requires validation at the target wind farm.

Keywords: wind power forecasting, artificial intelligence regression models, computational cost, input information

1. Introduction

Short-term wind power forecasts are used in grid dispatch and electricity trading. Wind power output is the electrical power generated by a turbine or wind farm. It changes with meteorological conditions. Forecast errors can affect subsequent operating decisions, so error reduction has received substantial attention [1]. Deployment also introduces time constraints. Frequent retraining consumes computing resources. A model may have a low test error but take too long to produce a forecast. Such a model may be unsuitable for the intended application. Accuracy and computational cost must therefore be considered together.

Artificial intelligence methods are widely used for wind power forecasting. Existing reviews discuss model structures, data processing, and computational efficiency [1, 2]. González Sopeña et al. proposed a benchmarking framework with consistent data, temporal resolution, and forecast

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.MELB38462

horizons [3]. These conditions matter when models are selected. General descriptions of algorithm strengths and weaknesses cannot establish which model should be deployed. In this review, numerical comparisons of error and runtime are restricted to comparable experiments within the same study.

The review examines how accuracy and computational cost are related in short-term wind power forecasting. It also considers how input conditions affect model performance. The task definition and cost boundaries are described first. Within-study comparisons are then examined, followed by evidence on meteorological inputs and data processing. No new forecasting algorithm is proposed. The analysis supports candidate screening under error requirements and resource limits. It also identifies reporting needs for more comparable experiments.

2. AI regression models and comparison criteria

2.1. Review scope and evidence organization

A focused narrative review was conducted using twelve research and review articles published between 2020 and 2026. The literature coverage is selective, and no statistical pooling was performed. The main scope is point forecasting of future wind power from hourly to day-ahead horizons. Some missing-data experiments include a 15-minute horizon. Those results are used only to discuss the effects of missing inputs. Wind speed forecasting and long-term energy forecasting are excluded from the main comparison. Studies that only fit contemporaneous inputs to power output are also excluded.

For empirical studies, the forecast target and input availability were examined. Data splitting, error metrics, and timing boundaries were also checked. Studies reporting both errors and runtimes were used to assess their relationship. Experiments with changed inputs were used to examine performance under different data conditions. Review articles were used to explain evaluation principles. Model descriptions and qualitative cost labels were excluded from numerical comparisons.

2.2. Representative regression models

Wind power forecasts are continuous values. The methods discussed here are therefore used for regression. Support vector regression (SVR) balances fitting error against model complexity. Nonlinear relationships can be represented through kernel functions. Random forests (RF) combine predictions from multiple trees. XGBoost adds trees sequentially to correct errors in the current ensemble. Both are tree ensemble methods, but their training procedures differ. Long short-term memory (LSTM) networks retain and update historical information through gated states. They belong to the recurrent neural network family [1].

These structures describe how information is processed. They do not establish a fixed performance ranking. Lagged power values can be supplied to SVR and tree models. LSTM networks have a sequence structure, but their results still depend on input windows, parameter settings, and training data. Linear regression and multilayer perceptrons trained by backpropagation can be used as simpler baselines. Backpropagation is a training method, so it does not define a single neural network architecture. LightGBM, DLinear, Transformer, and PatchTST are included in Table 1 as additional controls from one benchmark [4].

DOI: 10.54254/2755-2721/2027.MELB38462

2.3. Error metrics and computational cost

Mean absolute error (MAE) and root mean square error (RMSE) are common forecasting metrics. RMSE gives greater weight to large errors. For normalized RMSE (NRMSE), the denominator must be stated. Rated wind farm capacity is one possible denominator. Values cannot be pooled simply because the same metric name is used. Percentage errors can also become misleading when power output approaches zero [5]. Here, accuracy describes regression performance and should not be confused with classification accuracy.

Computational cost includes preprocessing, parameter search, final training, and inference. Training time measures the time needed to build the model. Inference time measures the time needed to generate predictions after training. These phases serve different operational needs. Parameter counts and operation counts describe complexity, but they cannot replace measured runtime, peak memory, or energy use. Runtime also depends on hardware, software, and batch size [2]. The timing boundary must be specified. Full-test-set inference time, for example, is different from per-sample prediction latency. Resource comparisons are meaningful only when the tasks and evaluation procedures are sufficiently comparable.

3. Literature comparison of accuracy and computational cost

3.1. Accuracy and runtime within the same task

The examined results do not show a fixed relationship in which more computation always produces lower errors. Liu et al. compared six models for day-ahead forecasting at the Sotavento wind farm [4]. Chronological data splits were used. Meteorological information came from the same source, and comparable tuning budgets were assigned. Table 1 presents hold-out NRMSE, final training time, and test-set inference time.

In this experiment, LightGBM had lower values than XGBoost for all three measures. DLinear required more training time than LightGBM but less inference time. Its prediction error was also higher. Thus, a runtime comparison must identify the computational phase being measured.

Table 1. Reported accuracy and computational time in a single-farm day-ahead comparison [4]

Model	Hold-out NRMSE (%)	Final training (s)	Test-set inference (s)
Ridge	12.8054	0.0047	0.0004 ± 0.0001
XGBoost	11.3381	2.060	0.0273 ± 0.0005
LightGBM	10.2326	0.563	0.0101 ± 0.0009
DLinear	11.1095	5.629	0.0079 ± 0.0008
Transformer	10.6688	8.894	0.0161 ± 0.0011
PatchTST	11.9398	8.945	0.0218 ± 0.0004

Note: Values are reproduced from Tables 1 and A7 of Liu et al. [4]. They were not measured in this review. Inference time covers the test set, not individual predictions. The reported ± values are retained. Their statistical definition and the number of repetitions are not specified in Table A7. Search time, peak memory, and energy use are excluded.

These results apply to the original experimental setting. Information sources were shared, but the inputs were organized differently across models. Historical power was not included. Model rankings also changed between the fixed hold-out split and rolling validation. Rolling validation reused

DOI: 10.54254/2755-2721/2027.MELB38462

parameters selected under the fixed split. The temporal independence of this tuning information requires further checking. Hardware and repeated timing procedures were insufficiently documented for full runtime reproduction. Different error metrics also selected different leading models [4]. Table 1 therefore supports a comparison for this single-farm case. It cannot establish a general advantage for tree models or deep learning models elsewhere.

Incomplete timing boundaries can change the interpretation. Jalali et al. reported optimization, training, and testing times separately [6]. Their hybrid method required 191 s for training, compared with 354 s for the reference LSTM. However, an additional 778 s was required for optimization. Reporting only training time would understate the full model-building cost. Hao et al. excluded classification time and total prediction time from their computational analysis of Poisson resampling RF [7]. Both studies show why timing boundaries must be checked. Their runtimes are not combined with Table 1 because the experimental tasks differ.

Particle swarm optimization (PSO) is an auxiliary search method, not a power forecasting model. An et al. used PSO to adjust the input weights and thresholds of an extreme learning machine [8]. AdaBoost was then used to construct the ensemble predictor. Performance changes reflect both the base model and the ensemble procedure. They cannot be attributed entirely to PSO. Search time should be recorded separately. An optimization method should be assessed through the accuracy gain and additional cost of its specific configuration.

3.2. Effects of input information and data processing

Model performance can change when input information is changed. Chen et al. studied five wind farms in the Norwegian Arctic region [9]. Power-only models were compared with models that also used numerical weather prediction (NWP) inputs. Errors were reduced overall when meteorological information was added at common forecast horizons. The improvements differed across farms and models. Performance gaps between models also changed [9]. Model families alone therefore cannot determine forecasting performance.

The power-only experiments covered horizons of 1 to 3 h. The multivariate experiments covered 1 to 6 h. Their mean errors across the full horizon ranges should not be compared directly. The study did not clearly establish whether time order was preserved throughout cross-validation. Costs before and after adding NWP were not reported [9]. The results support an assessment of input information, but they do not identify which model becomes computationally cheaper with NWP.

Input features must also be available when a forecast is issued. Chen et al. considered the NWP release schedule and an approximately two-hour publication delay [9]. Weather forecasts that are unavailable in operation can inflate offline performance if they are used during testing. Release times should therefore be checked before additional meteorological features are evaluated. More input columns do not necessarily provide usable new information.

Zhou et al. examined data enrichment for day-ahead forecasting at three wind farms [10]. Weather forecast errors, neighboring nodes, temporal weather features, and additional weather sources were incorporated. At one farm, XGBoost ACC increased from 69.1% to 72.6%. LSTM ACC increased from 62.0% to 72.3%. The gap between the models narrowed. ACC was defined as 100% minus capacity-normalized NRMSE, rather than classification accuracy [10]. These results show that an input change can alter the relative performance of models. A ranking obtained before enrichment may therefore be unsuitable after preprocessing is changed.

Training and inference times were not reported before and after enrichment. The computational cost of the accuracy gains cannot be quantified [10]. Full-dataset statistics were also used for standardization, so information from the test period may have entered preprocessing. Feature groups

DOI: 10.54254/2755-2721/2027.MELB38462

were added sequentially. Their independent contributions are therefore difficult to separate. These limitations affect the interpretation of the reported gains. The experiment links input changes to error changes, but it does not establish a complete relationship among data conditions, accuracy, and cost.

Missing inputs are another operational concern. Stratigakos and Andrianesis compared linear and neural predictors for a target wind farm [11]. Forecast horizons ranged from 15 min to 4 h, and missing-input durations were varied. Imputation and adaptive treatment had different relative performance under short and persistent interruptions. Results obtained with complete data cannot therefore represent performance during data outages. Measured computational costs were not provided for each condition. The neural predictor was also not an LSTM. These boundaries must be retained when the results are interpreted.

3.3. Conditional model selection and evidence limitations

Candidate models can first be screened against operational limits. Frequent model updates require measurement of the full retraining time. Tight forecast delivery deadlines require measurement of total prediction latency, including preprocessing. Models exceeding the error, runtime, or memory limits should be excluded first. The remaining options can then be compared. Under identical conditions, a model can be excluded if it has higher error and higher costs on every relevant measure. If different models lead on error, training time, or inference time, the strictest operational constraint should guide selection. This procedure is derived from the reviewed comparison conditions. It has not been validated through deployment across wind farms.

The selection should be checked again after input conditions change. More training samples do not guarantee useful accuracy gains. Sewdien et al. examined training data volume and artificial neural network configurations [12]. Within their tested range, larger training datasets did not produce consistent, substantial accuracy improvements. The finding is limited to those models and data. It cannot define a universal sample requirement for SVR, RF, or LSTM. It also cannot establish that smaller datasets always improve efficiency.

Only twelve publications were included, and their task settings differ considerably. Few experiments varied data conditions while measuring errors and complete computational costs together. The four main model families were not all compared under identical conditions in the reviewed evidence. Errors from different papers were therefore not pooled. Unreported runtimes were not inferred from model structures. The suggested screening procedure combines performance evidence with input availability and timing boundaries. It distinguishes findings suitable for preliminary screening from results that require new tests at the target farm.

4. Conclusion

This review examined prediction errors and computational costs of AI regression models for short-term wind power forecasting. In the within-task comparisons, lower errors did not always require higher costs. Runtime rankings could also differ between training and inference. Structural complexity alone was insufficient for model selection. Parameter settings and the computational phase being assessed both affected the comparison.

Meteorological inputs, enrichment, and missing-data treatment influenced model performance. Most relevant studies did not report complete costs under these changing conditions. A general relationship between resource investment and accuracy improvement cannot yet be established. Comparisons should use the same forecast task and inputs available in operation. Search, training, and inference costs should be reported separately. Several time windows should be tested to assess

DOI: 10.54254/2755-2721/2027.MELB38462

whether a ranking depends on one data split. These recommendations describe the usable scope of the evidence. They do not establish a universally superior model.

Future experiments should assess data conditions and computational costs together. Input configurations can be varied while forecast horizons and information availability are held constant. Errors, full training time, prediction latency, and peak memory should be recorded jointly. Tests should then be repeated at other wind farms. Such experiments can establish whether data improvements change the accuracy-cost relationship and provide firmer evidence for deployment decisions.

References

[1] Yang, Y., Lou, H., Wu, J., Zhang, S., & Gao, S. (2024). A survey on wind power forecasting with machine learning approaches. *Neural Computing and Applications, 36*, 12753–12773. https://doi.org/10.1007/s00521-024-09923-4

[2] Manzano, E. A., Nogales, R. E., & Rios, A. (2025). A systematic review of wind energy forecasting models based on deep neural networks. *Wind, 5*(4), art. 29. https://doi.org/10.3390/wind5040029

[3] González Sopeña, J. M., Pakrashi, V., & Ghosh, B. (2023). A benchmarking framework for performance evaluation of statistical wind power forecasting models. *Sustainable Energy Technologies and Assessments, 57*, art. 103246. https://doi.org/10.1016/j.seta.2023.103246

[4] Liu, J., Lu, Y., & Zou, G. (2026). A comparative study of day-ahead wind power forecasting models for a single wind farm under strict chronological splitting and unified hyperparameter tuning conditions. *Energies, 19*(12), art. 2784. https://doi.org/10.3390/en19122784

[5] González-Sopeña, J. M., Pakrashi, V., & Ghosh, B. (2021). An overview of performance evaluation metrics for short-term statistical wind power forecasting. *Renewable and Sustainable Energy Reviews, 138*, art. 110515. https://doi.org/10.1016/j.rser.2020.110515

[6] Jalali, S. M. J., Osório, G. J., Ahmadian, S., Lotfi, M., Campos, V. M. A., Shafie-khah, M., Khosravi, A., & Catalão, J. P. S. (2022). New hybrid deep neural architectural search-based ensemble reinforcement learning strategy for wind power forecasting. *IEEE Transactions on Industry Applications, 58*(1), 15–27. https://doi.org/10.1109/TIA.2021.3126272

[7] Hao, J., Zhu, C., & Guo, X. (2021). Wind power short-term forecasting model based on the hierarchical output power and Poisson re-sampling random forest algorithm. *IEEE Access, 9*, 6478–6487. https://doi.org/10.1109/ACCESS.2020.3048382

[8] An, G., Jiang, Z., Cao, X., Liang, Y., Zhao, Y., Li, Z., Dong, W., & Sun, H. (2021). Short-term wind power prediction based on particle swarm optimization-extreme learning machine model combined with Adaboost algorithm. *IEEE Access, 9*, 94040–94052. https://doi.org/10.1109/ACCESS.2021.3093646

[9] Chen, H., Birkelund, Y., Anfinsen, S. N., & Yuan, F. (2021). Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region. *Journal of Renewable and Sustainable Energy, 13*(2), 023314. https://doi.org/10.1063/5.0038429

[10] Zhou, Y., Ma, L., Ni, W., & Yu, C. (2023). Data enrichment as a method of data preprocessing to enhance short-term wind power forecasting. *Energies, 16*(5), 2094. https://doi.org/10.3390/en16052094

[11] Stratigakos, A., & Andrianesis, P. (2026). Learning data-driven uncertainty set partitions for robust and adaptive energy forecasting with missing data. *IEEE Transactions on Smart Grid, 17*(1), 806–817. https://doi.org/10.1109/TSG.2025.3620957

[12] Sewdien, V. N., Preece, R., Rueda Torres, J. L., Rakhshani, E., & van der Meijden, M. (2020). Assessment of critical parameters for artificial neural networks based short-term wind generation forecasting. *Renewable Energy, 161*, 878–892. https://doi.org/10.1016/j.renene.2020.07.117