

Received 29 October 2022, accepted 16 November 2022, date of publication 18 November 2022,
date of current version 30 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3223435

RESEARCH ARTICLE

A Hybrid Three-Stage, Short-Term Wind-Power Prediction Method Based on SDAE-SVR Deep Learning and BA Optimization

RUIQIN DUAN¹, XIAOSHENG PENG^{1,2}, (Member, IEEE), CONG LI^{1,2}, ZIMIN YANG²,
YAN JIANG¹, XIUFENG LI¹, AND SHUANGQUAN LIU¹

¹System Operation Department, Yunnan Power Grid Corporation Ltd., Kunming, Yunnan 650011, China

²State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Xiaosheng Peng (XiaoshengPeng@hust.edu.cn)


This work was supported in part by the Reserve Talents Program for Middle-Aged and Young Leaders of Disciplines in Science and Technology of Yunnan Province, China, under Grant 202105AC160014; and in part by the Science and Technology Program of Yunnan Power Grid Corporation Ltd., under Grant YNKJXM20210100.

ABSTRACT Wind power prediction (WPP) is necessary to the safe operation and economic dispatch of power systems. In order to improve the prediction accuracy of WPP, in this paper we propose a three-step model named SDAE-SVR-BA to be applied in short-term WPP based on stacked-denoising-autoencoder (SDAE) feature processing, bat algorithm (BA) optimization and support vector regression (SVR). First, we preprocessed the original NWP data input into the SDAE-SVR-BA model to adapt to the training and prediction of the proposed model. Second, we input the preprocessed features into the SDAE network, whose parameters are optimized by BA to obtain the depth-mapping features. Finally, we input the features of SDAE network mapping into SVR, whose parameters are optimized by BA for prediction, so as to obtain the SDAE-SVR-BA model. In this paper, we used BA during the training process to optimize the number of hidden layers and hidden layer nodes of SDAE, the penalty factor parameter C and the kernel function radius g of the SVR model. Additionally, we verified the model with a wind farm example and compared it to the traditional model. Based on the verification data applied in this article, in a forecast for the next twelve hours, the normalized root means square error (NRMSE) of SDAE-SVR was 11.97% and the NRMSE of SDAE-SVR-BA model was 11.54%, reduced by 1.24% compared with SDAE, which demonstrates the effectiveness of the proposed method.

INDEX TERMS Stack denoising autoencoder, bat optimization algorithm, wind power prediction.

I. INTRODUCTION

With the development of new energy, the installed capacity of wind power is increasing year by year worldwide [1]. Due to the intermittent, fluctuating and random characteristics of wind power generation, large-scale grid integration of wind power brings challenges to the safe and stable operation of the grid. Wind power prediction (WPP) is intended to predict the future output of wind power through weather forecast data, wind-farm operating status data and other parameters, which can improve the predictability of wind power, provide a basis

The associate editor coordinating the review of this manuscript and approving it for publication was Sajid Ali .

for grid operation and dispatch, and realize the safety and reliability of the grid run [2], [3], [4].

At present, there are two difficulties in WPP: 1) the volume of WPP input data is large and covers a large amount of information, making it difficult to fully mine effective information, and its feature mapping is required; 2) the WPP model is complex, and it is difficult to obtain the optimal model structure and parameters. It is necessary to apply efficient artificial intelligence algorithms to optimize the model structure and parameters.

In order to fully mine the effective information in the input data of wind power forecasting, deep neural networks need to be applied to extract features abstract.

The deep neural network adopts a multilayer structure that simulates the human brain, and it extracts features from the bottom to the top of the input data step by step, finally it forms more ideal features suitable for prediction. In addition, deep learning and machine learning algorithms have applications in the field of electricity markets prediction [5], [6].

In the literature [7], a convolutional neural networks (CNN) model is applied to predict wind energy, but the CNN model is more suitable for processing image information, and wind is typically time-series information. In the literature [8], a long short-term memory (LSTM) network is applied to analyze time-series data, which is more suitable for forecasting time-series data. LSTM has a memory function for historical data, so that this type of network can well fit the trend of wind power changes [9], [10], [11], [12], [13]. In the literature [10], specific methods of how to classify, identify, and predict are proposed, which are two-level clustering, CNN and LSTM, but the parameters of the neural network are not optimized and the accuracy of the model may not be the highest. In the literature [11], a LSTM recurrent neural network-based framework was proposed to predict the load and the accuracy is higher than other listed rival algorithms. However, the main part of the framework, LSTM, is not well described in this paper. In the literature [12], a novel reliability assessment strategy is proposed and the effectiveness of proposed assessment strategy is verified by real data. However, there is no comparison between the prediction results of other neural networks. To summarize, when the power sequence transitions from one trend to another trend or when predicting a longer time scale, the prediction results of the network will show a certain phase delay, resulting in errors [14].

In the literature [15], stacked autoencoders (SAE) are applied for feature dimensionality reduction. This method abstracts feature information through a layer-by-layer network. Through the unsupervised learning of data samples, the information contained in the original high-dimensional data can be restored to the maximum extent, which can effectively process nonlinear data and has stronger applicability [16]. This indicates that SDAE can be well used to process nonlinear time series.

Deep learning is currently a hot and cutting-edge method for power systems [17]. However, how to design the structure and parameters of the deep learning model has always been the bottleneck of its application. In the literature [18], the gradient-descent algorithm is applied to optimize the model parameters. The optimization idea of this method is to apply the negative gradient direction of the current position as the search direction. The closer the gradient-descent method is to the target value, the smaller the step size and the slower the progress. This method is simple to implement, but only is suitable for optimization problems where the objective function is convex.

In the literature [19], Newton's method is applied to optimize model parameters. Newton's method is a common method for solving unconstrained optimization problems.

This method is an iterative algorithm, and its convergence speed is faster than that of the gradient-descent algorithm, but each step needs to solve the inverse matrix of the Hessian matrix of the objective function, which makes the calculation more complicated.

In the literature [20], the bat algorithm (BA) is applied to optimize the model parameters. This algorithm is an optimization technique based on iteration. First, a set of random solutions are initialized, and then the optimal solution is searched through iteration, and a new local solution is generated by random flight around the optimal solution, which strengthens the local search [21]. Compared with other algorithms, BA is far superior in terms of accuracy and effectiveness, and not many parameters need to be adjusted.

Combining the advantages of deep learning methods and optimization algorithm, in this paper we propose a novel combined WPP model based on the SDAE deep learning method, BA optimization and support vector regression (SVR). We verified the model with data from one wind farm in a province of China. The results show that the proposed method has significant advantages in terms of the NRMSE and NMAE compared to traditional SDAE and SVR methods.

This paper first introduces the SDAE, SVR and BA algorithms applied in this paper, and what parameters of SDAE and SVR need to be optimized, and proposes the three-stage model SDAE-SVR-BA. This paper uses the wind farm data of a province, compares the prediction accuracy of different methods, different nodes, different hidden layers and different optimization methods, and finally proves that the proposed SDAE-SVR-BA has the highest accuracy.

II. THE COMBINATION METHOD OF SDAE, SVR AND BA

A. INTRODUCTION OF SVR APPLIED IN THIS PAPER

Support vector machines (SVM) are rooted in Vapnik-Chervonenkis and Structural Risk Minimization principles of statistical learning theory [22], [23], [24], which are widely applied in academia and industry. Compared to previous machine learning algorithms, SVM provides a more powerful method for dealing with nonlinear problems. The SVR applied in this paper refers to the application of SVM to linear regression. The difference between SVR regression and SVM classification is that the sample points of SVR have only one category in the end.

The penalty factor parameter C and kernel function radius g of the SVR model have great influence on the accuracy of the prediction model. The penalty factor C is applied to indicate the importance attached to individual points. The size of C should be moderately. If the penalty factor C is too large, SVR can process only linear samples, resulting in overfitting and poor generalization ability. Kernel function is a feature-transformation function, and for different distributed data, kernel radius g should be selected to achieve the best feature mapping. The comparison of model parameters before and after optimization is shown in Fig. 1. It can be seen from the figure that SVR after parameter optimization

has better linearity. At the same time, the model accuracy is also improved accordingly.

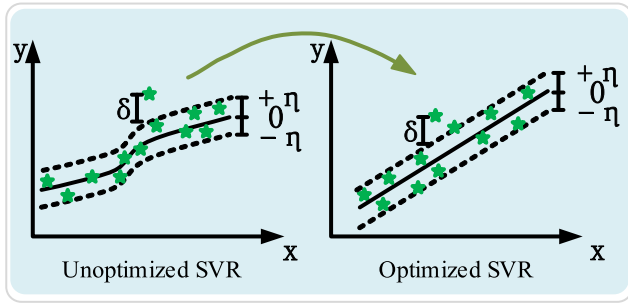


FIGURE 1. Comparison of model parameters before and after optimization.

B. INTRODUCTION OF SDAE APPLIED IN THIS PAPER

The SDAE proposed by Pascal Vincent et al. is an extension of the SAE. Its central idea is to add random noise at each input layer of the encoder model to train and learn more robust feature representation.

1) DENOISING AUTOENCODER

Autoencoder (AE) is an unsupervised learning neural network which reconstructs the target representation according to the input data; it is composed of an encoder and a decoder. The basic AE network structure is the same as the three-layer traditional neural network, including the input layer, the hidden layer and the output layer. The output layer and the input layer have the same dimension. The network structure is shown in the first part of Fig. 2. The learning goal of self-coding is $\tilde{x} = h_w, b(x) \approx x$, that is, to make the output vector \tilde{x} equal to the input vector x as far as possible. So AE tries to approximate the identity function to obtain a compressed representation of the input data at the hidden layer, which often better represents the original data.

Denoising autoencoder (DAE) adds noise to the training data of the AE. Noise can be added by using Gaussian noise or by randomly setting the input neuron value to zero (i.e., the dropout technique is applied). By adding random noise to the data applied in the AE training, the AE can be forced to learn how to remove the noise, thus obtaining data that is not polluted and destroyed. The DAE can find a more effective and stable feature representation in the case of corrupted or polluted input data, which is a more abstract and high-level representation of the original input data, thus enhancing the robustness of the whole model [10], [21].

The second part of Fig. 2 shows the principle of denoising training. The character x represents the original input data, x' represents the data obtained after being destroyed or polluted, y represents the characteristic representation obtained by encoding processing of x' , and z represents the output after the decoding of y .

The loss function is applied to represent the training effect of the DAE. To minimize the difference between the output

data and the original input, the loss function must be minimized. The DAE model enhances the robustness of the feature representation through the feature mapping of contaminated data during the training process.

2) THE BASIC STRUCTURE OF SDAE

SDAE are obtained by stacking the DAE together to obtain more advanced and abstract feature representations [16]. A model diagram of the SDAE is shown in the third part of Fig. 2. The structure of the AE is a neural network containing three layers, and the characteristic mapping between the input layer and the output layer is required after the training of a single AE. Multiple DAE are stacked to form SDAE with deep learning hierarchy structure for layer-by-layer training, and the outputs of the former DAE are applied as the input of the latter DAE.

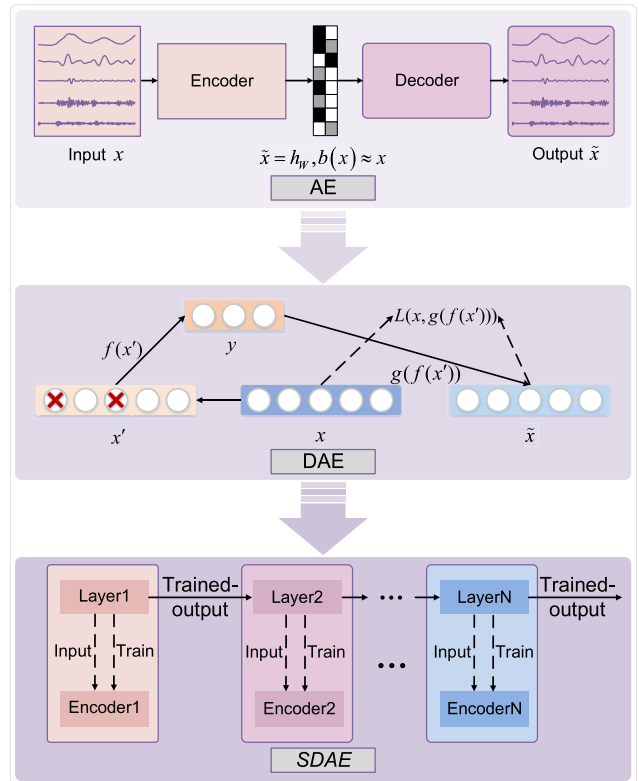


FIGURE 2. Principle of the AE/DAE/SDAE model.

C. INTRODUCTION OF BA APPLIED IN THIS PAPER

BA is a new meta-heuristic optimization algorithm proposed by Xin-She Yang [25]. The superiority of BA over other widely applied optimization methods, such as genetic algorithm and particle swarm optimization, has been proved by scholars in various research fields [26], [27].

This algorithm is based on the echolocation behaviour of microbats. When bats hunt and find prey, they change the frequency, loudness and pulse emissivity of transmitting signals to select the best solution until the target stops or the

conditions are met. In essence, tuning technology is applied to control the dynamic behaviour of a bat population and balance the parameters related to the algorithm to obtain the optimal BA [26]. The model optimization process based on BA is shown in Fig. 3.

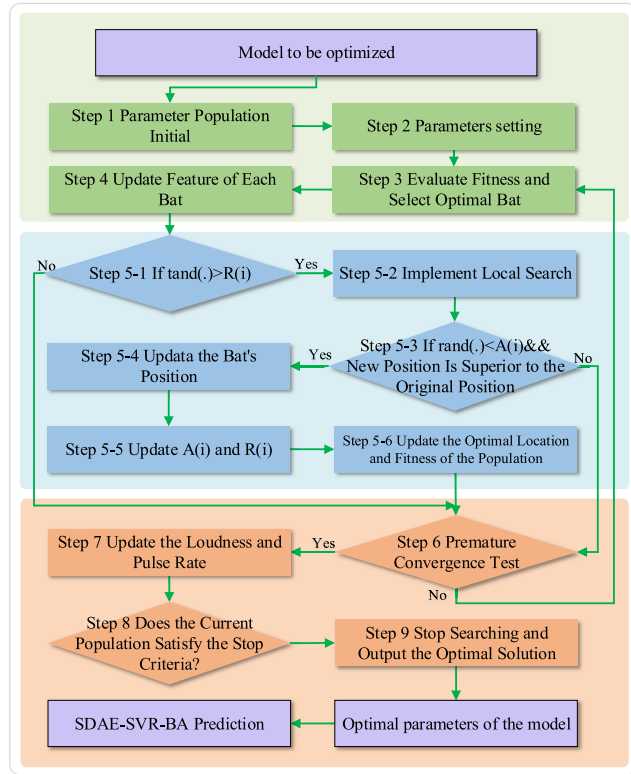


FIGURE 3. Model optimization process based on BA.

The BA optimization is applied as follows:

Step 1: Initialize the objective function and algorithm parameters of the optimization calculation. We set bat population size n_b , upper pulse frequency f_u , lower pulse frequency f_l , pulse loudness A_0 , pulse emittance R , dimension D of position vector, and maximum number of iterations N_1 .

Step 2: Randomly set the initial location and characteristic values of the micro bats. For cell bat i , a pulse frequency f_i and a D position vector X_i^0 should be randomly generated, and a D zero vector should be initialized to represent the initial velocity v_i^0 and the subsequent velocity after the update.

Step 3: Calculate the fitness of each bat in the initial population, and retrieve the bat X^* with the optimal fitness value of the initial generation.

Step 4: Update the eigenvalues of each bat in the population.

Step 5: In each iteration, if $rand1 > R(i)$ ($R(i)$ is the impulse emissivity of the i_{th} bat), the current optimal solution is selected to locally perturb the random number $rand1$ generated by the unit bat, and the new solution is judged to accept the disturbance. The judgment basis is to calculate the new fitness of bats after disturbance. If the new fitness is better than its own optimal fitness or random number $rand2 > R(i)$

($R(i)$ is the impulse loudness of the i_{th} bat), the new position after disturbance will be applied to replace the old position for storage.

Step 6: Determine whether there is a cell bat with better fitness than the global optimal fitness during this iteration. If there is, update the location and fitness value of the global optimal solution.

Step 7: Update loudness and pulse rate.

Step 8: Judge whether the end condition is met. If not, skip step 4; if it is met, skip step 9.

Step 9: The search stops and outputs the location and fitness of the bat corresponding to the global optimal solution.

BA is a new meta-heuristic optimization algorithm. Its superiority over other widely applied optimization methods, such as genetic algorithm and particle swarm optimization, has been proved by scholars in various research fields. The parameter settings of BA in this paper are shown in Table 1. The population size is twenty, the pulse transmission rate is 0.5 times/s, the maximum frequency is 2Hz, the minimum frequency is zero, the initial loudness is 0.5dB, and the number of iteration rounds is one hundred.

TABLE 1. Parameter setting of bat optimization algorithm.

Parameter	Description	Value
S	cluster size	20
r	pulse firing rate	0.5/s
f_{max}	maximum frequency	2Hz
f_{min}	minimum frequency	0
A	initial response	0.5dB
N	maximum number of iterations	100
d	parameter dimension of SDAE-SVR	2

D. INTRODUCTION OF BA APPLIED IN THIS PAPER

Combining the advantages of SVR, SDAE and BA optimization algorithms, we proposed an SDAE-SVR-BA short-term wind-power forecasting model. The overall flowchart of SDAE-SVR-BA is shown in Fig. 4. The process mainly consists of three parts: SDAE based on BA optimization, SVR based on BA optimization and prediction process based on SDAE-SVR-BA.

The BA optimization algorithm simply consists of four steps. The first is the initial population parameter, then evaluate fitness and select optimal bat, then update feature of each bat, and finally searching and output the optimal solution. The second step and the third step are in loop iteration. In part 1, as for SDAE, the number of hidden layers and the number of hidden layer nodes are both optimized by BA. In part 2, as for SVR, the penalty factor parameter C and kernel function radius g are both optimized by BA.

In part 3, the specific steps of short-term WPP based on SDAE-SVR-BA are as follows. First, the multidimensional NWP data and wind-farm historical power data in the original feature database are preprocessed. Then the BA optimized

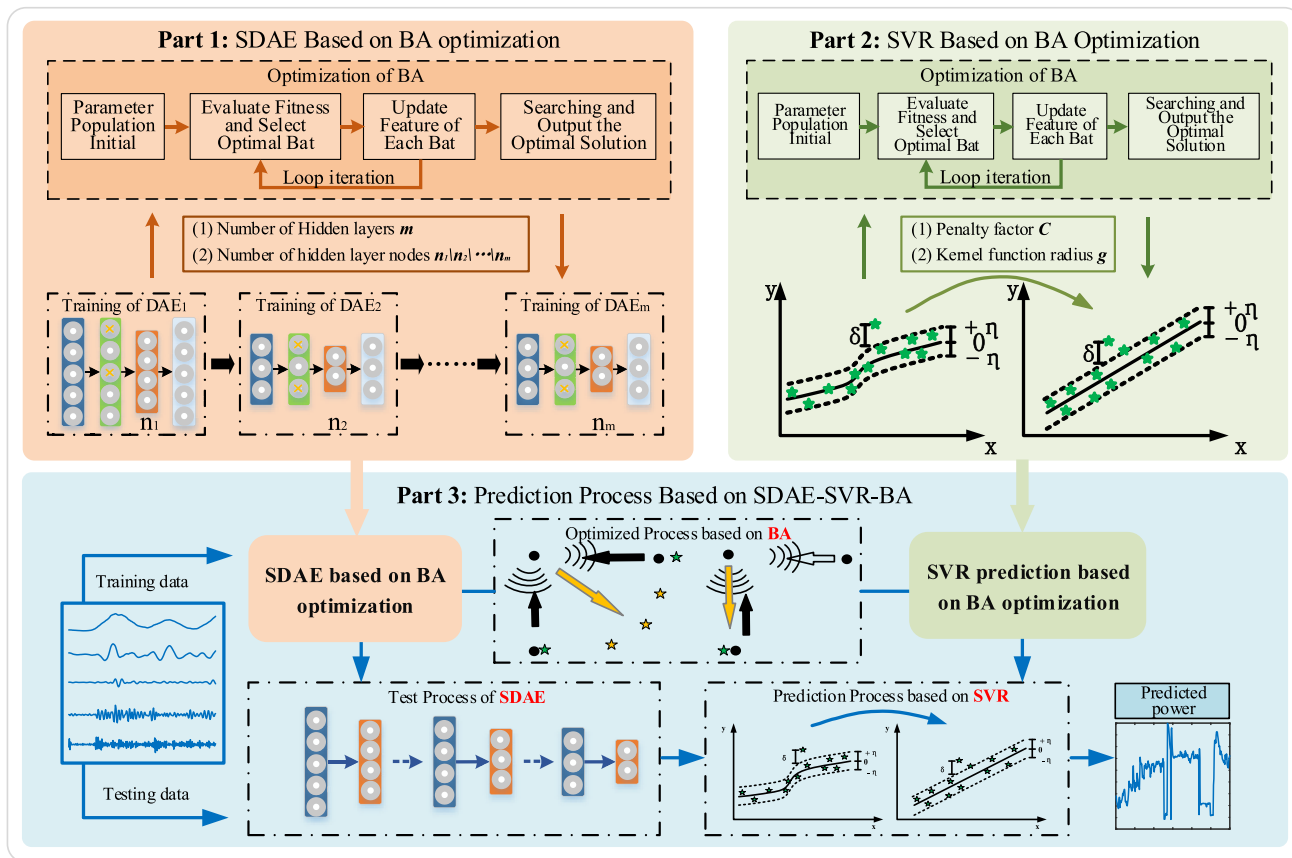


FIGURE 4. Flowchart of SDAE-SVR-BA.

stack denoising self-encoder is trained with the preprocessed data, and the low-dimensional feature data are abstracted from the high-dimensional feature data. Finally, the low-dimensional feature data are input into SVR optimized by BA for prediction.

III. CASE STUDY

In order to evaluate the performance of the proposed prediction model, we derived the data for the calculation example in this paper from a wind farm in a province, with a time resolution of one hour. We used the data from July 1, 2009, to January 1, 2011, for training, and the data from January 1, 2011, to February 1, 2011, for testing. The features used in this paper are wind speed and direction at four different heights. And We used the historical power and NWP data of the twelve points before the predicted point as the model input and the wind power for the next twelve hours as the model output.

To evaluate the prediction accuracy of the proposed prediction framework, we applied two error standards: normalized root mean square error (NRMSE) and normalized mean absolute error (NMAE). We applied E_{NRMSE} to evaluate the dispersion degree of power prediction error, and its expression is shown in (1), where f_i represents the predicted power of the first point, y_i represents the actual power of the first point,

y_{max} represents the maximum value of the actual power, and represents the number of samples. The smaller the E_{NRMSE} , the better the prediction effect.

$$E_{NRMSE} = \frac{1}{y_{max}} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2} \quad (1)$$

We applied NMAE to evaluate the average level of power prediction error; its expression is shown in (2), where f_i represents the predicted power of the first point, y_i represents the actual power of the first point, y_{max} represents the maximum value of the actual power, and represents the number of samples. The smaller the E_{NMAE} , the better the prediction effect.

$$E_{NMAE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - f_i}{y_{max}} \right| \quad (2)$$

A. COMPARISON OF IMPROVED SDAE-SVR DEEP LEARNING MODEL AND TRADITIONAL SINGLE MODEL

We compared SDAE-SVR with the deep learning model SDAE-NN and the shallow machine learning model SVR and BPNN. The results are shown in Fig.5 and Fig.6.

In Fig. 5 and Fig. 6, the standard root mean square error and standard mean absolute error of the four models are shown. The results of the prediction were as follows.

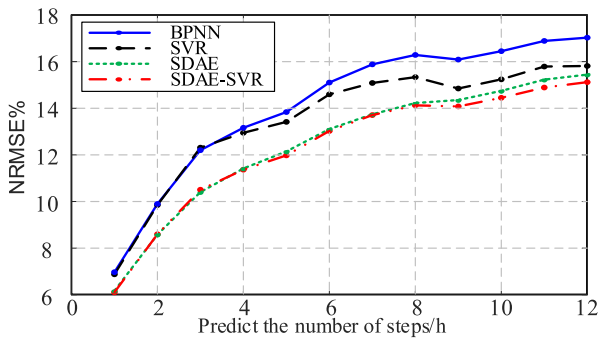


FIGURE 5. NRMSE of the four prediction models.

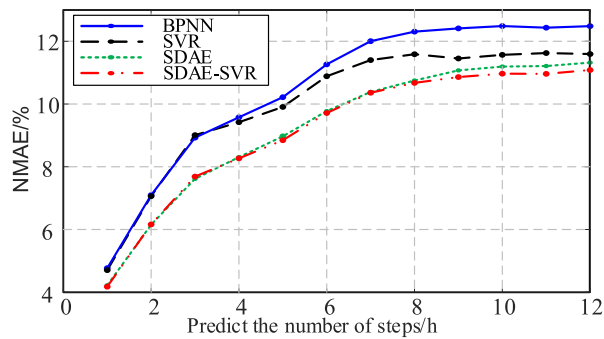


FIGURE 6. NMAE of the four prediction models.

The deep learning models SDAE-SVR and SDAE had excellent feature extraction and abstraction capabilities, and their prediction performance at all steps was significantly better than that of the other two shallow machine learning models, BPNN and SVR. Moreover, compared with the shallow model, the larger the number of prediction steps, the more obvious the prediction error of the deep learning model was reduced. This indicates that the deep learning model has more advantages in multistep prediction and confirms the robustness of the deep learning model for multistep prediction. For example, compared to BPNN and SVR, the NRMSE of the SDAE prediction model at steps 1, 4 and 12 was reduced by 0.80% and 0.71%, 1.75% and 1.53%, 1.59% and 0.37%, respectively; the SDAE-SVR prediction model was in one step, four steps, and twelve steps, of which NRMSE decreased by 0.86% and 0.77%, 1.80% and 1.58%, 1.91% and 0.69%, respectively.

The NRMSE of SDAE-SVR in the 12-step prediction was lower than that of SDAE and SVR. Compared with SDAE and SVR, the NRMSE of SDAE-SVR prediction model at 1, 4 and 12 steps decreased by 0.07% and 0.77%, 0.05% and 1.58%, 0.32% and 0.69%, respectively, which shows that, compared to the SDAE-NN and SVR when applied alone, the combined method we propose can better utilize the high-level abstract features extracted by SDAE.

B. VALIDATION OF BA OPTIMIZATION ALGORITHM

1) BA OPTIMIZED SDAE

The number of hidden layers of SDAE was set to 1–5, the BA was applied to optimize the number of hidden layer nodes of

each SDAE model, and the optimal structure of the SDAE model with different hidden layers and the corresponding of the average RMSE was predicted at 12h. The results are shown in Table 2.

TABLE 2. Optimal structure and prediction error of the SDAE model with different hidden layers optimized by BA.

SDAE layers		1	2	3	4	5
Optimal number of nodes	Hidden layer 1	20	32	25	29	10
	Hidden layer 2	/	19	31	45	45
	Hidden layer 3	/	/	16	35	49
	Hidden layer 4	/	/	/	5	46
	Hidden layer 5	/	/	/	/	8
NRMSE		0.124	0.118	0.116	0.120	0.125

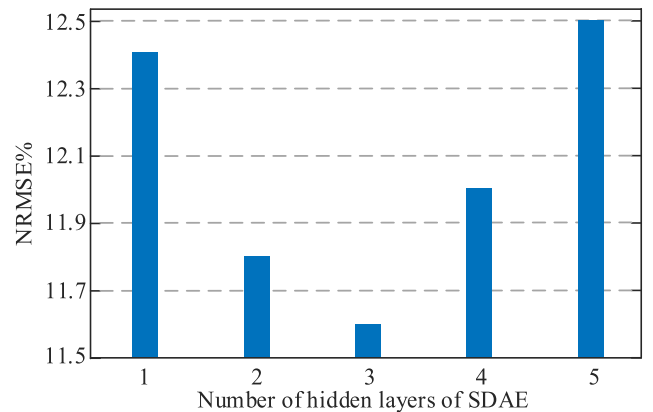


FIGURE 7. Curve of the average NRMSE predicted at 12h with the number of SDAE layers.

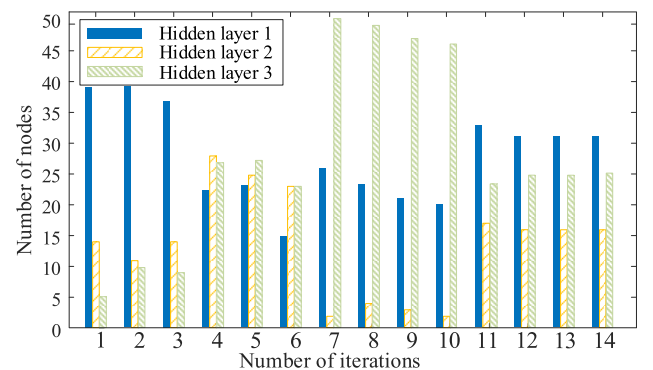


FIGURE 8. Number of nodes in each hidden layer of SDAE varies with the number of iterations.

The curve of the 12h predicted average NRMSE with the number of SDAE layers is shown in Fig. 7. It can be seen that when the hidden layer number of SDAE was three, the

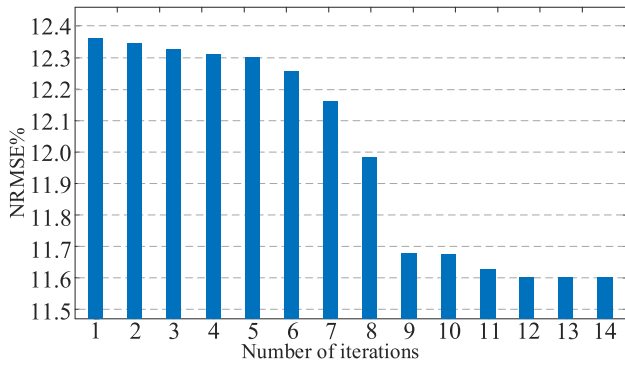


FIGURE 9. Average NRMSE predicted by 12h varies with the number of iterations.

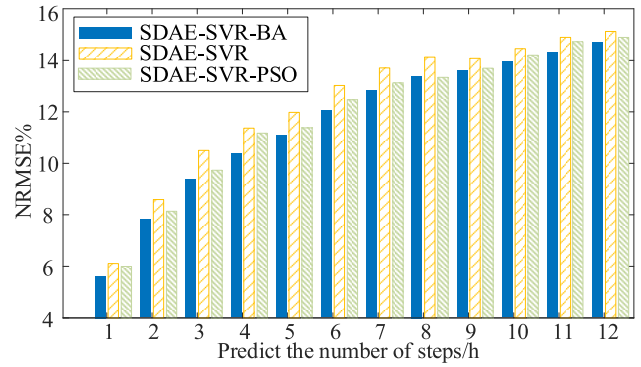


FIGURE 12. Comparison of NRMSE of the three prediction models at twelve prediction time steps.

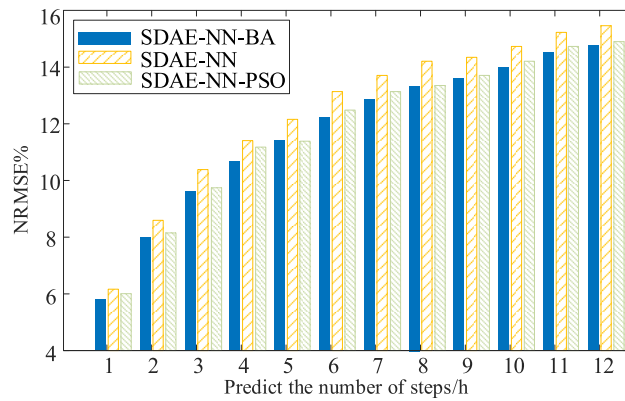


FIGURE 10. Comparison of NRMSE of SDAE model with and without BA optimization.

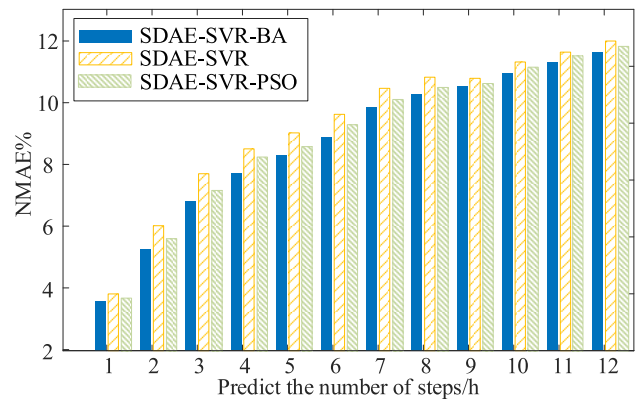


FIGURE 13. Comparison of NMAE of the three prediction models at twelve prediction time steps.

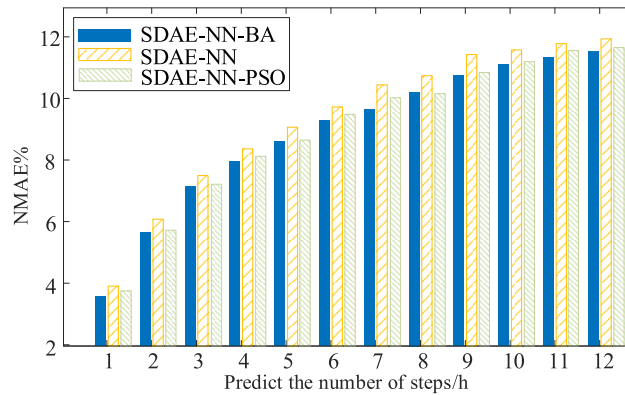


FIGURE 11. Comparison of NMAE of SDAE model with and without BA optimization.

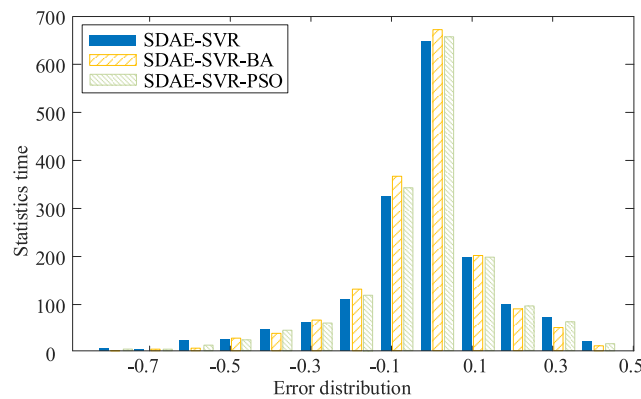


FIGURE 14. Error distribution statistics of the three prediction models.

average NRMSE of the corresponding 12h prediction was the smallest.

For the BA optimized SDAE model, when the number of hidden layers of SDAE is three, it has the highest prediction accuracy. Therefore, further study on the BA optimization process of SDAE with three hidden layers is needed. Fig. 8 shows the changes in the number of nodes in each hidden layer of SDAE as the number of iterations increases;

Fig. 9 shows the change in the average NRMSE predicted for 12h with the increase in the number of iterations.

It can be seen from the figure that as the number of iterations of the BA optimization algorithm increases, the prediction accuracy of the SDAE prediction model containing three layers gradually increases. When the number of iterations reaches twelve, the number of hidden layer nodes in SDAE 1–3 is twenty-five, thirty-one and sixteen, respectively, and

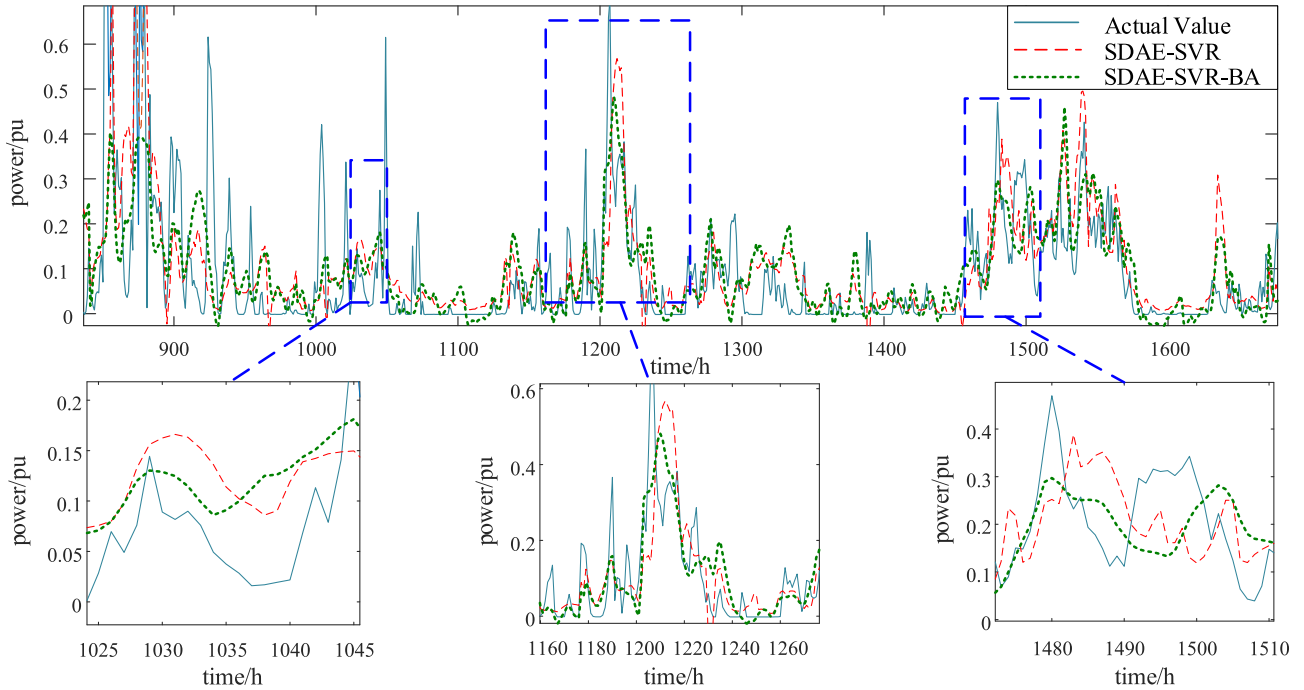


FIGURE 15. Comparison of the output power curve of the two prediction models and the actual power curve.

the prediction accuracy is optimal and no longer increases. The corresponding 12h prediction NRMSE is 11.6%.

The comparison of prediction accuracy among the three models which are BA-optimized, PSO-optimized and unoptimized SDAE (set as $n_1 = 9$, $n_2 = 9$, $n_3 = 8$ according to the empirical formula) is shown in Fig. 10 and Fig. 11. Compared to SDAE-NN without optimization, the NRMSE of the one-step, four-step and twelve-step predictions of the SDAE-NN prediction model optimized by BA was reduced by 0.35%, 0.76%, and 0.69%, respectively. Compared with SDAE-NN prediction model optimized by PSO, the NRMSE of the one-step, four-step and twelve-step predictions of the SDAE-NN prediction model optimized by BA declined 0.17%, 0.45%, and 0.21%, respectively.

2) BA OPTIMIZED AND IMPROVED SDAE-SVR DEEP LEARNING MODEL

In order to further illustrate the effectiveness of the bat optimization algorithm, we also compare SDAE-SVR-BA, SDAE-SVR-PSO and SDAE-SVR without optimization (two parameter penalty factor C and kernel function radius g are 7.4102 and 0.0016). The accuracy results of the WPP models are shown in Fig. 12 and Fig. 13. Fig. 14 shows the error distribution statistics of the three prediction models.

It can be seen from Fig. 14: Firstly, the RMSE of SDAE-SVR-BA at twelve-time steps was smaller than that of SDAE-SVR without optimization. Compared with SDAE-SVR, the NRMSE of SDAE-SVR-BA prediction model at three steps, 1, 4 and 12, was reduced by 0.48%, 0.98%, 0.43%, respectively. Compared with SDAE-SVR-PSO, the NRMSE

of SDAE-SVR-BA prediction model at three steps, 1, 4 and 12, was reduced by 0.41%, 0.77%, 0.23%, respectively. The results prove the effectiveness of BA optimization. Secondly, When the prediction time step was greater than 6, the effect of BA algorithm optimization became more and more obvious, which shows the advantages of the optimized model in multistep prediction.

Fig. 15 shows the statistics of output power error predicted by SDAE-SVR-BA with BA and SDAE-SVR without BA. The errors of the two methods roughly obeyed the normal distribution with a mean value of zero, but the variance of the model using BA was smaller, more at the time between -0.3 and 0.2 , while the variance of the model without BA was greater; there were more moments less than -0.3 and greater than 0.2 . That proves the advantages of the optimized model in prediction.

IV. CONCLUSION

In this paper we proposed a short-term WPP method based on SDAE-SVR-BA and validated it with data from a wind farm in a province of China. The conclusions are summarized as follows.

The model SDAE-SVR-BA proposed in this paper has greater accuracy than the deep learning model and traditional model. Based on the data in this paper, compared with SDAE and SVR, the NRMSE of the SDAE-SVR-BA prediction model at the first and twelfth step was reduced in accuracy by 1.05% and 1.75%, 1.72% and 1.99%, respectively, which shows that the deep learning model SDAE-SVR-BA has excellent feature extraction and abstraction capabilities.

Compared with shallow models, the larger the number of prediction steps, the more obvious the prediction error of the deep learning model is reduced, which demonstrates the robustness and stability of the deep learning model for multistep prediction.

The bat optimized model has greater accuracy than the PSO optimized model and the prediction model without optimization algorithm. The NRMSE of SDAE-SVR-BA in the twelve-step prediction is lower than that of SDAE-SVR-PSO and SDAE-SVR. Compared with SDAE-SVR-PSO and SDAE-SVR, the NRMSE of SDAE-SVR-BA prediction model at the first and twelfth step reduced by 0.15% and 0.30%, 0.97% and 1.94% in accuracy, respectively, which demonstrate that the accuracy of the optimization algorithm is greater, and the model optimized by BA is more accurate than the model optimized by PSO.

The hidden layers of SDAE have an impact on the prediction accuracy of the model. The SDAE method with three hidden layers is more accurate than that with one, two, four or five hidden layers. The NRMSE of the SDAE method containing three layers is 0.116. Compared to the SDAE method with one, two, four and five hidden layers, the NRMSE of SDAE method with three layers reduced accuracy by 6.90%, 1.72%, 3.45%, 7.76%, respectively.

In this paper, there are some questions that are not referred. Other deep learning methods, such as LSTM and BLSTM, are not discussed in this paper, because the model proposed in this paper is relatively complex and the sequence processing time is long. The relationship between each stage of the model proposed in this paper and error transmission is the focus of subsequent research in this paper.

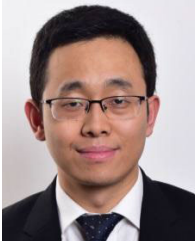
REFERENCES

- [1] X. Zhang, "Reactive voltage partitioning method for the power grid with comprehensive consideration of wind power fluctuation and uncertainty," *IEEE Access*, vol. 8, pp. 124514–124525, 2020.
- [2] X. Peng, Y. Chen, K. Cheng, H. Wang, Y. Zhao, B. Wang, J. Che, C. Liu, J. Wen, C. Lu, and W.-J. Lee, "Wind power prediction for wind farm clusters based on the multifeature similarity matching method," *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 4679–4688, Sep. 2020.
- [3] Y. Sun, Z. Li, X. Yu, B. Li, and M. Yang, "Research on ultra-short-term wind power prediction considering source relevance," *IEEE Access*, vol. 8, pp. 147703–147710, 2020.
- [4] Y. Zhang, H. Sun, and Y. Guo, "Wind power prediction based on PSO-SVR and grey combination model," *IEEE Access*, vol. 7, pp. 136254–136267, 2019.
- [5] I. Shah, S. Akbar, T. Saba, S. Ali, and A. Rehman, "Short-term forecasting for the electricity spot prices with extreme values treatment," *IEEE Access*, vol. 9, pp. 105451–105462, 2021.
- [6] N. Bibi, I. Shah, A. Alsubie, S. Ali, and S. A. Lone, "Electricity spot prices forecasting based on ensemble learning," *IEEE Access*, vol. 9, pp. 150984–150992, 2021.
- [7] H. Huang, M. Zhou, S. Zhang, L. Zhang, G. Li, and Y. Sun, "Exploiting the operational flexibility of wind integrated hybrid AC/DC power systems," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 818–826, Jan. 2021.
- [8] C. P. Mbo'o and K. Hameyer, "Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection," *IEEE Trans. Ind. Appl.*, vol. 52, no. 5, pp. 3861–3868, Sep./Oct. 2016.
- [9] B. Zhou, X. Ma, Y. Luo, and D. Yang, "Wind power prediction based on LSTM networks and nonparametric kernel density estimation," *IEEE Access*, vol. 7, pp. 165279–165292, 2019.

- [10] C. Miao, H. Li, X. Wang, and H. Li, "Ultra-short-term prediction of wind power based on sample similarity analysis," *IEEE Access*, vol. 9, pp. 72730–72742, 2021.
- [11] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
- [12] J. Zhang, H. Sun, Z. Sun, W. Dong, Y. Dong, and S. Gong, "Reliability assessment of wind power converter considering SCADA multistate parameters prediction using FP-growth, WPT, K-means and LSTM network," *IEEE Access*, vol. 8, pp. 84455–84466, 2020.
- [13] X. Peng, Q. Xu, H. Wang, J. Lang, W. Li, T. Cai, S. Duan, Y. Xie, and C. Li, "A novel efficient DLUBE model constructed by error interval coefficients for clustered wind power prediction," *IEEE Access*, vol. 9, pp. 61739–61751, 2021.
- [14] W. Huang and W. Zhang, "Research on distributed wind power reactive voltage coordinated control strategy connected to distribution network," in *Proc. 4th Int. Conf. Energy, Electr. Power Eng. (CEEPE)*, Apr. 2021, pp. 529–534.
- [15] X. Shi, F. Zhou, M. Tao, and Z. Zhang, "Human movements separation based on principle component analysis," *IEEE Sensors J.*, vol. 16, no. 7, pp. 2017–2027, Apr. 2016.
- [16] C. Jia, M. Shao, S. Li, H. Zhao, and Y. Fu, "Stacked denoising tensor auto-encoder for action recognition with spatiotemporal corruptions," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1878–1887, Apr. 2018.
- [17] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2017.
- [18] H. Zhang, Y. Liu, J. Yan, S. Han, L. Li, and Q. Long, "Improved deep mixture density network for regional wind power probabilistic forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 2549–2560, Jul. 2020.
- [19] H. He, W. Lee, D. Luo, Y. Cao, Z. Zhan, and T. Lu, "A contactless insulator contamination levels detecting method based on infrared images features and RBFNN," *IEEE Trans. Ind. Appl.*, vol. 55, no. 3, pp. 2455–2463, May/Jun. 2019.
- [20] S.-M. Baek and J.-W. Park, "Hessian matrix estimation in hybrid systems based on an embedded FFNN," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1533–1542, Oct. 2010.
- [21] E. L. Miotto, P. B. de Araujo, E. de Vargas Fortes, B. R. Gamino, and L. F. B. Martins, "Coordinated tuning of the parameters of PSS and POD controllers using bioinspired algorithms," *IEEE Trans. Ind. Appl.*, vol. 54, no. 4, pp. 3845–3857, Jul. 2018.
- [22] Y. Wang, D. Wang, and Y. Tang, "Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods," *IEEE Access*, vol. 8, pp. 17071–17079, 2020.
- [23] Y. Wei and X. Gao, "Transmission line galloping prediction based on GA-BP-SVM combined method," *IEEE Access*, vol. 9, pp. 107680–107687, 2021.
- [24] D. Y. Yang, L. Wang, G. Cai, Z. Chen, J. Ma, B. Wang, and Z. Sun, "Synchronized ambient output-only based online inter-area transfer capability assessment considering small signal stability," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 261–270, Jan. 2021.
- [25] H. Liang, Y. Liu, Y. Shen, F. Li, and Y. Man, "A hybrid bat algorithm for economic dispatch with random wind power," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5052–5061, Sep. 2018.
- [26] Y. Han and X. Tong, "Multi-step short-term wind power prediction based on three-level decomposition and improved grey wolf optimization," *IEEE Access*, vol. 8, pp. 67124–67136, 2020.
- [27] R. Eskandarpour and A. Khodaei, "Leveraging accuracy-uncertainty tradeoff in SVM to achieve highly accurate outage predictions," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1139–1141, Jan. 2018.



RUIQIN DUAN received the M.Sc. degree from the Beijing Institute of Technology, in 2016. He is currently working with Yunnan Power Grid Corporation Ltd., Kunming, Yunnan, China. His research interests include power generation planning, power generation reliability, power markets, pricing, supply, and demand.



XIAOSHENG PENG (Member, IEEE) received the B.Sc. and M.Sc. degrees from the Huazhong University of Science and Technology, China, in 2006 and 2009, respectively, and the Ph.D. degree in electrical engineering from Glasgow Caledonian University, in 2012, funded by EPSRC. He has worked as a Postdoctoral Research Fellow at Glasgow Caledonian University funded by EDF Energy. He is currently an Associate Professor with the School of Electrical and Electronic

Engineering, Huazhong University of Science and Technology. His research interests include big data, artificial intelligence and its application in power systems, new and renewable energy forecasting, and condition monitoring of power plant. He is a member of IET and IEC SC8A WG2.



CONG LI received the B.S. degree from the Department of Electrical Engineering, North China Electric Power University, Baoding, China, in 2020. He is currently pursuing the M.S. degree with the School of Electrical and Electronic Engineering, Huazhong University of Science and Technology. His research interests include deep learning, wind power prediction, and artificial neural networks.



ZIMIN YANG received the B.Eng. degree from the University of Liverpool and the M.Sc. degree from The University of Manchester. He is currently pursuing the Ph.D. degree with the Huazhong University of Science and Technology. His research interests include big data, artificial intelligence, and low carbon scenario.



YAN JIANG received the M.Sc. degree from Hohai University, in 2008. She is currently working with the Yunnan Power Dispatching and Control Center, Yunnan Power Grid Corporation Ltd., Kunming, Yunnan, China. Her research interests include power generation planning, power generation reliability, power markets, pricing, supply, and demand.



XIUFENG LI received the M.Sc. degree from Sichuan University, in 2008. He is currently working with the Yunnan Power Grid Corporation Ltd., Kunming, Yunnan, China. His research interests include power generation planning, power generation reliability, power markets, and pricing.



SHUANGQUAN LIU received the B.S. and M.S. degrees from the North China University of Water Resources and Electric Power, in 2002 and 2005, respectively, and the Ph.D. degree from the Huazhong University of Science and Technology, China, in 2009. He is currently a Senior Engineer with the Yunnan Power Grid Corporation Ltd., and a Former Research Fellow with the Belfer Center's Environment and Natural Resources Program, Harvard Kennedy School. His research interests include the generation scheduling, electricity market, and energy policy.

...