Ultra-Short-Term Wind Power Prediction Based on Bidirectional Gated Recurrent Unit and Transfer Learning

Wenjin Chen[1], Weiwen Qi[2], Yu Li[3]*, Jun Zhang[1], Feng Zhu[2], Dong Xie[2], Wei Ru[2], Gang Luo[2], Meiya Song[2] and Fei Tang[3]

[1]State Grid Zhejiang Electric Power Company, Ltd., Hangzhou, China, [2]State Grid Shaoxing Power Supply Company, Shaoxing, China, [3]School of Electrical and Automation, Wuhan University, Wuhan, China

Wind power forecasting (WPF) is imperative to the control and dispatch of the power grid. Firstly, an ultra-short-term prediction method based on multilayer bidirectional gated recurrent unit (Bi-GRU) and fully connected (FC) layer is proposed. The layers of Bi-GRU extract the temporal feature information of wind power and meteorological data, and the FC layer predicts wind power by changing dimensions to match the output vector. Furthermore, a transfer learning (TL) strategy is utilized to establish the prediction model of a target wind farm with fewer data and less training time based on the source wind farm. The proposed method is validated on two wind farms located in China and the results prove its superior prediction performance compared with other approaches.

Keywords: bidirectional gated recurrent unit, transfer learning, target domain, wind power, wind power forecasting

INTRODUCTION

The renewable energy problem is the focus of the 21st century (Zheng et al., 2017; Li et al., 2016). The transformation of the power grid is the key to solving this problem. The new form of the power grid with renewable energy as the main bulk is the ruling development trend of the future power grid (Li et al., 2021; Shen et al., 2021a). The Global Wind Energy Development Report 2019 shows that the newly installed capacity of global wind turbines in 2019 is 60.4 GW (Chen et al., 2021). However, the uncertainty existing in new energy, such as wind power, is not conducive to the safe and stable operation of the power grid. Therefore, accurate WPF is beneficial for enhancing system reliability (Shi et al., 2014).

There are three types of WPF methods including the physical method, statistical method, and artificial intelligence method. The first establishes a physical model that reflects the relationship between the wind power and numerical weather forecast (NWP) (Zhao et al., 2018), which is difficult to model and calculate. Yang proposes an expanded sequence-to-sequence (E-Seq2Seq) based data-driven SCUC expert system for dynamic multiple-sequence mapping samples, which is a pioneer study for SCUC problems (Yang et al., 2021a). The second (statistical method) is suitable for wind farms that have been built for a long time because it needs enough historical data. The representative algorithms of this method are Auto-Regression (AR) (Wu et al., 2014; Shen et al., 2021), Bayesian approach (Wang et al., 2019a), and Kalman filter (Yang et al., 2019). The final, AI method, such as support vector machine (Deo et al., 2016), artificial neural network (Wang et al., 2019b), extreme learning machine (Ali and Prasad, 2019), can deal with the complex nonlinear relationship between input and output and extract the deep features of input information, which has been widely used in recent years.

The ultra-short-term prediction of wind power is essentially a multi-variable time series prediction problem. In recent years, recurrent neural network (RNN) has developed rapidly. As the improved versions of RNN, long short-term memory (LSTM) network and gated recurrent unit (GRU) (Lin and Liu, 2020; Yang et al., 2021b) can efficiently extract the temporal correlation characteristics of wind power, and also mine the relationship between power and weather, which improves the performance of WPF. But there is a timing delay in actual prediction.

In addition, all deep learning approaches rely on a sufficient sample of data. However, newly built wind farms may not provide enough data, which makes WPF difficult. However, TL is a new method that breaks through traditional machine learning and is widely used in computer vision, text classification, and other fields (Wang et al., 2020; Shen and Raksincharoensak, 2021a; Yang et al., 2021; Yang et al., 2019; Shen et al., 2021b). It can finish pre-training of a model in the source domain with sufficient data and then transfer the pre-training model to the target domain after fine-tuning. On the one hand, TL can overcome the problem of few data, on the other hand, it can reduce the training time (Zhuang et al., 2020; Zhang et al., 2021). At present, there are few studies on the applications of TL in WPF.

In order to improve the prediction performance of RNN, the Bi-GRU method is proposed in this paper to enhance the and capacity and forecasting accuracy of the model by bidirectionality of the structure. The Bi-GRU enables the GRU to process the data in two directions including forward (future) and backward (past). Moreover, the TL strategy is used to forecast the wind power of newly built wind farms with few training data. The TL combined with Bi-GRU is used to ensure the power prediction accuracy and reduce the training time, which can guarantee model performance and reduce computational costs at the same time.

The rest of this paper is organized as follows. In *The Proposed BI-GRU Model and Transfer Learning Method*, the Bi-GRU model and transfer learning method are explained. Case studies and discussion are shown in *Case Studies*. *Conclusion* concludes this study by summarizing the key findings and contributions of this paper.

THE PROPOSED BI-GRU MODEL AND TRANSFER LEARNING METHOD

The Bi-GRU Prediction Model

RNN is widely used in time series prediction, but it has problems of gradients vanishing and exploding, and its memory ability for long series is limited (Liu et al., 2021). As the improved version of RNN, LSTM, and GRU effectively solve these problems and determine the sequential information to be forgotten and remembered through the gating mechanism. The gating mechanism of GRU is simpler than that of LSTM because it combines the forget gate and input gate of LSTM and reduces the computation while ensuring the prediction ability of the neural network. In addition, Bi-GRU is able to extract long-term dependencies before and after the current state, which means that Bi-GRU can extract more temporal features from sequential data, so Bi-GRU performs better than GRU. The structure diagram of Bi-GRU is shown in **Figure 1**.
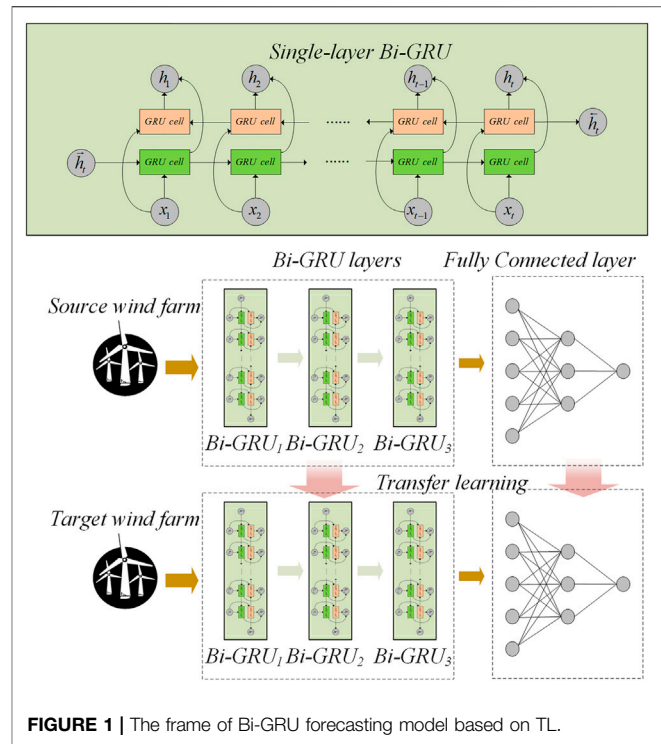

FIGURE 1 | The frame of Bi-GRU forecasting model based on TL.

The GRU cell has only two gates (an update gate z_t and a reset gate r_t). The update gate controls the extent to which the state information at the previous moment is retained into the current state, and the reset gate determines the extent to which the current state is combined with the previous information. The information flow is shown as follows in a GRU cell.

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{1}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{2}$$

$$\tilde{h}_t = tanh(W x_t + U(r_t \grave{o} h_{t-1})) \tag{3}$$

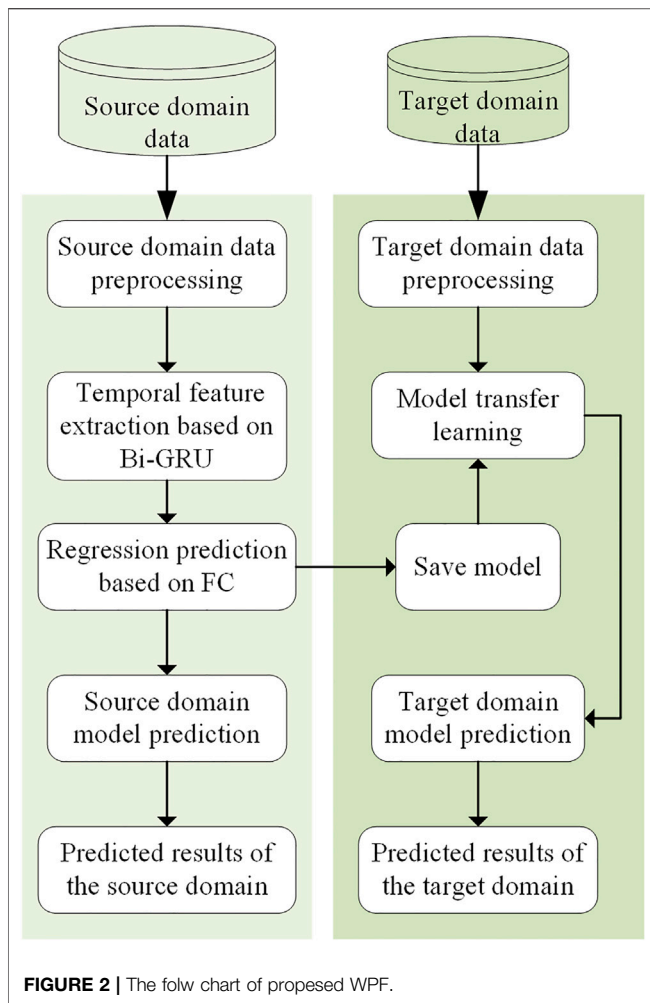$$h_t = z_t \grave{o} \tilde{h}_t + (1 - z_t) \grave{o} h_{t-1} \tag{4}$$

Where x_t, h_t are the input data and current state (also used as the output of a cell) at time t, respectively. h_{t-1} is the previous state. $\tilde{h}_t$ is the candidate state. W_r, U_r, W_z, U_z, W, U, and b_r, b_z represent weights and bias parameters, respectively. σ, tanh are activation functions and $\grave{o}$ denotes an element-wise product. But in Bi-GRU, the output h_t is concatenated by the outputs in two directions.

$$\vec{h}_t = \mathbf{GRU}\left(\vec{h}_{t-1}, x_t\right) \tag{5}$$

$$\overleftarrow{h}_t = \mathbf{GRU}\left(\overleftarrow{h}_{t-1}, x_t\right) \tag{6}$$

$$h_t = W_t \vec{h}_t + U_t \overleftarrow{h}_t + b_t \tag{7}$$

Where, $\vec{h}_t$, $\overleftarrow{h}_t$ represent the outputs in two directions, W_t, U_t, and b_t represent weights and bias parameters, respectively. In addition, FC neural network is used after Bi-GRU to fit the learned features to labels, which means achieving prediction by matching dimensions between inputs and outputs. The Rectified

FIGURE 2 | The folw chart of propesed WPF.

$$DS_s = \{F_s, L_s\} \tag{8}$$
$$DS_t = \{F_t, L_t\} \tag{9}$$

Where DS_s, DS_t represent the data space of the source domain and target domain, respectively. F_s, L_s are the features of the source domain and target domain data spaces, respectively. F_t, L_t are the labels of the source domain and target domain data spaces, respectively. The tasks of the source domain and target domain are to find the optimal parameters W_s and W_t, to make that the predicted values P_s and P_t are as close as possible to the labels L_s and L_t. TL is to fine-tune the source domain model parameter W_s to make the target domain parameter as close as possible to the optimal target domain parameter W_t.

$$P_s = f_s(F_s, W_s) \tag{10}$$
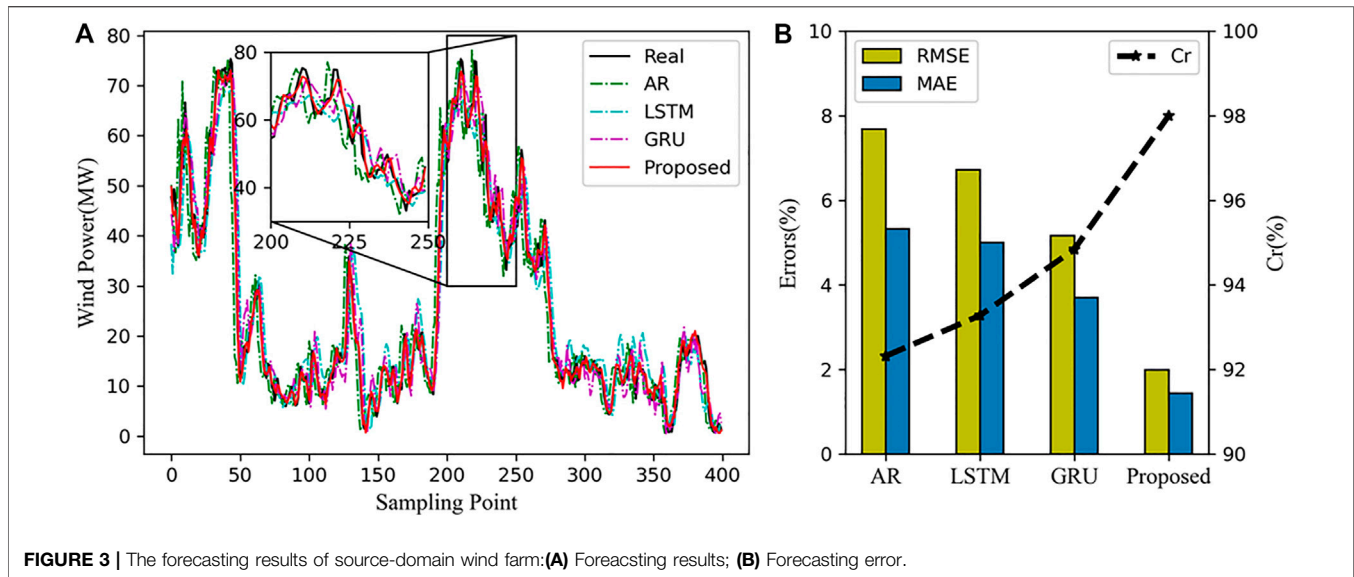$$P_t = f_t(F_t, W_t) \tag{11}$$

The prediction framework diagram of the method proposed in this paper is shown in **Figure 1**, and the processing flowchart is shown in **Figure 2**. The prediction process is mainly divided into two parts. In the first part, the wind farm power prediction model in the source domain is established. In the data pre-processing stage, the original data in the source domain are normalized to eliminate the scale difference of features and facilitate the use of gradient descent of loss function. First, the pre-processed data is fed into the three-layer Bi-GRU neural network. Then the FC layer matches the output dimension to achieve WPF to get the source-domain prediction results. The second part is to build the wind farm power prediction model in the target domain, and the data pre-processing is the same as the first part. The pre-trained source domain model is loaded and the parameters in the pre-trained model are transferred to the target domain as the initial parameters. Using a small amount of target-domain data to train the network, a fine-tuned target domain prediction model is obtained.

CASE STUDIES

In order to verify the effectiveness and superiority of the proposed prediction model and TL method, the experiment is divided into two parts. The first part compares the Bi-GRU with the AR, LSTM, and GRU. The second part uses the Bi-GRU prediction model and TL method to predict the power of wind farms in the target domain. The programming language used is Python3.8. The deep learning framework is PyTorch1.8.1.

Data Description

Two wind farms from Zhejiang Province in China are named ZJFD01 and ZJFD02 respectively. Each wind farm contains measured active power and meteorological data. The meteorological data contains wind speed, direction measured (sine and cosine of wind direction) at the hub, and air density. The time interval is 15 min. Since the running time of the two wind farms is different, the amount of historical data recorded is different. The wind farm ZJFD01 has recorded a large amount of data (including July 1, 2019–August 30, 2021) with an installed capacity of 90 MW, which is taken as the source-domain wind farm. The wind farm ZJFD02 (including January 1, 2021–August

Linear Unit (ReLU) activation function is utilized in the FC layer.

The Transfer Learning Method

The TL method is a machine learning concept that TL is used to improve the performance of target tasks on target domains by transferring the knowledge contained in different but similar source domains (Qureshi et al., 2017). Usually, the model is pre-trained in the source domain with sufficient data. Then the pre-trained model is fine-tuned in the target domain with small data, which makes full use of the source domain data to improve the performance of the model on the target domain data. TL methods can be divided into instance-based approach, feature-based approach, and parameter-based approach. The historical data of wind farms with short construction time or imperfect detection devices may not be enough to support the training of prediction models. In this paper, the parameter-based TL approach is used. The pre-training model trained by wind farms with sufficient data is fine-tuned by the target domain with insufficient data to accomplish the target tasks more efficiently. The basic idea of transfer learning can be expressed as follows.

FIGURE 3 | The forecasting results of source-domain wind farm: **(A)** Foreacsting results; **(B)** Forecasting error.

25, 2021) has recorded a small amount of data with an installed capacity of 200 MW, which is taken as the target-domain wind farm. The relationship between input and output of samples in the target domain and source domain in these datasets is similar because the relationship of wind power and meteorological variables in different wind farms is semblable. Therefore, data domains can be positively transferred.

Evaluation Metrics

In order to evaluate the prediction performance of the prediction model, the root mean square error (RMSE), mean absolute error (MAE), and accuracy (Cr) are taken as evaluation metrics according to international standards. They are defined as follows. In addition, training time is introduced as a new evaluation index in the experiment of the target-domain wind farm.

$$\mathbf{RMSE} = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^{n} \left(\frac{P_{real,i} - P_{pred,i}}{C_i} \right)^2} \times 100\% \qquad (12)$$

$$\mathbf{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{P_{real,i} - P_{pred,i}}{C_i} \right| \times 100\% \qquad (13)$$

$$C_r = 1 - \mathbf{RMSE} \qquad (14)$$

Where $P_{real,i}$, $P_{pred,i}$, and C_i are real output power, predicted output power, and capability of wind farm respectively. n is the total number of predicted samples.
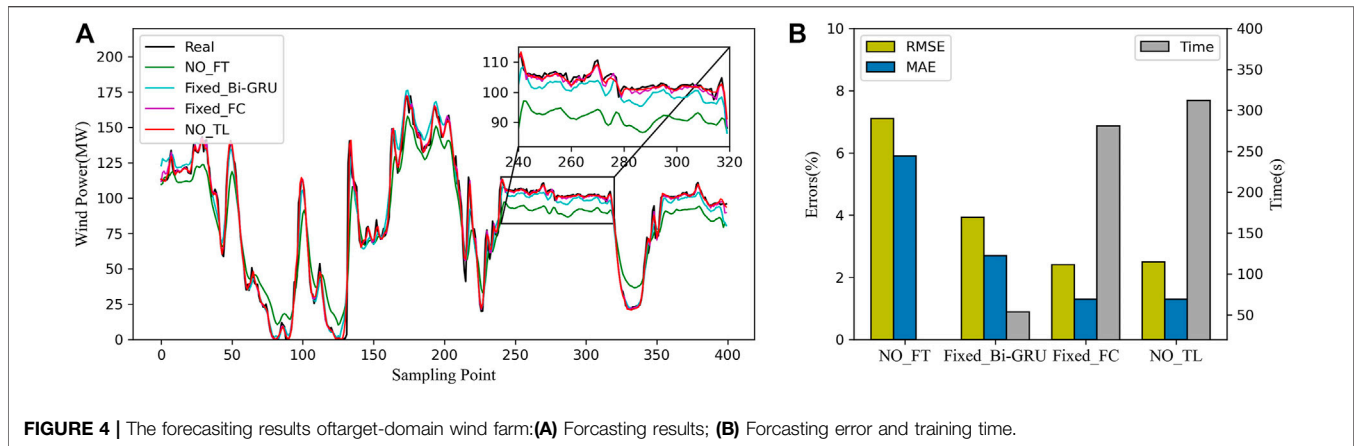
The Experiment of Source-Domain Wind Farm

The source-domain prediction model, the Bi-GRU method, is established for the ZJFD01 wind farm. In the data pre-processing stage, the supervised learning dataset is constructed. The output power of the current time step y_t is selected as the label. The previous four timesteps ($x_{t-1}, x_{t-2}, \ldots, x_{t-4}$) are selected as features.

A total of 70% of the dataset was used as the training set and the last 30% as the verification set.

For hyperparameters, set input size to 4, hidden size to 8, and the number of layers to 3. Then, the FC layer is connected to tag dimension matching, and the number of neuron nodes in the input layer is 64 (since the output of Bi-GRU is flattened), the hidden layer is 32, and the output layer is 1. In the model training stage, mean-squared loss is used as the loss function to measure the error between predicted power and actual output power, and the Adam optimization algorithm is used as the optimizer. In order to evaluate the superiority of the proposed method in wind farm prediction in the source domain, RMSE, MAE, and Cr are used as evaluation metrics.

The 400 sampling points of the test dataset are taken to verify the prediction effects of various methods. The power prediction results are shown in **Figure 3**. Compared with other methods, the power prediction curve of Bi-GRU is closer to the actual power output curve trend. As can be seen from **Figure 3**, the RMSE and MAE of the proposed method are significantly lower than those of other methods, and the accuracy is improved. Compared with LSTM and GRU, RMSE and MAE are reduced by 4.73 and 3.17% respectively. The prediction effect of GRU is better than that of LSTM because the same iteration times are set, but GRU has a simpler structure and fewer parameters to be optimized, so it has higher accuracy. There are two reasons why the proposed method is superior: 1) The Bi-GRU can excavate the relationship between historical meteorological data and current power data layer by layer through various gating mechanisms, and can also excavate the local and long-term correlation before and after the power data series; 2) The characteristics of both the forward and reverse time sequence of power and meteorological data are taken into account by the bidirectional mechanism, so it can effectively improve the accuracy of prediction. As seen from the local amplification figure, the forecasting curve trend of all methods is close to the actual power curve, but there are different levels of phase difference. However, the bidirectional mechanism of Bi-GRU solved this problem, making the prediction curve more closely fit the

FIGURE 4 | The forecasiting results oftarget-domain wind farm:**(A)** Forcasting results; **(B)** Forcasting error and training time.

actual power curve, which is the important reason for its better prediction performance.

The Experiment of Target-Domain Wind Farm

In order to ensure the prediction accuracy and reduce the training time, the power prediction of the ZJFD02 wind farm in the target domain is based on transfer learning. The parameters and structure of the pre-trained model from ZJFD01 are migrated to the ZJFD02. The preprocessing method of the dataset is the same as that of source wind farm. In order to explore and verify the advantages of using transfer learning to predict power, the following cases are compared:

a) The pre-trained model in the source domain is directly loaded, and denoted as NO_fine-tunning (NO_FT);

b) The pre-trained model in the source domain is loaded and the parameters in Bi-GRU layers are frozen and the parameters of the FC layer are fine-tuned with target-domain data, which is named Fixed_Bi-GRU;

c) The pre-trained model in the source domain is loaded and the parameters in the FC layer are frozen and the parameters of Bi-GRU layers are fine-tuned with target-domain data, which is named Fixed_FC;

d) Redefine a prediction model whose structure is the same as that of the source-domain model but whose parameters are not trained at all. Then train it with target-domain data, which is named NO_TL.

In addition to RMSE and MAE, training time is added to the evaluation metrics to measure the improvement of computing speed caused by TL. Except for case (a), the number of iterations in other cases is set to 200.

By taking 400 sampling points, the prediction results and performance of the above cases can be compared, as shown in **Figure 4**. From the perspective of prediction accuracy, the prediction accuracy of case (a) and case (b) is lower than that of case (c) and case (d). RMSE and MAE of case (c) are 4.705 and 4.607%, respectively, lower than that of case (a), because there are still differences in the dataset of the source domain and target domain. If there is no parameter fine-tuning, it will cause a large prediction error. Case (b) and (c) fixed different parameters of the

network layer, RMSE, and MAE were reduced by 1.533 and 1.404% respectively compared with case (b), because the number of three-layer network parameters of Bi-GRU was much more than that of FC layer. After fine-tuning in the target domain, case (b) changed the parameters of the model to a greater extent than case (b), so it is closer to the optimal target domain model; the prediction accuracy of case (c) and case (d) was similar, RMSE is 2.397 and 2.484%, MAE is 1.295%, and 1.298%, respectively. From the perspective of time-consuming, cases (b), (c), and (d) are compared. it is obvious that the training time of case (b) is less than that of (c). Most parameters of this prediction model are still Bi-GRU layers, so it saves training time to fine-tune FC layer parameters. The accuracy of (c) is similar to that of (d), but the training time of (c) is 9.9% shorter than that of (d). Therefore, using the transfer learning fine-tuning the pre-trained model can guarantee the prediction accuracy and save the training time to a certain extent compared with the training model starting from the beginning.

CONCLUSION

In this paper, a Bi-GRU prediction model based on the transfer learning method is presented for the ultra-short-term of wind power. According to the results of case studies, some conclusions are summed up as follows. On one hand, the Bi-GRU prediction model can extract the temporal features of wind power sequential data in two directions, which learns deeper historical information and realize higher accuracy of WPF than GRU and LSTM. On the other hand, the prediction model combined with the TL method saves training time and reduces the requirement for abundant data. In the future, more detailed research about how to balance training time and accuracy of prediction using TL will be completed. Moreover, more comprehensive evaluation metrics aimed at evaluating the TL method in WPF will be established (Shen and Raksincharoensak, 2021b).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

WC: Data provision, Project administration, Methodology. WQ: Data Processing, Project administration, Supervision. YL: Methodology, Data Processing, Writing- Original draft preparation, Software. JZ: Conceptualization, Writing-Reviewing and Editing, Supervision. FZ: Validation, Investigation, Supervision. DX: Visualization, Supervision. WR: Investigation, Supervision. GL: Data Processing, Supervision. MS: Data Processing, Supervision. FT: Supervision, Writing-Reviewing and revising grammar and correct expression.

FUNDING

REFERENCES

Ali, M., and Prasad, R. (2019). Significant Wave Height Forecasting via an Extreme Learning Machine Model Integrated with Improved Complete Ensemble Empirical Mode Decomposition. *Renew. Sustain. Energ. Rev.* 104, 281–295. doi:10.1016/j.rser.2019.01.014

Chen, X., Zhang, X., Dong, M., Huang, L., Guo, Y., and He, S. (2021). Deep Learning-Based Prediction of Wind Power for Multi-Turbines in a Wind Farm. *Front. Energ. Res.* 9, 723775. doi:10.3389/fenrg.2021.723775

Deo, R. C., Wen, X., and Qi, F. (2016). A Wavelet-Coupled Support Vector Machine Model for Forecasting Global Incident Solar Radiation Using Limited Meteorological Dataset. *Appl. Energ.* 168, 568–593. doi:10.1016/j.apenergy.2016.01.130

Li, Z., Jiang, W., Abu-Siada, A., Li, Z., Xu, Y., and Liu, S. (2020). Research on a Composite Voltage and Current Measurement Device for HVDC Networks. *IEEE Trans. Ind. Electron.* 68 (9), 8930–8941. doi:10.1109/TIE.2020.3013772

Li, Z., Ye, L., Zhao, Y., Song, X., Teng, J., and Jin, J. (2016). Short-term Wind Power Prediction Based on Extreme Learning Machine with Error Correction. *Prot. Control. Mod. Power Syst.* 1 (1), 1. doi:10.1186/s41601-016-0016-y

Lin, Z., and Liu, X. (2020). Wind Power Forecasting of an Offshore Wind Turbine Based on High-Frequency SCADA Data and Deep Learning Neural Network. *Energy* 201, 117693. doi:10.1016/j.energy.2020.117693

Liu, X., Lin, Z., and Feng, Z. (2021). Short-term Offshore Wind Speed Forecast by Seasonal ARIMA - A Comparison against GRU and LSTM. *Energy* 227, 120492. doi:10.1016/J.ENERGY.2021.120492

Qureshi, A. S., Khan, A., Zameer, A., and Usman, A. (2017). Wind Power Prediction Using Deep Neural Network Based Meta Regression and Transfer Learning. *Appl. Soft Comput.* 58, 742–755. doi:10.1016/j.asoc.2017.05.031

Shen, X., Ouyang, T., Khajorntraidet, C., Li, Y., Li, S., and Zhuang, J. (2021a). Mixture Density Networks-Based Knock Simulator. *Ieee/asme Trans. Mechatron.*, 1. doi:10.1109/TMECH.2021.3059775

Shen, X., Ouyang, T., Yang, N., and Zhuang, J. (2021b). Sample-based Neural Approximation Approach for Probabilistic Constrained Programs. *IEEE Trans. Neural Netw. Learn. Syst.* 1, 8. doi:10.1109/TNNLS.2021.3102323

Shen, X., and Raksincharoensak, P. (2021a). Pedestrian-aware Statistical Risk Assessment. *IEEE Trans. Intell. Transport. Syst.*, 1–9. doi:10.1109/TITS.2021.3074522

Shen, X., and Raksincharoensak, P. (2021b). Statistical Models of Near-Accident Event and Pedestrian Behavior at Non-signalized Intersections. *J. Appl. Stat.* 1, 21. doi:10.1080/02664763.2021.1962263

Shi, J., Ding, Z., Lee, W.-J., Yang, Y., Liu, Y., and Zhang, M. (2014). Hybrid Forecasting Model for Very-Short Term Wind Power Forecasting Based on Grey Relational Analysis and Wind Speed Distribution Features. *IEEE Trans. Smart Grid.* 5 (1), 521–526. doi:10.1109/tsg.2013.2283269

Wang, J., Zhang, N., and Lu, H. (2019a). A Novel System Based on Neural Networks with Linear Combination Framework for Wind Speed Forecasting. *Energ. Convers. Manage.* 181, 425–442. doi:10.1016/j.enconman.2018.12.020

Wang, Y., Wang, H., Srinivasan, D., and Hu, Q. (2019b). Robust Functional Regression for Wind Speed Forecasting Based on Sparse Bayesian Learning. *Renew. Energ.* 132, 43–60. doi:10.1016/j.renene.2018.07.083

Wang, Z., Zhang, J., Zhang, Y., Huang, C., and Wang, L. (2020). Short-term Wind Speed Forecasting Based on Information of Neighboring Wind Farms. *IEEE Access* 8, 16760–16770. doi:10.1109/access.2020.2966268

Wu, B., Song, M., Chen, K., He, Z., and Zhang, X. (2014). Wind Power Prediction System for Wind Farm Based on Auto Regressive Statistical Model and Physical Model. *J. Renew. Sustain. Energ.* 6 (1), 013101. doi:10.1063/1.4861063

Yang, D. (2019). On post-processing Day-Ahead NWP Forecasts Using Kalman Filtering. *Solar Energy* 182, 179–181. doi:10.1016/j.solener.2019.02.044

Yang, N., Huang, Y., Hou, D., Liu, S., Ye, D., Dong, B., et al. (2019). Adaptive Nonparametric Kernel Density Estimation Approach for Joint Probability Density Function Modeling of Multiple Wind Farms. *Energies* 12 (7), 1356. doi:10.3390/en12071356

Yang, N., Liu, S., Deng, Y., and Xing, C. (2021a). An Improved Robust SCUC Approach Considering Multiple Uncertainty and Correlation. *IEEJ Trans. Elec Electron. Eng.* 16 (1), 21–34. doi:10.1002/tee.23265

Yang, N., Yang, C., Wu, L., Shen, X., Jia, J., Li, Z., et al. (2021b). Intelligent Data-Driven Decision-Making Method for Dynamic Multi-Sequence: An E-Seq2Seq Based SCUC Expert System. *IEEE Trans. Ind. Inf.*, 1. doi:10.1109/TII.2021.3107406

Yang, N., Yang, C., Xing, C., Ye, D., Jia, J., Chen, D., et al. (2021c). Deep Learning-based SCUC Decision-making: An Intelligent Data-driven Approach with Self-learning Capabilities. *IET Gener. Transm. Distrib.*. doi:10.1049/gtd2.12315

Zhang, L., Xie, Y., Ye, J., Xue, T., Cheng, J., Li, Z., et al. (2021). Intelligent Frequency Control Strategy Based on Reinforcement Learning of Multi-Objective Collaborative Reward Function. *Front. Energ. Res.*, 1–12. doi:10.3389/fenrg.2021.760525

Zhao, X., Liu, J., Yu, D., and Chang, J. (2018). One-day-ahead Probabilistic Wind Speed Forecast Based on Optimized Numerical Weather Prediction Data. *Energ. Convers. Manage.* 164, 560–569. doi:10.1016/j.enconman.2018.03.030

Zheng, D., Eseye, A. T., Zhang, J., and Li, H. (2017). Short-term Wind Power Forecasting Using a Double-Stage Hierarchical ANFIS Approach for Energy Management in Microgrids. *Prot. Control. Mod. Power Syst.* 2 (1), 13. doi:10.1186/s41601-017-0041-5

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2021). A Comprehensive Survey on Transfer Learning. *Proc. IEEE* 109 (1), 43–76. doi:10.1109/JPROC.2020.3004555