2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe) | 979-8-3503-4743-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/EEEIC/ICPSEUROPE57605.2023.10194807

Learning from ChatGPT: A Transformer-Based Model for Wind Power Forecasting

1st Xiaoran Dai
School of Electrical Engineering and Automation
Wuhan University
Wuhan, China
dai_xiaoran@whu.edu.cn

2nd Guo-Ping Liu
Center for Control Science and Technology
Southern University of Science and Technology
Shenzhen, China
liugp@sustech.edu.cn

3rd Wenshan Hu
School of Electrical Engineering and Automation
Wuhan University
Wuhan, China
wenshan.hu@whu.edu.cn

4th Zhongcheng Lei
School of Electrical Engineering and Automation
Wuhan University
Wuhan, China
zhongcheng.lei@whu.edu.cn

5th Hong Zhou
School of Electrical Engineering and Automation
Wuhan University
Wuhan, China
hzhouwuhee@whu.edu.cn

***Abstract*—Wind power forecasting is a crucial aspect of renewable energy production, as it helps to optimize energy output and ensure grid stability. In recent years, Transformer-based language models such as ChatGPT have been successfully used in natural language processing tasks, but their application in wind power forecasting remains largely unexplored. In this article, we propose using a Transformer model, the core of ChatGPT, to improve the accuracy of wind power forecasting. Using the self-attention mechanism, the developed model can capture the complex temporal relationships in large-scale time series data. Furthermore, the proposed method is evaluated on a test set using various performance metrics. Results show that our model outperforms traditional forecasting models, achieving higher accuracy. Our findings suggest that Transformer-based models have significant potential for improving wind power forecasting accuracy and ultimately contributing to a more sustainable energy future.**

***Index Terms*—Wind power forecasting, ChatGPT, Transformer-based model, Time series prediction**

I. Introduction

Wind power has become an increasingly significant source of world energy in recent decades, which is green, inexhaustible, and installation-flexible [1]. According to the reports of world energy in 2022 [2], [3], the share of wind power generation and installed capacity can be depicted as Fig. 1. It can be seen that both of them maintain steady growth every year. By 2030, the wind power capacity in the net zero scenarios is expected to reach 3105.9 GW [3], accounting for 20% of global electricity generation.[1] However, the intermittent and fluctuated nature of wind is still a challenge, which may bring fluctuations or even disasters to power grids [4]. Therefore, an accurate and reliable wind power forecasting (WPF) technique is critical for the efficient operation and management of wind power systems, as it enables operators to better anticipate and respond to changes in wind conditions and optimize power output.

WPF is the task of predicting the power output of a wind farm over a given time horizon, typically ranging from a few minutes to several days. Accurate forecasting enables operators to optimize power system operations, reduce energy costs, and integrate wind energy into the grid, thus supporting the transition to a sustainable energy system [5]. Owing to the wide applications of WPF, various forecasting methods have been developed [6], [7]. In general, these forecasting models can be divided into the following three categories: physical models, statistical models, and machine learning models.

The physical forecasting model, which is developed based on numerical weather prediction (NWP) and aerodynamics, has demonstrated its effectiveness and interpretability in long-term WPF [8]. However, short-term NWPs are often less accurate and the resolution may not meet forecast needs. Statistical models mainly include the auto-regressive integrated

This work was supported in part by the special scholarship of Wuhan University Postgraduate Overseas Exchange Program.

979-8-3503-4743-2/23/$31.00 ©2023 IEEE

[1][Online]. Available: https://www.energy.gov/eere/wind/20-wind-energy-2030-increasing-wind-energys-contribution-us-electricity-supply

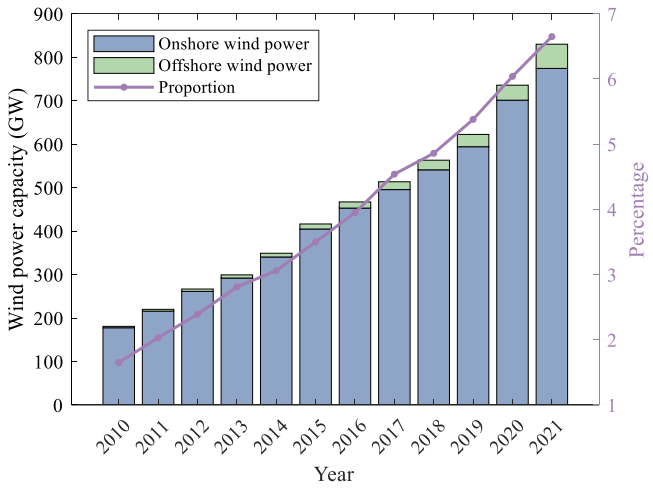

Fig. 1. Global installed wind power capacity and its share of electricity generation.

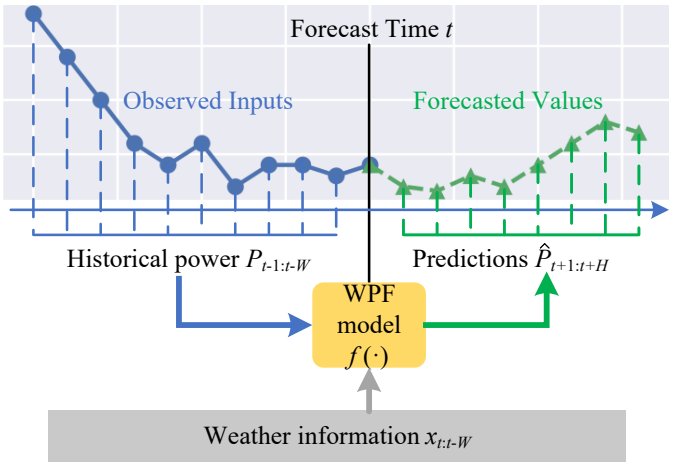

Fig. 2. Diagram of WPF process.

moving average (ARIMA) method, Markov models, and ridge regression methods. Earlier, statistical models were widely applied for time series power prediction [9]. However, with the rapid increase of data volume, statistical models gradually lose their advantages compared to machine learning-driven methods.

Machine learning models such as the conventional support vector machine (SVM), extreme learning machine (ELM), and K-nearest neighbor (KNN) have been extensively investigated for WPF and each has its own strengths and limitations [10]. In recent years, the application of deep learning techniques in time series prediction, including WPF, has gained much attention due to its impressive performance. As an effective solution for multivariate time series prediction, the long short term memory (LSTM) network is the most widely used technique in WPF, along with its various modifications. For example, the bidirectional LSTM is combined with deep concatenated residual networks for one-hour-ahead WPF [11]. In [12], the Gated Recurrent Unit (GRU) is improved using attention mechanism for WPF.

Although the aforementioned recurrent neural networks have shown promising results in processing temporal information, they have disadvantages in parallel processing and limited capabilities in handling long-term temporal dependencies. Recent developments in deep learning have produced potent language models like ChatGPT that can generate human-like responses to given prompts. While ChatGPT was originally designed for natural language processing (NLP) tasks, its underlying Transformer-based architecture has also been successfully applied to a variety of other sequence prediction tasks, such as image and speech recognition as well as weather forecasting [13].

Inspired by the above works on Transformer models for NLP and sequence prediction, in this paper, we propose a Transformer-based forecasting approach for WPF that leverages the powerful learning ability of Transformer to capture time series relations from large amounts of historical data and make accurate forecasting about future power sequences. We evaluate the performance of the proposed model on a real-world dataset and demonstrate that outperforms existing state-of-the-art methods.

The rest of this paper is structured as follows. Section II provides the problem definition of WPF and related works of Transformer-based neural networks and their applications. The developed forecasting architecture is presented in Section III. Experimental results and our conclusions are given in Sections IV and V respectively.

II. Preliminaries

A. Problem Definition

The task of WPF can be formulated as a multivariate time series prediction problem. It utilizes the temporal relationships within the power sequences and weather-related data, such as wind speed, direction, temperature, humidity, etc. from past observations to predict future power generation. In practice, the WPF is carried out as a rolling forward prediction over time, as shown in the following equation:

$$\hat{P}\left(t+a\right)=f\left(x_{t:t-b},P_{t-1:t-b}\right), \tag{1}$$

where $a = 1,2,\ldots,H$ is the predictive horizon, $b = 1,2,\ldots,W$ is the look-back window length, $\hat{P}$ and P are predictive and actual power respectively, x represents the weather information. Moreover, $f\left(\cdot\right)$ represents the nonlinear mapping relationship from historical data to predictions, which is also what we need to construct by neural networks subsequently.

To illustrate the forecasting process more graphically, the forecasting process corresponding to (1) is drawn in Fig. 2. It can be clearly seen that at time t, the model $f\left(\cdot\right)$ uses historical power and weather information to forecast the power generation at the future time.

B. Related Works

ChatGPT, a chatbot from Open AI, is recently making waves across the Internet. Its applications are not limited to NLP

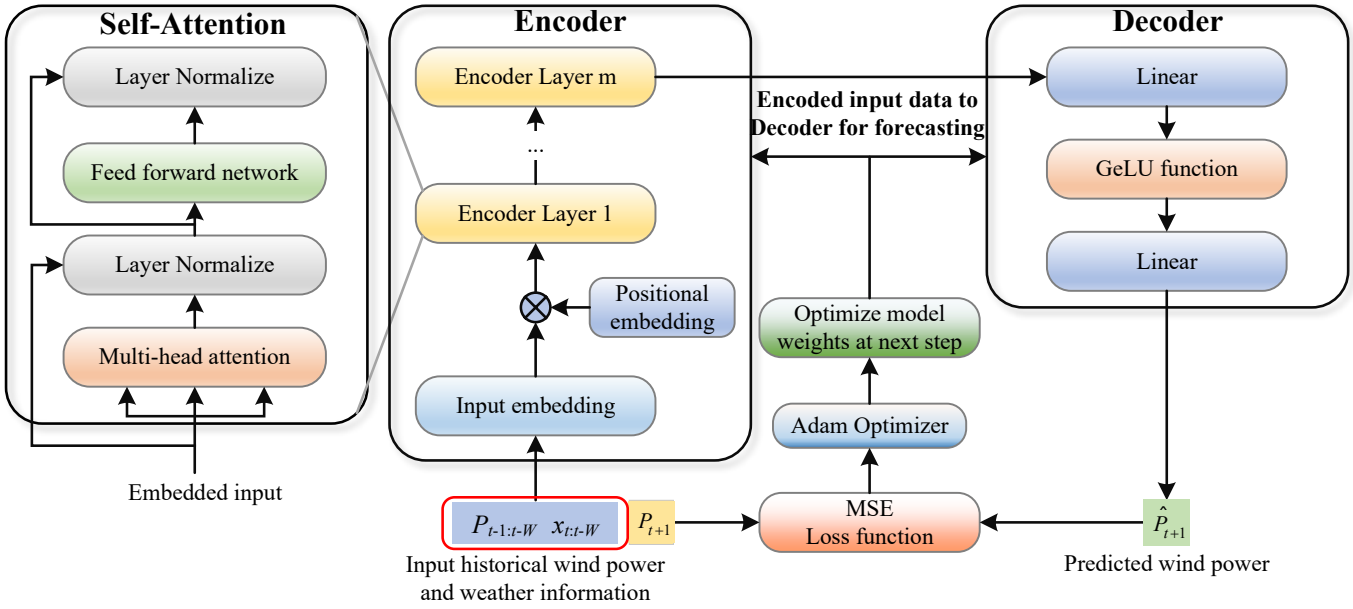

Fig. 3. Forecasting architecture of Transformer-based model

and human-like text generation, it has been used in a variety of fields such as transportation [14], healthcare [15], and algorithmic intelligence [16]. ChatGPT is a pretranining language model whose underlying technology is the Transformer architecture. For WPF tasks, although ChatGPT cannot be directly applied due to its limitation in processing time series data, its underlying Transformer architecture is a promising option for its ability to capture temporal dependencies and attention mechanisms.

The development of Transformer-based models began with the introduction of Transformer concept by *Vaswani et al.* in 2017 [17], which was initially designed for NLP tasks. The authors introduced a self-attention mechanism which enables the model to attend to all input positions when calculating the representation of each position. This architecture achieved state-of-the-art results on several NLP tasks and quickly became the preferred model for NLP applications [18].

Specifically, the Transformer model consists of an encoder and a decoder [17]. The encoder processes the input sequence and produces a hidden state representation, while the decoder generates the output sequence based on the encoded representation and the previous output. The attention mechanism allows the model to focus on different parts of the input sequence at different time steps, which is useful for modeling long-term dependencies.

As for Transformer-based models in WPF, one advantage is their ability to capture complex temporal patterns in the input sequence, such as the effects of weather conditions on wind power generation over time. Another advantage is their scalability, Transformer architectures are proven to be able to handle large amounts of data due to their parallel computation character, which makes them more efficient when dealing with large-scale data in real-world WPF tasks.

Motivated by the discussion above, Transformer-based models have shown promise in WPF due to their capability to capture complex and nonlinear relationships between input variables and output sequences. However, only a few research efforts have been devoted to Transformers for WPF. Considering the potential of the Transformer to improve the forecasting accuracy, a full procedure of a Transformer-based model regrading the designing, training, and implementation is to be explored subsequently.

III. Wind Power Forecasting Model Based on Transformer

The Transformer-based model employs the self-attention mechanism as its main model architecture, which can more effectively capture the temporal correlation in the wind power sequence and weather conditions compared to traditional recurrence structures. Based on the self-attention mechanism, the Transformer-based model is mainly composed of two parts: the Encoder module and the Decoder module. The overall structure of the tailored Transformer model is illustrated in Fig. 3 and described as follows:

Encoder : The Encoder module is a crucial component of the Transformer-based model, which comprises an input embedding layer, a positional encoding layer, and m encoding layers. The value of m is a hyperparameter that can be optimized. Corresponding the Encoder module in Fig. 3, the input historical sequences are processed as follows.

First, the input historical sequence is transformed into a higher-dimensional vector space through the input embedding layer. At time t, H input power values are converted into d-dimensional vectors through a fully connected layer. As the recurrence structure is discarded, the sequence information of the input elements is captured by the positional encoding (PE)

module. The PE module embeds order information using sine and cosine functions illustrated in (2). Here, pos represents the position of the elements in the time series sequence, i represents the dimension number of the embedded vector, and d_{model} represents the dimension of the input embedding layer. Finally, the position information and the input embedding values are added to the Embedding layer.

$$\begin{aligned} PE\left(pos, 2i\right) &= \sin\left(pos/10000^{2i/d_{model}}\right), \\ PE\left(pos, 2i+1\right) &= \cos\left(pos/10000^{2i/d_{model}}\right). \end{aligned} \tag{2}$$

The Embedding layer is composed of m stacks connected in a sequential order. Each stack consists of a multi-head attention layer and a feed forward network (FFN), where multi-head attention layer is on the basis of the self-attention module. In this module, three matrices W_Q, W_K, W_V are defined to perform a linear transformation of all input vectors as shown in (3). Hence, the query matrix Q, the key matrix K and the value matrix V can be obtained. Finally, the attention score can be calculated by Softmax function in (4) to characterize which data is more noteworthy (it has greatest impact on the predicted output), where $d_k = d_{model}$.

$$\begin{cases} Q = W_Q P_{input\_embedding}, \\ K = W_K P_{input\_embedding}, \\ V = W_V P_{input\_embedding}. \end{cases} \tag{3}$$

$$attention = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{4}$$

The multi-head attention mechanism aims to allow the model to pay attention to the characteristics of different features, where each head can learn different temporal relationships through the input data. Specifically, supposing there are n heads, the sequence is then divided into n groups and the matrices W_Q, W_K, W_V mentioned above transforms to $\left(W_1^Q, W_1^K, W_1^V\right), \left(W_2^Q, W_2^K, W_2^V\right) \ldots \left(W_n^Q, W_n^K, W_n^V\right)$. Further, Q, K, V matrices can be obtained by contacting $Q_i, K_i, V_i, i = 1, 2 \ldots n$. Thus, the attention score under the multi-head attention mechanism can be obtained by (4) as well.

Decoder : Unlike the complex structure of the Decoder in the original Transformer [17], the Decoder structure in this paper only obtains the final output through two linear layers and an activation function. Specifically, the input wind power sequence is encoded by Encoder and the encoded output enters the Decoder. First, it passes through a linear layer, and then the data is processed by the Gaussian error Linear Unit (GeLU) activation function [19]. Finally, the output is reshaped by a linear layer with a dimension of 1.

The output of the Decoder module is the predicted power value, which is used to compare with the actual value to update the parameters in the Encoder and Decoder. After certain training epochs, the Transformer-based model can be employed for WPF.

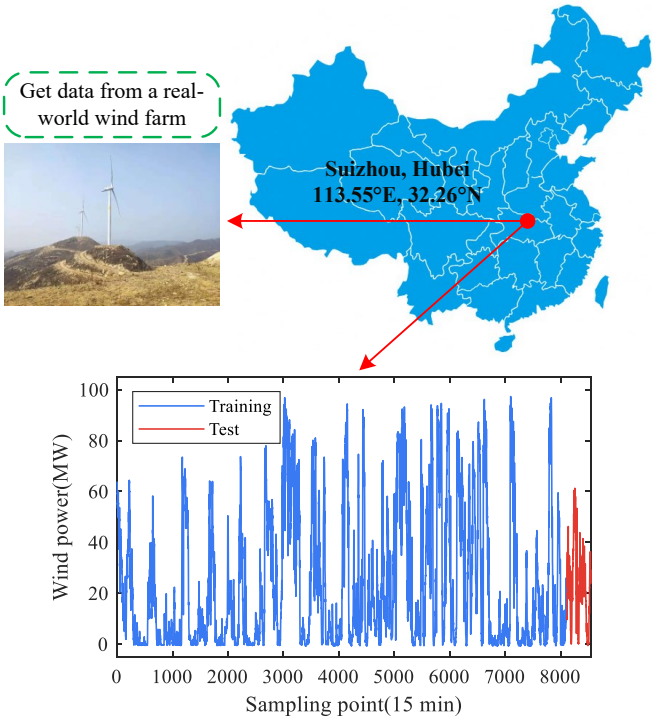

Fig. 4. Experimental study areas and dataset.

IV. Experimental Results

In this section, we demonstrate the forecasting performance between the proposed Transformer-based model and some other deep learning methods, including CNN-LSTM, AGRU, and MLP.

A. Data Description

In this study, a real-world wind farm power dataset is used for validation, which is sampled from the Wan-He wind farm in Suizhou, China. As shown in Fig. 4, Suizhou is located in the northern of Hubei Province, with hilly terrain and rich wind energy resources. Due to low-resolution NWP data, we carried out the one-hour-ahead and two-hour-ahead WPF as examples for algorithm comparisons.

Specifically, Wan-He wind farm consists of three main types of wind turbines, UP121 rated at 2.0 MW, GW131 rated at 2.2 MW and MySe145 rated at 3.0 MW. The total capacity of the wind farm is 99.5 MW. The wind power data is from March 1 to May 31, 2021, and all data are recorded every 15 minutes, as shown in Fig. 4.

B. Experimental settings

1) Reference Forecast Methods: We compare the proposed model with the following methods:

a) CNN-LSTM [20]: A hybrid model combining convolutional neural networks (CNN) and LSTM. Number of features of CNN is set to be 10, the layer of LSTM is set to be 4 with 10 neurons each layer.

TABLE I
PERFORMANCE EVALUATION OF ONE-HOUR-AHEAD WPF

Methods	MAE(MW)	MAPE(%)	RMSE(MW)
CNN-LSTM	6.34	6.37	8.21
AGRU	6.33	6.37	8.20
MLP	6.32	6.36	8.29
Our model	**6.12**	**6.15**	**8.01**

TABLE II
PERFORMANCE EVALUATION OF TWO-HOUR-AHEAD WPF

Methods	MAE(MW)	MAPE(%)	RMSE(MW)
CNN-LSTM	9.22	9.26	11.87
AGRU	9.06	9.11	11.67
MLP	9.22	9.26	11.80
Our model	**9.02**	**9.07**	**11.63**

b) AGRU [12]: An enhanced version of GRU with attention mechanism. The number of attention module is set to be 1 and the hidden neurons is set to be 50.
c) MLP [21]: The multilayer perceptron , an adaptive and highly parallel processing model. The hidden layer of MLP is set to be 2 with 30 neurons each layer.

In addition to the reference methods mentioned above, the hyperparameters of our method are set as follows: there are two encoders with 8 heads in total, resulting in 24 features.

2) Evaluation: In this paper, the performance of forecasting models is evaluated by the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean squared error (RMSE). The MAE, MAPE, and RMSE are defined respectively, as follows:

$$\begin{aligned} MAE &= \frac{1}{N}\sum_{i=1}^{N}\left|P\left(i\right)-\hat{P}\left(i\right)\right|, \\ MAPE &= \frac{1}{N}\sum_{i=1}^{N}\frac{\left|P(i)-\hat{P}(i)\right|}{P_{\max}}, \\ RMSE &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|P\left(i\right)-\hat{P}\left(i\right)\right|^{2}}, \end{aligned} \tag{5}$$

where N is the length of the test set, $P(i)$, $\hat{P}(i), P_{\max}$ represent the actual wind power,the forecasting wind power, and the maximum generating power, respectively.

3) Algorithm Implementation: The tested algorithms are implemented based on PyTorch 1.9.0 and trained on a single NVIDIA GeForce RTX 2070. In the training process, the optimizer chosen for all models is Adam, with a learning rate of 0.001.

C. Forecasting Results

Table I and II demonstrate the forecasting results of all the models on the two tests. The quantitive results show that the proposed model outperforms the other deep learning models in terms of all the three metrics. Specifically, for one-hour-ahead WPF shown in Table I, our model achieves an MAE of 6.12 MW, an MAPE of 6.15%, and an RMSE of 8.01 MW, about 3% higher accuracy than other models.

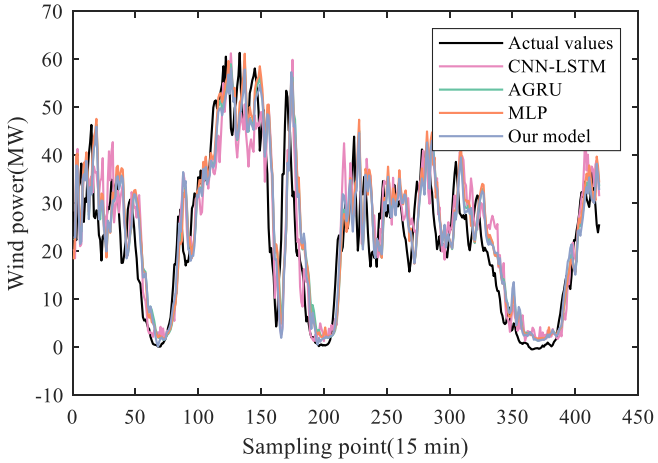

Fig. 5. Forecasting curves for one-hour-ahead WPF.

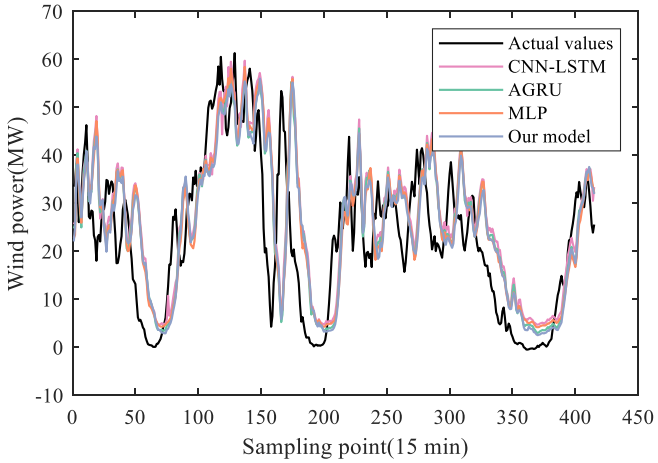

Fig. 6. Forecasting curves for two-hour-ahead WPF.

Similarly, in the two-hour-ahead WPF task, our model also outperforms the other models in all three evaluation metrics. It is noteworthy that another model using attention mechanism demonstrated its advantage over the other two models when predicting longer time horizons. This once again confirms the advantage of attention mechanism in capturing long-term dependencies.

To provide a more intuitive view of the WPF results, the forecasting curves of the two tests are presented in Fig. 5 and 6, respectively. It is evident that the lag phenomenon is more pronounced in the curve of Fig. 6, which is a challenge that cannot be adequately addressed in time series prediction. However, the advantages of the attention mechanism are already reflected in Table II.

D. Discussion

The results of our experiments show that the Transformer-based model for WPF outperforms traditional deep learning models such as LSTM and MLP. This can be attributed to the Transformer's ability to capture long-term dependencies in the

data and its parallelizable structure, which makes it efficient for processing large amounts of data.

However, it should be empirically noted that the performance of the Transformer-based model may be sensitive to the quality and quantity of input data. In addition, the model's interpretability is limited, as it operates as a "black box" like most deep learning models, with no clear understanding of the underlying features it uses to make predictions.

Overall, the proposed Transformer-based model shows great potential for WPF, but further research is needed to address its limitations and optimize its performance in real-world applications.

V. Conclusion

In this article, we propose a novel approach for WPF using a Transformer-based model learning from ChatGPT. Leveraging the potent learning capability of self-attention mechanism in Transformer model, the proposed method is proved to capture the temporal patterns in wind power information accurately. The experimental results demonstrate that the proposed Transformer-based model outperforms traditional forecasting models, achieving higher accuracy in predicting wind power generation. Similar to the ChatGPT to the NLP tasks, our findings suggest that Transformer-based models have significant potential for improving WPF accuracy and can contribute to a more sustainable energy future.

References

[1] A. Cherp, V. Vinichenko, J. Tosun, J. A. Gordon, and J. Jewell, "National growth dynamics of wind and solar power compared to the growth required for global climate targets," *Nature Energy*, vol. 6, no. 7, pp. 742–754, 2021.

[2] H. Ritchie, M. Roser, and P. Rosado, "Energy," *Our World in Data*, 2022, https://ourworldindata.org/energy.

[3] "Wind power capacity in the net zero scenario," 2022. [Online]. Available: https://www.iea.org/data-and-statistics/charts/wind-power-capacity-in-the-net-zero-scenario-2010-2030

[4] P. Roy, Y. Liao, and J. He, "Economic dispatch for grid-connected wind power with battery-supercapacitor hybrid energy storage system," *IEEE Transactions on Industry Applications*, vol. 59, no. 1, pp. 1118–1128, 2023.

[5] X. Dai, G.-P. Liu, and W. Hu, "An online-learning-enabled self-attention-based model for ultra-short-term wind power forecasting," *Energy*, vol. 272, p. 127173, 2023.

[6] Z. Li, L. Ye, Y. Zhao, M. Pei, P. Lu, Y. Li, and B. Dai, "A spatiotemporal directed graph convolution network for ultra-short-term wind power prediction," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 1, pp. 39–54, 2023.

[7] Y. Huang, G.-P. Liu, and W. Hu, "Priori-guided and data-driven hybrid model for wind power forecasting," *ISA Transactions*, vol. 134, pp. 380–395, 2023.

[8] M. Neshat, M. M. Nezhad, E. Abbasnejad, S. Mirjalili, L. B. Tjernberg, D. A. Garcia, B. Alexander, and M. Wagner, "A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the lillgrund offshore wind farm," *Energy Conversion and Management*, vol. 236, p. 114002, 2021.

[9] S. Sharda, M. Singh, and K. Sharma, "Rsam: Robust self-attention based multi-horizon model for solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1394–1405, 2021.

[10] K. L. Jørgensen and H. R. Shaker, "Wind power forecasting using machine learning: State of the art, trends and challenges," in *2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE)*. IEEE, 2020, pp. 44–50.

[11] M.-S. Ko, K. Lee, J.-K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur, "Deep concatenated residual network with bidirectional lstm for one-hour-ahead wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1321–1335, 2021.

[12] Z. Niu, Z. Yu, W. Tang, Q. Wu, and M. Reformat, "Wind power forecasting using attention-based gated recurrent unit network," *Energy*, vol. 196, p. 117081, 2020.

[13] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[14] J. Zhang, J. Pu, J. Xue, M. Yang, X. Xu, X. Wang, and F.-Y. Wang, "Hivegpt: Human-machine-augmented intelligent vehicles with generative pre-trained transformer," *IEEE Transactions on Intelligent Vehicles*, pp. 1–8, 2023.

[15] S. B. Patel and K. Lam, "Chatgpt: the future of discharge summaries?" *The Lancet Digital Health*, vol. 5, no. 3, pp. e107–e108, 2023.

[16] F.-Y. Wang, Q. Miao, X. Li, X. Wang, and Y. Lin, "What does chatgpt say: The dao from algorithmic intelligence to linguistic intelligence," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 3, pp. 575–579, 2023.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[18] K. S. Kalyan, A. Rajasekharan, and S. Sangeetha, "Ammu: a survey of transformer-based biomedical pretrained language models," *Journal of biomedical informatics*, vol. 126, p. 103982, 2022.

[19] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[20] Q. Wu, F. Guan, C. Lv, and Y. Huang, "Ultra-short-term multi-step wind power forecasting based on cnn-lstm," *IET Renewable Power Generation*, vol. 15, no. 5, pp. 1019–1029, 2021.

[21] S. Khazaei, M. Ehsan, S. Soleymani, and H. Mohammadnezhad-Shourkaei, "A high-accuracy hybrid method for short-term wind power forecasting," *Energy*, vol. 238, p. 122020, 2022.