ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

A Multi-Layer Adaptive Transformer Framework for Operational State Recognition in Wind Storage Systems

Shike Wang[1,*], *Bin Chen*[2], *Yabing Sun*[3] *and Guang Shi*[3]

[1] China Resources Power Technology Research Insitute Co., Ltd., Dongguan, Guangdong, China, 513808
[2] China Resources Power Holdings Co., Ltd, Shenzhen, Guangdong, China, 518000
[3] Rundian Energy Science and Technology Co., Ltd, Zhengzhou, Henan, China, 450000

SUMMARY: *The wind generation process is very sensitive to uncertainties based on environmental effects due to the rapid growing wind capacity. In the situation where no means of grid connection exist, the variation in power could compromise safety of power system operations. In this paper, a model of the wind-energy storage system with a doubly-fed generator model is presented. Seven groups of operating states are determined based on the frequency dynamic response performance of the wind-storage system. To enhance the recognition of wind-storage operating states, a novel model (CTM-Net) is proposed based on the combination of multilayer Transformer and Convolutional Neural Network (CNN). When a wind-power failure occurs, the probabilities of change in value of output between 0 and 4 MW and 76 and 80 MW are considered and energy storage devices are allocated considering the chargeable and dischargeable ratio of the system. It has been demonstrated that the wind-storage operating states have obvious variations at different sampling times. The faulty shutdown state (ZT2) identification rate is 99.57% which encourages the effectiveness of state recognition significantly. The operating state of the wind-storage system can be identified accurately using multilayer adaptive Transformer.*

KEYWORDS: *wind storage system; Transformer network; CNN; CTM-Net model; operation state recognition*

1 Introduction

The current energy situation in the global community, coupled with steady economic growth in China, has led to a significant rise in energy consumption. On the other hand, wide-scale use of fossil fuels has resulted in environmental issues like the greenhouse effect, smog, and acid rain [1, 2]. For resolving problems related to energy deficiency and environmental degradation, there is a need for a change in the energy consumption pattern and a reduction in dependence on conventional sources of energy. Using renewable energy in place of fossil energy is an important approach towards environmentally sustainable development [3].Energy sources in the world are slowly shifting from fossil energy sources to renewable ones such as solar energy [4], wind [5], geothermal [6] and biomass [7]. Out of all these types of renewable energy, wind energy has received much attention owing to its high capacity factor, cheap cost of production, and capability of generating electricity in varying climates [8].

The global wind power installed capacity increased by more than 743 GW (24 GW) during

*harmonynext2025@163.com
https://doi.org/10.65102/is2026493
Published: 30 April, 2026

the period 2001–2020. According to data released by the Global Wind Energy Council (2022), wind energy capacity that was newly developed in 2021 amounted to 93.6 GW. For achieving the goal of carbon emissions being equal to zero, the current installation rate needs to quadruple by the end of the century. The above trends mean that wind energy will play an increasingly significant role in future energy production. Wind energy storage can be charged and discharged as required, thus adding the generated wind power of wind farms. Energy storage can compensate for power variations due to rapid change in power generation of wind farms to make improvements in power quality of grid power and match the actual generated power with power demands on the grid side [9, 10]. Energy storage can reduce the intrinsic loss associated with traditional electric grids, increase the application of renewable energy sources, regulate energy supply and demand, suppress power fluctuations, decrease environmental pollution, improve the reliability and efficiency of electric grids [11, 12]. However, wind turbines exist in complicated natural environments throughout their service life and are affected by external disturbances (e.g., variation in wind velocity, temperature, humidity) and internal disturbances (e.g., sand, salt spray). In some cases, turbine health condition is affected. Therefore, it is necessary to conduct regular monitoring and maintenance [13]. In recent years, the advantages of deep learning in information recognition and feature extraction, especially the Transformer architecture [14, 15]. Nowadays, deep learning models show excellent performance in information recognition and feature extraction, especially the development of Transformer models. Existing whole turbine-based operational state recognition methods focus on joint single-component state assessment and mostly use physical modeling, which is computationally complex and extremely label-dependent. In contrast, deep learning is able to extract complex health features directly from the turbine's operational data, and the unsupervised approach can effectively reduce the label dependency, and still achieve efficient operational state identification in the absence of a large amount of labeled data [16, 17]. Thus, it can provide more promising support for abnormal state identification of the whole turbine, which has wide application potential and practical significance.

Deep learning models include long short-term memory networks, deep belief networks, convolutional neural networks, and Transformers, and such models have been widely used in state recognition and prediction of wind turbine. However, there has not been enough research in regard to the detection and evaluation of operation states for wind energy storage systems. In Tchakoua et al.'s paper [18] , the authors systematically analyzed condition-monitoring technologies for wind turbines and evaluated maintenance strategies, and then the authors came up with observations about the development of condition-monitoring technologies in wind energy generation and the issues associated with the development of these technologies. Stetco et al. [19] studied the application of machine learning in the state monitoring of wind turbines; the authors found that neural networks, support vector machines, and decision trees were common methods used for state recognition, especially for identifying blade faults and monitoring generator temperature of the wind turbines. Zhan et al. [20] used an active power prediction model to predict the normal active power value of wind turbines and evaluated the degree of discrepancy between real residuals and normal residuals by Mahalanobis distance, which can be considered as a good evaluation index because it could reflect the power-generating capacity of the wind turbines well. Li et al. [21] introduced squeeze-and-excite (SE) attention module into YOLOv5 to form a deep-learning model of monitoring blade condition of the wind turbines, and the improved convergence rate and efficiency were achieved; the deep learning algorithm was applied to improve the performance of the model. Zhu et al. [22] calculated multi-dimensional data by integrating the advantages of CNN and LSTM under the framework of kernel principle component analysis (PCA) to create a condition-monitoring model of the wind turbines that could supervise the wind farm operation and issue warnings

about the abnormal states of the wind turbines. Xiang et al. [23] developed a fusion model of CNN, bidirectional gated recurrent unit (BRU), and attention mechanism, where the salient information was extracted as high-dimensional features by CNN and delivered to BRU by the attention mechanism layer.

The faults were detected through the training of ResNet-50 and Inception-V3 models with respect to images associated with blade icing from online databases and magazine pictures. The process of detecting icing faults was conducted through the use of Grad-CAM and Score-CAM methods of locating icing and then enhanced U-Net algorithms to separate icing areas [24]. Kong et al. [25] introduced an approach for monitoring of wind-turbine status with the use of CNNs and GRUs. It was suggested to train a normal behavior model with respect to historical information with the use of control charts to detect any anomalies within the wind turbines. Park et al [26] utilized DNNs and ANNs to learn historic information regarding wind turbines. As a result, the effectiveness of the learning system of faults increased; the condition identification rate was higher than 90 percent in the test dataset. Fu et al [27] proposed a method of identifying the state of the wind turbines through deep-learning. By defining the observable variables and creating logical relationships among them by means of using CNN and LSTMs, the fast and reliable detection of fault status was reached. The method introduced by Mohammadi et al. [28] was concerned with the monitoring of wind turbines operating-status and faults using multi-channel CNNs and sensor technologies. The approach achieved an F1 score of 0.94 in the testing real-world data with the correct classification of turbine fault types. The approach of Lu et al. [29] suggested using the optimalization of DBN with the help of the chaotic quantum particle swarm optimization algorithm to derive deep and compressed-sample differences. With the help of such an optimalized DBN, the monitoring of the operating status of the planetary gearbox was achieved. Lyu et al. [30] combined unified and refined LSTM networks with CNNs for developing an interactive learning system to monitor operating status. Interactive learning enables accurate identification of turbine operation states and exceeds the comparative model's effectiveness in terms of fault detection.

Currently, the application of Transformer framework in wind power generation mainly includes abnormal state identification and power generation prediction. Alwadie et al [31] developed a wind turbine operational fault state monitoring system based on the visual Transformer framework, the system captures spatial and temporal characteristics of information such as rotational speed and generator series data through an attention mechanism, the system shows strong robustness on open datasets and improves the efficiency of wind turbine operation. The deep learning technology is also contributing to the efficiency of wind turbine performance. Dhiman et al. [32] applied a model based on Transformer with multi-head attention capability to identify blade icing in wind turbines and gathered data about the performance of turbines located on Hitra Island in Norway. The convolutional neural network and CNN-LSTM network models were used to validate the efficiency of the proposed model by learning and classifying abnormal conditions in wind turbines. Yin et al. [33] developed an FD-Transformer algorithm based on tensor product dimensionality reduction to analyze high-dimensional and unstable fault data of wind turbines. By adopting tensor product relative position matrix dimensional composition method, the high-dimension of data could be decomposed and the features were extracted through deep learning. It was shown that the proposed algorithm achieved 93.65% accuracy when predicting blade fault conditions of wind turbines. Wu et al. [34] built a Multiscale Attention Network with dual-branch Transformer to predict wind energy generation. Comparing with other predicting algorithms, the proposed model showed smaller mean squared error and mean absolute error. And has greater potential in the field of wind power generation prediction.

The paper considers the dynamics and frequency responses of wind-storage systems and

their operating state groupings using doubly-fed generators and wind-energy storage systems models. An identification model for the operating states of the wind-storage system is going to be designed using transformer structures and convolutional neural networks. In addition, a multilayer adaptive Transformer coding layer is set up to enhance the efficiency of identification by obtaining accurate temporal parameters associated with wind-storage operating states. The suggested approach provides the theoretical basis of ensuring the stable work of the wind-energy storage systems.

2 Wind storage system operation state response characteristics

The ongoing depletion of fossil energy supplies and the increasing pollution of the environment have made it a common cause and a realistic solution to accelerate the creation and use of wind power across nations. Nevertheless, wind power generation is quite irregular and unpredictable. With its growing penetration rate year after year, the issue of wind curtailment is becoming more acute. At the same time, there is a swift development of a new round of power system reforms that take the reform of the electricity prices as their central duty. In such a context, only wind parks with a high degree of schedulability and friendliness to the grid can be prioritized in dispatching and benefit from competitiveness in the power market.

2.1 Wind storage system topology

2.1.1 Doubly-fed generator model

The conversion of wind energy into mechanical energy by a wind turbine is a complex aerodynamic process, which if modeled accurately and in detail must take into account the theory of paddle modeling, involving complex geometric problems and lengthy computational problems. To avoid the complexity of the problem, a simplified wind turbine model is often used instead, and the simplified model is described as follows:

The mechanical torque T_m produced by the wind turbine is:

$$T_m = 0.5\rho\pi R^3 V_w^2 C_P(\beta,\lambda)/\lambda \tag{1}$$

The power absorbed by a wind turbine from the wind is:

$$P_m = T_m\omega_m = \frac{1}{2}\rho\pi R^2 C_P(\beta,\lambda)V_W^3 \tag{2}$$

where ρ is the air density, R is the blade radius of the fan, $V\omega$ represents the wind speed, β is the pitch angle, and λ is the tip speed ratio ($\lambda = \omega mR/V\omega$, the speed at the tip of the blade divided by the actual wind speed, the longer the blade, the greater the tip speed ratio at the same wind speed). C_p is the maximum wind energy utilization coefficient, if we want to get the maximum wind energy, we must make C_p reach the maximum value, there are many factors affecting C_p, but it is mainly determined by the pitch angle β and the blade tip speed ratio λ. The expressions are as follows:

$$C_p = 0.22\left(\frac{116}{\lambda_i} - 0.4\beta - 5\right)e^{-\frac{125}{\lambda_i}} \tag{3}$$

Eq. $\lambda_i = 1/\left(\frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1}\right)$.

Since the gearbox part is omitted from the doubly-fed generator, its transmission structure consists of three parts: the wind wheel, the low-speed shaft and the generator rotor. When establishing the mechanical transmission model of the doubly-fed generator, the gearbox can be regarded as a rigid gear, and the rotational inertia of the transmission shaft is equivalent to that of the generator rotor, so the first-order inertia link can be used to describe the transmission system. Namely:

$$\frac{d\omega_m}{dt} = \left(T_e - T_m - B_m\omega_g\right)/J_{eq} \tag{4}$$

where Te is the electromagnetic torque, B_m is the rotational viscosity coefficient, Jeq is the equivalent rotational inertia, and ω_g is the rotational speed of the generator rotor. The rotational speed of the wind turbine drive shaft ω_m is equal to the rotational speed of the generator rotor ω_g.

2.1.2 Structural framework for wind storage systems

The paper addresses the operational condition of the wind-storage system in different conditions and the precise structure will be illustrated in Fig. 1. Wind-power system comprises of direct-drive permanent magnet wind turbines and they are connected to the DC bus through rectifier link. Combination of battery storage and supercapacitor storage constitutes a wind-storage system. Every storage medium has access to the DC bus with a bidirectional DC-DC converter. It is supported by DC bus which is connected to AC bus through contact line. Then it is connected to AC bus with the help of DC-AC converter and finally to power grid with the help of another contact line.

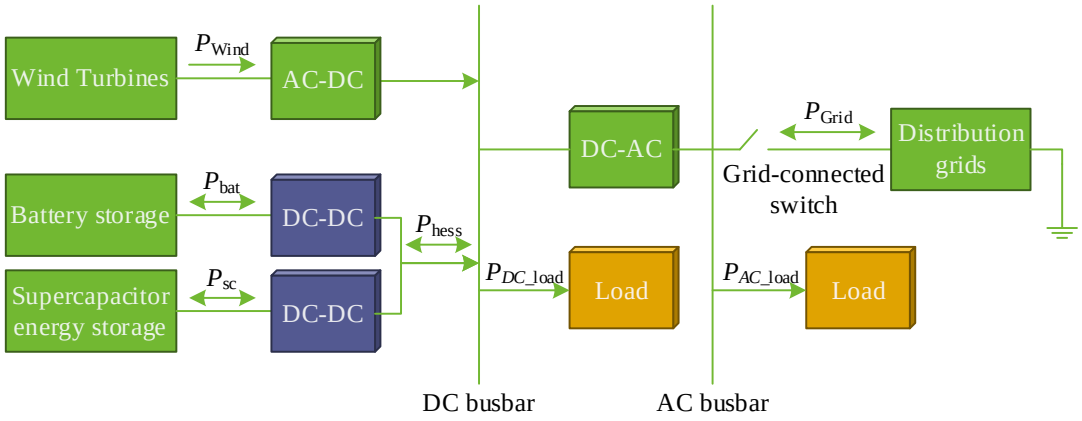


Figure 1: Structural framework of the wind storage system

Using the flow of energy in the system shown in the figure, ignoring all forms of energy loss in the system, we can derive the power balance equation:

$$P_{\text{Gind}} = P_{\text{Wind}} + P_{\text{bat}} + P_{\text{sc}} - P_{\text{DC\_load}} - P_{\text{AC\_load}} \tag{5}$$

where P_{Grid} is the wind power grid-connected power, P_{Wind} is the wind power output power, P_{bat} is the battery storage response power command, and P_{sc} is the supercapacitor storage response power command. $P_{\text{DC\_lood}}$ and $P_{\text{AC\_lood}}$ are the DC load power and AC load power, respectively.

The control of energy storage in the wind-storage system makes use of both battery energy storage and supercapacitor energy storage systems for managing the energy outputs of the wind farm on various frequency scales, thereby making up for the variations in energy output. This process ensures that the output of energy from the wind farm to the grid meets the requirements of the grid and delivers energy to the loads. The wind-storage system goes into islanding mode if the grid connection switch is turned off. In this case, wind energy mostly works according to the requirements of the load side, while the energy storage system acts as the backup power supply. The wind-storage system goes back into the grid connection mode if the grid connection switch is turned on.

2.1.3 Wind energy storage system modeling

Double electric layer capacitors, a relatively new type of energy storage device, are safer, can store more energy per unit capacity, and are able to respond quickly, with fast charging and discharging speeds, and are sustainable.

By analyzing the principle of capacitance, the relational equation of capacitor can be obtained as:

$$I = C\frac{\mathrm{d}U_{\text{dc}}}{\mathrm{d}t} \tag{6}$$

where I denotes the current flowing through the capacitor, C denotes the capacitor capacitance value, $\mathrm{d}U_{\text{dc}}$ denotes the amount of change in capacitor voltage, and $\mathrm{d}t$ denotes the amount of change in time.

In the equivalent circuit, the voltage change due to the series resistance is:

$$\Delta U_{R_{\text{s}}} = IR_{\text{s}} \tag{7}$$

The total voltage variation can be obtained as:

$$\Delta U = I\frac{\mathrm{d}t}{C} + \Delta U_{R_{\text{s}}} \tag{8}$$

The capacitance capacity value can be found from the above equation as:

$$C = \frac{I\mathrm{d}t}{\mathrm{d}U - IR_{\text{s}}} \tag{9}$$

Due to the existence of resistors R_{s}, R_{p}, the supercapacitor in the process of charging and discharging will produce loss, consume part of the energy, when the charging and discharging current is larger, the power consumed on the resistor is larger, which will affect the efficiency of the supercapacitor, so it is necessary to limit the charging and discharging currents within a

reasonable range.

The supercapacitor energy storage system mainly consists of the DC/DC converter and the supercapacitor where the latter acts as the main source of storage. The major role of the supercapacitor is to store excess energy generated by the grid and deliver it to the grid where necessary, thus balancing the amount of electricity being supplied and consumed. Through comparing the operation state of the converter with the reference point, the converter can be manipulated to control the charging/discharging of the supercapacitor. If the voltage of the DC-side is less than the reference voltage, the converter will operate in Boost Mode, whereby it allows the supercapacitor to release its energy to the DC side, balancing the voltage level. However, if the voltage of the DC-side is greater than the reference voltage, the converter operates in Buck Mode, thus allowing the supercapacitor to store excess DC energy.

2.2 Characteristics of wind storage system operation status

2.2.1 Wind storage system dynamic response

To the power grid view, whenever power imbalance happens like the sudden load increase and the decline of frequency, the synchronous generator will automatically react to the frequency change and increase electromagnetic power at an accelerated rate. Conversely, wind power and energy storage with supplementary artificial inertia control may readily offer frequency-modulation power to assist in restoring system power balance.

In order to examine the dynamic properties of system frequency, the present paper develops a frequency-response model of the wind-storage system. The model consists of synchronous generator, wind turbine, energy storage and the load power-response module. The inertia constant of synchronous generator is denoted by H_{g} of the synchronous machine taking into account the wind power penetration rate ρ, D_{sys} is the damping coefficient of the system, K_{L} is the load regulation coefficient, R is the tuning coefficient, T_{s} is the inertia time constant of the turbine, and F_{s} is the characteristic coefficient of the turbine.

The frequency response equation of the wind storage system is:

$$\left(2H_{\mathrm{g}}p + D_{\mathrm{sys}}\right)\Delta f(t) = \Delta P_{\mathrm{d}} - \Delta P_{\mathrm{m}} - \Delta P_{\mathrm{L}} - \Delta P_{\mathrm{WB}} \tag{10}$$

where ΔP_{m} is the synchronous generator power output, ΔP_{L} is the load regulation signal, $\Delta P_{\mathrm{L}} = K_{\mathrm{L}}\Delta f$ and ΔP_{d} is the perturbation power, and ΔP_{WB} is the power output of the WTGs and storage. Where H_{g} can be expressed as:

$$H_{\mathrm{g}} = H_{\mathrm{g1}} \cdot (1-\rho) \tag{11}$$

The power output ΔP_{m} of the synchronous generator is affected by the wind penetration rate, the turbine parameters, and the tuning coefficient, which can be expressed as:

$$\Delta P_{\mathrm{m}} = \frac{1-\rho}{R}\frac{1+F_{\mathrm{s}}T_{\mathrm{s}}s}{1+T_{\mathrm{s}}s}\Delta f \tag{12}$$

Wind-storage support power is time-varying in frequency-response analysis. This paper will initially assess the wind-storage inertia and determine the largest frequency deviation and inertia response time based on the virtual inertia magnitude of the wind-storage system to

achieve frequency safety warning. Once the results of the virtual inertia evaluation are acquired, they can be integrated with the synchronous-generator inertia calculation. Even though this reduced approach could contain some errors, it would be able to significantly lower the complexity of the computation induced by the time-varying study of wind-storage support power.

2.2.2 Operational status of the wind storage system

When distributing power among wind turbines in a wind farm, it should be taken into consideration that both the current operating state of any individual unit and its adjustable power range are important factors when determining power allocation in a reasonable way. The adjustable power range is different across the units and depends on the potential output of the unit in the subsequent cycle, its current generation power and fault condition. Thus, the units should be grouped based on the information about their operating state.

Apart from having a maximum power generation capacity, wind turbine generators have a minimum power generation capacity. In case the lower limit of the assigned power is considered to be the minimum capacity of the generator, frequent shutdowns will happen because of changes in power generation capacity. Thus, this paper restricts the set-power range as follows:

$$\begin{cases} P_{\min set} < P_{n-i}^{set}(t+1) < P_e \\ 0 < P_{stop} < P_{\min set} \end{cases} \tag{13}$$

where n is the category of the unit, i is the number of the unit in the category to which it belongs, P_e is the rated power of the unit, $P_{n-i}^{set}(t+1)$ is the output power of the power control system for the next control cycle set for the nth unit in the category i, and P_{stop} is the output power that triggers the wind turbine unit to The output power that triggers the wind turbine to stop is usually 150 kW in 1.5 MW wind turbines, and $P_{\min set}$ is the minimum value of the set output power.

In addition, the state quantity e_{n-i} is used to indicate whether the ith unit in the nth class is in a faulty state or not, i.e:

$$e_{n-i} = \begin{cases} 0 & \text{Normal operation} \\ 1 & \text{Failure shutdown} \end{cases} \tag{14}$$

In this paper, the wind turbines in wind farms are categorized into seven classes according to the operating states such as the output power in the current control cycle, the predicted output power in the next control cycle and the current fault conditions, and their specific classifications are shown in Table 1. In the table, t and $(t+1)$ denote the current control cycle and the next control cycle, respectively, $P_{n-i}^{real}(t)$ is the output power of the nth unit of the nth class at the current time, and $P_{n-i}^{forecast}$ is the predicted output power of the ith unit of the nth class at the next control cycle. The output power of the unit in the next control cycle is predicted.

Table 1: Run state classification

Code	Category	Current power	Predicted power	Fault
ZT1	Communication anomaly	-	-	-
ZT2	Fault shutdown	0	-	$e_{2-i}=1$
ZT3	Zero power	$0<P_{3-i}^{real}(t)$	$0(P_{stop}>P_{3-i}^{forecast}(t+1))$	$e_{3-i}=0$
ZT4	Low power	$0<P_{4-i}^{real}(t)$	$P_{stop}<P_{4-i}^{forecast}(t+1)<P_{minset}$	$e_{4-i}=0$
ZT5	Power decline	$P_{minset}<P_{5-i}^{real}(t)$	$P_{5-i}^{real}(t)<P_{5-i}^{forecast}(t+1)$	$e_{5-i}=0$
ZT6	Power increase	$P_{minset}<P_{6-i}^{real}(t)$	$P_{6-i}^{real}(t)\geq P_{6-i}^{forecast}(t+1)$	$e_{6-i}=0$
ZT7	Ready to start	0	$P_{stop}<P_{7-i}^{forecast}(t+1)$	$e_{7-i}=0$

In order to describe the various types of units are defined as follows, m_n is the number of units of the nth class, ΔP_n is the value of the power change of the nth class in the next cycle, $P_n^{forecast}(t)$ is the predicted value of the output power of the nth class in the current cycle, and $P_n^{real}(t)$ is the actual output power of the nth class in the current cycle.

3 Wind storage system operation state identification model

The policy incentives and increasing energy demands have facilitated the massive growth of wind-based renewable energy. Nevertheless, wind power is associated with intrinsic volatility, intermittency, and poor anti-peaking properties that create significant challenges to system dispatch planning, once wind power is incorporated into grid operation. In case of large-scale integration of renewable energy, power dispatch departments would typically use somewhat crude and cautious generation planning approaches. Hence, using the reserve capacity of other units after the intermittent power supply has been integrated into the regional grid can not completely absorb the wind-power output or overcome the combined effect of both system-load volatility and prediction errors. To decrease system-regulation pressure, more sophisticated energy storage technology needs to be used to even out renewable-energy output and enhance system flexibility and elasticity.

3.1 Deep Learning Related Technologies

3.1.1 Convolutional Neural Networks

The convolutional neural network refers to the neural network based on various convolution processes. The neural network architecture is inspired from biological neural network architecture and human neurons connection system. The basic components that constitute a CNN include the input layer, the convolution layers, pooling layers, and the fully connected layers. The connections between the data nodes in the network are used to handle many mapping functions.

(1) Convolutional layer. The convolutional layer is the first layer of a convolutional neural network since it primarily executes convolution operations on the raw input signal. To extract features of the input data, various convolution kernels are used. These feature maps created by the convolutional layer have certain combinations and the matrix-generation expression of the feature map is defined as:

$$a_j^l = f\left(z^l\right) = f\left(\sum_{i \in M_j} a_j^{l-1} * K_j^l + b_j^l\right) \tag{15}$$

where a is the feature map, M is the set of feature maps, K is the convolution kernel, b is the bias term corresponding to the feature map, l and i are the l th i th layer, respectively, and * is the convolution operator.

(2) Pooling layer. The pooling layer is placed after the convolutional layer. The role of this layer is to lower the dimension of the input data, improve the network's reliability, and prevent overfitting. Common types of pooling include down-sampling, which scans through the output from the convolutional layer using a specific stride and calculates the maximum or average value in a specified window. The equation for uniform sampling is illustrated below:

$$y_{mn} = \frac{1}{S_1 S_2} \sum_{j=0}^{S_2-1} \sum_{i=0}^{S_1-1} X_{m*S_1+i, n*S_2+j} \tag{16}$$

where X is the input feature map in two dimensions, y is the sampled feature map, and the size of the sampled region is $S_1 * S_2$.

(3) Fully connected layer. The fully connected layer acts as a relay to get the output by connecting different weights, and has the function of processing and integrating the obtained distributed feature representations and sending them to the specified space. It can output the features learned from the network model according to the specified dimensions.

(4) Classification Layer. The classification layer can process the groups of feature vectors extracted from the previous layers of the network with probabilities by means of the corresponding classification function for the purpose of classification and outputting classification probabilities. Convolutional neural networks usually use the SoftMax layer as the classification layer, and its activation function is the SoftMax function, which calculates the features input from the fully connected layer and converts them to other real vectors of the same dimension. The training set samples are:

$$T = \left\{\left(x^{(1)}, y^{(1)}\right), \left(x^{(m)}, y^{(m)}\right)\right\} \tag{17}$$

where $x^{(i)}$ is a random sample in the training set and $y^{(i)} \in (1, 2, 3, k)$ is the category label of the sample in the training set.

The SoftMax layer computes the conditional probability on the input vector, i.e:

$$P(y = j \mid x), (j = 1, , k) \tag{18}$$

The SoftMax function is represented as:

$$h_\theta\left(x^{(i)}\right) = \begin{bmatrix} P\left(y^{(i)} = 1 \mid x^{(i)}; \theta\right) \\ P\left(y^{(i)} = 2 \mid x^{(i)}; \theta\right) \\ P\left(y^{(i)} = k \mid x^{(i)}; \theta\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e_i^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{19}$$

where $\theta = \left[\theta_1^T, \theta_2^T, \theta_3^T, \theta_k^T\right]$ is the connection weight between the input of the network layer

and the k th layer, and it is a parameter of the model's training and learning.

3.1.2 Transformer model

The architecture of the classical Transformer model in natural language processing is depicted in Fig. 2. Both its encoder and decoder consist of Transformer blocks with a common structure. The major elements of every Transformer block consist of Multi-Head Attention (MHSA), Feedforward Network Layer (FFN), Layer Normalization (LN), and Residual Connection.

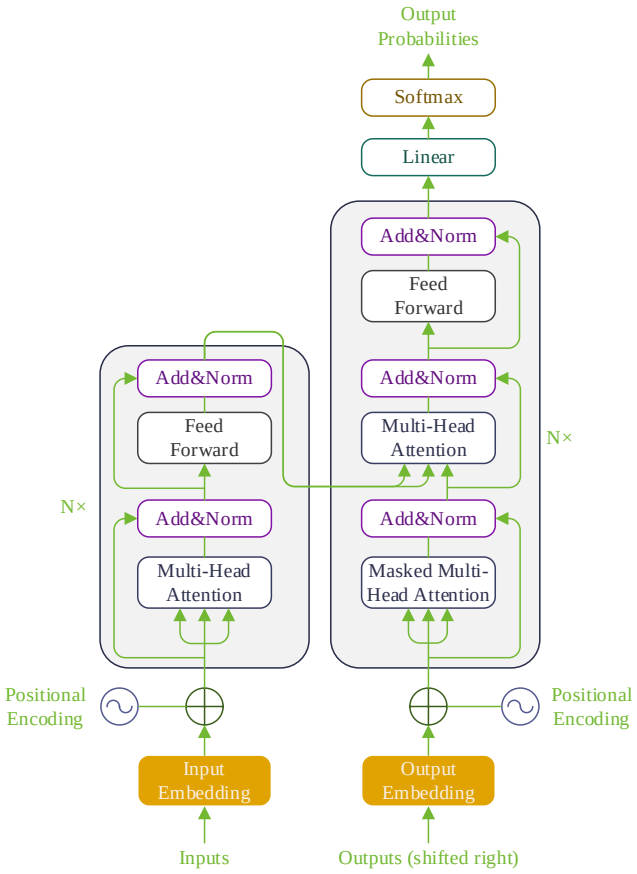


Figure 2: Transformer Model Structure Diagram

The purpose of the self-attention mechanism is to capture the correlation between all N input vectors, first, for the input sequence $\mathbf{X}$, by defining three learnable parameter matrices $\mathbf{W}^Q \in^{D\times D_q}$, $\mathbf{W}^K \in^{D\times D_k}$ and $\mathbf{W}^V \in^{D\times D_k}$ $\left(D_q = D_k\right)$, which are linearly mapped to a sequence of three vectors, the query vectors $\mathbf{Q} = \mathbf{X}\mathbf{W}^Q$, key-value vector $\mathrm{K} = \mathrm{X}\mathrm{W}^K$ and value vector $\mathrm{V} = \mathrm{X}\mathrm{W}^V$. Second, the scaled dot product is used to compute the attention weights $\mathbf{A} \in^{N\times N}$, which are given by the formula:

$$\mathbf{A} = softmax\left(\frac{\mathbf{Q}\mathbf{K}^{\mathrm{T}}}{\sqrt{D_q}}\right) \tag{20}$$

where D_q is the dimension of $\mathbf{Q},\mathbf{K}$. Finally, the output vector $\mathbf{Z} \in^{N\times D_v}$ of the SA is obtained by the following equation:

$$\mathbf{Z} = \mathrm{SA}(\mathbf{X}) = \mathbf{AV} \tag{21}$$

SA procedure is not concerned about the positional order of elements in the input sequence. Thus, position encoding is typically included in the initial input sequence such that the model can more easily learn long distance dependencies. Manual setting of positional encoding is also typical. One popular way of doing it is by creating position encodings of the same dimensionality through sine and cosine functions that may be formally represented like this:

$$PE_{(pos,2i)} = \sin\left(pos / 10000^{2i/d}\right) \tag{22}$$

$$PE_{(pos,2i+1)} = \cos\left(pos / 10000^{2i/d}\right) \tag{23}$$

where i denotes the dimension index in the position encoding and pos denotes the position of the element in the sequence.

The SA captures the correlation between all vectors of the input in a linear projection space. To enhance the representation of the model, the MHSA contains multiple independent SA layers (headers) that allow the model to model different interaction information in multiple different projection spaces simultaneously. Assuming that the MHSA has to learn the correlation of input sequences in h projection spaces, the MHSA defines its own learnable weight matrix for each header, i.e., $\left\{\mathbf{W}^{Q_i}, \mathbf{W}^{Z_i}, \mathbf{W}^{V_i}\right\}$. $i = 0(h-1)$. The mathematical definition of MHSA is denoted as:

$$\mathrm{MHSA}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \left[\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_{h-1}\right]\mathbf{W}^O \tag{24}$$

where [ ] denotes the splicing operation, $\mathbf{W}^O \in^{hD_v \times N}$ is the output projection matrix, and $\mathbf{Z}_i$ is the output vector of each SA, whose corresponding computational procedure is:

$$\mathbf{Z}_i = softmax\left(\frac{\mathbf{Q}\mathbf{W}^{Q_i}\left(\mathbf{K}\mathbf{W}^{K_i}\right)^{\mathrm{T}}}{\sqrt{D_q / h}}\right)\mathbf{V}\mathbf{W}^{\mathrm{V}_i} \tag{25}$$

Feedforward network FFN on the output of MHSA ($\mathbf{Z}$) undergoes further transformations to enhance the expressive power of the model. It mainly consists of two linear transformation layers and a nonlinear activation function. The computational process is represented as:

$$FFN(\mathbf{Z}) = F_2\sigma\left(F_1\mathbf{Z}\right) \tag{26}$$

where F_1 and F_2 are the parameter matrices of the two linear transformation layers and $\sigma(\cdot)$ represents the nonlinear activation function.

3.2 CTM-Net State Recognition Model

3.2.1 CTM-Net network model design

The proposed CTM-Net state recognition network can be effectively used to improve the recognition accuracy of operating states of wind-storage systems. The precise architecture is

represented in Fig. 3. The network applies a Transformer encoder to the sequential data of wind-storage operating parameters and pays attention to the features that occupy different positions within the sequence, thus, it is able to capture long-range dependencies. A CNN-based multiscale feature extraction module is introduced to fuse cross-scale features, which also enhances the network performance. Also, an enhanced residual module is employed to mitigate gradient vanishing, and a channel-attention mechanism is utilized to enhance the network perception of important information. The expressive capacity of the model is increased by increasing network depth, and the effectiveness of recognition accuracy is achieved.

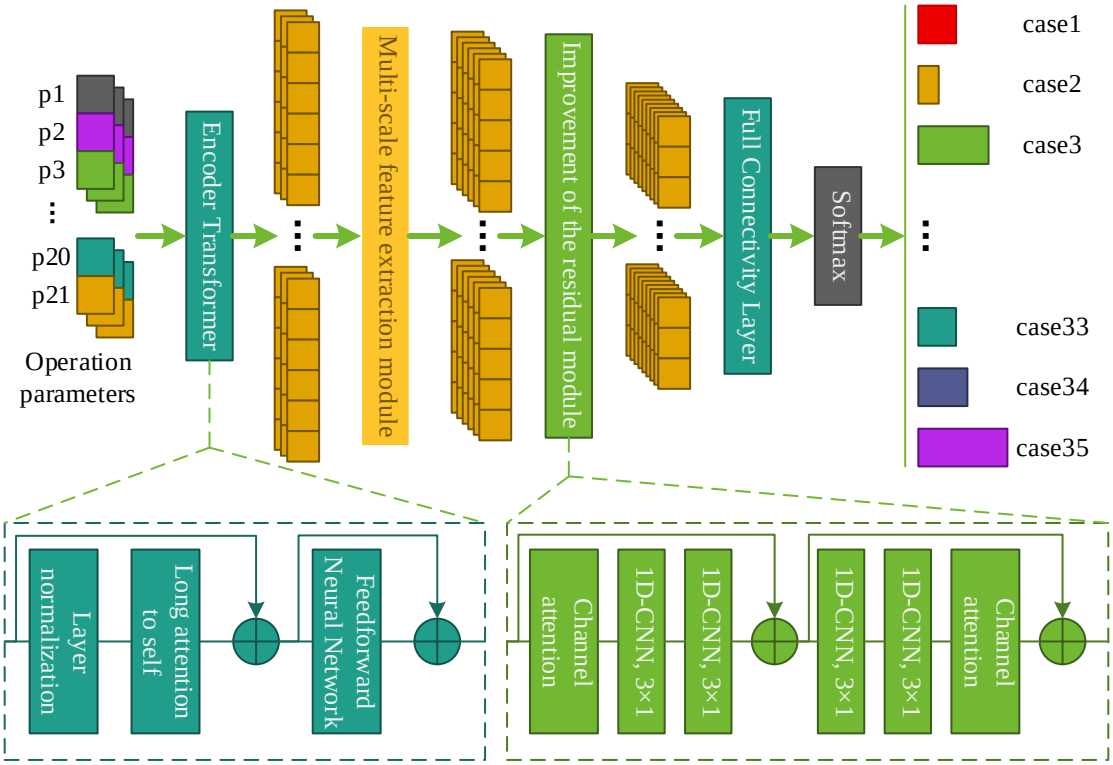


Figure 3: Design of CTM-Net Network Model

The CTM-Net model uses wind-storage operating-state parameters measured at the same sampling interval as the model input. The data are first entered into the Transformer encoder module which allows the model to learn longer range sequence features with an extremely high level of accuracy. Next, the data features are extracted using the multiscale feature extraction module at various scales. Thereafter, improved residual module further increases the depth of the network and channel-attention mechanism enhances the perception of important information which leads to enhancement in feature extraction ability. Finally, the fully connected layer integrates the extracted high-dimensional features where the SoftMax classifier yields the results of recognition of operating states of the wind-storage system.

3.2.2 Multilayer Adaptive Transformer Coding

Transformer model captures interactions between feature blocks across a global context using its multi-head attention mechanism, which allows better extraction of global context information. Additional spatial information may be stored in order to recognize operating-states, enhancing state-recognition accuracy and correctness. In order to take advantage of all Transformer benefits, this paper puts forward a light-weight version of Transformer Encoder into the CTM-Net model. The module has a global self-attention mechanism that processes and

aggregates information on the feature graph. Consequently, the amount of network parameters can be decreased without losing accuracy of detection and the efficiency and precision of determining operating states enhanced.

In the Transformer Encoder module, in order to process the large-scale data input, the input is first divided into multiple patches, and after adding positional encoding to each patch, the one-dimensional vectors are transformed into Embedded Patches, which are used as inputs to the Encoder module. After layer normalization in the Encoder module, multiple sets of input data are processed in parallel by the Multihead Attention Module, which employs multiple self-attention mechanism modules to perform matrix multiplication by transposing the input query matrix Q and key-value matrix K to obtain a correlation vector matrix, and then, after tensor scaling and masking operations, in order to avoid vanishing of the gradient due to the SoftMax activation function operation, the correlation vector matrix is divided by the correlation vector matrix. After tensor scaling and masking operations, in order to avoid gradient loss due to SoftMax activation function operations, the correlation vector matrix is divided by the square root of its dimension to complete the normalization operation. Then, the normalized result is multiplied with the value matrix V to obtain the weighted output result. The multi-head attention mechanism module does not only compute the current data in data processing, but also integrates the feature information of other data in the context.

When the data passes through the Multihead Attention Mechanism, it undergoes Dropout random deactivation processing, after which it is fused with the input data using residual connection, and again after layer normalization it enters into the MLP Multilayer Perceptron Machine module, and finally fuses with the input data using residual connection and then undergoes random deactivation to prevent overfitting.

3.2.3 Multi-scale feature extraction design

CNN extracts target features layer by layer through different receptive fields, and the size of the receptive field determines the effectiveness of the extracted features. When the receptive field is small, the observed features are localized and can retain the information in the input data better. When the receptive field is large, too many invalid features are obtained. Inspired by the structure of InceptionV1, this section designs a multi-scale feature extraction module, which samples the input operating state parameters at different granularities to obtain features at different scales, which enhances the diversity of features and facilitates the operating state identification.

Considering the problem of operating state parameter features and network parameters, a multi-scale feature extraction module is designed in this section. The module consists of three branches, given that the network input flight parameters are 20*1 one-dimensional vectors, each branch uses a 1*1 convolution kernel, a 3*1 convolution kernel and a 5*1 convolution kernel, respectively, and the different scale features obtained from the convolution of the three branches are fused through the method of feature fusion, which is used as the feature input for the next layer of the network.

4 Empirical analysis of the operational status of the wind storage system

The problem of energy crisis is becoming more acute, and environmental issues are being regarded as more important, making the development of renewable energy one of the most significant strategic decisions to create a resource-efficient and environmentally clean society. Wind power and other new energy outputs are characterized by strong volatility and uncertainty.

Therefore, configuring energy storage at new energy field stations can reduce the adverse influence of wind power and other intermittent sources on the grid from the generation side. This also provides an important way to shift new energy from passive grid acceptance to active grid participation.

4.1 Energy storage efficiency of wind energy storage systems

4.1.1 Probability distribution of different energy storage capacities

As an illustration, consider a wind farm in the province of N, which consists of 20 sets of 4 MW doubly-fed wind turbines. The cut-in and cut-out wind speeds are 4 m/s and 24 m/s respectively. The rated wind speed is 12 m/s and the wind-power loss rate is 0.05. All wind turbines are installed in regular formation with no blockage, so the scale parameter and shape parameter of the wind-speed distribution were chosen as 5.25 and 1.65 respectively.

The wind-farm output is separated into 20 intervals and the probability density of wind-farm output in each interval is computed as a result. Fig. 4 illustrates the probability distribution curves of wind-farm output in two conditions, i.e., without taking into account operational failure and taking into consideration operational failure. It can be seen on the figure that, given the consideration of wind-farm operational failure, the probability density of the output between 0-4 MW becomes higher and that of the output between 76-80 MW becomes lower than that of the no-failure scenario. The outcome indicates that wind-power operational failure has an impact on wind-farm output reliability. In particular, when wind-power failure is taken into account, the probability density attains the maximum in the 0-4 MW and 76-80 MW output ranges. The probability of wind-power output rises in the 0-4 MW range and falls in the 76-80 MW range. Based on this information, the output density of the wind farm is much lower than the demand of 12 MW of the auxiliary engine in the power plant, and it is hard to achieve the required output reliability to meet the startup demands.

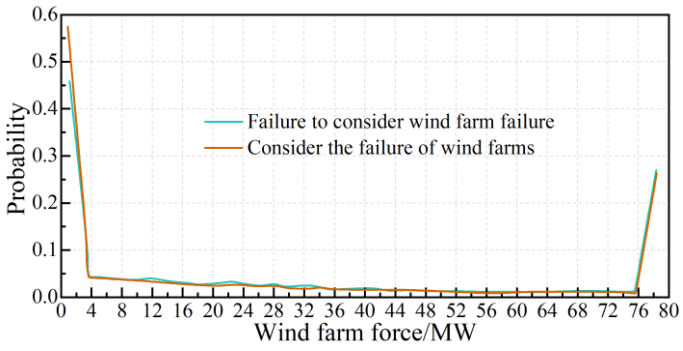


Figure 4: Yield probability distribution curve

Following the results of the analysis, it is determined that 3 MW of constant-output energy storage compensation is installed in order to improve the performance of wind farms, and subsequently, the probability density of the wind-storage output is analyzed. The results are presented in Figure 5. We can observe that after wind-energy storage has been introduced, the total horizontal coordinate of the output probability graph shifts slightly to the right. As a result, the portion of the curve with high probability density is inclined towards higher output-capacity scale.

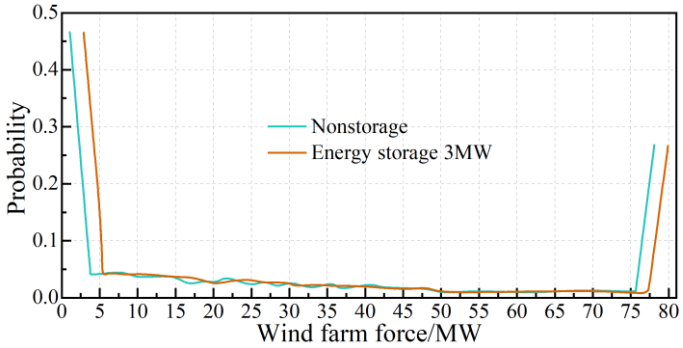


Figure 5: The output probability curve of the configuration energy storage

In view of the fact that the increment for energy storage capacity considered in this analysis is 3 MW, a separate energy storage capacity is considered for every possible wind farm output level. This probability distribution can be seen in Figure 6. As the energy storage capacity increases, the entire probability density curve for wind farm storage system output shifts horizontally, meaning the wind farm storage system output increases.

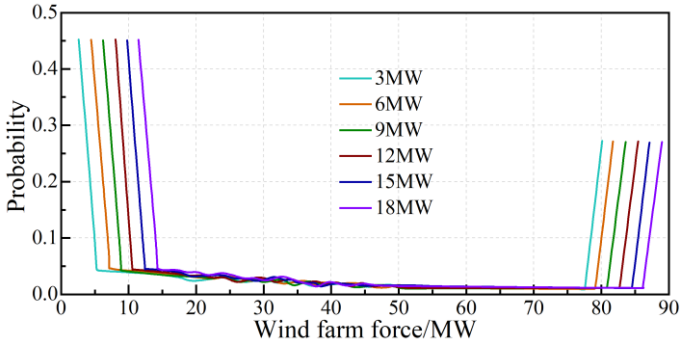


Figure 6: The probability of output under different energy storage capacity

4.1.2 Stable operational performance of wind energy storage

Wind-power production and battery-based storage need to act in concert to ensure that wind farms remain steady when starting up and connecting to the grid. To accomplish this purpose, this paper develops a response-oriented control approach to the integrated wind-storage system. The approach derives the optimal output command of each storage unit adaptively and uses it in the frequency regulation tracking. This is followed by the frequency-related dynamic adjustment controller which changes the outputs of various storage units as per the dispatching orders of the wind-power storage platform, creating a closed-loop regulation system. This would link frequency regulation to state switching in system operation in this manner. Accordingly, the reasoning leads to the conclusion that a coordinated regulation scheme of multi-storage units is set up and its implementation process is presented as follows:

(1) The general aim of the output is being computed with the help of the power data gathered on the wind-power storage platform to keep the frequency regulation balance as well as provide a steady operation in the start-up period.

(2) The adaptive storage regulation approach and the storage unit task allocation system enable determining the working state and output of each battery-storage unit.

(3) At startup, every single battery-storage unit performs local station-level regulation based on the startup conditions and the output-tracking algorithm. Next, the frequency regulation calculation module receives the current state of charging of the integrated wind-power storage unit. Such feedback mechanism goes on till the end of the start-up process.

The SOC deviation index α and the improvement index β of critical overcharge and overdischarge SOC are used as evaluation metrics to assess the effectiveness of the proposed coordinated regulation scheme during startup. The integrated wind-power storage unit startup simulation model is constructed in PSCAD/EMTDC and the main simulation parameters are given in Table 2. In this simulation environment, the desired operating level of storage is defined and the performance of the suggested wind-power storage coordination approach is studied further.

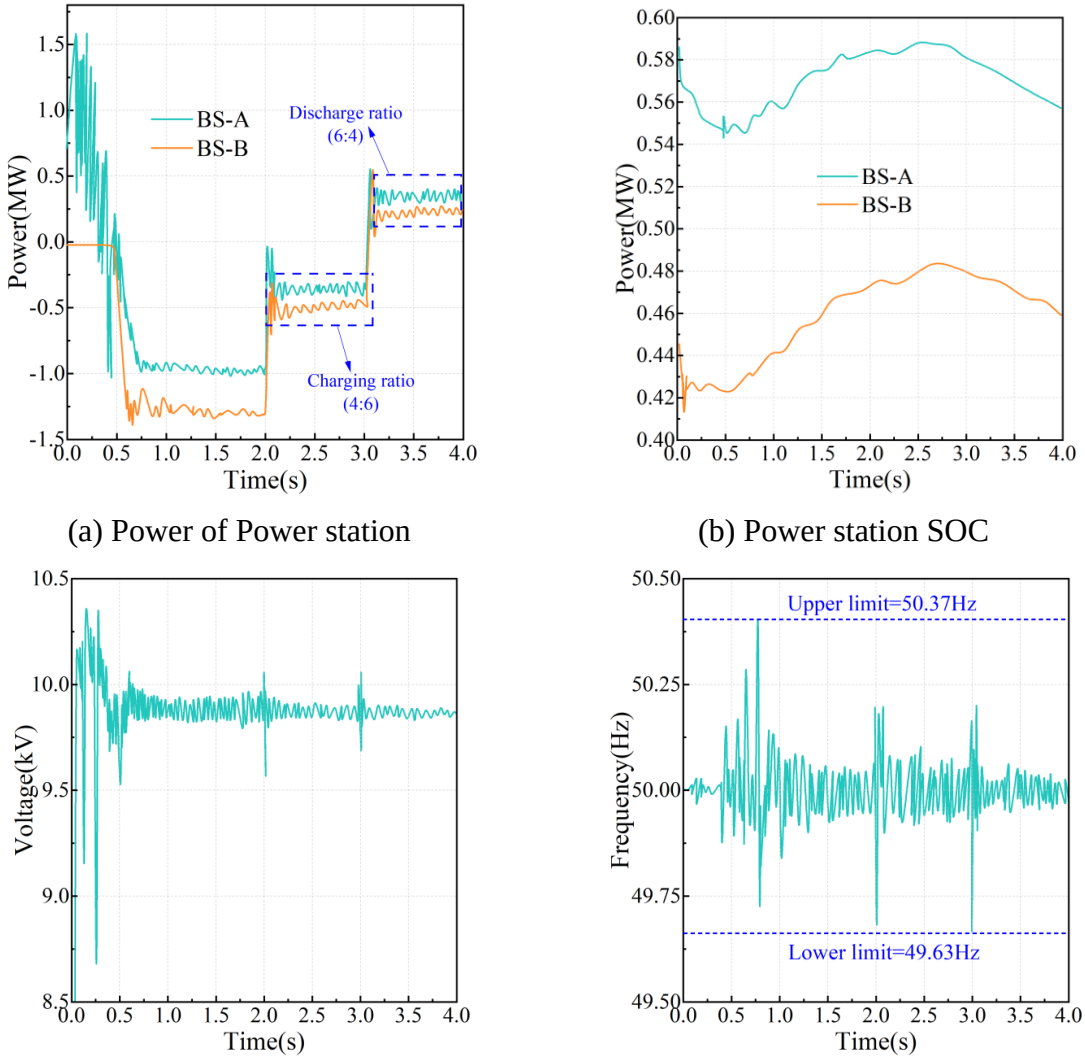
Table 2: Simulation parameters for energy storage coordinated control

Name	Parameters	Numerical value
Operating range of energy storage SOC	Stable interval	[0.25, 0.75]
	Critical interval	[0.75, 0.85], [0.25, 0.35]
	Pre-shutdown interval	[0.75, 1.00], [0.00, 0.25]
Battery energy storage BS-A	Rated capacity/MW·h	0.45
	Max charge and discharge power/MW	1.25
Battery energy storage BS-B	Rated capacity/MW·h	0.45
	Max charge and discharge power/MW	1.25
Wind turbine WT-A	Rated wind speed/$m \cdot s^{-1}$	12
	Rated capacity/MW	3
Wind turbine WT-B	Rated wind speed/$m \cdot s^{-1}$	12
	Rated capacity/MW	3
Busbars for wind farms	Bus voltag/kV	12
	Line impedance/$\Omega \cdot km^{-1}$	0.235+j0.125
	Frequency/Hz	60

At the steady state of energy-storage operation of the wind-storage system, the system activates itself within 02 seconds. Once the wind-power output stabilizes at 3 MW, the first load is fixed at 1 MW. Subsequently, the start-up load is increased by 1 MW after 2 seconds and again by 1 MW after 3 seconds. The initial SOC of the two groups of energy storage power stations is given as 0.6 and 0.5 respectively. If energy storage is kept in the normal stabilization range, there is no need to correct the stable operation strategy of the wind-storage system, and allocation can be done based on the chargeable and dischargeable fraction of each energy storage station. Figure 7 illustrates the start-up features of the wind-storage system in the steady-state mode of operation. Figures 7(a)-(d) demonstrate the power of each energy storage station, the variation of the SOC of each energy storage station, the bus voltage and the bus frequency of the wind farm respectively. Table 3 shows the start-up operation characteristics of the wind-storage system.

The wind-storage system can vary its output power based on the wind-power output and auxiliary-engine output as illustrated in the figure. In this case, energy storage BS-A uses V/T control and energy storage BS-B uses PQ control. When the load is loaded, there is a power-command delay and the output power is slightly oscillating. Between 2.0-3.0 seconds, the wind-power output is higher than the auxiliary power, and the energy storage system is being charged. It does so since the energy storage BS-B has a greater chargeable capacity than the BS-A, and the charging power of BS-B is also greater than the charging power of BS-A. Between 3.0-4.0 seconds, wind-power output is lower than auxiliary-machine power, and the energy storage system is discharged. Together with the information in the table, the wind-storage system created in this paper triggers the cooperative distribution strategy. This strategy may establish the charging and discharging power of every energy storage unit based on the charge-discharge extent of each unit.

When the traditional allocation strategy is adopted, each storage power station allocates power according to a 1:1 ratio. As a result, the charging time becomes too long, which may cause overcharging of BS-A, while the discharging time also becomes too long and may lead to overdischarging of BS-B. Therefore, the wind-storage system enables energy storage stations to allocate power according to the charge-discharge ratio. In this case, the charging and discharging ratios are 4:6 and 6:4, respectively, and the charging/discharging power can be adjusted in real time. This effectively prevents overcharging and overdischarging of energy storage.



(a) Power of Power station

(b) Power station SOC

(c) Bus voltage

(d) Bus frequency

Figure 7: The start-up characteristics of the wind storage system

Table 3: The start-up working characteristics of the wind storage system

Method	-	α/%	β/%	Startup result
This method	BS-A	3.52%	-	Success
	BS-B	2.16%	-	
Traditional method	BS-A	2.75%	-	Success
	BS-B	2.75%	-	

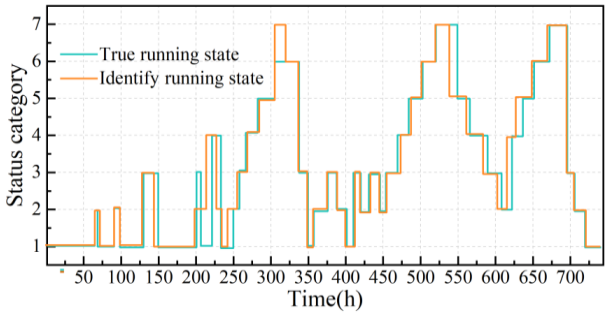
4.2 Identification of the operational status of the wind storage system

4.2.1 Operational status identification results

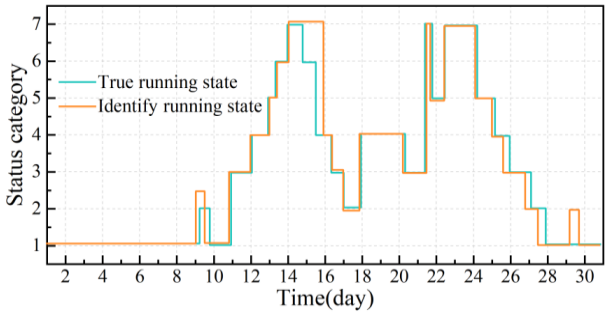
The CTM-Net model is created based on PyTorch in this paper. The Transformer is trained with 8 layers and multi-head attention module of 8 heads. The optimization function is chosen as Adam and Cross Entropy Loss Function (CE) is chosen as the loss function. The dimension of the pooled output is kept the same as that of the input. L2 regularization coefficient is 0.0001, batch size is 128 and initial learning rate is 0.005. At 50 training iterations, the learning rate is reduced to 0.0005. The fully connected module has 2 layers. Lastly, the SoftMax function is applied in classification. The value of the dropout is 0.6 and the activation function in the feed-forward neural network layer is ReLU. Wind-farm data in the Province of N in September 2025 are used to train the experiment, whereas data in October are used to validate it.

According to the wind-storage system operating-state recognition model developed in the last section, there are seven various types of states, i.e., ZT1 - ZT7, in the wind-storage system operating state. Operating states of the wind-storage system at various sampling intervals are subsequently identified. The results of the recognition are shown in Fig. 8, with Fig. 8(a) and Fig. 8(b) representing the sampling intervals of 1 hour and 1 day, respectively.

This figure illustrates the fact that the volatility of wind-storage operating parameters differs across various sampling intervals and this, in turn, leads to the changes in the volatility of operating-state identification. Nevertheless, the general tendency is the same. It means that the CTM-Net model, which is founded on the multilayer adaptive Transformer, can correctly identify operating states by using characteristic variables derived by wind-storage operating parameters. The model has a good identification performance and can be used to identify the operating states of the wind-storage apparatus. Thus, it may offer the reliable data support to ensure the stability of the wind-storage system operation.



(a) Sampling interval = 1hour



(b) Sampling interval = 1day

Figure 8: The wind storage system runs the status recognition result

4.2.2 Operational State Confusion Matrix

For the wind storage system operating states, a total of seven different state categories are defined in this paper, including communication anomaly, fault shutdown, zero power, low power, power drop, power rise, and pending startup. Taking the data of October 2025 as a test, the number of each operation state is different, and the CTM-Net model designed in this paper is used to develop the operation state identification. The confusion matrix is used as an evaluation index to obtain the confusion matrix for recognizing the operating states of the wind storage system as shown in Figure 9.

As shown in the figure, CTM-Net achieves the strongest recognition effect for the faulty shutdown state (ZT2). This state contains 470 fault samples, and the total recognition rate reaches 99.57%. The recognition rates of the other operating states range from 97.61% to 98.97%. These results demonstrate that the CTM-Net model established through the multilayer adaptive Transformer network can accurately identify wind-storage operating states and has clear advantages in this task. This advantage mainly comes from the use of the multilayer adaptive Transformer network as an encoder to extract time-series data of wind-storage operating states, together with the multiscale CNN for multiscale feature extraction, which further improves the recognition accuracy of wind-storage operating-state parameters.

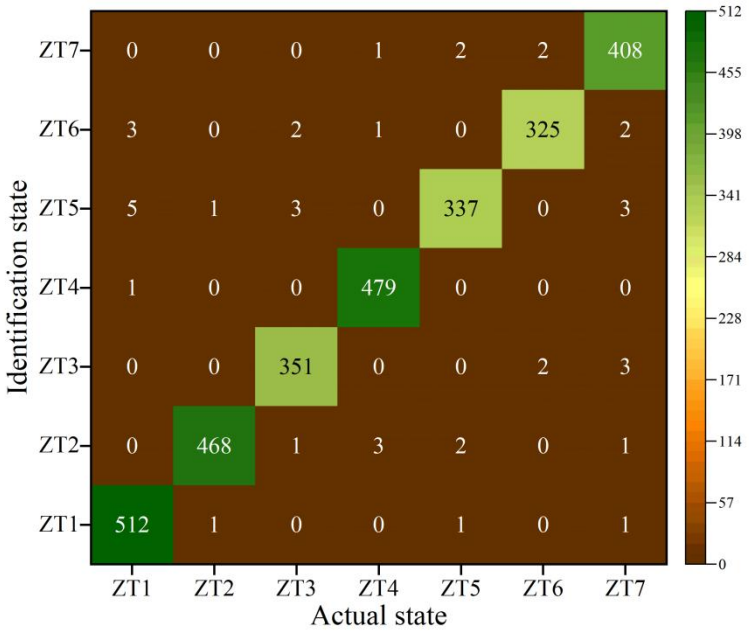


Figure 9: The identification confusion matrix of the running state

4.2.3 Model Recognition Efficiency Comparison

When using wind-storage operating-state identification in practice, it is important to not only consider recognition accuracy, but also the amount of time spent on state identification, i.e., recognition efficiency. Hence, in this section the traditional Transformer model is used as a baseline to further assess the performance of CTM-Net in identifying the wind-storage operating-states. There are ten measurement points spread across various parts of the wind-storage system. The time taken by Transformer and CTM-Net to recognize operating-state data of different measurement points is measured independently and the comparative outcomes are shown in Fig. 10.

On the whole, the figure shows that the average computing time of Transformer and CTM-Net are 73.46 sec and 67.09 sec, respectively. The average computing time of CTM-Net is 8.67

percent less than that of the Transformer model. The time taken by Transformer and CTM-Net to recognize the wind-storage operating states changes with the measurement points. At measurement point 6, the time taken by both the models to recognize the feature vectors gathered using the wind-storage operating states is lower compared to other measurement points. The comparison of computing time reveals that Transformer usually consumes more time than the CTM-Net at every measurement point. Although the Transformer algorithm has high feature-extraction ability and it will be able to find the final solution more quickly in the solution space due to the distributed computation it uses, the local search time is significantly higher than that of CTM-Net. Hence, its total performance is also not as good as that of CTM-Net. Due to the fact that although the Transformer algorithm has excellent feature extraction performance and can find the final solution in the solution space faster through distributed computation, the local search time of the Transformer algorithm is much longer than that of the CTM-Net model, which is not as good as the CTM-Net model. This is due to the fact that although the Transformer algorithm has excellent feature extraction performance, it can find the final solution in the solution space faster through distributed computing, but the local optimization performance of the Transformer algorithm is weak, resulting in a significant decrease in the efficiency of the algorithm, making the algorithm's computation time rise significantly.

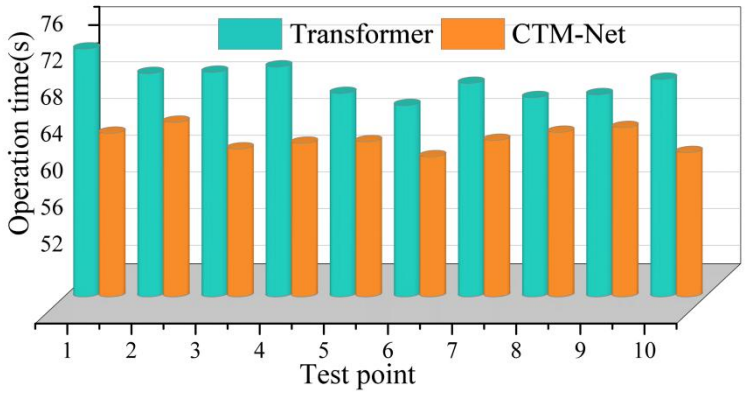


Figure 10: Comparison of model recognition efficiency

5 Conclusion

This paper introduces CTM-Net as a wind-storage system operating state recognition system based on Transformer and CNN. The model uses a multilayer adaptive Transformer to derive wind-storage operating-state parameters and incorporates CNN to produce deep-level operation-state features. Experimental analysis proves that the recognition rate of CTM-Net in wind-storage operating states is higher than 97 percent and its total recognition performance is higher than that of the conventional Transformer network. This paper presents a novel way of enhancing the identification of operating-states of wind-storage systems, which assists in achieving consistent operations of the systems and also increasing the economic profitability of wind-storage operation.

Funding

2023 Key R&D Program of China Resources Power, Research and Demonstration Application Project of Active Support Technology for Grid-Forming Energy Storage (CRP-R&D-LX-2023-

025).

About the Author

Shike Wang received the B.S. and Ph.D. degrees in electrical engineering from Xi'an Jiao Tong University (XJTU), Xi'an, China, in 2014 and 2020, respectively. She then joined China Resources Power Technology Research Institute Co., Ltd. as a R&D engineer in energy storage and renewable energy generation. Her research interests include power grid and source coordination, and small-signal stability of power electronics systems. Dr. Wang received the First Prize Paper Award in IEEE Transactions on Power Electronics in 2021, Science and Technology Progress Special Grade Award of China Power Supply Society in 2023. She is selected to Young Elite Scientists Sponsorship Program by China Association for Science and Technology.

Bin Chen, 1968.11, male, affiliated to Shanghai University of Electric Power, is the first director of CR Power Thermal Engineering and Automation Professional Committee. He has participated in promoting the demonstration and implementation of Liyujiang thermal-storage frequency regulation project and Xinjiang grid-forming energy storage project, focuses on the formulation of thermal control technical standards and professional management of thermal power plants, and presides over technical reviews of various large-scale thermal power and new energy infrastructure projects.

Yabing Sun (January 1990), Zhengzhou, Henan, graduated with a master's degree in Electrical Engineering from the School of Electrical Automation and Information Engineering at Tianjin University, currently works at China Resources Electric Power Technology Research Institute.In recent years, he is mainly engaged active support technology for new power systems, including grid construction technology, distributed phase-shifting camera, and coordinated control of new energy stations.

Guang Shi (1971), male, professor level senior engineer, graduated from Tsinghua University, mainly engaged in research on power system protection and control, presided over and participated in a number of projects at all levels.

References

[1] Wang, Y., Gu, A., & Zhang, A. (2011). Recent development of energy supply and demand in China, and energy sector prospects through 2030. Energy policy, 39(11), 6745-6759.

[2] Hassan, A., Ilyas, S. Z., Jalil, A., & Ullah, Z. (2021). Monetization of the environmental damage caused by fossil fuels. Environmental Science and Pollution Research, 28(17), 21204-21211.

[3] Jurasz, J., Canales, F. A., Kies, A., Guezgouz, M., & Beluco, A. (2020). A review on the complementarity of renewable energy sources: Concept, metrics, application and future research directions. Solar energy, 195, 703-724.

[4] Hosseini, A. A., & Hosseini, S. H. (2012). Utilizing solar energy instead of fossil fuels as domestic energy (case study: Dehloran city, Ilam province, Iran). Energy exploration & exploitation, 30(3), 389-401.

[5] Wolniak, R., & Skotnicka-Zasadzień, B. (2023). Development of wind energy in eu countries as an alternative resource to fossil fuels in the years 2016–2022. Resources,

12(8), 96.

[6] Yudha, S. W., Tjahjono, B., & Longhurst, P. (2022). Sustainable transition from fossil fuel to geothermal energy: A multi-level perspective approach. Energies, 15(19), 7435.

[7] Xingang, Z., & Pingkuo, L. (2013). Substitution among energy sources: An empirical analysis on biomass energy for fossil fuel of China. Renewable and Sustainable Energy Reviews, 18, 194-202.

[8] Radzka, E., Rymuza, K., & Michalak, A. (2019). Wind power as a renewable energy source. Journal of Ecological Engineering, 20(3).

[9] Hasan, N. S., Hassan, M. Y., Majid, M. S., & Rahman, H. A. (2013). Review of storage schemes for wind energy systems. Renewable and Sustainable Energy Reviews, 21, 237-247.

[10] Zhao, H., Wu, Q., Hu, S., Xu, H., & Rasmussen, C. N. (2015). Review of energy storage system for wind power integration support. Applied energy, 137, 545-553.

[11] Nyamdash, B., Denny, E., & O'Malley, M. (2010). The viability of balancing wind generation with large scale energy storage. Energy Policy, 38(11), 7200-7208.

[12] Yousif, M., Ai, Q., Wattoo, W. A., Jiang, Z., Hao, R., & Gao, Y. (2019). Least cost combinations of solar power, wind power, and energy storage system for powering large-scale grid. Journal of power sources, 412, 710-716.

[13] McKenna, R., vd Leye, P. O., & Fichtner, W. (2016). Key challenges and prospects for large wind turbines. Renewable and Sustainable Energy Reviews, 53, 1212-1221.

[14] Abdullahi, M., Oyelade, O. N., Kana, A. F. D., Bagiwa, M. A., Abdullahi, F. B., Junaidu, S. B., ... & Chiroma, H. (2025). A systematic literature review of visual feature learning: deep learning techniques, applications, challenges and future directions. Multimedia Tools and Applications, 84(19), 20439-20496.

[15] Lee, K., & Ram, S. (2024). Explainable deep learning for false information identification: An argumentation theory approach. Information Systems Research, 35(2), 890-907.

[16] Xiao, C., Liu, Z., Zhang, T., & Zhang, X. (2021). Deep learning method for fault detection of wind turbine converter. Applied Sciences, 11(3), 1280.

[17] Li, X., & Zhang, W. (2022). Physics-informed deep learning model in wind turbine response prediction. Renewable Energy, 185, 932-944.

[18] Tchakoua, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T. A., & Ekemb, G. (2014). Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. Energies, 7(4), 2595-2630.

[19] Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., ... & Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. Renewable energy, 133, 620-635.

[20] Zhan, J., Wang, R., Yi, L., Wang, Y., & Xie, Z. (2019). Health assessment methods for wind turbines based on power prediction and mahalanobis distance. International Journal of Pattern Recognition and Artificial Intelligence, 33(02), 1951001.]

[21] Li, W., Zhao, W., & Du, Y. (2025). Large-scale wind turbine blade operational condition monitoring based on UAV and improved YOLOv5 deep learning model. Mechanical Systems and Signal Processing, 226, 112386.

[22] Zhu, A., Zhao, Q., Yang, T., Zhou, L., & Zeng, B. (2023). Condition monitoring of wind turbine based on deep learning networks and kernel principal component analysis. Computers and Electrical Engineering, 105, 108538.

[23] Xiang, L., Yang, X., Hu, A., Su, H., & Wang, P. (2022). Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. Applied Energy, 305, 117925.

[24] Hacıefendioğlu, K., Başağa, H. B., Yavuz, Z., & Karimi, M. T. (2022). Intelligent ice detection on wind turbine blades using semantic segmentation and class activation map approaches based on deep learning method. Renewable Energy, 182, 1-16.

[25] Kong, Z., Tang, B., Deng, L., Liu, W., & Han, Y. (2020). Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. Renewable Energy, 146, 760-768.

[26] Park, J., Kim, C., Dinh, M. C., & Park, M. (2022). Design of a condition monitoring system for wind turbines. Energies, 15(2), 464.

[27] Fu, J., Chu, J., Guo, P., & Chen, Z. (2019). Condition monitoring of wind turbine gearbox bearing based on deep learning model. Ieee Access, 7, 57078-57087.

[28] Mohammadi, H. G., Arshad, R., Rautmare, S., Manjunatha, S., Kuschel, M., Jentzsch, F. P., ... & Schollbach, D. (2020, September). DeepWind: an accurate wind turbine condition monitoring framework via deep learning on embedded platforms. In 2020 25th IEEE international conference on emerging technologies and factory automation (ETFA) (Vol. 1, pp. 1431-1434). IEEE.

[29] Lu, L., He, Y., Ruan, Y., & Yuan, W. (2020). Wind turbine planetary gearbox condition monitoring method based on wireless sensor and deep learning approach. IEEE transactions on instrumentation and measurement, 70, 1-16.

[30] Lyu, Q., Liu, J., He, Y., Wang, X., & Wu, S. (2023). Condition monitoring of wind turbines with implementation of interactive Spatio–Temporal deep learning networks. IEEE Transactions on Instrumentation and Measurement, 72, 1-11.

[31] Alwadie, A. S., Yasin, S., Irfan, M., Draz, U., & Ali, T. (2025). TransWind: a vision transformer framework for wind turbine fault diagnosis on supervisory control and data acquisition system. RE&PQJ, 23(1), 19-24.

[32] Dhiman, H. S., Patil, S., Wagle, S., Soni, N., Kotecha, K., Selvachandran, G., & Abraham, A. (2025). Multi-head attention transformer and Bayesian inference recommendation engine-based blade icing detection framework for wind turbines. Neural Computing and

Applications, 1-20.

[33] Yin, L., Liu, Y., & Wang, N. (2025). Tensor product-fault diagnosis-Transformer based wind turbine blade fault prediction method. Engineering Applications of Artificial Intelligence, 162, 112535.

[34] Wu, J., Chang, Z., Zhang, L., Chen, M., Li, S., & Qiu, F. (2025). A Dual-Branch Transformer Network with Multi-Scale Attention Mechanism for Microgrid Wind Turbine Power Forecasting. Electronics, 14(13), 2566.