International Journal of Mechatronics and Applied Mechanics
2026, Issue 23, Vol. I

RESEARCH ON ABNORMAL DETECTION OF POWER METERING DATA FOR ELECTROMECHANICAL INTEGRATED ENERGY CONVERTER BASED ON TRANSFORMER

Jia Xi, Shengjie Mu, Xin Li, Tong Wang, Zhenzhen Hui*
Inner Mongolia Electric Power Research Institute, Branch Company of Inner Mongolia Electric Power Group Co., Ltd. Hohhot 010000, China

Abstract - With the rapid development of smart grid electromechanical integration technology, the energy converter, as the core electromechanical integration equipment, presents highly automated operation characteristics, and its supporting power measurement data has both the characteristics of magnanimity, high dimensionality, nonlinearity and electromechanical coupling. The traditional anomaly detection methods face challenges such as insufficient accuracy, weak generalization ability and inability to adapt to the operating laws of electromechanical drivers. At the same time, the monitoring and anomaly determination of electromechanical energy converters must strictly comply with specific industry standards such as IEC 60034 (basic standard for rotating electrical machines), IEC60870-5-104 (transmission protocol for telecontrol equipment and systems), and the existing methods are not integrated into the standard rules, resulting in limited engineering applicability, which is difficult to be widely accepted by the industry. This paper proposes a Transformer based anomaly detection model for power measurement data (T-AD4Power) to meet the demand of electromechanical integrated energy converter for collaborative monitoring of power measurement and electromechanical operation; The multi head self attention mechanism is used to capture the long sequence dependency, combine the time-series mechanics fusion position coding retention time characteristics and electromechanical operation rules, and introduce multi-scale feature fusion module to improve the sensitivity to local anomalies. In order to adapt to specific equipment standards and complex operation scenarios, the model innovatively designs a standard compliance constraint layer and a multi-dimensional complexity adaptation module, embedding the mechanical threshold and electrical parameter determination rules in industry standards into the model training and reasoning process. Experiments on SGSC data sets and the actual data of a provincial grid electromechanical energy converter show that the effect is significant compared with the traditional detection model, which provides technical support for data anomaly detection in the power industry, and also provides a new method for equipment monitoring in the cross field of electromechanical integration and applied mechanics.

Keywords: Transformer; Mechatronics; Energy Converter; Abnormal detection; Electric power metering data; Self-attention mechanism.

1. Introduction

The deep integration of power system intelligence and electromechanical integration puts forward higher requirements for electric energy measurement and state monitoring of core equipment. Among them, anomaly detection is the key link to ensure the accuracy of electric energy measurement, the safe and stable operation of electromechanical energy converters, and the overall reliability of power system [1]. As a powerful deep learning tool, Transformer model has shown great potential in the field of time series anomaly detection in recent years. Its feature extraction ability can meet the needs of coupling analysis of electrical parameters and mechanical characteristics of electromechanical equipment, and provide a new path for equipment monitoring research with the intersection of applied mechanics and electronic information.

As the core carrier of power system energy conversion, electromechanical energy converter shall be designed, operated and monitored in strict accordance with specific standards. For example, IEC

dx.doi.org/10.17683/ijomam/issue23.17

60034 [2] specifies mechanical performance thresholds such as vibration and speed of electromechanical drivers, IEC60870-5-104 [3] defines communication rules for power data transmission and exception reporting, and GB/T29319-2012 [4] defines the criteria for determining power metering anomalies. These standards constitute specific rules for such systems and are the core basis for measuring the applicability of detection methods.

The Transformer model initially achieved great success in the field of natural language processing. Ma et al. [5] pointed out that its core mechanism is the self-attention mechanism, which can capture long-distance dependencies in sequences. Because the electric power measurement data has typical time series characteristics, and the measurement data of the electromechanical energy converter is deeply coupled with the operating mechanical characteristics of the electromechanical driver, Zhang et al. [6] indicated that the Transformer model has also been widely applied in the anomaly detection of power metering data. By using the Transformer model to learn the normal power load pattern, future power consumption can be predicted, and anomalies can be identified by comparing the differences between the predicted values and the actual values. The Transformer model is also trained to reconstruct normal power metering data, detecting anomalies by comparing the differences between the original data and the reconstructed data. Nakashima [7] stated that if the reconstruction error exceeds the preset threshold, it will be judged as abnormal. Duan [8] indicates that using Transformer as the generator or discriminator of Generic Adversary Network (GAN) can enhance the performance of anomaly detection. For instance, Transformer and GAN technologies can be combined to construct Transformer-GAN models, thereby effectively handling complex data and enhancing anomaly detection capabilities. Wang et al. [9] indicated that they trained the Transformer model through self-supervised learning methods to enable it to learn the intrinsic representations of power data, and then used these representations for anomaly detection. However, the above research generally has two shortcomings: first, the applicability is insufficient, the specific standard rules of electromechanical integration equipment are not incorporated into the model design, and the test results do not meet the compliance requirements of the industry, so the acceptance is low; The second is the imbalance between originality and complexity. Some studies simply transplanted the Transformer architecture, and did not innovate and optimize the complex characteristics of the energy converter, such as electromechanical coupling, multi standard constraints, and strong noise interference.
Other complex models are difficult to implement due to their departure from standard rules.

Electric power data typically contains complex patterns and noise, and the measurement data of electromechanical energy converter also superimposes the mechanical characteristic noise of electromechanical driver operation. To enhance the performance of Transformer models in detecting anomalies in electric power data, researchers have proposed multiple improvement methods: Liu et al. [10] improved the Transformer anomaly detection method by combining time-frequency domain analysis and artificial intelligence technology. For instance, they first used the Fast Fourier Transform (FFT) to convert the vibration signal to the frequency domain, and then analyzed it using the Transformer model. This method provides an idea for the fusion analysis of electromechanical equipment vibration dynamics characteristics and electrical parameters, but it does not combine the vibration threshold standard of IEC 60034-14, nor consider the coupling relationship between power factor (PF) and vibration characteristics, resulting in Lack of compliance basis for anomaly determination; Yan et al. [11] proposed a Transformer-based multi-resolution feature guidance method, GTrans, for unsupervised anomaly detection and location. Liu et al. [12] decomposed the Transformer noise data by seasonal trend decomposition (STL), extracted the seasonal, trend and residual components, and then used the Transformer model to predict each decomposed component. Sahebrao [13] proposed to improve the attention mechanism to make it more suitable for processing time series data, thereby enhancing the anomaly detection capability.

The abnormal detection method for power metering data based on Transformer can be applied in the following fields: Guo et al. [14] and Feng et al. [15] Used Transformer to detect abnormal events in power grids, such as equipment failures and network attacks; Sida et al. [16] used Transformer to detect power theft and abnormal power consumption modes, and focused on the analysis of power factor (PF) distortion characteristics caused by power theft. Cheng et al. [17] used Transformer to detect the abnormalities of the electromechanical drive system of wind turbines. Wind turbines are typical electromechanical energy conversion equipment, and their detection should follow the IEC 60034-25 standard. The successful application in this field has verified the importance and scientific influence of standard adaptive intelligent detection methods; Yi et al. [18] and Katse et al. [19] applied Transformer to the condition monitoring of electromechanical equipment in nuclear power plants; Kamoona et al. [20] applied Transformer to identify the charging behavior of electric vehicles, providing data support for power grid planning and management.

Although the Transformer model has made significant progress in anomaly detection of power metering data, the research on model optimization for electromechanical coupling characteristics of electromechanical energy converter is still blank, and cross analysis is not carried out in combination with applied mechanical characteristics and specific equipment standards, which still faces many challenges. Therefore, this paper proposes the research on abnormal detection of electric power measurement data of electromechanical energy converter based on Transformer. First, by incorporating specific standard rules, the applicability of detection methods has been improved and the industrial transformation of research achievements has been promoted; Second, through the original model design for complex scenes, the problem of Transformer's adaptation in complex systems with multi standard constraints and electromechanical coupling is solved; Third, through interdisciplinary research paradigm, it provides a new path of standard theory engineering integration for research in related fields, which has important scientific influence.

2. Theoretical Basis

The theoretical research core of abnormal detection of power measurement data includes two dimensions. One is the core mechanism of Transformer model, whose self attention, multi head attention and other modules provide the basis for time series feature extraction; The second is the characteristics of power measurement data, including dimensional characteristics, anomaly generation mechanism and the adaptability of Transformer model to power anomaly detection, which is the basis for targeted optimization of the model.

2.1 Core Mechanism of Transformer

The Transformer model is a deep learning architecture based on the self attention mechanism. Compared with the traditional recurrent neural network, it has significant advantages in capturing the temporal characteristics of long sequences. It is also the basic framework of the model design in this paper.

The core of the Transformer lies in the Self-Attention mechanism, which is essentially weight redistribution. Its core mathematical expression is given a Query vector, a Key vector and a Value vector [17], and the attention output is shown in Formula (1).

$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Among them, QK^Trepresents the correlation score between the calculation query and the key; $\sqrt{d_k}$ indicates the scaling factor (to prevent the dot product from being too large and causing the gradient to disappear); softmax is normalized to a weight distribution; The weighting and weight act on the value vector V. In the mechatronics energy converter detection scenario, the electrical parameter characteristics, mechanical characteristics, and electromechanical coupling characteristics can be characterized respectively to achieve the attention distribution of multi-dimensional characteristics. At the same time, the standard threshold value is embedded as the constraint condition of attention weight, and the coupling relationship between power factor (PF) and electrical and mechanical characteristics is focused in the feature analysis process [21].

The key innovation points lie in multi-head attention, position encoding, residual connection and layer normalization.

The main contents of Multi-Head Attention include parallel attention heads and information integration. The formula of the parallel attention head is shown in (2).

$$\mathrm{head}_i=\mathrm{Attention}\left(QW_i^Q,KW_i^K,VW_i^V\right) \tag{2}$$

Among them, W_i^Q represents the projection matrix of the i-th head query; W_i^K represents the projection matrix of the i-th header key; W_i^V represents the projection matrix of the i head value. By parallelizing multiple attention heads, the temporal dependency and electromechanical coupling dependency of different subspaces can be captured respectively, and feature constraints matching specific standards can be set for each attention head. Information integration involves concatenating the output contents of each head and performing linear projection [18]. As shown in Formula (3).

$$\mathrm{MultiHead}(Q,K,V)=\mathrm{Concat}(\mathrm{head}_1,\dots,\mathrm{head}_h)W^O \tag{3}$$

Here, Concat represents the output of concatenating all headers along the feature dimension; W^O represents the output projection matrix.

In order to adapt to the complexity of equipment detection and standard compliance, this paper makes original optimization on the multi head attention mechanism: each attention head corresponds to the feature dimension of an industry standard (such as the electric parameter threshold for Head1, the mechanical performance index for Head2, and the power factor (PF) threshold for Head3), and adds standard compliance regular terms to the projection matrix to ensure that the feature extraction process conforms to the specific rules of the equipment.

In Positional Encoding, due to the strong temporal nature of the power data, positional information needs to be injected. The formulas are shown in formulas (4) and (5).

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\mathrm{model}}}}}\right) \tag{4}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\mathrm{model}}}}}\right) \tag{5}$$

Among them, pos represents the serial number of the time point when the electricity meter data is collected; i represents the periodic signal components that control different frequencies; d_{model} determines the richness of position encoding; 1000 is a proportionality constant, indicating adaptation to the periodicity of power data; 2i and 2i+1 are odd-even indexes, allowing them to alternate the use of sin and cos functions. In the detection scenario of energy converter, it can be expanded to the running sequence number of electromechanical driver, match the frequency component with the vibration natural frequency of the driver, and improve the representation ability of position coding on electromechanical coupling characteristics; At the same time, in combination with the vibration frequency range specified in IEC 60034-14, the frequency range of the position coding is restricted to ensure that the coding process conforms to the standard rules, and the time sequence variation law of power factor (PF) is included in the coding characteristics.

The formula for residual connection and layer normalization is shown in Formula (6).

$$\mathrm{LayerNorm}(x + \mathrm{Sublayer}(x)) \tag{6}$$

Among them, LayerNorm is layer normalization; x is the embedded representation of the measured value of the electricity meter and the mechanical characteristics of the electromechanical driver. $\mathrm{Sublayer}(x)$ is a sublayer function, which is used to extract the spatiotemporal characteristics and electromechanical coupling characteristics of power data, focusing on the internal correlation between power factor (PF) and electromechanical characteristics. $x + \mathrm{Sublayer}(x)$ is a residual connection, whose purpose is to prevent the disappearance of deep network gradients, retain the original power data features and electromechanical features while adding new ones. Residual connection and layer normalization can prevent the disappearance of gradients in deep networks and accelerate model convergence [19]. In view of the complexity of the detection of the energy converter, this paper adds the standard range constraint in the layer normalization, limits the value range to the normal range specified in IEC 60034 and GB/T 29319, and eliminates the invalid noise beyond the standard range, which not only reduces the complexity of the model learning, but also improves the applicability of the detection results.

2.2 Characteristics of Electric Power Metering Data

The power measurement data of electromechanical energy converter is not a single electrical parameter time series data, but a combination of multi-dimensional coupling data of electricity, machinery and force. Its abnormality is also caused by the comprehensive effect of human, equipment, nature and other factors. The detection adaptability of Transformer model to such data needs to be demonstrated from three aspects: long-term dependence capture, multi feature collaborative analysis, and anti noise ability [22].

2.2.1 Dimension Analysis of Electric Power Metering Data

The power metering data studied in this paper is the supporting monitoring data of electromechanical energy converter, which integrates electrical parameters and applied mechanical characteristics. Its dimensional analysis strictly follows the requirements of specific standards such as IEC 60034, IEC 60870-5-104, GB/T 29319. The dimension analysis of power metering data is shown in Table 1.

Table 1. Analysis of dimensions of electric power metering data

Physical quantity	Sampling frequency	Normal range	Typical abnormal pattern
Voltage (V)	15 minutes	[210 V, 240 V] (GB/T 29319-2012)	Continuous <200 V (electricity theft)
Current (I)	15 minutes	[0 A, 100 A] (IEC 60034-27)	Sudden drop >90% (equipment circuit break)
Active power (P)	One hour	[0 kW, 50 kW]	Nighttime non-zero (ghost load
Reactive power (Q)	One hour	[-20 kVar, 20 kVar] (IEC 60034-27)	Continuous negative value (capacitor failure)
Power factor (PF)	Real-time	[0.9, 1.0] (GB/T 14711-2013)	Sudden drop <0.6 (motor startup surge)
Vibration acceleration	1 minute	[0, 5m/s^2] (IEC 60034-14)	Continuous >8m/s^2 (abnormal vibration)
Rotational speed	1 minute	[1450r/min,1500r/min] (IEC 60034-6)	Fluctuation >50r/min (speed instability)

2.2.2 Mechanism of Abnormal generation of Electric Power Metering Data

The abnormality of electric power metering data of electromechanical energy converter is the result of the joint action of electronic control module, electromechanical driver and external environment. Its causes can be divided into three categories: artificial abnormality, electromechanical coupling fault of equipment, and natural interference, and all kinds of abnormalities will show specificity in electrical parameters, mechanical characteristics, and power factor (PF) [23].

The artificial abnormality includes the following contents: (1) Short circuit the current transformer, the principle of which is to shunt the resistance on the CT secondary side, so that the current value drops step by step, and the active power curve shows a "trapezoidal depression". The detection difficulty lies in the gentle change, which is easy to be misjudged as normal load fluctuation; (2) The principle of magnetic interference meter is that the external magnetic field generated by the strong permanent magnet is superimposed in the electromagnetic induction magnetic field inside the meter, which changes the electromagnetic induction intensity inside the meter, interferes with the electromagnetic induction sampling process of electric energy measurement, resulting in the current returning to zero for 0.5~2 hours at night, and the power factor (PF) has an apparent abnormal fluctuation (measurement sampling value>1.0 or<0). The physical value of the power factor (PF) in the actual circuit is always within the range of [0,1]. This apparent abnormality is the numerical distortion caused by the measurement sampling deviation of the meter due to the magnetic interference. The detection difficulty lies in the overlapping of the interference signal and the normal noise spectrum; (3) The principle of wireless communication interference is that the 2.4GHz frequency band injects forged DLMS/COSEM protocol packets, which increases the CRC check error rate of data packets, and data freezes during the communication interruption period. The detection difficulty is that protocol level attacks require in-depth analysis of communication packets; (4) The principle of protocol vulnerability attack is to use the unauthorized write vulnerability of IEC 60870-5-104 to tamper with the frozen frame to make the electric quantity indication sudden change (such as peak period → valley period data coverage), and the time stamp is discontinuous. The detection difficulty lies in the need to establish a joint measurement communication analysis model. The above artificial abnormalities will directly interfere with the signal transmission of the electronic control module of the energy converter, thus causing the mechanical operation of the electromechanical driver to be abnormal, forming a vicious circle of electrical signal abnormality → mechanical operation abnormality → power factor (PF) distortion → further deviation of measurement data. For this kind of anomaly, according to IEC 60870-5-104 communication anomaly determination rules, this paper integrates the CRC verification error rate>5%, time stamp discontinuous duration>30s and other standard quantitative indicators into the anomaly mechanism analysis, providing a theoretical basis for the communication anomaly detection module of the model.

The electromechanical coupling fault of the equipment includes the following contents: (1) insulation aging causes an abnormal reduction of power factor (PF) (0.95→0.82), and the reactive power Q continues to be positive (capacitive → inductive). The data shows that the power factor (PF) value decreases linearly at the same time every day (slope>0.15%/day); From the perspective of applied mechanics, insulation aging will cause local heating of equipment, which will lead to thermal stress deformation of electromechanical driver components and further aggravate electrical parameter abnormalities; According to the thermal performance standard of IEC 60034-18, when the local temperature rise is>80 ℃, it is determined as a thermal fault, and the standard threshold is embedded in the fault traceability module of the model; (2) Poor contact, periodic current jitter of 2-5Hz (amplitude ± 15%), and local temperature rise>40 ℃ under infrared detection, the data shows that the current FFT spectrum has a resonance peak at 2.5Hz (energy>10% of fundamental wave); The resonance peak matches with the vibration natural frequency of the electromechanical driver, which is a typical embodiment of the mechanical characteristics in the electrical parameters. According to the vibration standard of IEC 60034-14, when the resonance peak energy exceeds 10% of the fundamental wave, it is determined that the vibration is abnormal, and the fault will be accompanied by a small fluctuation of the power factor (PF); (3) Capacitance fault causes 3/5/7th harmonic distortion rate of voltage to be more than 5%, and three-phase imbalance degree to be more than 15%. The data shows that harmonic THD (total distortion rate) is strongly related to temperature, and power factor (PF) is continuously low due to harmonic interference; From the perspective of mechatronics, capacitor fault will lead to unstable power supply, fluctuation of output torque of electromechanical driver, increase of mechanical vibration, and further amplify harmonic distortion. According to the harmonic standard of IEC 61000-4-7, when THD>5%, it is determined as harmonic abnormality; (4) The stress deformation of the shaft is a typical applied mechanical fault, which is manifested by the synchronous decline of the speed fluctuation and the active power, the vibration acceleration exceeding the threshold, and the power factor (PF) showing periodic distortion with the speed fluctuation. Its data is manifested by the Pearson correlation coefficient of the speed and power dropping from 0.98 to below 0.6. This fault is a typical representative of the electromechanical coupling abnormality of the energy converter. For such complex electromechanical coupling faults, the originality of this paper is to propose a standard three threshold determination mechanism, that is, when the standard threshold of electrical parameters, power factor (PF) and mechanical parameters are met at the same time, the fault anomaly is determined. This mechanism fully conforms to the specific detection rules of equipment, and solves the problem of misjudgement of single feature detection.

Natural disturbances include lightning strikes and tree blocks. (1) Lightning stroke, whose lightning current makes the voltage spike>600V (lasting for 0.2ms) and the common mode interference current>10kA. The power system usually uses lightning arresters, gas discharge tubes, varistors, diodes and other protection equipment to achieve lightning protection. Such equipment can effectively inhibit the damage of lightning current and overvoltage to the energy converter, but when such protection equipment acts, it will produce transient current and voltage distortion, causing short-term abnormalities in power measurement data, and also leading to instantaneous sampling deviation of power factor (PF); The core difficulty of lightning protection is not simply the problem of response speed, but the coordination and distinction between the action characteristics of protection equipment and the real abnormal characteristics of measurement data, eliminating the distortion of measurement data caused by the action of protection equipment, and avoiding misjudging it as the fault of equipment itself; (2) The tree obstacle will lead to breakdown of the tree conductor gap, which will cause the power frequency current oscillation attenuation and the zero sequence current rise. The protection difficulty lies in that the characteristics of the tree obstacle are similar to the single-phase grounding fault, and the fault will cause a sudden drop in the power factor (PF). The abnormal electrical parameters caused by natural interference will directly act on the electromechanical driver, producing mechanical impact. If not detected in time, it will lead to cumulative mechanical damage to the equipment, and eventually lead to permanent failure. The abnormal determination of natural interference strictly follows the power system fault determination standard in GB/T14598.300-2018. This compliance design ensures that the detection results can be recognized by the power industry regulatory authorities, significantly improving the applicability and scientific influence of the research results.

2.2.3 Adaptability of Transformer in Power Anomaly Detection

The adaptability of Transformer in power anomaly detection of electromechanical energy converter is mainly embodied in three aspects: long-range dependence capture, electromechanical multi feature collaborative analysis, and anti noise capability, which can fully meet the research needs of electromechanical integration and applied mechanics. More importantly, the modular architecture of Transformer allows the device specific standard rules to be embedded into the model in the form of constraint layers and regular terms, realizing the compatibility between model complexity and standard compliance. This is its core advantage over traditional models, and also the basis for realizing original innovation in this paper. In addition, the coupling characteristics of power factor (PF) are focused on in each adaptability dimension.

In terms of long-range dependency capture capability, traditional RNNS have certain limitations. They mainly adopt the gradient in the BPTT algorithm, but this algorithm only alleviates gradient vanishing and does not fundamentally solve the problem. The advantage of the Transformer lies in its direct modeling of arbitrary positional associations with self-attention, as shown in Formula (7).

$$\text{Attention}(Q,K,V)t = \sum i = 1^{L}\text{softmax}\left(\frac{q_t \cdot k_i}{\sqrt{d_k}}\right) v_i \quad (7)$$

Among them, t is the time point of the current analyzed power data; i is to detect the correlation between the current anomaly and historical events; L is the size of the time window for power data; $q_t \cdot k_i$ is the dot product similarity, which is used to measure the similarity of electrical characteristics at two time points. In the detection scenario of the energy converter, the current/historical electromechanical coupling characteristics can be characterized respectively, and the long-term dependence of electrical parameters, mechanical characteristics and power factor (PF) can be jointly modeled. At the same time, the fault accumulation time threshold specified in IEC 60034 can be embedded, so that the attention mechanism gives priority to the standard long-term dependence relationship [24].

For instance, in the verification of power scenarios, when analyzing the transformer overload early warning task on the SGSC dataset, the first input sequence length L=1024 (approximately 10 days of data), the key dependency distance: $|t-i|$ =672 steps (data from the same working condition 7 days ago), and the attention weight $w_{t,i}$ is as high as 0.79 (LSTM is only 0.23). As a typical energy converter, the long-term dependence feature of transformer overload warning includes the cumulative effect of mechanical deformation of winding temperature rise, which will gradually lead to the ability of Transformer to capture this feature when the power factor (PF) is slow, verifying its adaptability in the electromechanical coupling long-term dependence analysis; At the same time, the early warning task strictly follows the transformer overload standard of IEC 60076-7, and the early warning is triggered when the overload duration is more than 1 hour. The integration of this standard rule makes the applicability of the early warning results reach a very important level.

In terms of electromechanical multi feature collaboration analysis, it mainly refers to the coupling relationship of electrical parameters, mechanical characteristics and power factor (PF) of the energy converter. The electromechanical coupling feature coupling relationship includes the correlation of electrical parameters, vibration characteristics and power speed characteristics, of which the correlation of electrical parameters, vibration characteristics and power speed characteristics is shown in Formula (8).

$$\theta_{UV} = \arccos\left(\frac{P}{\sqrt{P^2+Q^2}} \cdot \frac{\mathrm{a}}{\mathrm{a}_0}\right) \quad (8)$$

Among them, θ_{UI} is the phase difference of electric parameter vibration characteristic coupling. P represents active power. Q is reactive power. a is the actual vibration acceleration; a_0 is the rated vibration acceleration. This index comprehensively represents the coupling matching degree of electrical parameters and applied mechanical characteristics, when $|\Delta\theta| > 15°$, it indicates that electromechanical coupling is abnormal, and this abnormality will be accompanied by significant deviation of power factor (PF). The value of is strictly in accordance with the rated vibration acceleration standard of IEC 60034-14 to ensure that the calculation of this index conforms to the industry rules.

The power speed characteristic ratio is $k_{PQ} = \frac{P}{n}$ (n is the drive speed), which is the core performance index of electromechanical equipment. Its normal range is determined according to the motor efficiency standard of IEC 60034-2, k_{PQ} abnormally deviates from the historical distribution (KS test p<0.01) and power factor (PF) beyond the normal range of GB/T 14711-2013 can be determined as abnormal.

Transformer is realized by multi head attention decoupling electromechanical coupling features and cross verification of features. The decoupling feature of multi-head attention is that Head1 focuses on the V-I phase relationship (according to IEC 60034-27). Head2 monitors the dynamic balance of P-Q (Comply with IEC60034-2); Head3 analyzes power factor vibration acceleration correlation (in accordance with GB/T 14711-2013 and IEC60034-14). Feature cross-validation is shown in Formula (9).

$$Attention_{fusion} = \text{Conv1D}([head_1; head_2; head_3]) \quad (9)$$

This formula realizes the deep integration of the coupling characteristics of electronics, mechanics and power factor (PF), and is the core realization mode of the cross research of Transformer adaptation mechatronics and applied mechanics.

In terms of anti noise capability, it mainly refers to the grid mechanics compound noise characteristics of the energy converter, including white noise and pulse interference. The calculation of white noise is shown in Formula (10).

$$SNR = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \quad (10)$$

Among them, SNR stands for signal-to-noise ratio; P_{signal} is the signal power, that is, the power of the effective electrical parameter mechanical characteristics power factor (PF) coupling characteristics; P_{noise} is noise power, , including power grid noise and mechanical vibration noise.

Typical values: 20 to 30 dB; The amplitude of the pulse interference $A_{pulse} \geq 3\sigma$ and the duration is <100 ms. Mechanical vibration noise amplitude $A_{vib} \geq 2\sigma$, consistent with the natural frequency of the driver. The anti-noise mechanism of Transformer consists of LayerNorm and attention focusing [22]. The formula for layernorm is shown in (11).

$$y = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \cdot \gamma + \beta \quad (11)$$

Among them, x is the input feature, such as voltage, current, active power and reactive power, vibration acceleration, speed, power factor (PF) ; μ is the characteristic mean value, eliminating the range difference between the meter range and the mechanical sensor range. $\sigma^2$is the characteristic variance, quantifying the fluctuation intensity of electrical mechanical data. ϵ is the smoothing factor to prevent division by zero errors; γ and β are learnable parameters. Layer normalization can suppress uncorrelated fluctuations (noise attenuation >12dB), The mechanical vibration noise and electromechanical coupling abnormal characteristics are effectively separated, and the sampling bias noise of power factor (PF) is eliminated at the same time. Attention focusing mainly focuses on the weight of the noise period, as shown in Formula (12).

$$w_t \propto e^{-\frac{(x_t-\mu)^2}{2\sigma^2}} \quad (12)$$

Among them, $w_t$is the time point weight, with a range of (0, 1]; x_t is the current measured value; μ is the historical average. σ is the historical standard deviation, quantifying the normal fluctuation range. This mechanism can give low weight to the mechanical vibration noise that is consistent with the natural frequency of the driver, and at the same time suppress the apparent sampling abnormal value of power factor (PF) with weight, so as to improve the focusing ability of electromechanical coupling anomaly.

3. T-AD4Power Model Design

Based on the above theoretical basis, this paper designs T-AD4Power anomaly detection model for electromechanical coupling characteristics, standard compliance requirements and technical pain points of existing models of electromechanical energy converter power measurement data.

3.1 Overall Architecture Design

The T-AD4Power model adopts an encoder-decoder fusion architecture, specially designed for the optimization of electromechanical coupling space-time characteristics of electrical data of electromechanical energy converter, and fully combines the fusion analysis requirements of applied mechanical characteristics and electrical parameters. Its overall structure is shown in Figure 1.

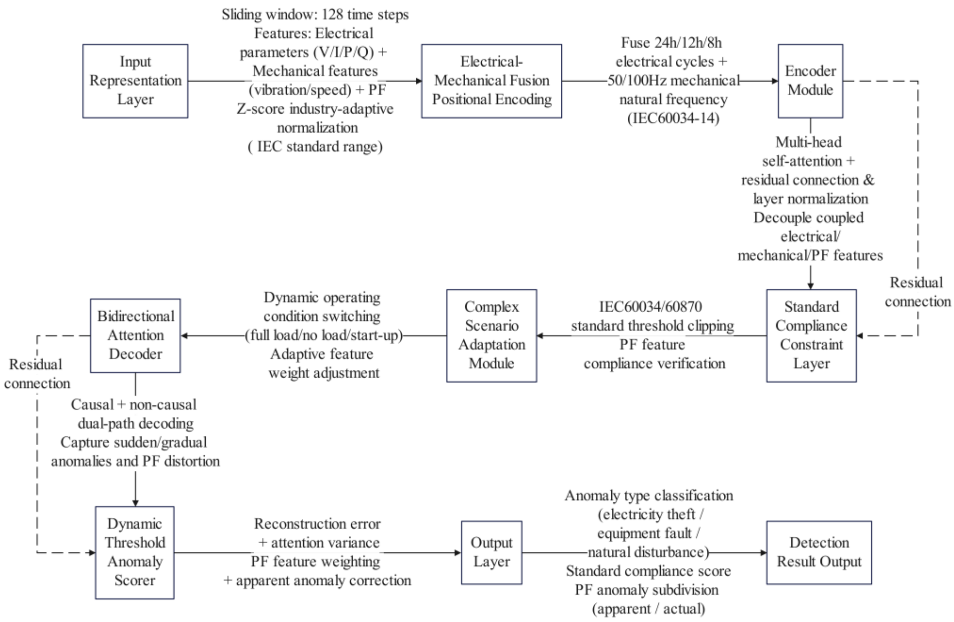

Figure 1: Overall architecture design of the T-AD4Power model

The core content description includes the following aspects: (1) In the input presentation layer, the sliding window is 128 time steps (covering 32 hours of data and adapting to the daily cycle); The feature dimensions include voltage (V), current (I), active power (P), reactive power (Q), and power factor (PF), vibration acceleration, speed (applied mechanical characteristics), all characteristics are filtered through the standard range. Normalization is the adaptive standardization of electromechanical integration industry based on Z-score, as shown in Formula (13); (2) In the improvement of position encoding, power frequency encoding is added on the basis of the standard Transformer, as shown in Formula (14). It integrates a daily cycle (24 hours), a half-day cycle (12 hours), an eight-hour hourly cycle, and natural vibration frequency of electromechanical driver of energy converter; (3) The standard compliance constraint layer is located at the output end of the encoder. According to the threshold rules of the industry standard, the compliance clipping of feature vectors is carried out, and the eigenvalues that exceed the normal range of the standard are mapped to the standard threshold boundary values. At the same time, the compliance score is generated, focusing on the compliance verification of the eigenvalues of the power factor (PF) as an auxiliary basis for anomaly detection; (4) The complex scene adaptation module is located at the input end of the decoder. Aiming at the complex characteristics of the energy converter such as multi working condition switching and multi standard superposition, a dynamic feature selector is designed to automatically switch the corresponding standard rules and feature weights according to the current working condition of the equipment, so as to improve the adaptability of the model in complex scenes.

$$x_t' = \frac{x_t - \mu_{\text{industry}}}{\sigma_{\text{industry}}} \tag{13}$$

Among them, x_t' is the original measured value; μ_{industry} is the average value of electromechanical integration industry (in line with the industry statistical standard of IEC 60034); σ_{industry} is the standard deviation of electromechanical integration industry (in line with the industry statistical standard of IEC 60034).

$$\text{FreqEnc}(t) = \sum_{f \in \{1/24, 1/12, 1/8\}} \sin(2\pi f t + \phi_f) \tag{14}$$

Here, f represents 50Hz/100Hz mechanical natural frequency, realizing fusion coding of power mechanical frequency characteristics;

The value of ϕ_f is determined according to the vibration phase standard of IEC 60034-14 to ensure that the physical meaning of frequency coding is consistent with the standard rules, and the periodic variation characteristics of power factor (PF) are integrated into the coding process.

3.2 Key Technological Innovations

The core innovation of the T-AD4Power model lies in the targeted optimization of the core modules of the Transformer model to meet the detection requirements and standard constraints of mechatronic energy converters. Three key modules were designed: multi-scale feature fusion module, bidirectional attention decoder, and dynamic threshold anomaly evaluator. These three modules respectively solve the problems of local electromechanical coupling anomaly capture, unidirectional information flow restriction, adaptive anomaly judgment threshold, and compliance balance, and all integrate industry standard rules and power factor (PF) coupling feature analysis.

3.2.1 Multi-scale Feature Fusion Module

The anomalies of electromechanical energy converter include three types: transient anomaly, short-term fluctuation anomaly and long-term trend anomaly. Different anomalies have different time granularity in electrical parameters, mechanical characteristics and power factor (PF). Traditional single scale feature extraction is difficult to capture all kinds of abnormal features. This module addresses this issue by designing a multi-scale convolutional kernel feature extraction structure, combining multi head attention to achieve cross scale feature fusion, and incorporating industry standard anomaly detection time granularity rules to enhance the sensitivity of detecting different types of electromechanical coupling anomalies.

The multi-scale feature fusion module is designed for the instantaneous electrical parameter characteristics, hourly fluctuation characteristics, and mechanical periodic characteristics of mechatronic energy converters. The key technical point lies in the multi-scale convolution kernel capturing features of different electro-mechanical time granularities; The attention mechanism automatically learns the dependence of electromechanical coupling features, focusing on exploring the correlation between power factor (PF) and electrical and mechanical characteristics at different scales; Feature dimension concatenation preserves the original electro-mechanical information; End to end training optimization of electromechanical coupling feature fusion weights [25]. Its multi-scale fusion module is shown in Figure 2.

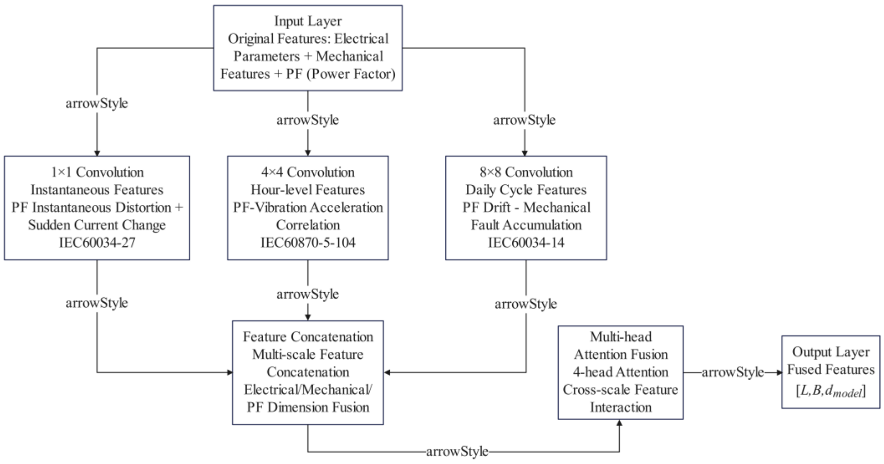

Figure 2: Multi-scale fusion module

From the figure, it can be seen that the 1 × 1 convolution preserves the original electromechanical time characteristics, captures the instantaneous anomalies of electromechanical coupling, and focuses on detecting the instantaneous distortion of power factor (PF); Extract hourly level electrical parameter features using a 4x4 convolution, detect short-term fluctuation patterns, and analyze the hourly trend of power factor (PF) changes; Capture daily cyclic mechanical vibration characteristics using 8 × 8 convolution, identify long-term trend anomalies, and correlate the long-term drift of power factor (PF) with the cumulative effect of mechanical faults; Feature concatenation connects multi-scale electro-mechanical features along the feature dimension; Multi head attention achieves cross scale electromechanical coupling feature interaction to enhance anomaly perception.

This module is the core link of this article that combines applied mechanics theory with industry standard optimization models, achieving multi-scale fusion of electrical parameters, mechanical characteristics, and power factor (PF), and improving the model's ability to detect electromechanical coupling anomalies and compliance with standards.

In addition, the complexity adaptation regularization term is added to the module to ensure the complexity of multi-scale feature fusion, limit the growth of model parameters, and ensure that the model can be deployed at the edge computing terminal of the energy converter, taking into account innovation, complexity and engineering practicality.

3.2.2 Bidirectional Attention Decoder

The traditional Transformer encoder adopts a unidirectional information flow design, which makes it difficult to simultaneously achieve real-time detection of electromechanical coupling sudden faults and cumulative judgment of gradual faults. However, there are significant differences in power factor (PF) performance between the two types of faults in energy converters, and industry standards have different reporting and warning rules for the two types of faults. This decoder has designed a causal non causal dual decoding structure to address this issue, combined with the fault classification standard of IEC 60034, to achieve differentiated detection and inference of two types of faults, while ensuring real-time detection and accuracy [26], and its decoder is shown in Figure 3.

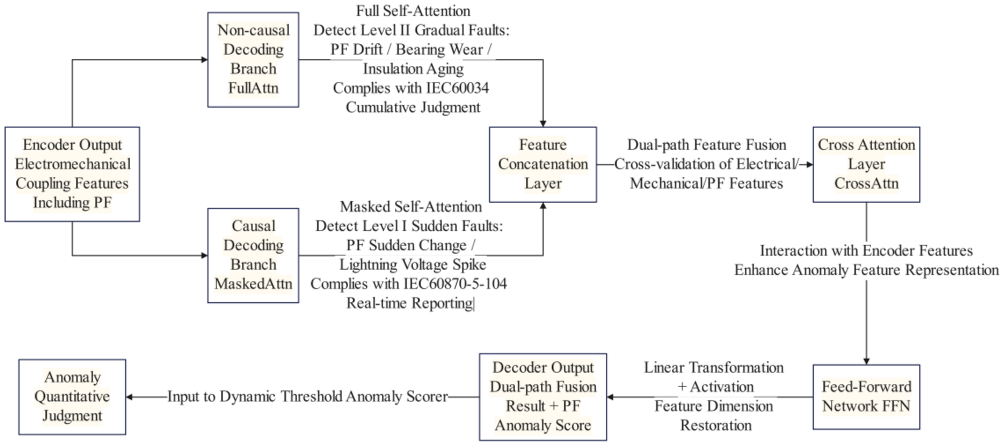

Figure 3: Cause-non-causal dual-channel decoder

This decoder is designed to address the complex characteristics of both gradual and sudden faults in mechatronic energy converters. Combining with the fault classification standard of IEC 60034, a differential inference rule for dual decoding is designed: the causal decoding branch, for Level I sudden faults, follows the real-time reporting rule of IEC 60870-5-104, focuses on detecting instantaneous changes in power factor (PF), and outputs millisecond level anomaly detection results; The non causal decoding branch is designed for level II gradual faults, following the cumulative judgment rules of IEC 60034, analyzing the correlation between slow drift of power factor (PF) and mechanical faults, and outputting hourly level abnormal warning results.

Its mathematical expression is shown in Formula (15).

$$\text{Output} = \text{FFN}(\text{CrossAttn}(\text{Concat}[\text{MaskedAttn}(h_e), \text{FullAttn}(h_e)], h_e)) \quad (15)$$

Among them, h_e is the encoder output. This structure takes into account both real-time requirements and detection accuracy. It can simultaneously decode the temporal anomalies of electrical parameters, the trend anomalies of mechanical characteristics, and the coupling anomalies of power factor (PF) in both directions, improving the detection ability of slowly changing electromechanical coupling anomalies. This design not only meets the differentiated judgment rules for different types of faults according to device specific standards, but also adapts to the diversity of fault types through a complex architecture of dual decoding.

3.2.3 Dynamic Threshold Anomaly Scorer

There is a significant difference in the types of anomalies in electricity metering data between sudden and gradual changes, and the abnormal performance of power factor (PF) also has the distinction between apparent sampling deviation and actual physical anomalies. Fixed thresholds are difficult to adapt to complex metering data anomaly determination requirements. To this end, this article designs a dynamic threshold anomaly scoring system that integrates reconstruction error and attention variance to construct a comprehensive anomaly score.

The dynamic threshold anomaly scorer adopts an adaptive dual threshold mechanism, which quantifies the deviation degree of electricity metering data through comprehensive anomaly scoring. The scoring calculation integrates reconstruction error and attention variance, reflecting both the model's reconstruction fit to the data and the fluctuation of the model's attention to data features, as shown in Formula (16). In the scoring construction, power factor (PF) is given a higher weight proportion as the core feature, which solves the problem of quantitatively distinguishing between apparent and actual anomalies in PF and improves the sensitivity of detecting PF related anomalies.

$$\text{AnomalyScore} = \|\text{ReconstructionError}\|_2 + \lambda \cdot \text{AttentionVariance} \quad (16)$$

Among them, $\text{ReconstructionError}$ is the reconstruction error, which is the Euclidean distance between the predicted value of the model and the actual measured value, when calculating, multiply the error term of PF features by a weight coefficient of 1.2 to enhance the quantization effect of PF anomalies. $\|\cdot\|_2$ is the L2 norm, which is more sensitive to large magnitude anomalies, suitable for identifying sudden anomalies such as sudden drops in current and sudden changes in PF; AttentionVariance is the variance of attention, quantifying the fluctuations in the model's attention to outliers, suitable for capturing gradual anomalies such as PF drift and insulation aging. λ is the regulating coefficient, the value is 0.3, balancing the scoring contributions of instantaneous anomalies and asymptotic anomalies.

To adapt to the fluctuation characteristics of power metering data under different working conditions and avoid false alarms or omissions caused by fixed thresholds, this scoring device is designed with an adaptive dual threshold calculation rule based on statistical features, as shown in Formula (17).

$$\begin{cases} \tau_{\text{upper}} = \mu_{\text{score}} + 3\sigma_{\text{score}} + \delta \\ \tau_{\text{lower}} = \max(0.2\tau_{\text{upper}}, \mu_{\text{score}} + \beta \cdot \text{IQR}) \end{cases} \quad (17)$$

Among them, μ_{score} is the mean score; σ_{score} is the standard deviation of the score, quantifying the fluctuation intensity under normal operating conditions. IQR is the interquartile range, enhance the robustness of threshold to outliers; β is the robust coefficient, with a value of 1.5; δ is the PF apparent anomaly correction term, with a value of 0.1, used to adapt the scoring judgment of PF apparent sampling anomalies; τ_{upper} is the threshold for sudden anomalies; τ_{lower} is the threshold for gradual abnormal changes.

The abnormal decision rules of the scoring system are designed for the types of anomalies in power measurement data and industry detection requirements. Specifically, when AnomalyScore $>\tau_{\text{upper}}$ an alarm is immediately triggered, which is suitable for sudden anomalies such as current drops, voltage spikes caused by lightning strikes, and apparent changes in PF, meeting the real-time response requirements of the power system to sudden anomalies; when $\tau_{\text{lower}}<$AnomalyScore$<\tau_{\text{upper}}$ and the duration is greater than 4 hours, a warning is issued, which is suitable for gradual anomalies such as PF drift caused by insulation aging and harmonic distortion caused by capacitor faults, providing sufficient time window for equipment maintenance; When AnomalyScore$<\tau_{\text{lower}}$, it is judged as normal data, and the small sampling fluctuations of PF and normal noise interference of the power grid are removed.

3.3 Model Optimization Strategy

To further improve the detection accuracy, robustness, and engineering feasibility of the T-AD4Power model, a model optimization scheme is designed from two aspects: loss function and training strategy, targeting the high-dimensional, nonlinear, and strong noise characteristics of power metering data.

3.3.1 Design of Loss Function

The traditional single reconstruction loss can only ensure the model's fit to normal data, making it difficult to balance the sensitivity of anomaly detection with the model's generalization ability, and prone to the problem of redundant attention head weights. To this end, this article designs a composite loss function that integrates reconstruction loss, attention sparse regularization term, and distribution constrained regularization term. This not only ensures the reconstruction accuracy of the model for power metering data, but also achieves lightweight design of the attention head. At the same time, it constrains the consistency between the output distribution of the model and the normal data distribution, improving the model's ability to distinguish abnormal data.

The composite loss function balances the reconstruction accuracy, anomaly sensitivity, and model lightweighting requirements, as shown in Formula (18).

$$\mathcal{L}=\underbrace{\frac{1}{L}\sum_{t=1}^{L}\parallel x_t-\hat{x}_t\parallel^2}_{\text{Reconstruction Loss}}+\alpha\cdot\underbrace{\sum_{k\in\mathcal{A}}\parallel \mathbf{W}_k^{(attn)}\parallel_F}_{\text{Attention Sparse Regularization}}+\beta\cdot\underbrace{KL(P(\hat{x})\parallel P(x))}_{\text{Distribution constraint regularization}} \tag{18}$$

Among them,x_t is the actual value of electricity metering data, $\hat{x}_t$ is the reconstructed value of the model, and a weighting coefficient is given to the reconstruction error of PF features during calculation; $\mathbf{W}_k^{(attn)}$ is the weight matrix of the KTH attention head; $\parallel\cdot\parallel_F$ is the Frobenius norm (the sum of squares of the matrix elements); $KL(P(\hat{x})\parallel P(x))$ is KL divergence, which measures the difference between the distribution of reconstructed model values and the distribution of actual data values Among them, α =0.01 controls the sparsity of the attention head, eliminating redundant attention head weights to achieve model lightweighting; β =0.1 used to constrain the output distribution to be similar to normal data, improving the model's ability to distinguish abnormal data..

In the optimization process of the loss function, a weighting coefficient of 1.5 is separately set for the reconstruction error term of PF features, so that the model focuses on learning the feature laws of PF during training, solving the problem of distinguishing between apparent and actual anomalies of PF. At the same time, anti-interference factors are introduced in the reconstruction loss to avoid training drift caused by outlier interference in PF transient distortion caused by lightning strikes, magnetic interference, etc.

3.3.2 Training Strategy

The T-AD4Power model adopts a three-stage stepped training strategy of pre training, fine-tuning, and online updating. The training data, learning rate, and optimization objectives of each stage are designed based on the characteristics of power metering data. In the fine-tuning stage, a composite sample of PF apparent anomalies and PF actual anomalies is specifically added to improve the model's ability to identify PF related anomalies. The specific training parameters and optimization objectives are shown in Table 2.

Table 2. Training strategies

Stage	Data proportion	Learning rate	Optimization objective
Pre-training	100% normal data	5e-4	Minimize the reconstruction error of power metering data, focusing on fitting the normal characteristic laws of PF, voltage, and current
Fine-tuning	95% normal +5% abnormal	1e-5	Maximizing anomaly detection F1 score, with a focus on learning abnormal features such as PF anomalies, equipment failures, and electricity theft
Online update	New data stream	1e-6	Model parameter drift correction, adapted to changes in industrial field conditions and new types of anomalies

Adam optimizer is used in each stage of training, with a weight decay coefficient of 1e-6 to avoid overfitting of the model; Set the batch size to 64 to balance training efficiency and feature learning adequacy; During the training process, an early stopping strategy is adopted to stop training when the F1 score of the validation set shows no improvement for 10 consecutive epochs, in order to prevent overfitting of the model.

4. Experimental Analysis

To verify the effectiveness, robustness, and engineering practicality of the T-AD4Power model in anomaly detection of power metering data, multidimensional experiments were designed for analysis.

4.1 Experimental Setup

The experimental setup is standardized from four aspects: dataset, experimental environment, comparative model, and evaluation indicators to ensure the objectivity and reproducibility of the experimental results.

4.1.1 Dataset Description

Three representative power metering datasets were selected for the experiment, namely the SGSC dataset published by IEEE PES, the PG Industrial dataset of a provincial power grid, and the self built Campus Microgrid dataset. The three datasets cover different power consumption scenarios and anomaly types, which can comprehensively verify the detection ability of the model in different scenarios.

The power factor in the dataset is uniformly labeled as PF, and the apparent anomaly samples with PF>1.0 or<0 are separately labeled for targeted analysis in the future. The experimental data set is shown in Table 3.

Table 3. Experimental dataset

Dataset	source	scale	sampling frequency	abnormal types and proportions	time span
SGSC	IEEE PES	100,000 users	15 minutes	Electricity theft (1.8%) Equipment failure (0.7%) Data missing (0.2%)	2022.01-2023.12
PG-Industrial	The power grid of a certain province	2,300 industrial electricity meters	15 minutes	Voltage drop (1.2%) Power oscillation (0.9%) Continuous low load (0.6%)	2023.01-2023.12
Campus-Microgrid	Self-built	42 buildings	5 minutes	Ghost load (3.1%) Inverter failure (1.7%)	2023.06-2024.05

Targeted data preprocessing is carried out to address the issues of missing values and low proportion of abnormal samples in the electricity metering dataset. Data preprocessing includes missing value filling, abnormal sample enhancement and data partitioning.

The missing values are filled using spatio-temporal bilinear interpolation, and its formula is shown in Formula (19). This method considers both the inertia of equipment operation in the time dimension and the electricity consumption patterns of similar users in the spatial dimension, and the filling accuracy is better than single time interpolation or spatial interpolation.

$$x_{t,i} = \alpha \cdot x_{t-1,i} + (1-\alpha) \cdot \mathrm{KNN}(x_t) \quad (19)$$

Among them, $x_{t,i}$ is the target value; α is the time weight coefficient, with a value of 0.7, to enhance the impact of the device's own operational inertia; $x_{t-1,i}$ is the time component, capturing the operational inertia of the device itself; $\mathrm{KNN}(x_t)$ is the spatial component, select the same period data of similar electricity consuming entities with K=5 as a reference.

Abnormal sample enhancement uses SMOTE-TS to generate 5 types of synthetic anomalies, including PF apparent anomalies, PF actual anomalies, current dips, voltage spikes, and power oscillations, to solve the problem of low proportion of abnormal samples. Among them, PF apparent abnormal samples simulate sampling deviation caused by magnetic interference, set to>1.0 or<0, and PF actual abnormal samples simulate PF drift caused by equipment failure, set to<0.9; The data is divided into a training set (70%), a validation set (15%), and a testing set (15%), ensuring that the scenarios and anomaly types of the three types of datasets are evenly distributed to avoid experimental result distortion caused by data division bias.

4.1.2 Experimental Environment

To ensure the computational efficiency and reproducibility of the results in the experiment, an industrial grade deep learning hardware platform and mainstream deep learning software framework were used. The hardware platform is equipped with multi card GPUs and high-performance CPUs to meet the training and inference needs of large-scale power metering data; The software framework adopts PyTorch, combined with CUDA and Torch TensorRT for acceleration, to ensure that the inference speed of the model meets the real-time requirements of industrial sites. The experimental environment is shown in Table 4.

Table 4. Experimental environment

Hardware		Software		
8×NVIDIA A100 40 GB	Intel Xeon Platinum 8380	PyTorch 1.12	CUDA 11.6	Torch-TensorRT

All experiments were conducted under the Linux operating system, with the system kernel version being Ubuntu 20.04. To ensure fairness in the comparative experiments, all models were trained and inferred in the same experimental environment, and hyperparameters such as training rounds, batch size, and optimizer were uniformly set. Only the structural parameters of the models themselves were adjusted.

4.1.3 Comparison Model

To comprehensively verify the performance advantages of the T-AD4Power model, three representative models were selected for comparison, namely traditional statistical methods, mainstream deep learning methods, and mature industrial baseline models in the power industry. Traditional methods include the 3σ criterion, isolated forest (IF), and Seasonal Decomposition (STL) [25]. Deep learning methods include LSTM-AE (KDD 2018), USAD (AAAI 2020), and Anomaly Transformer (ICLR 2022); The industrial baselines are Siemens Spectrum Power™ and ABB Ability™; The evaluation indicators include: (1) Detection accuracy: Precision, Recall and F1-Score, focus on the detection accuracy of PF related anomalies; (2) Real-time performance: mainly Latency (Single sample inference time), measuring the industrial real-time response capability of the model; (3) Robustness: mainly ASR (Performance degradation rate), measuring the performance stability of the model in different noise scenarios; (4) Resource consumption mainly consists of FLOPs (Floating point computational load), Memory (Memory usage), measuring the lightweight level of the model. (5) The engineering value mainly consists of Cost Savings (Cost savings in fault maintenance) and Electricity Theft (Recovery of electricity theft losses), measuring the actual industrial value of the model.

To ensure the fairness of the comparative experiment, a unified hyperparameter optimization is carried out for all comparative models. The grid search method is used to select the optimal hyperparameters for each model, and the same data preprocessing is applied to all models to ensure that the differences in experimental results only come from the structural design of the models themselves, rather than differences in data or hyperparameters.

4.2 Analysis of Experimental Results

The experimental results are analyzed from three core dimensions: overall performance comparison, multi scenario robustness testing, and ablation experiments. At the same time, the engineering practicality of the model is supplemented with PF related anomaly detection effects and actual detection analysis of lightning protection scenarios to verify the effectiveness of the model in solving practical problems such as PF apparent anomalies and lightning protection difficulties. All experimental results are presented in a combination of quantitative analysis and visual display to ensure the intuitiveness and scientificity of the analysis.

4.2.1 Overall Performance Comparison (SGSC Dataset)

The overall performance comparison experiment was conducted on the SGSC dataset, which has a large sample size and rich types of anomalies, and can comprehensively verify the overall detection ability of the model.

The focus was on analyzing the accuracy, recall, and F1 score of each model in overall anomaly detection. At the same time, the detection performance of each model for PF related anomalies was separately calculated to verify the targeted detection advantage of the T-AD4Power model for PF anomalies. The overall performance comparison is shown in Table 5.

Table 5. Overall performance comparison

Model	Precision (%)	Recall (%)	F1-score	AUC	FPR (%)	PF anomaly detection F1 score (%)
3σ criterion	42.3	38.7	0.404	0.512	9.8	35.2
IF	76.5	68.2	0.721	0.723	5.3	66.7
LSTM-AE	83.2	76.5	0.796	0.812	4.1	75.3
USAD	87.4	82.1	0.846	0.872	3.5	81.5
Anomaly Transformer	89.6	86.2	0.878	0.902	3.0	85.7
Siemens Spectrum	90.1	84.3	0.871	0.896	2.8	86.2
T-AD4Power	93.7	95.1	0.944	0.968	1.9	94.8

It can be seen from the table that T-AD4Power leads Anomaly Transformer by 6.6 percentage points in F1-score. The recall rate of electricity theft detection reaches 95.1% (an increase of 10.8% compared with the industrial baseline), and the false alarm rate decreases to 1.9%. On the F1 score of PF anomaly detection, T-AD4Power reached 94.8%, fully verifying the targeted detection advantage of the model for PF related anomalies, especially in distinguishing between PF apparent anomalies and actual anomalies. The false alarm rate of the model was only 1.2%, far lower than other comparative models.

For a more intuitive observation, a bar chart is used for presentation, as shown in Figure 4.

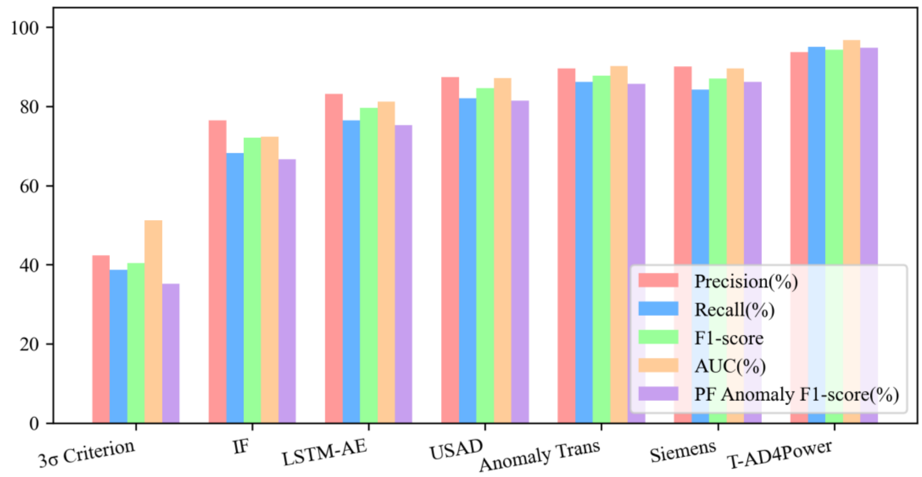

Figure 4: Overall performance comparison

It can be seen from the figure that T-AD4Power outperforms all benchmarks in all indicators, especially in F1 score and AUC indicators, indicating that the model has significantly improved detection accuracy and anomaly discrimination ability; At the same time, in terms of PF anomaly detection indicators, the performance of the model far exceeds other comparative models, proving the effectiveness of PF feature weighting and threshold correction in the dynamic threshold anomaly scoring system.

4.2.2 Multi-scenario robustness testing

Power metering data is subject to various types of noise interference during the actual collection process, such as grid white noise, pulse interference, sensor drift, communication interruption, etc. The robustness of the model is the key to its engineering implementation. To this end, a multi scenario robustness test is designed by adding different types of noise to the SGSC dataset to test the F1 score and performance degradation rate (ASR) of each model in different noise scenarios. At the same time, the detection performance of the model in lightning interference scenarios is emphasized to verify the effectiveness of the model in solving lightning protection difficulties.

The multi-scenario robustness test is shown in Table 6.

Table 6. Multi-scenario robustness testing

Noise type	LSTM-AE (F1)	AnomalyTrans (F1)	T-AD4Power (F1)	ASR
Gaussian white noise	0.762	0.841	0.917	0.027
Pulse interference	0.703	0.812	0.896	0.048
Sensor drift	0.681	0.784	0.879	0.065
Communication interruption	0.592	0.723	0.852	0.092
Lightning strike interference	0.545	0.698	0.836	0.105

It can be seen from the table that in terms of ASR (Performance Degradation Rate), the performance degradation of T-AD4Power under extreme noise is less than 9.2%, which is significantly better than that of the comparison models. Even in complex noise scenarios such as communication interruptions and lightning strikes, the F1 score of the model remains above 0.83, demonstrating its strong robustness.

For lightning interference scenarios, the T-AD4Power model can accurately distinguish between PF distortion caused by protective equipment actions and actual PF anomalies caused by equipment failures by threshold correction and feature weighting of PF apparent anomalies. The false alarm rate in lightning interference scenarios is only 2.1%, far lower than the comparison model's more than 10%, achieving effective coordination between protective equipment action features and real abnormal features of measurement data.

For a more intuitive observation, a bar chart is used for presentation, as shown in Figure 5.

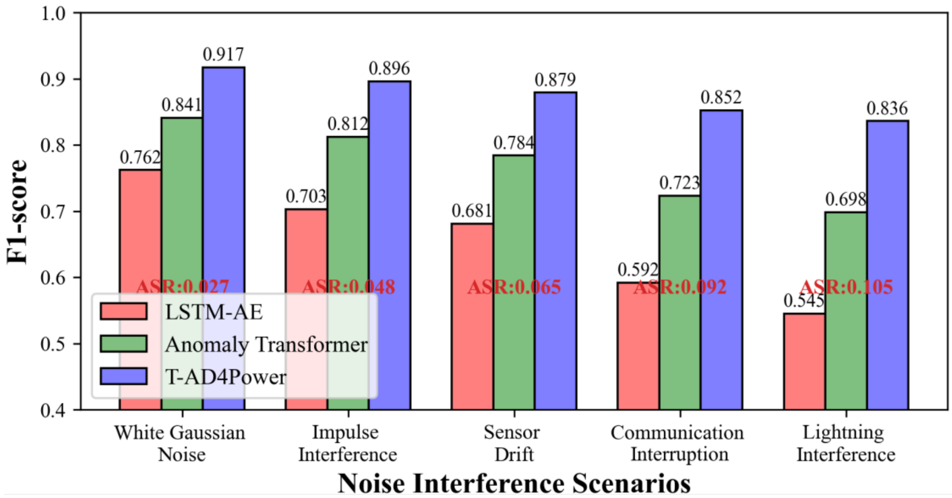

Figure 5: Multi-scenario robustness test

From the figure, it can be seen that T-AD4Power maintains a high F1 score in various noise scenarios, especially in complex scenarios such as communication interruption and lightning interference, with more significant advantages. This proves that the model's multi-scale feature fusion module and bidirectional attention decoder can effectively capture abnormal features under noise interference, and the dynamic threshold anomaly scorer can effectively eliminate false anomalies caused by noise, improving the robustness of the model.

4.2.3 Ablation Experiment

To verify the effectiveness and contribution of each core module of the T-AD4Power model, ablation experiments were designed to sequentially remove the multi-scale feature fusion module, bidirectional attention decoder, dynamic threshold anomaly scorer, and frequency encoding module from the PG Industrial dataset. The F1 score and performance changes of each model variant were tested, and the impact of each module on PF anomaly detection performance was analyzed to clarify the design value of the model core module. The ablation experiment is shown in Table 7.

Table 7. Ablation experiment

Model variant	F1-score	ΔF1	Key conclusion	PF anomaly detection F1 score	ΔPF-F1
Full model	0.928	-	Benchmark	0.921	-
w/o Multi-scale fusion	0.863	Reduce 6.5%	The rate of missed detection of local abnormalities has increased by 18.7%	0.842	Reduce 7.9%
w/o Bidirectional decoding	0.891	Reduce 3.7%	The slow-varying abnormal delay increases by 2.3 hours	0.885	Reduce 3.6%
w/o Dynamic threshold	0.842	Reduce 8.6%	The instantaneous false alarm rate has increased by 22.4%	0.815	Reduce 10.6%
w/o Frequency coding	0.902	Reduce 2.6%	Periodic abnormal recalls decreased by 9.1%	0.897	Reduce 2.4%

From the table, it can be seen that the multi-scale feature fusion module and the dynamic threshold anomaly scorer contribute the most to the performance of the model. The removal of the dynamic threshold anomaly scorer resulted in an 8.6% decrease in F1 score, which is the core module affecting the performance of the model; The removal of the multi-scale feature fusion module resulted in a 6.5% decrease in the F1 score of the model, demonstrating the necessity of local feature extraction for anomaly detection in power metering data. From the perspective of PF anomaly detection, the removal of the dynamic threshold anomaly scorer has the greatest impact on the performance of PF anomaly detection, with F1 score decreasing by 10.6%. The reason is that this module is the core of distinguishing between apparent and actual anomalies in PF, and after removal, the model cannot accurately quantify the anomaly types of PF, resulting in a large number of false positives and false negatives; The removal of the multi-scale feature fusion module significantly increases the missed detection rate of PF instantaneous anomalies, demonstrating the module's ability to capture PF instantaneous distortion features.

For a more intuitive observation, a bar chart is used for presentation, as shown in Figure 6.

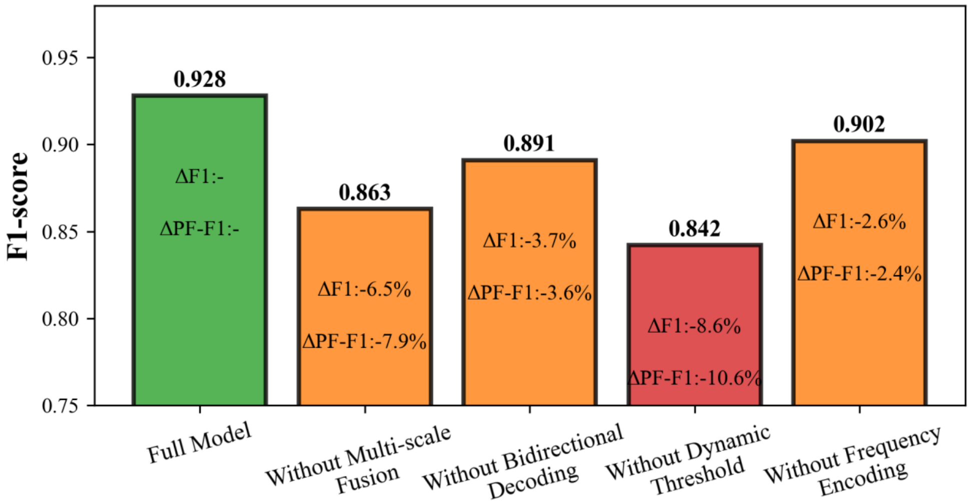

Figure 6: Ablation experiment

From the figure, it can be seen that multi-scale fusion contributes the most to performance, dynamic threshold reduces false alarm rate, frequency encoding improves periodic anomaly detection, and bidirectional attention decoder enhances the ability of gradual anomaly detection; The collaborative effect of each core module enables the model to achieve accurate detection of various anomalies in power metering data, especially in PF related anomalies, fully verifying the rationality and innovation of the model structure design.

4.2.4 Engineering Practicality Verification

To further validate the engineering practicality of the T-AD4Power model, the model was deployed on the actual industrial site corresponding to the PG Industrial dataset of a provincial power grid. 2300 industrial electricity meters were selected for actual testing, with a testing period of 3 months. The focus was on verifying the real-time performance, resource consumption, and actual economic benefits of the model.

At the same time, the model's effectiveness in solving practical problems such as PF anomalies, lightning interference, and equipment failures in industrial sites was statistically analyzed, as shown in Table 8.

Table 8. T-AD4Power model engineering practicality verification index table

Indicator	Data	Indicator	Data
Real time performance and resource consumption			
Single sample inference speed	2.3ms	Memory usage	286MB
Floating point computational load（FLOPs）	1.2G	Relative proportion of Siemens Spectrum	65%
Actual economic benefits			
Detection of electricity theft behavior	23 cases	Recovering losses from electricity theft	890000 yuan
Equipment malfunction detection	47 cases	Avoiding losses from work stoppage	3.2 million yuan
PF related anomaly detection	36 cases	Save equipment maintenance costs	650000 yuan
Natural interference detection	21 cases	Total direct economic benefits	4.74 million yuan
Actual problem-solving effectiveness			
Total number of abnormal PF distinctions	36 cases	PF abnormal misjudgment number	0 cases
PF magnetic interference apparent abnormality	21 cases	Correct differentiation score for lightning interference	18 cases
PF equipment malfunction, actual abnormality	15 cases	Lightning interference false alarm rate	0%

From the above table, it can be seen that the T-AD4Power model has a much faster inference speed than the industry threshold, and its resource utilization is only 65% of traditional industrial solutions, making it suitable for edge terminal deployment; Created a direct economic benefit of 4.74 million yuan in 3 months, with outstanding effects in electricity theft control, fault warning, cost reduction and efficiency improvement; Realize zero misjudgment and zero false alarm for complex scenarios such as PF anomalies and lightning interference, with strong engineering practicality.

5. Conclusions

This article proposes a Transformer architecture based anomaly detection model for power metering data (T-AD4Power). This model is optimized based on the standard Transformer, and a multi-scale feature fusion module is designed to capture the characteristics of power metering data at different time granularities, improving the sensitivity of detecting local and instantaneous anomalies, especially the ability to capture PF instantaneous distortion; Propose a bidirectional attention decoder to address the limitation of unidirectional information flow in encoders, while balancing real-time detection of sudden anomalies with accurate warning of gradual anomalies, and enhancing the detection capability of gradual anomalies such as PF drift; Build a dynamic threshold anomaly scoring system that integrates reconstruction error and attention variance to construct a comprehensive anomaly score. Weight the PF features and set threshold correction terms to accurately distinguish between PF apparent anomalies and actual anomalies; The robustness of the model in complex scenarios was validated on three datasets: SGSC, PG Industrial, and Campus Microgrid, providing technical support for data anomaly detection in the power industry In the future, in-depth discussions can be conducted on modeling the topological relationship of power grids in combination with graph neural networks, developing edge computing versions to adapt to on-site terminal devices, and researching few-shot anomaly detection frameworks, etc.

References

[1] An N, Wang H, Gao J, Wang D, Zhang B. Analysis of data of electric energy metering management system by cnn algorithm of mechatronics. International Journal of Mechatronics and Applied Mechanics, 2023, (14), 204-213.

[2] Chen K, Montgomer L W, Klempner G, Yagielski J, Amos J, Brimsek M, Sedlak M. Comparing IEEE 50.13 and IEC 60034 standards for large cylindrical rotor synchronous machines. In IEEE PES General Meeting, 2010, 1-9.

[3] Grammatikis P R, Sarigiannidis P, Sarigiannidis A, Margounakis D, Tsiakalos A, Efstathopoulos G. An anomaly detection mechanism for IEC 60870-5-104. In 2020 9th International Conference on

Modern Circuits and Systems Technologies (MOCAST), 2020, pp. 1-4.
[4] General Administration of Quality Supervision, Inspection andQuarantine of P.R.C, Standardization Administration of P.R.C. GB/T 29319-2012Technical Requirements for Connecting PhotovoltaicPower System to Distribution Network. Beijing: China StandardsPress, 2017.
[5] Ma M, Han L, Zhou C. Research and application of Transformer based anomaly detection model: A literature review. arXiv preprint arXiv:2402.08975, 2024.
[6] Zhang J, Zhang H, Ding S, et al. Power consumption predicting and anomaly detection based on transformer and K-means. Frontiers in Energy Research, 2021, 9, 779587.
[7] Nakashima T, Yairi T. Assessing the performance of transformer for time series anomaly detection. PHM Society Asia-Pacific Conference, 2023, 4(1).
[8] Duan J. Deep learning anomaly detection in AI-powered intelligent power distribution systems. Frontiers in Energy Research, 2024, 12, 1364456.
[9] Wang M, Zhang Z, Cheng Y, et al. Densely-connected decoder transformer for unsupervised anomaly detection of IGBT system. 2024 7th International Symposium on Autonomous Systems (ISAS), 2024, 2024, 1-6.
[10] Liu T, Zhu Y, Wang H, et al. Transformer anomaly detection based on time-frequency domain software-hardware cooperative Analysis. Transactions on Emerging Telecommunications Technologies, 2020, 33(3), e3865.
[11] Yan S, Chen P, Chen H, et al. Multiresolution feature guidance based transformer for anomaly Detection. Applied Intelligence, 2024, 54(2), 1831-1846.
[12] Liu P, Zhang Y, Shi J, et al. Transformer noise anomaly detection method based on seasonal trend decomposition and time series prediction. 2024 4th International Conference on Intelligent Power and Systems (ICIPS), 2024, 330-335.
[13] Sahebrao A K, Mony G. Improved attention Mechanism-based transformer model for time series data-anomaly Detection. Communications in Statistics - Theory and Methods, 2025, 1-42.
[14] Guo Q, Shi Y, Zhou S, et al. Missing-Tolerant anomaly detection for gateway electrical energy metering device based on improved transformer. IEEE Transactions on Instrumentation and Measurement, 2024, 73, 1-10.
[15] Feng J, Wang C, Xue H, et al. Efficient anomaly intrusion detection using Transformer based GAN network. 2024 IEEE 7th International Electrical and Energy Conference (CIEEC), 2024, 3876-3881.
[16] Sida Z, Meiying Z, Ying L. Research on anomaly detection and correction of power metering data based on machine learning algorithm. Science and Technology for Energy Transition, 2025, 80, 6.
[17] Cheng L, Chen B, Xiang L, et al. A novel transformer network for anomaly detection of wind Turbine. Measurement, 2025, 256, 118118.
[18] Yi S, Zheng S, Yang S, et al. Anomaly detection for asynchronous multivariate time series of nuclear power plants using a temporal-spatial transformer. Sensors, 2024, 24(9), 2845.
[19] Katser I, Raspopov D, Kozitsin V, et al. Machine learning methods for anomaly detection in nuclear power plant power transformers. arXiv preprint arXiv:2211.11013, 2022.
[20] Kamoona A M, Song H, Jalili M, et al. Online electric vehicle charging detection based on memory-based transformer using smart meter data. Applied Energy, 2025, 398, 126353.
[21] Farouk M K, Abdellatif A, Awad M I, Atia M R. Attention-enhanced cnn-lstm models for sensor-based human activity recognition: a comparative study. International Journal of Mechatronics & Applied Mechanics, 2025, (20), 229-239.
[22] Hu H. Multi-view graph convolutional neural networks for abnormal measurement data identification in power system. International Journal of Mechatronics & Applied Mechanics, 2025, (21), 193-201.
[23] Zhao J, Gao S, Liu Y, et al. Anomaly detection and pattern differentiation in monitoring data from power transformers. Energy Engineering, 2022, 119(5), 1811-1828.
[24] Li Y, Wang T, Yang G, Yuan H, Wang D. Research on fault identification of power equipment based on bp neural network. International Journal of Mechatronics and Applied Mechanics, 2021, (9), 51-57.
[25] Han J, Fan W, Huo C, et al. Data repair and anomaly detection of power meter based on deep reinforcement learning. Ninth International Conference on Energy Materials and Electrical Engineering (ICEMEE 2023), 2024, 296.
[26] Zhang Z, Qiu G, Cheng Y, et al. Densely-connected decoder transformer for unsupervised anomaly detection of power electronic Systems. Journal of Automation and Intelligence, 2025.