ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on multidimensional data mining and analysis technology in condition monitoring of regional centralized control hydropower equipment

Renshan Xu[1,*], *Shulin Deng*[1], *Yao Yu*[1], *Shikai Lu*[1], *Ziwan Huang*[1] *and Wei Zhou*[1]

[1] Hubei Energy Group Loushui Hydropower Co., Ltd., Enshi, Hubei,445000

SUMMARY: *Regional centralized control of hydropower operation requires continuous condition monitoring and anomaly analysis of multi-power station units and their auxiliary machines. Focusing on the hydroelectric generator set, speed regulation system, main transformer and auxiliary equipment, this paper constructs a data mining and analysis framework, and maps vibration, swing, temperature, head, flow, active power, pressure pulsation and event quantity into a four-dimensional working condition portrait of "physics-environment-health-business". In terms of method, statistical features, short-time Fourier transform features and wavelet packet process features were fused, and support vector machine, random forest and long short-term memory network were combined to complete steady-state identification, transient tracking and joint discrimination, and index deviation analysis, abnormal pattern extraction, early warning classification and feedback write-back were realized. Experimental results show that the Accuracy and F1 of the test set reach 95.6%and 95.1%respectively, and the average Accuracy and F1 of the field pilot are 95.3%and 94.7%respectively, which can support fine-grained state discrimination, anomaly interpretation analysis and operation and maintenance decision support under regional centralized control conditions.*

KEYWORDS: *Regional concentrated hydropower; Condition monitoring; Multidimensional data mining; Anomaly analysis*

1 Introduction

In the regional centralized control mode, the unit operation, auxiliary machine coordination and monitoring information aggregation of hydropower stations are unified into the centralized monitoring and analysis link. The state fluctuations of the auxiliary systems such as hydraulic generator set, speed regulation system, main transformer and cooling and lubrication are directly related to the unit output stability, peak regulation and frequency modulation response and regional production control accuracy. The unit state is not only determined by a single measurement point. Vibration, swing, temperature, head, flow, active power, pressure pulsation, environment and business process information participate in the state evolution expression. In the regional centralized control scenario, the data scale is larger, the update frequency is higher, and the working condition switching is more intensive. The condition monitoring has shifted from single equipment threshold comparison to multi-dimensional heterogeneous data-driven continuous identification and analysis.

The state interpretation of hydropower equipment is based on clear boundary of operating

*13886783310@163.com
https://doi.org/10.65102/is2026547 Published: 30 April, 2026

conditions and accurate data representation. There are significant differences in the signal distribution of the unit in the process of starting up, idling, excitation, no-load, simultaneous grid connection, load increase after grid connection, stable operation, load reduction, normal shutdown, fast shutdown and load rejection. The same equipment will also show different vibration and temperature rise characteristics under different head, load and environmental conditions. It is difficult to maintain a fine distinction between steady state, transient state, test state and fault state by interpreting monitoring values only with fixed thresholds. The technical data you provided summarized the unit working condition identification as a four-dimensional portrait of "physics-environment-health-business", and gave the integration ideas of steady-state distribution identification, transient process identification, index calculation, index analysis and early warning feedback, which provided a clear technical basis for multi-dimensional data mining in the condition monitoring of regional centralized control hydropower equipment. At the same time, regional centralized control is not a simple superposition of single station monitoring, but also requires the establishment of a consistent data interface between unified coding, second sampling, rule arrangement, horizontal comparison and historical backtracking, so as to make the comparison of similar equipment with the same working condition and cross-unit trend analysis computable, and support closed-loop processing and tracking analysis of early warning events.

Kumar and Saini studied the monitoring system of hydropower stations supported by the Internet of Things and cloud computing, and proposed a data-driven remote sensing and platform monitoring path [1]. Huang et al. studied the condition monitoring and fault diagnosis of hydropower units based on LSTM correction model, and proposed a timing correction mechanism to enhance the stability of fault recognition [2]. Zhang et al. studied the quality enhancement method of hydropower unit monitoring data, and proposed a hierarchical clustering and generative completion scheme for missing and noise data [3]. Al-Hardanee and Demirel studied the state prediction of hydropower stations and proposed a fault detection method combining RNN and LSTM [4]. Dao et al. studied the fault diagnosis of hydropower units and proposed a deep learning model with chaotic quadratic interpolation optimization [5]. Li et al. studied the application of digital twin technology in reversible hydraulic turbine monitoring system, and proposed a digital mirror framework for state mapping and monitoring feedback [6]. Hajimohammadali et al. studied the performance difference between deep network and hybrid network in intelligent monitoring of hydropower stations, and proposed the monitoring and evaluation idea under multi-model comparison [7]. Dao et al. also studied hydraulic turbine wear fault diagnosis and proposed IWSO optimized CNN-LSTM network for complex wear state identification [8]. Existing research has covered the direction of perception, prediction, diagnosis and digital mapping, but a unified method chain still needs to be formed around the working condition constraints, multi-dimensional feature coupling, stable and transient joint identification and anomaly analysis closed loop of regional centralized control scenarios.

Based on this, this paper focuses on the multi-dimensional data mining and analysis technology in the condition monitoring of regional centralized control hydropower equipment, and constructs the overall architecture for multi-source monitoring data, the feature construction and state representation method, the multi-dimensional monitoring data mining module, and the equipment condition discrimination and anomaly analysis method. The main body will focus on three tasks: (1) Combining the regional centralized control business logic and unit working condition rules, the state expression framework covering steady state, transient state, test state and abnormal state is established; (2) Fusing statistical features, time-frequency features and time series features to form multi-dimensional data mining links adapted to SVM, random forest and LSTM models; (3) Around the key indicators such as vibration, temperature, swing, pressure pulsation and power response, a computerized analysis method that can be used for state identification, anomaly analysis and operation and maintenance decision support is formed. It provides direct support for subsequent experimental verification and result analysis.

2 Related work

Condition monitoring of regional centralized control hydropower equipment involves multiple calculation steps, such as unit steady-state identification, transient process tracking, equipment health assessment, abnormal early warning feedback and so on. The research perspective has expanded from single measurement point threshold interpretation to multi-source heterogeneous data modeling, unified coding of state representation, and online analysis link construction. In this direction, the model is required to be able to describe the coupling relationship between vibration, swing, temperature, head, flow, active power and pressure pulsation, and the algorithm is also required to adapt to the sector-level sampling, condition switching, rule configuration and cross-unit lateral comparison in regional centralized control scenarios. The condition monitoring under the condition of regional centralized control no longer stays at the single equipment alarm level, but gradually shifts to the continuous recognition and analysis driven by the four-dimensional working condition portrait of "physics-environment-health-business".

Esmaeili Nezhad and Samimi studied the application of machine learning in transformer condition monitoring, and proposed a systematic application framework from feature extraction, pattern recognition to health assessment [9]. Yang et al. studied the predictive maintenance method of dry-type transformer based on vibration signal, and proposed an intelligent maintenance path for equipment life management [10]. Mei et al. studied fault diagnosis of relay protection system in smart substation and proposed a recognition model combining Transformer structure and transfer training [11]. Kang et al. studied multi-sensor information fusion technology in fault warning of smart grid equipment, and proposed a processing scheme for the linkage of fusion perception and warning [12]. Shao et al. studied the fault tracking of relay protection system and circuit breaker, and proposed an improved random forest driven fault chain tracking method [13]. Li et al. studied the mechanical fault diagnosis of high-voltage circuit breakers and proposed a state recognition method based on multi-modal data fusion [14]. Xu et al. studied multi-sensor hybrid multi-modal feature fusion and proposed a fusion structure for adaptive bearing fault diagnosis [15].

Table 1: Comparison of related studies

Reference	Research Object	Core Method	Main Contribution	Applicable Characteristics
[9]	Transformer condition monitoring	Machine learning review framework	Summarizes the full chain of feature extraction, modeling, and evaluation	Suitable for methodological systematization
[10]	Dry-type transformer maintenance	Vibration-driven predictive maintenance	Establishes a life-oriented identification path	Suitable for maintenance decision-making scenarios
[11]	Relay protection system	Transformer + transfer learning	Enhances complex fault identification capability	Suitable for transfer modeling tasks
[12]	Smart grid equipment warning	Multi-sensor information fusion	Strengthens warning linkage and fusion perception	Suitable for heterogeneous sensing scenarios
[13]	Relay protection–circuit breaker fault tracking	Improved random forest	Improves fault-chain localization capability	Suitable for correlation tracking analysis
[14]	Mechanical faults of high-voltage circuit breakers	Multimodal data fusion	Enhances the accuracy of mechanical state identification	Suitable for multi-source state recognition
[15]	Bearing fault diagnosis	Hybrid multimodal feature fusion	Optimizes fused feature representation capability	Suitable for high-dimensional feature construction

Zhang et al. studied fault diagnosis of hydraulic components and proposed an improved TSO-CNN-BiLSTM multi-sensor fusion model [16]. Bampoula et al. studied condition monitoring and predictive maintenance of manufacturing assets, and proposed a collaborative modeling method of LSTM autoencoder and Transformer encoder [17]. Surucu et al. studied the theory and application progress of machine learning condition monitoring, and proposed a review framework covering modeling processes, application scenarios and algorithm evolution [18]. Tama et al. studied deep learning fault diagnosis of rotating machinery based on vibration signals and proposed a technical summary around signal representation, network design and deployment adaptation [19]. Matania et al. studied the deep learning literature in vibration fault diagnosis of key rotating machinery, and proposed a systematic combing around data dependence, generalization ability and interpretability [20].

From the existing results, it can be seen that the related research of condition monitoring has formed a relatively complete technology spectrum: one kind of work emphasizes the discrimination ability of machine learning and deep learning in monitoring and identification, another kind of work emphasizes the expression ability of multi-sensor and multi-modal fusion in complex working conditions, and another kind of work emphasizes the collaborative value of predictive maintenance, fault tracking and interpretable analysis. These results provide a reference calculation method for the monitoring of regional centralized control hydropower equipment. However, the research directly oriented to the cooperative operation conditions of hydraulic turbine generator set, speed regulation system, main transformer and auxiliary equipment still needs to be further consolidated into the integrated calculation chain of joint steady-state and transient identification, rule configuration, index analysis, anomaly tracking and early warning feedback. Based on this, on the basis of related research, this paper introduces the idea of four-dimensional condition portrait and multi-dimensional feature joint modeling, and constructs a data mining and state discrimination method for the centralized control scene of the adaptation area, so that the subsequent experimental verification and anomaly analysis and evaluation can be established on a unified calculation expression.

3 Method and system design

3.1 Overall framework of condition monitoring of regional centralized control hydropower equipment

In order to realize the continuous monitoring, working condition recognition, anomaly analysis and early warning feedback of the hydroelectric generator set and its key auxiliary system under the condition of regional centralized control, this paper constructed the overall framework for multi-dimensional monitoring data mining. The architecture organizes data links around four types of information: unit operation status, hydraulic boundary, equipment health and business process, and can uniformly access, encode and analyze vibration, swing, temperature, head, flow, active power, pressure ripple and switching events. Compared with the single-site monitoring mode, the regional centralized control scenario emphasizes more on the data consistency between multiple units, multiple plants and multiple time scales. Therefore, the overall architecture not only assumes the signal acquisition task, but also supports the horizontal comparison across units, the vertical tracking of the same working condition and the closed-loop disposal of abnormal events. In order to ensure that multi-source monitoring data can maintain a continuous flow among collection, processing, analysis and feedback under the condition of regional centralized control, this paper constructs the overall architecture of hierarchical condition monitoring, as shown in Figure 1.

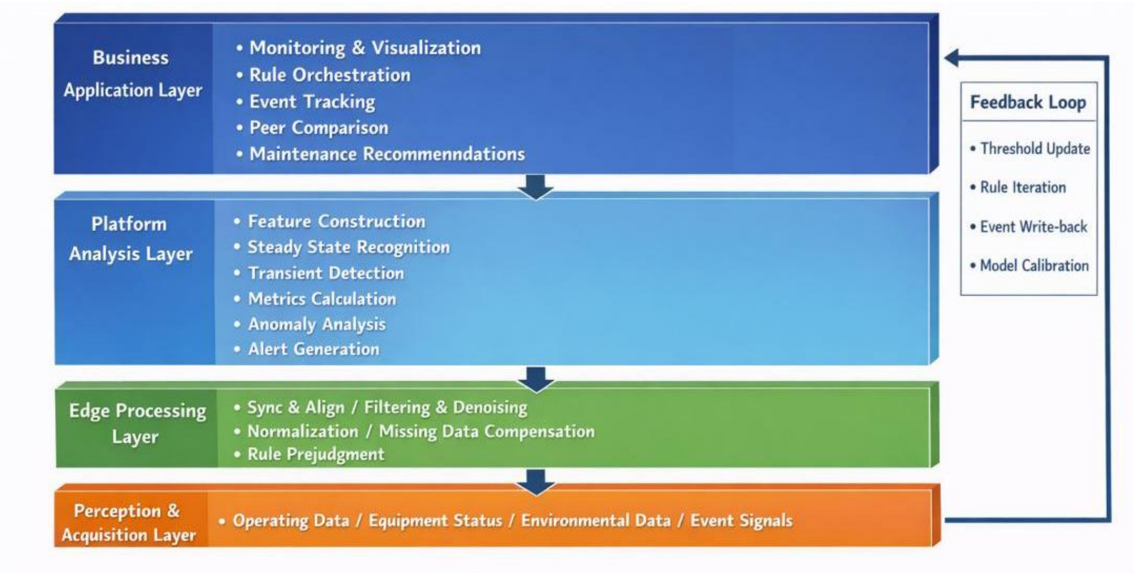

Figure 1: Overall architecture of condition monitoring of regional centralized control hydropower equipment

The sensing access level is responsible for collecting the unit working condition data, equipment status data and environmental data according to the second frequency to the field equipment and monitoring terminal, and completing the tag alignment, time synchronization and data caching. In order to map multi-source monitoring data into a unified input vector, the system uses the following state encoding formula to complete the data organization of the architecture entry:

$$x_t = \left[\alpha_p \tilde{p}_t, \alpha_e \tilde{e}_t, \alpha_h \tilde{h}_t, \alpha_b \tilde{b}_t\right] \tag{1}$$

where x_t represents the unified monitoring state vector at time t; $\tilde{p}_t$ represents the coding results of physical operation parameters, corresponding to speed, head, flow, output, vibration and swing. $\tilde{e}_t$ represents the environmental quantity coding result, which corresponds to information such as upstream and downstream water level, water temperature and sediment. $\tilde{h}_t$ represents the coding result of equipment health quantity, corresponding to bearing temperature, seal state and insulation state; $\tilde{b}_t$ represents the business process quantity coding results, corresponding to the starting and stopping, grid connection, peak shaving and frequency modulation and other operating conditions events. α_p,α_e,α_h,α_b are the weight coefficients of different dimensions. The function of Equation (1) is to compress the four types of heterogeneous data into a unified input space, which is convenient for the subsequent edge processing layer and platform analysis layer to call directly.

The edge processing layer is responsible for noise filtering, dimension normalization, missing measurement compensation and rule prediction, so that the monitoring data entering the upper layer can be calculated. In order to describe the standardized aggregation process of edge processing layer for multi-dimensional monitoring data, this paper uses the following formula to complete cross-index fusion calculation and time series alignment:

$$z_t = \sum_{i=1}^{m} w_i\, \hat{x}_{i,t}, \quad w_i = \frac{\exp(\beta_i)}{\sum_{j=1}^{m} \exp(\beta_j)} \tag{2}$$

Here, z_t represents the aggregated features output by the edge processing layer. $\hat{x}_{i,t}$ denotes the i normalized monitored quantity; m represents the number of indicators involved in the fusion; w_i represents the index weight; The β_i represents the weight parameter jointly determined by the historical volatility, the condition sensitivity and the stability of the missing test. Equation (2) is used to convert the monitoring quantities of different dimensions and different sampling sensitivities into unified and comparable fusion results, so that the output of the edge layer not only retains the key state information, but also weakens the traction effect of a single high fluctuation index.

The platform analysis layer is the computing center of the whole system, which undertakes the core tasks of multi-dimensional feature construction, steady state and transient state identification, working condition index calculation, abnormal pattern extraction and early warning event generation. When the unit is in the state of stable operation, load regulation, shutdown load reduction, concurrent grid connection or fast shutdown, the data distribution and change trajectory are not the same. The system needs to consider two paths of distribution identification and process identification at the same time. In order to express the joint discrimination mechanism of the platform analysis layer for the working state, the system defines the following state probability fusion model:

$$P(c_k|t) = \lambda P_s(c_k|z_t) + (1-\lambda) P_l(c_k|Z_{t-L:t}) \tag{3}$$

Here, $P(c_k|t)$ represents the comprehensive probability that time t belongs to the k type of working condition. $P_s(c_k|z_t)$ represents the discrimination probability given by the steady-state recognition model based on the current aggregated features. $P_l(c_k|Z_{t-L:t})$ represents the process probability given by the transient recognition model based on the sequence characteristics in the time window $[t-L,t]$. λ is the fusion coefficient of the steady state and transient paths. Let $Z_{t-L:t}$ denote a temporal feature fragment of length L. Equation (3) reflects the idea that the platform analysis layer uses distribution recognition and process recognition to complete joint discrimination at the same time, which is suitable for multi-state switching scenarios under the condition of regional centralized control.

The business application layer is responsible for monitoring display, rule orchestration, event tracking, comparison of similar equipment and operation and maintenance advice output, so that the status recognition results can directly serve the regional centralized control production business. The monitoring results not only correspond to single alarm output, but also precipitate into traceable working condition portraits, index trajectories and event evidence. In order to realize the graded early warning and feedback output of the business application layer, this paper further constructs the comprehensive risk scoring formula as follows:

$$R_t = \eta_1 A_t + \eta_2 D_t + \eta_3 C_t + \eta_4 G_t \tag{4}$$

Here, R_t represents the comprehensive risk score at time t. A_t represents the abnormal amplitude index, which is used to describe the deviation degree of vibration, temperature or pressure pulsation. D_t stands for anomaly duration index, which is used to measure the anomaly holding time. C_t represents the conflict degree of working condition transformation, which is used to describe the inconsistency degree between the recognition result and the rule chain. G_t represents the number of rule violations, which is used to count the number of abnormal rules triggered in the current time window. η_1,η_2,η_3 and η_4 are the risk weights. Equation (4) transforms the platform side analysis results into executable business warning basis, so that the monitoring display, event tracking and maintenance suggestions can correspond to uniform scoring rules.

In general, the architecture organizes the perception access, edge processing, platform analysis and business application into a continuous data and computing chain, so that the multi-source heterogeneous monitoring data in the regional centralized control scene can achieve unified access, unified expression, unified discrimination and unified feedback. The perception layer ensures the integrity of field data collection, the edge layer ensures the computability of input data, the platform layer undertakes the core tasks of state recognition and anomaly analysis, and the business layer is responsible for transforming analysis results into monitoring, early warning and maintenance information.

3.2 Feature construction and state representation method for multi-dimensional heterogeneous data

The state recognition of regional centralized control hydropower equipment does not rely on a single measurement point, but on the collaborative expression of unit operating parameters, environmental boundaries, health states and business process information. Vibration, swing, temperature, head, flow, power and pressure pulsations show different distribution characteristics and evolution trajectories under the conditions of stable operation, thermal reserve, load regulation, start and stop, load rejection and so on. Therefore, the core of feature construction is not simply splicing measurement points. Instead, it maps data from different sources, sampling frequencies and semantic levels into a computable state representation.

In order to convert the four types of heterogeneous monitoring quantities into comparable unified inputs, the system first completes the fractal dimension normalization process, and then constructs the weighted state encoding vector as follows.

$$x_t = \left[\omega_p \frac{p_t - \mu_p}{\sigma_p + \varepsilon}, \omega_e \frac{e_t - \mu_e}{\sigma_e + \varepsilon}, \omega_h \frac{h_t - \mu_h}{\sigma_h + \varepsilon}, \omega_b \frac{b_t - \mu_b}{\sigma_b + \varepsilon}\right] \tag{5}$$

Here, x_t denotes the uniform state input vector at time t; p_t, e_t, h_t, b_t represent physical operation parameters, environmental parameters, health state parameters and business process parameters respectively. μ_p, μ_e, μ_h, μ_b represent the mean vector of each dimension parameter in the history window. Let σ_p , σ_e , σ_h and σ_b denote the corresponding standard deviation vectors. ω_p , ω_e , ω_h , ω_b represent the weight coefficients of the four types of information. The tiny constant ε is set to prevent the denominator from being zero. The function of Equation (5) is to eliminate the offset caused by different dimensions and different fluctuation ranges, and then encode the four types of data into a unified input, so that the subsequent statistical features, time-frequency features and state embedding are established on a consistent data scale.

Steady state condition recognition requires the continuous monitoring values to be converted into comparable statistical expressions, so the system extracts the basic distribution characteristics of the data in the time window. To transform the continuous monitoring values in the time window into a comparable steady-state distribution representation, the system extracts statistics such as mean, standard deviation and kurtosis, which are calculated in the following form.

$$\phi_s = [\mu_t, \sigma_t, \kappa_t], \quad \mu_t = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_t)^2}, \quad \kappa_t = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu_t}{\sigma_t}\right)^4 \tag{6}$$

Here, ϕ_s denotes the steady-state statistical eigenvector. μ_t represents the mean value in the time window, which is used to describe the central position of the monitored quantity. σ_t represents the standard deviation, which is used to describe the fluctuation intensity. κ_t is the kurtosis, which describes the kurtosis of the distribution. N denotes the number of window samples. Equation (6) converts the original monitoring value into the steady-state representation oriented to distribution identification, which is suitable for the working conditions with clear boundaries such as hot standby state, grid-connected state and stable operation state.

In addition to statistical features, vibration, swing and pressure pulsation also have obvious time-frequency coupling characteristics. It is difficult to reflect the change of frequency structure before and after load switching with only time-domain quantity description, so the system continues to construct time-frequency representation. In addition to the basic statistical features, in order to further describe the joint variation relationship between vibration, swing and pressure pulsation signals in time domain and frequency domain, the system constructs a short-time Fourier time-frequency representation, which is expressed in the following form.

$$S_x(\tau, f) = \sum_{n=-\infty}^{+\infty} x(n)w(n - \tau)e^{-j2\pi fn} \tag{7}$$

Here, $S_x(\tau, f)$ represents the time-frequency representation of the signal $x(n)$ at time position τ and frequency f. $w(n-\tau)$ is the sliding window function; Let $e^{-j2\pi fn}$ denote the complex exponential basis functions. Equation (7) is used to capture the local variation of energy in different frequency bands with time under steady-state operating conditions, so that the periodic structure of vibration and pressure pulsation can enter the state representation process.

In order to understand the organizational relationship of multi-dimensional heterogeneous features in this section, the link of feature construction and state representation is summarized as the process shown in FIG. 2.

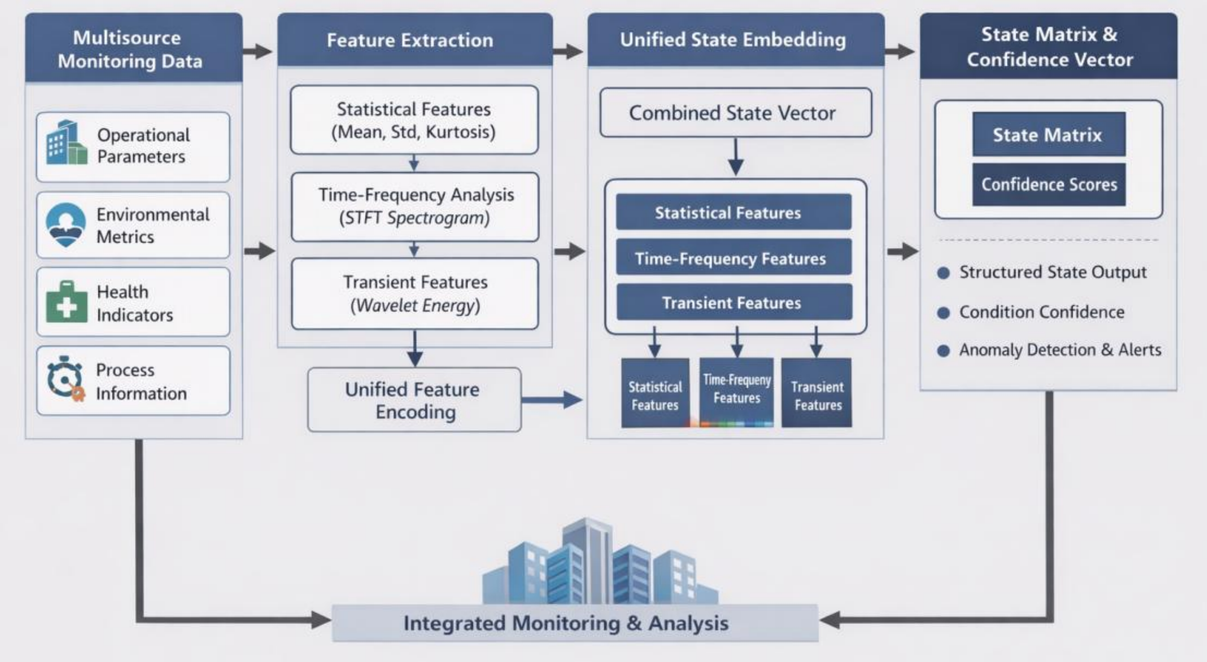

Figure 2: Multi-dimensional heterogeneous data feature construction and state representation process

For transient processes such as startup and shutdown, load increase and decrease, and load dump, feature construction cannot stay at the static distribution level, but also needs to identify the mutation location and energy transfer in the time series. In view of the energy redistribution phenomenon in the transient processes such as starting and stopping, increasing and decreasing load and load rejection, the system further uses the wavelet packet decomposition to quantify the relative energy of each frequency band, and its calculation form is as follows.

$$E_r = \frac{\sum_{q=1}^{Q} \left|d_{r,q}\right|^2}{\sum_{u=1}^{U} \sum_{q=1}^{Q} \left|d_{u,q}\right|^2} \tag{8}$$

where, E_r represents the relative energy of wavelet packet in the r frequency band. $d_{r,q}$ denotes the wavelet packet coefficients of the r frequency band at the q sampling point. U denotes the total number of frequency bands; Q represents the number of samples in a single frequency band. Equation (8) quantizes the energy redistribution in the transient process, which is convenient to distinguish between the normal start-stop process and the unexpected load mutation process.

After extracting steady-state features, time-frequency features and transient features, the platform also needs to organize these results into a unified representation, which can be used by the rule engine, statistical learning model and temporal network. After the steady-state statistical features, time-frequency features and transient features are extracted, the platform maps the feature vectors from different sources into a unified space and constructs a joint state embedding representation, which is calculated in the following form.

$$h_t = \tanh\left(W_s\phi_s + W_f\phi_f + W_p\phi_p + b\right) \tag{9}$$

Here, h_t denotes the joint state embedding vector at time t. Let ϕ_s denote the steady-state statistical characteristics; ϕ_f is the time-frequency feature; Let ϕ_p denote the characteristics of the transient process; W_s,W_f,W_p represent the mapping matrix of different feature branches; b represents the bias term; $\tanh(\cdot)$ is used to compress the output range and enhance the nonlinear representation. Equation (9) is used to compress the features from different sources into a unified embedding space, so that the state representation has a consistent dimensional structure.

State representation requires not only service model input, but also service condition rule configuration, indicator analysis, similar equipment comparison and exception interpretation. Therefore, the platform also needs to convert the embedding results into a structured representation that can be read by the business side. In order to form the representation results that can be directly invoked by the rule engine and business applications, the platform further constructs the state matrix and the condition confidence vector as shown below.

$$M_t = [h_t, r_t, q_t], \quad q_t = \mathrm{softmax}\left(W_h h_t + W_r r_t + b_q\right) \tag{10}$$

where M_t represents the state representation matrix at time t; h_t represents the joint state embedding obtained by the fusion of steady-state statistical features, time-frequency features and transient process features. r_t represents the rule encoding vector generated according to the working condition rules, pre and post state relationships and event constraints. q_t represents the confidence distribution corresponding to each candidate condition. W_h and W_r represent the mapping matrix between feature embedding and rule encoding, respectively. b_q represents the bias vector; $\mathrm{oftmax}(\cdot)$ is used to convert the linear combination into a comparable condition confidence. The function of Equation (10) is to compress the model features and rule features into the unified representation result, so that the platform can not only output the structured state description, but also give the credibility of various working conditions synchronously, which is used for subsequent state discrimination, anomaly analysis and early warning feedback.

In general, the method in this section transforms multi-dimensional heterogeneous monitoring quantities into computable, comparable and interpretable state representations, which not only retains the distribution differences of steady-state operating conditions, but also retains the evolution characteristics of transient processes. The feature construction and state representation methods thus formed provide direct input for subsequent multi-dimensional monitoring data mining modules and equipment state discrimination in regional centralized control scenarios, and also maintain consistent data semantics between rule configuration, indicator analysis and anomaly tracking.

3.3 Design of multi-dimensional monitoring data mining module

The monitoring data mining module of regional centralized control hydropower equipment is not a single model calling process, but a combined computing unit built for multi-source heterogeneous data, different operating stages and different analysis goals. In the process of stable operation, thermal standby, load regulation, start-up and shutdown, and load rejection, the data distribution of the unit is obviously different. A single algorithm is difficult to simultaneously take into account the steady-state identification accuracy, transient process tracking ability and rule interpretation consistency. Therefore, the data mining module is divided into sample organization unit, steady state mining unit, transient mining unit and result fusion unit, so that multi-dimensional monitoring data can be processed continuously along the path of "input sorting, feature calling, state output-abnormal feedback".

In order to ensure that the second sampling data can be correctly sent to the corresponding module in different operating stages, the system first completes the time window organization and regular mask constraints. In order to organize the continuous monitoring sequence into input fragments that can be directly called by each branch module, the system defines the following rules to constrain the window construction formula.

$$X_t^* = [x_{t-L+1}, x_{t-L+2}, \dots, x_t] \odot m_t \tag{11}$$

where X_t^* denotes the valid input window at time t; x_t represents the monitoring state vector at a single time; L denotes the length of the time window; m_t represents the constraint mask composed of working condition rules, pre and post state relations and event tags. $\odot$ denotes the bitwise constraint operation. The function of Equation (11) is to arrange the original continuous sequence into computable fragments according to the working condition boundary, so as to avoid aliasing of start-stop and steady-state samples in the input stage.

After the sample organization is completed, different branch modules undertake different data mining tasks. The steady-state mining unit is oriented to the working conditions with clear distribution boundaries, and is suitable for calling statistical features, time-frequency features, SVM, random forest and other models. The transient mining unit is oriented to the continuous evolution process, focusing on the time related states such as start and stop, load increase and decrease, and load dump, which is suitable for the call sequence fragment and LSTM structure. The result fusion unit is responsible for unifying the two types of output and rule information into the same representation space, so that the subsequent index analysis and abnormal feedback are consistent.

Table 2: Multidimensional monitoring data mining module and functional description

Module Name	Input Content	Core Processing	Output Result	Functional Role
Sample Organization Unit	Second-level monitoring sequences, event annotations, operating condition rules	Time alignment, window segmentation, rule-mask constraints	Computable sample segments	Ensures input consistency
Steady-State Mining Unit	Statistical features, time-frequency features	SVM and RF-based distribution recognition	Steady-state condition identification results	Identifies stable operating states
Transient Mining Unit	Time-series segments, process labels	LSTM-based process modeling	Transient state trajectories	Tracks dynamic processes
Result Fusion Unit	Steady-state outputs, transient outputs, rule vectors	Confidence fusion, conflict correction	Integrated state results	Ensures result consistency
Indicator Analysis Unit	State results, key indicator sequences	Fluctuation analysis, deviation analysis	Indicator anomaly descriptions	Supports interpretive analysis
Warning Feedback Unit	Integrated state results, rule thresholds	Hierarchical warning, event write-back	Warning events and feedback information	Forms a closed-loop process

In order to unify the output of steady state branch, transient branch and regular branch into a single result, the platform continues to construct a multi-branch joint mining model. In order to express the joint output relationship of multi-branch data mining module, this paper uses the following branch fusion formula to complete the comprehensive state calculation process.

$$y_t = \gamma_1 f_{svm}(\phi_s) + \gamma_2 f_{rf}(\phi_s) + \gamma_3 f_{lstm}(X_t^*) \tag{12}$$

where y_t represents the joint mining output at time $t. f_{svm}(\phi_s)$ is the svm output based on the steady-state statistical features ϕ_s. $f_{rf}(\phi_s)$ represents the output of random forest based on the same steady-state feature. $f_{lstm}(X_t^*)$ represents the lstm output based on the time window samples. γ_1,γ_2,γ_3 denote the weight coefficients of the three classes of branches. Equation (12) combines distribution identification and process identification, so that the module design can cover both steady-state identification requirements and transient tracking requirements.

Joint mining results also need to be coordinated with rule configurations and business thresholds to form interpretable state results. Only model scores are not enough to support regional centralized control services. It is also necessary to let the comprehensive output correspond to the rule chain, indicator chain and early warning chain. In order to form the module output that can be directly invoked by index analysis and early warning feedback, the system further defines the state confidence update formula as follows.

$$c_t = \mathrm{softmax}\left(W_y y_t + W_r r_t + b_c\right) \tag{13}$$

where, c_t represents the condition confidence vector; r_t represents the condition rule encoding vector; W_y and W_r denote the mapping matrix between the model output and the rule output, respectively. b_c represents the bias term;$\mathrm{softmax}(\cdot)$ is used to convert the synthesized results into a comparable confidence distribution. Equation (13) is used to compress data mining results and rule information into a unified output space, so that index analysis, anomaly interpretation and early warning classification can directly read the same set of state results.

From the perspective of module collaboration, the sample organization unit is responsible for segmenting the continuous monitoring sequence into effective inputs that meet the boundary constraints of the operating conditions, the steady state mining unit is responsible for extracting the stable state difference from the statistical distribution, the transient mining unit is responsible for identifying the state transition from the time evolution process, and the result fusion unit is responsible for compressing the multi-model output and rule information into a unified confidence result. The indicator analysis and early warning feedback unit continues to follow this result backward to complete deviation identification, anomaly marking, and event writeback. After this setting, each module forms a clear division of labor in the three links of input, calculation and output, and the multi-dimensional monitoring data can also enter the subsequent equipment state discrimination and anomaly analysis process along a unified link.

3.4 Equipment state discrimination and anomaly analysis method in regional centralized control scenario

The equipment state discrimination under the condition of regional centralized control cannot stay at the single threshold alarm level. When the unit is in different working conditions, the same vibration amplitude, temperature deviation or pressure pulsation do not correspond to the same risk meaning, so the condition identification results, index deviation degree and rule constraints must be included in the analysis chain at the same time. In this paper, the equipment state discrimination is divided into four links: condition confidence determination, index deviation analysis, early warning level generation and feedback correction, so that the regional centralized control platform can convert real-time monitoring results into structured state conclusions.

In order to unify the model output and the working condition rules into the same discrimination result, the system first constructs the comprehensive judgment function of the working condition state to complete the first-layer state recognition process as follows.

$$S_k(t) = \alpha_1 P_k(t) + \alpha_2 R_k(t) + \alpha_3 H_k(t) \tag{14}$$

where, $S_k(t)$ represents the comprehensive judgment score of the k type of working condition at time t. $P_k(t)$ represents the operating probability output by the steady-state or transient identification model. $R_k(t)$ is the rule matching score; $H_k(t)$ represents the similarity score of historical working condition samples of the same class. α_1,α_2, and α_3 are the weight coefficients. Equation (14) is used to compress the data-driven and rule-driven results into a unified operating condition score, so that the platform can first give the current state of the equipment before proceeding to the next step of anomaly analysis.

After the condition is determined, the anomaly analysis does not directly use the original value, but calculates the deviation around the baseline of the same condition, so as to avoid misjudgment between different operating states. In order to quantitatively describe the deviation degree of key indicators relative to the baseline of the same working condition, this paper further defines the calculation formula of abnormal deviation of indicators as follows.

$$D_t = \sum_{i=1}^{m} \omega_i \frac{\left| x_{i,t} - \bar{x}_{i,c} \right|}{\sigma_{i,c} + \varepsilon} \tag{15}$$

Here, D_t represents the comprehensive deviation index at time t; $x_{i,t}$ denotes the i real-time index value; $\bar{x}_{i,c}$ denotes the historical mean value of the i index under the current working condition c; $\sigma_{i,c}$ denotes the corresponding standard deviation; Let ω_i denote the index weights; m represents the number of indicators involved in the analysis; ε is a tiny constant set to prevent the denominator from being zero. Equation (15) can be used to uniformly characterize the deviation intensity of vibration, temperature, swing, efficiency and stability under the current working condition.

After the comprehensive deviation is formed, the system also needs to continue to convert the abnormal amplitude, duration and event attributes into executable warning levels. To map the abnormal deviation, duration and event attributes into hierarchical warning results, the platform continues to construct the comprehensive warning score formula as follows.

$$W_t = \beta_1 D_t + \beta_2 T_t + \beta_3 E_t + \beta_4 N_t \tag{16}$$

Among them, W_t represents the comprehensive warning score; D_t is the comprehensive deviation index. T_t represents the anomaly duration normalization. E_t stands for event severity coefficient, which is used to distinguish general abnormal, early warning and fault level events. N_t represents the penalty term for unexpected condition conversion; β_1,β_2,β_3, and β_4 are the scoring weights. Equation (16) is used to transform multi-source abnormal factors into scalable early warning results, so that the platform can output state conclusions and treatment suggestions at different response levels.

After the early warning output is completed, the regional centralized control system also needs to dynamically modify the threshold according to historical events and seasonal changes to ensure that the discriminant model remains stable under different environmental conditions. In order to achieve dynamic threshold correction and feedback update, the system uses the following adaptive threshold update formula to complete the closed-loop feedback process.

$$\theta_{t+1} = \lambda\theta_t + (1-\lambda)(\mu_a + \gamma\sigma_a) \tag{17}$$

Here, θ_{t+1} represents the updated early warning threshold in the next period. Let θ_t denote the current threshold; μ_a represents the mean value of recent abnormal samples of the same class; σ_a represents the fluctuation degree of abnormal samples of the same class. γ is the threshold sensitivity coefficient; Let λ denote the historical threshold retention weight. Equation (17) is used to smoothly update the threshold according to the event accumulation results, so that the platform can still maintain the consistency of the discrimination results when the water temperature changes in summer, the load fluctuation increases, or the operation mode is adjusted.

Through the above four links, state discrimination and anomaly analysis are no longer a single output, but a continuous analysis chain of "working condition determination - deviation calculation - early warning grading - threshold correction". After setting in this way, the model results, rule results and historical event results can be consistent under the same logic, and the regional centralized control platform can further give similar equipment comparison, trend analysis and maintenance suggestions.

4 Experiment and result analysis

4.1 Validation of multi-dimensional monitoring data mining effect

In order to verify the usability of the multi-dimensional monitoring data mining module in the regional centralized control scenario, this paper systematically tests the steady state recognition, transient recognition and fusion discrimination effects in the offline computing environment. The second level monitoring data of six units were selected for the experimental object, covering eight types of inputs such as vibration, swing, temperature, head, flow, active power, pressure pulsations and event quantities. According to the above definition and rule configuration of working conditions, the sample segmentation of eight typical working conditions including shutdown standby, hot standby, grid-connected state, stable operation state, load regulation process, normal shutdown process, fast shutdown process and load rejection process was carried out, and 28800 valid sample segments were finally formed. Among them, there are 20160 training sets, 4320 validation sets, and 4320 test sets. The steady-state recognition branch uses statistical features and STFT features, the transient recognition branch uses wavelet packet features and LSTM sequence as input, and the fusion layer uses rule constraints and probability weighting mechanism to output the comprehensive state results.

As shown in FIG. 3, the proposed method is compared with single SVM, single RF, single LSTM and rule-free fusion models. On the test set, the Accuracy and F1 of the proposed method are 95.6% and 95.1%, which are higher than 89.4% and 88.7% of SVM, 91.2% and 90.4% of RF, and 93.1% and 92.5% of single LSTM. The Accuracy of the rule-free fusion model is 94.0%, indicating that the multi-branch output will still be confused when the state boundary is close if there is no rule constraint of the working condition. Figure 3 reflects that the overall stability of the condition recognition is higher after statistical features, timing features and rule information are cooperated.

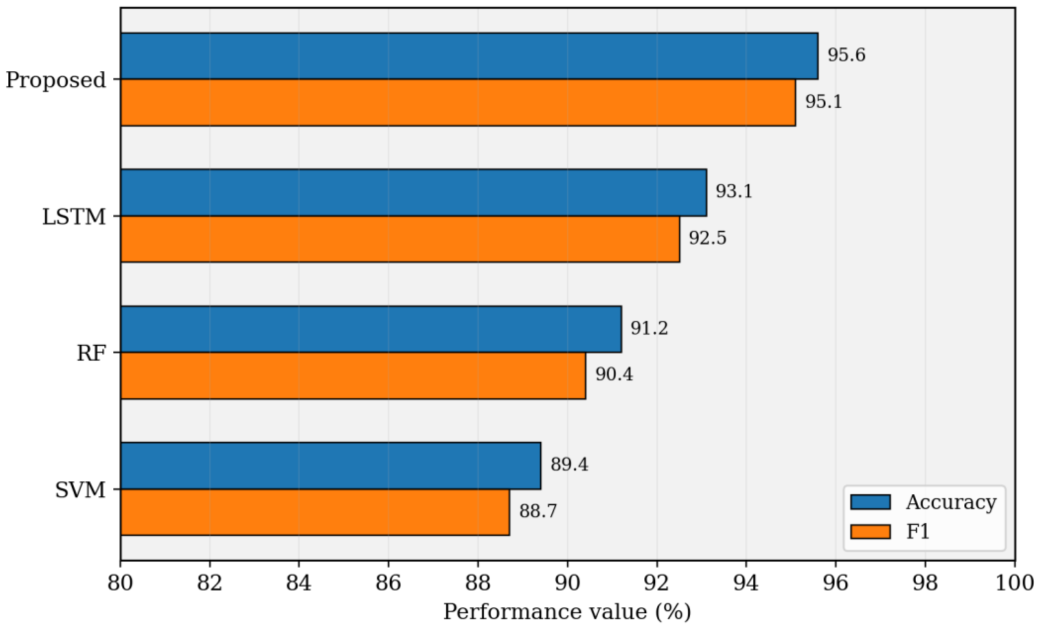

Figure 3: Accuracy versus F1 comparison of different models on the test set

To further illustrate the adaptability of the multidimensional data mining module under different condition types, Table 3 presents the recognition results on the set of four types of conditions. The recognition accuracy of steady condition is 96.3%, transient condition is 94.8%, test condition is 93.7%, and abnormal condition is 95.0%. The steady-state performance is the highest, because the statistical distribution boundary is clear. The transient condition is slightly lower, mainly from the similar dynamic characteristics of fast shutdown and load rejection in the local stage. The recall rate of abnormal working conditions remains at 94.2%, which indicates that the fusion module can better identify unexpected states.

Table 3: The recognition effect on the set of different working conditions

Operating Condition Set	Accuracy / %	Precision / %	Recall / %	F1 / %
Steady-State Conditions	96.3	96.7	95.8	96.2
Transient Conditions	94.8	95.1	94.0	94.5
Test Conditions	93.7	94.2	92.6	93.4
Abnormal Conditions	95.0	95.5	94.2	94.8
Average	95.0	95.4	94.2	94.7

Figure 4 continues with the confusion distribution for each condition category. The recognition accuracy of grid-connected state and stable operation state are 96.8% and 97.1%, 95.4% for hot standby state, 94.2% for load regulation process, 93.8% for normal shutdown process, 92.9% for fast shutdown process, and 93.1% for load rejection process. The cross misjudgment rate of fast shutdown process and load rejection process is relatively high, but the misjudgment rate decreases from 8.6% to 5.1% after the fusion rule is added, which indicates that the rule chain has an obvious correction effect on the transient boundary.

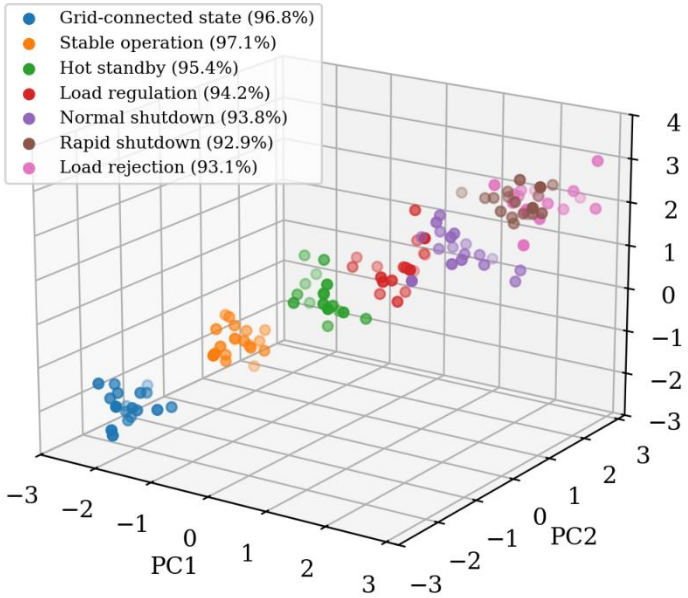

Figure 4: Distribution of recognition accuracy for each typical working condition category

To evaluate the impact of the feature construction strategy on the mining effect, Figure 5 compares three feature combinations. When using only statistical features, the F1 is 90.6%; When adding STFT, it increased to 93.4%. After adding the wavelet packet process features on this basis, F1 reaches 95.1%. This result shows that the steady-state features can support the basic classification, the time-frequency features enhance the description of vibration and pressure pulsations, and the process features further improve the discrimination ability of transient operating conditions such as start-up and shutdown, load increase and decrease. On the whole, the combination of multi-dimensional features has a direct support role in the effect of data mining in regional centralized control scenarios.

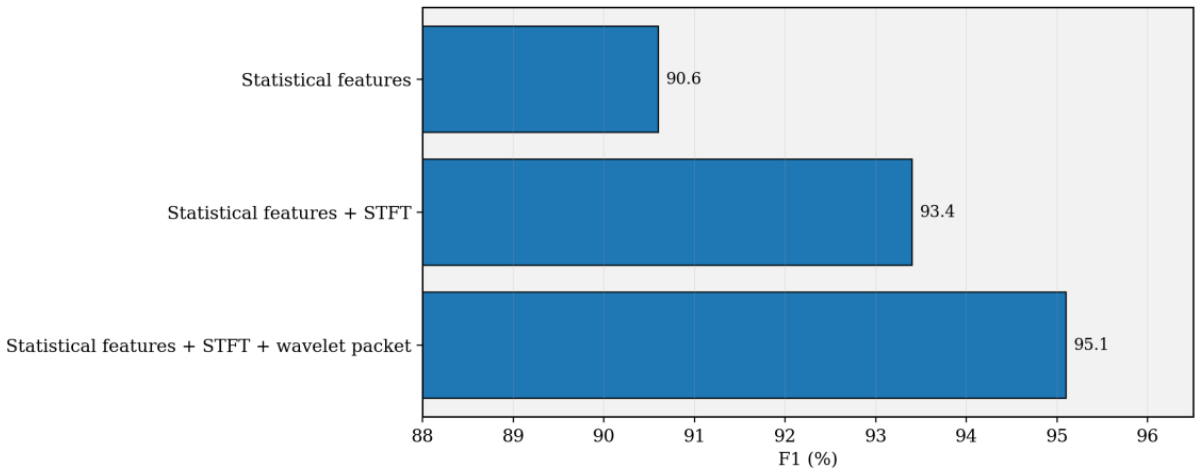

Figure 5: F1 comparison for different combinations of features

Synthesizing the above verification results, it can be seen that the multi-dimensional monitoring data mining module has better recognition stability and feature adaptation ability in the regional centralized control scenario. The distribution recognition of steady-state conditions maintains a high accuracy, and the process discrimination of transient conditions is also significantly improved after rule constraints and multi-branch fusion. The statistical features, time-frequency features and process features form a complementary relationship, so that the model can still maintain consistent output under different operating conditions.

4.2 Equipment state identification and anomaly analysis and evaluation under regional centralized control scenario

In order to evaluate the practical applicability of the multi-dimensional monitoring data mining module under the regional centralized control condition, this paper selected 4 hydropower stations, 12 units and 84 key measurement points to carry out field pilot verification, and continuously connected to the vibration, swing, temperature, head, flow, active power, pressure ripple and switching data event data, and the pilot period was 21 days. All data are collected at the frequency of seconds, and online identification and anomaly analysis are carried out according to the above definition of working conditions, rule configuration and index system. The field platform also records the status category, warning level, response delay and event writeback results for comprehensive evaluation of recognition accuracy and service availability.

As shown in Figure 6, the recognition accuracy of the system in the grid-connected state, stable operation state, load regulation process, normal shutdown process and fast shutdown process reaches 96.4%, 97.0%, 94.6%, 93.8% and 92.7%, respectively. The stable operation state and grid-connected state are the most stable, because the boundary of steady state distribution is clear. The fast shutdown process is slightly lower, which is mainly affected by process mutation and overlapping of adjacent operating conditions. In abnormal events, the recognition agreement rates of vibration exceeding the limit and high temperature reached 95.1% and 94.4%, respectively, indicating that the multi-source features still maintain good discrimination ability under field conditions.

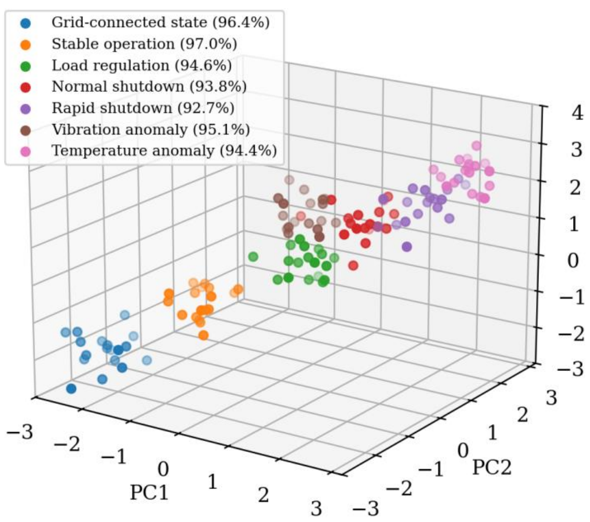

Figure 6: Distribution of typical working conditions and abnormal event recognition accuracy in the field pilot

To further illustrate the comprehensive performance of the system in the regional centralized control site, Table 4 presents the key evaluation results in four pilot scenarios. The

Accuracy of each scene is maintained above 94.9%, and the highest is 95.8%, which shows that the method has a relatively stable recognition ability under the conditions of different stations and units. The Recall is maintained between 93.8% and 94.5%, which indicates that the abnormal state and the transition process of the operating condition can be captured completely. The response delay is controlled between 0.58 s and 0.66 s, which meets the real-time analysis requirements in the regional centralized control scenario. The false alarm rate is always lower than 4.1%, and the warning advance amount remains between 17.8 min and 19.4 min, indicating that the state discrimination results can support on-site early warning and operation and maintenance decisions.

Table 4: Comparison of comprehensive evaluation results of the field pilot

Pilot Scenario	Accuracy / %	Recall / %	F1 / %	Response Latency / s	False Alarm Rate / %	Early Warning Lead Time / min
Unit Cluster No. 1 at Station A	95.6	94.5	95.0	0.58	3.5	19.4
Unit Cluster No. 2 at Station B	94.9	93.8	94.3	0.64	4.1	17.8
Unit Cluster No. 3 at Station C	95.1	94.0	94.6	0.66	3.9	18.1
Unit Cluster No. 4 at Station D	95.8	94.3	95.0	0.60	3.6	19.1
Average	95.3	94.1	94.7	0.62	3.8	18.6

In order to analyze the actual contribution of each module to the field effect, this paper continues the ablation experiment, and the results are shown in Table 5. The Accuracy of the full model is 95.3% and F1 is 94.7%. After removing the rule constraint branch, the Accuracy drops to 92.8%, which indicates that the working condition rule has an obvious correction effect on the boundary of adjacent states. After removing the transient identification branch, F1 drops to 91.9%, indicating that the start-stop and load change processes need independent timing modeling. After removing the index analysis branch, the false alarm rate increases from 3.8% to 6.4%, which indicates that the index deviation analysis of the same working condition has direct value for anomaly interpretation and alarm screening.

Table 5: Results of ablation experiments

Model Configuration	Accuracy / %	F1 / %	False Alarm Rate / %	Response Latency / s
Full Model	95.3	94.7	3.8	0.62
Without Rule-Constraint Branch	92.8	92.1	5.9	0.57
Without Transient Recognition Branch	93.4	91.9	4.8	0.49
Without Indicator Analysis Branch	94.0	93.1	6.4	0.55

The comprehensive field results show that the multi-dimensional monitoring data mining module not only maintains a high recognition accuracy, but also forms a stable anomaly analysis ability under the condition of regional centralized control. The rule branch, transient branch and indicator analysis branch show complementary relationships, so that the state recognition results can be directly transformed into early warning levels and operation and maintenance tips. This shows that the feature construction, module design and state discrimination path proposed in the previous section have been able to adapt to the real hydropower operation environment, and provide a reliable basis for the discussion in the next section.

4.3 Discussion

The state recognition and anomaly analysis method of regional centralized control hydropower equipment constructed in this paper concatenates operating condition rules, steady-state distribution identification, transient process identification and index deviation analysis into a continuous data processing chain. Compared with the schemes relying on single-class statistical characteristics or single time series model, this combination mode is more suitable for the operation characteristics of hydropower units with frequent switching conditions. The above results show that the rule constraint can tighten the discrimination boundary between adjacent working conditions, the transient identification branch can enhance the discrimination effect of the start-stop and load adjustment process, and the index analysis branch can improve the consistency of anomaly interpretation. Such a processing path is consistent with the four-dimensional condition portrait, condition rule configuration and early warning feedback logic proposed in the document. It should also be noted that the monitoring data in the regional centralized control scenario will be jointly affected by the fluctuation of water head, the change of environmental temperature, the switch of operation mode and the integrity of measurement points. When the distribution of samples in the same working condition drips, the model output will appear boundary compression. In the future, the historical alignment of the same working condition, dynamic threshold update and lightweight timing modeling can be further strengthened, so that the state discrimination results can maintain more stable performance under cross-site, cross-season and cross-load conditions, and provide a finer basis for maintenance advice output. Overall, the method in this paper unifies condition recognition, probability output, rule matching and event feedback into the same calculation chain, which makes the results easier to track and more suitable for software analysis and intelligent operation and maintenance in regional centralized control scenarios

5 Conclusion

This paper focuses on the multi-dimensional data mining and analysis technology in the condition monitoring of regional centralized control hydropower equipment, and constructs a complete calculation chain covering the overall architecture, feature construction, module design, state discrimination and anomaly analysis. Based on the four-dimensional portrait of "physics-environment-health-business", the system organizes the steady-state distribution identification, transient process identification, index deviation analysis and early warning feedback into a unified process, so that the regional centralized control platform can output continuous and interpretable state results under the conditions of multiple units, multiple stations and multiple working conditions switching. The experimental results show that the proposed method maintains a good performance in terms of the accuracy of working condition recognition, the stability of anomaly analysis and the effectiveness of on-site early warning, which can provide data support for operation monitoring, trend analysis and maintenance advice. It should be noted that the current method is still sensitive to sample distribution drift under the same working condition, extreme environmental disturbance and cross-site measurement point differences. When the water head change, seasonal temperature change or sampling quality fluctuation are superimposed, the state boundary will shrink, which will affect the consistency of anomaly interpretation. Subsequent research can continue to strengthen the historical alignment of the same working condition, dynamic threshold update, lightweight time series modeling and cross-station migration mechanism, so as to maintain stronger robustness of condition monitoring results in regional centralized control scenarios, and further support intelligent early warning and auxiliary decision-making. From the perspective of engineering implementation, this method integrates rule configuration, indicator analysis, event tracking and feedback writeback into the same software framework, which reserved extensible interfaces for the intelligent operation and maintenance of equipment under the condition of regional centralized control, and also made the whole calculation chain more complete, which was convenient for subsequent function expansion and actual deployment.

Funding

Science and technology project of Hubei Energy Group Lou Shui Hydropower Co., LTD., project name Research and Application of Key Technologies for Regional Hydropower Intelligent Remote Centralized Control, project No. : KJCX202500274

References

[1] Kumar K, Saini R P. Data-driven internet of things and cloud computing enabled hydropower plant monitoring system[J]. Sustainable Computing: Informatics and Systems, 2022, 36: 100823.

[2] Huang J, Guo B, Dian S. Condition monitoring and fault diagnosis of hydropower generator based on LSTM correction model[J]. Journal of Intelligent & Fuzzy Systems, 2023, 45(2): 2419-2436.

[3] Zhang F, Guo J, Yuan F, et al. Enhancement methods of hydropower unit monitoring data quality based on the hierarchical density-based spatial clustering of applications with a noise–Wasserstein slim generative adversarial imputation network with a gradient penalty[J]. Sensors, 2023, 24(1): 118.

[4] Al-Hardanee O F, Demirel H. Hydropower station status prediction using RNN and LSTM algorithms for fault detection[J]. Energies, 2024, 17(22): 5599.

[5] Dao F, Zeng Y, Zou Y, et al. Fault diagnosis method for hydropower unit via the incorporation of chaotic quadratic interpolation optimized deep learning model[J]. Measurement, 2024, 237: 115199.

[6] Li Q, Xin L, Li R. Application of digital twin technology in monitoring system of pump turbine[J]. Discover Mechanical Engineering, 2024, 3(1): 30.

[7] Hajimohammadali F, Crisostomi E, Tucci M, et al. Evaluating deep learning networks versus hybrid network for smart monitoring of hydropower plants[J]. Energies, 2024, 17(22): 5670.

[8] Dao F, Zeng Y, Zou Y, et al. Wear fault diagnosis in hydro-turbine via the incorporation of the IWSO algorithm optimized CNN-LSTM neural network[J]. Scientific Reports, 2024, 14(1): 25278.

[9] Esmaeili Nezhad A, Samimi M H. A review of the applications of machine learning in the condition monitoring of transformers[J]. Energy Systems, 2024, 15(1): 463-493.

[10] Yang X, Zhao X, Shen X, et al. Intelligent Predictive Maintenance of Dry-Type Transformer Based on Vibration[J]. Wireless Communications and Mobile Computing, 2022, 2022(1): 2796167.

[11] Mei Y, Ni S, Zhang H. Fault diagnosis of intelligent substation relay protection system based on transformer architecture and migration training model[J]. Energy Informatics, 2024, 7(1): 120.

[12] Kang Z, Zhang Y, Du Y. Application of multi-sensor information fusion technology in fault early warning of smart grid equipment[J]. Energy Informatics, 2024, 7(1): 123.

[13] Shao N, Chen Q, Yu C, et al. Fault tracing method for relay protection system–circuit breaker based on improved random forest[J]. Electronics, 2024, 13(3): 582.

[14] Li T, Xia Y, Pang X, et al. Mechanical fault diagnosis of high voltage circuit breaker using multimodal data fusion[J]. PeerJ Computer Science, 2024, 10: e2248.

[15] Xu Z, Chen X, Li Y, et al. Hybrid multimodal feature fusion with multi-sensor for bearing fault diagnosis[J]. Sensors, 2024, 24(6): 1792.

[16] Zhang D, Zheng K, Liu F, et al. Fault diagnosis of hydraulic components based on multi-sensor information fusion using improved TSO-CNN-BiLSTM[J]. Sensors, 2024, 24(8): 2661.

[17] Bampoula X, Nikolakis N, Alexopoulos K. Condition monitoring and predictive maintenance of assets in manufacturing using LSTM-autoencoders and transformer encoders[J]. Sensors, 2024, 24(10): 3215.

[18] Surucu O, Gadsden S A, Yawney J. Condition monitoring using machine learning: A review of theory, applications, and recent advances[J]. Expert Systems with Applications, 2023, 221: 119738.

[19] Tama B A, Vania M, Lee S, et al. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals[J]. Artificial Intelligence Review, 2023, 56(5): 4667-4709.

[20] Matania O, Dattner I, Bortman J, et al. A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges[J]. Journal of Sound and Vibration, 2024, 590: 118562.