International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 6715-6721

https://doi.org/10.70517/ijhsa464573

Deep Learning Algorithms for Fault Analysis and Preprocessing of Hydroelectric Power Plant Equipment in Ensuring Sustainable Development of Housing Power Systems

Yan Ma[1], Jun Si[2], Qiuzhen Yan[1] and Jun Wang[1,*]
[1] School of Computer Science and Technology, Zhejiang University of Water Resources and Electric Power, Hangzhou, Zhejiang, 310018, China
[2] Zhejiang Provincial Energy Group Company Ltd., Hangzhou, Zhejiang, 311500, China
Corresponding authors: (e-mail: alexis_my@163.com).

Abstract This paper proposes a method for standardized collection of hydropower plant equipment data, and establishes a bi-directional long and short-term memory network (Bi-LSTM) model applying the attention mechanism. After the standardized collection of hydropower plant operation data and feature processing, an equipment fault diagnosis process is established, and a variety of fault pre-processing schemes are formulated according to the actual situation, such as adjusting parameters, distributing loads, hierarchical response and closed-loop feedback, etc. The Bi-LSTM model is also used in the experiments to verify the accuracy of the data collected. The experiments verified that the Bi-LSTM model surpasses the classical algorithms such as SVM, BP and CNN in fault identification accuracy, and its accuracy can reach 92.14% when the training set has 1000 samples. Moreover, the performance test of the system also shows stable response time, high transmission efficiency, and possesses good real-time and scalability. The proposed research can supply theoretical basis and technical route for constructing intelligent and solid housing power system, and promote the management of hydroelectric power station equipment in the direction of intelligent forecasting and automatic maintenance.

Index Terms hydropower plant equipment, equipment fault diagnosis, Bi-LSTM model, housing power system, automatic maintenance

I. Introduction

Accelerated urbanization has led to an increasing demand for electricity in residential houses, which has brought certain challenges to the operation of power systems [1]. Moreover, people have more stringent requirements on the stability and continuity of power supply [2]. As the main provider of clean energy, whether the hydroelectric power station can operate normally is directly related to the reliability of electricity for residents. Nowadays, in order to ensure the safety of traditional facilities, hydroelectric power plant enterprises pay great attention to the management of internal operations, and has taken a number of perfect measures to strive for economic interests and social benefits go hand in hand. However, the current fault diagnosis technology of hydropower plant equipment still exists the phenomenon of slow response and lack of accuracy, which cannot meet the standard of sustainable development of the housing power system [3]. Utilizing the excellent fault recognition function and fast response mechanism of deep learning algorithms, it is promising to be popularized and applied in the protection of the power system, which in turn promotes the sustainability of the housing power system [4].

The purpose of this paper is to study the use of deep learning algorithms to carry out intelligent analysis and processing, improve the speed of fault diagnosis and develop corresponding preventive measures, such as changing the operating parameters of the machine, making maintenance schedules, etc., to minimize the damage of faults to the residential power system, and put forward feasible solutions for the development of hydropower plants.

II. Applications of deep learning

II. A. Data Acquisition and Processing

During the operation of hydropower plants, a wide variety of data from different sources are generated, and each type of data is stored with a unified sub-table format and configuration. For example, real-time data of water level and time period statistics, including hourly and monthly tables, contain sequences of features within different time periods. For the problems of insufficient calculation of various types of data of hydropower station equipment and different sequences of eigenvalues, this paper is based on the eigenvalue calculation and system operation calculation, the data acquisition and processing flow is shown in Figure 1, and the main steps include:

(1) In accordance with the type of operational data to categorize and organize the data, and clarify the specific data requirements in each type of data collection.

(2) Preliminarily determine the processing standards for each type of data set by combining the prescribed standards for hydropower plant equipment, industry norms and work experience.

(3) To strengthen communication and collaboration with residential, power grid and higher management units, and to uniformly determine data collection and processing rules.

(4) To strengthen the communication and collaboration with the residence, the power grid and the higher management units to reach the consistency of the data collection and processing specifications.

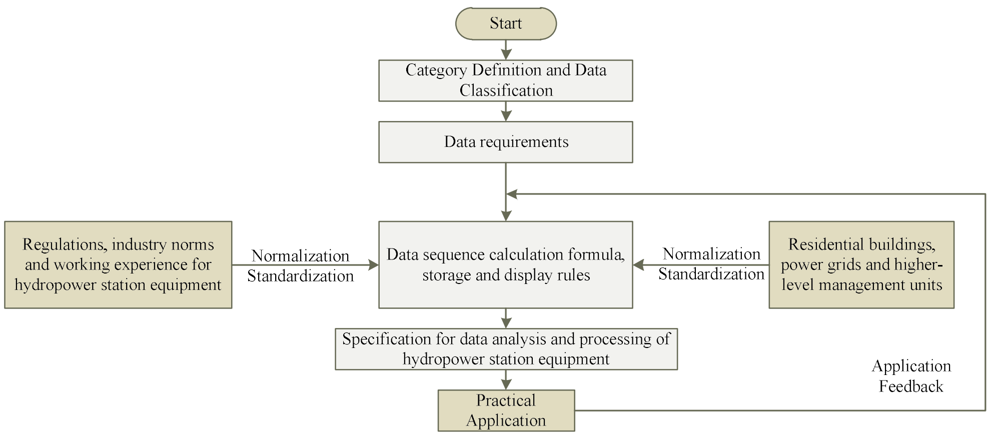


Figure 1: Data collection and processing flow

II. B.Model Selection and Establishment

Housing power system, hydroelectric equipment will generate a large amount of data, deep learning algorithms for a sufficient amount of data samples, with a certain effectiveness [5]. Deep learning in the long short-term memory network (LSTM) can not only predict the unidirectional text sequence information, but also over the hierarchical structure, extract sequence features, the introduction of the attention mechanism can accelerate the model training [6]. Therefore, in this paper, the attention mechanism and bidirectional long and short-term memory network (Bi-LSTM) model is constructed, and Figure 2 shows the Bi-LSTM network structure. The results of the collected and processed housing power data are used as an initialized weight matrix to analyze the dependencies before and after the sequence of hydroelectric power plant equipment in the word sequence encoding layer.

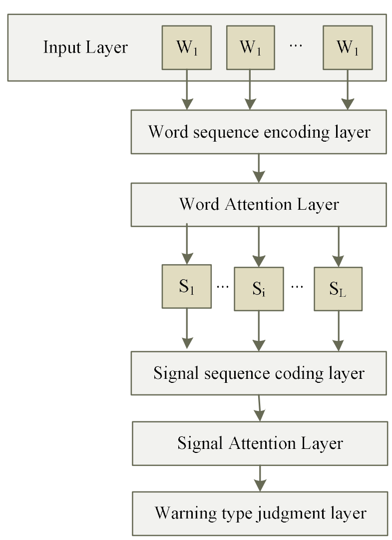


Figure 2: Bi-LSTM network structure

$w_t\left(t \in [1, N]\right)$ denotes the word vectors of the faulty signal i at time step t, and the hidden state of each word vector in the output signal of the Bi-LSTM after learning by preprocessing h is:

$$h_i = Bi - LSTM + +\left(w_t\right) \quad t \in [1, N] \tag{1}$$

Considering that only some of the words in the fault signals are informative for fault type diagnosis, the attention mechanism assigns weight α_{it} to the word hidden states according to the importance difference and weights all the hidden states h_{it} of a single warning signal i to generate an average generated signal vector s_i with the following formula:

$$u_{it} = \tanh\left(W_w h_{it} + b_w\right) \tag{2}$$

$$\alpha_{it} = \frac{\exp\left(u_{it}^T u_w\right)}{\sum_t \exp\left(u_{it}^T u_w\right)} \tag{3}$$

$$s_i = \sum_t \alpha_{it} h_{it} \tag{4}$$

where W_w and b_w are the weight parameter matrix and the bias vector of h_{it}, respectively, u_{it} is the output of h_i after the signal-attention layer, and u_w is the randomly initialized sequence based on the housing power system vector [7].

In the signal sequence coding layer, the input to each time step is the signal vector $s_i\left(i \in [1, L]\right)$. Since the timing of fault signals in equipment data acquisition is implicitly correlated, the sequence correlation between fault signals can be learned through this layer, and the output signal coding vector h_i:

$$h_i = Bi - LSTM + +\left(w_t\right) \quad t \in [1, L] \tag{5}$$

Fault warning events are multi-signal sequential textual data, and redundant signals have little effect on the type judgment, so it is important to distinguish different signals by the signal attention layer. Firstly, the key semantic features of the signals are extracted, then the implicit state weights of each time step output from the signal sequence coding layer are calculated to generate the event feature vector v_E, E denotes the warning event, and the formula is as follows:

$$u_i = \tanh\left(W_s h_i + b_s\right) \tag{6}$$

$$\alpha_i = \frac{\exp\left(u_i^T u_s\right)}{\sum_i \exp\left(u_i^T u_s\right)} \tag{7}$$

$$v_E = \sum_i \alpha_i h_i \tag{8}$$

where s denotes the warning signal, W_s and b_s are the weight parameter matrix and the bias vector of h_i, respectively, u_i is the output of h_i through the single-layer sensing machine, and u_s is the output of the single-layer sensing machine based on housing power system initialized signal vector.

The warning type determination layer then generalizes the deep features learned by the upstream network, and the softmax classifier is used to output the warning event type label with the following formula:

$$p = \operatorname{soft}\max\left(W_E v_E + b_E\right) \tag{9}$$

where W_E and b_E are the weight parameter matrix and bias vector of the layer, respectively, and p is the probability of the label of the equipment warning type, and the warning type corresponding to the largest element of p is selected by the softmax function as the final fault location, cause, etc. Based on the Bi-LSTM model, the hydroelectric power plant equipment fault categories are identified, so as to target the treatment strategy.

II. C.Model Training and Validation

Using the eigenvalues and system operation data, the operation data of the hydropower plant equipment are categorized and divided into a training set and a validation set in the ratio of 7:3 or 8:2 [8]. The purpose of the training set is to adjust the parameters and learn the features of the Bi-LSTM model, so that the model can recognize the hidden abnormal tendency and the precursor of faults in the operating conditions of the equipment. The validation set is used to test the performance of the model during the training process, to examine the ability to predict the unknown data, and to avoid overfitting in the training set. Through the dynamic observation of validation error, loss function and accuracy, the model parameters are adjusted in real time to ensure that the deep learning algorithms have good generalization and provide technical support for the normal operation and continuous progress of the residential power system.

II. D.Fault pre-processing strategy

When deep learning models accurately detect anomalies or hidden faults in the operation of the equipment, immediate and appropriate pre-developed countermeasures are necessary to minimize the negative impact on the normal operation of the residential power system and energy supply. Pre-established countermeasures can be implemented from the following perspectives:

(1) In the early stages of a fault, immediate analysis of the model output data can be used to quickly detect any tendency of abnormal operation of the equipment. Once abnormalities are detected, the operating parameters are immediately adjusted, such as reducing load, current, voltage and other important parameters, so that the equipment does not remain in an abnormal operating state for a long period of time, which in turn reduces the accumulation of stress on the main components, and achieves the purpose of slowing down the aging or re-damage of the machine. And this dynamic change of parameters can also achieve the effect of energy saving, and increase the stability and safety of the whole system [9].

(2) Once a hydroelectric power plant equipment failure occurs, it is necessary to rely on the central dispatching system to quickly dynamically adjust the energy load. With the combination of deep learning technology and big data control platform, it can intelligently analyze the current operating conditions of nearby hydropower plants, grid nodes or standby energy facilities, and make reasonable load distribution, so that the load in the fault area can be transferred in a timely manner. This not only maintains the uninterrupted power supply to residents, but also improves the resistance and adaptability of the residential power network, prevents the cascading effect caused by localized failures, and reduces the risk of large-scale power outages.

(3) Based on the results of the deep learning model's prediction of fault types, fault probabilities and development trends, the fault risk level can be accurately assessed. Accordingly, operation and maintenance personnel can deploy maintenance plans in advance for high-risk areas, and carry out key inspections and repairs for potential fault points. For the mild faults detected, adjustments can also be made through remote control or local intervention to avoid further expansion. Precision maintenance not only improves maintenance efficiency, but also effectively shortens downtime and resource loss [10].

(4) It is important to formulate hierarchical countermeasures for different levels of faults. For example, ordinary software failures can use remote reboot, such as low-cost methods to quickly restore the system to normal. Electrical system emergencies use medium-intensity methods such as isolation and temporary power supply. Major equipment damage requires the activation of back-up units and the replacement of parts or equipment. The multi-tiered response plan allows for flexibility and efficiency, minimizing the impact on residential electrical system operations.

(5) In order to achieve closed-loop management, all troubleshooting tools and their effectiveness should be uploaded to the central control platform immediately. The comparison of the model's determination with the results of actual repairs can improve the parameters of the model as well as the algorithmic construction, so that the prediction accuracy and reaction speed can be improved. Moreover, the feedback system is helpful for the establishment of a fault database, which is conducive to the rapid response to similar problems in the future [11], [12]. In this way, an intelligent cycle system of “prediction-intervention-feedback-optimization” is formed, which greatly improves the level of intelligent maintenance management of hydropower stations and the sustained security of the power supply system in residential areas.

The above multi-level fault pre-processing measures can significantly improve the safety and response sensitivity of hydropower plant equipment operation, thus laying a solid foundation for the continuity of power supply and intelligent management in residential areas.

III. Analysis of application results

III. A. Experimental setup

In this paper, the arithmetic data are derived from the operation data of a hydropower plant centralized control center for the years 2020-2024, as well as fault warning signals, which are stored in the form of a database. In the preparation stage of the experiment, the operational data of the hydropower plant involved in the training is divided into two groups, of which the training stage and the testing stage are each used in general. The data of both phases were expanded by data acquisition and processing, and the length of the sampling window was 1024, and the offset was set to 60. According to the training of the Bi-LSTM, there were 500 training samples for each case as well as 20 test samples, and there were no repetitions between these samples. In the experimental process, the Bi-LSTM model calculation method is compared with other methods to demonstrate its advantages in solving the fault diagnosis of hydropower plant equipment.

III. B. Performance evaluation

In order to verify the effectiveness of the improved Bi-LSTM algorithm, the other fault analysis and processing models were set up as a control group, from which 100, 200, 300, 400, and 500 samples were randomly selected and divided into five groups, and comparison experiments were conducted. The sample selection process was repeated five times during training, using five different training sets generated to reduce the bias of small sample data due to randomness. Table 1 shows the results of the comparison of the accuracy of different models, comparing with machine learning algorithms (SVM), neural networks (BP), and convolutional networks (CNN), the accuracy of deep learning Bi-LSTM is more prominent, and has a greater advantage in fault diagnosis. Taking the training set of 500 samples as an example, it has an accuracy rate of 92.14%, which is far more than other models.

Table 1: Comparison of accuracy of different models

Model Name	100 samples	200 samples	300 samples	400 samples	500 samples
Bi-LSTM	94.51	94.47	94.36	93.25	92.14
SVM	87.85	85.47	84.32	83.38	82.26
BP	88.54	86.48	86.21	82.47	80.65
CNN	86.25	86.14	85.40	84.23	83.31

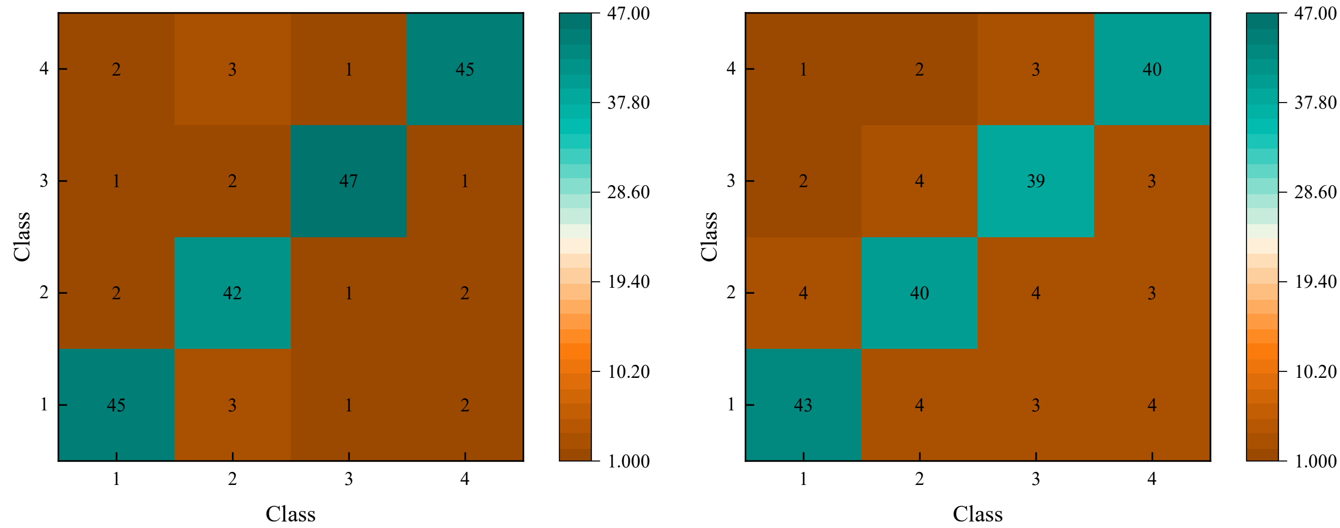
Performance testing of software or hardware platforms is essential when analyzing hydropower plant equipment failures, and is important to ensure the continued development of the housing power system. Performance testing is a continuous task throughout the fault analysis cycle. JMeter, an open source Java-based performance testing tool, is utilized to evaluate the performance and stability of the model. Table 2 shows the results of the performance test. The average response time obtained from the platform test is 1800ms, and the median response time is 1910ms. It can be seen that the response rate of Bi-LSTM is quite stable during the operation of the hydropower plant equipment. 90% of the requests have a response time of no more than 2506ms, while 99% of the requests are completed within 6413ms. It can be seen that the response time of the majority of requests is acceptable, only a few requests have long response time, and the study of such abnormal requests is helpful to find out the improvement measures. In addition, the throughput is 99.8 requests per second with 325.65 receive KB/sec and 46.20 transmit KB/sec, which indicates that the Bi-LSTM has a stable processing efficiency for the housing power system, and it can select appropriate solutions according to the different nature of the equipment faults.

Table 2: Platform test results

Performance indicators	Index value	Performance indicators	Index value	Performance indicators	Index value
Number of samples	1000	90% response time percentile	2506ms	Data receiving rate	325.65KB/sec
Average response time	1800ms	99% response time percentile	6413ms	Data sending rate	16.20 KB/sec
Median response time	1910ms	Minimum/maximum response time	15 / 6026ms	Throughput (processing rate)	99.88 / sec

III. C. Troubleshooting

Category 1 is the normal state of hydropower equipment operation, and categories 2-category 4 are bearing wear, voltage instability and signal loss, respectively. Fig. 3 shows the confusion matrix of fault diagnosis results, Fig. 3(a) shows the Bi-LSTM algorithm, which can be visualized that Bi-LSTM has higher prediction accuracy and performs well on fault 2 and fault 3 categories. Fig. 3(b) shows the SVM algorithm, in contrast this model has more classification errors on categories with a higher level of confusion.



(a) Bi LSTM algorithm (b) SVM algorithm

Figure 3: Confusion matrix of fault diagnosis results

IV. Conclusion

In this paper, deep learning techniques are used to explore the ways in which fault analysis and pre-processing of hydroelectric power plant equipment can be applied to the sustainable development of power systems in residential areas. With the design and implementation of Bi-LSTM model with attention mechanism, the accuracy of fault detection as well as the immediate response of the system are improved. Experiments have demonstrated that this model has excellent performance in categorizing the operating conditions of equipment, especially in identifying important fault types such as bearing wear and voltage fluctuation, which significantly outperform the traditional approach. The systematic integration of fault pre-treatment schemes, including dynamic parameter adjustment, load distribution, hierarchical countermeasures, and immediate feedback mechanisms, makes the whole system operation more robust and control more flexible. In terms of performance tests, the model demonstrates good performance in terms of response speed and other criteria, thus proving its engineering utility. In conclusion, the deep learning algorithm proposed in this paper is applied to the monitoring of hydropower plant equipment, which not only makes the fault disposal intelligent, but also provides scientific and technological support for the safe, stable, and low-emission operation of the power system in residential areas, and has a positive impact on the promotion of the construction of smart grids and sustainable energy systems.

Funding

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LTGG24F020004.

References

[1] Ahmad, M. N., Cheng, Q., & Luo, F. (2022). Dynamic linkage between urbanization, electrical power consumption, and suitability analysis using remote sensing and GIS techniques. Photogrammetric Engineering & Remote Sensing, 88(3), 171-179.
[2] Pan, H., & Liu, R. (2024). Mining research on correlation factors of residential electricity stability based on improved FP-growth algorithm. Management of Environmental Quality: An International Journal, 35(3), 547-566.
[3] Betti, A., Crisostomi, E., Paolinelli, G., Piazzi, A., Ruffini, F., & Tucci, M. (2021). Condition monitoring and predictive maintenance methodologies for hydropower plants equipment. Renewable Energy, 171, 246-253.
[4] Aksoy, B. (2021). Estimation of energy produced in hydroelectric power plant industrial automation using deep learning and hybrid machine learning techniques. Electric Power Components and Systems, 49(3), 213-232.
[5] Michalski, M. A., Melani, A. H., da Silva, R. F., de Souza, G. F., & Hamaji, F. H. (2021). Fault detection and diagnosis based on unsupervised machine learning methods: A kaplan turbine case study. Energies, 15(1), 80.
[6] Guo, B., Wang, Y., Pan, W., & Sun, Y. (2023). Fault diagnosis method for hydro-power plants with Bi-LSTM knowledge graph aided by attention scheme. Journal of Vibroengineering, 25(8), 1629-1641.
[7] Bradde, T., Chevalier, S., De Stefano, M., Grivet-Talocia, S., & Daniel, L. (2021). Handling initial conditions in vector fitting for real time modeling of power system dynamics. Energies, 14(9), 2471.
[8] Sahin, M. E., & Ozbay Karakus, M. (2024). Smart hydropower management: utilizing machine learning and deep learning method to enhance dam's energy generation efficiency. Neural Computing and Applications, 36(19), 11195-11211.

[9] Han, F., Zhao, Y., Zeng, J., Zhang, S., & Wu, T. (2024). Uncertain parameters adjustable two-stage robust optimization of a rural housing integrated energy system considering biomass on-site utilization. Energy, 296, 131106.
[10] Belagoune, S., Bali, N., Bakdi, A., Baadji, B., & Atif, K. (2021). Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. Measurement, 177, 109330.
[11] Wu, J., Li, Q., Chen, Q., Peng, G., Wang, J., Fu, Q., & Yang, B. (2022). Evaluation, analysis and diagnosis for HVDC transmission system faults via knowledge graph under new energy systems construction: A critical review. Energies, 15(21), 8031.
[12] Ren, H., Ma, H., Li, H., & Lu, R. (2023). A disturbance observer based intelligent control for nonstrict-feedback nonlinear systems. Science China Technological Sciences, 66(2), 456-467.