International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 2306-2320

https://doi.org/10.70517/ijhsa463193

Automatic labeling of audio information for acoustic diagnosis of mechanical faults based on artificial intelligence algorithms

Bin Lv[1], Zengwei Zhou[2], Yong Xu[2], Peng Luo[1], Feng Yan[1] and Yijian Wang[2,*]
[1] CSSC Jiujiang Jingda Technology Co., Ltd., Jiujiang, Jiangxi, 332000, China
[2] Jiujiang Precision Measurement and Testing Technology Research Institute, Jiujiang, Jiangxi, 332000, China
Corresponding authors: (e-mail: LeonChang0226@163.com).

Abstract In the field of acoustic diagnostic fault audio monitoring of mechanical equipment, the fault audio sample data plays a crucial role in enhancing the operational safety of mechanical equipment. To this end, this paper designs a mechanical fault audio acquisition framework based on a high-resolution A/D digital-to-analog converter, and performs noise reduction on the acquired acoustic audio signals by improving the variable-step-length LMS adaptive filtering algorithm. Then the compression-aware algorithm is utilized to realize the sampling data compression, transmission and recovery of the mechanical fault audio information, and the sampling accuracy of the mechanical fault audio information is improved by the additive random sampling algorithm, which further reduces the power consumption of the sampling of the acoustic audio information of mechanical faults. The transformer equipment fault audio information of a power grid is selected as the research object, the BIOES annotation specification is adopted to automatically annotate the mechanical fault acoustic automatic audio information, and the mechanical fault audio information knowledge base is established. The BiLSTM+CRF and Transformer models are utilized for named entity recognition and entity relationship extraction of mechanical fault diagnosis audio information knowledge respectively. The results show that the audio information acquisition method designed in this paper for acoustic diagnosis of mechanical faults has higher efficiency and lower power consumption, and the F1 values of entity recognition and entity relationship extraction for BiLSTM+CRF and Transformer models reach 95.87% and 86.31%, respectively.

Index Terms variable step size LMS algorithm, compressed perception, additive random sampling, BiLSTM+CRF, mechanical fault diagnosis, Transformer model

I. Introduction

With the increasing complexity, large-scale, high-speed and automation of modern industrial production equipment, condition monitoring and fault diagnosis of mechanical equipment is also receiving more and more attention. In recent decades, with the continuous development and progress of science and technology, especially the rapid development of computer technology, mechanical fault diagnosis technology has formed a more complete emerging comprehensive engineering disciplines, the content involves fault mechanisms, sensors and measurements, data acquisition, signal processing and control, artificial intelligence, expert systems, computer hardware and software, digital communications and other technical fields [1]-[4]. Practice has proved that the research on mechanical fault diagnosis methods and their applications has important social and economic values, mainly in the following aspects: (1) preventing serious accidents and avoiding social hazards [5]; (2) realizing the visual maintenance, reducing the maintenance cost, and obtaining the best economic benefits [6]; and (3) supervising the quality of the maintenance, and improving the safety of equipment operation [7]. Due to the gradual increase in the degree of interrelated coupling between the elements of the mechanical equipment system, the factors affecting the operation of the system also increase, the possibility of mechanical equipment failure or failure is increasing, the failure mechanism of the mechanical system is also becoming more and more complex, and the traditional fault diagnosis technology is facing a new challenge [8], [9].

Mechanical equipment operation, its vibration signal contains a lot of state information, especially when the equipment failure, the vibration characteristics of the signal within the vibration will produce an obvious impact component, the resulting acoustic signal characteristics will be changed [10], [11]. Whether mechanical fault diagnosis can be carried out successfully depends largely on whether it can extract a sufficient number of mechanical state signals from a large number of complex mechanical state signals, and can truly and objectively reflect the diagnosis of the object's working conditions of the information [12], [13]. Acoustic signal diagnosis has

many advantages, such as non-contact measurement, simple testing method, online testing and no attachment influence, which is especially suitable for industrial field condition monitoring and fault diagnosis [14], [15]. However, the actual acoustic field environment is complex and variable, and the signal to be recognized (fault source signal) is often mixed with various interference signals or noise, which cannot be recognized effectively [16], [17]. Therefore, in order to accurately extract the mechanical fault characteristics, these noises need to be suppressed or excluded, and the cause of the fault is subsequently recognized based on the fault characteristics.

In acoustic diagnosis, information labeling is one of the necessary prerequisites for achieving accurate diagnosis. Traditional information labeling is manual, and the efficiency and accuracy of labeling are difficult to guarantee, and automated information labeling is developed. However, as far as the existing automatic labeling methods for the coupled noise and fault features recognition accuracy is not high, and all mechanical failure causes are not instantaneous, some of the faults exist cyclical, correlation, spatio-temporal characteristics, in some of the existing methods of audio segmentation, only the current audio features are considered, further increasing the error of acoustic diagnosis [18]-[21]. And artificial intelligence algorithms provide a solution for this, in the field of fault diagnosis, audio signals can be transformed and audio features can be learned, providing a new direction for the optimization of signal annotation for acoustic diagnosis technology [22], [23].

A single mechanical equipment fault audio feature may have key information lost in the extraction process, which greatly limits the effectiveness of existing equipment fault audio algorithms. In this paper, starting from the acquisition of audio signals for acoustic diagnosis of mechanical faults, a corresponding audio signal acquisition framework is established, and its acquired signals are processed with noise reduction through the improvement of the variable step size LMS algorithm. Then, additive random sampling algorithm is selected for real-time acquisition of mechanical fault acoustic diagnostic audio signals, and compression-aware algorithm is introduced to realize sampling data compression, transmission and recovery. Based on the fault audio data of transformer equipment in a power grid, the fault audio information is automatically labeled and formed into a fault audio information knowledge base through the BIOES labeling specification. BiLSTM and CRF are combined to establish a named entity recognition model for mechanical fault diagnosis audio information, and the Transformer model is introduced to extract the entity relations of mechanical fault diagnosis audio information. For the application effectiveness of the above methods, this paper carries out simulation experiments and quantitative data analysis.

II. Acoustic Diagnosis of Mechanical Failures Audio Information Acquisition

Mechanical equipment is to ensure the normal operation of all types of mechanical power system is one of the important components, and mechanical failure often occurs in the early stage of failure, the degree of failure is light and not easy to be accurately identified. Once not dealt with in time, the fault will expand rapidly, thus causing serious safety operation problems. At this stage, the automatic diagnosis of mechanical equipment acoustic faults has the problem of untimely recognition and poor recognition effect, how to assist in realizing the accurate recognition of mechanical faults through more accurate audio information acquisition results, which has become a research hotspot nowadays.

II. A.Audio Acquisition for Acoustic Diagnosis of Mechanical Failures

II. A. 1) Audio Acquisition of Mechanical Failures

In order to better realize the real-time acquisition of audio data for acoustic diagnosis of mechanical faults, this paper designs an audio acquisition framework for mechanical faults based on a high-resolution A/D digital-to-analog converter, so as to ensure the accuracy of the acquired data and improve the reliability of subsequent fault analysis. The DSC digital signal controller is selected as the CPU processor, which has a high speed main frequency and can realize the real-time acquisition of the operating status data of mechanical equipment, while its powerful signal processing capability can also pre-process the acquired data to improve the efficiency of fault analysis. Figure 1 shows the framework of acoustic audio data acquisition for mechanical faults. Temperature analysis and vibration analysis and other methods are used to monitor the working status of the main shaft, gear box, generator and other mechanical components of the mechanical equipment, which is divided into three parts: sensor signal conditioning module, signal processing module, and data communication transmission module.

The system will sensor (temperature, vibration, rotational speed) collected data, through the signal pre-processing circuit conditioning (amplification, shaping, filtering, etc.) and converted to a suitable ADC acquisition of signals, and then ADC will be converted from these analog signals to digital signals, the CPU control board vibration, temperature, rotational speed converted by the ADC of the digital signals and audio signals with the monitoring center in the CPU control panel cache and Communication. Ethernet and GPRS can be selected as the transmission method to meet the needs of complex and changing field environment.

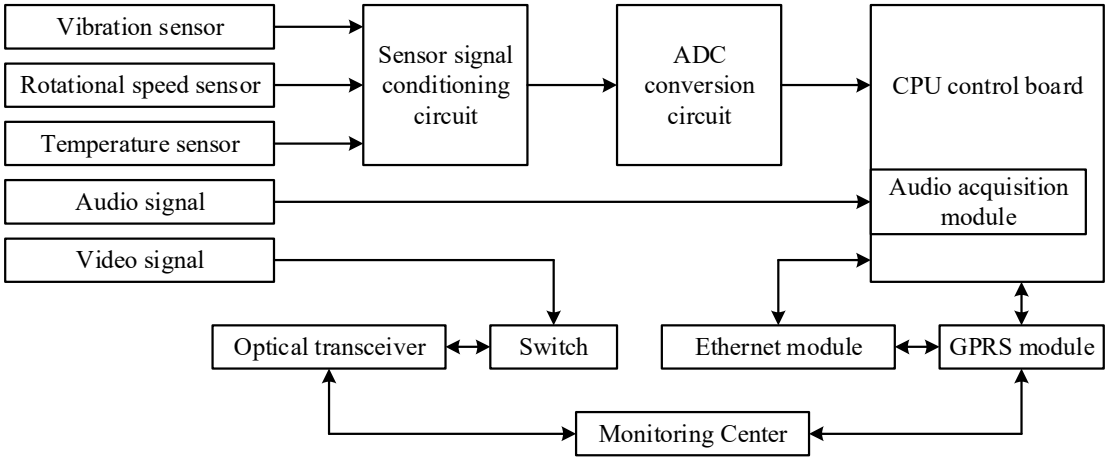


Figure 1: The design diagram of data acquisition system

II. A. 2) Acoustic audio signal noise reduction

In order to better realize the acquisition of acoustic audio signals of mechanical failures, this paper proposes a noise reduction process for the acquired acoustic audio signals based on an improved variable step size least mean square (LMS) adaptive filtering algorithm [24].

In the LMS adaptive filtering algorithm, $x(n)$ denotes the input of the state quantity of the adaptive filter at the current n th moment, $y(n)$ denotes the output of the result of the state quantity at the n th moment after the estimation of the adaptive filter, and $d(n)$ is the desired response at the n th moment. $e(n)$ is the auto-tuning parameter of the filter, the main meaning is the estimation error of the filter at the n th moment, and $w(n)$ is the weighting coefficient of the filter derived by the adaptive algorithm at the n th moment.

The basic formula of LMS algorithm is as follows:

(1) The output result after filter estimation is:

$$y(n) = W^T(n)X(n) = a_{i=1}^{M} w_i(n)x_i(n) \tag{1}$$

(2) The estimation error is:

$$e(n) = d(n) - y(n) \tag{2}$$

(3) The weight coefficients are updated to:

$$w(n+1) = w(n) + me(n)x(n) \tag{3}$$

where $X(n)$ is the state vector of the filter at n moments, $W(n)$ is the weight vector of the adaptive filter at n moments, M is the number of iterations of the adaptive filter, and μ is the step factor for the update of the weight coefficients. Where:

$$X(n) = [x(n), x(n-1), \cdots, x(n-M+1)]^T \tag{4}$$

$$W(n) = [w(n), w(n-1), \cdots, w(n-M+1)]^T \tag{5}$$

The condition for convergence of the algorithm is known from related studies to be satisfied by the step factor μ:

$$0 < \mu < \frac{1}{\gamma_{\max}} \tag{6}$$

where $\gamma_{\max}$ is the maximum eigenvalue of the correlation matrix $R\{E[W(n) - W_0]\}$ and W_0 is the Wiener solution.

In the LMS algorithm, the selection of the step length μ is crucial, if the value of μ is too large, the convergence speed becomes faster, but when the weight coefficients are close to the ideal value, because the step length is fixed, it is possible to miss the ideal value, and the weight coefficients are hovering around the ideal value. And when the step length μ is very small, the algorithm is slow, but when the weight coefficients are close to the ideal value will be close to the ideal value, and will not miss the ideal value.

Based on this, this paper chooses the hyperbolic tangent function $f(x)=1-\frac{2}{1+\exp(2x)}$ as the step factor improvement function, and introduces the parameters α,β to jointly control the change of the step factor curve, and obtains the relationship between the step size and the error as follows:

$$\mu(n)=\alpha\left(1-\frac{2}{\exp(\beta e(n))}\right) \tag{7}$$

Since the step factor should be positive, the function reads:

$$\mu(n)=\alpha\left(1-\frac{2}{\exp(\beta\,|\,e(n)\,|)}\right) \tag{8}$$

In the variable step size LMS algorithm, the value of the step size μ changes with the convergence of the weight coefficients, and when the error is large, the step size μ is larger, when the convergence speed is faster. When the error decreases, the step size μ becomes smaller to ensure the convergence accuracy.

II. B.Acquisition algorithm for low power consumption of audio signals

II. B. 1) Data compression perception algorithm

In order to meet the design requirements of audio data acquisition for acoustic diagnosis of mechanical faults, the acquisition node has to compress the high-flow sampled data in order to ensure real-time transmission. In addition, this practice of increasing the computational volume in exchange for decreasing the communication volume can effectively reduce the energy consumption of the acquisition node and extend the life cycle of the node. Therefore, the core content in the design of audio information acquisition for acoustic diagnosis of mechanical faults is the design of data compression and recovery algorithms. As a result, this paper uses compression-aware technology to realize sampled data compression, transmission and recovery on the basis of existing related research [25].

Consider a finite-length one-dimensional discrete-time signal x over the real number field, x is said to be K -sparse over the basis ψ if there exists some set of sparse bases ψ such that the following equation holds. Namely:

$$x=\psi z \tag{9}$$

where $x\in R^{N\times 1}$, $z\in R^{N\times 1}$, and there are only K non-zero or smaller values in z, $\psi\in R^{N\times N}$, and N is the signal dimension.

On this basis, if a measurement matrix ϕ can be constructed such that the following equation holds. i.e:

$$y=\phi x=\phi\psi z \tag{10}$$

$$\hat{z}=\min\|z\|_0 \quad s.t. \quad y=\phi\psi z \tag{11}$$

An exact or approximate approximation $\hat{x}$ of x can be obtained by solving the optimization problem in the sense of the l_0 paradigm shown in Eq. (11), viz:

$$\hat{x}=\psi\hat{z} \tag{12}$$

where $\varphi\in R^{M\times N}$ the measurement matrix or projection matrix and M is much smaller than N, $y\in R^{M\times l}$ the projection or measurements of x on the measurement matrix ϕ, M the dimensions of the measurements, and $\hat{x},\hat{z},x,z$, The exact or approximate approximation of the

The measurement value dimension M is much smaller than the signal dimension N and the optimization problem (11) is an underdetermined problem, however, by choosing the measurement matrix ϕ and the sparse basis ψ wisely and making them satisfy the finite isometric characteristic criterion or uncorrelated, (11) can be converted into an optimization problem in the sense of the l_1 paradigm:

$$\hat{z}=\min\|z\|_1 \quad s.t. \quad y=\phi\psi z \tag{13}$$

Since problem (13) is a convex optimization problem, it can be conveniently transformed into a linear programming problem, which is solved by using the interior point method, the gradient projection method, the second-order conic programming, and the matched tracing method.

In the design of this paper, the sparse basis ψ selects the fast Fourier transform, the measurement matrix ψ selects the Gaussian random matrix, the signal recovery algorithm selects the interior point method, the signal

dimension N, the dimension of the measurement value M is determined according to the experiment, the calculation of the formulae (9) and (10) is carried out in the acquisition node, and the calculation of the formulae (12) and (13) is completed in the remote computer.

II. B. 2) Audio Signal Sampling Algorithm

Random sampling of audio signals for acoustic diagnosis of mechanical faults only requires adjusting the timing of A/D sampling. If you want to realize random sampling of audio signals on the audio acquisition node, the key is to set the timing of random sampling of signals.

The key to the realization of random sampling is to generate a sequence of random sampling moments to control the action of the sampling device, and the commonly used method is the simple random sampling method. First of all, define the sampling time window, average sampling rate, calculate the number of points to be sampled in each sampling time window, and set up a random number generator to generate random sampling moments. And then in accordance with the time sequence for sorting to get the sampling time sequence, the use of the time sequence to control the hardware sampling device, so that it is in the prescribed moment on the analog audio signal sampling. However, purely random sampling sampling moment generation method is not the most ideal in practical applications, for its shortcomings this paper proposes another sampling moment generation method, that is, additive random sampling algorithm.

The core idea of the additive random sampling algorithm is to ensure the randomness of the sampling moment on the basis of the next sampling moment according to the previous sampling moment for calculation. That is, on the basis of the previous sampling moment plus a reasonable time period, so as to form a new sampling moment, and finally constitute a complete sampling moment sequence [26].

The first step is to calculate the length of the sampling time for a fixed number of points, which can be called the sampling time window. The determination of the sampling time window is based on the sparsity of the audio signal, the signal reconstruction accuracy to be achieved during reconstruction, which can be expressed as:

$$T_w = \frac{n}{f_s} \tag{14}$$

where T_w is the length of the sampling time window, f_s is the sampling frequency that satisfies the Nyquist Sampling Theorem $(f_s > 2f\max)$, and n the number of sampling points. After determining the sampling time window, the sampling moments within the sampling time window can be determined by equations (15) and (16), and the number of sampling moments is the number of sampling points n. That is:

$$\alpha = m \tag{15}$$

$$t_i = t_{i-1} + \tau_i \quad t_0 = 0 \tag{16}$$

$$\tau_i = N\left(\frac{T_w}{\alpha}, \mu^2 \frac{T_w^2}{\alpha^2}\right) + S \times N(0,1) \tag{17}$$

where α is the number of samples, t_i is the i th sampling moment, $i = 1,2,3,\cdots\alpha$, and τ_i is the sampling interval. The average sampling frequency of the samples of additive random sampling can be expressed as:

$$f_r = \frac{\alpha}{T_w} \tag{18}$$

II. B. 3) Audio Signal Acquisition Process

In wireless sensor networks, resource-limited sensor nodes sense the acoustic diagnostic audio signals of mechanical faults and send them to resource-rich aggregation nodes. In order to prolong the system lifetime and improve the service quality, the resource asymmetry between sensor nodes and aggregation nodes must be taken into account in the design of the mechanical fault acoustic diagnosis audio signal acquisition method. Fig. 2 shows the resource-efficient audio signal acquisition algorithm based on compressed perception theory and additive random sampling, which can be realized by programming the sensor nodes and aggregation nodes.

In the signal acquisition algorithm for acoustic diagnosis of mechanical faults, sensor nodes and convergence nodes collaborate to realize audio signal acquisition. Among them, the sensor nodes only need to complete the low-frequency random sampling of audio signals with small computational amount and complexity, while the convergence nodes need to complete the signal reconstruction and sparse transformation with large computational amount and involving high-dimensional matrix operations. This signal acquisition method fully takes into account

the different roles of sensor nodes and aggregation nodes in the system and the difference in the number of resources they possess.

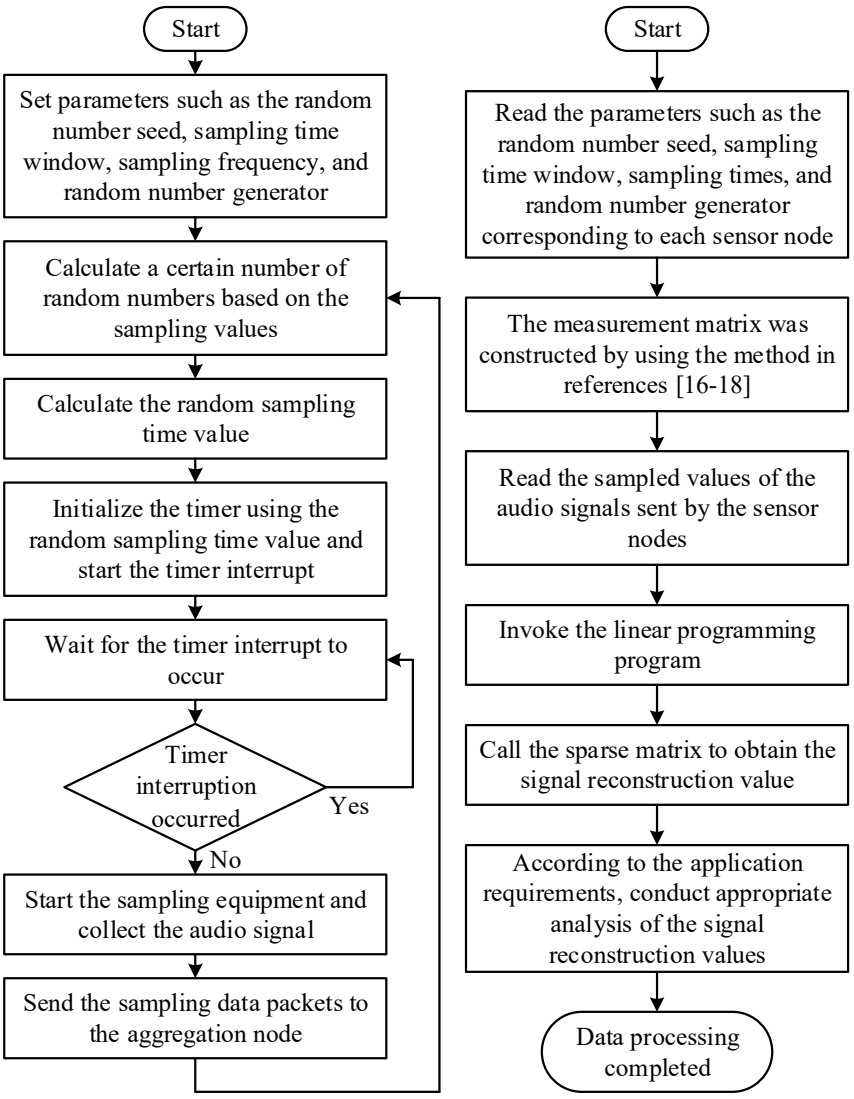


Figure 2: Signal acquisition algorithm process

In order to ensure that the aggregation node can accurately reconstruct the original signal from a small amount of randomly sampled data, the aggregation node and the sensor node must use exactly the same parameters such as the random number generator and the random number seed, the sampling time window, and the number of samples, which can be realized by pre-programming the sensor node and the aggregation node. Among them, to ensure uniqueness, the random number seed can choose the sensor node MAC address, and the random number generator can directly call the related software library function.

II. C.Simulation of Acoustic Audio Acquisition of Mechanical Failure

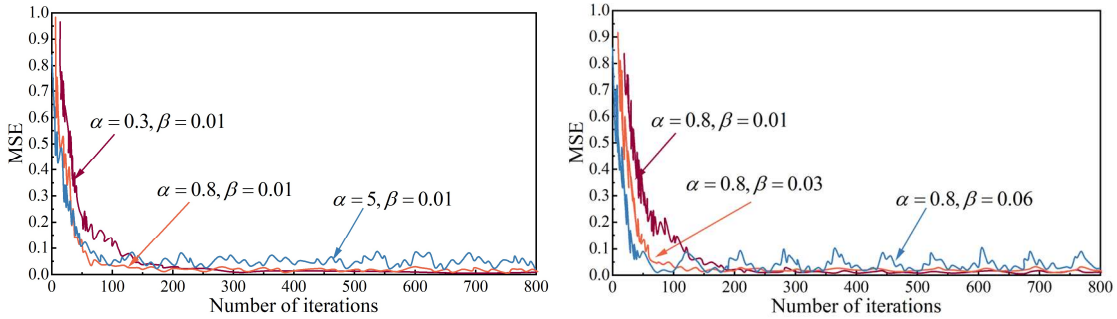
II. C. 1) Filtering performance impact analysis

In order to enhance the accuracy of the acquisition of acoustic diagnostic audio signals for mechanical faults, this paper introduces an improved variable step size LMS algorithm for noise reduction of acoustic diagnostic audio signals. In the improvement of the variable step size LMS algorithm, this paper introduces the parameter α, β to jointly control the change of the step factor curve, and the change of the value of the step size μ will have an

impact on the convergence speed of the algorithm. Based on this, in order to verify the effect of the value of parameter α,β on the filtering performance, this paper sets up a simulation performance analysis.

In the simulation conditions set up in this paper, a sinusoidal signal with an amplitude and frequency of 2 is input as the audio signal for acoustic diagnosis of mechanical faults, and a Gaussian white noise with a mean of 0 and a variance of 1 is superimposed. Second-order linear adaptive prediction is used to estimate the effect of the parameter α,β in the variable-step-size LMS algorithm on the performance of noise filtering of the audio signal. The number of sampling points is set to 800, and 100 independent simulations are done respectively, and then their statistical averages are obtained to get the learning curve. Figure 3 shows the simulation results of the filtering performance, in which Figures 3(a)~(b) show the convergence curves of the algorithm when β is fixed α takes different values and α is fixed β takes different values, respectively.

From the simulation results, it can be seen that the selection of α,β is crucial to the filtering performance. If the parameters are selected appropriately ($\alpha=0.8,\beta=0.01$ in this section), the filtering effect is ideal. On the contrary, if $\alpha=0.3,\beta=0.01$ or $\alpha=0.8,\beta=0.06$, the system will be slow to converge or have a large steady-state error, or even unstable sometimes. Therefore, in practical applications, for different data, the existing methods can be simulated and experimented to obtain the ideal α,β, so as to ensure the practicality of the algorithm. On the whole, after setting a reasonable on, the convergence speed of the algorithm in this paper is significantly improved and the steady state error is smaller. Thus further illustrating the superiority of this paper's algorithm in terms of convergence speed and steady state error.



(a) β fixed α takes different values (b) α fixed β takes different values

Figure 3: Simulation results of filtering performance

II. C. 2) Comparison of signal acquisition algorithms

In the audio information acquisition framework for acoustic diagnosis of mechanical faults established in this paper, the compression perception algorithm is mainly utilized to compress and recover the data, and then combined with the additive random sampling algorithm to establish the audio signal acquisition process. In order to prove the effectiveness of the algorithm in performing mechanical fault acoustic diagnosis audio information acquisition, the FPGA acquisition system and dual ARM acquisition system are selected as the comparison of the algorithm in this paper. Record 20 segments of mechanical failure acoustic diagnostic audio signals with a duration of 15s, and after transmitting them to the convergence node, compare the deviation between the three algorithms and the actual audio signals, and the results are shown in Figure 4. The actual mechanical fault acoustic diagnosis audio signal data is obtained by professional measurement equipment.

By observing Fig. 4, it can be seen that among the three algorithms, the deviation between the value of the acoustic diagnostic audio signal of mechanical failure obtained through the method of this paper and the actual audio signal is the smallest (0.38%), with only a very small deviation, and many times there is a perfect match. On the contrary, the other two methods showed large deviations from the actual audio signal values (2.89% vs. 3.73%). It is concluded that the acquisition process of this paper's method based on a single WSN node downplays the noise characteristics on different routes and reduces the acquisition deviation of audio signals for acoustic diagnosis of mechanical faults. Therefore, the method in this paper has a high mechanical fault acoustic diagnosis audio signal acquisition accuracy, which can ensure a small deviation between the acquisition results and the actual results.

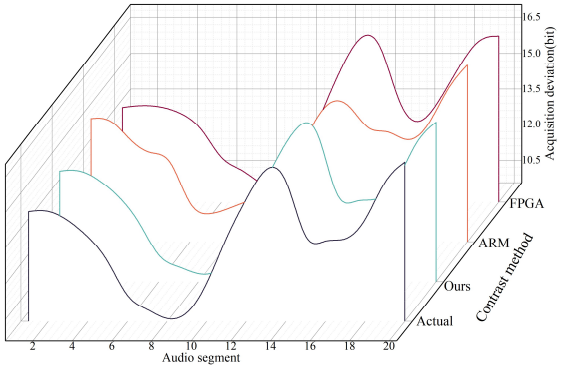


Figure 4: Comparison of Audio signal acquisition deviations

In addition, this paper also selects the packet loss rate as an evaluation index to compare the performance of the three algorithms again. The packet loss rate selected in this paper is defined as the packet loss rate by taking the number of data collected and sent out by the sensor nodes in total per unit time - the number of data successfully sent from the sensor nodes to the aggregation nodes per unit time as the divisor, and selecting the number of data collected and sent out by the sensor nodes in total per unit time as the divisor, and the ratio obtained is defined as the packet loss rate. During the experiment, the number of audio segments is set to 20, and the packet loss rate of multiple audio sensing signal acquisition within 20 min is tested, and the comparison results are shown in Fig. 5.

From the figure, it can be clearly seen that the three algorithms in this paper's method to achieve the lowest packet loss rate, the highest value of the packet loss rate of 1.98%, the dual ARM acquisition system followed by the highest packet loss rate of the FPGA acquisition system, the lowest packet loss rate of the two is 2.89%, compared to this paper's method of packet loss rate is higher. This is due to the method in this paper in the acquisition before the use of improved variable step size LMS algorithm for mechanical fault acoustic diagnosis audio signal noise reduction processing, can effectively mine the audio signal information, remove the audio signal noise, eliminate the audio signal that affects the acquisition effect, improve the probability of effective audio signals are captured at the same time, optimize the acquisition effect, reduce the acquisition packet loss rate.

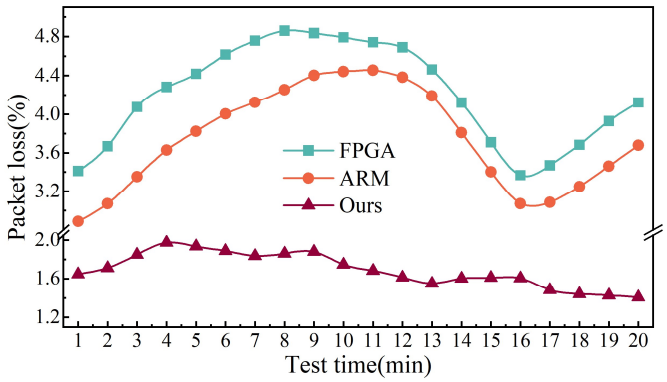


Figure 5: Comparison results of packet loss rates of algorithms

II. C. 3) Algorithmic Sampling Power Analysis

In this paper, the mechanical failure of a power grid transformer is simulated as an experimental object. The audio signals of the 1st, 2nd and 3rd branches have signal-to-noise ratios of 6.13dB, 6.56dB and 7.34dB, respectively, before filtering, and the first two branches are filtered to obtain the signal-to-noise ratios of the audio signals of the 1st and 2nd branches of 10.57dB and 10.84dB, respectively, and the data reachability of the audio signals of the two branches is 85.73% and 79.24%, respectively. The signal-to-noise ratio of the audio signal in the third branch is unchanged, and the data reachability is 64.24%.

In this paper, additive random sampling is utilized for data acquisition of audio signals for acoustic diagnosis of mechanical faults, and in order to verify the acquisition power consumption under this method, heuristic sampling and oversampling are selected as comparisons, and the critical value of data loss rate [0.01,0.1] is used as the horizontal coordinate to obtain the sampling power consumption comparison results of different algorithms as shown in Fig. 6.

As can be seen from the figure, in the case of small data loss rate, compared with heuristic sampling and oversampling algorithms, the power consumption of the audio signal acquisition system for acoustic diagnosis of mechanical faults supported by the additive random sampling algorithm proposed in this paper is about 73.24% and 53.48% of it. This fully demonstrates the feasibility of the additive random sampling algorithm in mechanical fault acoustic diagnostic audio signal acquisition, which can also effectively meet the accuracy and low power consumption requirements of the power grid for mechanical equipment fault diagnostic signals.

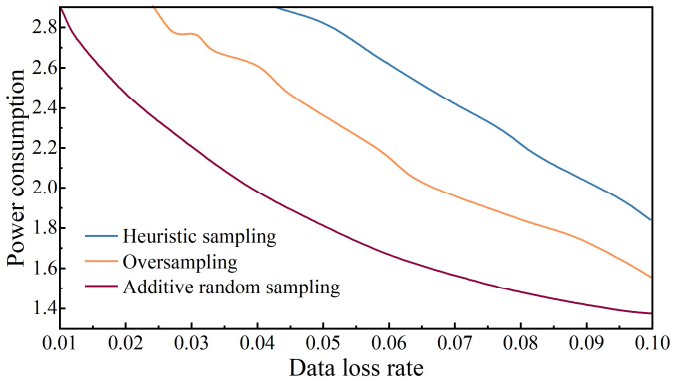


Figure 6: Sampling power consumption comparison results

III. Labeling and identification of audio for mechanical troubleshooting

Due to the complexity and variability of the working environment in which most of the mechanical equipment is located, it is often faced with a variety of faults and abnormalities, resulting in a threat to the safety and stability of the mechanical system. Therefore, mechanical equipment fault diagnosis has been a hot issue in the field of mechanical system research. By fully combining the acoustic audio features generated during the occurrence of mechanical equipment, realizing the conversion of audio to text through automatic annotation methods, and then identifying and extracting its entity relationship by combining with deep learning technology, the accurate identification of mechanical faults can be better realized.

III. A. Data Sources and Labeling of Mechanical Failures

III. A. 1) Sources of data on mechanical failures

This paper takes a power grid transformer equipment as the object of study, the acoustic audio information data of transformer equipment failure, from a power grid intelligent monitoring system and personnel to organize the fault information, the relevant fault text stored in the form of Excel, a more comprehensive record of the mechanical equipment failure occurs in the form of acoustic audio data of detailed information. Among them, the cause of the failure analysis data, in the form of unstructured text describes the failure of the model of the equipment, the cause of the failure, corrective measures and other valuable information.

The valuable value of the mechanical failure knowledge is that it can effectively assist the staff to analyze the cause of the failure, and can give advice on how to deal with the failure. The acoustic fault knowledge of transformer equipment is aimed at extracting fault causes and treatment methods, and the fault knowledge structure is defined according to the summarization of acoustic audio feature information and text data features for fault cause analysis. In the mechanical fault knowledge structure, through the equipment phenomenon caused by the equipment phenomenon of this closed-loop structure, can effectively extract the causal chain of causing accidents, and the rest of the knowledge structure can effectively express the fault treatment measures.

III. A. 2) Labeling of mechanical failure signals

In this project, the collected data is labeled according to the BIOES labeling specification, i.e., the first word of each entity is marked as "B- (entity name)", the middle word is marked as "I- (entity name)", the last word is marked as "E- (entity name)", the single character is marked as "S- (entity name)", and for irrelevant words, all words are marked as O. Table 1 describes the data labeling.

Table 1: Data Annotation Instructions

No.	Entity name	Code	Annotation instructions
1	Parameter	PAR	Various technical parameters in transformer design
2	Index	IND	Transformer performance indicator
3	Structure	STR	The characteristics of the transformer component
4	Factor	FAC	Transformer parts design factors
5	Design	DEN	The design method of the transformer
6	Compute	CMP	The calculation method of transformer
7	Standard	STA	Guide, standard, specification other design constraints
8	Material	MAT	Materials used in transformer parts
9	Craft	CRA	The process of the transformer manufacturing
10	Component	COP	Combination of multiple parts of transformer
11	Element	ELE	The smallest unit of the transformer unit
12	Faults	FAU	Abnormal job of transformer

III. B. Audio Knowledge Extraction for Mechanical Fault Diagnosis

III. B. 1) BiLSTM+CRF-based Named Entity Recognition

Aiming at the characteristics of entity objects in the acoustic diagnostic information text of transformer mechanical faults, the paper adopts the method based on BiLSTM+CRF framework for fault entity word identification and extraction [27]. The model framework is shown in Fig. 7, which mainly includes embedding layer, BiLSTM layer, output layer and CRF layer.

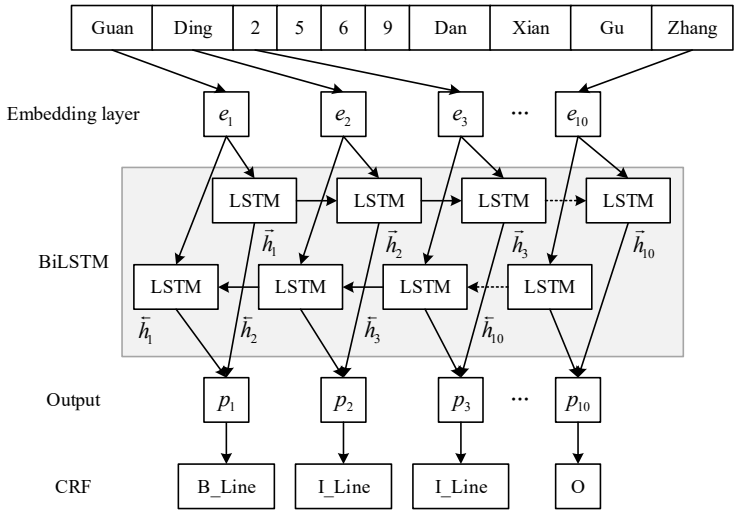


Figure 7: The process of NER based on BiLSTM+CRF

(1) Bidirectional Long-Short-Term Memory Neural Network (BiLSTM) is the application of bidirectional structure on Long-Short-Term Memory Neural Network (LSTM), and the structure of each of its units is the same as that of the conventional LSTM, except that there is an additional hidden layer that processes sequences according to the reverse direction on the whole.

The training method of BiLSTM also uses the backpropagation through time algorithm, and its forward and backpropagation process is similar to that of a conventional RNN. At moment t, the forward computation process of BiLSTM generally follows the following equation:

$$\sigma(x) = 1/(1+e^{-x}) \tag{19}$$

$$\tanh x = (e^{x} - e^{-x})/(e^{x} + e^{-x}) \tag{20}$$

$$\vec{i}(t) = \sigma(W_{ei}e_t + W_{hi}\vec{h}_{t-1} + W_{ci}\vec{c}_{t-1} + \vec{b}_i) \tag{21}$$

$$\vec{f}(t) = \sigma(W_{\sigma f}e_t + W_{bf}\vec{h}_{t-1} + W_{cf}\vec{c}_{t-1} + \vec{b}_f) \tag{22}$$

$$\vec{c}_t = \vec{f}_t \vec{c}_{t-1} + \vec{i}_t \tanh(W_{sc} e_t + W_{bc} \vec{h}_{t-1} + \vec{b}_c) \tag{23}$$

$$\vec{o}(t) = \sigma(W_{co} e_t + W_{ho} \vec{h}_{t-1} + W_{co} \vec{c}_{t-1} + \vec{b}_o) \tag{24}$$

$$\vec{h}_t = \vec{o}_t \tanh \vec{c}_t \tag{25}$$

where e_t is the model input at moment t, $\vec{i}_t, \vec{f}_t, \vec{o}_t$ are the outputs of the forward network input gates, oblivion gates, and output gates at moment t, respectively. $\vec{c}_i, \vec{h}_i$ are the cellular state, the hidden state of the forward network at moment t, $\vec{c}_{t-1}, \vec{h}_{t-1}$ are the cellular state, the hidden state of the forward network at moment $t-1$, respectively, and W is the respective corresponding weight matrix. b is the respective corresponding bias.

The computation of the reverse network is similar to that of the forward network, except that the computation of the hidden state $\overleftarrow{h}_i$ of the reverse network at moment t depends on the hidden state $\overleftarrow{h}_{t+1}$ at moment $t+1$.

Splice the hidden states of the forward and backward network outputs by position to get $h_i = \left[\vec{h}_i \oplus \overleftarrow{h}_i\right]$, and the outputs at each moment constitute the complete sequence of hidden states $\{h_1, h_2, \cdots, h_n\}$, n is the number of input words, and then through linear computation, the hidden state vectors are mapped to the k-dimension, and k is the number of tags in the labeled set, which results in the features extracted by the BiLSTM network automatically, notated as $P = \{p_1, p_2, \cdots, p_n\} \in R_{n \times k}$ (R is the vector space of size $n \times k$), as the final output of BiLSTM.

(2) CRF layer. the CRF is able to obtain a globally optimal sequence of labels by considering the relationship between neighboring labels. For the word vector sequence $S = \{e_1, e_2, \cdots, e_n\}$ of text S, let $\tilde{y} = \{y_1, y_2, \cdots, y_n\}$ be the predicted sequence corresponding to S, where each tag $y_i \in [1, k]$ is the index number of the category in the labeled set, for a total of k^n possible sequence combinations. $\tilde{y}^* = \{y_1^*, y_2^*, \cdots, y_n^*\}$ denotes the labeled sequence corresponding to S.

Define the score of each predicted sequence:

$$s(S, \tilde{y}) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{p_i, y_i} \tag{26}$$

where A is the matrix of $(k+2) \times (k+2)$, plus 2 is added at the beginning and end of the sentence to improve robustness, with start and end states. $A_{y_i, y_{i+1}}$ is the transfer score from category y_t to y_{t+1}, which represents the dependency between the labeled categories of the entity, y_0, y_{n+1} are the start state and the termination state, respectively, and P is BiLSTM's output matrix, and P_{p_i, y_t} is the output vector p_t at the tth moment with the score corresponding to the class y_t.

Define the SoftMax probability score of the labeled sequence $\tilde{y}^*$ in the set of all possible sequences $\tilde{Y}$ to be denoted as:

$$p(\tilde{y}^* \mid S) = e^{s(S, \tilde{y}^*)} / \sum_{\tilde{y} \in \tilde{Y}} e^{s(S, \tilde{y}^*)} \tag{27}$$

The goal of model training is to make $p(\tilde{y}^* \mid S)$ converge to 1. Trained with BiLSTM, it continuously optimizes each parameter matrix of A and BiLSTM.

The trained model predicts the test data to get the best labeled sequence, which is given by the formula:

$$\tilde{y}_m = \arg\max s(S, \tilde{y}) \tag{28}$$

III. B. 2) Transformer-based Entity Relationship Extraction

Transformer is implemented based on the Encoder-Decoder framework, where the left half is the encoder and the right half is the decoder. In the structure of Transformer-based relational extraction network, the feature vectors of multidimensional participle representations are transformed by the Embedding layer and fed into the Encoder network unit.The Encoder network unit consists of 2 sub-layers, the multi-head attention mechanism and the fully-connected position-based feed-forward network, and the outputs of each layer undergoes a normalization computation [28].

The multi-head attention mechanism is an improvement of the original attention mechanism. There are three juxtaposed arrows underneath the Multihead Attention Layer, which represent Value, Key and Query from left to right, and the technique can be represented as:

$$Y = MultiHead(Q, K, V) \tag{29}$$

where Y represents the result of the multi-head attention calculation.

When the input enters the multi-head attention layer, Q, K, and V are transformed by the parameter matrix for mapping the fully connected layer, and then the three results of the transformation are done as a dot product operation, and finally the process is repeated h (the number of heads) times, and the results after calculating the h times are spliced together with the concat() function. The formula is as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{30}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{31}$$

$$MultiHead(Q, K, V) = Concat(head_1, head_2, ..., head_h)W^O \tag{32}$$

After the input is computed by the multi-head attention layer, the result goes to the “Sum and Normalize Layer”. In the Transformer model, the “Sum and Normalize Layer” actually adopts the Residual Connection method, i.e., the outputs of different attention heads are summed up, and then the summed results are normalized by the layer. Through this process, the Transformer model can capture the residuals more easily to extract the key information in the data more effectively, and also improve the generalization ability of the model, so that it can better adapt to different scenarios. After the “sum and normalization layer” processing, it enters the fully-connected neural network layer, where the word embeddings at each position are transformed individually to unify the dimensions of their contexts. Finally, the final output is processed by the Summing and Normalization Layer again.

III. C. Mechanical Fault Diagnosis Audio Recognition Validation

III. C. 1) Comparison of Recognition Model Performance

In this paper, the BiLSTM+CRF model is used to recognize the named entities of the information labeled by the audio signals of mechanical fault diagnosis, with the purpose of helping the operation and maintenance personnel to better grasp the possible fault conditions of the transformer equipment, so as to formulate the relevant overhauling strategies in line with the fault repair needs. In order to verify the recognition performance of BiLSTM+CRF model for mechanical fault audio named entities, BiLSTM, LSTM+CRF, BiGRU+CRF, BERT+LSTM are selected as the comparison models, and the accuracy, recall and F1 value are selected as the evaluation indexes, and the results of the named entity recognition of different models are obtained as shown in Table 2.

The outputs of BiLSTM model are independent of each other, and the labeling bias problem occurs when label prediction is performed, so its recognition effect is relatively poor, and the F1 value of the model is 89.42%.The LSTM+CRF and BERT+LSTM models add a CRF layer and a BERT layer to the LSTM model, respectively, and obtain the labeling constraint rules by learning the labeling constraints rules and utilizing the information of neighboring labels. a globally optimal labeling sequence to alleviate the labeling bias problem.Compared with the BiLSTM model, the model recognition effect is improved with the addition of the CRF layer and the BERT layer, with the F1 values of 91.06% and 91.58%, respectively.The BiGRU+CRF model, on the other hand, is adequately trained with the knowledge information of the audio of the mechanical failure by using the BiGRU model, and combines with the BERT pre-training language model as the embedding layer, so that the model more fully considers the location information and contextual semantic information of the characters, and its recognition effect is improved compared with BiLSTM, LSTM+CRF and BERT+LSTM models, and the F1 value reaches 93.45%.

While the BiLSTM+CRF named entity recognition model designed in this paper utilizes BERT in the embedding layer to provide a deep bi-directional representation of the mechanical failure feature knowledge, and also adds the start state and termination state in the CRF layer, which is conducive to better analyzing the failure signal information and further optimizing the model's various parameter matrices through the dependency relationship between the entity's various label categories, and then obtaining the optimal label sequence of mechanical fault signals. This makes the F1 value of the BiLSTM+CRF model designed in this paper reach 95.87%, which can provide reliable data support for the accurate identification of mechanical failure entities.

III. C. 2) Entity Relationship Extraction Validation

After realizing the named entity recognition of mechanical failure audio signal knowledge, this paper introduces the Transformer model combined with the multi-head attention mechanism for extracting the entity relations of mechanical failure audio signal knowledge, to further enhance the relationship mining capability of mechanical failure audio signal. Based on the entity annotation results of mechanical failure audio signals given in the previous

paper, the specific effect of Transformer model in extracting 12 different types of entity relationships is verified. Table 3 shows the extraction results of the 12 types of entity relations.

Table 2: The recognition results of named entities of different models (%)

Model	Precision	Recall	F1 value
BiLSTM	85.93	93.51	89.42
LSTM+CRF	89.24	92.87	91.06
BERT+LSTM	88.75	94.38	91.58
BiGRU+CRF	91.64	94.75	93.45
BiLSTM+CRF	94.58	96.98	95.87

As can be seen from the table, the average accuracy, recall rate and F1 value of the mechanical fault audio information entity relation extraction method based on Transformer model proposed in this paper are 86.97%, 86.17% and 86.31%, respectively. It is not difficult to find that the entity relationship extraction method adopted in this paper achieves the lowest F1 value of 80.42% on the "standard specification" and the highest F1 value of 93.17% on the "anomaly". Through the analysis of the text of power equipment fault defects and the experimental results, it is known that the standard specifications for transformer machinery are rich and diverse, and some entity pairs with "standard specification" relationship are divided into "ownership relationship", "generic relationship" and "location relationship". Moreover, there are some entity pairs that do not have a "standard specification" relationship that are incorrectly classified into this category, resulting in the worst extraction results on the entity relationship of "standard specification" in the entire transformer mechanical equipment failure audio information knowledge text.

On the one hand, in the transformer power equipment fault audio information text used in this paper, the abnormal audio information of transformer equipment only contains three types: surge, crack and abrasion, and the content is relatively simple, and on the other hand, because there are not many interference terms between the entity of "abnormal" and the transformer equipment name, most of the entity pairs with "abnormal" relationship are correctly classified.

In general, this paper achieves good experimental results in the extraction relationship of 12 types of entities in the knowledge text of acoustic fault audio information of transformer power equipment, which basically meets the requirements of use.

Table 3: The extraction results of 12 types of entity relationships (%)

No.	Code	Precision	Recall	F1 value
1	PAR	85.24	83.18	84.28
2	IND	83.17	83.45	83.64
3	STR	84.05	83.64	83.85
4	FAC	88.04	83.26	85.78
5	DEN	86.72	84.75	84.92
6	CMP	81.94	85.64	83.61
7	STA	84.41	85.96	80.42
8	MAT	91.42	84.57	81.95
9	CRA	93.58	83.02	90.48
10	COP	82.07	90.35	91.24
11	ELE	88.75	91.48	92.35
12	FAU	94.26	92.71	93.17
13	Average	86.97	86.17	86.31

IV. Conclusion

Effective acquisition and annotation of audio information for acoustic diagnosis of mechanical faults is a necessary path to realize accurate identification of mechanical faults. To this end, this paper designs a mechanical fault audio acquisition and a low-power acquisition algorithm for multi-channel mechanical fault audio signals. Then the transformer equipment fault data of a power grid is selected, and its fault audio information is textually labeled using the BIOES labeling specification. BiLSTM+CRF model is introduced for named entity recognition, combined with the Transformer model to extract the entity relations of fault audio information. From the simulation results, the low-

power acquisition algorithm of audio signal for acoustic diagnosis of mechanical failure established in this paper has high data acquisition effectiveness, and the overall power consumption is relatively low, which can meet the requirements of audio information extraction of mechanical failure. In addition, after the audio information of mechanical failure is represented by knowledge annotation, entity recognition and entity relationship extraction of audio information of mechanical failure can be realized by BiLSTM+CRF and Transformer model. This helps the mechanical equipment operation and maintenance personnel to better grasp the mechanical failure problems in order to develop targeted operation and maintenance strategies, which also helps to better protect the safe operation of mechanical equipment.

Although this paper has achieved certain research results, but there are still some shortcomings in the research process. For example, the research sample size of the audio information of mechanical failure is relatively small, and the textual annotation method lacks a certain degree of intelligence. In the subsequent research process, we will further explore the new direction of intelligent labeling of audio information of mechanical failure, and establish a new method of directly converting audio signals into text information.

Funding

This work was supported by Jiangxi Province "Ganpo Juncai" Support Program-Leading Academic and Technological Leaders in Key Disciplines (Project Number: 20232BCJ23064).

References

[1] Liu, C., Cichon, A., Królczyk, G., & Li, Z. (2021). Technology development and commercial applications of industrial fault diagnosis system: a review. The International Journal of Advanced Manufacturing Technology, 1-33.

[2] Fernandes, M., Corchado, J. M., & Marreiros, G. (2022). Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. Applied Intelligence, 52(12), 14246-14280.

[3] Gangsar, P., & Tiwari, R. (2020). Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. Mechanical systems and signal processing, 144, 106908.

[4] Singh, V., Gangsar, P., Porwal, R., & Atulkar, A. (2023). Artificial intelligence application in fault diagnostics of rotating industrial machines: A state-of-the-art review. Journal of Intelligent Manufacturing, 34(3), 931-960.

[5] Peeters, C., Guillaume, P., & Helsen, J. (2018). Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy. Renewable Energy, 116, 74-87.

[6] Kopbayev, A., Khan, F., Yang, M., & Halim, S. Z. (2022). Fault detection and diagnosis to enhance safety in digitalized process system. Computers & Chemical Engineering, 158, 107609.

[7] Wei, Y., Li, Y., Xu, M., & Huang, W. (2019). A review of early fault diagnosis approaches and their applications in rotating machinery. Entropy, 21(4), 409.

[8] Hou, N., Ding, N., Qu, S., Guo, W., Liu, L., Xu, N., ... & Wu, C. M. L. (2022). Failure modes, mechanisms and causes of shafts in mechanical equipment. Engineering failure analysis, 136, 106216.

[9] Sun, X., Wang, M., Zhan, B., Xiong, Y., & Yu, W. (2023). An intelligent diagnostic method for multisource coupling faults of complex mechanical systems. Shock and Vibration, 2023(1), 6928871.

[10] Tang, L., Tian, H., Huang, H., Shi, S., & Ji, Q. (2023). A survey of mechanical fault diagnosis based on audio signal analysis. Measurement, 220, 113294.

[11] Huang, S., Lin, Y., Tang, W., Deng, R., He, Q., Gu, F., & Ball, A. D. (2022). Sensing with sound enhanced acoustic metamaterials for fault diagnosis. Frontiers in Physics, 10, 1027895.

[12] Zhou, Q., Yan, P., Liu, H., & Xin, Y. (2019). A hybrid fault diagnosis method for mechanical components based on ontology and signal analysis. Journal of Intelligent Manufacturing, 30(4), 1693-1715.

[13] Lin, L., Wang, B., Qi, J., Chen, L., & Huang, N. (2019). A novel mechanical fault feature selection and diagnosis approach for high-voltage circuit breakers using features extracted without signal processing. Sensors, 19(2), 288.

[14] AlShorman, O., Alkahatni, F., Masadeh, M., Irfan, M., Glowacz, A., Althobiani, F., ... & Glowacz, W. (2021). Sounds and acoustic emission-based early fault diagnosis of induction motor: A review study. Advances in Mechanical Engineering, 13(2), 1687814021996915.

[15] Wijaya, H., Rajeev, P., Gad, E., & Vivekanantham, R. (2022). Automatic fault detection system for mining conveyor using distributed acoustic sensor. Measurement, 187, 110330.

[16] Yin, H., Xu, H., Zhang, H., Wang, J., & Fang, X. (2025). MC-ABDS: A system for low SNR fault diagnosis in industrial production with intense overlapping and interference. Applied Acoustics, 227, 110217.

[17] Li, Y., Cheng, G., & Liu, C. (2021). Research on bearing fault diagnosis based on spectrum characteristics under strong noise interference. Measurement, 169, 108509.

[18] Yuan, X., Zhang, Z., Liang, P., Zheng, Z., & Zhang, L. (2024). A fusion TFDAN-Based framework for rotating machinery fault diagnosis under noisy labels. Applied Acoustics, 219, 109940.

[19] Li, G., Wu, J., Deng, C., Wei, M., & Xu, X. (2022). Self-supervised learning for intelligent fault diagnosis of rotating machinery with limited labeled data. Applied Acoustics, 191, 108663.

[20] Mühlbauer, M., Würschinger, H., Polzer, D., Ju, S., & Hanenkamp, N. (2020). Automated data labeling and anomaly detection using airborne sound analysis. Procedia CIRP, 93, 1247-1252.

[21] Korkos, P., Linjama, M., Kleemola, J., & Lehtovaara, A. (2022). Data annotation and feature extraction in fault detection in a wind turbine hydraulic pitch system. Renewable Energy, 185, 692-703.

[22] Mofleh, A. F., Shmroukh, A. N., & Ghazaly, N. (2020). Fault detection and classification of spark ignition engine based on acoustic signals and artificial neural network. International Journal of Mechanical and Production Engineering Research and Development, 10(3), 5571-5578.
[23] Bhuiyan, M. R., & Uddin, J. (2023). Deep transfer learning models for industrial fault diagnosis using vibration and acoustic sensors data: A review. Vibration, 6(1), 218-238.
[24] Yan Ma,Hao Zhang & Ke Yang. (2025). A novel adaptive direct signal interference cancellation method for underwater active electromagnetic detection systems using an error signal power-based variable step-size Fast Block Least Mean Square algorithm. Signal Processing,234,109957-109957.
[25] Geethu Joseph,Venkata Gandikota,Ayush Bhandari,Junil Choi,In soo Kim,Gyoseung Lee... & Guoxin Zhang. (2025). Low-resolution compressed sensing and beyond for communications and sensing: Trends and opportunities. Signal Processing,235,110020-110020.
[26] Xie Siyu,Zhang Shujun,Wang Ziming & Gan Die. (2024). Compressed Least Squares Algorithm of Continuous-Time Linear Stochastic Regression Model Using Sampling Data. Journal of Systems Science and Complexity,37(4),1488-1506.
[27] Cai Buqing,Tian Shengwei,Yu Long,Long Jun,Zhou Tiejun & Wang Bo. (2024). ATBBC: Named entity recognition in emergency domains based on joint BERT-BILSTM-CRF adversarial training. Journal of Intelligent & Fuzzy Systems,46(2),4063-4076.
[28] Wanlong Liu,Yichen Xiao,Shaohuan Cheng,Dingyi Zeng,Li Zhou,Weishan Kong... & Wenyu Chen. (2025). Document-level relation extraction with structural encoding and entity-pair-level information interaction. Expert Systems With Applications,268,126099-126099.