**Time Series Prediction based Fault Severity Estimation in Rotating Machinery using Predictive Accuracy**

Abstract

Monitoring the health of rotating machinery plays a pivotal role in ensuring the reliability of industrial processes. Bearings, being significant sources of failure in such equipment, necessitate the implementation of effective fault diagnosis methods. While traditional diagnostic methods involve classifying signals based on different fault types and levels, it's important to note that fault occurrence is a continuous process rather than a discrete one. As a result, the adoption of continuous indicators that accurately reflect fault severity is imperative for timely and precise detection. This paper proposes a novel approach that leverages vibration signals and time series prediction method for fault detection in rotating machinery. Through the extraction of pertinent features from vibration signals and the utilization prediction model, the severity of faults can be effectively quantified. The experimental results reveal a gradual reduction in prediction accuracy with increasing fault severity. By establishing appropriate thresholds, the method enables the early identification of incipient fault indications, facilitating timely machinery shutdown as a preventive measure. Additionally, experiments demonstrate that this method requires only a few shots for training and is applicable to different types of faults, thus exhibiting its applicability as a comprehensive solution for mechanical system fault detection.

*Keywords*: fault severity estimation; LSTM; time series prediction; vibration signal; rotating machinery

**1. Introduction**

Monitoring the health condition of rotating machinery stands as a pivotal endeavor in ensuring the reliability of industrial processes. Among the components of rotating machinery, rolling element bearings emerge as particularly critical. Thus, effective diagnosis of bearing faults assumes a pivotal role in curtailing maintenance expenses with-in manufacturing systems.

Modern industries heavily rely on rotating machinery, which operate under harsh conditions for extended periods, rendering them susceptible to component failures that entail safety risks and economic losses. Consequently, the vigilance over the well-being of rotating machinery becomes indispensable. Notably, bearings, extensively deployed mechanical elements in rotating systems, stand out as significant sources of failures within such equipment. Bearing faults alone can constitute up to 44% of the total fault occurrences in certain devices [1]. Consequently, research in fault detection and diagnosis within rotating machinery predominantly revolves around bearing faults.

When it comes to fault diagnosis, a multitude of parameters can be monitored, such as vibrations, acoustic emissions, currents, flow, speed, pressure, temperature, lubricant conditions, strain, wear, and rotor-to-stator rubbing, among others. Among these parameters, vibration emerges as the widely acknowledged, extensively utilized, and highly effective condition monitoring technique in the realm of rotary machinery. As evidenced by Malla and Panigrahi's research in 2019 [2], vibration-based condition monitoring has demonstrated the ability to detect ap-proximately 90% of machine faults or failures.

Drawing from prior research, techniques for assessing fault severity can be broadly categorized into signal processing-based methods and learning-based methods [3]. Traditional signal processing techniques involve extracting features from signals obtained from various sources linked to fault severity estimation and bearing wear assessment. This is achieved through the analysis of specific frequencies or the computation of indicators that characterize the faults. Esteemed methods in this category encompass the Fourier transform, Hilbert transform [4], wavelet transform [5], empirical mode decomposition[6], and related methodologies.

Since its inception, deep learning has achieved remarkable feats across numerous domains and has found extensive applications in engine fault diagnosis. Deep learning methods for diagnosing faults in rotating machinery can be broadly classified into two categories.

The first category concentrates on detecting fault locations or types, with fault classification as the primary goal. This domain has yielded highly promising results. Noteworthy achievements include CNN-LSTM (Convolutional Neural Network-Long Short-Term Memory) achieving 99.6% accuracy [7], SDAE (Stacked Denoising Autoencoder) attaining 99.83% accuracy [8], and EDAE (Ensemble Deep Autoencoder) reaching 99.15% accuracy [9], among other notable outcomes.

The second category involves computing a degradation indicator that quantifies machinery health, facilitating the assessment of fault severity. Regression models are often employed for this purpose. Shen et al.[10], for instance, utilized an SVR (Support Vector Regression) model to quantitatively estimate fault sizes. Additionally, it's possible to treat varying levels of fault severity as distinct categories, leveraging classifiers to achieve the task. Lei et al. [11] employed WNN (Wavelet Neural Networks) as a severity classifier, with input features selected from the most sensitive Intrinsic Mode Function (IMF) obtained via Empirical Mode Decomposition (EMD). The selection criteria were grounded in the mean and standard deviation of kurtosis values for data samples of each IMF.

This study argues that in real operational contexts, the severity of mechanical faults inherently exhibits continuous variation, spanning from mild to severe. Categorizing faults based on predetermined severity levels becomes impractical when applied to real-world scenarios. However, a significant portion of existing research in this field predominantly employs artificial intelligence methods tailored to classification. Thus, this paper endeavors to introduce an artificial intelligence-driven approach for establishing a degradation indicator for rotating machinery faults, with a particular focus on engine vibration signals. This indicator embodies the universality and context-independence inherent in artificial intelligence, while also maintaining continuity, rendering it suitable for practical applications.

In real operational environments, mechanical faults inherently manifest as a continuous spectrum of severity, ranging from mild to severe. Attempting to categorize these faults into predefined severity levels becomes impractical when applied to real-world scenarios. Despite this, a significant portion of the existing research in this field primarily relies on artificial intelligence methods that are centered around fault classification. Consequently, the objective of this paper is to introduce an artificial intelligence-based approach for establishing a degradation indicator for faults in rotating machinery, with a specific focus on the utilization of engine vibration signals. This indicator combines the universality and robustness inherent in artificial intelligence with the crucial aspect of maintaining continuity, thereby rendering it highly suitable for practical applications.

To attain this goal, the study employs the extraction of time and frequency domain features from the original vibration signals. These features are subsequently utilized in conjunction with an LSTM network to predict sequences of the selected features. In instances involving fault-free signals, the predictions are anticipated to closely align with the actual observations. Conversely, when forecasts deviate inaccurately from the observed values, it indicates the presence of unidentified faults. The degree of this inaccuracy corresponds to the magnitude of the fault. The efficacy of this methodology has been verified across a range of datasets.

This paper follows the subsequent structure: Section 1 introduces the context and current research landscape pertaining to fault detection and estimation in rotating machinery. Section 2 elucidates fundamental concepts, encompassing time and frequency domain features, LSTM networks, and the Dynamic Time Warping algorithm. Within Section 3, we propose a novel fault detection approach. Section 4 offers an account of the experimental validation of the proposed method using simulated vibration signals, coupled with a comprehensive analysis of the outcomes. Finally, Section 5 draws conclusions grounded in the insights garnered from the study's findings.

**2. Theoretical fundamental**

This section offers insights into the theorems employed in the proposed diagnosis approach.

*2.1. Features extraction*

As bearing faults progress, the mechanical system undergoes an increase in vibration intensity, manifesting in the vibration signals. To avert machinery breakdown and ensure optimal production efficiency, the challenge lies in real-time monitoring of defect severity, enabling timely interventions guided by defect trends. Hence, it becomes imperative to extract sensitive features that precisely mirror the current health status of the bearing. This step holds a pivotal role in subsequent procedures.

As demonstrated in [10], [12], a comprehensive array of features was scrutinized for analysis, as detailed in Tables 1 and 2. Table 1 delineates the definitions of ten time-domain features, encompassing root mean square (RMS), square root of amplitude (SRA), kurtosis value (KV), skewness value (SV), peak-to-peak value (PPV), crest factor (CF), impulse factor (IF), margin factor (MF), shape factor (SF), and kurtosis factor (KF). Table 2 expounds on the definitions of three frequency-domain features: frequency center (FC), RMS frequency (RMSF), and root variance frequency (RVF).

Moreover, traditional statistical metrics such as mean, variance, maximum value, and minimum value were also integrated as features, outlined in Table 3.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1** Time domain features | | | |
| Feature | Definition | Feature | Definition |
| RMS |  | CF |  |
| SRA |  | IF |  |
| KV |  | MF |  |
| SV |  | SF |  |
| PPV |  | KF |  |

|  |  |
| --- | --- |
| **Table 3** Statistical domain features | |
| Feature | Definition |
| Mean |  |
| Var |  |
| Std |  |
| Max |  |
| Min |  |

|  |  |
| --- | --- |
| **Table 2** Frequency domain features | |
| Feature | Definition |
| FC |  |
| RMSF |  |
| RVF |  |

*2.2. Dynamic time warping*

Dynamic Time Warping (DTW) [13] is a non-linear warping technique that employs dynamic programming to perform time warping and calculate distance measures. By establishing a matching path between data points within two correlated time series of arbitrary lengths, the algorithm assesses the similarity of these sequences. DTW excels at handling time misalignments between sequences and demonstrates notable fault tolerance and robustness. Hence, it has found application in the feature selection phase of rotating bearing fault detection. Experimental outcomes indicate that DTW can enhance the accuracy of fault diagnosis while minimizing the feature count, thereby enhancing the efficiency and efficacy of fault diagnosis [14].

Assuming that the corresponding lengths of the two time series and are and respectively. The core principle of DTW revolves around finding an optimal alignment path between and . DTW aims to discover an alignment path that minimizes the total distance between corresponding elements in and. To construct , we define a set of conditions: the path starts at the first elements of both sequences , ends at the last elements, and adheres to monotonically increasing indices, where and. The objective is to find the alignment path that satisfies these conditions while minimizing the accumulated cost along the path.

To compute the DTW distance between sequences 𝑋 and 𝑌, the algorithm employs a dynamic programming approach. It constructs a matrix with dimensions , initialized with infinity values. The elements represent the cumulatived cost up to the element in the alignment path. The computation involves pairwise comparison elements in the sequences and updating the matrix based on the following formula:



Upon the completion of the matrix, the DTW distance is calculated as , which signifies the optimal accumulated cost of the alignment path.

*2.3. Long Short-term Memory*

Recurrent Neural Networks (RNNs) have gained extensive popularity in the realm of sequence learning. As depicted in Fig. 1, at a given time , the recurrent hidden layer neurons receive input not just from the input layer at time but also from their own state at time, denoted as.

图示

描述已自动生成

Fig. 1 Architecture of RNN.

As a result, the output is a blend of the present and prior states, and this mechanism can be represented by Equation (1):



Where,

represents the input to the RNN at time ,   
 denotes the state of the hidden layer at time ,   
 represents the state of the neural network at,  
 is the bias of the hidden layer,   
 is the bias of the output layer,   
 denotes the weights between the hidden and input layer,   
 represents the weights of the hidden layer at and the hidden layer at .

Despite encountering the vanishing gradient problem during the training backpropagation phase, which can impede the performance of traditional RNNs, progress has been made to address this limitation and capture long-term dependencies within data features. An exemplar of such progress is the Long Short-Term Memory (LSTM) architecture introduced by Hochreiter and Schmidhuber in 1997 [15]. LSTM has demonstrated superior classification and regression capabilities on time series datasets, including voice and natural language processing datasets, in contrast to conventional RNNs. LSTM incorporates a memory cell that supersedes the role of hidden RNN units. This memory cell encompasses four gates: the forget gate (dictating the extent of information retention from the previous cell state), the input gate (determining whether to incorporate new information into the cell), the input modulation gate (regulating the volume of information to be written into the cell), and the output gate (managing the quantity of information to be extracted from the cell) [16]. LSTMs proficiently maintain error propagation during back-propagation across time and layers.

日程表

描述已自动生成

Fig. 2 LSTM memory cell.

Fig.2 depicts the typical LSTM cell, where (sigmoid) represents the gate activation function, and (tanh) denotes the input or output node activation. The LSTM model presented in Fig.2 is described by Equation (2).



Where,

and are the weights at time between the input and hidden layer,

and are the weights at time and between the hidden layers,

and are the biases of the gates,

is the value of the hidden layer at time ,

and are the output values of the forget gate, input modulation gate, input gate, output gate respectively,

and are the current state at time and respectively.

2.4 Modified Z-score

In the conducted study concerning the application of neural network methodologies for signal analysis, the imperative of outlier identification and rectification was paramount to assure the analytical robustness and integrity of the findings. The adoption of the Modified Z-Score method, as elucidated in [17], provided a rigorous framework for this purpose. Diverging from conventional norms, this method employs the median and median absolute deviation (MAD) as its foundational estimators, circumventing the limitations posed by extreme observations inherent in traditional mean and standard deviation-based approaches. The formulation of the Modified Z-Score is mathematically represented as:

where denotes the median of the dataset, and . A pivotal criterion for identifying an observation as an outlier within this framework is when the absolute value of exceeds , expressed as . This threshold criterion underscores the method's capacity to discern outliers with heightened sensitivity, especially in datasets exhibiting skewness or constrained by size. The election to utilize the Modified Z-Score in the outlier detection process was informed by its empirical robustness and methodological soundness, thereby significantly elevating the credibility and methodological integrity of the investigative outcomes.

**3. Proposed method**

This article evaluates the severity of faults based on the following idea: A normally operating engine has inherent regularity in its vibration signals, which allows for the establishment of a time series prediction model for normal signals. Unknown signals that match the predictions of this model are considered to originate from a machine operating normally; those that do not match the predictions indicate the occurrence of some faults.

The delineation of the proposed method for recognizing bearing fault severity is illustrated in Fig.3.

Fig. 3 Flowchart of proposed method.

We aim to establish a time series prediction model to learn the regularity of normal vibration signals. In principle, this model can be any regression model. Due to the powerful multi-input and multi-output capabilities of LSTM, we choose LSTM as our prediction model in this article. However, the original vibration signals are complex and subject to a lot of noise interference, which makes models that directly learn from the original signals difficult to converge during the training phase. Therefore, we opt to extract features from the original signals and use these feature sequences for training.

To construct the feature time series, the signal is first converted into multiple samples using a sliding window. Let the original signal be represented by , where is a vector of dimension . This is represented as a 2D array of shape in the program. The goal is to transform it into multiple signal segments, with each segment calculating a feature vector.

A sliding window of lengthis used to sequentially extract portions from the original signal. To obtain as many samples as possible, the window's step size is set to . Thus, a maximum of signal segments can be extracted from the original signal, and the number of segments to be extracted is set as . Each signal segment is of length , with the *-*th sample (signal segment) represented as , which in the program is a 2D array of size . All signal segments together form a 3D array of size .

As mentioned in Section 2.1, we consider 18 common features from the field of signal analysis. It's important to note that the feature calculation formulas mentioned in the paper are only applicable to one-dimensional signals. However, real signal collection involves multiple measurement points, with the data from each measurement point constituting a signal channel. Therefore, in practical processing, features should be extracted separately for each channel of the signal. The final feature time series will have dimensions that are the product of the number of original signal channels and the number of features.

For the input signal, after it is converted into multiple signal segments using a sliding window, each channel of each signal segment can calculate features, that is, each signal segment matrix produces a feature vector with dimensions of . Thus, each original signal (N rows by channel columns) is transformed into a feature time series (piece rows by columns).

If all the 18 dimensions of the feature time series are considered, the data dimensionality becomes too high, making the model difficult to converge. Therefore, it is desirable to select a subset of features for use. Since the goal is to reflect the presence of faults through prediction accuracy, the appropriate features should exhibit significant differences between the feature time series generated by normal signals and those generated by fault signals. Assuming we have normal signals and some known "reference fault signals" from faulty machinery, we extract all possible features from both normal and fault signals, calculating the similarity for each corresponding feature one by one. The feature with the smallest similarity is the appropriate feature.

In the selection of similarity, the Dynamic Time Warping (DTW) algorithm is used to calculate the DTW score for each column between two feature matrices. It's noteworthy that for a certain feature (such as the mean), since a sequence will be calculated on each channel (assuming there are three channels, there would be chn1\_Mean, chn2\_Mean, chn3\_Mean), each channel will have a corresponding DTW score. We take the sum of the scores from each channel as the final score for that feature.

By sorting the features in descending order of scores, we obtain the order in which features should be used. Generally speaking, selecting the first feature is sufficient to distinguish between normal and faulty signals.

Assuming only one feature is selected, the next step is to construct a feature sequence sample of normal signals for the sequence prediction model to learn the temporal information of this feature sample. Here, LSTM is used as the sequence prediction model.

A sliding window is applied again to extract feature sequence sample matrices from the feature matrix of normal signals for training the LSTM. The LSTM model aims to predict the values of the next points (feature vectors) based on the previous points (feature vectors) in the sequence. Thus, the length of the sliding window is set to . The sliding step size remains at . The training dataset must contain at least samples, so the length of the feature time series, , must be greater than or equal to .

Each feature sequence sample is a matrix of rows by columns (if two features are selected, then it would be columns), with the first rows by *channel* columns serving as the input and the subsequent rows by *l* columns as the output for that sample. Training is conducted with a total of samples.

After training is completed, the feature sequence samples of normal signals used for training are input into the model, yielding a set of normal MAE, which represents the distribution of MAE generated by normal signals after prediction. The same method is applied to other normal signals, and the MAEs obtained should conform to this distribution. The MAEs generated by fault signals do not conform to this distribution, which means, the MAEs of fault signals are considered outliers comparing to normal MAEs.

The modified Z-score is considered an effective method for detecting outliers, especially when dealing with non-normal data. Experiments indicate that the normal MAEs follow a non-normal distribution, making the modified Z-score suitable for our method. However, our objective is not to identify outliers within a dataset, but rather to determine whether a set of unknown MAEs conforms to the distribution of another set of MAEs. Therefore, we need to adapt the modified Z-score method, calculating a threshold determined by the normal MAEs. By comparing the unknown MAEs to this threshold, we can determine whether they are outliers.

The original definition of the modified Z-score has already been described in Section 2.4. A feasible criterion for being classified as an outlier is , which is equivalent to , where is the unknown MAE to be judged, and is the median of the normal MAEs. Considering that only an excessively large unknown MAE is deemed a fault, it can be asserted that . Therefore, the threshold given by the modified Z-score is:

where is the median of the normal MAEs, and MAD is the median absolute deviation of the normal MAEs, calculated as .

For the unknown signals awaiting testing, feature sequences are extracted in the same way as during the training phase and fed into the time series prediction model to obtain the corresponding MAE. Each fault signal is in fact divided into multiple samples, yielding a set of MAEs. This set of MAEs is compared with the threshold provided by the modified Z-score. If the average of these MAEs exceeds the threshold, or if the proportion of MAEs exceeding the threshold reaches a certain level (such as 50%), then it is asserted that the unknown signal is abnormal. Otherwise, it is considered normal.

**4. Experiment validation**

This section introduces the dataset employed in experiments and validates the efficacy of the proposed method through experimental assessments.

*4.1. Data description*

In this research, we utilized three widely recognized bearing fault datasets to support our experimental analysis. First, the dataset provided by Case Western Reserve University (CWRU) Bearing Data Center focuses on the fault diagnosis of motor bearings, offering rich laboratory-level fault diagnosis data by simulating various types of bearing defects, such as inner race, outer race, and rolling element defects, serving as a vital resource in the field of bearing fault analysis. Second, the bearing fault dataset from the Intelligent Maintenance Systems (IMS) Center at the University of Cincinnati collects extensive data on bearing performance under various operational conditions, widely used in bearing fault prediction and health management research. Lastly, the XJTU-SY dataset, provided by the Institute of Design Science and Basic Component at Xi’an Jiaotong University (XJTU) and the Changxing Sumyoung Technology Co., Ltd. (SY), represents a bearing fault dataset contain complete run-to-failure data of 15 rolling element bearings that were acquired by conducting many accelerated degradation experiments, which is designed to provide more detailed and comprehensive fault mode analysis data, aimed at advancing fault diagnosis technologies.

4.1.1. Case Western Reserve University Bearing Data (CWRU)

The data used in this research work come from two bearings installed in a motor driven mechanical system, one at the drive end of the motor and the other at the fan end. In both bearings three types of faults (outer race, inner race and ball faults) were introduced using electro-discharge machining with various fault diameters. For the case of the outer race faults, experiments were conducted for both fan and drive end bearings with outer raceway faults located at 3 o’clock (directly in the load zone), at 6 o’clock (orthogonal to the load zone), and at 12 o’clock. Each bearing was tested under four different loads, 0, 1, 2, and 3 hp.

After collecting data with a 16-channel DAT recorder, it is processed in the MATLAB environment, and all data files are saved in the MATLAB (.mat) format. Each file includes one or more recordings of DE (Drive-end), FE (Fan-end), and BA (Normal-baseline) acceleration data, collected at sampling frequencies of 12kHz and 48kHz, respectively. In our experiments, only DE and FE data are used, treating them as two channels of a test dataset.

A more detailed description of the experimental set-up and the apparatus involved can be found in the Case Western Reserve University’s website [18].

4.1.2 Dataset of Intelligent Maintenance Systems (IMS), University of Cincinnati.

The IMS bearing dataset has been collected on an endurance test rig of the University of Cincinnati and released in 2014[19]. The test rig (shown in Fig.4) has the following characteristics:

• 4 double row bearings of type Rexnord ZA-2115,

• 2000 rpm stationary speed,

• 6000 lbs load applied onto the shaft and bearing by a spring mechanism,

• PCB 253B33 High sensitivity Quart ICP® accelerometers.

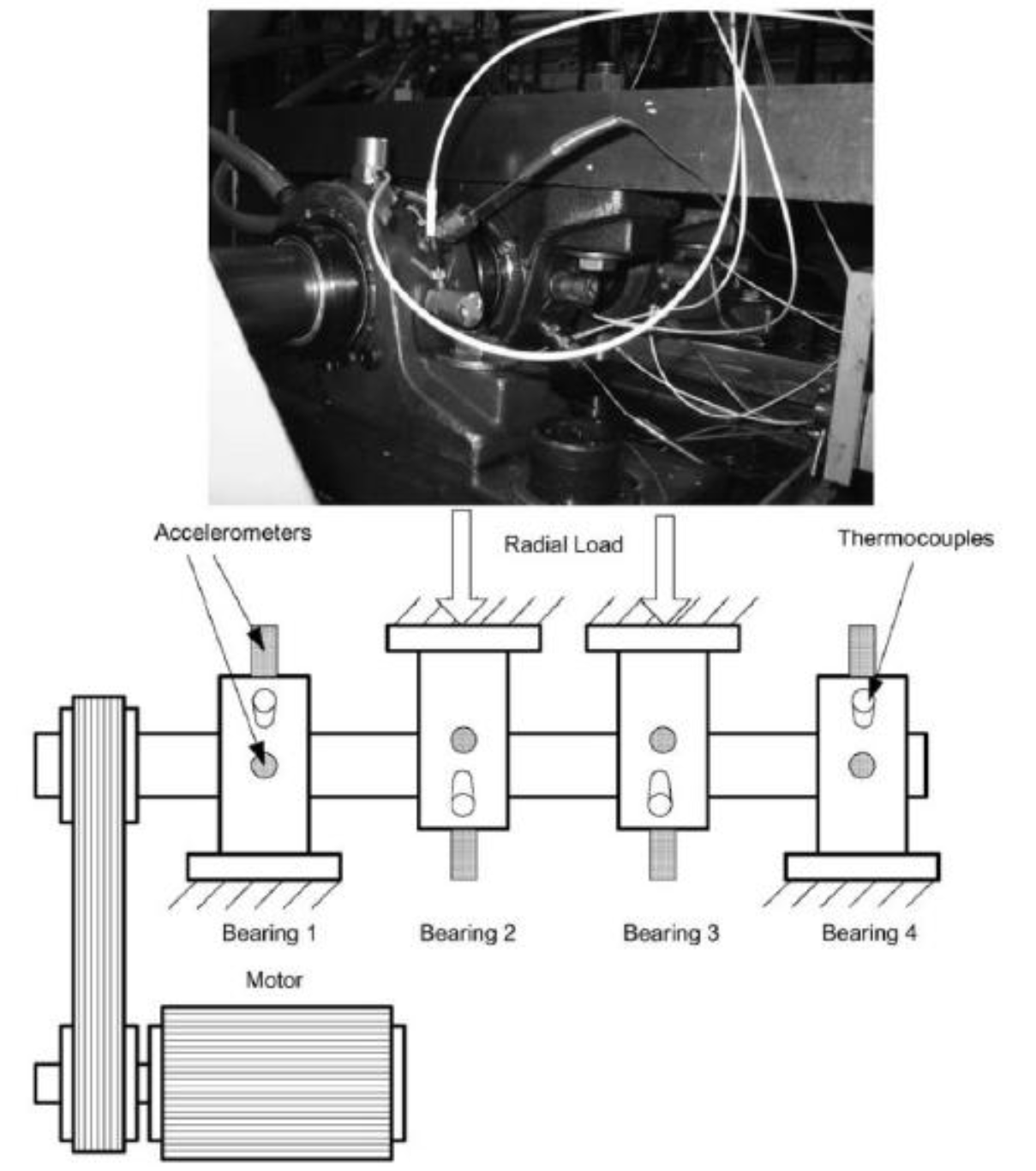


Fig. 4 Bearing test rig and sensor placement illustration [19]

An AC motor, coupled by a rub belt, keeps the rotation speed constant. The four bearings are in the same shaft and are forced lubricated by a circulation system that regulates the flow and the temperature. It is announced on the provided “Readme Document for IMS Bearing Data” in the downloaded file, that the test was stopped when the accumulation of debris on a magnetic plug exceeded a certain level indicating the possibility of an impending failure.

Three datasets are provided on the downloaded file, composed of numerous files of one second each. Each file is made of 20,480 samples. Although it is mentioned that the sampling frequency is 20 kHz, it is thus believed that it was actually 20.48 kHz (and the spectral analyses presented in this paper all seem to support this assumption, since more consistent results are then obtained with respect to expected fault frequencies). A one second acquisition has been made every 5 (for the first 54 acquisitions) or 10 min, but it is sometimes subjected to a series of interruptions that make the time history not continuous (Figure 5).

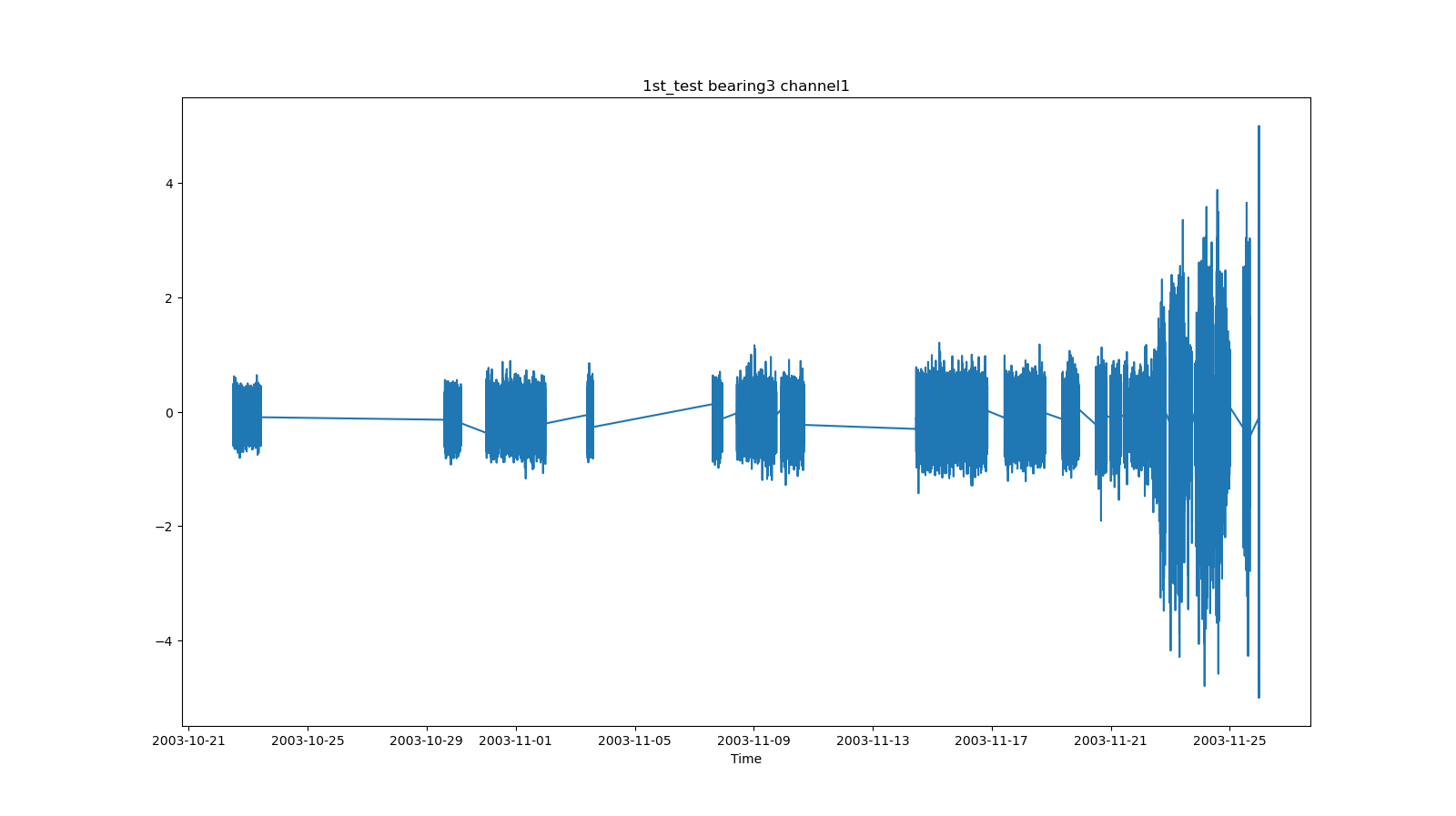


Fig. 5 Raw signal and its actual time history for IMS dataset 1, the first channel of bearing3

The experiment was stopped conventionally when the accumulation of debris on a magnetic plug exceeded a certain level, indicating the possibility of an impending failure. The resulting endurance duration was equal to 49,680 min (i.e., 34 days and 12 h), exceeding the bearing designed lifetime, which was more than 100 million revolutions. It should be stressed that this dataset is particularly valuable because the bearing degradation was left to evolve naturally and was not artificially induced, as may often happen to speed up the experimental test.

4.1.3. Xi'an Jiaotong University Bearing Fault Dataset (XJTU-SY)

The bearing dataset from XJTU-SY has been made public which could be download at http://biaowang.tech/xjtu-sy-bearing-datasets. And the outer race fault data was used to demonstrate the efficacy of the suggested method in this fault diagnosis[20].

As shown in the following figure, the bearing testbed is composed of an alternating current (AC) induction motor, a motor speed controller, a support shaft, two support bearings (heavy duty roller bearings), a hydraulic loading system and so on. This testbed is designed to conduct the accelerated degradation tests of rolling element bearings under different operating conditions (i.e., different radial force and rotating speed). The radial force is generated by the hydraulic loading system and applied to the housing of tested bearings, and the rotating speed is set and kept by the speed controller of the AC induction motor.

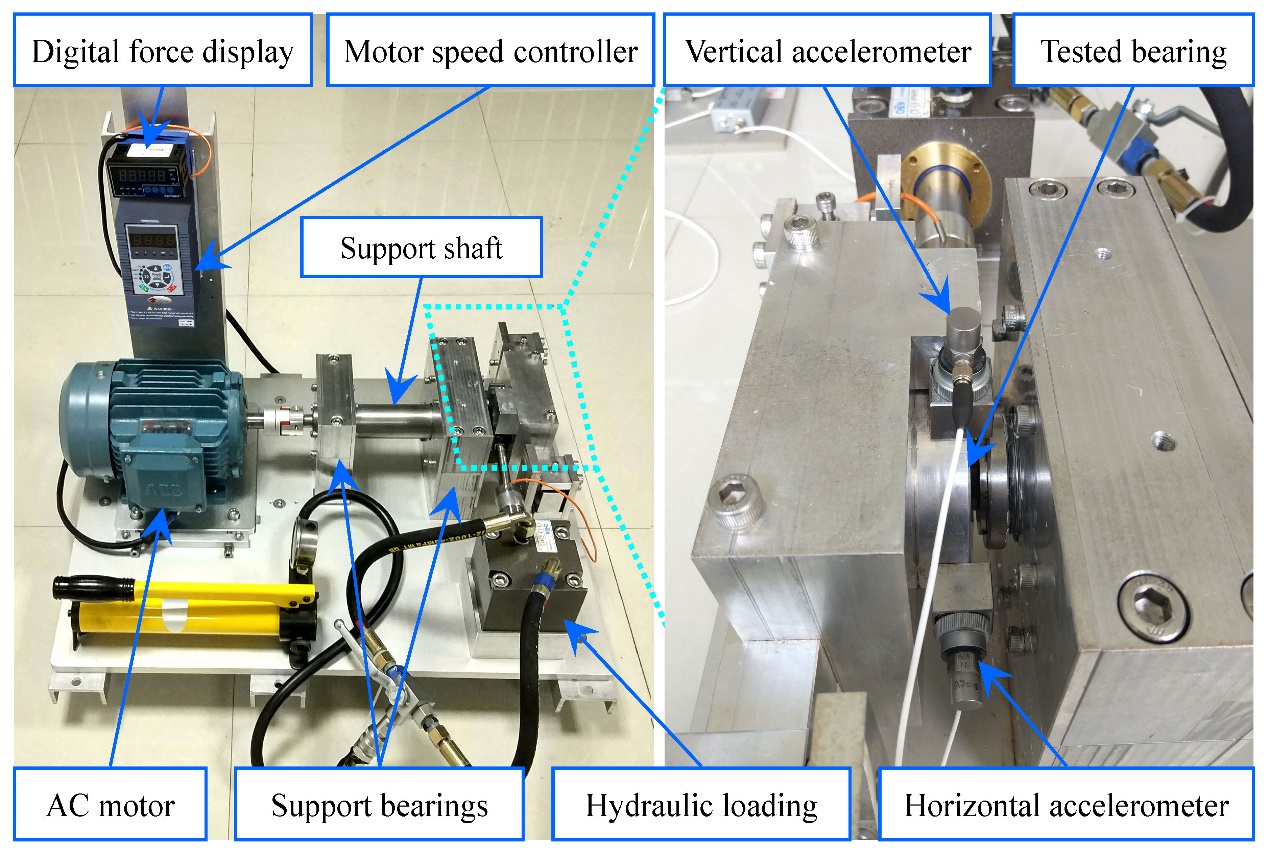


Fig. 6 Bearing Accelerated Life Test Bench

The dataset employed in this study utilized LDK UER204 rolling bearings as the test subjects. The relevant parameters of these bearings can be found in Table 4. To comprehensively evaluate their performance, experiments were structured around three distinct operational conditions, as outlined in Table 5. Across each operational condition, multiple sets of vibration signals were meticulously collected throughout the complete lifecycle of the bearings. Throughout the experiments, two PCB 352C33 unidirectional acceleration sensors were affixed to the test bearings using magnetic mounts, ensuring secure attachment in both horizontal and vertical orientations. A DT9837 portable dynamic signal acquisition system was employed to record the vibration signals. The sampling frequency was meticulously set to 25.6 kHz, accompanied by a 1-minute sampling interval. Each individual sampling session spanned a duration of 1.28 seconds. For a detailed interpretation of the dataset, please refer [21].

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4** LDK UER204 Bearing parameters | | | |
| Parameter | Value | Parameter | Value |
| Inner race diameter(mm) | 29.30 | Ball diameter(mm) | 7.92 |
| Outer race diameter(mm) | 39.80 | Number of balls | 8 |
| Bearing median diameter(mm) | 34.55 | Contact angle(°) | 0 |
| Basic rated dynamic load(N) | 12820 | Basic rated static load(N) | 6.65 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 5** Bearing Accelerated Life Test Conditions | | | |
| Number | 1 | 2 | 3 |
| Rotational Speed | 2100 | 2250 | 2400 |
| Radial Force | 12 | 11 | 10 |

To obtain a complete time series, the data obtained from each sampling is concatenated, resulting in run-to-failure vibration signals. The first two signals are shown in Fig.7. The label Bearing 1\_2 means the data is the second test in work condition 1, and so on.

|  |  |
| --- | --- |
| 图表  描述已自动生成 | 图表  描述已自动生成 |
| Fig. 7 The run-to-failure signals for the first two experiments under test condition 1. | |

*4.2. Experiment and result*

4.2.1 CWRU dataset

This dataset provides signals under different operating conditions and fault situations, with each signal measured at a constant fault level. To verify the effectiveness of the method proposed in this paper in diagnosing faults, normal signal with a motor load of 0 (Approx. Motor Speed 1797 rpm) numbered as 97 and signals in different fault situations from the drive end with the same motor load and a sampling frequency of 48k were employed, as listed in Table 6.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 6** 48k Drive End Bearing Fault Data | | | | | |
| Fault Diameter | Inner Race | Ball | Outer Race (Position 6:00) | Outer Race (Position 9:00) | Outer Race (Position 12:00) |
| 0.007" | 109 | 122 | 135 | 148 | 161 |
| 0.014" | \* | 189 | 201 | \* | \* |
| 0.021" | 213 | 226 | 238 | 250 | 262 |

The important hyper parameter settings in the algorithm are shown in Table 7:

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 7** Hyper Parameter Setting | | | |
| Parameter | Value | Parameter | Value |
|  | 2048 |  | 100 |
|  | 22000 |  | 10 |
|  | 110 |  | 200 |

Taking signal 97 as the normal signal and 226 as the reference fault signal, feature sequences are extracted separately. The DTW scores for each column of the feature sequences are calculated and arranged in descending order. The higher the ranking, the greater the difference between the feature in the normal and fault signals, making it suitable for subsequent processing. It is important to note that these scores are only reference values, and the determination of the actual features to be used should also consider other factors, such as the convergence of the prediction model

The top 10 ranked features are shown in Table 8. Ultimately, the mean value *Mean* would be selected as the feature used in the prediction model. Since 226 represents the highest level of ball failure, this score ranking implies that Mean is suitable for use in this method to detect various levels of ball failure.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 8** DTW scores of different features | | | |
| Feature | DTW score | Feature | DTW score |
| Mean | 29.77 | Var | 17.63 |
| RVF | 19.87 | KF | 17.51 |
| RMSF | 19.70 | Min | 17.31 |
| SRA | 18.11 | FC | 17.09 |
| RMS | 17.85 | MF | 16.99 |

Use the feature *Mean* sequence extracted from normal signals to train an LSTM as a time series prediction model. After obtaining the model, for each type of fault signal, extract the Mean feature sequence as well, input it into the LSTM to obtain the corresponding MAE (Mean Absolute Error), and compare it with the MAE distribution of normal signals and the calculated threshold, drawing the comparison into histogram as Figure 8.

|  |
| --- |
|  |
|  |
|  |

Fig. 8 Ball failure signals test MAE distribution histograms

In the graph, the blue and green bins represent the distributions of MAE (Mean Absolute Error) for normal and test signals, as determined by the algorithm, respectively. The red dashed line indicates the threshold calculated based on the MAE of normal signals, with MAEs exceeding this threshold considered anomalies, suggesting that the signals from which these MAEs arise are abnormal compared to normal signals. The black dashed line represents the average MAE of the test signals, with values above the threshold serving as a criterion for asserting a fault.

From the graph, it is evident that the MAE distribution of signals from any fault level is generally much higher than that of normal signals and naturally exceeds the threshold set based on the Z-score. Although only the highest fault level signal 226 was used during training, signals 122 and 189 are completely unknown to the algorithm, yet it can still accurately detect faults.

The same measures were applied to inner race faults and outer race faults in three different directions, and the results are presented in Table 9.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 9** CWRU failure signals test result | | | | |
| Fault type | Inner Race | Outer Race (Position 6:00) | Outer Race (Position 9:00) | Outer Race (Position 12:00) |
| Reference fault signal | 213 | 238 | 250 | 262 |
| Best feature | KV | Min | Mean | Mean |
| DTW score of the best feature | 17.67 | 18.50 | 17.07 | 16.11 |
| MAE histogram | 109    213 | 135    201    238 | 148    250 | 161    262 |

It is easy to observe that under each fault type, there is a significant difference in MAE between normal and faulty signals. Normal and fault signals can be clearly distinguished. The only exception is signal number 250, which has 6% of its MAE below the threshold. However, considering that the average MAE is above the threshold, it can still be asserted that the algorithm has identified the fault in this signal.

4.2.2 IMS dataset

IMS is a full lifecycle experimental dataset, with signals collected from machinery operating normally until it stops due to a fault, during which the fault level gradually increases. For signals from a full lifecycle experiment, it is not possible to conduct a binary diagnosis of normal or fault as with the CWRU data. Instead, we hope to calculate the critical point at which the mechanical state deteriorates from normal to fault. The guidance significance is that, in real industrial processes, machinery should be shut down for maintenance upon reaching this critical point to prevent more severe damage.

During the IMS dataset experiment, multiple signal samples are taken in one test, obtaining several segments of vibration signals. The aforementioned algorithm is applied to each segment of signal in turn, resulting in a corresponding set of MAEs. If the proportion of MAEs exceeding the threshold reaches a certain level, for example, if 50% of the sample MAEs are above the threshold, then the segment of signal is considered abnormal. The collection time of this segment of signal is then identified as the critical point of fault occurrence.

To demonstrate the reliability of our algorithm, our experimental results will be compared with existing studies on the IMS dataset.

Sacerdoti and colleagues' research[22] utilized the first test data from the IMS dataset to evaluate various signal processing techniques for detecting and locating rolling element bearing faults. By comparing the Kurtosis values of diagnostic techniques, they identified that Cepstrum Pre-Whitening (CPW) and Improved Envelope Spectrum (IES) performed best in localizing and identifying faults. Fault detection focused on bearings 3 and 4, with an inner race fault in bearing 3 located between the 33rd and 34th days of the experiment, while rolling element and outer race faults in bearing 4 were detected starting from the 26th day.

Gousseau et al.[23] utilized the dataset provided by the Center for Intelligent Maintenance Systems (IMS) at the University of Cincinnati, employing a comprehensive suite of signal processing techniques such as time-domain analysis, spectral analysis, blind deconvolution, spectral coherence, and envelope spectrum for in-depth diagnostics and prognosis of vibrations in rolling element bearings. Their analysis successfully diagnosed inner race damage in bearing 3 and ball damage in bearing 4 within dataset 1, as well as outer race damage in bearing 1 in dataset 2, while dataset 3 failed to reveal the anticipated fault characteristics. For dataset 1's bearing 3, time-frequency analysis detected damage from 33.8 days, envelope spectrum analysis identified it slightly earlier on day 32, and spectral coherence analysis earliest on day 29.2. For dataset 1's bearing 4, time-frequency analysis detected damage on day 18, with spectral coherence analysis confirming its presence on day 23. In Test 2, all methods (time-frequency analysis, envelope spectrum, spectral coherence) accurately diagnosed outer race damage in bearing 1 starting from day 3.5.

To compare results with the two aforementioned literature, our experiment employs only Dataset 1, excluding bearings 1 and 2 which exhibited no faults, and focuses solely on the data from four channels corresponding to bearings 3 and 4 (as what [22] done). Considering that the machine is certainly normal at the start of operation and that the fault level increases during operation, the signals obtained from the initial few samplings are regarded as normal signals, and the signal from the last sampling is used as a reference fault signal for validation. Given the long duration of the experiment and the large volume of data, considering only the first normal signal could lead to overfitting. Therefore, during feature selection, the first four signal segments are all considered as normal signals, with DTW scores calculated and summed for each. The hyperparameter is set to 100, with other settings consistent with Section 4.2.1 of the CWRU data experiment.

The experiment results is shown as Table 10.

|  |  |  |
| --- | --- | --- |
| **Table 10** IMS dataset 1st\_test result | | |
| Bearing | 3 | 4 |
| Best feature | SV | MF |
| DTW score of the best feature | 7.16 | 6.16 |
| MAE graph |  |  |
| Faul emerge time | 11/22 23:56 | 11/14 15:32 |

When more than 50% of the MAE calculated for a segment of signal exceeds the threshold based on the Z-score, it is considered that a fault has occurred. Using the method proposed in this paper, the fault occurrence date for bearing 3 was detected to be November 22, equivalent to the 32nd day after the start of the experiment; the fault occurrence date for bearing 4 was detected to be November 14, equivalent to the 24th day after the start of the experiment. These results are close to those given in literature [22] and [23], demonstrating that the method proposed in this paper can indeed effectively detect faults.

4.2.3 XJTU-SY dataset

XJTU-SY is another open-source dataset that encompasses the entire lifespan of experimental data. The experimental methodology applied to the XJTU-SY dataset is entirely identical to that used for the IMS dataset. Consider follows experiments:

(1) Utilizing the data from the first three experiments under operating condition 1, and given that these experiments exhibited the same type of fault, the data from the third experiment (Bearing1\_3) is used for feature selection, while the other two experiments are employed for testing.

After calculation, the most suitable feature was found to be *Min*, with a DTW score of 7.98. The experimental results for Bearing1\_1 and Bearing1\_2 are shown in Table 11.

|  |  |  |
| --- | --- | --- |
| Table 11 XJTU-SY dataset Bearing1 result | | |
| Data | Bearing1\_1 | Bearing1\_2 |
| MAE graph |  |  |
| Faul emerge time | 9th | 14th |

From the graph, it can be seen that Bearing1\_1 exhibited abnormalities as early as the 9th sampling, with the severity of the abnormalities fluctuating initially. By the 73rd sampling, the proportion of abnormal MAE reached 1, at which point it can be considered a complete failure. Bearing1\_2 showed abnormalities by the 14th sampling and quickly progressed to complete failure.

(2) Using the data from three outer race fault experiments under operating condition 2, namely Bearing2\_2, Bearing2\_4, and Bearing2\_5, the feature selection was performed using Bearing2\_5, while the other two were used for testing.

After calculation, the most suitable feature was found to be Meanf, with a DTW score of 8.33. The experimental results for Bearing2\_2 and Bearing2\_4 are shown in Table 12.

|  |  |  |
| --- | --- | --- |
| **Table 12** XJTU-SY dataset Bearing2 result | | |
| Data | Bearing2\_2 | Bearing2\_4 |
| MAE graph |  |  |
| Faul emerge time | 48th | 15th |

From the graph, it can be seen that Bearing2\_2 exhibited abnormalities by the 48th sampling, while Bearing2\_4 showed abnormalities by the 15th sampling and quickly progressed to complete failure.

(3) Using Bearing1\_3 (with an outer race fault) for feature selection and Bearing1\_5 (with both inner and outer race faults) for testing; and using Bearing3\_3 (with an inner race fault) for feature selection, with Bearing3\_2 (with inner race, rolling element, cage, and outer race faults) for testing. This experiment aims to demonstrate that the algorithm proposed in this paper remains effective for data with mixed fault types.

The selected features and the results of anomaly detection on the test data are shown in Table 13.

|  |  |  |
| --- | --- | --- |
| **Table 13** XJTU-SY dataset Bearing1\_5 and Bearing3\_2 result | | |
| Data | Bearing1\_5 | Bearing3\_2 |
| Best feature | Min | PPV |
| DTW score of the best feature | 7.98 | 8.62 |
| MAE graph |  |  |
| Faul emerge time | 34th | 1961st |

Bearing1\_5 is considered to have developed a fault at the 34th sampling; Bearing3\_2 first showed an abnormal MAE ratio exceeding 50% at the 497th sampling, which, from the perspective of the entire experiment, could be seen as a false alarm. This is because the abnormal MAE ratio remained relatively low for a long period after the 1120th sampling, not rising back to 100% until the 1961st sampling.

*4.3. Results analysis*

Validation on three different datasets demonstrates that the algorithm proposed in this paper can effectively detect anomalies in vibration signals, and it has the following advantages:

I) Experiments on the CWRU dataset show that the method can accomplish fault detection tasks with a small number of samples. Despite only using fault data with a diameter of 0.021 inches for training, it can detect faults with diameters of 0.007 and 0.014 inches. This characteristic is especially suitable for real industrial environments. In many practical application scenarios, acquiring a large amount of labeled data is often costly and time-consuming. Therefore, an algorithm that can effectively utilize a limited number of samples for accurate detection has significant practical value. Moreover, this method is versatile, capable of detecting a variety of faults, including inner race faults, outer race faults, and rolling element faults, all using the same approach.

II) Experiments on the IMS dataset demonstrate that this method can be used for the health monitoring of rotating machinery, providing timely warnings when the machine exhibits abnormalities. Compared to existing work, our method does not use complex signal analysis techniques, nor does it require manual identification of characteristic lines in spectrograms, yet it detects faults at times similar to or even earlier than existing methods. This indicates that the method is more suitable for unsupervised scenarios.

III) The experiments conducted on the XJTU-SY dataset corroborate the conclusions of the previous two experiments. In experiments (1) and (2) on the XJTU-SY dataset, the data used for testing and the data used for feature selection were not the same, which once again demonstrates that this method can accomplish fault detection tasks with a small number of samples. Additionally, despite the different operating conditions in (1) and (2), the method proposed in this paper was able to detect the occurrence of faults, showcasing its universality. Experiment three (3) shows that even when the unknown signal's fault mode involves a mixture of multiple faults, this method remains effective.

**5. Conclusions**

This paper presents a novel method for detecting faults based on engine vibration signals. The method involves extracting time and frequency domain features from normal signals to establish an LSTM time series prediction model. It utilizes the prediction accuracy as a continuous indicator for unknown signals to indicate the presence of faults. And the effectiveness of the method is validated through experiment.

The advantages of this method are as follows:

(1) Overcoming the limitations of traditional strategies that categorize fault severity in a discrete manner, this method utilizes a continuous indicator to indicate the severity of faults. As a result, it provides greater flexibility in setting threshold values to differentiate faults.

(2) Thanks to a well-structured algorithm framework, this method boasts strong scalability and customizability. In the feature selection phase, in addition to the 18 features preset in this paper, other new features can also be incorporated. Apart from the DTW algorithm used in this paper, other similarity measurement algorithms can be employed. During the training phase, instead of the LSTM network used in this paper, other time series prediction models can be utilized. Furthermore, for outlier detection, algorithms other than the Z-score can also be adopted.

(3) During training, only a small amount of fault data was used (specifically in the feature selection process), yet the method can still effectively detect faults in unknown signals. Given that collecting fault data in real-world engineering environments can be challenging, and typically only a small amount of fault data can be obtained, this method is expected to perform well in real production scenarios with limited samples.

The limitations of this work include:

(1) Training deep learning model requires a large number of feature time series samples, and obtaining these feature samples necessitates slicing the original signals. Therefore, implementing this method requires a sufficient amount of original data, and it can consume significant memory resources during computation.

(2) Manually selecting features lacks intelligence. If features could be extracted adaptively using neural network methods, it would save time resources while also increasing generality.

(3) The process of feature selection and threshold determination lacks standardized criteria and relies on the subjective decisions of researchers. There is a need for more universally applicable and robust methods for adaptive decision-making.

Acknowledgements

This work was supported by the National Science and Technology Major Project, China, under Grant J2019-I-0019-0018 and Grant J2019- I-0001-0001.

Weirui Zhang and Haihui Wang is with the School of Mathematical Sciences, Beihang University (BUAA), Beijing 100191, China (Corresponding author: Haihui Wang e-mail: whhmath@buaa.edu.cn).

**References**

[1] G. Georgoulas, T. Loutas, C. D. Stylios, and V. Kostopoulos, “Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition,” *Mechanical Systems and Signal Processing*, vol. 41, no. 1–2, pp. 510–525, Dec. 2013, doi: 10.1016/j.ymssp.2013.02.020.

[2] C. Malla and I. Panigrahi, “Review of Condition Monitoring of Rolling Element Bearing Using Vibration Analysis and Other Techniques,” *J. Vib. Eng. Technol.*, vol. 7, no. 4, pp. 407–414, Aug. 2019, doi: 10.1007/s42417-019-00119-y.

[3] M. Cerrada *et al.*, “A review on data-driven fault severity assessment in rolling bearings,” *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, Jan. 2018, doi: 10.1016/j.ymssp.2017.06.012.

[4] C. Bujoreanu, R. Monoranu, and N. D. Olaru, “Study on the Defects Size of Ball Bearings Elements Using Vibration Analysis,” *Applied Mechanics and Materials*, vol. 658, pp. 289–294, 2014, doi: 10.4028/www.scientific.net/AMM.658.289.

[5] M. Singh and R. Kumar, “Thrust bearing groove race defect measurement by wavelet decomposition of pre-processed vibration signal,” *Measurement*, vol. 46, no. 9, pp. 3508–3515, Nov. 2013, doi: 10.1016/j.measurement.2013.06.044.

[6] S. Zhao, L. Liang, G. Xu, J. Wang, and W. Zhang, “Quantitative diagnosis of a spall-like fault of a rolling element bearing by empirical mode decomposition and the approximate entropy method,” *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 154–177, Oct. 2013, doi: 10.1016/j.ymssp.2013.04.006.

[7] “An Improved Bearing Fault Diagnosis Method using One-Dimensional CNN and LSTM,” *SV-JME*, Jun. 2018, doi: 10.5545/sv-jme.2018.5249.

[8] X. Guo, C. Shen, and L. Chen, “Deep Fault Recognizer: An Integrated Model to Denoise and Extract Features for Fault Diagnosis in Rotating Machinery,” *Applied Sciences*, vol. 7, no. 1, Art. no. 1, Jan. 2017, doi: 10.3390/app7010041.

[9] H. Shao, H. Jiang, Y. Lin, and X. Li, “A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders,” *Mechanical Systems and Signal Processing*, vol. 102, pp. 278–297, Mar. 2018, doi: 10.1016/j.ymssp.2017.09.026.

[10] C. Shen, F. Hu, F. Liu, A. Zhang, and F. Kong, “Quantitative recognition of rolling element bearing fault through an intelligent model based on support vector regression,” in *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, Beijing, China: IEEE, Jun. 2013, pp. 842–847. doi: 10.1109/ICICIP.2013.6568189.

[11] Y. Lei, Z. He, and Y. Zi, “EEMD method and WNN for fault diagnosis of locomotive roller bearings,” *Expert Systems with Applications*, vol. 38, no. 6, pp. 7334–7341, Jun. 2011, doi: 10.1016/j.eswa.2010.12.095.

[12] T. W. Rauber, F. De Assis Boldt, and F. M. Varejao, “Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 637–646, Jan. 2015, doi: 10.1109/TIE.2014.2327589.

[13] E. Keogh and C. A. Ratanamahatana, “Exact indexing of dynamic time warping,” *Knowl Inf Syst*, vol. 7, no. 3, pp. 358–386, Mar. 2005, doi: 10.1007/s10115-004-0154-9.

[14] G. Wang, T. Wang, J. Chen, and S. Zhao, “Bearing fault feature selection method based on dynamic time warped related searches,” *J. vibroeng.*, vol. 25, no. 2, pp. 311–324, Mar. 2023, doi: 10.21595/jve.2022.22863.

[15] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[16] R. Sabir, D. Rosato, S. Hartmann, and C. Guehmann, “LSTM Based Bearing Fault Diagnosis of Electrical Machines using Motor Current Signal,” in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA: IEEE, Dec. 2019, pp. 613–618. doi: 10.1109/ICMLA.2019.00113.

[17] S. Saleem, M. Aslam, and M. R. Shaukat, “A REVIEW AND EMPIRICAL COMPARISON OF UNIVARIATE OUTLIER DETECTION METHODS”.

[18] “Welcome to the Case Western Reserve University Bearing Data Center Website | Case School of Engineering | Case Western Reserve University,” Case School of Engineering. Accessed: Mar. 12, 2024. [Online]. Available: https://engineering.case.edu/bearingdatacenter/welcome

[19] J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services (2007). IMS, University of Cincinnati. "Bearing Data Set", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA

[20] B. Wang, Y. Lei, N. Li, and N. Li, “A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings,” *IEEE Trans. Rel.*, vol. 69, no. 1, pp. 401–412, Mar. 2020, doi: 10.1109/TR.2018.2882682.

[21] Yaguo L., Tianyu H., Biao W., Naipeng L., Tao Y., and Jun Y., “XJTU-SY Rolling Element Bearing Accelerated Life Test Datasets: A Tutorial,” *Journal of Mechanical Engineering*, vol. 55, no. 16, p. 1, 2019, doi: 10.3901/JME.2019.16.001.

[22] D. Sacerdoti, M. Strozzi, and C. Secchi, “A Comparison of Signal Analysis Techniques for the Diagnostics of the IMS Rolling Element Bearing Dataset,” *Applied Sciences*, vol. 13, no. 10, p. 5977, May 2023, doi: 10.3390/app13105977.

[23] W. Gousseau, J. Antoni, F. Girardin, and J. Griffaton, “Analysis of the Rolling Element Bearing data set of the Center for Intelligent Maintenance Systems of the University of Cincinnati”.