

Research Article

Health Diagnosis of Roadheader Based on Reference Manifold Learning and Improved *K*-Means

Xiaodong Ji^(b),¹ Yang Yang^(b),² Yuanyuan Qu^(b),¹ Hai Jiang,¹ and Miao Wu¹

¹China University of Mining and Technology (Beijing), Beijing 100083, China ²Shijiazhuang Coal Mining Machinery Co., Ltd., Shijiazhuang 050000, China

Correspondence should be addressed to Yuanyuan Qu; 201419@cumtb.edu.cn

Received 3 July 2021; Revised 20 September 2021; Accepted 1 November 2021; Published 30 November 2021

Academic Editor: Nibaldo Rodríguez

Copyright © 2021 Xiaodong Ji et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The safe and stable operation of roadheader is of great significance to the efficient and rapid production of a coal mine. Health diagnosis based on vibration signals has been studied in bearings and motors. Complex geological conditions and bad working environment lead to the characteristics of nonlinear and time-varying vibration signals of a roadheader. In this paper, a health state analysis method based on reference manifold (RM) learning and improved *K*-means clustering analysis was proposed; the method was verified by using the real-time collected roadheader cutting reducer fault signal. Firstly, the comparison signal and analysis signal were extracted from the actual collected vibration data of the roadheader, and the referential analysis samples were constructed through time domain and wavelet packet energy analysis. Then, the characteristic structure of the low-dimensional space of the referential analysis samples is obtained by Locally Linear Embedding (LLE), which is a method of manifold learning. Through the improved *K*-means clustering analysis method, the low-dimensional structure parameters were analyzed and the clustering effect index was obtained, which was used as the health evaluation index (HEI). Finally, the normal distribution model of the health evaluation index is established, and the confidence interval of the health evaluation index is determined, so as to realize the health state analysis of the roadheader and realize the fault warning function. Through the analysis of data of three sensors, the results show that the roadheader failed on the 15th day, which is consistent with the actual working condition. Through practical analysis, the effectiveness of the method was verified and provided a kind of fault analysis idea and method for equipment working under complex working conditions and the theoretical basis for fault type analysis.

1. Introduction

In the past 20 years of coal mining in China, the development of roadheader has experienced the introduction of foreign technology, the self-development of small roadheader, the application of high-power multiauxiliary function roadheader, and the research and development of intelligent roadheader [1, 2]. After years of experience accumulation, the reliability and automation of the roadheader has reached a certain height. However, due to the different physical characteristics of coal seams in different regions of China, the different geological conditions, and the different operating habits of operators, there will still be some failures in the process of using the roadheader: hydraulic system blockage and heating, electrical equipment failure, transmission parts, and hydraulic pump and motor damage, which lead to reduced excavation efficiency and increased the excavation time and economic cost. In order to ensure the normal use of the roadheader, the important parts of the roadheader will be replaced and periodically overhauled regularly in the coal mine, which leads to the transition maintenance of the roadheader and the waste of resources.

The working environment of the coal roadway heading belongs to group operation in the narrow and closed space, and the working conditions are complicated and changeable [3]. The cutting head of the roadheader directly cuts the coal rock, so its load has the characteristics of nonlinear, timevarying, and strong coupling, coupled with the physical characteristics of the coal seam, roof pressure, and other factors of uncertainty so that the actual load is often fluctuating violently, and the predictability is very low. Therefore, under the complex impact and forced vibration of the tunneling machine, the pick gear will accelerate the wear and tear, the fatigue life of the cutting reducer and the supporting parts of the machine body will be reduced, and even lead to the fuselage swing, the wrong action of the electronic control system, and other situations, affecting the normal production. The health status assessment of the roadheader should take full account of the structural damage caused by vibration and carry out multiangle assessment [4]. So, it has become an important research content of tunneling production to evaluate the health status of the roadheader effectively.

At present, some research has been done on the performance analysis of the roadheader at home and abroad. Piotr Cheluszka designed a method for the vibration identification of tunneling machine cutting head using a high-speed camera. The effectiveness of this method is verified by the data collection [5]. Acaroglu compared and analyzed the vibration characteristics and stability of the cutting head of the roadheader under different cutting modes. According to stability analysis for the roadheader with point attack cutters, there might be a stability problem at turning around the vertical axis stability state in arcing the cutting mode [6]. Huang established a nonlinear dynamic model and carried out vibration analysis of the bearing under time-varying loads in order to study the dynamic characteristics of the cutting head and the cantilever beam system [7]. Sadi Evren Seker focuses on roadheader performance prediction using six different machine learning algorithms and a combination of various machine learning algorithms via ensemble techniques. Algorithms are ZeroR, random forest (RF), Gaussian process, linear regression, logistic regression, and multilayer perceptron (MLP). As a result, MLP and RF give better results than the other algorithms. Also, the best solution achieved was the bagging technique on RF and principle component analysis (PCA) [8]. Qiang Liu used four machine learning tools (the backpropagation neural network based on genetic algorithm optimization, the Naive Bayes based on genetic algorithm optimization, the support vector machines based on particle swarm optimization, and the support vector machines based on dynamic cuckoo) to analyze the vibration data of the tunneling machine and complete the incipient fault detection and identification [9].

The analysis of the vibration signal of the roadheader is carried out on the basis of theoretical analysis and simulation test. Although it provides some theoretical support for the fault analysis of the roadheader, it lacks the actual data of tunneling operation as support. Zhang analyzes the vibration of a rotary table by combining the finite element model with tested data from an underground coal mine. The vibration mode, frequency, and damping ratio of the actual results are consistent with the simulation results, which verify the validity of the simulation data [10]. Yang proposed a roadheader anomaly detection method based on VSAPSO-BP under single class learning, aiming at the problem of missing fault data and fault degree division in abnormal detection of the roadheader turntable [11]. Qu analyzed the measured vibration data of the roadheader turntable by DTCWT and obtained the natural frequency of the turntable under the actual working state of the roadheader, which verified the feasibility of applying the modal recognition theory to the vibration signal processing of the roadheader under complex working conditions [12].

To sum up, the common analysis methods of roadheader vibration were time-frequency analysis and wavelet analysis. The characteristic parameters of roadheader were extracted by these methods and then compared with the simulation data to realize fault identification of roadheader. These analyses can only be carried out on the roadheader working under a single working condition, which cannot meet the nonlinear and time-varying characteristics of the roadheader load. They were not very universal on the other condition. In order to deeply understand the relationship between the vibration and running state of the roadheader, the problem of resource waste caused by excessive maintenance in the actual production of the roadheader was solved, the lowproduction efficiency caused by faults was reduced, and a feasible fault state evaluation method of the roadheader was explored; in this paper, health vibration data and analysis data are fused, and the relationship between the health data and the analysis data is compared through reference manifold learning, and the structural characteristic parameters of the samples in the low-dimensional space are mapped out. Then, the clustering effect of low-dimensional characteristic parameters was obtained through the improved K-means analysis, and the clustering effect was taken as the health evaluation index (HEI) of the roadheader to complete the evaluation of the health state of the roadheader.

2. Analysis Scheme

The technical scheme of this health state analysis method is shown in Figure 1. First, the health data and analysis data of different time periods are separated from vibration data of the roadheader, and characteristic parameters are extracted by time domain analysis and wavelet packet analysis. The characteristic parameters of health data and the characteristic parameters of analysis data were fused into the reference analysis samples. After that, dimensionality reduction processing by manifold learning (ML) was carried out on the referential analysis samples. Then, improved K-means clustering analysis was carried out on the dimensionality reduction data, and the clustering effect was used as the health evaluation index of the samples. Finally, the normal distribution model of the evaluation index was established by using the Gaussian model, and the confidence interval of the evaluation index is determined.

3. Constructing the Reference Analysis Sample

3.1. Vibration Signal Collection. The vibration signal collection site of the roadheader is Xingdong Mine in Hebei Province, China. The coal in this mine is mainly gas-fat coal with high calorific value, and the coal seam is deep between 580 and 1200 m underground. The roadway is 1100 lane in the southern mining area of the Xingdong mine; the

Shock and Vibration



FIGURE 1: The technical scheme of this health state analysis method.

roadway section and the coal seam section are the same, as shown in Figure 2. The tunneling length is 200 m, and there is no geological change such as fault and gangue in the tunneling process. The roadway trend is shown in Figure 3.

The experimental object is the EBZ160 boom-type roadheader with the largest cutting surface of 5300×4850 mm. The Boom-type roadheader is a comprehensive tunneling equipment with cutting, shipping, and walking, which is composed of cutting section, loading section, frame, walking section, back support, electric cabinet, and other sections, as shown in Figure 4. And, the cutting section is mainly composed of cutting head, telescopic section, cutting motor, reducer, lifting cylinder, rotary cylinder, and turntable, as shown in Figure 5. The reducer is a two-stage planetary gear reducer, which drives the cutting head to rotate through the output shaft to realize its function, and its transmission ratio is 31.03. [17].

According to the working environment and load conditions of the roadheader, a total of 5 measuring points are arranged near the cutting part of the roadheader, and each measuring point has 2 vibration sensors in different directions, thus with a total of 10 sensors. The specific location of the sensor is shown in Figure 6 and Table 1. The sampling frequency of the vibration data is 10 kHZ, and the sampling time is 32 days.

Roadheader cutting section is not only the main working part of the roadheader but also the main vibration impact bearing part. Sensor 3 is the closest to the failure part of this experiment. Therefore, Sensor 3 was selected as the data analysis point, as shown in Figure 7.

As shown in Figure 8, it shows the vibration data of the roadheader sensor 3 working for 3 hours on the 16th day.



FIGURE 2: Schematic of section.



3.2. Constructing Samples. According to the working records of the roadheader, a mechanical failure occurred in the cutting-arm of the roadheader during the 32-day data collection process, and the gear of the reducer of the cutting-



FIGURE 4: The roadheader.



FIGURE 5: Cutting section of the roadheader.

arm was worn and the teeth were broken due to long-term overload. The fault caused the roadheader to be unable to run, so it was reused after replacing the reducer.

In traditional vibration signal analysis, the sensitive characteristic is extracted from vibration signals for comparison and experience summary, based on some data processing methods and statistical methods [18–20]. In this paper, the characteristic extracted by the traditional analysis (time-frequency characteristic frequencies and band energy) is used as the original analysis data.

The time-frequency characteristic parameters of the vibration signal are extracted, including peak-to-peak, effective value, absolute mean value, impulse factor, kurtosis value, margin factor, peak factor, waveform factor, and 8 frequency band energy eigenvalues of three-layer wavelet

packet analysis (as Table 2), which are composed into the characteristic parameter set of the vibration signal.

The steps of frequency band energy feature extraction for wavelet packet decomposition are as follows:

(1) 3-layer wavelet packet decomposition are performed for each group of vibration data, and the wavelet packet decomposition coefficients of 8 subbands from low frequency to high frequency of the 3rd layer are obtained:

$$(X_{31}, X_{32}, X_{33}, X_{34}, X_{35}, X_{36}, X_{37}, X_{38}).$$
 (1)

(2) The wavelet component analytical coefficients are reconstructed, and the signals in each subband range are extracted:

$$S = S_{31} + S_{32} + S_{33} + S_{34} + S_{35} + S_{36} + S_{37} + S_{38}.$$
 (2)

(3) The energy of each subband is calculated:

$$Q_{3i} = \int |S_{3i}(t)|^2 dt = \sum_{h=1}^{\infty} |x_{ih}|^2, \qquad (3)$$

where i = 1, 2, ..., 7, 8 is the subband and h = 1, 2, ..., n is the data of S_{3i}

(4) The energy of each subband is normalized:

$$Ei = \frac{Q_{3i}}{Q},\tag{4}$$

where $Q = \sum_{i=1}^{i=8} Q_{3i}$.



FIGURE 6: Arrangement drawings of the roadheader's measuring point.

Measuring point	Measuring point position	Sensor direction	Sensor	Coordinate
1	Lifting cylinder support	Perpendicular to the cutting arm (X)	1	
1	point	Horizontal to the cutting arm (<i>Y</i>)	2	
2	Cutting georbox	Perpendicular to the cutting arm (X)	3	
2	Cutting gearbox	Horizontal to the cutting arm (<i>Y</i>)	4	
2	Hinge point between	Perpendicular to the cutting arm (X)	5	Establishing the coordinate system according to the roadheader: <i>X</i> : perpendicular to the side of the cutting part;
5	cutting part and turntable	Horizontal to the cutting arm (<i>Y</i>)	6	<i>Y</i> : horizontal to the heading direction of the roadheader; <i>Z</i> : perpendicular to the upper plane of the roadheader
4	Fixed points of lifting	Perpendicular to the cutting arm (<i>X</i>)	7	
4	cylinder and turntable	Horizontal to the cutting arm (<i>Y</i>)	8	
F	Trumtable conton	Horizontal roadheader direction (Y)	9	
3	furniable center	Vertical roadheader upper plane (Z)	10	

TABLE 1: Corresponding relation of vibration sensor.



FIGURE 7: Arrangement drawing of vibrating sensor.



FIGURE 8: Vibration data.

TABLE 2: Characteristic parameters.

Time-frequency characteristics	Markup symbols	Energy ratios of wavelet analysis	Markup symbols
Peak-to-peak	pk	Ratio 1	E1
Effective value	st	Ratio 2	E2
Absolute mean	me	Ratio 3	E3
Impulse factor	va	Ratio 4	E4
Kurtosis value	Kr	Ratio 5	E5
Margin factor	L	Ratio 6	E6
Peak factor	С	Ratio 7	E7
Waveform factor	S	Ratio 8	E8

On the first day, the vibration signal was collected, and the roadheader ran normally. On the 20th day, a broken tooth occurred, which led to the roadheader could not run, and the reducer was replaced.

According to the running condition of the roadheader and the work log, 3rd, 6th, 9th, 11th, 13th, 14th, 15th, 16th, 17th day, and AM 18th day, BM 18th day, and AM 19th day were taken as the health state analysis time points of the roadheader. 9 segments of vibration data with a length of 180 seconds were extracted from each time point as the analysis samples.

Through signal analysis, the characteristic parameter sample of each segment of data is obtained:

$$x_{ti} = [pk_{ti}, st_{ti}, me_{ti}, ...E7_{ti}, E8_{ti}]^{T},$$
(5)

where t = 1, 2, 3, ..., 12 are the analysis time points and i = 1, 2, 3, ..., 9 are the 9 segments vibration data of a point.

Then, 9 characteristic parameter samples of the analysis time point composed the analysis sample:

$$X_t = \{x_{t1}, x_{t2}, x_{t3}, \dots x_{t9}\}.$$
 (6)

According to the above method, the sample of the 1st day is obtained, which was used as the comparison sample:

$$X_0 = \{x_{01}, x_{02}, x_{03}, \dots x_{09}\}.$$
 (7)

Finally, the reference analysis sample is constructed, which is the analysis object of LLE:

$$N_t = \{X_0, X_t\}, t = (1, 2, 3, ..., 11, 12).$$
(8)

4. Reference Manifold Learning

The manifold learning method can be divided into linear learning and nonlinear learning [14]. Its essence is that the sample points in the high-dimensional space are a manifold stretched by a few major independent variables simultaneously acting on the measurement space, and the manifold learning is to pick up the low-dimensional manifold embedded in the high-dimensional observation space and to find the essential characteristics of the data in the observation space and establish a new mapping relationship [14, 21]. Manifold learning will maintain the local structure of the high-dimensional original data to the maximum extent and realize the learning and enhancement of data characteristics and essential information, when it obtains the low-dimensional embedded feature data that is most similar to the original data, which is different from other feature extraction methods and the greatest advantage in the process of mining and enhancing the essential information of data [22–24].

Linear manifold learning is mainly to extract the mechanism modal response of long-order signals, remove the interference of irrelevant noise, learn the signal modal information, and mine the feature information of low-dimensional manifold embedded in high-dimensional space. Its analysis pays more attention to the global changes and ignores the local relationships of samples. Nonlinear manifold learning can express and reconstruct the local information of samples globally and express better the internal relationship of small data space.

Considering the following aspects, the paper adopts a nonlinear manifold learning method—Locally Linear Embedding (LLE) to mine the inherent characteristic of the two kinds of data embedded in the sample species:

(1) In this paper, a reference comparison analysis method is used to analyze a comparison model composed of the analysis samples and health samples in local space. The nonlinear manifold learning method is more suitable for the clustering distribution law of the feature space of the referenced model. The local structure relationship of the original sample can be maintained to the maximum extent.

- (2) The complexity and diversity of load lead to the nonlinear of the vibration signal of the roadheader.
- (3) In this paper, the reference analysis sample is 18 × 16 small sample, which is more suitable for the nonlinear analysis method.

Therefore, in this paper, Locally Linear Embedding (LLE) is used to analyze the referential analysis samples and find the low-dimensional mapping relationships. LLE describes the local geometric features of the data by the locally linear projection information, taking advantage of the difference of the local neighborhood of each point and the weight information of the neighboring points, and finally realizes the expression and reconstruction of the local information of the sample on a global scale.

LLE constructs the weight subspace of the data group by relying on the neighborhood of each group data, by using the weight information of data, which is different at different locations or in different periods of time on the same equipment. Then, through dimensionality reduction analysis of dataset, the information of multidimensional dataset can be expressed and reconstructed in a low-dimensional space [25–27].

For the high-dimensional feature set,

 $N = \{n_1, n_2, n_3, \dots, n_i\}, n_i \in \mathbb{R}^{D \times m}, D = 16, m = 18$ sensitive features are manifested in a low-dimensional (d-dimensional) space:

$$Y = \{y_1, y_2, y_3, \dots, y_i\}, y_i \in \mathbb{R}^{d \times m}.$$
 (9)

LLE is divided into three steps [28, 29] and is discussed in the below sections.

4.1. Construction of Neighborhood Space. According to the high-dimensional features and the Euclidean distance between each sample point n_i in N, m nearest neighbors of each sample point is found as follows:

$$N_i = knn(n_i, m), N_i = [n_{1i}, \dots n_{ik}].$$
 (10)

4.2. Calculation of Local Weight. The nonlinear relationship between each sample and its neighbor's subspace is calculated. The local error function $\varepsilon(W)$ is minimized, and the local weight matrix W is constructed as follows:

$$\varepsilon(W) = \sum_{i=1}^{m} \left\| n_{i} - \sum_{j=1}^{k} w_{ij} n_{j} \right\|^{2}$$

$$= \sum_{i=1}^{m} \left\| (n_{i} - n_{ji}) w_{ji} \right\|^{2}$$

$$= \sum_{i=1}^{m} \left\| (N - N_{i}) w_{i} \right\|^{2}$$

$$= \sum_{i=1}^{m} w_{i}^{T} (N - N_{i})^{T} (N - N_{i}) w_{i},$$

(11)

where *N* is the high-dimensional feature set, N_i is the *m* neighbor of n_i , W_{ij} is the weight between n_i and n_j ; if they are not neighbors, then $w_{ij} = 0$. And the weight of neighbor's subspace of n_i meets the following:

$$\sum_{i=1}^{n} w_{ij} = \sum_{j=1}^{k} w_i = 1.$$
(12)

Setting

$$S_i = (N - N_i)^T (N - N_i),$$
 (13)

and substituting equation (13) into equation (11),

$$\varepsilon(W) = w_i^T S_i w_i. \tag{14}$$

In order to get the optimal weight matrix, according to the population dimension reduction analysis method, the Lagrange multiplier method is adopted. So,

$$L(w_i) = \sum_{i=1}^{n} w_i^T S_i w_i + \lambda (w_i^T 1_m - 1).$$
(15)

Taking the derivative of equation (15)

$$\frac{\partial L(w_i)}{\partial w_i} = 2S_i w_i + \lambda 1_m = 0.$$
(16)

Therefore,

$$w_i = \frac{S_i^{-1} 1_m}{1_m^T S_i^{-1} 1_m},\tag{17}$$

where 1_m is $m \times 1$ column vectors with all 1.

4.3. Embedded Coordinate Projection. Embedded coordinate projection is to solve the mapping of the low-dimensional space. W is an $n \times n$ matrix. And for the sake of clarity, we expand the matrix w as $w^{n \times n}$.

$$W = \{w_i = [w_{i1}, w_{i2}, \dots w_{in}]\}^T.$$
 (18)

Set:

$$\sum_{j=1}^{n} w_{ji} y_{ji} = \sum_{j=1}^{k} w_{ji} y_{ji} = Y W_{i}.$$
(19)

After calculation, we obtain

$$\psi(Y) = \sum_{i=0}^{n} \left\| \left(Y \left(I_i - W_i \right) \right) \right\|^2.$$
 (20)

According to the matrix equation

$$\sum_{i} (a_{i})^{2} = \sum_{i} a_{i}^{T} a_{i} = tr(AA^{T}).$$
(21)

Substituting equation (21) into equation (20),

$$\psi(Y) = tr(Y(I - W)(I - W)^{T}Y^{T})$$

= tr(YMY^T), (22)

where $M = (I - W)(I - W)^{T}$.

$$L(Y) = tr(YMY^{T}) + \lambda(YMY^{T} - nI).$$
(23)

Taking the derivative of equation (23)

$$\frac{\partial L}{\partial Y} = 2YM + 2\lambda Y = 0. \tag{24}$$

Therefore,

$$M^T Y^T = -\lambda Y^T, \tag{25}$$

where $M^T = M$.

So, we can obtain

$$MY^T = -\lambda Y^T.$$
(26)

So, matrix Y^T is composed of eigenvectors of the matrix M. In order to reduce the dimension of data to d, we only need to get the eigenvectors corresponding to the smallest d nonzero eigenvalues of M. In LLE analysis, the smallest eigenvalue is generally discarded because it is too close to 0. Therefore, the eigenvectors of $[2, \ldots, d+1]$ eigenvalues from small to large are selected.

Through the above methods, 12 reference analysis samples of 12 time points were analyzed by LLE, where d = 2, k = 6. The analysis results are shown in Figure 9.

The results of LLE analysis show that

- (1) The analysis results of the comparison samples and the analysis samples show that the dimension reduction features at the 3rd, 6th, 9th, 11th, 13th, and 14th day were crossed together without segmentation. It indicates that the roadheader runs normally and is not damaged.
- (2) The analysis results of the comparison sample and the analysis sample began to separate from each other on the 15th and 16th day, and the data of the comparison sample and the analysis sample appeared obvious segmentation phenomenon on the 17th day.
- (3) As can be seen from the analysis results on the 16th, 17th, 18th, and 19th day, as time goes by, the separation distance between the health comparison samples and the dimension reduction characteristics of the analysis samples is getting farther and farther, indicating that the roadheader faults are becoming more and more serious. And as the fault becomes more and more serious, the gap within the class becomes smaller and smaller, and the gap between the classes becomes larger and larger.

Through the analysis, the roadheader failed on the 15th day, and with the passage of time, the failure became more and more serious, and eventually led to the damage of the roadheader reducer.

5. Clustering Analysis

Cluster analysis is one of the important research fields of data mining and pattern recognition. It plays an extremely important role in identifying the internal structure of data. Commonly used clustering analysis algorithms include *K*-means, *K*-modes, PAM (partitioning around medoid), and CLARA (clustering large applications) algorithm.

K-means algorithm is a classical algorithm used to solve clustering problems. Compared with other algorithms, *K*-means is simpler and faster. In terms of data processing, the *K*-means algorithm has better scalability. Meanwhile, the *K*-means algorithm tries to find *K* partitions that minimize the squared error function value. When the difference between clusters is obvious, its clustering effect is better.

5.1. Improved K-Means Analysis. The basic idea of K-means clustering is as follows: the sample set is divided into K clusters according to the size of the sample distance, so as to ensure that the points within the cluster are closely related to each other, and the distance between the clusters is as large as possible. The degree of density within clusters and the degree of dispersion between clusters can evaluate the clustering effect [30–32].

In the construction of referential analysis samples, comparison samples and analysis samples have been separated, so when K-means clustering processing is carried out, there is no need to divide samples into clusters. Through the improved K-means data processing, the clustering center U of comparison samples and analysis samples and the Calinski–Harabasz (CH) index of the K-means clustering were finally obtained [33–35].

CH index is a parameter to evaluate the effect of clustering. The effect of clustering the evaluation can be reflected by the degree of density within clusters and the degree of dispersion between clusters. The number of analysis samples or comparison samples in each time period is i = 9, and the data category K=2. Therefore, the CH index of K-means clustering is used as the Clustering effect indicator (CEI) in this paper, which is used as the health evaluation indicator (HEI) of the roadheader.

The improved K-means algorithm is as follows:

- (1) The comparison data and analysis data in dimensionality reduction space (result of LLE) are divided into two clusters $C = (C_1, C_2)$
- (2) The centroid of sample C_{λ} is calculated
- (3) The centroid matrix for sample *C* is obtained:

$$U = (u_1, u_2). \tag{27}$$

(4) Calinski-Harabasz (CH) index is calculated:

$$CH(k) = \frac{trB(k)/(k-1)}{trW(k)/(n-k)},$$
(28)

where n = 18 is the number of sample points, k is the cluster of samples, trB(k) is the trace of the class deviation matrix, and trW(k) is the trace of the inclass deviation matrix.

After the improved *K*-means clustering analysis, the centroid of the two clusters in the dimensioned reduced space is shown in Figure 10 and Table 3. Calinski–Harabasz



FIGURE 9: LLE analysis result. \Diamond : low-dimensional spatial characteristics of health comparison signals after LLE processing; \bigstar : low-dimensional spatial characteristics of analysis signals after LLE processing.

Shock and Vibration



FIGURE 10: Centroid of clusters.

TABLE 3: Centroid of the two clusters.

Day	3rd	6th	9th	11th	13th	14th	15th	16th	17th	AM18th	BM18th	AM19th
Haalth data	-0.0107	0.0115	0.0232	0.01	0.0203	0.0352	-0.101	-0.139	-0.355	-0.603	-0.454	-0.727
Health data	0.0366	-0.0221	-0.0898	-0.0131	-0.0559	0.129	0.0999	-0.0451	-0.586	-0.586	-0.892	-1.053
A	-0.0069	0.0567	0.098	0.0496	0.0857	0.0465	0.404	0.552	0.230	0.326	0.381	0.509
Analysis data	-0.105	0.0267	0.0163	0.0199	0.0112	-0.03	0.0864	0.178	0.648	0.624	0.604	0.875

TABLE 4: Health evaluation indicator (HEI) of each time point.

Day	3rd	6th	9th	11th	13th	14th	15th	16th	17th	AM18th	BM18th	AM19th
HEI (CH)	13.28	14.41	18.98	22.97	15.45	24.34	29.08	38.15	50.43	99.2	143.17	297.87

(CH) index, namely, health evaluation indicator (HEI), is shown in Table 4 and Figure 11.

After the improved K-means clustering analysis, we found that

- (1) When the roadheader was faultless, the distance between the centroid points of the healthy sample and the analysis sample was very close, and the value of the HEI (health evaluation index) was also maintained within a stable value range.
- (2) After the failure of the roadheader, the distance between the centroid points of the healthy sample and the analysis sample became farther and farther with the severity of the failure, and the HEI increased exponentially.

5.2. Health Assessment Based on SGM. Generally, the distribution of health status assessment parameters of equipment conforms to the normal distribution, and the nonhealth data will deviate from this distribution by the degree of failure [36]. In this paper, the single Gaussian model (SGM) is used to analyze the HEI distribution of the roadheader. In the SGM model, HEI conforms to the normal probability distribution function (pdf), and the equation is

$$p(s/u,\delta) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left[-\frac{(s-u)^2}{2\delta^2}\right],$$
 (29)

where *s* is the HEI, *u* is the mean value of HEI of health signal, and δ is the variance of HEI of health signal [37].

According to the normal distribution, the 95% confidence level of the model was selected as the health interval.

P of normal distribution corresponding to the HEI of the equipment at each time point can indicate the close level of the equipment status with the health status. The corresponding value of the equipment failure status signal will exceed the given health confidence level interval.

When the equipment fails, P of the vibration signal will exceed the given health confidence interval.

Health assessment indicators (HEI) of the roadheader at 3rd, 6th, 9th, 11th, 13th, and 14th day were selected as normal distribution parameters of the health data.

By calculation, u = 18.2436; $\delta = 17.8799$.

Subsituting it into equation (29),



FIGURE 11: The trend of HEI. ◊: low-dimensional spatial characteristics of health comparison signals after LLE processing; ★: lowdimensional spatial characteristics of analysis signals after LLE processing; ■: the center of low-dimensional characteristics of health comparison signals; •: the center of low-dimensional characteristics of analysis signals.

$$p = \frac{1}{44.807} \exp\left[-\frac{(s - 18.2436)^2}{639.3835}\right].$$
 (30)

The 95% confidence interval corresponds to $s_p = 25.8627$. When the HEI value $s > s_p$, the roadheader failure occurs. If the value $s - s_p$ is larger, then the fault is more serious.

6. Data Analysis

In order to verify the noncontingency and effectiveness of the method, we analyze the vibration signals of sensors 1 and 5, which are more advanced in the distance cutting reducer, according to the above method. Finally, HEI of the two sensors was obtained, as shown in Figure 12 and Table 5, and the corresponding confidence interval of two sensors is shown in Table 6.

Based on the above analysis results, it can be concluded that the HEI of the two sensors exceeds the confidence interval on the 15th day of the roadheader operation, and the HEI increases with the severity of the fault. The results are consistent with those of sensor 3. It is proved that this method can evaluate the health status of the roadheader.



FIGURE 12: The HEI trend of two sensors, (a) the HEI trend of sensor 1, and (b) the HEI trend of sensor 5.

TABLE 5:	HEI	of	two	sensors	in	each	time	point.
----------	-----	----	-----	---------	----	------	------	--------

Day	3rd	6th	9th	11th	13th	14th	15th	16th	17th	AM18th	BM18th	AM19th
HEI of sensor 1	6.29	8.41	7.99	10.98	11.45	15.20	21.31	30.25	37.59	67.41	120.51	198.62
HEI of sensor 5	9.48	15.58	10.82	13.54	11.54	14.85	24.09	28.15	36.43	70.20	113.28	207.16

TABLE 6: The confidence interval of two sensors.

Sensor	Sensor 1	Sensor 5
Confidence interval	$S_{P1} = 15.2765$	$S_{P5} = 16.5844$

7. Conclusion

Through the analysis results of sensors 1, 3, and 5, it was concluded that from the 15th day: $s > s_p$, which mean the roadheader was starting to malfunction. And with the aggravation of the failure, the HEI increased exponentially. Eventually, the reducer was scrapped and could not be used at all. The analysis results of the three sensors were consistent. The research shows that the low-dimensional sensitive features characteristics based on reference manifold learning can reflect the health status of roadheader, and the health evaluation index (HEI) by cluster analysis can identify the roadheader faults.

- (1) In view of the characteristics of nonlinear and timevarying vibration signals of the roadheader caused by unstable loads, the health data and analysis data were fused. Through LLE learning, the characteristic parameters of data in the low-dimensional structure were constructed by taking advantage of the different weight of data in the local neighborhood. It solves the difficult problem that some or several characteristic parameters cannot identify the health state of equipment.
- (2) Since the number of clusters and the data in each cluster has been determined, the CH value judging the clustering effect was used as the evaluation index to judge the health data and analysis data of the lowdimensional characteristic parameters of the vibration signal of the roadheader and take it as the health evaluation index (HEI) of the roadheader.

(3) The normal distribution model of the HEI on healthy condition is established, and the warning line of HEI is determined by confidence interval, which can judge whether the roadheader fails.

Through the analysis of this paper, it not only provides a new diagnosis idea for roadheader health diagnosis but also provides a technical support for fault diagnosis of large equipment. At the same time, it also provides reference value for the identification of equipment fault types [2, 13, 15, 16].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been conducted by the College of Intelligent Mines of the China University of Mining and Technology Beijing in collaboration with Jizhong Energy Group. This work was financially supported by the National Basic Research Program of China (973 Program), (Grant no. 2014CB046306) and the National Natural Science Foundation of China (Grant nos. 51874308 and 61803374).

References

- P. Zhang, S. Li, S. Qiu, L. Guo, and X. Hu, "Advance detection technology and development of fast intelligent roadway tunneling," *Journal of China Coal Society*.
- [2] S. Deshmukh, A. K. Raina, V. M. S. R. Murthy, R. Trivedi, and R. Vajre, "Roadheader—a comprehensive review," *Tunnelling* and Underground Space Technology, vol. 95, p. 1, Article ID 103148, 2020.
- [3] A. Asadi, A. Abbasi, and A. Bagheri, "Prediction of roadheader bit consumption rate in coal mines using artifificial neural networks," in *Proceedings of the 3rd Nordic Rock Mechanics Symposium-NRMS International Society for Rock*

Mechanics and Rock Engineering, p. 12, Helsinki, Finland, October 2017.

- [4] O. Acaroglu and C. Erdogan, "Stability analysis of roadheaders with mini-disc," *Tunnelling and Underground Space Technology*, vol. 68, pp. 187–195, 2017.
- [5] Z. Huang, Z. Zhang, Y. Li, G. Song, and Y. He, "Nonlinear dynamic analysis of cutting head-rotor-bearing system of the roadheader," *Journal of Mechanical Science and Technology*, vol. 33, no. 3, pp. 1033–1043, 2019.
- [6] P. Cheluszka and R. Mann, "Vibration identification of the roadheader cutting head using high-speed cameras," *MATEC Web of Conferences*, vol. 252, Article ID 03018, 2019.
- [7] S. E. Seker and I. Ocak, "Performance prediction of roadheaders using ensemble machine learning techniques," *Neural Computing & Applications*, vol. 31, no. 4, pp. 1103–1116, 2017.
- [8] Q. Liu, S. Liu, Q. Dai, X. Yu, D. Teng, and M. Wei, "Datadriven approaches for diagnosis of incipient faults in cutting arms of the roadheader," *Shock and Vibration*, vol. 2021, pp. 1–15, Article ID 8865068, 2021.
- [9] M. Zhang, F. Lyu, X. Tang, and Y. Yang, "Analysis of vibration of roadheader rotary table based on finite element method and data from underground coalmine [J]," *Shock and Vibration*, vol. 2018, Article ID 4396520, 10 pages, 2018.
- [10] J. J. Yang, Z. Tang, X. Wang, Z. Wang, and M. Wu, "Roadheader anomaly detection method based on VSAPSO-BP under the single category learning," *Joural of Vibration Measurement & Diagnosis*, vol. 39, no. 1, pp. 130–135, 2019.
- [11] Y. Qu, X. Ji, and F. Lv, "Operating modal analysis of the turntable on a working road-header using DTCWT," in *Proceedings of the International Conference on Intelligent and Interactive Systems and Applications*, Bangkok, Thailand, November 2019.
- [12] Li Xu, Research on Mechanical Properties of Key Structures on Boom-Type Roadheader, China University of Mining & Technology, Beijing, China, 2017.
- [13] W. Wei Fan and P. Pizhong Qiao, "Vibration-based damage identification methods: a review and comparative study," *Structural Health Monitoring*, vol. 10, no. 1, pp. 83–111, 2011.
- [14] S. Cao and W. Zhang, Modal Analysis of Vibrating Structures—Theoretical Experiments and Applications, pp. 65–68, Tianjin University Press, Tianjin, China, 2001.
- [15] Z. Zhao, Research on Vibration Signal Based Machinery Fault Feature Extraction and Diagnosis, Beijing Jiaotong University, Beijing, China, 2012.
- [16] X. Ding, Research on Theory and Methodology of Manifold Feature Enhancement for Machine Condition Monitoring and Diagnosis, University of Science and Technology of China, Hebei, China, 2017.
- [17] M. Gashler, D. Ventura, and T. Martinez, "Manifold learning by graduated optimization," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 6, pp. 1458–1470, 2011.
- [18] Q. He, F. Xie, L. Hong, and L. Lan, "Feature extraction of vibration of turbine unit based on manifold learning method," *Journal of Vibration, Measurement & Diagnosis*, vol. 34, no. 4, pp. 705–708, 2014.
- [19] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [20] L. Hajderanj, D. Chen, and I. Weheliye, "The impact of supervised manifold learning on structure preserving and classification error: a theoretical study," *IEEE Access*, vol. 9, pp. 43909–43922, 2021.
- [21] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms

(PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Computer Science Review*, vol. 40, Article ID 100378, 2021.

- [22] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [23] X. Cui, Q. Wang, W. E. I. Kai, T. E. N. G. Geer, and X. U. Xiangjun, "Laser-induced breakdown spectroscopy for the classification of wood materials using machine learning methods combined with feature selection," *Plasma Science* and Technology, vol. 23, no. 5, pp. 1–9, 2021.
- [24] L. Dai, Q. Li, Y. Chen, X. Ding, W. Huang, and Y. Shao, "Complex scale feature extraction for gearbox via adaptive multi-mode manifold learning," *Measurement*, vol. 174, Article ID 108688, 2021.
- [25] Z. Li, Y. Cao, H. Peng, and F. Meng, "Smart grid fault location based on manifold theory and L1+L2 constraints," *Power System Protection and Control*, vol. 48, no. 18, pp. 48–57, 2020.
- [26] D. Camarena-Martinez, M. Valtierra-Rodriguez, J. P. Sanchez, D. Lieberman, R. J. Troncoso, and A. Garcia-Perez, "Shannon entropy and K-means method for automatic diagnosis of broken rotor bars in induction motors using vibration signals," *Shock and Vibration*, vol. 2016, Article ID 4860309, 10 pages, 2016.
- [27] X. Zhang, J. Zhao, X. Ni, H. Li, and W. Song, "Fault Diagnosis of rolling bearing based on multifractal descending algorithm and improve K means clustering," *Transacting of Beijing Institute of Technology*, vol. 39, no. 5, pp. 473–479, 2019.
- [28] M. Kozak, ""A dendrite method for cluster analysis" by caliński and harabasz: a classical work that is far too often incorrectly cited," *Communications in Statistics - Theory and Methods*, vol. 41, no. 12, pp. 2279-2280, 2012.
- [29] T. Calinski and J. Harabasz, "A dendrite method for cluster Analysis," *Communications in Statistics - Simulation and Computation*, vol. 3, no. 1, pp. 1–27, 1974.
- [30] S. Lukasik, P. A. Kowalski, M. Charytanowicz, and P. Kulczycki, "Clustering using Flower Pollination Algorithm and Calinski-Harabasz Index," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation*, pp. 24–29, Vancouver, Canada, July 2016.
- [31] Xu Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," *Material and Science Series*, vol. 569, no. 5, 2019.
- [32] W.-J. Huang and N.-C. Su, "A study of generalized normal distributions," *Communications in Statistics—Theory and Methods*, vol. 46, no. 11, pp. 5612–5632, 2017.
- [33] S. Alavi and M. Tarhani, "On a skew bimodal normal-normal distribution fitted to the old-faithful geyser data," *Communications in Statistics - Theory and Methods*, vol. 46, no. 15, pp. 7301–7312, 2017.
- [34] G. Wang and D. Zhang, "Innovation practice and development prospect of intelligent fully mechanized technology for coal mining," *Journal of China University of Mining and Technology*, vol. 47, no. 3, pp. 459–467, 2018.
- [35] D. G. ChiragMongia and ShankarSehgal, "Vibration response-based condition monitoring and fault diagnosis of rotary machinery," *Neurocomputing*, vol. 434, pp. 98–125, 2021.
- [36] Y. Hu, Study on Roadheader's Vibration Characteristics Based on Working Conditions in Coal Mine, China University of Mining & Technology, Beijing, China, 2015.
- [37] S. Wang, "Design of EBZ160T type digging and exploration machine," *Taiyuan Research Institute China Coal Technology* and Engineering Group, vol. 12, no. 40, pp. 110–112, 2019.