

Research Article

Health Assessment of High-Speed Train Running Gear System under Complex Working Conditions Based on Data-Driven Model

Chao Cheng ^{1,2,3}, Ming Liu ¹, Bangcheng Zhang ⁴, Xiaojing Yin,⁴ Caixin Fu,² and Wanxiu Teng²

¹School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China

²National Engineering Laboratory, CRRC Changchun Railway Vehicles Co., Ltd., Changchun 130062, China

³Department of Automation, Tsinghua University, Beijing 100084, China

⁴School of Mechatronic Engineering, Changchun University of Technology, Changchun 130012, China

Correspondence should be addressed to Bangcheng Zhang; zhangbangcheng@ccut.edu.cn

Received 13 April 2020; Revised 15 June 2020; Accepted 23 June 2020; Published 23 July 2020

Guest Editor: Qian Zhang

Copyright © 2020 Chao Cheng 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.

It is very important for the normal operation of high-speed trains to assess the health status of the running gear system. In actual working conditions, many unknown interferences and random noises occur during the monitoring process, which cause difficulties in providing an accurate health status assessment of the running gear system. In this paper, a new data-driven model based on a slow feature analysis-support tensor machine (SFA-STM) is proposed to solve the problem of unknown interference and random noise by removing the slow feature with the fastest instantaneous change. First, the relationship between various statuses of the running gear system is analyzed carefully. To remove the random noise and unknown interferences in the running gear systems under complex working conditions and to extract more accurate data features, the SFA method is used to extract the slowest feature to reflect the general trend of system changes in data monitoring of running gear systems of high-speed trains. Second, slowness data were constructed in a tensor form to achieve an accurate health status assessment using the STM. Finally, actual monitoring data from a running gear system from a high-speed train was used as an example to verify the effectiveness and accuracy of the model, and it was compared with traditional models. The maximum sum of squared resist (SSR) value was reduced by 16 points, indicating that the SFA-STM method has the higher assessment accuracy.

1. Introduction

With the continuous improvement of the safety and stability of high-speed trains, the study of the assessment of the health status of the running gear system has received extensive attention in recent years [1–3]. As a critical component to withstand and transmit various loads from the vehicle body and route, while mitigating its dynamic effect, the running gear system is quite prone to failure after a long period of high-speed operation. This requires a health status assessment of the running part to improve its safety and reliability. However, there are many unknown interferences and many noises in the working environment of the running

gear system; thus, it is difficult to accurately assess the status of the running gear system by only using the original monitoring data. Therefore, to reduce the influences of unknown interference and noise under complex conditions and to enhance the accuracy of the assessment, in this study, we designed a data-driven health assessment model to ensure the safety, stability, and reliability operation of running gear systems.

The running gear system of high-speed trains is a complicated electromechanical system composed of many components. Any component may suffer from cracks, corrosion, leakage, and other faults, which can cause its faster degradation [4–6]. The extrusion wear of the wheel

pair and the stiffness degradation of the steel spring will also lead to the degradation of the running gear system. Additionally, the monitoring data collected by the sensor under interference conditions contain noise, which will have a great impact on the health status assessment of the running gear system. Therefore, a health assessment of the running gear system has the following two characteristics:

- (1) There are a lot of random noises caused by sensor factors and unknown interferences caused by internal or external environmental factors in the monitoring data of the running gear system
- (2) Structural complexity due to the close coupling of components in the running gear system is difficult to be described by precise mathematical models through mechanism analysis, which will limit the use of analytical models

Currently, the common methods for health status assessment of complex electromechanical systems are mainly divided into three categories: the method based on semi-quantitative information [7–9], data-driven method [10–13], and model-based method [14–16]. With an increase in the number of sensors in the running gear system, it has become very easy to obtain a large amount of data that can reflect the actual status of the system. Additionally, with the continuous development of feature extraction methods in recent years, data-driven methods have been able to extract feature information from massive amounts of data. Currently, many experts and scholars use data-driven methods to assess the health status of complex electromechanical systems. In [17], studies have combined the odd-even space method with the recursive least square algorithm to propose a fault detection and assessment method for a quadrotor UAV based on a linear time-varying system. With the analytic hierarchy process (AHP), which is a kind of traditional decision-making assessment method [18], Qian et al. [19] have suggested a kind of electric power dispatching control scheme based on health status assessments, and through the AHP, this was carried out to perform a wind turbine health status assessment. This method can improve the operation efficiency of a wind farm to the greatest extent and reduce the fatigue load of the fan fault. Through continuous improvement of the hidden Markov model (HMM) [20, 21], Liu et al. [22] proposed a discrete hidden Markov model (DHMM) based on K-means clustering, and the K-means clustering algorithm was adopted to filter the sample points that are inconsistent with the actual class labels, which could better detect and isolate faults. Recently, deep learning technology has been widely used in status assessment [23, 24], and Liang et al. [25] proposed a convolutional neural network (CNN) method, which can be applied for status assessment of a gearbox. Using a massive amount of training data, a high assessment accuracy can be achieved. However, none of the above methods take into account the problem of accurate extraction of original data features. Too many data features will generate redundant variables and increase the complexity of the algorithm, thus affecting the accuracy of the assessment model.

For this reason, the Hilbert–Huang transform (HHT) and the support vector machine (SVM) have been used in the engine fault intelligent diagnosis method (EFD), which uses HHT to extract features and effectively and provides engine fault diagnosis [26]. Song et al. [27] combined statistical filtering (SF) and the wavelet packet transform (WPT) to propose a new method of signal feature extraction and fault diagnosis for a low-speed mechanical system. Principal component analysis (PCA) is the most classical method used for feature extraction [28], and one study constructed subspaces in different directions using PCA analysis of the principal components, to divide the original feature space into several subspaces automatically, and developed a monitoring and assessment scheme for the model. Moreover, Jiang et al. [29] presented a distributed fault detection and isolation method based on fault-related variable selection and Bayesian reasoning. They use an optimization algorithm to determine the optimal subset of variables for each fault, build a sub-PCA model in each subset, and combine the monitoring results of each subset through Bayesian reasoning. This method significantly reduces the redundancy and complexity and thus improves the monitoring performance. To sum up, there are many status assessment methods based on data that have improved the assessment accuracy after feature extraction, but, for actual work environments, there are many unknown disturbances combined with the considerable noise from the complex electromechanical systems. The traditional feature extraction method is difficult to carry out accurately for feature extraction, which affects the accuracy of the assessment model.

Therefore, a new process monitoring method based on SFA was applied in one study [30, 31] to extract the feature with the slowest change from the original monitoring data, and the process monitoring based on slow data was used to distinguish the deviation between dynamic anomalies and normal operating conditions. The purpose of the SFA method in data processing is to find mapping functions of instantaneous scalar input and output from multidimensional monitoring data so that the output slowness data changes as slowly as possible. Additionally, it carries information reflecting the general trend of the system change and filters the slowness data according to the slowness. The filtered data often represents the short-term noise of complex system changes, and the remaining data more accurately reflect the general trend of system change. Therefore, this paper proposes a SFA-STM method to assess the health status of the running gear systems of high-speed trains. The STM is a classification algorithm based on tensor data, which is very suitable for dealing with nonlinear and non-Gaussian problems. In [32, 33], bearing monitoring data was constructed into a tensor matrix and the method based on STM was applied to fault diagnosis of the bearings. Because the tensor data form retains the space and time form of the original data, the data are fully utilized, which can prevent information loss caused by the multidimensional feature of vectorization, providing a good diagnosis accuracy. In [34], the authors proposed a method based on a hybrid support tensor for the diagnosis and positioning of open circuit faults

of modular multilevel converters, and the classification accuracy was better than that of the support vector machine. Therefore, in this study, we used SFA to extract the slowness data from the monitoring data of the running gear system and to eliminate the noise data. Then, the slowness data reflecting the general trend of the process change was constructed into a tensor matrix and input into the STM model, to enhance the accuracy of STM for the health status assessment of the running gears of high-speed trains. This solves the problem of the many unknown interferences and random noises in monitoring data under actual working conditions, which affect assessment results.

This paper is arranged as follows. Section 2 introduces the structure of the running gear system of a high-speed train and describes the problem. In Section 3, the health assessment model for the running gear system in a high-speed train based on SFA-STM is proposed. Section 4 describes a practical case to verify the method proposed in this paper. Section 5 provides the conclusions of this paper.

2. Preliminaries

This part briefly introduces the running department and describes the problems to be solved.

2.1. Description of Running Gear System. As shown in Figure 1(a), the running gear system is a complex coupling system composed of multiple components that is located between the high-speed train body and the track and pulls the vehicle along the track. As the core component of a high-speed train, it is mainly composed of a frame, axle box, suspension device, driving device, brake device, and sensor. During the manufacturing phase, many sensors are preinstalled in the running gear system to monitor the status of the running gear system. Each sensor is integrated with multiple sensor units to monitor different physical quantities and collect different types of data, such as temperature, vibrations, and impact. Therefore, a single sensor is also called a composite sensor component. A total of 11 composite sensors are installed in the monitoring bearing support area, and the installation direction should be consistent with the direction of the impact signal. The specific distribution location is shown in Figure 1(b). A1–A4: measurement point of the axle box bearing. B1–B3: motor bearing measurement points (2 is the measurement point of the motor drive bearing and 3 is the measurement point of the motor rotor). C1–C4: gearbox measurement point (3 and 4 are measuring points of small and large gears at the motor end). The distribution of sensors is very complex, and the many sensors can easily obtain considerable monitoring data reflecting the status of the running gear system during the process of train operation. However, because of the complex monitoring environment, the considerable noise and interference factors cannot be ignored; thus, it is particularly important to remove noise and interference from the monitoring data.

To better determine the health status of a running gear system, it is necessary to monitor the temperature, vibrations, impact, speed, and other physical quantities of the

running gear system. For this reason, in this study, we analyzed three subsystems of the running gear system, including the axle box, gearbox, and motor, and we evaluated the status of the running gear system by using the monitoring data from these three subsystems.

2.2. Problem Description. When a high-speed train runs at high speed for a long time, the gearbox bearing is prone to cracks and deformation. Additionally, the wheelset of the running gear system will be squeezed and worn when it touches the ground for a long time, and this will be corroded by different environments while staying at a high-temperature status [35]. The air spring also has the possibility of gas leakage, and the stiffness degradation of the shock absorber steel spring can occur. All these factors will cause degradation of high-speed trains. From the distribution of a large number of complex sensors, the running gear system status can be determined based on monitoring data. However, because the complex engineering monitoring environment and running environment are complex, the monitoring data is mixed with large amounts of random noise and uncertainty owing to disturbances. This seriously affects the monitoring data, affecting the accuracy of direction of actual running status, and there are two key reasons for this situation [34]:

- (1) Environmental factors: this is mainly due to the operation environment of high-speed trains and the sensor environment. In the high-speed running of high-speed trains, it is inevitable that they will be affected by road condition factors, interior factors, climate factors, and other disturbances, which cause the sensor monitoring data to fail to reflect the actual status of the running gear system over a short period. This can affect the accuracy of the status assessment of a running gear system.
- (2) Sensor factor: this is mainly determined by the sensor's principle, material, manufacturing process, and other physical characteristics. After long working hours, the quality and function of a sensor will be reduced to different degrees, and a large amount of random noise will be generated, which will continuously affect the monitoring performance. As a result, some errors in the information will be recorded, and accurate monitoring data cannot be obtained, which will seriously affect the accuracy of the status assessment of the running gear system.

According to the above analysis, monitoring data are mainly affected by these two factors, while the traditional feature extraction method has difficulty in removing the influence of uncertainty disturbances and random noise. Thus, a new approach is therefore required to address this issue.

3. Methodology

To accurately assess the health status of high-speed train running gear system, a health status assessment model based

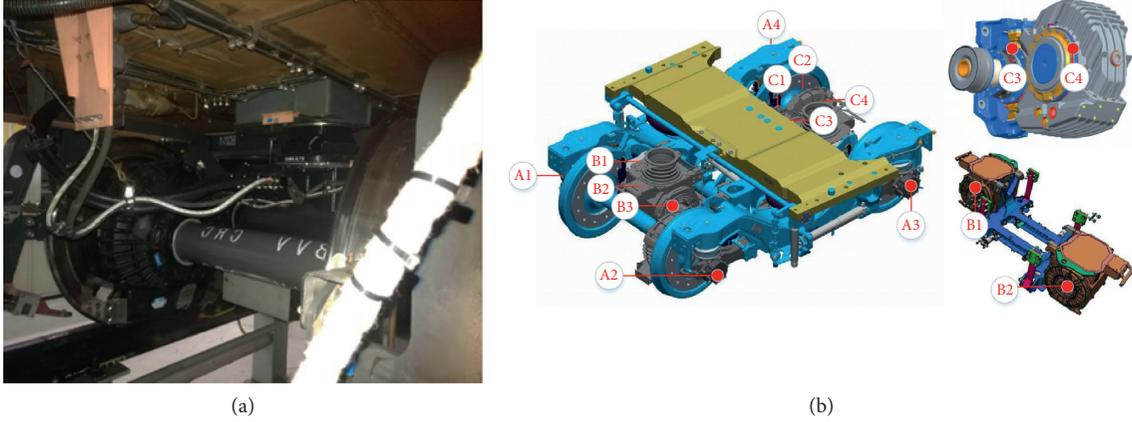


FIGURE 1: Structure diagram of high-speed train running gear system. (a) Real structure diagram of high-speed train running gear and (b) CATIA (computer-aided three-dimensional interactive application) structure diagram of high-speed train running section.

on SFA-STM is proposed. As shown in Figure 2, the model structure is mainly divided into three parts. First, the features reflecting the general trend of system changes are selected as the input for the SFA. Second, the monitoring data processed by SFA is constructed into a tensor matrix. Third, the STM health status assessment model is constructed using the tensor matrix.

3.1. SFA. Data reflecting the changing trend of high-speed train running gear systems is covered by random noise and unknown interference, and when these data are used to assess the health status of a running gear system, it will seriously affect the accuracy of the assessment results. Thus, SFA is the most effective method to reduce noise and interference.

The purpose of SFA data processing is to determine the instantaneous scalar input-output mapping function from the multidimensional time input signal so that the change in the output signal is as slow as possible while carrying the information reflecting the general trend of the system change. Mathematically, the purpose of an SFA is for a given m -dimensional input signal:

$$\mathbf{x}(t) = [x_1(t), \dots, x_m(t)]^T. \quad (1)$$

Find the eigenfunction $\{g_j(\cdot)\}_{j=1}^m$ of a series of slow feature sets so that

$$\min_{g_j(\cdot)} \langle \dot{s}_j^2 \rangle_t, \quad (2)$$

under constraint

$$\langle s_j \rangle_t = 0 \text{ (zero mean)}, \quad (3)$$

$$\langle s_j^2 \rangle_t = 1 \text{ (unit variance)}, \quad (4)$$

$$\forall i \neq j: \langle s_i s_j \rangle_t = 0 \text{ (decorrelation and order)}, \quad (5)$$

to minimize, where $\{s_j(t) = g_j(x(t))\}_{j=1}^m$ is a series of slow features, and $\Delta(s_j) \doteq \langle \dot{s}_j^2 \rangle_t$ is a measure of slowness of $s_j(t)$.

$\langle \cdot \rangle_t$ is the time average, and \dot{s} is the first derivative of s with respect to time.

For SFA, each slow feature $s_j(t)$ is a linear combination of input variables:

$$s_j = g_j(\mathbf{x}) = \mathbf{w}_j^T \mathbf{x}, \quad (6)$$

where $\{\mathbf{w}_j\}_{j=1}^m$ represents the coefficient vector. The mapping from $\mathbf{x}(t)$ to $\mathbf{s}(t)$ can be abbreviated as follows:

$$\mathbf{s} = \mathbf{W}\mathbf{x}, \quad (7)$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_m]^T$ is the coefficient matrix to be optimized by SFA.

Substitute equation (6) into constraint (3), and we obtain

$$\langle s_j \rangle_t = \mathbf{w}_j^T \langle \mathbf{x} \rangle_t = 0. \quad (8)$$

If the input variable \mathbf{x} is scaled to a zero mean ahead of time, constraint (3) is automatically satisfied.

For the solution of slow feature s , the following steps are required, after the singular value decomposition (SVD) of the covariance matrix $\langle \mathbf{x}\mathbf{x}^T \rangle_t$, and it can be obtained:

$$\langle \mathbf{x}\mathbf{x}^T \rangle_t = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T. \quad (9)$$

Next, the whitening transformation can be expressed as

$$\mathbf{z} = \mathbf{\Lambda}^{-1/2} \mathbf{U}^T \mathbf{x} \doteq \mathbf{Q}\mathbf{x}, \quad (10)$$

where $\mathbf{Q} = \mathbf{\Lambda}^{-1/2} \mathbf{U}^T$ is the whitening matrix. You can easily get $\langle \mathbf{z}\mathbf{z}^T \rangle_t = \mathbf{Q} \langle \mathbf{x}\mathbf{x}^T \rangle_t \mathbf{Q}^T = \mathbf{I}$ and $\langle \mathbf{z} \rangle_t = \mathbf{0}$, so the purpose of the SFA is to further translate into finding a matrix $\mathbf{s} = \mathbf{P}\mathbf{z}$ that satisfies $\mathbf{P} = \mathbf{W}\mathbf{Q}^{-1}$ because

$$\mathbf{s} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{Q}^{-1} \mathbf{z} \doteq \mathbf{P}\mathbf{z}. \quad (11)$$

Then, it can be seen that constraints (4) and (5) can be simply written as follows:

$$\langle \mathbf{s}\mathbf{s}^T \rangle_t = \mathbf{I}. \quad (12)$$

Take equation (11) and bring it to equation (12):

$$\langle \mathbf{s}\mathbf{s}^T \rangle_t = \mathbf{P} \langle \mathbf{z}\mathbf{z}^T \rangle_t \mathbf{P}^T = \mathbf{P}\mathbf{P}^T = \mathbf{I}, \quad (13)$$

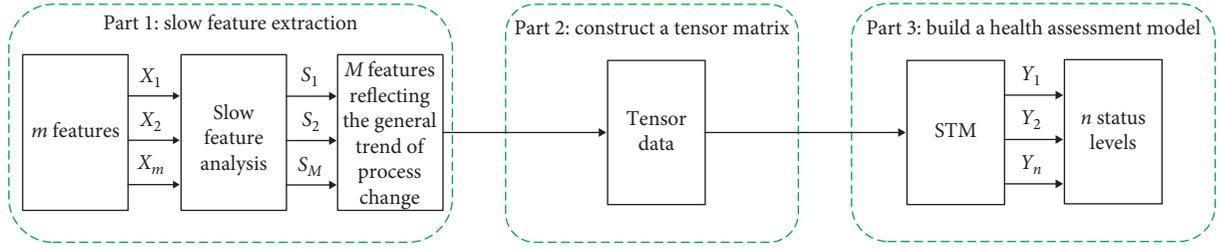


FIGURE 2: Flow chart of the health status assessment model of high-speed train running gear system.

so we know that \mathbf{P} is an orthogonal matrix. Therefore, the optimization problem of SFA can be simplified to find the orthogonal matrix \mathbf{P} , thus minimizing $\mathbf{s} = \mathbf{Pz}$ to $\langle \dot{s}_j^2 \rangle_t$. Such problems can be solved by SVD of the covariance matrix $\langle \dot{z}\dot{z}^T \rangle_t$:

$$\langle \dot{z}\dot{z}^T \rangle_t = \mathbf{P}^T \mathbf{\Omega} \mathbf{P}. \quad (14)$$

Therefore, the orthogonal eigenvector $\{P_j\}_{j=1}^m$ and the corresponding eigenvalue $\{\omega_j\}_{j=1}^m$ can be obtained and verified:

$$\begin{aligned} \langle \mathbf{ss}^T \rangle_t &= \mathbf{P} \langle \mathbf{zz}^T \rangle_t \mathbf{P}^T = \mathbf{I}, \\ \langle \dot{s}_j^2 \rangle_t &= \mathbf{p}_j^T \langle \dot{z}\dot{z}^T \rangle_t \mathbf{p}_j = \omega_j. \end{aligned} \quad (15)$$

Finally, the change matrix \mathbf{W} can be calculated as

$$\mathbf{W} = \mathbf{P} \mathbf{\Lambda}^{-1/2} \mathbf{U}^T, \quad (16)$$

so the slow feature s can be calculated as

$$\mathbf{s} = \mathbf{Pz} = \mathbf{P} \mathbf{\Lambda}^{-1/2} \mathbf{U}^T \mathbf{x} = \mathbf{Wx}. \quad (17)$$

The statistical characteristics of the slow feature can be expressed by the following equation:

$$\begin{aligned} E\{\mathbf{s}\} &= \mathbf{0}, \\ E\{\mathbf{ss}^T\} &= \mathbf{I}, \\ E\{\dot{s}\} &= \mathbf{0}, \\ E\{\dot{s}\dot{s}^T\} &= \mathbf{\Omega}. \end{aligned} \quad (18)$$

According to different slowness features, the slowness features are extracted according to the following equation:

$$M = \text{card}\{s_i | \Delta(s_i) < \max_j^q \{\Delta(x_j)\}\}, \quad (19)$$

where $\text{card}\{\cdot\}$ represents the number of elements in a set and $\Delta(s_i) = \langle \dot{s}_j^2 \rangle_t$ is the $s_i(t)$ slowness measure. Set $q = 0.1$, and $\max_j^q \{\Delta(x_j)\}$, here as the upper quantile of the set $\{\Delta(x_j)\}$, is an evaluation criterion for the slowness of $s_i(t)$, which can

filter out slowness of the M largest $s_i(t)$, expressed as $s_d = [s_1, \dots, s_M]^T$. These slow features reflect the general trend of system changes in the running gear system. For the $M_e = m - M$ slow features screened out, they represent the short-term noise and unknown disturbance of the system changes. The above proofs have been given in [36].

3.2. STM Model. Using the SFA model in the previous section, data reflecting the general trend of the system change of the running gear system was successfully extracted. Then, the nonlinear model between the input and output needs to be built by the STM, to achieve an accurate assessment of the health status of running gear systems of high-speed trains.

Consider a dichotomy problem, where the training set is

$$\begin{aligned} T &= \{(\chi_1, y_1), (\chi_2, y_2), \dots, (\chi_l, y_l)\}, \\ y_i &\in \{+1, -1\}, \end{aligned} \quad (20)$$

where l represents the number of samples, $\chi_i \in R^{I_1 \times I_2 \times \dots \times I_N}$ represents the i input data, y_i represents the class label of the i data, and $y_i \in \{+1, -1\}$. The optimization problem of the STM model is

$$\begin{aligned} \min_{w^{(n)}|_{n=1}, b, \xi} J(w^{(n)}|_{n=1}, b, \xi) &= \frac{1}{2} \left\| \sum_{n=1}^N w^{(n)} \right\|_F^2 + C \sum_{i=1}^l \xi_i, \\ \text{s.t. } y_i \left(\chi_i \prod_{n=1}^N \times_n w^{(n)} + b \right) &\geq 1 - \xi_i, \end{aligned} \quad (21)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l.$$

Based on the idea of supervised tensor learning framework, the STM dichotomy model is decomposed into N suboptimization problems, and the expression form of one of its suboptimization problems after fixing $\{w^{(m)}\}_{m=1}^N$ and $m \neq n$ is

$$\begin{aligned} \min_{w^{(n)}, b^{(n)}, \xi^{(n)}} J(w^{(n)}, b^{(n)}, \xi^{(n)}) \\ = \frac{1}{2} \left\| w^{(n)} \right\|_F^2 \prod_{1 \leq m \leq N}^{m \neq n} \left\| w^{(m)} \right\|_F^2 + C \sum_{i=1}^l \xi_i^{(n)}, \text{ s.t. } y_i \left(\left(w^{(n)} \right)^T \left(\chi_i \prod_{1 \leq m \leq N}^{m \neq n} \times_m w^{(m)} \right) + b^{(n)} \right) \geq 1 - \xi_i^{(n)}, \xi_i^{(n)} \geq 0, \quad i = 1, 2, \dots, l. \end{aligned} \quad (22)$$

In equation (22) above, $w^{(n)} \in R^{I_n}$ ($1 \leq n \leq N$) is the weight of the n hyperplane, $b^{(n)}$ is the offset, $b^{(n)} \in R$ and $\xi^{(n)}$ are the relaxation variables, and C is the penalty factor. What we need to solve are N optimization problems with a similar form to equation (22); for these N problems, we can solve them through the optimization method of alternating projection and finally obtain the decision function of the model as

$$D(\chi) = \operatorname{sgn} \left(\chi \prod_{n=1}^N w^{(n)} + b \right). \quad (23)$$

The alternating projection algorithm requires the input tensor to train the sample set $\{\chi_i \in R^{I_1 \times I_2 \times \dots \times I_N}, y_m \in \{1, -1\}\}_{i=1}^l$, the iterative control threshold ε , and the output $w^{(n)} \in R^{I_n}$ ($1 \leq n \leq N$), $b \in R$.

Step 1: $w^{(n)}$ is initialized as the unit vector in R^{I_n} , $n = 1, 2, \dots, N$.

Step 2: iterate through steps 3 and 4 until the algorithm converges.

Step 3: iterate $w^{(n)}$ for $n = 1, 2, \dots, N$, fix $\{w^{(m)}\}_{m=1}^N$, $m \neq n$, solve equation (22), and get w_n .

Step 4: stop the calculation if

$$\sum_{n=1}^N \left(\left| (w_t^{(n)})^T w_{t-1}^{(n)} \right| \left\| w_t^{(n)} \right\|_F^{-2} - 1 \right) < \varepsilon. \quad (24)$$

is met. Otherwise, go to step 2, where $w_i^{(n)}$ and $w_{t-1}^{(n)}$ represent the values of the previous iteration step and the current iteration step corresponding to the n projection weight vector, respectively.

For the STM multiclassification model, the running gear system status label $L = \{L_1, L_2, \dots, L_s\}$ is defined, and the label is composed of s information classes. One-against-one (OAO) strategy is adopted to construct $S(S-1)/2$ binary STM models to model all possible paired classifications. Then, the decision function $D_{i,j}(\chi)$ between each possible pair of classes L_i and L_j ($i \neq j$) is obtained through dichotomy, and the fractional function

$$\text{Score}_j = \sum_{i=1, i \neq j}^s D_{i,j}(\chi). \quad (25)$$

of the sum of the number of labels allocated to the category by the sample is calculated. The classes with the highest scores were considered predictive labels for unclassified samples.

3.3. The Steps of Health Status Assessment Model for High-Speed Train Running Gear System. As shown in Figure 3, according to the algorithm described above, the steps to summarize the health status assessment algorithm of the high-speed train running gear system are summarized as follows:

Step 1: use SFA to extract the slow feature from the high-speed train running gear system monitoring data.

Step 1.1: normalize the sample data.

Step 1.2: whiten the data through equations (9) and (10).

Step 1.3: take the derivative of the matrix z .

Step 1.4: equations (14) and (16) are used to obtain the mapping matrix.

Step 1.5: slowness data s is obtained by equation (17).

Step 1.6: the slow feature is extracted according to equation (19).

Step 2: the filtered slowness data matrix of the running gear system is constructed into the fourth-order tensor form, as shown in Figure 4.

Step 3: the health status assessment model of the high-speed train running gear system based on STM was constructed.

Step 4: the STM multiclassification model is transformed into multiple binary classification models.

Step 4.1: the optimization problem of STM is divided into N suboptimization problems.

Step 4.2: the N suboptimization problems were solved by alternating projection algorithm, and the decision function $D_{i,j}(\chi)$ was obtained.

Step 5: through equations (23) and (25), the system was assessment.

In the assessment model, the motor temperature, gearbox vibration, and impact of external ring friction of the retaining frame in the axle box were set as the features to assess the health of the running gear system. The specific assessment process will be described through cases in Section 4.

4. Case Study

To verify the accuracy of the SFA-STM assessment model proposed in this paper, this section will describe the high-speed train running gear system as an example for experimental verification. For data collection of the running gear system, the axle box, gearbox, and motor monitoring data of No. 2 carriage and position running gear system of the train during a certain month were selected. To ensure that the collected data was the data when the high-speed train was in operation, the monitoring data with speeds above 1000 r/min were screened to verify the training data and test data for the model, and a total of six monitoring indicators, including temperature, vibrations, and impact were selected for positions 1 and 2, respectively. Owing to the influence of the operating environment on the running gear system of high-speed trains during actual operation and considering the complexity of parameters, such as weight, center of gravity, and the suspension of each component, the health status assessment criteria for the running gear system in the high-speed train were set as “normal,” “general,” and “bad,” as shown in Table 1.

Normal: under this condition, all parts of the running gear system worked normally, fasteners were stable and not loose, and all indicators were within factory

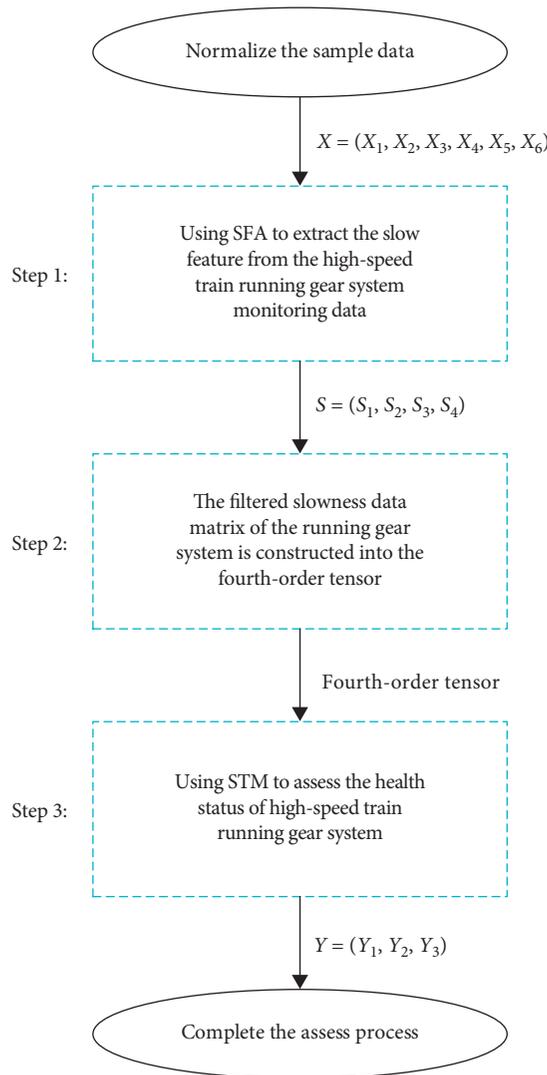


FIGURE 3: The steps of the health status assessment model for the high-speed train running gear system.

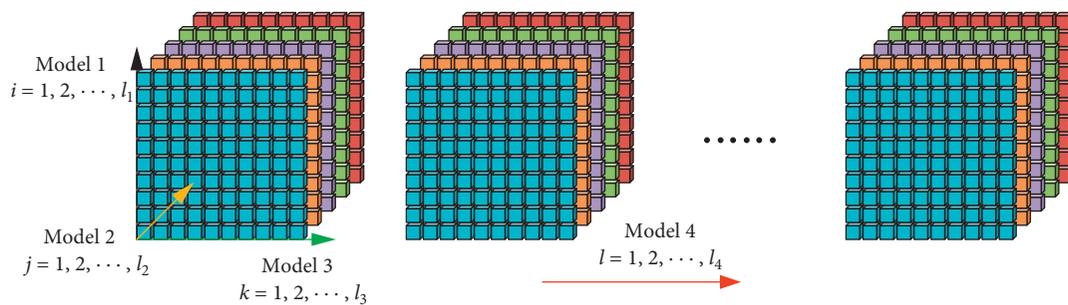


FIGURE 4: Schematic diagram of the fourth-order tensor model.

requirements. The average temperature was less than 30°C, the average vibration peak was less than 22 Hz, the average temperature and amplitude were at a very low levels, and the impact data were also at normal levels, thus ensuring the safe and smooth operation of the high-speed train.

General: under this condition, all parts of the running gear system still worked normally, but the parts were slightly worn and deformed, and all indicators were slightly higher than the factory requirements. The average temperature was in the range of 30–35°C, and the average vibration peak was in the range of 22–25 Hz. Under this condition, the

TABLE 1: The referential points.

Status	Normal	General	Bad
Temperature	<30°C	30–35°C	>35°C
Vibration	<22 Hz	22–25 Hz	>25 Hz

operation of the high-speed train is still unaffected, but the appearance, working status, and overall performance of each part should be checked, and preventive and corrective maintenance work should be performed promptly.

Bad: under this condition, although all parts of the running gear system could still maintain normal operation, the status of some parts was not good, the operation function decreased, and the critical point of health and failure had been reached. The average temperature was higher than 35°C, and the average vibration peak was above 25 Hz. Under this condition, the system should be completely overhauled to replace or repair the parts with a higher degree of damage to improve the safety of the high-speed train.

4.1. Data Preprocessing. Because there are a large number of abnormal environmental noise points in the actual monitoring data of high-speed train running gear systems, it is necessary to preprocess the data. As shown in Figure 5, among them, the short-term fluctuations of temperature, vibration, and shock data are large, and there is much repeated data and abnormal data. Therefore, as shown in Figure 6, the temperature vibration and impact monitoring data of the moving parts at certain positions on the train I and II were preprocessed. The relevant outliers were filtered via mean filtering, and the data were finally compressed to 800 groups. The trend chart is shown in Figure 7.

The temperature sensor was in close contact with the motor element; thus, the external interference was relatively small, whereas the vibration and impact sensors showed more obvious uncertainty in disturbances and random noise owing to the influence of the train speed and acceleration. Although the monitoring data was preprocessed to filter out some of the abnormal value points, the data still included the noise and disturbances as adverse factors, and this could not simply solved by pretreatment. Thus, we needed the following SFA method for further data processing so that the data actually reflected the high-speed train health status.

4.2. Assessment Model. The monitoring data of temperature, vibration, shock, and a total of six characteristics at position I and II were normalized. The slowness data is obtained through the SFA, and the coefficient matrix of SFA optimization is

$$W = \begin{pmatrix} -10.37 & -8.42 & -0.64 & 0.79 & 7.28 & -3.35 \\ 1.65 & 1.66 & 0.58 & -3.15 & -6.92 & 4.06 \\ -13.65 & 4.02 & -0.47 & -0.37 & -0.11 & -0.13 \\ 27.59 & 4.14 & 0.36 & -2.91 & -0.03 & -0.22 \\ -4.78 & -0.34 & -8.85 & 2.69 & 0.39 & 0.34 \\ -0.60 & -0.98 & 9.04 & 2.94 & -0.12 & 0.18 \end{pmatrix}, \quad (26)$$

through equation (19), the feature, $\Delta(s_j)$, was smaller than the maximum $\Delta(x_j)$ of 0.1 times, which was selected to screen out four slow features that reflected the overall changing trend of high-speed trains.

After SFA, 800 sets of monitoring data of six features were transformed into 800 sets of slowness data of four features; 800 sets of data were constructed into the fourth-order tensor model of 400 sets of data. Among them, mode 1 of the tensor is the first slow feature and the second slow feature. The mode 3 of the tensor is the third slow feature and the fourth slow feature. Mode 2 is the slowness data of the adjacent two mileage. Mode 4 is 400 sets of the slowness data. The abovementioned information is shown in Figure 8.

The 400 groups of the fourth-order tensor data, which have been constructed, were divided into 150 training data and 250 test data; then, 150 groups of training data were divided into three groups according to the actual status to construct the training set of the STM model. Finally, the decision functions $D_{1,2}(\chi)$, $D_{1,3}(\chi)$, and $D_{2,3}(\chi)$ between the status of each class were obtained through the STM dichotomy model. The corresponding minimization of energy function is $J_{1,2} = 0.00029988$, $J_{1,3} = 0.00029980$, and $J_{2,3} = 0.00029987$, respectively.

To verify the accuracy of the model, 250 groups of samples to be tested were input into the three decision functions obtained in the STM model, and the class with the highest score for each of them, namely, the prediction label, was calculated through equation (25), to complete the health status assessment of the running gear system.

As shown in Figure 9, the test samples can be accurately classified. Although few samples were misclassified, it can be concluded that the SFA-STM model can accurately assess the health status of the high-speed train running gear system.

4.3. Simulation. To more intuitively represent the reliability of the model presented in this paper, four classical data-driven methods (i.e., SVM, naive Bayes (NB), BP neural network, and hidden Markov model (HMM)) are used to compare and analyze this method. The data adopted by the three methods are the same as the data applied by the method proposed in this paper. All of them are 800 groups of monitoring data processed by mean filtering. However, because the training samples and test samples used in the method proposed in this paper are a total of 400 sets of data, to reflect the authenticity of comparative simulation, it is necessary to take the mean value of each adjacent mileage point in the 800 sets of monitoring data and reduce it to approximately 400 sets of training data. This process does not affect the accuracy of the evaluation model because, in engineering practice, two adjacent mileage points almost reflect the health status of the same running gear system. At the same time, the assessment grades of health status of the three models were also “normal,” “general,” and “bad,” with 150 sets of training data and 250 sets of test data.

SVM is used for comparative validation, as shown in Figure 10(a). In the beginning, the green line fits the blue line well, but in the middle, a large number of green points leave the blue line; in the end, the green line fits the blue line well.

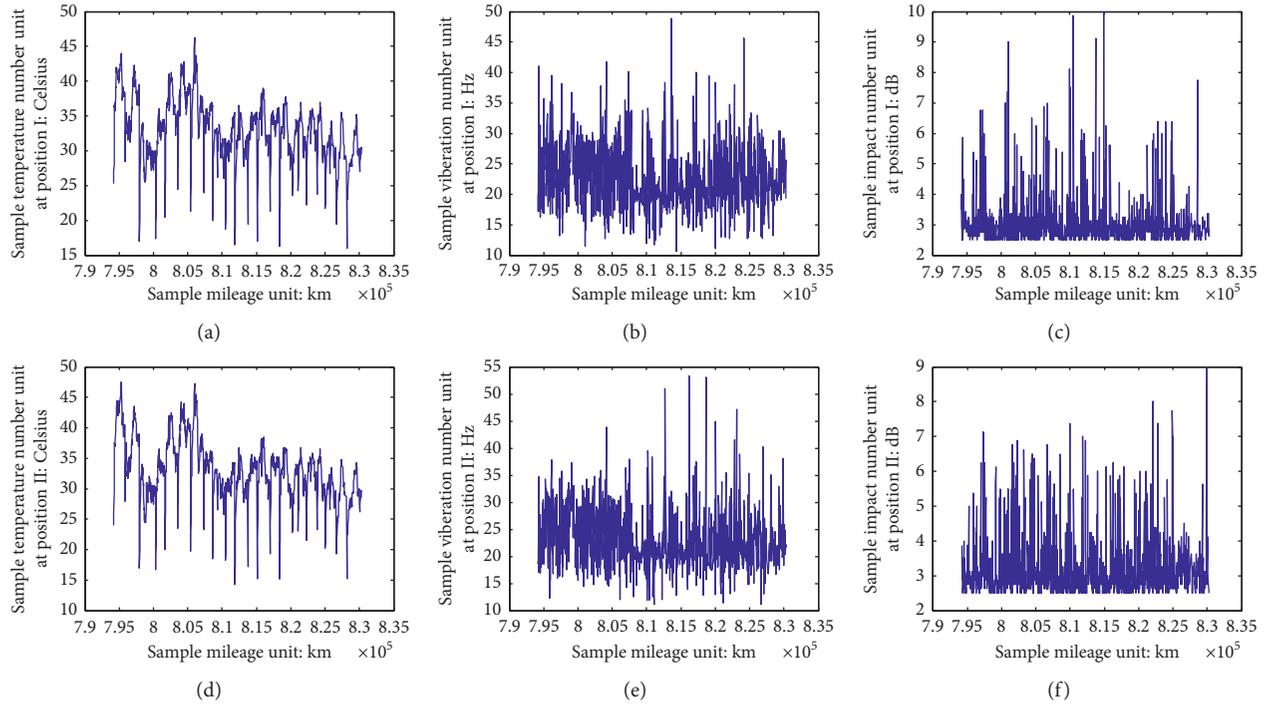


FIGURE 5: Trend chart of original monitoring data of the running gear system. (a) Position I temperature data trend chart. (b) Position I vibration data trend chart. (c) Position I impact data trend chart. (d) Position II temperature data trend chart. (e) Position II vibration data trend chart. (f) Position II impact data trend chart.

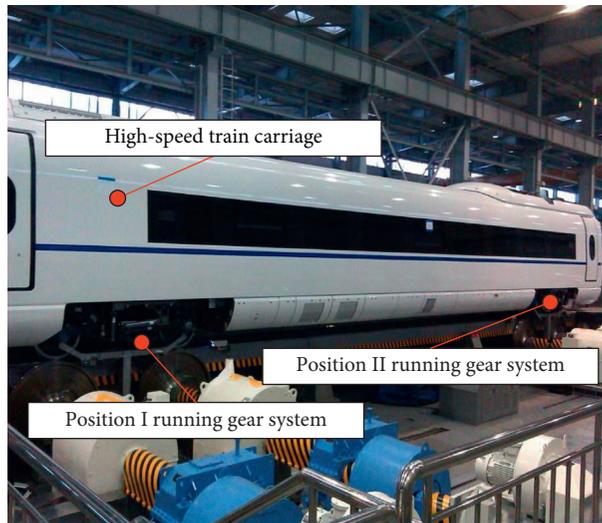


FIGURE 6: Actual position of the running gear system.

This indicates that the running gear system can be accurately assessed by the SVM model under normal and bad statuses, while the running gear system cannot be accurately assessed by the model in general status and at the beginning of bad status, which is a considerable limitation.

NB is used for comparative validation, as shown in Figure 10(b). From the fitting degree of the green line to the blue line in the figure, the NB model is still the same as the SVM model, which has many misjudgments in the middle status of the system and also has a certain degree of

misjudgments in good and bad statuses. Compared with the SFA-STM model, the NB model still has difficulty accurately assessing the health status of the system.

The BP neural network is used for comparative validation, as shown in Figure 10(c). The fluctuation of the green line always follows that of the blue line. Although there is a certain degree of abnormal point fluctuation, the overall level of the BP neural network model is more accurate than that of the SVM model and the NB model in assessing the health status of the running gear system. However, the

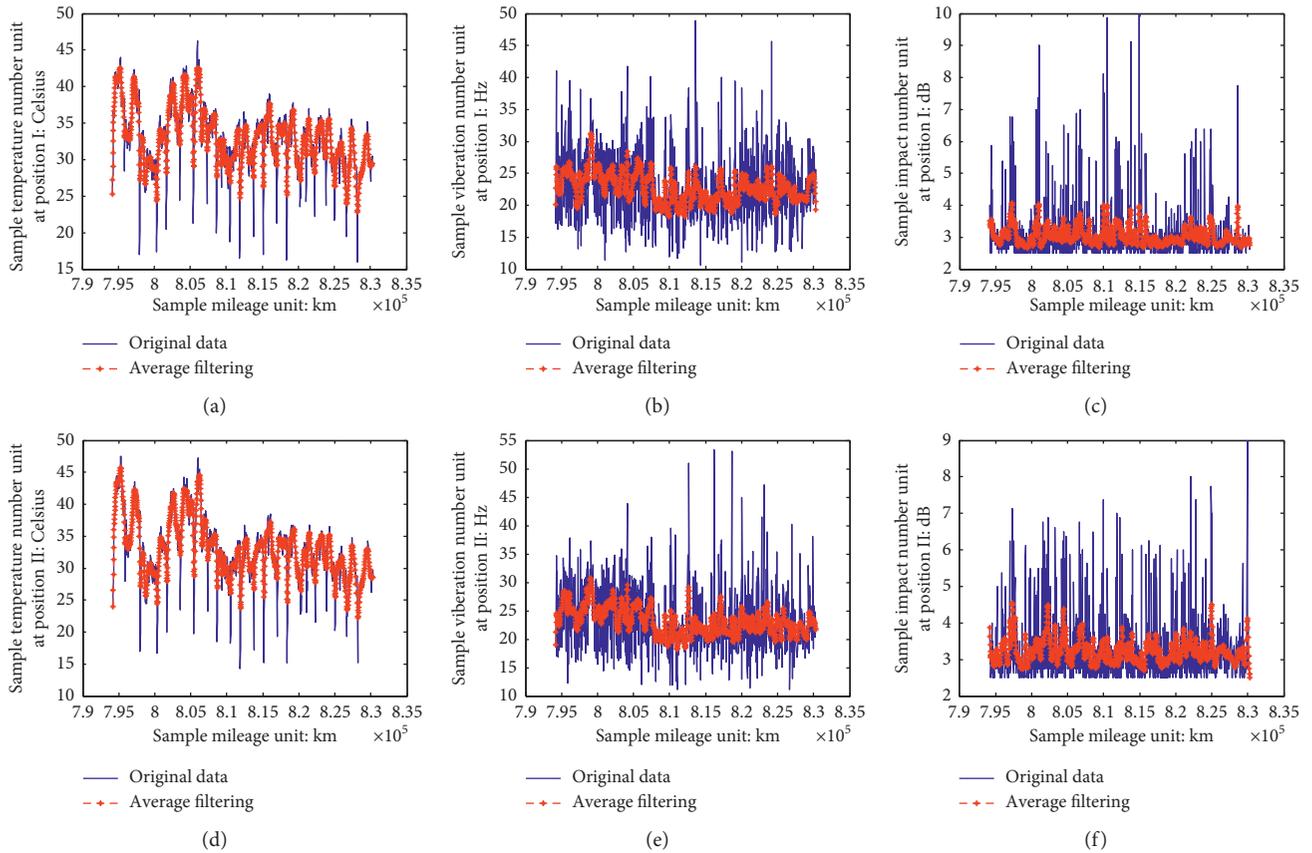


FIGURE 7: Trend chart of mean filter monitoring data. (a) Average filtering position I temperature data trend chart. (b) Average filtering position I vibration data trend chart. (c) Average filtering position I impact data trend chart. (d) Average filtering position II temperature data trend chart. (e) Average filtering position II vibration data trend chart. (f) Average filtering position II impact data trend chart.

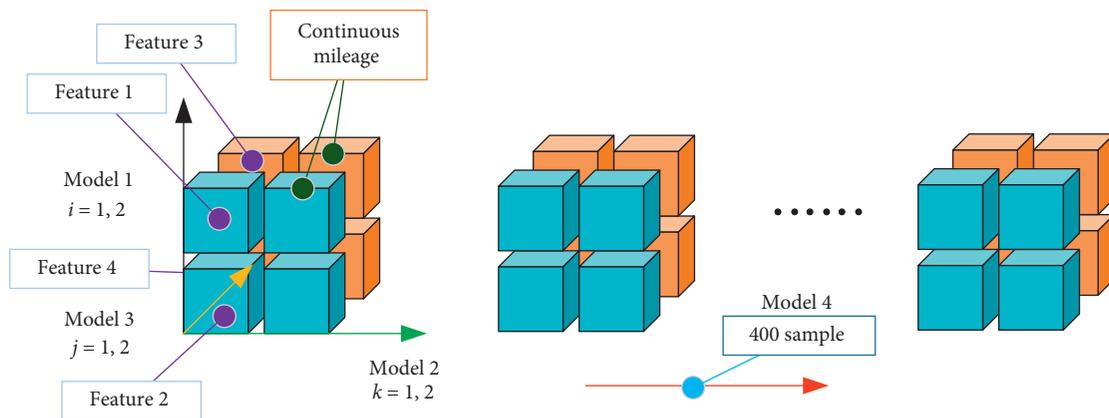


FIGURE 8: Schematic diagram of the tensor model.

abovementioned figure clearly shows the red line of the SFASTM model fits the blue line to a higher degree, which more accurately and intuitively reflects the health status of the system.

HMM is used for comparative validation, as shown in Figure 10(d). From the fitting degree of the green line to the blue line in the figure, the HMM model is still the same as the SVM model, which has a large number of misjudgments in the middle status of the system and also has a certain

degree of misjudgments in good and bad statuses. Because both the NB and HMM models are based on Bayesian classification algorithms, it is not surprising to obtain the same classification results as NB. In general, the HMM model still has difficulty accurately accessing the health status of the running gear system compared with the SFASTM model.

To directly compare the accuracy of the five methods, three indexes are used for comparison. First, the maximum

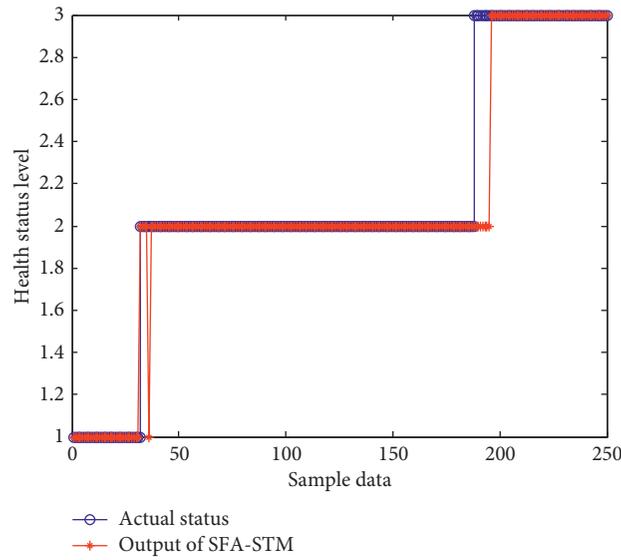


FIGURE 9: The SFA-STM model.

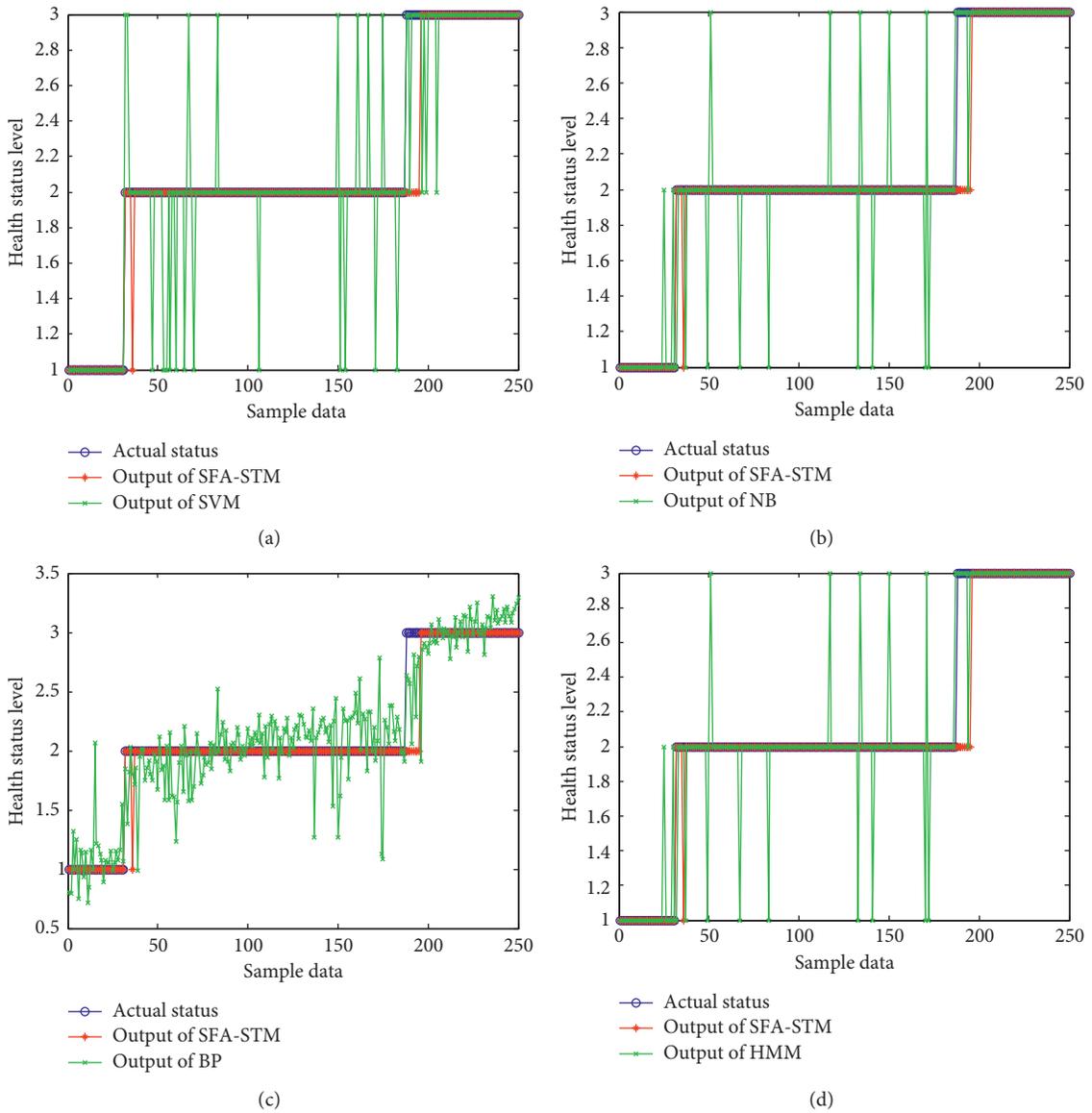


FIGURE 10: Simulation results of the (a) SVM model, (b) NB model, (c) BP model, and (d) HMM model.

TABLE 2: SSR values of the four models.

Model	SSR
SFA-STM	9
SVM	18
NB	25
BP	18.9135
HMM	25

TABLE 3: Table of the training time.

Model	Time (s)
SFA-STM	7.55
SVM	6.43
NB	0.73
BP	0.43
HMM	0.30

TABLE 4: TPR and FPR indicators.

Method	Normal		General		Bad	
	TPR (%)	FPR (%)	TPR (%)	FPR (%)	TPR (%)	FPR (%)
SFA-STM	100	0.47	99.36	8.51	87.3	0
SVM	93.55	4.25	90.38	3.19	98.41	3.41
NB	100	2.02	86.54	4.26	93.65	4.6
HMM	100	2.02	86.54	4.26	93.65	4.6

sum of the squared residuals (SSR) index is adopted, as shown in Table 2. According to equation (27), where y_i represents the status value obtained from the assessment of i sample according to the model, \hat{y}_i represents the actual status value of the i sample, and n represents the number of samples. The SSR index of the four methods is given, among which the SSR value of SFA-STM is the lowest, only 9; the SSR value of SVM is 18; the SSR values of NB and HMM are 25, and the SSR value of BP is 18.9135 [37]. Second, the true positive rate (TPR) and false positive rate (FPR) indexes were used to assess the accuracy of classification. Table 3 shows that the SFA-STM model still has the best effect [38]. In the end, as shown in Table 4, compared with the training time of each model, the training time is relatively long because both SFA-STM and SVM adopt the optimization algorithm. In general, it is seen that the SFA-STM model has the highest accuracy in assessing the health status of the high-speed train running gear system.

$$\text{SSR} = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (27)$$

5. Conclusion

In this study, the SFA-STM model is proposed to assess the health status of the high-speed train running gear system, which solves the problem that there are many unknown

interferences and random noise in the monitoring data under complex working conditions, which affect the assessment results. Using an example, it is shown that the SFA-STM system proposed in this study can accurately reflect the actual health status of the running gear system. Compared with the four types of traditional data-driven models, this method has higher applicability to practical engineering problems and provides a new solution to the problem that the general trend of system changes cannot be extracted by the traditional feature extraction method to assess the health status of the complex system under high noise and multiple disturbances.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 61903047, 61973046, and 61803044 and Jilin Province Development and Reform Commission under Grant no. 2019C040-3.

References

- [1] M. Torabi, S. G. M. Mousavi, and D. Younesian, "A high accuracy imaging and measurement system for wheel diameter inspection of railroad vehicles," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 10, pp. 8239–8249, 2018.
- [2] H. Chen and B. Jiang, "A review of fault detection and diagnosis for the traction system in high-speed trains," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 450–465, 2020.
- [3] H. Chen, B. Jiang, N. Lu, and W. Chen, *Data-driven Detection and Diagnosis of Faults in Traction Systems of High-Speed Trains*, Springer Nature, New York, NY, USA, 2020.
- [4] C. Yang, C. Yang, T. Peng, X. Yang, and W. Gui, "A fault-injection strategy for traction drive control systems," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 7, pp. 5719–5727, 2017.
- [5] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Gui, "Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1559–1567, 2018.
- [6] H. Chen, B. Jiang, W. Chen, and Z. Li, "Edge computing-aided framework of fault detection for traction control systems in high-speed trains," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1309–1318, 2020.
- [7] L. Jin, C. Peng, and T. Jiang, "System-level electric field exposure assessment by the fault tree analysis," *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no. 4, pp. 1095–1102, 2017.

- [8] F. Sihombing and M. Torbol, "Parallel fault tree analysis for accurate reliability of complex systems," *Structural Safety*, vol. 72, pp. 41–53, 2018.
- [9] Y. Dong, J. Zhang, Z. Li, Y. Hu, and Y. Deng, "Combination of evidential sensor reports with distance function and belief entropy in fault diagnosis," *International Journal of Computers Communications & Control*, vol. 14, no. 3, pp. 329–343, 2019.
- [10] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient fault detection with smoothing techniques in statistical process monitoring," *Control Engineering Practice*, vol. 62, pp. 11–21, 2017.
- [11] H. Chen, B. Jiang, and N. Lu, "Data-driven incipient sensor fault estimation with application in inverter of high-speed railway," *Mathematical Problems in Engineering*, vol. 2017, Article ID 8937356, 13 pages, 2017.
- [12] H. Chen, B. Jiang, W. Chen, and H. Yi, "Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4716–4725, 2019.
- [13] W. Deng, R. Yao, H. Zhao, X. Yang, and G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm," *Soft Computing*, vol. 23, no. 7, pp. 2445–2462, 2017.
- [14] C. Cheng, W. Wang, H. Luo, B. Zhang, G. Cheng, and W. Teng, "State-degradation-oriented fault diagnosis for high-speed train running gears system," *Sensors*, vol. 20, no. 4, p. 1017, 2020.
- [15] H. Luo, M. Krueger, T. Koenings, S. X. Ding, S. Dominic, and X. Yang, "Real-time optimization of automatic control systems with application to BLDC motor test rig," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 4306–4314, 2017.
- [16] Y. Hagh, R. Asl, and V. Cocquempot, "A hybrid robust fault tolerant control based on adaptive joint unscented Kalman filter," *ISA Transactions*, vol. 66, pp. 262–274, 2017.
- [17] W. Han, Z. Wang, and Y. Shen, "Fault estimation for a quadrotor unmanned aerial vehicle by integrating the parity space approach with recursive least squares," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 232, no. 4, pp. 783–796, 2018.
- [18] Y. Liu, W. Ni, and Z. Ge, "Fuzzy decision fusion system for fault classification with analytic hierarchy process approach," *Chemometrics and Intelligent Laboratory Systems*, vol. 166, pp. 61–68, 2017.
- [19] P. Qian, X. Ma, D. Zhang, and J. Wang, "Data-driven condition monitoring approaches to improving power output of wind turbines," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 8, pp. 6012–6020, 2019.
- [20] S. Wang, J. Xiang, Y. Zhong, Y. Zhou, and Y. Zhou, "Convolutional neural network-based hidden Markov models for rolling element bearing fault identification," *Knowledge-Based Systems*, vol. 144, pp. 65–76, 2018.
- [21] Y. Gao, F. Vilecco, M. Li, and W. Song, "Multi-scale permutation entropy based on improved LMD and HMM for rolling bearing diagnosis," *Entropy*, vol. 19, no. 4, p. 176, 2017.
- [22] J. Liu, Q. Li, W. Chen, and T. Cao, "A discrete hidden Markov model fault diagnosis strategy based on K-means clustering dedicated to PEM fuel cell systems of tramways," *International Journal of Hydrogen Energy*, vol. 43, no. 27, pp. 12428–12441, 2018.
- [23] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, 2017.
- [24] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.
- [25] K. Liang, N. Qin, D. Huang, and Y. Fu, "Convolutional recurrent neural network for fault diagnosis of high-speed train bogie," *Complexity*, vol. 2018, Article ID 4501952, 13 pages, 2018.
- [26] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert-Huang transform and support vector machine," *Applied Acoustics*, vol. 75, pp. 1–9, 2014.
- [27] L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 4, pp. 1887–1899, 2018.
- [28] Z. Ge and Z. Song, "Distributed PCA model for plant-wide process monitoring," *Industrial and Engineering Chemistry Research*, vol. 52, no. 5, pp. 1775–2194, 2013.
- [29] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 1, pp. 377–386, 2016.
- [30] J. Huang, O. K. Ersoy, and X. Yan, "Slow feature analysis based on online feature reordering and feature selection for dynamic chemical process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 169, pp. 1–11, 2017.
- [31] S. Zhang and C. Zhao, "Slow-feature-analysis-based batch process monitoring with comprehensive interpretation of operation condition deviation and dynamic anomaly," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3773–3783, 2019.
- [32] Z. He, J. Cheng, J. Li, and Y. Yang, "Linear maximum margin tensor classification based on flexible convex hulls for fault diagnosis of rolling bearings," *Knowledge-Based Systems*, vol. 173, pp. 62–73, 2019.
- [33] L. Bo, G. Xu, X. Liu, and J. Lin, "Bearing fault diagnosis based on subband time-frequency texture tensor," *IEEE Access*, vol. 7, pp. 37611–37619, 2019.
- [34] C. Li, Z. Liu, Y. Zhang, L. Chai, and B. Xu, "Diagnosis and location of the open-circuit fault in modular multilevel converters: an improved machine learning method," *Neurocomputing*, vol. 331, pp. 58–66, 2019.
- [35] J. Yang, Y. Yue, R. Zhu, W. Chen, and M. Li, "Dynamic characteristics of encased differential gear train with journal bearing," *Mathematical Problems in Engineering*, vol. 2020, Article ID 2436191, 15 pages, 2020.
- [36] C. Shang, F. Yang, X. Gao, X. Huang, J. A. K. Suykens, and D. Huang, "Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis," *AIChE Journal*, vol. 61, no. 11, pp. 3666–3682, 2015.
- [37] F. Vilecco, "On the evaluation of errors in the virtual design of mechanical systems," *Machines*, vol. 6, no. 3, p. 36, 2018.
- [38] A. Formato, D. Guida, D. Ianniello, F. Vilecco, T. Lenza, and A. Pellegrino, "Design of delivery valve for hydraulic pumps," *Machines*, vol. 6, no. 4, p. 44, 2018.