



Deep learning-based fault diagnostic network of high-speed train secondary suspension systems for immunity to track irregularities and wheel wear

Yunguang Ye¹ · Ping Huang^{2,3} · Yongxiang Zhang²

Received: 7 March 2021 / Revised: 3 August 2021 / Accepted: 4 August 2021 / Published online: 20 October 2021
© The Author(s) 2021

Abstract Fault detection and isolation of high-speed train suspension systems is of critical importance to guarantee train running safety. Firstly, the existing methods concerning fault detection or isolation of train suspension systems are briefly reviewed and divided into two categories, i.e., model-based and data-driven approaches. The advantages and disadvantages of these two categories of approaches are briefly summarized. Secondly, a 1D convolution network-based fault diagnostic method for high-speed train suspension systems is designed. To improve the robustness of the method, a Gaussian white noise strategy (GWN-strategy) for immunity to track irregularities and an edge sample training strategy (EST-strategy) for immunity to wheel wear are proposed. The whole network is called GWN-EST-1DCNN method. Thirdly, to show the performance of this method, a multibody dynamics simulation model of a high-speed train is built to generate the lateral acceleration of a bogie frame corresponding to different track irregularities, wheel profiles, and secondary suspension faults. The simulated signals are then inputted into the diagnostic network, and the results show the correctness and superiority of the GWN-EST-1DCNN method. Finally, the 1DCNN method is further validated using tracking data of a CRH3 train running on a high-speed railway line.

Keywords High-speed train suspension system · Fault diagnosis · Track irregularities · Wheel wear · Deep learning · Literature review

1 Introduction

As railway transportation is developing at a considerable speed in many regions worldwide, condition monitoring of high-speed trains is receiving increasing attention, in which, failures of suspension systems will increase the vibration of vehicle components and reduce the running stability, and may even lead to severe accidents, such as derailment [1, 2]. Therefore, it is of critical importance to diagnose the faults of railway vehicle suspension systems.

1.1 Literature review

Currently, the approaches for fault detection or isolation (FD/I) of railway vehicle suspension systems can be classified into two main categories, i.e., the model-based approach and data-driven approach [3].

1.1.1 Model-based approach

The first reported approaches for railway vehicle suspension fault diagnosis (RVSFD) are mainly model-based approaches. This type of approach usually requires the development of a sophisticated dynamic model to determine the relationship between faulty states and vehicle responses. Data collected by the sensors are then fed into such models to predict the corresponding vehicle dynamic

✉ Ping Huang
ping.huang@swjtu.edu.cn

¹ Institute of Land and Sea Transport Systems, Technical University of Berlin, Berlin 10587, Germany

² National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China

³ Institute for Transport Planning and Systems, ETH Zurich, 8093 Zurich, Switzerland

responses. The outputs of the model are compared with the real-time measurement data, and the residual between the measured data and the prediction data is designated to identify failures [4]. As shown in Table 1, the model-based approaches with different strategies for FD/I of railway vehicle suspensions over the past two decades can be mainly classified into four sub-categories [5]: (1) Kalman filter-based (KF-based); (2) interacting multiple models-based (IMM-based); (3) Rao–Blackwellised particle filter-based (RBPF-based); and (4) recursive least squares-based (RLS-based).

(1) KF-based [10–18]: Kalman filter (KF) is an effective tool to estimate state variables for a dynamic system. An earlier study using KF in fault detection and isolation (FD&I) of vehicle suspension systems is the literature [10]. In this work, a 2D half-vehicle model, which considered the lateral and yaw motions of two wheelsets and one bogie as well as the lateral motion of the carbody, was built. Based on the vehicle model with 7 degrees of freedom (DOFs), a KF-based method was proposed to detect and isolate the faults of the secondary suspension (anti-yaw damper and secondary lateral damper). The results showed that the KF-based method was computationally efficient and could identify the abrupt faults of vehicle suspension systems. In [11], a light rail vehicle (LRV) model consisting of three wagons, which considered two DOFs of each carbody (bounce and pitch motions) and one DOF of each bogie (bounce motion), was built. Based on the 9-DOF-LRV model, a KF-based method was used to generate residuals for fault isolation of the primary and secondary suspensions (dampers and springs). In [12], based on a vehicle model with the same structure as in [10], a

hybrid extended Kalman filter-based (HEKF-based) approach combined with a nonlinear residual generator was proposed for FD&I of the secondary suspension (anti-yaw damper and secondary lateral damper). In [13, 14], a three-dimensional vehicle model with 46 DOFs was built, and a multiple Kalman filter-based (MKF-based) approach was applied for FD&I of the secondary suspension system (anti-yaw damper, secondary vertical damper, and secondary lateral damper). With this MKF-based method, high robustness against track uncertainties can be achieved. In [15], based on a 7-DOF-model of ERRI B176 benchmark vehicle, a linear KF scheme was employed to diagnose faults of the secondary vertical suspension. Moreover, it was stated that this method can be used for condition monitoring of secondary suspension instead of calendar-based maintenance. In [16], based on a two-mass oscillator with 2 DOFs (i.e., one-eighth of the entire vehicle model), a cubature Kalman filter-based (CKF-based) approach was proposed to diagnose the faults of the secondary vertical suspension.

(2) IMM-based [19–21]: In [19, 20], the IMM approach was proposed to detect the failure of the secondary suspension (secondary lateral damper and secondary lateral spring) of the same vehicle as described in [10]. This approach is similar to the KF-based approaches, while it additionally includes mode mixing. Parallel KFs are no longer separated but interact with each other. The input state-space vector at each time step for a given filter is a combination of the output state-space vector of all filters at the previous time step. This combination is based on the mode likelihood and given transition probabilities

Table 1 Summary of approaches used in RVsFD

Categories	References	Sub-categories
Model-based	Li et al. [10], Wei et al. [11], Jesussek et al. [12], Jesussek et al. [13], Jesussek et al. [14], Onat et al. [15], Zoljic-Beglerovic et al. [16], Wei et al. [17], and Xu et al. [18]	KF-based
	Hayashi et al. [19], Hayashi et al. [20], and Mori et al. [21]	IMM-based
	Li et al. [22], Li et al. [23], and Li et al. [24]	RBPF-based
	Liu et al. [25], and Liu et al. [26]	RLS-based
	Lebel et al. [5], Wei et al. [6], Xue et al. [7], Wei et al. [8], and Liu [9]	Others
Data-based	Wei et al. [8], Mei et al. [29], Mei et al. [30], Dumitriu et al. [31], Aravanis et al. [32], Aravanis et al. [33], Aravanis et al. [34], Aravanis et al. [35], Sakellariou et al. [36], Sakellariou et al. [37], and Rossouw et al. [38]	SM-based
	Qin et al. [39], Kulkarni et al. [40], Wei et al. [41], and Hong et al. [42]	ML-based
	Ye et al. [3], and Gasparetto et al. [27]	HM-based
	Wu et al. [43]	DL-based

between modes. Based on the above studies, a model updating procedure was proposed in [21] to adapt the baseline model when a fault was detected, and to allow for the identification of simultaneous faults. In this work, a 2D vehicle model with 7 DOFs (the same as described in [10]) and a three-dimensional (3D) vehicle model with 34 DOFs were, respectively, built, and the IMM approach was used to detect the faults in the secondary suspension (secondary lateral damper and secondary lateral spring).

- (3) RBPF-based [22–24]: Another classical model-based approach reported in the earlier literature is the RBPF approach. For instance, an RBPF-based approach was proposed in [22] to estimate the secondary suspension parameters (anti-yaw damper and secondary lateral damper) of a half-vehicle model with 7 DOFs. The experimental results showed that the RBPF-based method was more promising than the traditional EKF-based approach. The RBPF-based method, however, is usually computationally expensive, and it is thus more suitable to be used in cases where the detection time is of minor importance. Similar studies were presented in [23, 24]. Although this approach relies on MKF, it is different from the previous approaches since the associated model is not selected in advance to represent a fixed fault type and magnitude [5].
- (4) RLS-based [25, 26]: The above model-based approaches used in RVSFD can find traces of KF. More precisely, they are more or less based on KF. In [25], KF was not used, but a closely related time-domain filter known as RLS was adopted instead. RLS is an algorithm equipped with the memory and machine learning features and has the capacity to identify multiple parameters simultaneously from an input–output linear system by filtering the error signal between the measured and simulation outputs. In this work, a 3D vehicle model with 42 DOFs was simulated. The field test data from an E464 locomotive were adopted to validate the feasibility of this approach. The results showed that this approach was promising in RVSFD. A similar study was presented in [26].

One of the biggest merits of the aforementioned model-based approaches is that through mathematical modeling, and the relationship between the input as a faulty state and the output that can reflect the system dynamic behavior can be clearly established [27]. This can help researchers or even field staff to clearly understand the diagnostic model, which is helpful for engineering applications. However, the following issues currently limit the development of these model-based approaches:

- *High modeling difficulty.* Vehicle dynamics models are challenging to be accurately built, mainly due to two causes [8]: (1) Train suspension systems are often nonlinear, and it is usually extremely difficult to obtain the detailed and accurate parameters of the nonlinear elements, such as dampers, and springs; (2) In train dynamics simulation, it is difficult to consider the elasticity of the carbody, bogie, wheelset, etc.
- *High hardware cost.* The above model-based approaches all require the use of a relatively large number of sensors, which makes the hardware used in RVSFD rather expensive and raises concerns about the reliability of the transducers. For instance, the minimum number of sensors used in [24] is 3, and more in [28].
- *Low computing efficiency.* Dynamics simulation involves a large number of nonlinear force calculations and iterative computations, especially when a complicated vehicle–track coupling system needs to be considered, resulting in low calculation efficiency.

In conclusion, the model-based approaches have great potential in RVSFD, but the corresponding vehicle model needs to be accurately established, and the calculation efficiency should also be improved.

1.1.2 Data-driven approach

The data-driven approach does not rely on vehicle simulation models but requires historical tracking data and prior training. As shown in Table 1, the data-driven approaches with different strategies for FD/I of railway vehicle suspensions over the past two decades can be mainly classified into four sub-categories: (1) statistics model-based (SM-based); (2) traditional machine learning-based (ML-based); (3) hybrid model-based (HM-based); and (4) deep learning-based (DL-based).

- (1) SM-based [29–35]: A classical SM-based approach used in RVSFD is based on the cross-correlation function. In [29], the cross-correlation function between the accelerations of two bogies was applied to determine the health conditions of vehicle suspension systems. The basic idea of this approach is that a faulty element in vehicle suspensions can alter the symmetry of a vehicle with a symmetrical configuration, which results in a coupling relationship between motions that can be observed in the cross-correlation function. Aiming at identifying the failure of the vertical primary suspension, this work analyzed the impact of the faulty damper on the correlation between the bounce, pitch, and roll acceleration signals. Similar studies were presented in [29, 30]. Actually, if this approach is used to identify different

failures of vehicle suspension systems, such as the primary vertical spring failure and secondary vertical damper failure, different cross-correlation functions corresponding to different faults must be observed and counted. This approach, therefore, is also a kind of statistical method. Another interesting method based on a stochastic functional model (FM) of the system dynamics under varying payload was postulated in [32]. Using the model-induced parameter space, the healthy system state under variable operating conditions was represented by a certain parameter subspace, which was constructed in an initial learning phase. In the inspection phase, fault detection was achieved by checking whether the current system dynamics belonged to the healthy parameter subspace or not. Moreover, a conventional statistical time series detection method was also introduced in [32] for comparison purposes, and the experimental results showed the superiority of the FM-based method. Similar studies were presented in [33–37]. In [38], a nonparametric, statistical time-series-based method was proposed to characterize the primary and secondary suspension faults of a self-steering three-piece MKV bogie. The method made use of changes in the vibration signal spectrum, and a verified dynamic simulation model was developed to generate vehicle suspension acceleration response for the healthy and faulty states. The result showed that, with this method, the damage of the primary, secondary, and stabilizer springs could be detected.

- (2) ML-based [39–42]: Traditional ML methods are mainly composed of two steps: (a) feature extraction; and (2) pattern recognition. In [39], a wavelet entropy-support vector machine (SVM)-based approach was proposed to diagnose the faults in high-speed train suspension systems. More specifically, the wavelet entropy of the bogie frame accelerations was extracted as the feature that reflects the fault states of suspension systems, and SVM was then adopted to classify these different faults. In this work, three types of vehicle suspension faults, including yaw dampers removal (YDR), lateral dampers removal (LDR), and air springs removal (ASR), were considered. In [40], the features used to characterize different signals were dominating frequency along with the corresponding relative damping coefficient, root mean square (RMS) of lateral bogie frame acceleration and mean ratio of axlebox acceleration and bogie frame acceleration. Two classifiers (i.e., linear SVM and Gaussian SVM) were used for the FD&I of yaw dampers of high-speed trains. The simulation results showed that both of these classifiers could identify the faulty yaw dampers

well. Moreover, the Gaussian SVM classifier performed slightly better in the training and testing phases, while it had a higher risk of overfitting to the current dataset. In [41], to diagnose the faults of the lateral suspension system of railway vehicles, four time-domain features (mean, standard deviation, skewness, and kurtosis) and three frequency-domain features (frequency center (FC), root mean square frequency (RMSF), and root variance frequency (RVF)) of the bogie lateral accelerations were extracted. After that, three classifiers (Dempster–Shafer (D-S) evidence theory, Fisher discrimination analysis (FDA), and SVM) were applied to the fault classification, respectively. The results showed that the three classifiers all could classify the faults with a high accuracy, in which, the D-S evidence theory outperformed the other two classifiers. In [42], to monitor the stiffness and damping coefficients of the vehicle suspension systems of high-speed trains in real-time, the position, height, and width of the largest peak in magnitude frequency of the axlebox accelerations were considered as the input features. The classifier used in this work was a multi-output support vector machine (MSVR). Besides, it was also stated that unlike the model-based approaches, this data-driven approach did not rely on accurate dynamics models.

- (3) HM-based [3, 27]: HM-based method refers to the combination of signal processing methods and ML-based methods. For instance, in order to diagnose the faults in the secondary suspensions system of high-speed trains, a feature extraction method based on multiscale permutation entropy and linear local tangent space alignment (MPE-LLTSA) was proposed in [3]. More specifically, a preliminary high-dimensional feature matrix was constructed using MPE, and LLTSA was then used to reduce the dimensionality for obtaining a low-dimensional feature matrix. The classifier used in this work was a multi-class SVM. The results showed that the MPE-LLTSA-SVM method could accurately recognize the secondary suspension faults when the track irregularities and wheel profiles were relatively constant. However, the robustness against track irregularities and wheel wear was not well solved. In [27], the random decrement technique (RDT) was used to extract the free response of the bogie frame lateral accelerations. The output of the RDT was then analyzed using the Prony method to identify the characteristic exponents of the system. In the fault classification step, two classifiers were compared, i.e., artificial neural networks (ANN), and k-nearest

neighbor (k -NN), and the k -NN classifier were proved to be more reliable than the ANN classifier.

- (4) DL-based [43] In recent years, DL methods have begun to be applied to various industries, but research on RVSFD is relatively scarce. In [43], to diagnose the faults of suspension systems of high-speed trains, three synchrony measurements (instantaneous phase synchrony, amplitude envelope synchrony, and composite synchrony) were applied to estimate the similarity between bogie acceleration signals, and a synchrony group convolutional network was proposed for feature extraction and pattern classification of the multichannel monitoring system. The effectiveness of the method was validated by a simulation dataset.

One of the biggest merits of the aforementioned data-driven approaches is that they do not rely on sophisticated dynamics models or high-fidelity simulations. In particular, the data-driven approaches combine the power of data analysis and engineering domain knowledge to generate a model that can be trained quickly and adapted easily to different vehicle suspension systems [42]. Moreover, two model-based approaches (robust observer, and the KF combined with the generalized likelihood ratio test (GLRT)) and two data-driven approaches (dynamical principal components analysis (DPCA), and the dynamical canonical variate analysis (CVA)) were, respectively, introduced in [8] to detect the faults in the primary and secondary suspensions of an urban rail vehicle. The comparison results showed that the data-driven approaches outperformed the model-based approaches from the perspective of modeling, computational efficiency, and accuracy. However, data-driven approaches currently used in RVSFD still face the following challenges:

- *Database establishment.* No matter from the perspective of data statistics or from the perspective of model training, data-driven approaches require a massive amount of historical tracking data, which is a common problem facing the entire industry.
- *Development of adaptive fault feature extraction method.* As far as ML-based approaches are concerned, the first step is to extract features that reflect the fault status of vehicle suspension systems. In reality, the vehicle suspension systems have many nonlinear components [44–47], including springs, dampers, etc. The nonlinear factors of the vehicle components, usually, result in acquired signals that contain multiple natural oscillation modes, especially when multi-faults are coupled together [3]. As a result, it is difficult to characterize these nonlinear signals by using traditional single time-domain or frequency-domain feature extraction methods [48–51].

- *High sensitivity of collected signals to track irregularities and wheel wear.* The running of a vehicle on a track is achieved through the wheel–rail contact. Track irregularities will seriously affect wheel–rail contact, such as contact area, contact force, and affect the vibration signals used for condition monitoring [52]. More importantly, the wheel profile will continuously change as the mileage increases due to the presence of wear [53, 54], which will seriously affect the vibration signals. In short, the high sensitivity of the collected signals to track irregularities and wheel wear could affect the robustness of data-driven approaches.

1.2 Motivation

With the advent of the era of big data, railway companies have started to establish related databases. Specifically, high-speed trains are usually equipped with a large number of sensors, and it is easy to acquire tracking data from these trains, which lays a solid foundation for the study, application, and promotion of data-driven approaches in RVSFD.

The train suspension system is a highly nonlinear system [3]. As described in Sect. 1.1.2, it is difficult to accurately characterize these nonlinear signals with traditional single feature extraction methods [55]. To overcome this problem, we have proposed a feature extraction method of MPE-LLTSA in [3] to RVSFD, which can realize the feature extraction of signals at multiple scales. However, a deep understanding of objects, as well as the corresponding signals, is still a prerequisite, and extensive expertise and data analysis capabilities are required for building this method. Therefore, a simple and adaptive feature extraction method that can overcome the nonlinear interference of the vehicle system is required. Under the background of the era of big data, DL is a powerful tool that has been successfully applied in many industries, such as image recognition [56], earthquake prediction [57], transportation planning [58], fault diagnosis of rotating machinery [59], multibody dynamics simulation (MBS) [59]. Exploring its possibility in fault diagnosis of railway vehicles is a topic of big interest. Motivated by this, this paper aims at developing a DL-based fault diagnosis method for RVSFD.

As described in Sect. 1.1.2, track irregularities and wheel wear will affect the vibration signals used for condition monitoring of railway vehicles. Therefore, the developed fault diagnosis method for train suspension systems must be guaranteed to be immune to changes in track irregularities and wheel wear before being put into use. In our previous work [3], we have briefly discussed the robustness of the diagnostic method caused by wheel wear, but it was not studied in-depth. In addition, the interference

caused by track irregularities was also not analyzed. To improve the previous work, these two issues are the main subjects of this study.

1.3 Contribution and structure of this paper

The main contribution of this work is summarized as follows:

- (1) A vehicle–track model considering track flexibilities, track irregularities, and wheel wear is built.
- (2) Two strategies for improving the robustness of the diagnostic network of train suspension systems are proposed, i.e., a Gaussian white noise strategy (GWN-strategy) and an edge sample training strategy (EST-strategy).
- (3) A DL-based fault diagnostic network for RVSF is established. The diagnostic network consists of three phases: In the first phase (data preprocessing), the GWN-strategy is used to the acceleration signal to make the diagnosis network immune to relatively high-frequency impacts caused by track irregularities. In the second phase (training dataset building), an EST-strategy is proposed to improve the robustness of the diagnostic network against wheel wear. Finally, in the third phase (training and recognition), a GWN-EST-1DCNN-based fault diagnostic network of high-speed train suspension systems is built.

The rest of this paper is structured as follows. In Sect. 2, an MBS model of a high-speed train is built, where three measured track irregularities and seven tracking wheel profiles are presented. In Sect. 3, a DL-based diagnosis network for RVSF is presented. In Sect. 4, the simulation result is presented. In Sect. 5, the presented diagnosis network is validated using the tracking data of a CRH3 train running on a high-speed railway line. Finally, concluding remarks are briefly given in Sect. 6.

2 Vehicle–track coupled model

Here, the description of the vehicle–track coupled model is divided into two parts: vehicle–track model (Sect. 2.1) and track irregularities and wheel wear (Sect. 2.2).

2.1 Vehicle–track model

The vehicle model built in our work consists of three sub-structures, one for the carbody and two for the bogies, where each bogie consists of one bogie frame, two wheelsets, and four axleboxes. These rigid bodies are assembled by primary and secondary suspensions. By assuming a constant running speed of the carbody, the

carbody is considered as 5 DOFs. The bogie and wheelset can be characterized by 6 DOFs each, and each axlebox only rotates relative to the corresponding wheelset, i.e., with one DOF. Finally, the MBS model of the vehicle has 49 DOFs. The final vehicle model simulated in SIMPACK is shown in Fig. 1a. For more information, as well as the main parameters, of the vehicle model, see Ref. [3].

The wheelset is supported by two rails, where Hertzian contact [61] and FASTSIM [62] algorithms are used. Simulating the track structure according to the realistic condition (e.g., ‘rail + rail slab + concrete base + sub-structure’ [63]) would involve a large number of DOFs, thus increasing the computational effort considerably. Therefore, referring to Refs. [64, 65], the track model is simplified as a co-running track with a form of ‘rail + track slab + ground’ [66]. The stiffness and damping of the fastener system are considered between the rail and the track slab, and the stiffness and damping of the cement–asphalt mortar are considered between sleeper and ground, as shown in Fig. 1b. Some parameters of the track model are listed in Table 2.

2.2 Track irregularities and wheel profiles

2.2.1 Track irregularities

The operating environment of a train is complex and changeable, in which track irregularities are often not constant [67–69]. The developed FD&I method must be immune to the disturbance of track irregularities before it is implemented in actual engineering. To investigate the impact of track irregularities on the robustness of the fault diagnosis method, track irregularities measured on three different high-speed railway lines are introduced, namely Wuhang–Guangzhou railway line (WG-line), Beijing–Tianjin railway line (JJ-line), and Zhengzhou–Xian railway line (ZX-line). Figure 2 shows the track irregularities of 1000 m in the whole line.

2.2.2 Wheel wear

When a train is running, the wheel profile will change continuously due to wear, which will always affect wheel–rail contact, including contact force, contact patch size, etc., and further affects the dynamic characteristics of the bogie, including the bogie frame acceleration. The FD&I method may incorrectly attribute the change in the bogie acceleration to a failure of the vehicle suspension system. To analyze the impact of wheel wear on the robustness of the diagnostic method, the wheel profile evolution of a CRH3 high-speed train running on the WG-line is introduced in our work for analysis. The total length of the WG-line is 1,068.8 km, the minimum radius of curve of the line

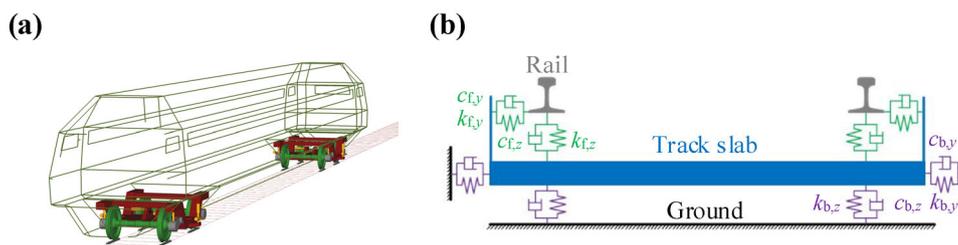


Fig. 1 Models for simulation: **a** vehicle model and **b** track model

Table 2 Primary parameters of the track model

Parameter	Value	Unit
Fastener stiffness (lateral k_{fy} , vertical k_{fz})	30,000, 1,50,000	kN/m
Cement–asphalt mortar stiffness (lateral k_{by} , vertical k_{bz})	70,000, 1,40,000	kN/m
Fastener damping (lateral c_{fy} , vertical c_{fz})	150, 100	kNs/m
Cement–asphalt mortar damping (lateral c_{by} , vertical c_{bz})	350, 1,400	kNs/m
Wheel–rail contact damping	100	kNs/m
Wheel–rail contact algorithm	Hertzian + FASTSIM	–
Wheel–rail friction coefficient	0.35	–
Poisson ratio	0.28	–
Rail cant	1:40	–
Rail profile	CHN60	–

is 7,000 m, the gauge is 1,435 mm, and the rail cant is 1/40. This line situation is a typical example of China high-speed railway lines. In Fig. 3, the new S1002CN profile, as well as the worn S1002CN profiles due to wear as the mileage increases, are presented. It can be seen that the wear at the flange root is mild, and the wear volume is mainly distributed in the range of -25 – 25 mm relative to the nominal rolling circle of the wheel. Eventually, a “hollow wear” that commonly occurs in high-speed trains is developed. More information concerning the tracking measurement of the wheel profile evolution can be found in [53].

3 Diagnostic network design

The description of the diagnostic network is divided into three subsections. In Sect. 3.1, the design of the deep neural network is described. Two strategies to improve the robustness against track irregularities and wheel wear are described in Sect. 3.2, i.e., a GWN-strategy and an EST-strategy. The whole structure of the diagnostic network for railway vehicle suspension systems is described in Sect. 3.3.

3.1 Design of deep neural network

3.1.1 One-dimensional CNN (1DCNN)

A traditional two-dimensional CNN (2DCNN) is designed to take advantage of the spatial features in 2D images by using locally connected and tied-weights convolutional filters that operate on multiple pixels simultaneously rather than a single pixel [56, 70], and this approach can better detect the dependencies between pixels. In a 2DCNN, 2D input data are first converted into 3D data (width, height, and depth), where the depth of 1 for a one-band image and 3 for a three-band image (red, green, blue). Next, a feature map is obtained by multiple applications of convolution operators across sub-regions of the entire image, which first add a bias term and then apply a nonlinear activation function. If the k th feature map at a given layer is represented as h^k , whose filters are determined by weights W_k and bias b_k , the feature map h^k is then expressed by

$$h_{ij}^k = \sigma((W_k * x) + b_k), \quad (1)$$

where $*$ is the convolution operator, $W_k * x = \sum_m \sum_n x(i-m, j-n)W_k(m, n)$, and σ is the activation function.

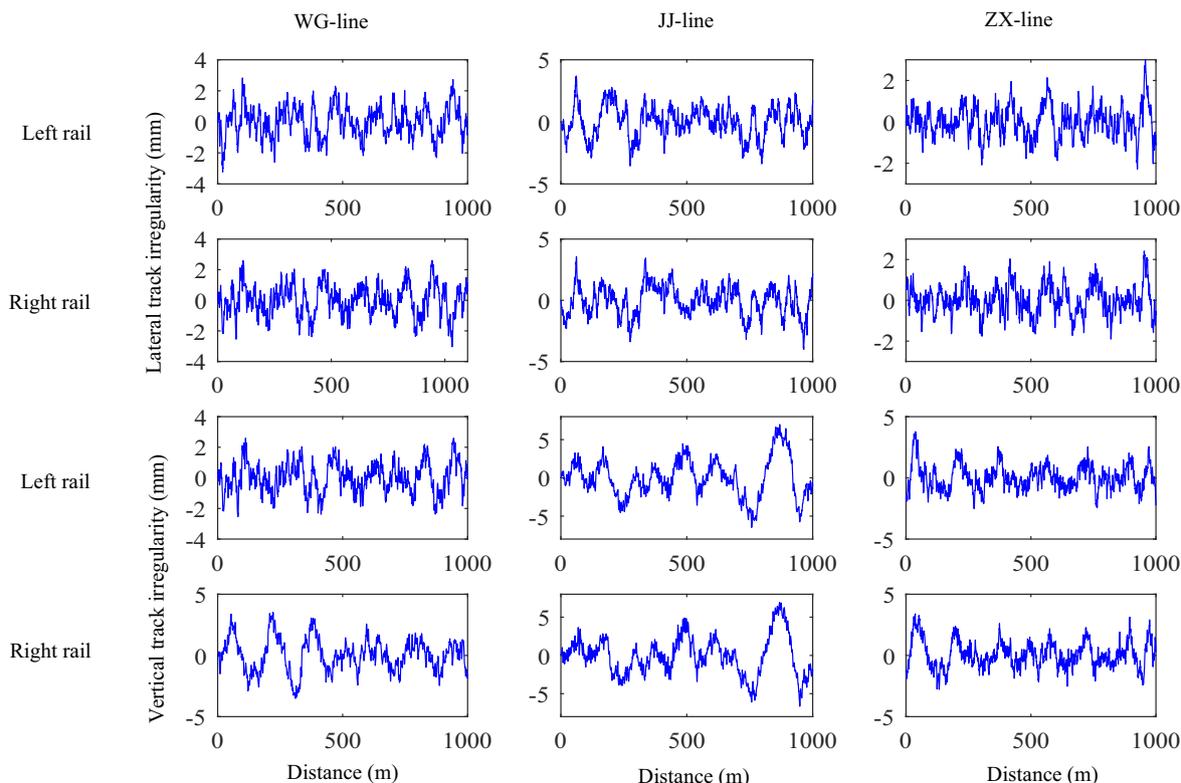


Fig. 2 Three measured track irregularities on WG-line, JJ-line, and ZX-line

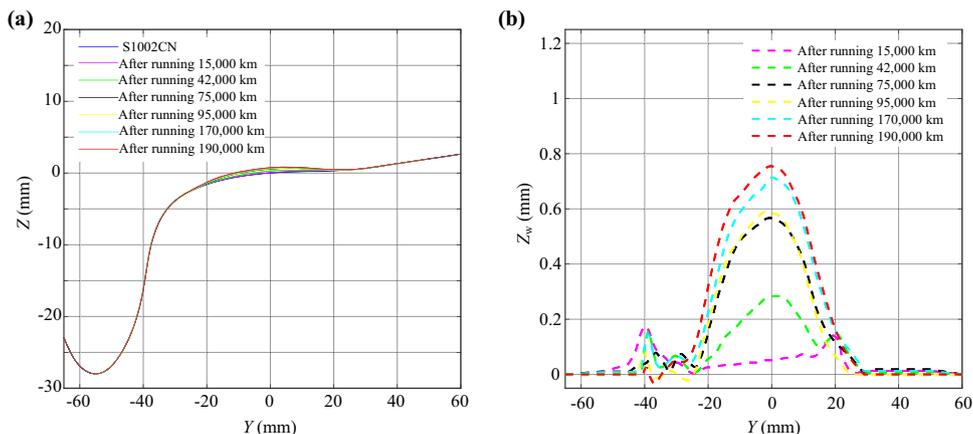


Fig. 3 Measurement of the wheel profile evolution: a measured wheel profiles and b their wear depth distributions with respect to mileage [53]

However, for 1D time-series data, such as the acceleration data used in this paper, 1DCNN is usually a more ideal choice [71]. Figure 4 shows the difference between the 2DCNN and 1DCNN, where applying 2DCNN to a 2D image will generate a 2D image, whereas applying 1DCNN to a 1D image will generate a 1D image. The convolutional filter of the 1DCNN is one-dimensional, which enables it to detect the interdependencies in 1D data.

3.1.2 The Architecture of the designed 1DCNN

The architecture of the 1DCNN model designed in our work is shown in Fig. 5. The proposed model includes 7 main blocks, the first 5 blocks are designed for feature extraction, among which each block consists of a 1DCNN layer, an advanced activation function (AAF) layer, and a max-pooling layer. The last two blocks are designed for

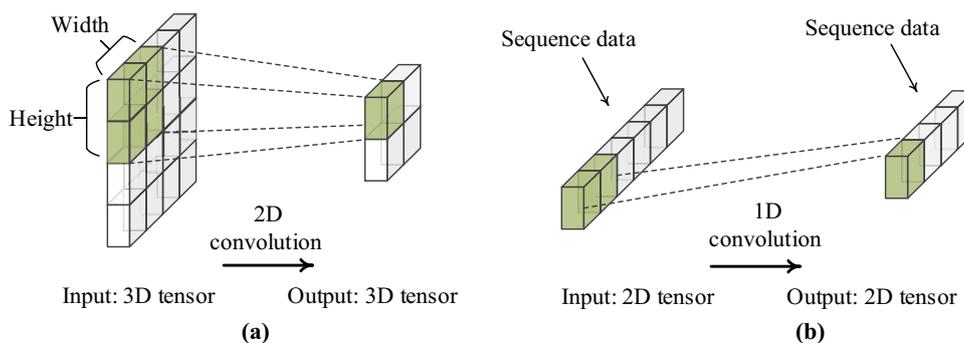


Fig. 4 The differences between convolution networks: **a** 2DCNN and **b** 1DCNN

classification, where each includes a fully connected layer and an activation layer.

For the design of the 1DCNN layer, existing studies [56, 70] have shown that the feature maps should change from wide and shallow to narrow and deep from the input layer to the output layer. This rule has proven to be very effective in many successful CNN models, such as the classic AlexNet [56] and VGGNet [70]. This article, therefore, follows this rule to adjust the number of the convolution kernels in the CNN layer part, i.e., the number of CNN convolution kernels in each subsequent layer is twice that of the previous layer. This strategy can increase the depth of the feature maps from the first block to the last block. In our model, the number of convolution kernels in the first CNN layer (block 1) is set to 32, and the number of convolution kernels in the last CNN layer (block 5) is 512.

For the selection of the activation function, most studies on CNN models use the rectified linear unit (ReLU) function [72]. This function, however, has a disadvantage, that is, a too-large learning rate or gradients could easily lead to the “death” of neurons, and the ReLU function often cannot perform well when the nonlinear relationship of the input dataset is very complicated [73]. The vehicle suspension system is a highly nonlinear system. For signals from such systems, this activation function is obviously not an ideal choice. Therefore, an advanced activation function, parametric rectified linear unit (PReLU) proposed in [73], is used in our work, and the expression is given:

$$f(y_i) = \begin{cases} y_i & \text{if } y_i > 0 \\ a_i y_i & \text{if } y_i \leq 0 \end{cases}, \tag{2}$$

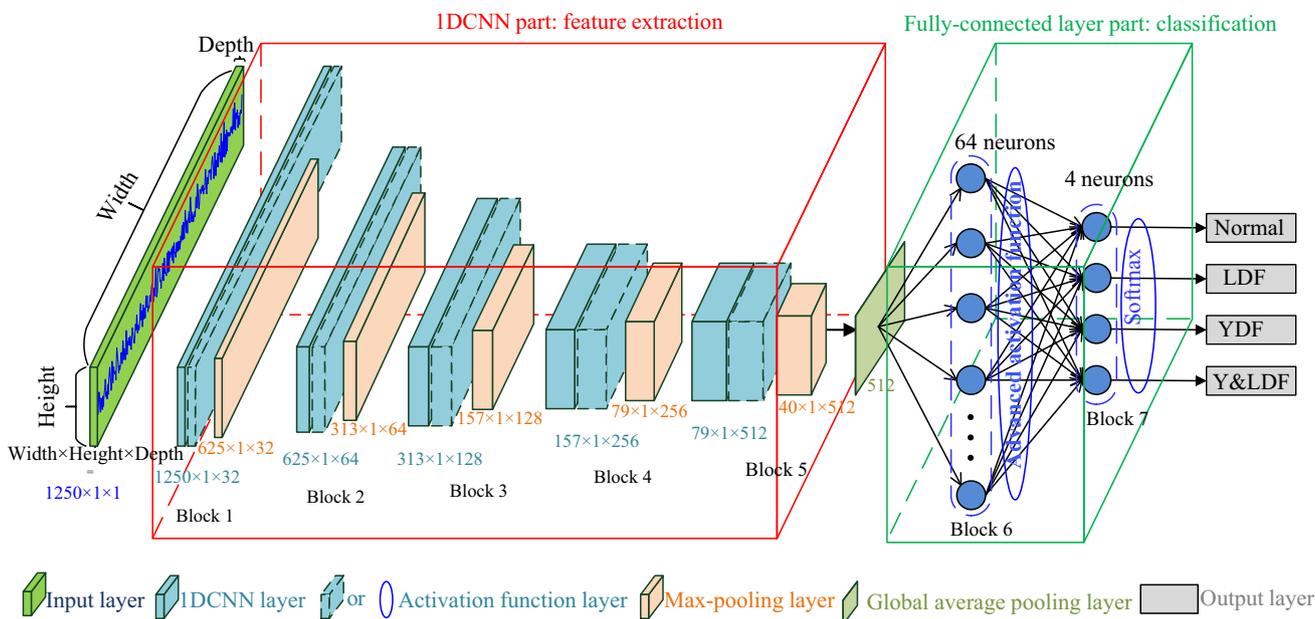


Fig. 5 The designed architecture of the proposed 1DCNN model

where y_i is the input of the activation function on the i th channel, which is the output of 1DCNN in this work; a_i is a coefficient controlling the slope of the negative part, whose value can be automatically learned from the data during the training phase to meet the dataset of different nonlinear relationships.

The design of the pooling layer is critical for the CNN model since it can significantly reduce the model parameters and the time required for training without sacrificing model accuracy [55, 60]. Therefore, after the 1DCNN and activation function layers of each block in the first 5 blocks, a local max-pooling layer is added to extract the key features of the 1DCNN layer output and reduce the model parameters. In addition, the stride of each max-pooling layer is set to 2, which can reduce the width of the feature maps. Further, in order to allow the output results of the 1DCNN model to be the input of the fully connected layer later (block 6), it is also necessary to transform the dimension of the output results of 1DCNN. This article uses the global average pooling layer [74] to replace the original flatten layer, which can further reduce the model parameters and increase efficiency.

In the first fully connected layer (block 6), the number of neurons is set to 64, and the advanced activation function PReLU is used as the activation function. The number of neurons in the last fully connected layer (block 7) is set to 4 since the number of fault categories in this work is 4 (see Sect. 4), and then, the Softmax [75] classification function is used as the activation function to output the model's predicted probability for each type of failure.

3.2 Two strategies for increasing robustness against track irregularities and wheel wear

3.2.1 Gaussian white noise strategy against track irregularities

Track irregularities affect the dynamic characteristics of railway vehicles, including the signals required for suspension fault diagnosis. Under different track irregularity conditions, the amplitude and frequency distribution of acceleration are often different, which may affect the robustness of the fault diagnosis method. Figure 6 shows the lateral acceleration distributions of the bogie frame under three different track irregularity conditions when the yaw damper fails (YDF). It can be clearly seen that although the three signals have roughly the same trend distribution, such as the frequency of the peaks, there are some different relatively high-frequency and low amplitude impact components between the two peaks.

In order to make the fault diagnostic network immune to these relatively high-frequency and low amplitude impact

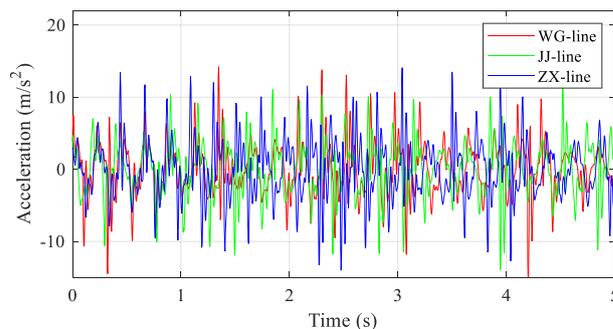


Fig. 6 Lateral acceleration of the bogie frame under three different track irregularity conditions (YDF, S1002CN profile)

components, a strategy of adding Gaussian white noise to the original signal is proposed to overcome these impact components caused by the track irregularities, i.e., the GWN-strategy. This strategy of adding GWN to the raw signal is often used in signal processing and pattern recognition. For instance, it is commonly used in empirical modal decomposition (EMD), and the method named ensemble empirical modal decomposition (EEMD) has been developed based on this [76]. It should be noticed that the amplitude of noise affects the diagnosis accuracy. However, there is no specific equation reported in the literature to guide the choice of the noise amplitude until now. Thus, for an investigated signal, different noise levels should be tried to select the appropriate one. In this paper, after many trials, it is suggested that the amplitude of the added white noise is about 0.2 times the standard deviation of the investigated signal. This value is also suggested when using EEMD proposed in Ref. [76]. It is important to note that such an approach of adding white noise does not eliminate the high-frequency components of the signal, rather it makes all signals have high-frequency components and thus, the diagnostic method is immune to these high-frequency components. Such a similar approach is also commonly used in CNN-based image recognition [77]. The feasibility of this strategy is demonstrated in Sect. 4.

3.2.2 Edge sample training strategy against wheel wear

As described in Sect. 2.2.2, the wheel profile will change as the train running mileage increases. This process is a continuous process, which will always affect the bogie frame acceleration. Figure 7 shows the bogie frame lateral accelerations for a new S1002CN profile, a worn S1002CN profile after running 95,000 km (S1002CN-W95K), and a worn S1002CN profile after running 1,95,000 km (S1002CN-W190K) when the yaw damper fails; it can be clearly seen that as the wear increases, the vibration amplitude of the bogie acceleration also increases. Therefore, using the dataset trained under the new wheel profile

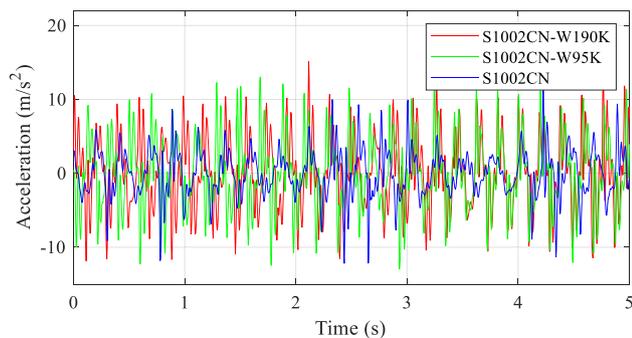


Fig. 7 Lateral acceleration of the bogie frame under three different wheel profiles (YDF, WG-line)

(S1002CN) to identify the dataset under the worn wheel profiles (S1002CN-W95K and S1002CN-W190K) may cause misidentification, i.e., the fault diagnosis method may incorrectly attribute the change in the bogie acceleration to a failure of the vehicle suspension system.

To overcome the above problem, an EST-strategy is proposed in this paper, i.e., during the phase of training dataset establishment, the dataset corresponding to the new wheel profile (S1002CN) and the dataset corresponding to the most worn wheel profile (S1002W190K) are used as the training dataset for the fault diagnosis method. With the EST-strategy, the interference of the wheel wear on the robustness of the fault diagnosis method can be suppressed, and it is not necessary to train the dataset corresponding to each worn wheel profile evolved during the running of the wheel (of course, it is also unrealistic). The feasibility of this strategy is demonstrated in Sect. 4.

3.3 Diagnostic network of railway vehicle suspension systems

Finally, the architecture of the designed diagnostic network of train suspension systems is illustrated in Fig. 8. The whole process is called GWN-EST-1DCNN method, and it consists of the following three phases:

Phase I Data preprocessing. In this phase, firstly, the bogie frame accelerations concerning different faults are collected, and the GWN-strategy described in Sect. 3.2.1 is then applied to the original acceleration signals.

Phase II Training dataset establishment. Based on the ETS-strategy described in Sect. 3.2.2, the samples corresponding to the new wheel profile (S1002CN) and the samples corresponding to the most worn wheel profile (S1002CN-W190K) are chosen as the training dataset for the diagnostic network, and their upper envelopes are extracted.

Phase III Fault diagnosis and visualization. Using the 1DCNN designed in Sect. 3.1.2 to train and classify different kinds of faults, and the final results are visualized by Andrews curve [78].

Andrews curve is a method for visualizing high-dimensional datasets by mapping each observation onto a function. For a k -dimensional dataset, $\xi_i = (x_{i1}, x_{i2}, \dots, x_{ik})$. The Andrews curve is a plot of (t, y_{it}) in the range of $t \in [-\pi, \pi]$, where y_{it} is given by

$$y_{it} = \frac{x_{i1}}{2} + x_{i2} \cos(\lambda_1 t) + x_{i3} \sin(\lambda_1 t) + x_{i4} \cos(\lambda_2 t) + x_{i5} \sin(\lambda_2 t) + \dots \quad (3)$$

where $\lambda_i = i, i = 1, 2, \dots$. It indicates that Andrews curves that are represented by functions close together suggest that the corresponding data points will also be close together, and thus, Andrews curve is suitable for visualizing the clustering and classification of high-dimensional datasets. More information concerning Andrews curve can be found in [79].

4 Simulation

In the simulation experiment, a normal state and three failure states are constructed. Firstly, the secondary lateral dampers, the yaw dampers, and both the yaw dampers and the secondary lateral dampers of the front bogie are, respectively, removed (abbreviated as LDF, YDF, and Y&LDF, respectively). Secondly, simulation experiments are performed under the normal state and these three different failure states (i.e., normal, LDF, YDF, and Y&LDF), respectively. Besides, a tri-axial accelerometer is installed on the bogie frame (see Fig. 8). The reason for the accelerometer position is that we plan to use the data collected by only one sensor in the future to monitor the faults including primary suspension faults (only secondary suspension faults are tested in this paper), and using the vibration acceleration data from the bogie frame is a compromise choice. The sampling frequency is selected as 250 Hz. The acceleration signal used in this work is the lateral acceleration signal from the tri-axial accelerometer. The vehicle speed is equal to 250 km/h. The specific fault construction process was described in detail in the authors' previous work [3].

The simulation experiment consists of 4 cases. Case I (Sect. 4.1.1) shows the feasibility of using the designed 1DCNN method for train suspension systems in the case of the same railway line and the same wheel profile. Through the univariate analysis method, Case II (Sect. 4.1.2) and

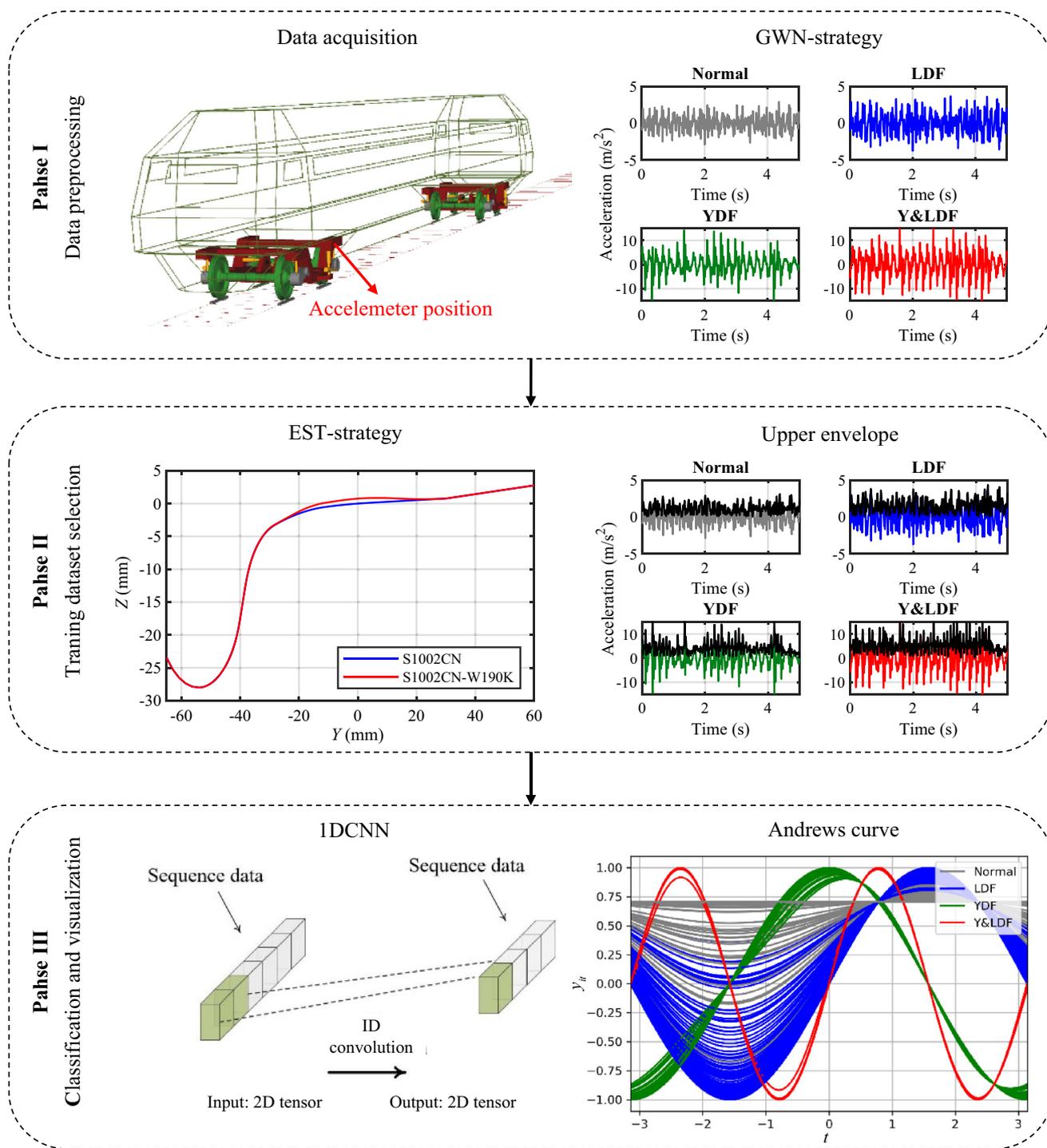


Fig. 8 The architecture of the fault diagnostic network (GWN-EST-1DCNN) of railway vehicle suspension systems

Case III (Sect. 4.1.3), respectively, demonstrate that the GWN-strategy can overcome the interference caused by track irregularities, and the EST-strategy can overcome the interference caused by wheel wear. Through the multivariate analysis method, Case IV (Sect. 4.1.4)

demonstrates that the GWN-EST-1DCNN method is not disturbed by simultaneous changes in track irregularities and wheel wear.

Table 3 The number of samples in different cases

Case	Training dataset and the number of trained samples (normal, LDF, YDF, Y&LDF)	Testing dataset and the number of tested samples (normal, LDF, YDF, Y&LDF)
I	WG-line: 864, 864, 864, 864	WG-line: 352, 352, 352, 352
II	WG-line: 864, 864, 864, 864	ZX-line: 1,248, 1,248, 1,248, 1,248 JJ-line: 1,248, 1,248, 1,248, 1,248
III	WG-line: 864, 864, 864, 864	WG-line: 352, 352, 352, 352
IV	WG-line: 864, 864, 864, 864	ZX-line: 1,248, 1,248, 1,248, 1,248 JJ-line: 1,248, 1,248, 1,248, 1,248

4.1 Case I (same line and same wheel profile)

A vehicle model, with S1002CN wheel profiles, running on the WG line is taken as an example to illustrate the feasibility of using the 1DCNN method for train suspension systems in the case of the same railway line and the same wheel profile. The training dataset and testing dataset are, respectively, composed of 3,456 samples and 1,408 different samples, respectively (see Table 3). The length of each sample is $t = 5$ s. According to the designed network of the 1DCNN described in Sect. 3.1, the final results,

including the convergence rate, confusion matrix, and visualization features, are shown in Fig. 9. Figure 9a indicates that the designed 1DCNN method with 30 epochs is convergent. The confusion matrix for the testing samples is shown in Fig. 9b, and it can be seen that all the states can be totally distinguished by the 1DCNN method (100%). To visualize the classification and clustering results, Andrews plot is presented in Fig. 9c, which shows that the four states can be completely separated and the clustering result is excellent. Overall, Fig. 9 indicates that the designed

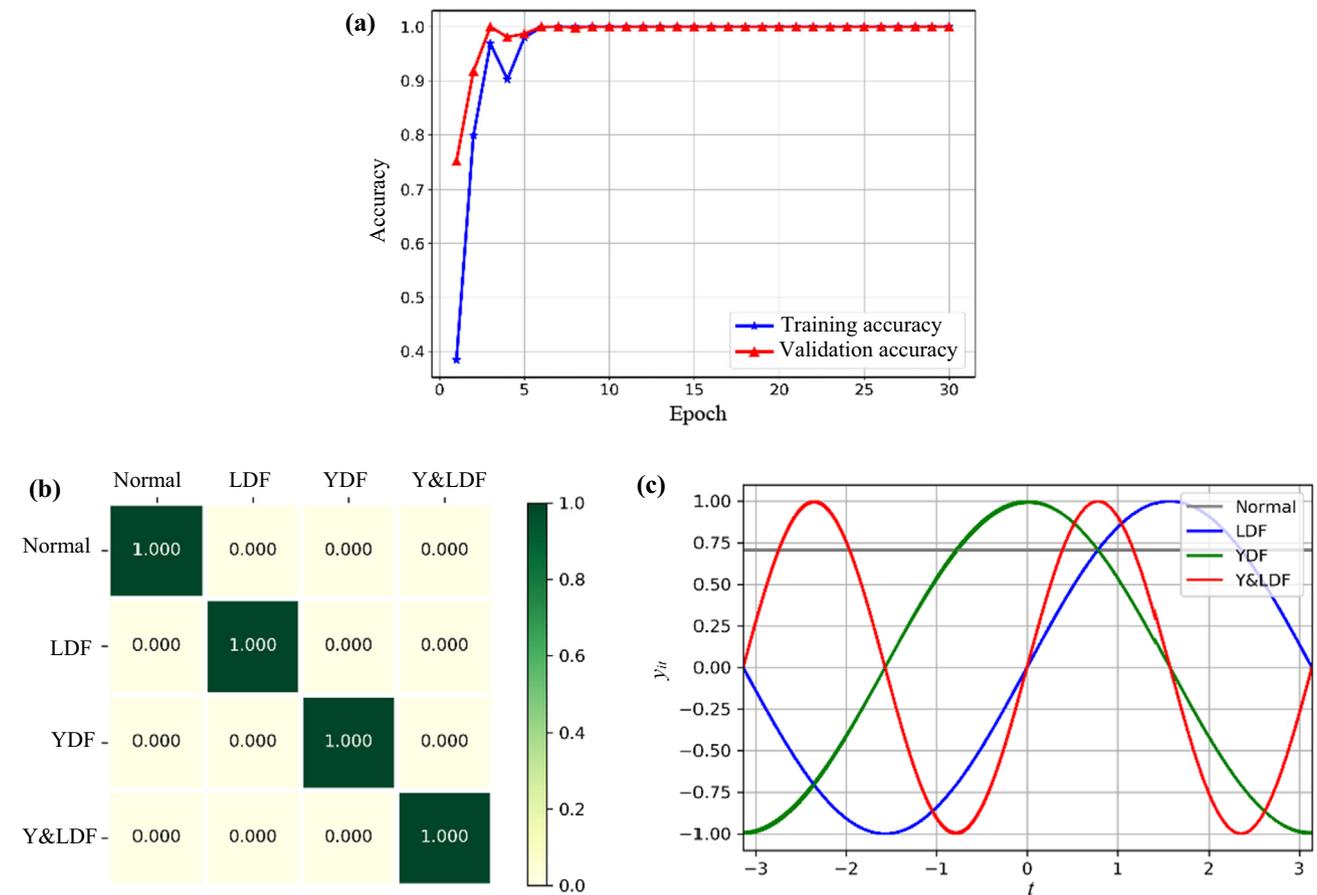


Fig. 9 Results for S1002CN profile and WG-line: **a** convergence rate, **b** recognition, and **c** visualization

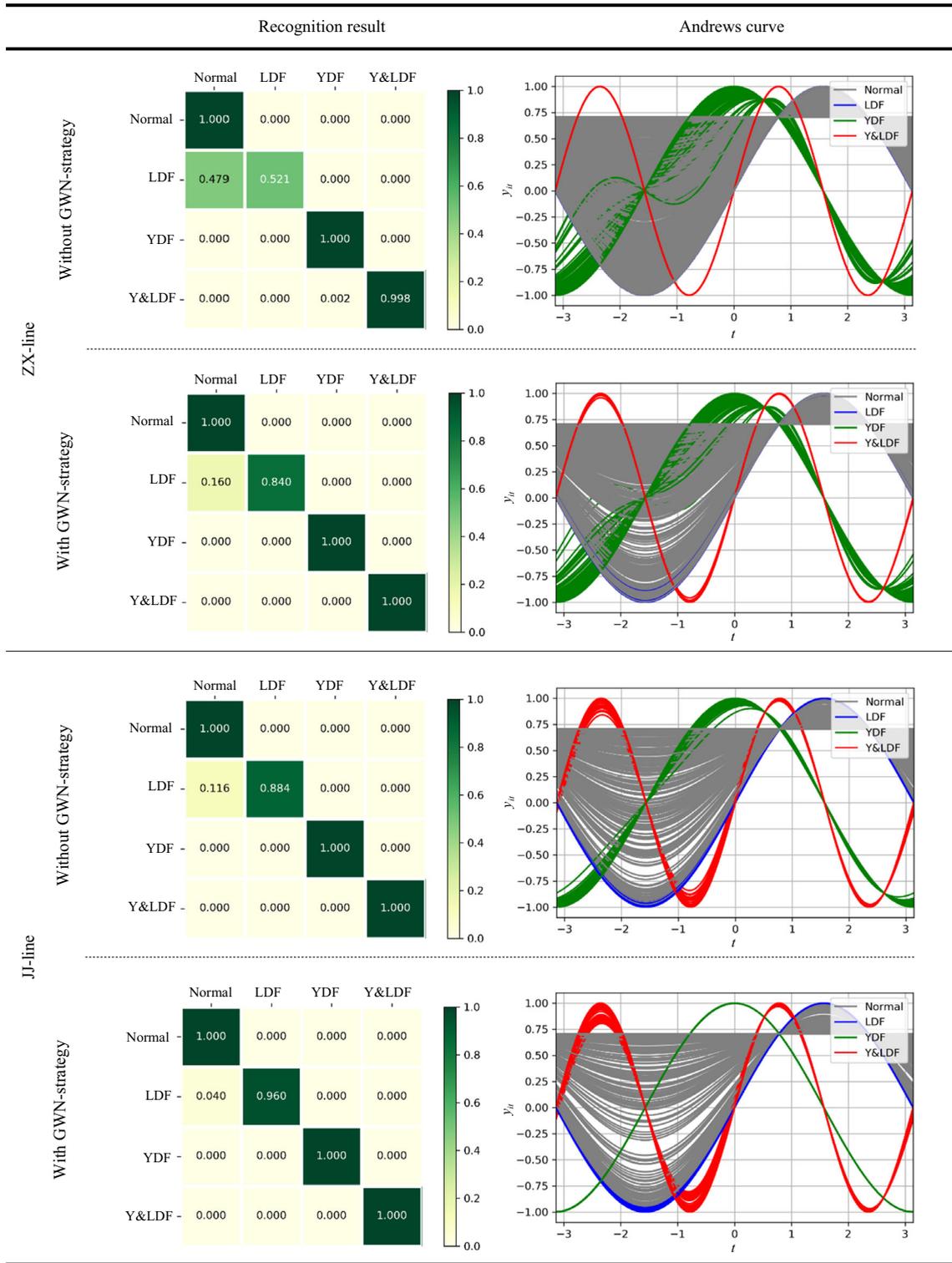


Fig. 10 Recognition and visualization results (S1002CN profile, ZX-line, and JJ-line)

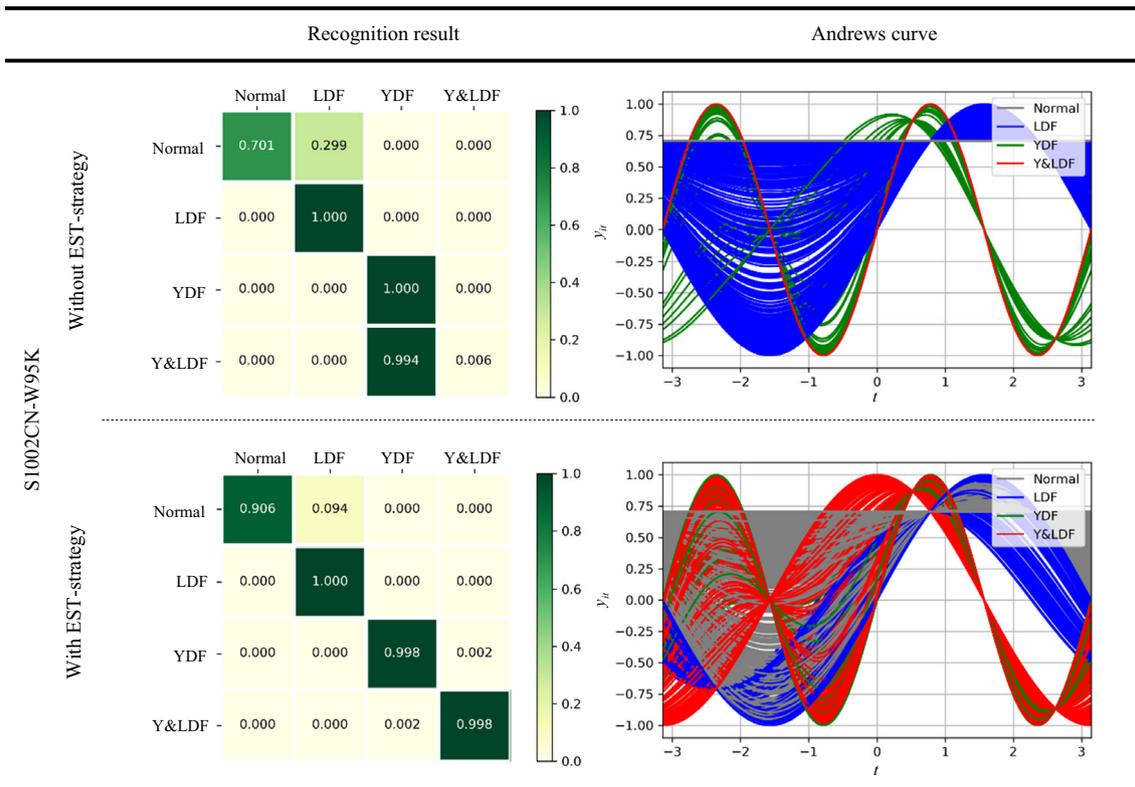


Fig. 11 Recognition and visualization results (WG-line, and S1002CN-W95K profile)

1DCNN can diagnose the secondary suspension faults and has the potential to be applied in RVSFD.

4.2 Case II (same wheel profile, and different railway lines)

To demonstrate that the GWN-strategy can improve the robustness of the diagnostic method against track irregularities, three measured track irregularities plotted in Fig. 4 are introduced here. The wheel profile used here is S1002CN. The training dataset and testing dataset are composed of 3,456 samples and 4,992 different samples, respectively (see Table 3). Finally, the results are shown in Fig. 10. It can be clearly seen that the LDF state cannot be well distinguished without the GWN-strategy. On the ZX-line, 47.9% of the LDF state was incorrectly identified as the normal state, and the corresponding incorrect recognition rate is 11.6% on the JJ-line. However, with the GWN-strategy, the corresponding incorrect recognition rates decrease from 47% and 11.6% to 16% and 4%, respectively; and the classification and clustering effect become better since it can be clearly seen that for the case of ZX-line and Without GWN-strategy, the Andrew curve cannot show the LDF state (blue curve) at all. The simulation

experiment proves that the GWN-strategy can improve the robustness against track irregularities.

4.3 Case III (same railway line, and different wheel profiles)

To demonstrate that the EST-strategy can improve the robustness of the diagnostic method against wheel wear, a new S1002CN wheel profile, the most worn wheel profile (S1002CN-W190K), and a wheel profile with a degree of wear between the two profiles (S1002CN-W95K) are introduced here (See Fig. 3). The rail line used here is the WG-line. The training dataset and testing dataset are composed of 3,456 samples and 1,480 different samples, respectively (see Table 3). Note that when using the EST-strategy, the samples of the training dataset are randomly selected from the dataset corresponding to the S1002CN wheel profile and the dataset corresponding to the S1002CN-W190K wheel profile, while without the EST-strategy, the training dataset is only the dataset corresponding to the S1002CN wheel profile; the testing dataset is the dataset corresponding the S1002CN-W95K. Figure 11 shows the recognition results obtained by these two methods. It can be seen that, without the EST-strategy, 29.9% of the normal state is incorrectly identified as LDF,

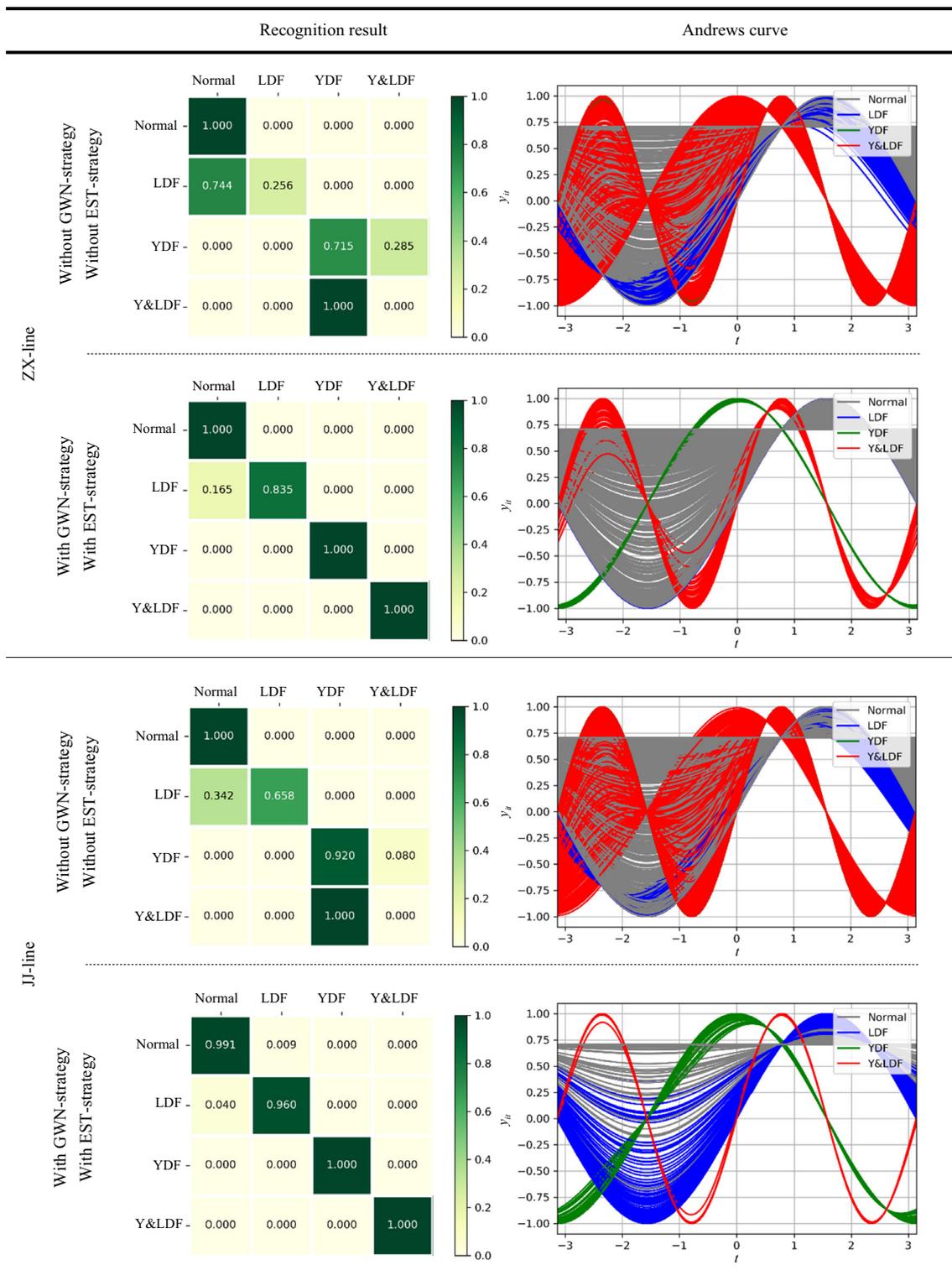


Fig. 12 Recognition and visualization results (ZX-line and JJ-line; S1002CN-W95K profile)

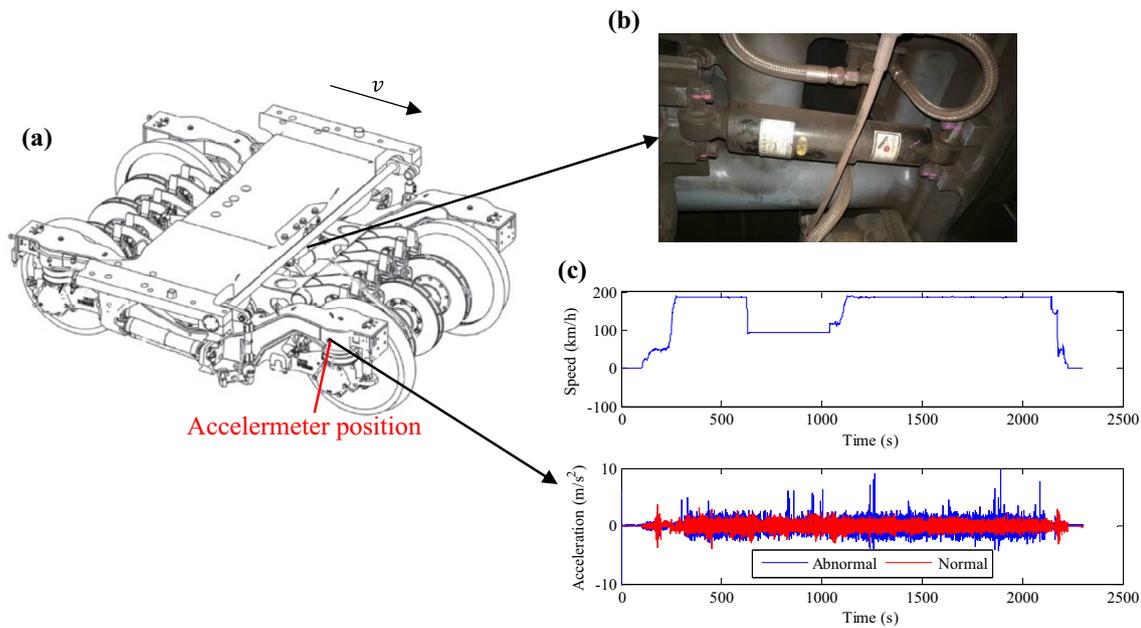


Fig. 13 Measurements of vehicle speed and front bogie acceleration: **a** the 3D drawing of the bogie; **b** the secondary lateral damper; **c** the vehicle speed, the lateral acceleration of the front bogie at the third car (abnormal), and the lateral acceleration of the front bogie at the second car (normal)

and all the Y&LDF samples are incorrectly recognized as YDF. By contrast, when using the EST-strategy, the accuracy is significantly improved. Besides, the Andrews curve also shows that when the EST strategy is applied, the difference in the curve distribution of these four faults becomes obvious. The simulation experiment proves that the EST-strategy can improve the robustness against wheel wear.

4.4 Case IV (different railway lines, and different wheel profiles)

In this subsection, we test the fault states under different track irregularities and different wheel profiles, and the training dataset and testing dataset are composed of 3,456 samples and 4,992 different samples, respectively (see Table 3). According to the technique route of the GWN-EST-1DCNN method described in Sect. 3.3, the final results in Fig. 12 show that, compared with simply using the 1DCNN method, the GWN-EST-1DCNN method can classify fault states regardless of whether the profile changes or the railway line changes except for a slightly worse prediction of the LDF samples on the ZX-line. Overall, it can be concluded that the recognition result is greatly improved.

4.5 Discussion

The advantages of the proposed GWN-EST-1DCNN method mainly arise from the following two aspects:

- (1) Track irregularities affect the bogie accelerations required for train suspension fault diagnosis. Under different track irregularities, there are some different relatively high-frequency and low amplitude impact components in these acceleration signals. The strategy of adding Gaussian white noise (GWN-strategy) to the original acceleration signals can improve the immunity of the diagnostic method to track irregularities since this strategy reduces the sensitivity of diagnostic methods to changes in track spectrum.
- (2) The wheel profile will change as the train running mileage increases. As the mileage increases, the amplitude of the bogie acceleration also increases. Therefore, using the dataset trained under the new wheel profile to identify the dataset under the worn wheel profiles may cause misidentification, i.e., the fault diagnostic method may incorrectly attribute the change in the bogie acceleration to a failure of the vehicle suspension system. The EST-strategy can improve the immunity of the diagnostic method to wheel wear mainly due to two reasons: (I) The training dataset of the diagnostic method covers a wider range of samples, which can identify the testing dataset to a certain extent more accurately. (II)

Training the diagnostic method with the datasets corresponding to the new wheel profile and the most worn wheel profile makes the method immune to the changes in the acceleration amplitude caused by wheel wear to a certain extent.

5 5. Experimental verification

5.1 Experiment setup

The actual operating conditions of railway vehicles are more complicated. To verify the performance of the 1DCNN method in real-life conditions, the field tracking data of a CRH3 train running on a high-speed railway line are applied. More information concerning the monitoring system can be found in [3]. The acceleration was measured through a tri-axial accelerometer mounted on the bogie frame (see Fig. 13), and the original sampling frequency was equal to 2 kHz. In our work, the lateral acceleration signal from the tri-axial accelerometer with a resampling frequency 250Hz is used.

When processing the tracking data, it was found that the acceleration signal of front bogie at the third car was abnormally vibrating, and its amplitude, for most of the time, was usually greater than that of the acceleration signal of other front bogies. Upon inspection, it was found that a hydraulic cylinder of the secondary lateral damper at the front bogie of the third car was short of oil (see Fig. 13b). Figure 13c shows the vehicle speed, the lateral acceleration of the front bogie at the third car (abnormal), and the lateral acceleration of the front bogie at the second car (normal).

5.2 Fault diagnosis

Three states, including the normal (normal), the lateral damper failure at the speed of 80–100 km/h (LDF80–100), and the lateral damper failure at the speed of 180–200 km/h (LDF180–200), are selected for analysis. The training dataset and testing dataset are, respectively, composed of 2,560 samples (normal: 1000; LDF80–100: 780; LDF180–200: 780) and 1024 samples (normal: 500; LDF80–100: 262; LDF180–200: 262), and the length of each sample is $t = 5$ s. The designed 1DCNN method described in Sect. 3.1 is applied to these signals, and the final results, including the convergence rate, confusion matrix, and Andrews curve, are shown in Fig. 14. Figure 14a indicates that the proposed 1DCNN method with 30 epochs is convergent. The confusion matrix for the testing samples is provided in Fig. 14b, and it can be seen that all the states can be totally distinguished (100%). Andrews plot

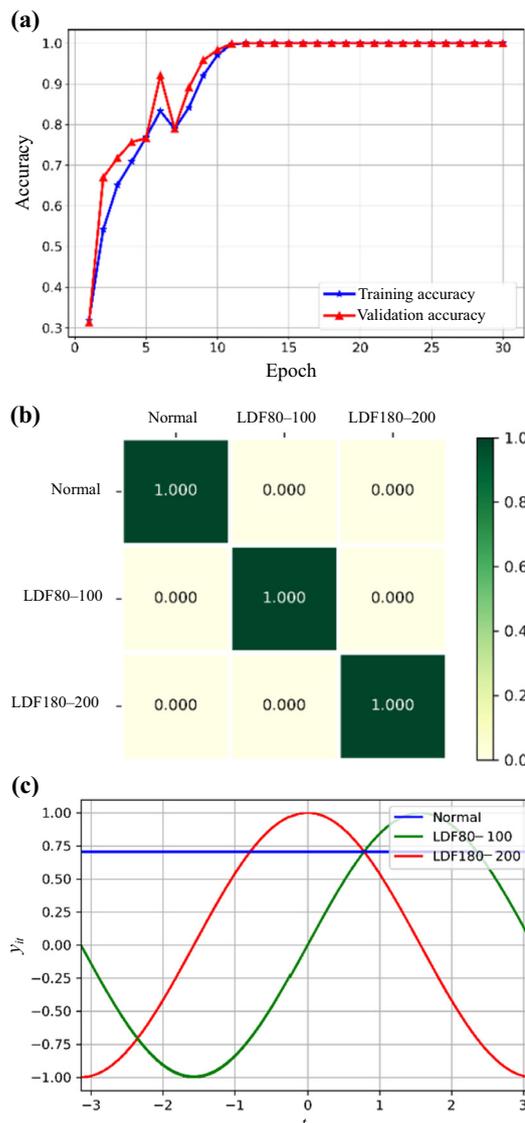


Fig. 14 Fault diagnosis results: a convergence rate, b recognition and c visualization

presented in Fig. 14c shows that the three states can be completely separated and the clustering result is excellent. The experiment results obtained using the field tracking data further verify that the proposed 1DCNN-based method can accurately identify the faults of the vehicle suspension systems at different speeds.

Due to the limitation of experimental resources, this paper only verified the method of 1DCNN to identify the secondary lateral damper failure using the field tracking data.

6 Conclusion

Track irregularities and wheel wear will affect the vibration signals used for condition monitoring of railway vehicles. The developed fault diagnosis method for train suspension systems, therefore, must be guaranteed to be immune to changes resulting from the track irregularities and wheel wear before being put into use. Aiming at solving this issue, a GWN-EST-IDCNN-based method for high-speed train suspension systems is proposed. This method consists of three phases. In the first phase (data preprocessing), a strategy of adding Gaussian white noise (GWN-strategy) is applied to the original signal, making the diagnostic method be immune to the interference caused by track irregularities. In the second phase (training dataset establishment), an EST-strategy is proposed to improve the robustness of the diagnostic network against wheel wear. In the third phase (training and recognition), a IDCNN-based fault diagnostic network of high-speed train suspension systems is built. Simulation experiments show the superiority and correctness of the proposed method. In addition, the field tracking data of a CRH3 train running on a high-speed railway line are used to further verify the effectiveness of the IDCNN method. The test results show that the method has the potential to be applied in the field of railway engineering.

This paper ends with the following notes. (1) It should be noted that the trained DL algorithm is extremely sensitive to the vehicle speed because the axlebox acceleration caused by different suspension faults varies at different vehicle speeds. Therefore, during on-board monitoring, the suspension status can be determined by obtaining the axlebox acceleration at a constant speed (e.g., 200 km/h or 250 km/h). However, to achieve real-time monitoring, more velocity conditions need to be further analyzed. (2) In the simulation experiments, only the complete damage of the dampers in the secondary suspension system is simulated. The degradation of suspension systems, including dampers, will be studied in the following-up work. (3) In the field experimental part, due to the limitation of experimental resources, this paper only verified the method of IDCNN to identify the secondary lateral damper failure using the field tracking data.

Acknowledgements This work is supported by the National Nature Science Foundation of China (No. 71871188), the Fundamental Research Funds for the Central Universities (No. 2682021CX051), and the first author is also supported by China Scholarship Council (No. 201707000113).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate

if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Iwnicki S (2006) Handbook of railway vehicle dynamics. CRC Taylor & Francis, Boca Raton, FL
- Bruni S, Goodall R, Mei TX, Tsunashima H (2007) Control and monitoring for railway vehicle dynamics. *Veh Syst Dyn* 45(7–8):743–779
- Ye Y, Zhang Y, Wang Q, Wang Z, Teng Z, Zhang H (2020) Fault diagnosis of high-speed train suspension systems using multiscale permutation entropy and linear local tangent space alignment. *Mech Syst Sig Process* 138:106565
- Li C, Luo S, Cole C, Spiriyagin M (2017) An overview: modern techniques for railway vehicle on-board health monitoring systems. *Veh Syst Dyn* 55(7):1045–1070
- Lebel D, Soize C, Funfschilling C, Perrin G (2020) High-speed train suspension health monitoring using computational dynamics and acceleration measurements. *Veh Syst Dyn* 58(6):1–22
- Wei X, Lin S, Liu H (2012) Distributed fault detection observer for rail vehicle suspension systems. In: 24th chinese control and decision conference (CCDC), Taiyuan. doi: <https://doi.org/10.1109/ccdc.2012.6244541>
- Xue P, Chai X-D, Zheng S-B (2015) Research on vehicle diagnosis based on state-space method. *Artif Intell Res* 4(2):55–60
- Wei XK, Jia LM, Liu H (2013) A comparative study on fault detection methods of rail vehicle suspension systems based on acceleration measurements. *Veh Syst Dyn* 51(5):700–720
- Liu F, Zhang H, He X, Zhao Y, Gu F, Ball AD (2020) Correlation signal subset-based stochastic subspace identification for an online identification of railway vehicle suspension systems. *Veh Syst Dyn* 58(4):569–589
- Li P, Goodall R (2004) Model based condition monitoring for railway vehicle systems, Control 2004, University of Bath, 2004, ID-508
- Wei X, Liu H, Qin Y (2011) Fault isolation of rail vehicle suspension systems by using similarity measure. In: International conference on intelligent railway transportation. Beijing, China, pp 391–396
- Jesussek M, Ellermann K (2013) Fault detection and isolation for a nonlinear railway vehicle suspension with a hybrid extended kalman filter. *Veh Syst Dyn* 51(10):1489–1501
- Jesussek M, Ellermann K (2014) Fault detection and isolation for a full-scale railway vehicle suspension with multiple Kalman filters. *Veh Syst Dyn* 52(12):1695–1715
- Jesussek M, Ellermann K (2015) Fault detection and isolation for a railway vehicle by evaluating estimation residuals. *Procedia IUTAM* 13:14–23
- Onat A, Kılınc O, Lata M (2016) A linear kalman filtering scheme for estimation of secondary vertical suspension of railway vehicles. *Vibroeng PROCEDIA* 7:124–128
- Zoljic-Beglerovic S, Stettinger G, Lubner B, Horn M (2018) Railway suspension system fault diagnosis using cubature kalman filter techniques. *IFAC-Papers Online* 51(24):1330–1335
- Wei X, Liu H, Jia L (2012) Fault detection of urban rail vehicle suspension system based on acceleration measurements. In: 2012

- IEEE/ASME international conference on advanced intelligent mechatronics (AIM), Kaohsiung. pp. 1129–1134
18. Xu B, Zhang J, Guan X (2014) Estimation of the parameters of a railway vehicle suspension using model-based filters with uncertainties, proceedings of the institution of mechanical engineers. Part F J Rail Rapid Transit 229(7):785–797
 19. Hayashi Y, Kojima T, Tsunashima H, Marumo Y (2006) Real time fault detection of railway vehicles and tracks. In: IET international conference on railway condition monitoring, Birmingham. pp. 20–25
 20. Hayashi Y, Tsunashima H, Marumo Y (2006) Fault detection of railway vehicles using multiple model approach. In: 2006 SICE-ICASE international joint conference, Busan. pp. 2812–2817
 21. Mori H, Tsunashima H (2010) Condition monitoring of railway vehicle suspension using multiple model approach. In: International conference on control, automation and systems, Gyeonggi-do. pp. 584–589
 22. Li P, Goodall R, Kadiramanathan V (2003) Parameter estimation of railway vehicle dynamic model using rao-blackwellised particle filter. In: 2003 European control conference (ECC), Cambridge. pp. 2384–2389
 23. Li P, Kadiramanathan V, Goodall R (2004) Estimation of parameters in a linear state space model using a rao-blackwellised particle filter. IEE Proceed Control Theory Appl 151(6):727–738
 24. Li P, Goodall R, Weston P, Ling CS, Goodman C, Roberts C (2007) Estimation of railway vehicle suspension parameters for condition monitoring. Control Eng Pract 15(1):43–55
 25. Liu X, Alfi S, Bruni S (2016) An efficient recursive least square-based condition monitoring approach for a rail vehicle suspension system. Veh Syst Dyn 54(6):814–830
 26. Liu X, Alfi S, Bruni S (2016) An efficient condition monitoring strategy of railway vehicle suspension based on recursive least-square algorithm. In: The dynamics of vehicles on roads and tracks, Boca Raton, CRC Press
 27. Gasparetto L, Alfi S, Bruni S (2013) Data-driven condition-based monitoring of high-speed railway bogies. Int J Rail Transp 1(1–2):42–56
 28. Alfi S, Bruni S, Mazzola L, Bionda S (2011) Model-based fault detection in railway bogies. In: 22nd international symposium on dynamics of vehicles on roads and tracks, Manchester, UK
 29. Mei T, Ding X (2008) A model-less technique for the fault detection of rail vehicle suspensions. Veh Syst Dyn 46(sup1):277–287
 30. Mei TX, Ding XJ (2009) Condition monitoring of rail vehicle suspensions based on changes in system dynamic interactions. Veh Syst Dyn 47(9):1167–1181
 31. Dumitriu M (2019) Fault detection of damper in railway vehicle suspension based on the cross-correlation analysis of bogie accelerations. Mech Ind 20(1):1–14
 32. Aravanis TCI, Sakellariou JS, Fassois SD (2017) Vibration based fault detection under variable non measurable operating conditions via a stochastic functional model method and application to railway vehicle suspensions. In: proceedings of the surveillance 9 international conference, Morocco
 33. Aravanis TCI, Sakellariou JS, Fassois SD (2018) Railway suspension fault detection under variable operating conditions via random vibration signals and the stochastic functional model based method. In: international conference on noise and vibration engineering (ISMA2018), Leuven
 34. Aravanis TCI, Sakellariou JS, Fassois SD (2018) On the problem of random vibration based fault detection in railway vehicle suspensions under variable and non-measurable operating conditions. In: AVT-305 research specialists' meeting on sensing systems for integrated vehicle health management for military vehicles, Athens
 35. Aravanis TCI, Sakellariou J, Fassois S (2020) A stochastic functional model based method for random vibration based robust fault detection under variable non-measurable operating conditions with application to railway vehicle suspensions. J Sound Vib 466:115006
 36. Sakellariou JS, Petsounis KA, Fassois SD (2002) On board fault detection and identification in railway vehicle suspensions via a functional model based method. In: Proc. international conference on noise and vibration engineering (ISMA), Leuven
 37. Sakellariou JS, Petsounis KA, Fassois SD (2015) Vibration based fault diagnosis for railway vehicle suspensions via a functional model based method: a feasibility study. J Mech Sci Technol 29(2):471–484
 38. Rossouw CCE, Fröhling RD, Nel AL (2019) Fault diagnosis in railway vehicle suspensions. In: International heavy haul association conference (IHHA), Narvik
 39. Qin N, Jin WD, Huang J, Jiang P, Li ZM (2013) High-speed train bogie fault signal analysis based on wavelet entropy feature. In: Proceedings of the international conference on advanced engineering materials and technology, Zhangjiajie. pp.2286–2289
 40. Kulkarni R, Qazizadeh A, Berg M, Stichel S (2019) Fault detection and isolation (FDI) method for vehicle running instability from vehicle dynamics response using machine learning. In: 11th international conference on railway bogies and running gears (BOGIE'19), Budapest
 41. Wei X, Jia L, Guo K, Wu S (2014) On fault isolation for rail vehicle suspension systems. Veh Syst Dyn 52(6):847–873
 42. Hong N, Li L, Yao W, Zhao Y, Yi C, Lin J et al (2019) High-speed rail suspension system health monitoring using multi-location vibration data. IEEE Trans Intell Transp Syst 21(7):2943–2955
 43. Wu Y, Jin W, Ren J, Sun Z (2019) Fault diagnosis of high-speed train bogie based on synchrony group convolutions. Shock Vib Spec Issue. <https://doi.org/10.1155/2019/7230194>
 44. Yabuno H, Takano H, Okamoto H (2008) Stabilization control of hunting motion of railway vehicle wheelset using gyroscopic damper. J Vib Control 14(1–2):209–230
 45. True H, Asmund R (2003) The dynamics of a railway freight wagon wheelset with dry friction damping. Veh Syst Dyn 38(2):149–163
 46. Bruni S, Vinolas J, Berg M, Polach O, Stichel S (2011) Modelling of suspension components in a rail vehicle dynamics context. Veh Syst Dyn 49(7):1021–1072
 47. Braghin F, Bruni S, Diana G (2006) Experimental and numerical investigation on the derailment of a railway wheelset with solid axle. Veh Syst Dyn 44(4):305–325
 48. Jia F, Lei YG, Lu N, Xing SB (2018) Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. Mech Syst Signal Process 110:349–367
 49. Ali JB, Fnaiech N, Saidi L, Chebel-Morello B, Fnaiech F (2015) Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. Appl Acoust 89:16–27
 50. Worden K, Staszewski WJ, Hensman JJ (2011) Natural computing for mechanical systems research: a tutorial overview. Mech Syst Signal Process 25(1):4–111
 51. Zhang L, Xiong GL, Liu HS, Zou HJ, Guo WZ (2010) Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. Expert Syst Appl 37(8):6077–6085
 52. Rosa AD, Alfi S, Bruni S (2019) Estimation of lateral and cross alignment in a railway track based on vehicle dynamics measurements. Mech Syst Signal Process 116:606–623
 53. Tao GQ, Du X, Zhang JH, Wen ZF, Jin XS, Cui DB (2017) Development and validation of a model for predicting wheel wear in high-speed trains. J Zhejiang Univ Sci A 18(8):603–616

54. Cui D, Wang H, Li L, Jin X (2013) Optimal design of wheel profiles for high-speed trains, proceedings of the institution of mechanical engineers. Part F J Rail Rapid Transit 229(3):248–261
55. Ye YG, Ning J (2009) Small-amplitude hunting diagnosis method for high-speed trains based on the bogie frame's lateral–longitudinal–vertical data fusion, independent mode function reconstruction and linear local tangent space alignment, proceedings of the institution of mechanical engineers. Part F J Rail Rapid Transit 233(10):1050–1067
56. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
57. Devries PMR, Viégas F, Wattenberg M, Meade BJ (2018) Deep learning of aftershock patterns following large earthquakes. Nature 560:632–634
58. Huang P, Wen C, Fu L, Peng Q, Tang Y (2020) A deep learning approach for multi-attribute data: a study of train delay prediction in railway systems. Inf Sci 516:234–253
59. Chang X, Tang B, Tan Q, Deng L, Zhang F (2019) One-dimensional fully decoupled networks for fault diagnosis of planetary gearboxes. Mech Syst Signal Process 141:106482
60. Ye Y, Huang P, Sun Y, Shi D (2021) MBSNet: a deep learning model for multibody dynamics simulation and its application to a vehicle–track system. Mech Syst Signal Process 157:107716
61. Meymand SZ, Keylin A, Ahmadian M (2016) A survey of wheel–rail contact models for rail vehicles. Veh Syst Dyn 54(3):386–428
62. Kalker JJ (1982) A fast algorithm for the simplified theory of rolling contact. Veh Syst Dyn 11(1):1–13
63. Zhai W (2020) Vehicle–track coupled dynamics models. Springer, Singapore
64. Chaar N, Berg M (2006) Simulation of vehicle–track interaction with flexible wheelsets, moving track models and field tests. Veh Syst Dyn 44(sup1):921–931
65. Smith ASJ, Odolinsk K, Niac SH, Jönssonc PA, Stichel S, Iwnicki S, Wheat P (2016) Estimating the marginal cost of different vehicle types on rail infrastructure. Working papers in transport economics, https://ideas.repec.org/p/hhs/ctswwps/2016_026.html
66. Bezin Y, Pålsson BA (2021) Multibody simulation benchmark for dynamic vehicle–track interaction in switches and crossings: modelling description and simulation tasks. Veh Syst Dyn. <https://doi.org/10.1080/00423114.2021.1942079>
67. Ning J, Lin J, Zhang B (2016) Time–frequency processing of track irregularities in high-speed train. Mech Syst Signal Process 66–67:339–348
68. Xu L, Zhai W, Chen Z (2018) On use of characteristic wavelengths of track irregularities to predict track portions with deteriorated wheel/rail forces. Mech Syst Signal Process 104:264–278
69. Xiao X, Sun Z, Shen W (2020) A Kalman filter algorithm for identifying track irregularities of railway bridges using vehicle dynamic responses. Mech Syst Signal Process 138:106582
70. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>
71. Shi D, Ye Y, Gillwald M, Hecht M (2020) Designing a light-weight 1D convolutional neural network with bayesian optimization for wheel flat detection using carbody accelerations. Int J Rail Transp 9(4):1–31
72. Nair V, and Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning, Haifa. pp. 807–814
73. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: IEEE international conference on computer vision (ICCV), Santiago. pp. 1026–1034
74. Keras: Deep learning library for theano and tensorflow, <https://www.datasciencecentral.com/profiles/blogs/keras-deep-learning-library-for-theano-and-tensorflow>
75. Liu W, Wen Y, Yu Z, Li M, Raj B, Song L (2017) SphereFace: deep hypersphere embedding for face recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), Honolulu. pp. 6738–6746
76. Wu ZH, Huang NE (2009) Ensemble empirical mode decomposition: a noise assisted data analysis method. Adv Adapt Data Anal 1(1):1–41
77. Shi D, Sabanovic E, Rizzetto L, et al. (2021) Deep learning based virtual point tracking for real-time target-less dynamic displacement measurement in railway applications. <https://arxiv.org/abs/2101.06702>
78. Andrews DF (1972) Plots of high-dimensional data. Biometrics 28:125–136
79. Moustafa RE (2011) Andrews curves. Wiley Interdiscip Rev Comp Stat 3(4):373–382