# A strategy for out-of-roundness damage wheels identification in railway vehicles based on sparse autoencoders

Jorge Magalhães<sup>1</sup> · Tomás Jorge<sup>1</sup> · Rúben Silva<sup>2</sup> · António Guedes<sup>1</sup> · Diogo Ribeiro<sup>1</sup> · Andreia Meixedo<sup>2</sup> · Araliya Mosleh<sup>2</sup> · Cecília Vale<sup>2</sup> · Pedro Montenegro<sup>2</sup> · Alexandre Cury<sup>3</sup>

Received: 25 October 2023 / Revised: 6 April 2024 / Accepted: 8 April 2024 / Published online: 19 June 2024 © The Author(s) 2024

## Abstract

Wayside monitoring is a promising cost-effective alternative to predict damage in the rolling stock. The main goal of this work is to present an unsupervised methodology to identify out-of-roundness (OOR) damage wheels, such as wheel flats and polygonal wheels. This automatic damage identification algorithm is based on the vertical acceleration evaluated on the rails using a virtual wayside monitoring system and involves the application of a two-step procedure. The first step aims to define a confidence boundary by using (healthy) measurements evaluated on the rail constituting a baseline. The second step of the procedure involves classifying damage of predefined scenarios with different levels of severities. The proposed procedure is based on a machine learning methodology and includes the following stages: (1) data collection, (2) damage-sensitive feature extraction from the acquired responses using a neural network model, i.e., the sparse autoencoder (SAE), (3) data fusion based on the Mahalanobis distance, and (4) unsupervised feature classification by implementing outlier and cluster analysis. This procedure considers baseline responses at different speeds and rail irregularities to train the SAE model. Then, the trained SAE is capable to reconstruct test responses (not trained) allowing to compute the accumulative difference between original and reconstructed signals. The results prove the efficiency of the proposed approach in identifying the two most common types of OOR in railway wheels.

**Keywords** OOR wheel damage  $\cdot$  Damage identification  $\cdot$  Sparse autoencoder  $\cdot$  Passenger trains  $\cdot$  Wayside condition monitoring

## 1 Introduction

In recent times, the field of structural health monitoring (SHM) has noted the emergence of application artificial intelligence (AI) techniques [1–3], to predict future events based on historical data. In civil engineering field, these techniques are in the background of current approaches for damage detection [4–6], fatigue life prediction [7, 8], crack damage detection and evaluation [9], and autonomous

structural visual inspection to detect various types of damage [10].

Within the railway field, a key application of these techniques involves increasing operational safety and proactively addressing maintenance needs. The railway wheels are not perfectly circular, and their surfaces are not perfectly smooth, even immediately after manufacturing [11]. Wheel OOR represents a significant challenge within wheel-rail interaction, inducing substantial fluctuations in normal forces, vibrations, rolling noise, and impact noise between the wheel and rail. As a result, it substantially affects passenger comfort and influences the railway system. This can lead to phenomena such as wheel axle instability, causing bending, damaged rolling bearings, cracks on the wheels, rails, and sleepers.

OOR wheels are typically categorized into two types of defects: wheel flat (Fig. 1a), a common tread defect mainly caused by repeated wheel/rail abrasion during the braking and the rolling of wheels over a long period of time [12];



Jorge Magalhães jorbm@isep.ipp.pt

<sup>&</sup>lt;sup>1</sup> CONSTRUCT-LESE, School of Engineering, Polytechnic of Porto, Porto, Portugal

<sup>&</sup>lt;sup>2</sup> CONSTRUCT-LESE, Faculty of Engineering, University of Porto, Porto, Portugal

<sup>&</sup>lt;sup>3</sup> Graduate Program in Civil Engineering, Federal University of Juiz de Fora, Juiz de Fora, Brazil



Fig. 1 Illustration of  $\mathbf{a}$  wheel flat and  $\mathbf{b}$  wheel polygonal wear

and wheel polygonal wear (Fig. 1b), defined as a periodic irregularity around the wheel circumference from the mean wheel radius [13].

In recent studies, several forms of OOR wheel conditions have been measured through experiments by assessing the structural implications arising from the dynamic phenomena [12, 14, 15]. In the works of Wu et al. [16] and Cai et al. [17], a detailed investigation is conducted via field experiment about the mechanism of high-order polygonal wear of wheels in China high-speed trains. According to the studies, the basic condition for the polygon generation of wheels depends on the operating speed, the excited resonant frequency, and the current characteristics of the wheels. In the results of Wu et al. [16], by changing the operating speed, the basic condition for polygon generation of wheels is changed and polygonal wear increases. For Cai et al. [17], the increase in the vehicle speed shifts the higher order of wheel polygonization to a lower order due to the "fixedfrequency" mechanism. On the wheel flat cases, Chang et al. [12] conducted an experimental investigation on the wheel/ rail impact based on wheel flats with various characteristics. These wheel flats were deliberately positioned around the rolling circle of the wheel tread, with testing conducted across speeds ranging from 0 to 400 km/h. The researchers observed that by increasing the speed, the wheel flat induced maximum wheel/rail dynamic impact force experienced a rapid rise, reaching its peak around 35 km/h. Subsequently, the force gradually declined as the speed continued to increase. This aspect was also identified numerically by Vale [18] on both ballasted and non-ballasted tracks.

These unusual physical phenomena can be managed through appropriate measures. The installation of sensors is the most common solution for this and can be done by incorporating onboard systems [14, 19–21] or by setting up wayside systems, currently standing out as an optimal solution for acquiring dynamic responses [1, 2, 22]. Furthermore, some researchers have formulated mathematical models and conducted numerical simulations to replicate train passages involving OOR wheels. These numerical simulations require the modeling of the different subsystems, i.e., track, vehicles and eventually bridges, which are typically calibrated based on modal parameters, namely frequencies and modal configurations [23, 24]. The methodologies for forecasting wheel/rail wear assessment involve the integration of a dynamic vehicle/track model and a wheel/rail damage model within a feedback loop. This entails a dynamic model to establish wheel/rail normal forces and contact patch creepage, and a pre-modeling of wear so that to iteratively update the wheel/rail profile.

Several authors implemented methods for damage detection based on dynamic responses and using different types of machine learning (ML) algorithms, such as artificial neural networks (ANN) [25], deep neural networks (DNN) [14, 26], principal component analysis (PCA) [27, 28], wavelet continuous transform (CWT) [29] and autoregressive (AR) models [30]. Among them, artificial neural network and deep neural networks algorithms have been applied in diverse areas though the years. Often, these ML techniques are used in combination with other techniques for structural damage detection, i.e., a combination of a deep autoencoder with a one-class support vector machine (OC-SVM), proposed by Wang and Cha [4] which enables to detect future structural damage, and an ANN with a Gaussian process developed by Gonzalez and Karoumi [31] for detect damage on railway bridges.

The difference between ANN and DNN techniques is in the quantity of hidden layers, as DNN represents a more intricate network characterized by simultaneous combinations of various ANNs. Being that, typically an ANN is configured in three layers: The first one is the input layer and does not receive input from any previous layer; the second is called the hidden layer and takes as input the output of the input layer; and the third layer, the output layer, takes its input from the hidden layer and performs an analogous operation [31]. The DNNs are composed by multiple hidden layers and are capable to extract damage-sensitive features from the input data without any pre- or post-processing of them. Compared to ANN with a single hidden layer, the multiple hidden layers enable the DNN to learn mathematically more complex underlying feature representations of the input data [4].

423

In an early application exploring various neural networks architectures, Kudva et al. [25] devised a method to identify damage in small structures using measured strain values. After trying out several alternatives, they established the optimal number of hidden layers and nodes per layer, which allowed them to train the neural network to deduce the damage size and location from measured strain values at discrete locations. Nowadays, the application of deep learning has been commonly used due to the main advantage to extract damage-sensitive features during their training processes and the proficiency to capture nonlinear relationships and intricate patterns in the input data [7]. Cha [9] introduced a vision-based approach employing a convolutional neural network (CNNs) with a deep architecture for identifying concrete cracks in images without the need for computing defect features. The study demonstrated notable efficacy, particularly in detection thin cracks under challenging lighting conditions, where traditional methods struggle. Nonetheless, implementing such techniques requires substantial training data to ensure the classifier's robustness.

Among the various techniques in ANN and DNN, autoencoders have been widely used in the detection of structural damage. An autoencoder comprises an encoder and a decoder, which work together to map input variables. According to Lee et al. [7], an autoencoder with more than one hidden layer is called a DAE, and the additional encoding and decoding processes are performed in each added hidden layer. In the standard approach, autoencoder-based anomaly detection techniques acquire an understanding of typical, unaffected behavior during the training phase. This encompasses characteristics like wave patterns and their associated amplitudes under undamaged circumstances. Subsequently, the anomaly detection process entails assessing whether the test data align with the acquired model or not [32, 33]. In work developed by Wang and Cha [34], a comparative study is carried out between different machine learning and deep learning techniques for detecting structural damage in a steel bridge model using acceleration data. Among the techniques compared on the work stands out the deep autoencoder with Mahalanobis distance, where only the acceleration data measured from the intact structural scenarios are used to train the deep autoencoder. After test procedure, three indexes were used to quantify the reconstruction losses and the Mahalanobis distance metric is applied to measure the similarity of testing data points to the training matrix. The method proposed by authors indicated a highly performance for global health conditions of structures. Pathirage et al. [35] developed an unsupervisedlearning framework for structural damage assessment, which consists of a deep autoencoder for structural characteristics dimension reduction, and a simple autoencoder for a regression task of predicting structural stiffness reduction. Likewise, Sarwar et al. [5] developed a method with a deep autoencoder to detect damage in a road bridge with acceleration responses from various types of vehicles. The method consists in training the autoencoder for feature extraction, calculating the mean absolute error (MAE) and a statistical distribution. The results presented by the authors indicate that the method is capable of detecting damage effectively, producing robust results even when subject to multivariate operational conditions, such as variations in road profiles, vehicle properties and measurement noise. On the other hand, autoencoders can also be applied in a classification procedure, which requires encompassing all possible scenarios (damaged and undamaged) within the training process [14, 36].

Typically, the steps for damage identification methods are related to data collection, pre-processing data, feature extraction, feature normalization, data fusion, and feature classification [2, 5, 14]. The data collection can be evaluated either with experimental or numerical data and its preprocessing can be done by transferring variables to another spatial domain [14, 21]. The transformation of the data record into alternative information, where the correlation with the damage is easily visible, is called feature extraction [37, 38]. Feature normalization plays a vital role in preventing false alarms, since several environmental effects, such as temperature and operational factors like the speed of a train, can influence infrastructure response more than damage. Data fusion techniques allow dimensional reduction while preserving the relevant information contained in the data, characterized by combining information from several indicators, of the same or different natures, to increase the reliability of the measured phenomenon. Mahalanobis distance is widely implemented to fuse all damage-correlated information [4, 29]. The classification process typically comprises outlier analyses, where a threshold is predicted based on the damage-sensitive features [14, 27] and a cluster analysis for automatic grouping [27, 39].

Given these aspects, the main goal of the present study is to develop a hybrid unsupervised ML strategy to detect OOR, namely wheel flats and polygonized wheels in passenger trains, identifying the type of the damage and the respective level of severity of the defect. The strategy proposed is validated based on a 3D numerical simulation of the train–track dynamic response for vehicle crossings. This model encompasses various vehicle properties and speeds, along with track irregularity profiles and noise. The core of the methodology involves training an autoencoder to obtain a damage index. For this purpose, the sparse autoencoder (SAE) was adopted, as it allows for obtaining better results, compared to common automatic autoencoders, due to sparsity restrictions [40]. The input data for the autoencoder comprise the vertical accelerations experienced by the vehicles while crossing the track. Once trained, the autoencoder is applied to predict the subsequent vehicle responses. The disparity between the modelbased predictions and the original vehicle responses gives rise to the prediction error, defined as DI. To increase the sensitivity of the DI, the Mahalanobis distance between the DI obtained from each sensor is evaluated. An outlier analysis is applied to detect damage, and a clustering technique is used to classify, both, the type of damage and the severity of each type of damage. The numerical implementation assesses the effectiveness of the proposed strategy across a range of simulated damage scenarios considering different geometric characteristics and defect amplitudes, as well as circulation speeds. Nevertheless, the architecture of the proposed methodology exhibits sufficient flexibility to incorporate a damage location stage.

The main contributions of the present work in relation to the existing bibliography can be summarized as follows:

- Enable converting the challenge of monitoring OOR wheels into a hybrid unsupervised ML approach.
- Define an SAE architecture with a combination of hidden layers and hyperparameters that allows the best information gain from baseline responses.
- Detect two types of OOR wheel damage scenarios on different wheels and on distinct sides of the train.

## 2 Sparse autoencoder

An autoencoder (AE) is an unsupervised neural network model and is used to estimate input variables (reconstruction) by learning the relationships and statistical patterns between the input variables [7]. The term "autoencoder" comes from the model trying to encode and then decode the input data, aiming to reconstruct the original data as accurately as possible [5]. The encoder module maps the input data x (original acceleration response) into arbitrary lower dimensional space z as an output  $\hat{x}$  (reconstructed acceleration response). The autoencoder process for each k neuron is expressed as follows:

$$\boldsymbol{z}_{k} = \varphi \left( \sum_{j=1}^{J} \boldsymbol{w}_{kj} \cdot \boldsymbol{x}_{j} + \boldsymbol{b}_{k} \right), \tag{1}$$

$$\hat{\boldsymbol{x}}_{j} = \varphi' \left( \sum_{j=1}^{J} \boldsymbol{w}_{kj}' \cdot \boldsymbol{z}_{k} + \boldsymbol{b}_{k}' \right),$$
(2)

where *j* is the number of acceleration response vectors,  $x_j$  is the *j*th element of the input data, and  $\hat{x}_j$  is the *j*th element of the output data;  $w_{kj}$ ,  $w'_{ki}$  and  $b_k$ ,  $b'_k$  are the weight matrices

and bias vectors for encoder and decoder modules, respectively, while  $\varphi$  and  $\varphi'$  are the activation functions of encoder and decoder, which can be linear or nonlinear. The number of epochs (iterations) of a training process allows for adjusting the weights and biases of the encoder and the decoder. In that period, the autoencoder tries to learn a compact representation in the hidden layer, enabling it to reconstruct the input data with minimal error [38]. A sparse autoencoder is a variant of the standard autoencoder that includes a sparsity constraint on the activation functions of the hidden layer. The sparsity constraint encourages the autoencoder to learn a more concise and sparse representation of the input data. Mathematically, the main difference in a sparse autoencoder lies in the regularization term added to the loss function to impose the sparsity constraint [41]. The cost function (E) for training a sparse autoencoder is an adjusted mean squared error function as follows:

$$E = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{J} (\mathbf{x}_{j_n} - \hat{\mathbf{x}}_{j_n})^2 + \lambda \cdot \frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{J} \left[ (\mathbf{w}_{kj})^2 + (\mathbf{w}'_{kj})^2 \right] + \beta \cdot \sum_{j=1}^{J} \text{KL}(\rho | \hat{\rho}_j),$$
(3)

where  $x_{j_n}$  is the *n*th element of  $x_j$ ;  $\hat{x}_{j_n}$  is the *n*th element of  $\hat{x}_j$ ;  $\lambda$  is the coefficient for regularization term;  $\beta$  is the coefficient for the sparsity regularization term;  $\rho$  is the average desired information gain, the sparsity proportion. These terms can be specified while training an autoencoder; and  $\hat{\rho}_j$  is the average information gain in the train process. The Kullback–Leibler (KL) divergence is a function for measuring how different two distributions are. In this case, it takes the value zero when  $\rho$  and  $\hat{\rho}_j$  are equal and become larger as they diverge from each other. Minimizing the cost function forces this term to be small; hence  $\rho$  and  $\hat{\rho}_j$  to become close to each other [41].

With this type of tool, depending on its architecture and the given input, it is possible to develop a network with the necessary characteristics to solve the problem for which it was created. The SAE models are skilful at accurately estimating intricate patterns and nonlinear connections within input variables. Therefore, the SAE model serves as a valuable tool for anomaly detection, identifying unusual instances through substantial reconstruction errors [7]. In detecting structural damage, the input can consist of the dynamic responses of the structure or images. Given the various applications of neural networks in the field of engineering, the present work is based on a traditional SAE that uses a training process to extract features from response measurements during the crossing of vehicles with healthy wheels.

## 3 Numerical modeling

This section is dedicated to assessing the data utilized in the ongoing study. The vehicle–track interaction is detailed in Sect. 3.1 along with the description of numerical models and the software used to extract dynamic responses. In Sect. 3.2, the virtual wayside system is presented. The simulated scenarios are shown in Sect. 3.3 along with theorical background of OOR defects, and Sect. 3.4 comprises the vertical acceleration responses obtained for each simulated scenario.

## 3.1 Vehicle-track interaction

For this study, the numerical simulations of train-track dynamic interaction were conducted using an in-house software called vehicle-structure interaction analysis (VSI). The analysis of vehicle-structure interaction is thoroughly explained and validated in the work of Montenegro et al. [42, 43], and it has been successfully applied in various other applications [27–30]. A 3D model of wheel-rail contact integrates the train and track through Hertzian theory [44]. It employs the USETAB routine [45] to calculate normal contact and computes tangential forces resulting from rolling friction creep. While these subsystem models were initially constructed separately, the VSI program interconnects them through a comprehensive coupling approach [43]. The graphical illustration of this process is presented in Fig. 2.

The numerical tool for these computations is implemented in MATLAB<sup>®</sup> [46] and imports the structural

matrices from both the vehicle and track previously modeled in ANSYS<sup>®</sup> [47]. The 3D ballast track numerical model employed in this study is a simplified version derived from the model validated with modal parameters as presented in Ribeiro et al. [48]. The vehicle adopted in this work consists on the Alfa Pendular train, which operates in the Portuguese Northern Railway line connecting Porto to Lisbon at the maximum speed of 220 km/h. The vehicle was also modeled in ANSYS<sup>®</sup> [47], utilizing a simplified model derived from the experimentally calibrated model based on modal parameters outlined in work of Ribeiro et al. [49]. A comprehensive description of both track and train model characteristics can be found in Mosleh et al. [50].

Rail irregularities in real-track conditions exist even in a healthy condition, and their effects on wheel–rail contact cannot be neglected [51]. At regular intervals of six months, the Railway Network Administration conducts assessments of track irregularities along the northern line of the Portuguese railway network. Moreover, power spectral density (PSD) curves are constructed using empirical data, and synthetic profiles of unevenness were generated. Consequently, rail surface irregularity patterns are generated for wavelengths spanning from 1 to 75 m with a maximum amplitude of 6 mm [28]. The wavelengths and amplitudes represent a good track condition as specified by the European Standard EN 13848-2 [52]. More details about the generation of unevenness profiles are originally provided by Mosleh et al. [53] and subsequently applied in numerous studies [27–30].

## 3.2 Virtual wayside monitoring system

A virtual wayside monitoring system is defined to measure rail accelerations due to the passage of a train. The system is composed of a set of 6 accelerometers mounted on the rail at mid-span between two sleepers, as illustrated in Fig. 3.



Fig. 2 Vehicle-track interaction model schematization



Fig. 3 Virtual wayside monitoring system: a back view; b top view

The numbers 1-to-6 in Fig. 3 represent the positions of the measurement points, in the right (1–3) and left (4–6) rails. Acceleration signals are assessed at a sampling frequency of 10 kHz. Subsequently, a low-pass Chebyshev type II digital filter [27, 28] with a cut-off frequency of 1500 Hz is applied to filter all-time series. This sampling high frequency can thus increase the variance of the subsequently extracted damage features [54]. Additionally, an artificial noise equivalent to 5% of the amplitude is incorporated into the numerical signal for a more realistic representation of the measured rail response [27, 28].

## 3.3 Simulated scenarios

The several train crossings simulated are classified into two groups, as shown in Table 1. The first group represents the baseline condition, composed of 120 undamaged scenarios, corresponding to the train passage with healthy wheels. The second group represents the passage of the train with a defective wheel, composed of two subgroups, the wheel flats and the polygonal wheels, in a total of 30 and 40 cases for each speed, respectively. These two types of defects were modeled by transforming the wheel defect into an equivalent and spaced rail defect, over which runs a perfect wheel [18], such as realized in many studies [28–30].

Within vibration-based damage detection methodologies, the sensitivity to damage depends on the location. To this end, two different types of damage were simulated on different sides of different wagons of the Alfa Pendular train, as shown in Fig. 4. The simulation of defects is guaranteed by superimposing them on the track, according to recent studies [28, 30].

#### 3.3.1 Baseline

To establish a solid groundwork that addresses a broad spectrum of situations aimed at identifying instabilities, multiple baseline simulations are performed. These simulations cover diverse load configurations, track condition variations, and

| Table 1 Baseline and damage scena | rios |
|-----------------------------------|------|
|-----------------------------------|------|

|                           | Baseline scenario     | Damage scenario  |
|---------------------------|-----------------------|--|
| Train type                | Alfa Pendular         | Alfa Pendular  |
| Loading schemes           | 3 (full, half, empty) | 1 (full)   |
| Irregularities profile    | 1-4                   | 1  |
| Speeds                    | 40–220 km/h           | 120 and 200 km/h   |
| Flat characteristics      | _                     | Flat lengths/depths (mm):<br>L1 [10–20]/D1 [0.02–0.06]<br>L2 [25–35]/D2 [0.09–0.16]<br>L3 [40–50]/D3 [0.23–0.36] |
| Polygonal characteristics | _                     | Harmonic order: H6-8, H12-14, H19-20, H29-30   |
|                           | _                     | Amplitude range (mm): A1[0.25–0.35], A2[0.55–0.65]   |
| Total number of analyses  | 120                   | 140  |



Fig. 4 Localization of damage

vehicle speeds. The assumptions that form the basis of these fundamental scenarios are succinctly presented in Table 1. This table outlines three distinct loading setups, four profiles of track irregularities (designated as 1 to 4), and a range of ten varying speeds (ranging from 40 to 220 km/h in increments of 20 km/h). The loading scenarios examined cover empty, half-load, and fully loaded conditions.

#### 3.3.2 Wheel flat

As a result of frequent and force braking in urban traffic conditions, railway wheels often exhibit a propensity to develop flat spots [18]. As shown in Table 1, for the wheel flats scenarios, 10 cases are considered for each severity group (L1–L3) making up a total of 30 passages for each speed. According to Chang et al. [12], the L1 group (low) presents a range of defect geometries that are admitted into circulation, denominated early flat. The L2 group (moderate), on the other hand, is characterized by a geometric range that includes flats in a more advanced state compared to L1, but which are still within the admissible range. Group L3 (severe) comprises wheel flat situations considered as damage. In this case, the flat is located on the left wheel of the last wheel set of the third vehicle, according to Fig. 4. The characteristics of the flats were selected according to several studies from the bibliography [12, 29, 50]. The wheel flat depth (D) is defined by the following expression [29]:

$$D = \frac{L^2}{16R_{\rm w}},\tag{4}$$

where *L* is the flat length, and  $R_w$  is the radius wheel (equal to 0.45 m). The vertical profile deviation (*Z*) of the wheel flat is characterized as follows [29]:

$$Z = \frac{D}{2} \left( 1 - \cos \frac{2\pi x_{w}}{L} \right) \cdot h \left[ x_{w} - \left( 2\pi R_{w} - L \right) \right],$$
  
$$0 \le x_{w} \le 2\pi R_{w},$$
 (5)

where *h* represents the Heaviside periodic function, and  $x_w$  is the coordinate aligned with the track longitudinal direction.



Fig. 5 Wheel flat characteristics

Figure 5 shows one example of the wheel flat profile for each simulated case.

#### 3.3.3 Polygonized wheel

In the railway context, these irregularities typically manifest themselves in distinct wavelengths varying from 10 cm to over 3 m corresponding to high-order polygonal OOR down to lower order or eccentricity around the rim's circumference, presenting amplitudes of the order of 1 mm [13]. Research articles in this domain detail the harmonic elements of these OOR irregularities, with their wavelengths ( $\Lambda$ ) determined by

$$A = \frac{2\pi R_{\rm w}}{\Theta},\tag{6}$$

where  $\Theta = 1, 2, 3, ..., n$  (harmonic components) and  $R_w$  is the radius wheel (equal to 0.45m). The polygonal wheel profiles (Fig. 6b, c) are defined based on experimentally measured profiles (Fig. 6a) with dominant harmonic orders of H6–8 [55], H12–14 [56], H19–20 [57], and H29–30 [17]. The lower orders (H6–8 and H12–14) are obtained for a speed circulation of 120 km/h; the higher orders (H19–20 and H29–30) are acquired for the vehicle's circulation of 200 km/h. For the amplitude of defects, two ranges are considered based on the study of Nielsen and Johansson [13], making up forty passages for each speed. According to Peng



Fig. 6 Characteristics of polygonal wheels: a experimental spectra of measured irregularities; b, c one example of each wheel profile simulation based on the respective spectra for amplitude A1 and A2, respectively

[58] and Iwnicki et al. [15], the range A1 is characterized by an amplitude in initial format of wear and the range A2 is a type of higher wear where the wheel should be re-profiled. In this case, the polygonal wheel is in the right wheel of the first wheelset of the first vehicle, according to Fig. 4.

The wheel profiles are characterized by the wavelengths (*w*) in the first 30 harmonics [28], based on the sum of sine functions (H = 30) as follows:

$$w(x_{w}) = \sum_{\Theta=1}^{H} A_{\Theta} \cdot \sin\left(\frac{2\pi}{\Lambda}x_{w} + \psi_{\Theta}\right),$$
  
$$0 \le x_{w} \le 2\pi R_{w},$$
  
(7)

where  $x_w$  is the distance along wheel circumference;  $\psi_{\Theta}$  is phase angle; and  $A_{\Theta}$  is the amplitude of the sine function for each  $\Lambda$ , which is calculated by

$$A_{\Theta} = \sqrt{2} \cdot 10^{L_{\rm w}/10} \cdot w_{\rm ref},\tag{8}$$

where  $w_{\text{ref}} = 1 \,\mu\text{m}$ . The wheel irregularity level ( $L_w$ ) values are selected based on the irregularity spectrums (Fig. 5a) for

all scenarios. By assigning phase angles  $(\psi_{\Theta})$  to sine functions in a uniformly and randomly distributed manner within the range of  $0-2\pi$ , five cases for each amplitude of wear of wheel irregularities are generated based on each spectrum.

Table 1 compiles all information relative to the simulated scenarios, covering the range of operating conditions that was examined.

#### 3.4 Track accelerations responses

Figure 7 presents the baseline time-series of the accelerometer installed on the rail in position 1. These plots show the influence of different loading schemes (Fig. 7a) and irregularity profiles on the track (Fig. 7b) for the speed of 160 km/h. Independently of the type of load considered during vehicle operation, this does not induce changes in the dynamic response. On the other hand, the results show some variations in the dynamic responses for different irregularity profiles. Figure 7c shows acceleration responses for three distinct speeds, highlighting the significant impact of train speed.



Fig. 7 Acceleration responses measured in position 1 for baseline scenarios: **a** comparison between empty, half, and full load schemes; **b** comparison between track irregularities; **c** comparison between speeds of 120, 160 and 200 km/h



Fig.8 Comparison between acceleration responses measured in position 1, considering wheel flat scenarios and the wagon crossing at a 120 km/h and b 200 km/h



Fig. 9 Comparison between acceleration responses measured in position 1, considering polygonized wheels scenarios and the wagon crossing at a 120 km/h and b 200 km/h

For the damage scenarios, Fig. 8 illustrates the wheel flat scenarios, while Fig. 9 presents the polygonal wheel scenarios, both captured by accelerometers positioned at location 1 of the rail. These plots depict various simulated scenarios, showing the effect of amplitude and speed on each type of damage. Regarding wheel flats (Fig. 8), the different peaks resulting from the impact of the flat are visible according to the respective severity. In the polygonal wheels (Fig. 9), the periodicity of the defect produces more evident impact along the dynamic response. According to simulated polygonized wheel profiles (Fig. 6), there is a noticeable impact from the H29–30 harmonic order on the dynamic response for 120 km/h, and from the H12–14 order for 200 km/h. This observation highlights the greater sensitivity of the last harmonic order to changes in speed.

## 4 Proposed methodology

The current section initially presents an overview of the proposed methodology. Then, specific aspects regarding the model's architecture as well as the proposed damage index are presented.

### 4.1 Overview

The proposed methodology for damage detection and classification is presented in Fig. 10. First, the vertical acceleration responses of all accelerometers are evaluated through numerical simulations, using only data obtained from the baseline scenarios (undamaged scenarios) for training the sparse autoencoder (SAE). The selection of the best hyperparameters of the training process consisted in a sensitivity analysis with 16 types of traditional SAE from MATLAB<sup>®</sup> [46]. With the prediction in the SAE of the baseline (ones not trained) and damage cases, the damage index (DI) is calculated by some metrics of the reconstructed losses. However, the new damage index, the natural logarithmic mean squared error (ln(MSE)) and a mean absolute error (MAE) were the most accurate metrics in the present work. Furthermore, the Mahalanobis distance is applied to fuse the damage index with ln(MSE) of all six sensors to increase the damage sensitivity. Finally, a statistical threshold for automatic damage detection is applied, and a cluster analysis is performed in two steps. The first step of cluster analysis consists in evaluating the type of damage using the features achieved after the fusion, with ln(MSE). The second one enables classification in terms of severity of each damage identified using only the MAE.

#### 4.2 Data collection

The proposed strategy for damage detection is numerically evaluated using simulated data generated through the vehicle–track dynamic interaction outlined in Sect. 3. The acceleration responses are obtained through a virtual wayside monitoring system with six sensors localized in the rail at mid-span between two sleepers, as represented in Fig. 3. All acceleration responses with a time step of  $10^{-4}$  s are converted as a function of the track position (with a step of



Fig. 10 Overview of the methodology for damage detection and classification

0.0062 m) to uniformize all data. The vehicle model has a total of 158.9 m long, whereby the dimension of the acceleration vectors comprises a maximum of 165 m of track.

## 4.3 SAE model

The baseline scenarios, constituting 80% for training and 20% for testing [14], include all speeds and load conditions to

ensure the SAE model's independent from track conditions. Consequently, the SAE model is trained on passages containing three of the four types of track irregularities (comprising 80% of the data, making a total of 95 crossings), while the remaining irregularity is reserved for testing (the remaining 20% of the data, making a total of 25 crossings). All damage scenarios are included in test procedure. Table 2 summarizes all information used for training and testing the SAE model.

| Table 2 | Characteristics | of data fo | r SAE model |
|---------|-----------------|------------|-------------|
|         |                 |            |             |

| Characteristics of operability | Data for training (baseline sce- | Data for testing              |   |  |
|--------------------------------|----------------------------------|-------------------------------|---|--|
|                                | nario)                           | Baseline scenario             | Damage scenario                                   |  |
| Train type                     | Alfa Pendular                    | Alfa Pendular                 | Alfa Pendular                                     |  |
| Loading schemes                | 3 (full, half, empty)            | 3 (full, half, empty) 1 (full |   |  |
| Irregularities profile         | 1–3                              | 4                             | 1   |  |
| Speeds                         | 40–220 km/h                      | 40–220 km/h                   | 120 and 200 km/h                                  |  |
| Record duration                | 2.5–16 s                         | 2.5–16 s                      | 3 and 5 s   |  |
| Wheel flat                     | _                                | _                             | L1/L2/L3  |  |
| Polygonized wheels             | -                                | -                             | H6–8, H12–14, H19–20,<br>H29–30 with A1 and<br>A2 |  |
| Total number of analyses       | 95                               | 25                            | 140   |  |

#### 4.3.1 Configuration of SAE

The architecture of SAE is designed using the 'trainAutoencoder' algorithm from MATLAB [46]. According to Wang et al. [40], the sparsity constraints must be determined to obtain the best results. By analogy with the mechanism of the human brain, when the brain is stimulated by a given stimulus, most neurons are inhibited, so it becomes evident that a small number of neurons can lead to a better selection of the essential characteristics of the data. Table 3 displays the 16 alternative types of SAE models considered for model selection using sigmoid activation functions across all instances. The different types of SAE models considered are divided into four groups (A-D) and each of these into four subgroups (1-4). Firstly, hidden size (number of neurons in the hidden layer) and epoch values are fixed, establishing the four different groups. This allows to examine the connection between the increase in hidden units and the increase in the number of iterations. Even so, the results within each subgroup were analyzed to understand the relationship between the reduction in the coefficient of the regularization term  $(\lambda)$ as sparsity regularization ( $\beta$ ) and the percentage of activation of the hidden unit  $(\rho)$  increase. The order of magnitude of each one parameter was stipulated considering the algorithm's default values.

Each SAE model's training process involved the utilization of the scaled conjugate gradient algorithm (SCG), with a stopping criterion of either achieving a loss function (*E*) value of  $10^{-6}$  or reaching the maximum number of epochs (iterations). All models training and numerical computations were performed on PC with AMD Ryzen<sup>TM</sup> 7 3700U Mobile with Radeon<sup>TM</sup> RX Vega 10 Integrated Graphics, R7 processor and 16 GB RAM.

Each individual model was employed within the damage detection methodology, evaluating all the results at each stage of the procedure. With that, the best SAE model found was B4, which comprises the following hyperparameters:  $\lambda = 10^{-5}$ ,  $\beta = 15$  and  $\rho = 0.9$ , 6 hidden layers (k=6), sigmoid function for activation functions and with a maximum of 500 epochs. This selection is explained in the next steps of the proposed methodology. A schematic representation of the SAE architecture used is presented in Fig. 11.

#### 4.3.2 Prediction of responses

The SAE model maps the feature space into a continuous domain, enabling accurate predictions of acceleration responses even for diverse circulation characteristics. After the training process, all test responses are reconstructed with the SAE model. For a more comprehensive assessment of signal reconstruction loss, three instances for each simulated scenario of two sensors were investigated, just to see the difference between the original and reconstructed response. In Fig. 12, the recorded acceleration responses from the first pair of accelerometers (1 and 4, Fig. 3) are compared with the reconstructed responses using SAE model, highlighting the difference between both, herein called error. Errors observed during baseline passages remain consistently minimal and nearly identical, as would be expected, given that SAE training is only performed with baseline responses. However, in the event of damage, the model could not reproduce the response with the same level of accuracy.

| Group | Hidden size | Epoch | λ                | β  | ρ    | SAE model |
|-------|-------------|-------|------------------|----|------|-----------|
| A     | 3           | 250   | 10 <sup>-2</sup> | 6  | 0.15 | A1        |
|       |             |       | $10^{-3}$        | 9  | 0.30 | A2        |
|       |             |       | $10^{-4}$        | 12 | 0.60 | A3        |
|       |             |       | $10^{-5}$        | 15 | 0.90 | A4        |
| В     | 6           | 500   | $10^{-2}$        | 6  | 0.15 | B1        |
|       |             |       | $10^{-3}$        | 9  | 0.30 | B2        |
|       |             |       | $10^{-4}$        | 12 | 0.60 | B3        |
|       |             |       | 10 <sup>-5</sup> | 15 | 0.90 | B4        |
| С     | C 9         | 750   | $10^{-2}$        | 6  | 0.15 | C1        |
|       |             |       | $10^{-3}$        | 9  | 0.30 | C2        |
|       |             |       | $10^{-4}$        | 12 | 0.60 | C3        |
|       |             |       | 10 <sup>-5</sup> | 15 | 0.90 | C4        |
| D     | 12          | 1000  | $10^{-2}$        | 6  | 0.15 | D1        |
|       |             |       | $10^{-3}$        | 9  | 0.30 | D2        |
|       |             |       | $10^{-4}$        | 12 | 0.60 | D3        |
|       |             |       | $10^{-5}$        | 15 | 0.90 | D4        |

Table 3 Different network architectures and hyperparameters used for model selection



Fig. 11 Architecture of a sparse autoencoder model



Fig. 12 Difference between original and reconstructed responses at the speed of 120 km/h for  $\mathbf{a}$  baseline,  $\mathbf{b}$  wheel flat L2, and  $\mathbf{c}$  polygonal wheel H12–14

433

The increased reconstruction loss is attributed to wheel damage, which influences the dynamic response of the track. That fact introduces inaccuracies in the reconstruction of the acceleration response. Since the SAE is exclusively trained for the healthy condition, its ability to accurately reconstruct responses is compromised when confronted with data from a damaged scenario.

## 4.4 Damage index

The damage index (DI) is computed individually for each passage and each sensor, quantifying the disparity between the measured response and the response reconstructed by the trained SAE model. This computation establishes a direct correlation: higher errors reflect more pronounced accelerations generated by the vehicle on the track. It is relevant that the two types of damage are distinct in nature and were simulated on different sides of the train. This distinction reinforces the significant impact of damage location on the dynamic response obtained and, subsequently, on the resulting DI. To compute the DI, the responses that were not part of the training process are predicted using the best SAE model.

Firstly, four indexes are used to quantify the reconstruction losses between the inputs and outputs of the sparse autoencoder, the original response  $x_j$  and reconstructed response  $\hat{x}_j$ , respectively. These indexes are mathematically expressed as follows:

ORSR = 
$$10 \log_{10} \frac{\sum_{j=1}^{n} x_{j}^{2}}{\sum_{j=1}^{n} \hat{x}_{j}^{2}},$$
 (9)

$$DAI = \frac{\pi}{2g} \left( \int_{0}^{T} x_{j}(t)^{2} dt - \int_{0}^{T} \hat{x}_{j}(t)^{2} dt \right),$$
(10)

$$MAE = \frac{1}{n} \cdot \sum_{j=1}^{n} (\mathbf{x}_j - \hat{\mathbf{x}}_j), \qquad (11)$$

$$\ln(\text{MSE}) = \ln\left[\frac{1}{n} \cdot \sum_{j=1}^{n} \left(\boldsymbol{x}_{j} - \hat{\boldsymbol{x}}_{j}\right)^{2}\right],$$
(12)

where n is the number of vertical acceleration response points in a sampling time period T, and g denotes the gravitational acceleration.

Three of these demonstrated indexes were applied in damage detection works using autoencoders. The overall reconstruction signal ratio (ORSR) and the difference of Arias intensity (DAI) were damage-sensitive features evaluated in work of Wang and Cha [34] on a steel bridge model and the mean absolute error (MAE) was used in work of Sarwar et al. [5] to detect damage in a road bridge. The natural logarithmic of mean squared error (ln(MSE)) is a new damage index proposed in the current work.

Figure 13 visually presents the outcomes achieved in each specified DI, considering accelerometer number 1 with damage cases considering a train speed of 120 km/h. This graphical presentation enables to take conclusions regarding the DI from various viewpoints, given that the spatial domain is different in each index. From the graphical aspect, it is visible some similarities between ln(MSE) and ORSR, and between MAE and DAI.

With a focus on the results archive from ORSR and ln(MSE), it becomes evident that the DI obtained with ln(MSE) exhibits a more noticeable differentiation among various scenarios (undamaged, wheel flat, and polygonal wheel). Additionally, the impact of speed in the undamaged scenarios is more pronounced in the results obtained through ORSR. When comparing the results acquired using the DAI and the MAE, a resemblance is observed in the divergence of the DI across all scenarios. However, it is remarkable that the disparity among different DI is significantly greater in the case of the DAI, as opposed to from the MAE. This discrepancy may inhibit the application of cluster analysis. Due to these minor distinctions, resulting in the exclusion of the DI derived from ORSR and DAI, the assessments focused on the DI obtained from ln(MSE) and MAE across all sensors.

As illustrated in Figs. 14 and 15, estimating the DI with ln(MSE) and MAE, respectively, the impact of the damage in the track responses is more pronounced in cases involving wheel flats. However, in scenarios involving polygonal wheels, the effects of the damage are felt in both tracks, even with reduced intensity. In the case of wheel flats, the sensors positioned on the rail opposite to the damaged side (1–3) exhibit DI values relatively lower, which indicates the need for all sensors to contribute to the damage detection methodology. These results were determined with the best SAE model (B4), although at this stage of the methodology, all autoencoders presented equivalent results only with decimal changes in both DI values.

#### 4.5 Data fusion

After computing the DI, the ln (MSE) is chosen, and a data fusion technique is applied to enhance the sensitivity of the damage index. Consequently, a new damage index (DI) is obtained for each simulation. The primary goal of data fusion is to condense the extracted data while retaining the most pertinent information, specifically, to improve the ability to characterize OOR damage wheels [2]. To achieve this, the Mahalanobis distance (MD) is



Fig. 13 Comparison between different damage indices (DIs)

used to transform the multivariate data into a single DI, as applied in previous works due to its simplicity and computational efficiency [27–30]. The MD calculates the distance between the damage and baseline scenarios, thereby quantifying their similarities. Smaller MD values indicate stronger similarities between the scenarios. The Mahalanobis distance is applied to merge the ln(MSE) values of all the sensors per each passage, increasing the damage index as follows:

$$MD = \sqrt{\left(\boldsymbol{x}_{i} - \overline{\boldsymbol{x}}\right) \cdot \boldsymbol{S}_{x}^{-1} \cdot \left(\boldsymbol{x}_{i} - \overline{\boldsymbol{x}}\right)^{\mathrm{T}}},$$
(13)

where  $x_i$  is the matrix with MSE of potential damage cases,  $\bar{x}$  is the matrix with the mean of estimated MSE in the baseline scenario, and  $S_x$  is the covariance matrix of the baseline simulations. Figure 16 shows the fusion of the damage index from

the SAE model B4, highlighting the formation of two different damage groups. The idea behind merging data from all sensors is also to consider the possibility of damage on both sides.

#### 4.6 Damage detection

The present stage of the ML-based methodology for automatically detecting OOR damage wheels involves data discrimination. In this proposed approach, the outlier analysis is employed for damage detection, utilizing the damage index obtained through data fusion. To distinguish between baseline and damage scenarios, a confidence boundary (CB) is implemented. The CB is calculated using the Gaussian inverse cumulative distribution function (ICDF), considering the mean value ( $\overline{\mu}$ ) and standard deviation ( $\sigma$ ) of the baseline feature vector:



Fig. 14 Damage index ln(MSE) considering test passages with damage scenarios at a 120 km/h and b 200 km/h



Fig. 15 Damage index MAE considering test passages with damage scenarios at a 120 km/h and b 200 km/h



Fig. 16 Data fusion-merging of all sensors for damage detection for a speed of a120 km/h and b 200 km/h

$$CB = inv F_x(1 - \alpha), \tag{14}$$

where

$$F(x|\overline{\mu},\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\alpha}^{x} \exp\left[-\frac{1}{2}\left(\frac{x-\overline{\mu}}{\sigma}\right)^{2}\right] dy, \ x \in \mathbb{R}.$$
(15)

Consequently, when DI is equal to or higher than CB, the feature is an outlier. The chosen significance level is set at 1%, in line with common practices in various structural health monitoring studies to identify damage [2, 27–30]. Figure 15 illustrates the efficiency of the proposed strategy in the best SAE models of each group (A4, B4, C1, D1), demonstrated through a comparison between the CB (depicted as a red line) and different damage indexes for each of the 70 train crossings with damage. These selected models were the best ones of each group according to low number of false positives. In Fig. 17a, b, a false positive is visible, exhibited by a passage at 220 km/h with the vehicle operating at half load. For Fig. 17c, d the identification of OOR wheels is accomplished perfectly.

#### 4.7 Damage classification

Subsequently, for damage classification, a clustering process is proposed to split datasets into distinct clusters that are both compact and well-separated. In this study, the *k*-means clustering technique is adopted, utilizing the city-block distance metric. The *k*-means clustering operates as a vector quantization technique, with the objective of separate a set of *n* data points into *k* clusters, with each data point assigned to the nearest cluster center [59–61]. This automatic classification technique is widely used in damage detection works [27, 62, 63].

#### 4.7.1 Damage identification

The clustering process is automated by implementing the global 'Silhouette' index (SIL) for finding *k* clusters [2]. Based on the achieved results with data fusion matrix, Fig. 16 shows the clusters obtained from the same SAE models evaluated with the outlier analysis. Figure 18a, d shows the cluster for all polygonal wheels (cluster P), and seven misclassifications on the wheel flat cases (cluster F). Figure 18b presents the best results achieved, where the *k*-means method can cluster the two different OOR damaged wheels and the undamaged (cluster B) perfectly, which justifies the choice of model B4 as the best SAE model. The worst results are in Fig. 18c, presenting nine misclassifications on both OOR scenarios.

Additionally, for the best SAE model (B4), when implementing automatic clustering process based on the results obtained through data fusion into a single vector that combines the two speeds of each type of damage (120 and 200 km/h), *k*-means algorithmic demonstrates an effective classification performance, with only four misclassifications in wheel flats scenarios (cluster F), as shown in Fig. 19.

#### 4.7.2 Severity of damage

After the classification of type of damage, it is possible to identify the severity of each one with the best SAE model. In this step, the clustering process is defined with a matrix composed by the mean values of the MAE for all sensors (Fig. 12) and k=3 clusters to obtain only three severity levels: low (cluster 1), medium (cluster 2) and high (cluster 3), using the global validation index 'CalinskiHarabasz' [64, 65]. The main objective of this step is to show the influence of each damage on the track. As the amplitude of the dynamic response increases, so does the damage index. This



Fig. 17 Outlier analysis in SAE model: a A4; b B4; c C1; d D1

leads to the observation that higher DI values correspond to higher levels of damage severity.

Figure 20 shows the severity of wheel flats with a single misclassification in cluster 2. This event is due to the value of the amplitude of acceleration response obtained for that passage, which is of the same order of magnitude as the passages in cluster 3. The classification for severity levels was the same for both crossing speeds, clearly identifying the three simulated wheel flat scenarios.

On the polygonized wheels, the objective is to understand which harmonic orders cause greater accelerations in the track. According to polygonal profiles shown in Sect. 3.3, the classification is shown in Fig. 21, where it is easily observed that the harmonics of H12–14 and H29–30 are more harmful than those of H6–8 and H19–20 for the two speeds studied. Each level in each type of harmonic demonstrates each amplitude of defect considered (Table 1). This means that the presence of a polygonal effect of order H6–8 and H19–20, with an amplitude of defect A2, displays the same level of severity as a polygonal effect of order H12–14 and H29–30 with an amplitude A1 (cluster 2). It should be remembered that irregularity profiles of order H6–8 and H12–14 were experimentally measured for a circulation speed of 120 km/h and those of H19–20 and H29–30 for a speed of 200 km/h. Nevertheless, the classification for the severity levels was the same for both crossing speeds, which allows identifying the harmonics that exert more significant impact forces on the track.



Fig. 18 Cluster analysis in SAE model: a A4; b B4; c C1; d D1



Fig. 19 Cluster analysis in SAE B4 for all scenarios together

# **5** Conclusions

This paper introduced an automated unsupervised strategy that employs hybrid machine learning techniques for detecting and identifying damage. In a broader context, the proposed strategy involves several steps: (1) pre-process the acquired data with the space transformation of accelerations response, (2) predict the responses in SAE model, (3) determine the DI with the ln(MSE) and MAE between original and reconstructed response, (4) merge DI from all sensors, and (5) discriminate the DI through the implementation of outlier analysis for damage detection and cluster analysis for classification by type and severity of damage.



Fig. 20 Cluster analysis of wheel flat at a 120 km/h and b 200 km/h



Fig. 21 Cluster analysis of polygonal wheels at a 120 km/h and b 200 km/h

The pre-processing of the data allowed the standardization of the dynamic responses and guaranteed the validity of using ln(MSE) and MAE as DI. These two different damaged indices extracted from a SAE model comprise two objects, one consisting of identifying the type of damage (ln(MSE)) and the other based on identifying the severity of each damage (MAE). To capture the range of variability within the DI, a Mahalanobis distance metric was employed across the values of ln(MSE) associated with each sensor. This analysis revealed that various sensors exhibited varying degrees of sensitivity, contingent upon the location of the damage. This approach enabled an enhanced assessment of wheel damages, improving the overall effectiveness of the damage detection system. This step was crucial for the selection of the best SAE model. The choice of SAE B4 as the best model was due to the performance acquired in the identification phase of the type of damage; however, with this model, a false positive occurs in the detection phase. All simulated SAE models present a logic in the combination of hyperparameters to understand their impact with the number of neurons in hidden layer and the number of epochs. The B4 model corresponds to the optimal parameters, according to the purpose for which it was developed, for the autoencoder training process. With this selection, it was possible to classify the severity for each OOR scenario.

The observed lower damage index for wheel flat, in comparison with polygonal wheel, is attributed to the relatively lesser impact of flat damage on the opposing rail, as opposed to polygonal defects. This physical phenomenon created a great challenge in damage detection, given that the objective was to know about the presence and type of damage, regardless of its location. To confrontation that, some damage indexes were evaluated before the step of data fusion.

These results demonstrate the immense potential of this novel technique in the railway sector, especially concerning infrastructure management. Although the proposed methodology was specifically designed to assess singular damage in a particular railway vehicle, it could also show good performance across various vehicles if the damage index is adjusted to account for the absence of vehicle dimensions. To eliminate these gaps and enhance the current methodology, potential further developments must include different types of vehicles with multiple OOR defects and the possibility of precisely localizing damage. The main challenge of damage localization consists in the type of monitoring system that was considered, i.e., the wayside system. This adds complexity to the issue, as the goal is to localize damage on vehicle wheels based on the dynamic responses of the track. Also, as a future research work, it will be planned that a dedicated experimental campaign in which vehicles with predefined and well-characterized OOR defects will pass on a specific instrumented track section. In case of wheel flat, the defects are previously introduced on the wheels. In the case of polygonal wheels, the dominant harmonic order will be measured after a series of kilometers traveled at different speeds. This will allow to precisely validate the methodology proposed on this work.

Acknowledgements This paper is a result of project WAY4Safe-Rail—Wayside monitoring system FOR SAFE RAIL transportation, with reference NORTE-01-0247-FEDER-069595, co-funded by the European Regional Development Fund (ERDF), through the North Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement. This work was financially supported by Base Funding-UIDB/04708/2020 and Programmatic Funding-UIDP/04708/2020 of the CONSTRUCT—Instituto de Estruturas e Construções, funded by national funds through the FCT/ MCTES (PIDDAC). The authors acknowledge Grant No. 2021.04272. CEECIND from the Stimulus of Scientific Employment, Individual Support (CEECIND) - 4th Edition provided by "FCT – Fundação para a Ciência, DOI : https://doi.org/10.54499/2021.04272.CEECIND/ CP1679/CT0003".

**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

- Meixedo A, Santos J, Ribeiro D et al (2021) Damage detection in railway bridges using traffic-induced dynamic responses. Eng Struct 238:112189
- Meixedo A, Santos J, Ribeiro D et al (2022) Online unsupervised detection of structural changes using train–induced dynamic responses. Mech Syst Signal Process 165:108268
- Cury A, Ribeiro D, Ubertini F et al (2021) Structural health monitoring based on data science techniques. Structural integrity, vol 21. Springer, Cham
- Wang Z, Cha YJ (2021) Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage. Struct Health Monit 20(1):406–425
- Sarwar MZ, Cantero D (2021) Deep autoencoder architecture for bridge damage assessment using responses from several vehicles. Eng Struct 246:113064
- Yessoufou F, Zhu J (2023) Deep autoencoder model for direct monitoring of bridges subjected to a moving vehicle load under varying temperature conditions. Structures 52:752–767

- Lee H, Lim HJ, Skinner T et al (2022) Automated fatigue damage detection and classification technique for composite structures using Lamb waves and deep autoencoder. Mech Syst Signal Process 163:108148
- Wang H, Li B, Gong J et al (2023) Machine learning-based fatigue life prediction of metal materials: perspectives of physics-informed and data-driven hybrid methods. Eng Fract Mech 284:109242
- Cha YJ, Choi W, Büyüköztürk O (2017) Deep learning-based crack damage detection using convolutional neural networks. Comput Aided Civil Eng 32(5):361–378
- Cha Y, Choi W, Suh G et al (2018) Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Comput Aided Civ Infrastruct Eng 33(9):731–747
- 11. Staśkiewicz T, Firlik B (2018) Out-of-round tram wheels–current state and measurements. AoT 45(1):83–93
- Chang C, Cai Y, Chen B et al (2022) Experimental study of the wheel/rail impact caused by wheel flat within 400 km/h using full-scale roller rig. Railw Sci 1(1):76–89
- Nielsen JCO, Johansson A (2000) Out-of-round railway wheelsa literature survey. Proc Inst Mech Eng Part F J Rail Rapid Transit 214(2):79–91
- 14. Ye Y, Zhu B, Huang P et al (2022) OORNet: a deep learning model for on-board condition monitoring and fault diagnosis of out-of-round wheels of high-speed trains. Measurement 199:111268
- Iwnicki S, Nielsen JCO, Tao G (2023) Out-of-round railway wheels and polygonisation. Veh Syst Dyn 61(7):1787–1830
- Wu Y, Du X, Zhang HJ et al (2017) Experimental analysis of the mechanism of high-order polygonal wear of wheels of a high-speed train. J Zhejiang Univ Sci A 18(8):579–592
- Cai W, Chi M, Wu X et al (2019) Experimental and numerical analysis of the polygonal wear of high-speed trains. Wear 440–441:203079
- Vale C (2021) Wheel flats in the dynamic behavior of ballasted and slab railway tracks. Appl Sci 11(15):7127
- Amini A, Entezami M, Papaelias M (2016) Onboard detection of railway axle bearing defects using envelope analysis of high frequency acoustic emission signals. Case Stud Nondestruct Test Eval 6:8–16
- Wu Y, Wang J, Liu M et al (2022) Polygonal wear mechanism of high-speed wheels based on full-size wheel-rail roller test rig. Wear 494–495:204234
- Ye Y, Wei L, Li F et al (2023) Multislice time-frequency image entropy as a feature for railway wheel fault diagnosis. Measurement 216:112862
- 22. Mosleh A, Montenegro PA, Costa PA et al (2021) Railway vehicle wheel flat detection with multiple records using spectral kurtosis analysis. Appl Sci 11(9):4002
- Costa C, Ribeiro D, Jorge P et al (2015) Calibration of the numerical model of a short-span masonry railway bridge based on experimental modal parameters. Procedia Eng 114:846–853
- Meixedo A, Ribeiro D, Calçada R, et al (2014) Global and local dynamic effects on a railway viaduct with precast deck. In: The second international conference on railway technology: research, development and maintenance, Ajaccio, Corsica, France, 8–11 April 2014
- Kudva JN, Munir N, Tan PW (1992) Damage detection in smart structures using neural networks and finite-element analyses. Smart Mater Struct 1(2):108–112
- Lieu QX (2023) A deep neural network-assisted metamodel for damage detection of trusses using incomplete time-series acceleration. Expert Syst Appl 233:120967
- 27. Silva R, Guedes A, Ribeiro D et al (2023) Early identification of unbalanced freight traffic loads based on wayside monitoring and artificial intelligence. Sensors 23(3):1544

- Guedes A, Silva R, Ribeiro D et al (2023) Detection of wheel polygonization based on wayside monitoring and artificial intelligence. Sensors 23(4):2188
- 29. Mosleh A, Meixedo A, Ribeiro D et al (2023) Early wheel flat detection: an automatic data-driven wavelet-based approach for railways. Veh Syst Dyn 61(6):1644–1673
- Mohammadi M, Mosleh A, Vale C et al (2023) An unsupervised learning approach for wayside train wheel flat detection. Sensors 23(4):1910
- Gonzalez I, Karoumi R (2015) BWIM aided damage detection in bridges using machine learning. J Civ Struct Health Monit 5(5):715–725
- 32. Yang K, Kim S, Harley JB (2023) Unsupervised long-term damage detection in an uncontrolled environment through optimal autoencoder. Mech Syst Signal Process 199:110473
- 33. Li L, Morgantini M, Betti R (2023) Structural damage assessment through a new generalized autoencoder with features in the quefrency domain. Mech Syst Signal Process 184:109713
- Wang Z, Cha YJ (2022) Unsupervised machine and deep learning methods for structural damage detection: a comparative study. Eng Rep 2022:e12551
- Pathirage CSN, Li J, Li L et al (2018) Structural damage identification based on autoencoder neural networks and deep learning. Eng Struct 172:13–28
- 36. Finotti RP, de Souza BF, Cury AA et al (2021) Numerical and experimental evaluation of structural changes using sparse autoencoders and SVM applied to dynamic responses. Appl Sci 11(24):11965
- 37. Ye Y, Huang C, Zeng J et al (2023) Shock detection of rotating machinery based on activated time-domain images and deep learning: an application to railway wheel flat detection. Mech Syst Signal Process 186:109856
- Meng Q, Catchpoole D, Skillicom D, et al (2017) Relational autoencoder for feature extraction. In: 2017 international joint conference on neural networks (IJCNN). Anchorage, AK, USA. IEEE, pp 364–371
- Mosleh A, Meixedo A, Ribeiro D et al (2023) Automatic clustering-based approach for train wheels condition monitoring. Int J Rail Transp 11(5):639–664
- 40. Wang P, Li C, Liang R et al (2023) Fault detection and calibration for building energy system using Bayesian inference and sparse autoencoder: a case study in photovoltaic thermal heat pump system. Energy Build 290:113051
- Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? Vis Res 37(23):3311–3325
- 42. Montenegro PA, Calçada R (2023) Wheel-rail contact model for railway vehicle-structure interaction applications: development and validation. Railw Eng Sci 31(3):181–206
- 43. Montenegro PA, Neves SGM, Calçada R et al (2015) Wheelrail contact formulation for analyzing the lateral train-structure dynamic interaction. Comput Struct 152:200–214
- Hertz H (1882) Ueber die Berührung fester elastischer Körper. Journal für die reine und angewandte Mathematik (Crelles Journal) 1882(92):156–171
- 45. Kalker JJ (1996) Book of tables for the Herzian creep-force law. Faculty of Technical Mathematics and Informatics, Delft University of Technology, Delft
- The MathWorks Inc (2018) MATLAB<sup>®</sup>, R2018a. Natick, Massachusetts

- 47. Ansys Inc (2018) ANSYS<sup>®</sup>, Release 19.2. Academic Research, Canonsburg P
- Ribeiro D, Calçada R, Brehm M et al (2021) Calibration of the numerical model of a track section over a railway bridge based on dynamic tests. Structures 34:4124–4141
- Ribeiro D, Calçada R, Delgado R et al (2013) Finite-element model calibration of a railway vehicle based on experimental modal parameters. Veh Syst Dyn 51(6):821–856
- Mosleh A, Montenegro P, Alves Costa P et al (2021) An approach for wheel flat detection of railway train wheels using envelope spectrum analysis. Struct Infrastruct Eng 17(12):1710–1729
- 51. Vale C, Calçada R (2010) Dynamic response of a coupled vehicle-track system to real longitudinal rail profiles. In: Proceedings of the tenth international conference on computational structures technology, Valencia, Spain, 14–17 Sept. 2010
- 52. European Committee for Standardization (2006) Railway applications—track—track geometry quality—part 2: measuring systems—track recording vehicles (EN 13848-2)
- Mosleh A, Costa PA, Calçada R (2020) A new strategy to estimate static loads for the dynamic weighing in motion of railway vehicles. Proc Inst Mech Eng Part F J Rail Rapid Transit 234(2):183–200
- Casas JR, Moughty JJ (2017) Bridge damage detection based on vibration data: past and new developments. Front Built Environ 3:4
- Mu J, Zeng J, Huang C et al (2022) Experimental and numerical investigation into development mechanism of wheel polygonalization. Eng Fail Anal 136:106152
- 56. Tao G, Xie C, Wang H et al (2021) An investigation into the mechanism of high-order polygonal wear of metro train wheels and its mitigation measures. Veh Syst Dyn 59(10):1557–1572
- Zhang J, Han GX, Xiao XB et al (2018) Influence of wheel polygonal wear on interior noise of high-speed trains. J Zhejiang Univ Sci A15:1002–1018
- 58. Peng B (2020) Mechanisms of railway wheel polygonization. Dissertaion, University of Huddersfield
- Hu H, Liu J, Zhang X et al (2023) An effective and adaptable K-means algorithm for big data cluster analysis. Pattern Recognit 139:109404
- Andrade Nunes L, Piazzaroli Finotti Amaral R, de Souza BF et al (2021) A hybrid learning strategy for structural damage detection. Struct Health Monit 20(4):2143–2160
- Alves V, Cury A, Cremona C (2016) On the use of symbolic vibration data for robust structural health monitoring. Proc Inst Civ Eng Struct Build 169(9):715–723
- Barile C, Casavola C, Pappalettera G et al (2022) Laplacian score and K-means data clustering for damage characterization of adhesively bonded CFRP composites by means of acoustic emission technique. Appl Acoust 185:108425
- 63. Meixedo A, Ribeiro D, Santos J et al (2022) Real-time unsupervised detection of early damage in railway bridges using trafficinduced responses. Structural health monitoring based on data science techniques. Springer, Cham, pp 117–142
- Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. Commun Stat Theory Meth 3(1):780133830
- Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF (2016) A new hybrid filter–wrapper feature selection method for clustering based on ranking. Neurocomputing 214:866–880