

# ***IET Image Processing***

## **Special issue Call for Papers**

---

**Be Seen. Be Cited.  
Submit your work to a new  
IET special issue**

Connect with researchers and  
experts in your field and share  
knowledge.

Be part of the latest research  
trends, faster.

**Read more**



The Institution of  
Engineering and Technology

# Method for automatic railway track surface defect classification and evaluation using a laser-based 3D model

ISSN 1751-9659

Received on 29th November 2019

Revised 12th May 2020

Accepted on 29th May 2020

E-First on 8th September 2020

doi: 10.1049/iet-ipr.2019.1616

www.ietdl.org

Jiaqi Ye<sup>1</sup>, Edward Stewart<sup>1</sup>, Dingcheng Zhang<sup>1</sup> ✉, Qianyu Chen<sup>1</sup>, Clive Roberts<sup>1</sup><sup>1</sup>Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham B15 2TT, UK

✉ E-mail: railCM@contact.bham.ac.uk

**Abstract:** Inspection of physical surface defects is a significant concern in many industrial areas. In railway systems, this process mainly includes the detection and classification of defects in rails and wheels, for which laser-based optical inspection technologies have gradually been applied in the form of 2D profile measurement, benefiting from its high precision and robustness to surface conditions. However, defect classification and evaluation after the initial detection works still rely heavily on human inspectors to make maintenance suggestions. The linear nature of rails makes it possible to increase the dimension of rail measurement data from 2D to 3D by aligning 2D profiles along the rail, from which more comprehensive diagnosis information becomes available. In combination with appropriate artificial intelligence algorithms, this approach can potentially replace human-dominated defect classification and evaluation work. This study presents a 3D model-based railway track surface defect classification and evaluation method. A set of geometrical features are extracted from the 3D model of track surface defects to describe a distinguishable pattern for each category of defect. Multi-class classifiers are then tested and have shown promising results on a group of artificial track surface defects, giving a systemic solution for 3D model-based automatic track surface defect inspection.

## 1 Introduction

The rapid development of high-speed railway has greatly increased the volume of railway transportation but has also resulted in demands for more efficient and intelligent rail inspection methods to ensure the safety and performance of the railway network. Rail inspection is a process of testing rails to determine and report the presence of defects in accordance with relevant standards (e.g. Network Rail's NR/L2/TRK/001 Inspection and Maintenance of Permanent Way) [1]. Specific works mainly include defect detection and classification, based on which corresponding remedial actions can be taken to prevent further incidents. Traditionally, human inspectors are the main actors for rail inspection, which can be inefficient and subjective, and also have safety concerns. In recent decades, efforts have been made in both universities and railway organisations to develop non-destructive testing (NDT) technologies. These methods are mainly based on advanced sensory and measuring technologies such as ultrasonic, magnetic, eddy current, visual, and laser inspection. As one of the NDT technologies, laser inspection refers to the use of laser instruments based on the principles of coherent light reflection, and has been widely used for inspection of rail surface defects and geometry due to its high precision and truly non-contact property.

Rail surface defects can be generated by manufacturing faults, improper handling, and rolling contact fatigue (RCF), among which RCF defects are difficult to control due to their complex mechanism of formation and long-term impact on the rail, and have become the major concern to the railway industry especially under the shorter maintenance window of high-speed rail [2]. RCF defects normally originate from rolling contact pressure, shear, and bending forces from the vehicle weight, and thus they are mostly cantered on the rail head regions that make contact with the wheel. Based on their geometries, RCF defects can be mainly divided into three categories: cracks, shelling, and squats [3]. Further propagation of RCF defects can result in rail breakages, which are still one of the main hazards of major train accidents (i.e. derailments) [4]. Therefore, developing advanced inspection technologies to inspect RCF defects more efficiently and accurately is always significant to improve the safety and performance of railway operations.

In railway systems, laser-based inspection is commonly implemented in the form of 2D profile measurement. With laser sensors mounted on carts or trains, high-resolution profiles can be captured to represent the geometry of the rail surface, on which the surface conditions of the rail can be evaluated [5, 6, 7, 8]. These profile-based defect detection methods are generally based either on a comparison with pre-measured standard profiles or on thresholding techniques. They have shown good experimental results in the laboratory, but complex inspection cases in the real world are still difficult to consider using these methods. For example, the range of tolerance for a comparison-based method is difficult to choose due to unstable measuring conditions, and it is also impossible to create a diversity of thresholding techniques covering the defects that have a wide variety of sizes in the real world [9]. The substantial improvement of 3D technologies has pushed their application in many areas such as virtual reality, robot navigation, medical imaging, and architectural restoration of cultural relics [10, 11, 12, 13]. Accordingly, the authors' previous research explored the feasibility of applying related laser-based 3D techniques into railway systems for detecting physical surface defects. The 2D profiles-based point cloud data is reconstructed to a 3D model, on which multiple geometrical features of the rail surface [depth gradient (DGD), face normal (FND), and face normal gradient (FNGD)] are generated. Each geometrical feature has been shown to align with the detection of a specific type of RCF defect (cracks, squats, or shelling) [14]. Compared to the defect detection methods mentioned above, 3D model-based defect detection neither relies on comparisons with standard models nor uses thresholds, but analyses the geometrical features of the defects in 3D, and explores their different characteristics such as the depth gradient of the nearest points and the face normal vector of each mesh element in different categories of defect, which are not accessible from 2D data sets. The promising results from both rail samples in the laboratory and the field test track demonstrated that appropriate geometrical features from the 3D model can provide sufficient information for RCF defect detection. As a further step of rail inspection, defect classification is often considered afterwards since the features that help to achieve defect detection can potentially provide effective information for defect classification as well.

## 2 Related works

In recent years, the use of machine learning and pattern recognition technologies has gradually replaced many conventional human-dominated object classification works. In machine learning and pattern recognition areas, classification is a process to find the mathematical regularity between the feature vectors ( $x$ ) and classes ( $y$ ) of the data set with a certain amount of pre-defined or undefined training data  $\{(x, y)\}$ . The function  $y' = f(x)$  that defines the regularity is called the classifier, which can then be used for the classification of an unknown data set [15]. Generally, this process includes data collection and pre-processing, feature extraction and selection, designing and training the classifier or clustering, and validation of the classifier or clusters.

For physical surface defect classification, images or video frames are the most commonly used data sources since they are relatively easy to access, from which informative features, such as the edges and textures of the object that are similar to the features we use to distinguish different types of defect with the human eye, are accessible [16]. From Bayesian models to neural networks, a large number of surface defect classification models have been developed and applied in other domains for a variety of surfaces such as steel, wood, wafers, ceramics, and fabric, to replace the process of manual operation [9, 17, 18]. Xue-Wu *et al.* developed a classification method for surface defects in strongly reflective metals. Five features were extracted from spectral texture measures of copper strips using wavelet transformations and a Sobel filter. Definition of the hyperplane of a support vector machine (SVM) and its high generalisation performance make it a good candidate for linearly non-separable and unknown data. A radial basis function kernel SVM-based classifier was trained with an accuracy of up to 91.3% [17]. A similar method developed by Chondronasios *et al.* focuses on extruded aluminium profiles and uses a Sobel edge detector to obtain the gradient magnitude from a data set of 145 extruded aluminium surface images. The features extracted from the gradient magnitude are then classified by a two-layer feed-forward artificial neural network, with which three classes, non-defective, blister, and scratches, can be defined with a best testing accuracy up to 98.6% [9]. However, the uncertainty of neural networks makes it difficult to further analyse the high bias of their performance with different topologies. A more comprehensive survey of vision-based defect detection and classification has been done by Kumar, in which hundreds of sources were reviewed to give an objective introduction to different classification methods [19]. Vision-based defect classification works rely heavily on image quality. High-quality images can give a clear representation of defects, from which specific mathematical features or feature vectors can be derived. The informative features can then be used to build the qualified classifier for objects.

However, due to the fact that rail faults are randomly distributed, and the in-service rail is currently mostly operated under the healthy condition with a rapid response to faults, it is impossible to collect a sufficient number of defective samples as training data to build an automatic defect classification model. In addition, the rail surface in the real world can be dark and oxidised, which makes some surface defects difficult to distinguish from 2D images. Furthermore, the running surface of the rail in use is usually reflective and can present different conditions in different weathers, which leads to problems with illumination [1, 20]. Overall, limited by both the quantity and quality of targeted rail samples, there have been few mature automatic vision-based rail surface defect classification approaches. Related works are mainly based on periodic inspections implemented by competent engineers using either mechanical contact measuring instruments or NDT technologies, and then manual generation of the defect classification report indicating the key information of the defects such as their position, exact type, and maintenance suggestions. The huge railway network and the randomly distributed surface defects cost a lot of manpower and resources. Therefore, developing an automatic defect classification method is meaningful to improve the performance of rail inspection from economic, efficiency, and objectivity points of view.

Benefiting from the development of laser measuring instruments, high-precision point cloud data can be captured and applicable for surface measurement. Meanwhile, the highly collimated laser beam makes it possible to keep high precision without the influence of surface conditions such as a shiny or oxidised rail surface in the real world [21]. Further research on laser-based surface inspection shows that those characteristics commonly used for vision-based defect classification such as edges, size, and textures are also accessible or analogically accessible from the 3D model reconstructed from laser sensor data [22, 23, 24]. At the same time, previous research has demonstrated that the geometrical features derived from artificial defects in the laboratory can provide sufficient diagnostic knowledge for the data from real cases. That is because artificial defects are made on a real rail section and mimic the geometrical characteristics of real defects in 3D. Thus, there are relatively small discrepancies between the geometrical features extracted, respectively, from artificial defects in the laboratory and defects in the real world, which makes artificial defects act as alternative research objects. For example, Papaelias *et al.* used a spinning rail containing artificially induced notches to test an alternating current field measurement technique [25]. Therefore, laser-based 3D models can potentially overcome the lack of defective samples in practice and provide sufficient attributes for developing automatic classification models.

Inspired by this, to pursue the previous research, a 3D model-based classification method to classify and evaluate RCF defects is proposed. The research aims to fill the gaps in the existing RCF defect classification method from the following points: (i) increasing the dimension of data sources from 2D images to laser-based 3D models to provide a more reliable and comprehensive description for physical surface defects; (ii) extracting appropriate geometrical features to describe dedicated patterns for RCF defects; (iii) training automatic classifiers to reduce human involvement in RCF defect classification and evaluation; (iv) continuously improving 3D model-based rail inspection by linking the detection and classification of RCF defects as a complete approach.

Section 3 introduces the profile laser-based 3D perceptual system, the process of data acquisition, 3D rail head surface reconstruction, and RCF defect detection and extraction. Six geometrical features are extracted from the 3D model of RCF defects, which are then used to indicate and build appropriate classifiers. Section 4 further evaluates these features, validates multi-class classifiers, and analyses the classification results. Finally, Section 5 summarises the work.

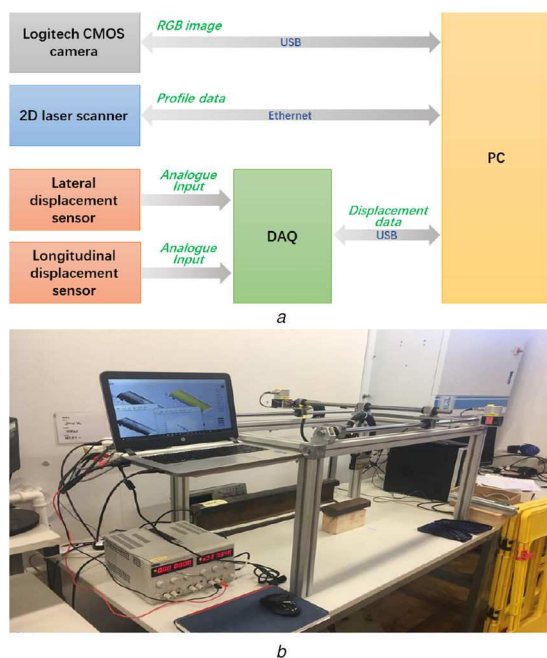
## 3 Methodology

### 3.1 Laser-based 3D perceptual system

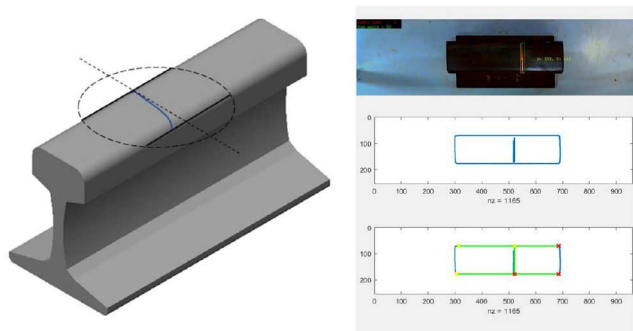
Fig. 1a shows the structure of the 3D perceptual system developed in this project for surface measurement. It consists of a low-cost 2D blue laser sensor from Micro-Epsilon which offers more precision due to a smaller spot size ( $\sim 12 \mu\text{m}$ ), high resolution (maximum 1280 points/profile), robustness against reflective and ambient light distortion, two draw-wire displacement sensors (WDS-1500-P60), a data acquisition device giving a resolution of the displacement measurement of up to 0.023 mm (NI USB-6210), and a self-designed data buffer and processing software interface running on a separate PC in a C++ environment. All these devices are integrated into a four-footed mechanical frame to perform surface profile measurement over the object with two degrees of freedom (Fig. 1b). Other specifications have been introduced in previous explorative research [14], so it is unnecessary to go into details here.

One of the significant improvements of the system is the integration of a camera module, added to the system to solve the laser sensor alignment problem which refers to the fact that the relative angle between the laser line and the rail edge is unknown at the time of the scan. However, it is expected that the laser line is perpendicular to the rail edge, so that the profiles can be aligned along the longitudinal direction of the rail to generate a coherent rail surface in the data registration section. This presently relies on





**Fig. 1** Rail head surface 3D perceptual system  
(a) System mechanism, (b) Completed system in the lab



**Fig. 2** Identification of laser sensor alignment angle



**Fig. 3** Defective rail samples with artificial RCF defects

manual calibration by a skilled operator and is imprecise. However, in practice, the laser line can be invisible to the human eye due to working conditions and partial direct reflection from the rail surface, especially with the influence of ambient light. It is therefore difficult and time consuming to ensure the correct localisation of the laser sensor relative to the rail.

To achieve automatic laser line alignment identification, a camera module is fixed to the laser sensor, similar to the built-in camera of the laser sensor itself which receives the diffusely reflected light of the laser line onto a highly sensitive sensor matrix to calculate the position of each measuring point [26]. A CMOS camera, which is more sensitive to the diffusely reflected laser light

than the human eye, is used to assist the localisation of the laser sensor. By extracting the blue colour space of the RGB image from the camera, the blue components that mainly contain the pixels of the blue laser line can be extracted. Then, Otsu's threshold method is used to extract the laser line from the background since the luminance of the pixels constituting the laser line is higher than that of other pixels [27]. To get the relative angle of the laser line to the rail edge, the rail edge needs to be identified as well. Thanks to the optical coherence of laser light [21], luminance-based identification of the laser line is more robust and accurate even with the disturbance of external light. After the laser line is identified, rather than applying edge detection to the whole image, the region around the laser line is extracted to reduce the computational complexity. The region of interest (ROI) is cropped based on the specification of the rail and the region highlighted by the laser to ensure the inherent rail edge can be included. From this, the rail edge can be identified easily by Canny edge detection and Hough transform methods. Fig. 2 shows the basic principles behind the automatic laser line alignment identification, and example results for the camera module. The integrated camera module with dedicated object identification algorithms helps to indicate and calibrate the position of the laser sensor before carrying out the measurement. This eliminates the error from manual operation, and thus improves the efficiency and accuracy of the original system.

### 3.2 Data acquisition and pre-processing

With the dedicated 3D perceptual system, the next step is to collect an adequate number of rail samples containing RCF defects from which the training data can be generated. As a research based on the geometrical characteristics of defects, collecting a sufficient number of defective samples from in-service rails to represent the geometrical characteristics of specific types of defect is impossible; the small discrepancy between the artificial defects and real defects makes the artificial defects act as alternative research objects.

Accordingly, a defect-free UIC 60 standard rail section [1000 mm ( $L$ )  $\times$  72 mm ( $W$ )  $\times$  172 mm ( $H$ )] was used in this research. The rail section was firstly cut into six subsections. Based on analysis of the geometrical characteristics of RCF defects (cracks, squats, shelling), 20 artificial defects were made for each category, thus giving 60 artificial RCF defects in total. Consequently, defects belonging to the same category have similar geometrical characteristics, and at the same time each artificial defect has different dimensions from the others. This ensures that all the samples are independent and identically distributed in statistics. Fig. 3 shows the completed defective rail samples, with the quantitative information for rail samples (length and width) and defects shown in Fig. 4.

These subsections can then be measured by the 3D perceptual system. The point cloud collected from the system represents the surface of objects with scattered data points. To give them a computer-based description to represent the geometrical characteristics of the rail, the same point cloud reconstruction method that has been verified in previous research is applied [14]. Fig. 5 shows the 3D reconstruction process for one defect of each type, with the raw data points in the first row. The customised method then filters the noisy points such as the rail web and ground, interpolates the raw data points, and builds the external surface of objects with quadrilateral mesh elements considering the symmetry of the rail surface about both  $X$  and  $Y$  axes in 3D space. The second row of Fig. 5 shows the reconstructed 3D models. It can be seen that defects correspond to regions where the surface points do not align with the general plane. This discrete change in the depth of points can be visualised as colour differences compared to the surrounding points as shown in the depth-coloured 3D models in the third row of Fig. 5. A simple, and therefore fast, thresholding solution is used to define ROI that encompass these deviations from the general plane. These ROI, indicated by a rectangular box in the figure, help reduce the computational loads of the following defect extraction procedure. The fourth row shows the extracted defects in 3D by using dedicated defect detection algorithms (DGD, FND, FNGD) [14]. They are the targeted objects

Sample	Defect	Surface length <sup>1</sup> (mm)	Maximum depth <sup>2</sup> (mm)	Surface angle <sup>3</sup> (°)/ width <sup>4</sup> (mm)	Cross-section view (approx. geometry)
1 164.0 (L) × 60.0 (W)	Crack × 10	10.0~15.0	1.60~2.40	90 (1.50 mm)	
2 164.0 (L) × 60.0 (W)	Crack × 10	17.0~27.0	3.00~5.60	90 (1.50 mm)	
3 164.0 (L) × 60.0 (W)	Squat × 10	6.10~7.0	1.10~1.60	6.0~7.0	
4 164.0 (L) × 60.0 (W)	Squat × 10	8.80~9.70	1.30~1.70	8.40~9.20	
5 164.0 (L) × 60.0 (W)	Shelling × 10	1.95~3.30	2.80~4.30	1.80~3.10	
6 164.0 (L) × 60.0 (W)	Shelling × 10	3.70~5.30	3.50~5.30	3.50~5.30	

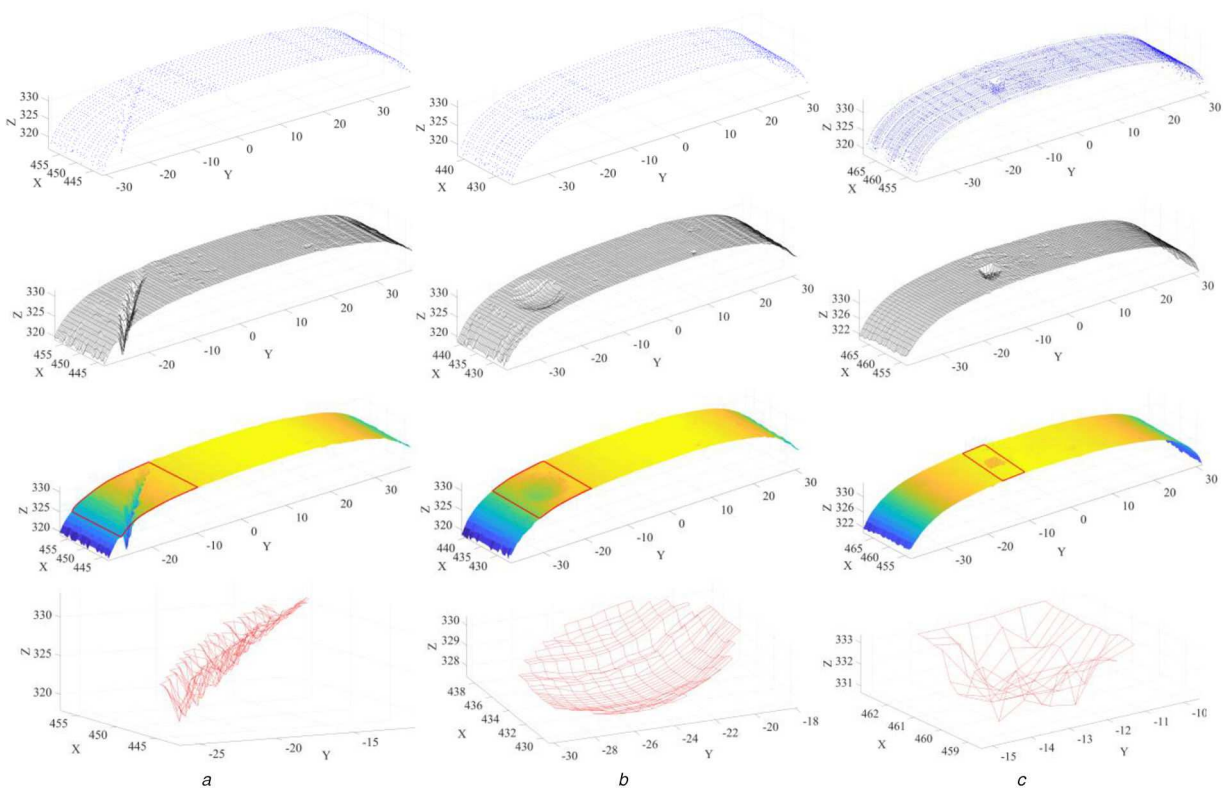
<sup>1</sup>Surface length refers to the size of the defect in the longitudinal direction of the rail.

<sup>2</sup>Maximum depth is the maximal distance between the bottom of the defect and the original surface of the rail.

<sup>3</sup>Surface angle is the cutting angle when making the artificial defects.

<sup>4</sup>Surface width refers to the size of the defect in the lateral direction of the rail.

**Fig. 4** Quantitative information of rail samples and artificial defects



**Fig. 5** Reconstructed 3D subsections and extracted 3D surface defects

to be classified in this research. With the controllable interpolation density, the size of each mesh element can be set to around  $1 \text{ mm}^2$  to recover the geometrical characteristics of the defect in as much detail as possible. Further informative features of the defects can be derived from them and can then be fitted into the following model learning and validation procedures.

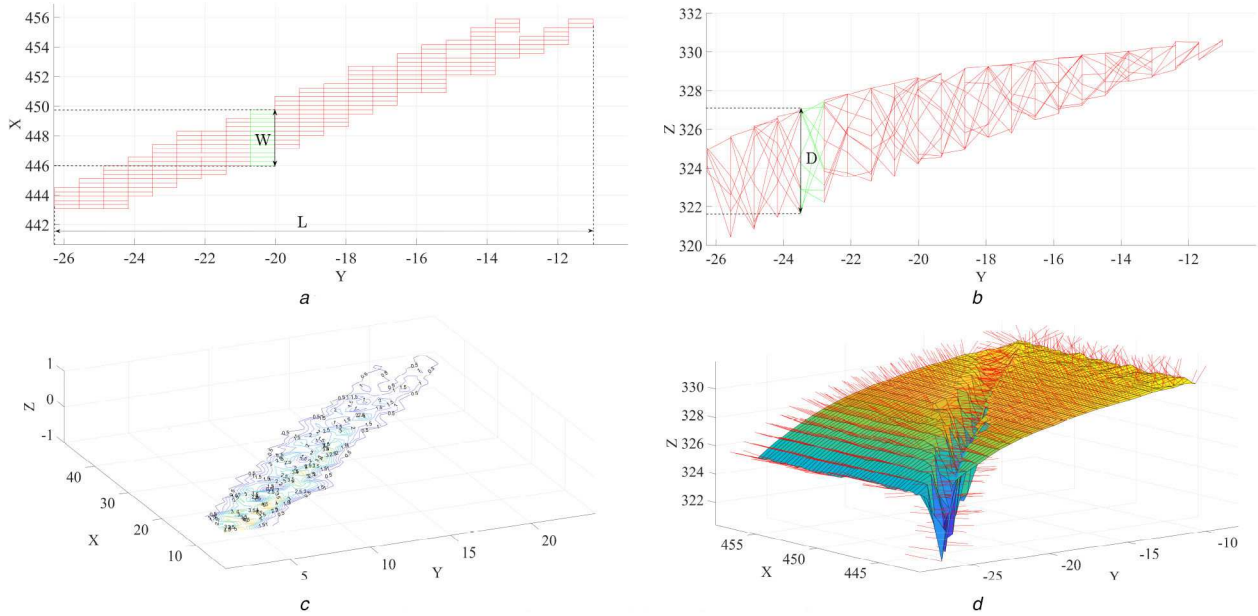
### 3.3 Quantified geometrical feature extraction

By this stage, the targeted objects are still some abstract geometries in 3D (the fourth row of Fig. 5). To build the specific pattern or description for each of them to perform classification, feature extraction is needed. In machine learning, feature extraction is a process of evaluating and extracting informative features from the original data set. These features can be a set of variables or vectors that contribute to the distributional difference of samples among each other. The correlation and difference of the pattern can be associated with the corresponding class, from which the classifiers can be trained [28]. To obtain effective quantified features from the 3D model of defects, in accommodation with the geometrical

features used for 3D model-based RCF defect detection [14], the features commonly used for vision-based object classification have been referenced, considering the similarity between the pixels of a 2D image and mesh elements of a 3D model. Features of vision-based object classification are generally some measurable characteristics of the object such as shape, colour, and texture, which can be quantified and then used to categorise the object into the corresponding class [29].

This paper firstly considers the features including the length and width that are related to the shape of the defects. In vision-based object classification, colour-based features generally refer to the colour information of each pixel and the variabilities of neighbouring pixels [19]. In a 3D point cloud, each mesh element is similar to a pixel in a 2D image containing the coordinate information; the difference is that the colour information of each pixel is replaced by the depth of each mesh element in 3D space. As such, the depth of each mesh element and the depth gradient magnitude that quantify the variability of neighbouring mesh elements are taken into consideration. The neighbouring mesh elements can be selected with a pre-defined sliding window with a





**Fig. 6** Visualisation of the informative features

(a) Top view of the 3D model, (b) Side view of the 3D model, (c) Contour plot containing the isolines of matrix  $Z$ , (d) Surface normal of mesh elements as radiating vectors

minimum size of  $2 \times 2$  to ensure at least one element (consisting of 4 points) is covered each time. The depth gradient magnitude indicates the greatest rate of depth increase in the region covered, which can be expressed as

$$|\nabla H(x_i, y_i)| = \max(H(x_i, y_i)) - \min(H(x_i, y_i)) \quad (1)$$

where  $H(x_i, y_i)$  represents the depth of points constituting the mesh element, and max and min are the values of maximum depth and minimum depth, respectively.

Similar to the texture of a 2D image, which is the spatial arrangement of the colour or intensities in a region of an image, the mesh element refers to a set of neighbouring points and the surface comprised by these points that give the spatial characteristics of the 3D model. In vision-based classification, textural features derived from the spatial domain of an image, such as co-occurrence matrix-based features, are widely used [30]. Co-occurrence matrices characterise the stochastic properties of the spatial distribution (e.g. contrast, homogeneity, and correlation) of an image. Similarly, face normal vectors describe the spatial orientation of mesh elements of the 3D model, which provide similar information as co-occurrence matrix-based features do for 2D images. Thus, the face normal vector of each mesh element is analogically considered as well. The quantified face normal vectors are the angle between the face normal of the reference plane ( $X$ - $Y$  plane) and the face normal of the mesh element in 3D space, which indicate fluctuation in the spatial orientation of each mesh element caused by defects, and can be represented as

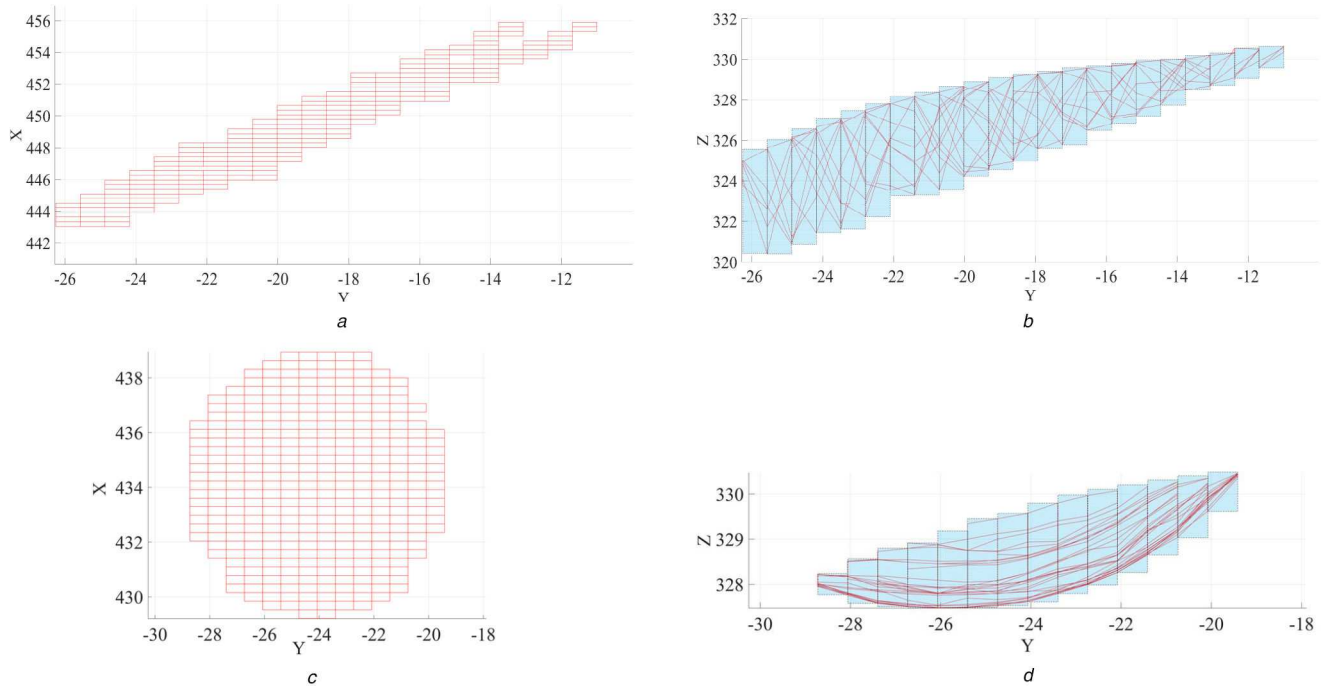
$$\begin{aligned} \beta &= 90^\circ - \gamma \\ &= 90^\circ - \tan^{-1}\left(\frac{\Delta z}{\sqrt{(\Delta x)^2 + (\Delta y)^2}}\right) \end{aligned} \quad (2)$$

where the face normal angle of the reference plane is defined as  $90^\circ$ ,  $\gamma$  is the face normal angle of the mesh element,  $\tan^{-1}$  is the arc-tangent function, and  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  are the differences between the two vertices constituting the surface normal vector in each axis in 3D space.

The aforementioned geometrical features including length ( $L$ ), width ( $W$ ), depth ( $D$ ), depth magnitude (DM), and face normal (FN) vectors of the 3D model of surface defects provide a simplified description of the defects, which can be used as candidate descriptors for 3D model-based defect classification. For visual ease and understanding, Fig. 6 visualises these features in the 3D model of a surface crack. Fig. 6a is the top view of the 3D

model. It can be seen that the nearest-neighbour interpolation applied in the process of 3D reconstruction distributes the data points uniformly about the  $X$ - $Y$  plane, and thus the quadrilateral mesh elements have the same size, which is decided by the interpolation density. Consequently, the  $L$  and  $W$  of the defect can be easily calculated by counting the number of mesh elements in the corresponding direction. The side view of the 3D model in Fig. 6b indicates the  $D$  of the defect. It is worth noting that the maximal  $L$ ,  $W$ , and  $D$  in the corresponding directions ( $Y$ ,  $X$ ,  $Z$ ) are used to ensure the uniformity of feature measurement for all the samples. Fig. 6c is the contour plot containing the isolines of matrix  $Z$ , where  $Z$  contains the DM values of the corresponding mesh elements relative to the  $X$ - $Y$  plane. Fig. 6d displays the surface normal of mesh elements as radiating vectors, where the surface normal of mesh elements around the defect is displayed for easy viewing, but not used when generating the corresponding quantified feature. It can be seen that each of the extracted 3D defects has multiple DM and FN values. To generate a value to represent the average level of these two features in each defect, the average values of the DMs and FNs are used.

In addition to the classification of defects, the evaluation of defect severity is also significant for guiding further maintenance operations such as rail grinding or replacement. In practice, related works generally rely on the experience of human observers to evaluate the size of the detected defect, which is not efficient and can be subjective [1]. 3D characterisation of the defect gives the opportunity to quantify the size of the defect in 3D. Consequently, the volume of defects is taken into consideration to reflect their severity, which is not accessible with other inspection methods. The difficulty is that the spatial structure of the defect is irregular. To calculate the volume of irregular geometry, the most commonly used method is to dissect the object into a set of simpler parts whose volume is easy to compute or approximate [31]. The volume of the entire object can then be obtained by summing up the volumes of a stack of these simpler dissections. Taking the volume calculation of a rail surface crack and a rail surface squat as examples, Figs. 7a and c show the top view of 3D models of the defects about the  $X$ - $Y$  plane. The quadrilateral mesh elements make it possible to divide the volume of defects into a set of rectangular solids. Figs. 7b and d are the side view of the extracted defects, from which the depths of corresponding mesh elements indicated with rectangular boxes are accessible. The linear nature of the rail makes all the mesh elements in the same column about the  $Y$ -axis have the same reference height to compare, which can be generated from the edge mesh elements of the 3D model. Consequently, the volume of a defect with irregular geometry can be approximated by rectangular solids of given bottom areas and depths.



**Fig. 7** Volume approximation of the defect

(a), (c) Top view of 3D models of the defects about the X-Y plane, (b), (d) Side view of the extracted defects

**Table 1** Subset of the full feature set for 3D model-based RCF defect classification

Class	Length, mm	Width, mm	Depth, mm	Feature	Face normal (avg.)	Volume, mm <sup>3</sup>
1. Crack_L1	15.34	1.70	1.80	Depth magnitude (avg.)	31.23	23.47
2. Crack_L2	24.96	2.70	4.30		34.13	144.89
3. Squat_L1	6.36	6.20	1.0		17.59	8.82
4. Squat_L2	9.91	9.30	1.60		17.89	34.33
5. Shelling_L1	3.54	3.30	1.90		33.39	17.43
6. Shelling_L2	5.62	5.10	4.60		51.26	103.55

If we assume the volume of a 3D surface defect consists of (or can be refilled with)  $k$  rectangular solids, the total volume of these solids can be expressed as  $V_{\text{solids}}$ . Based on the principle of a definite integral, the volume of the 3D defect  $V$  can be approximated when  $\Delta l \rightarrow 0$  and  $\Delta w \rightarrow 0$  with the increase of  $k$

$$V_{\text{solids}} = \lim_{\substack{\Delta l \rightarrow 0 \\ \Delta w \rightarrow 0}} \sum_k \Delta l \times \Delta w \times D_k = V \quad (3)$$

where  $\Delta l$  and  $\Delta w$  are the length and width of the mesh element, and  $D_k$  is the depth of the corresponding mesh element.

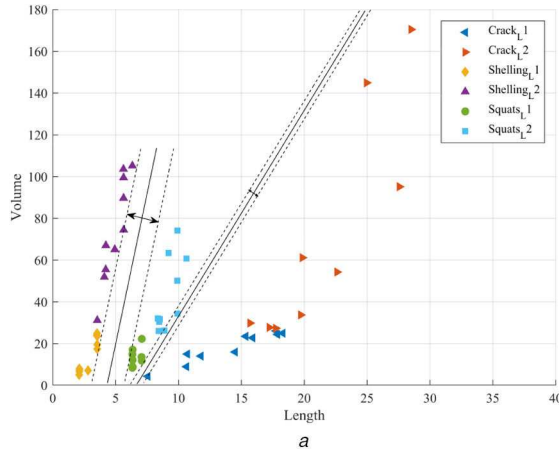
In summary, each 3D model of the defect is described by six quantified geometrical features ( $L$ ,  $W$ ,  $D$ , FN, DM, and  $V$ ), and thus all 60 artificial RCF defects can generate a  $60 \times 6$  matrix, which is used as the full feature set. Compared to existing 2D profile-based methods, from which only  $W$  and  $D$  are available, the proposed 3D model-based method provides a more informative feature set. Table 1 shows a subset of the full feature set which covers the full range of features and classes. Six pre-defined classes include three types of RCF defect and two levels of severity [L1 (minor) and L2 (more significant)] based on expert knowledge. Further processing of the training set before they are fitted into the learning process is introduced in the following section.

### 3.4 Design of the classifier

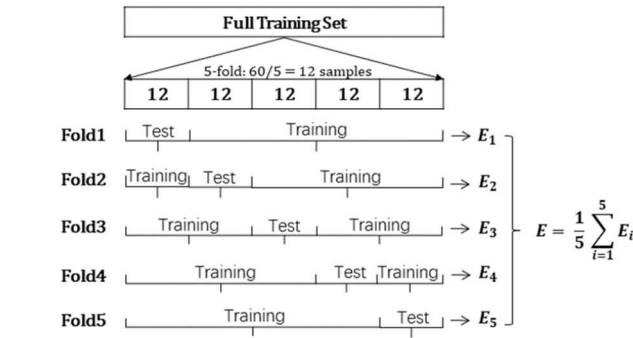
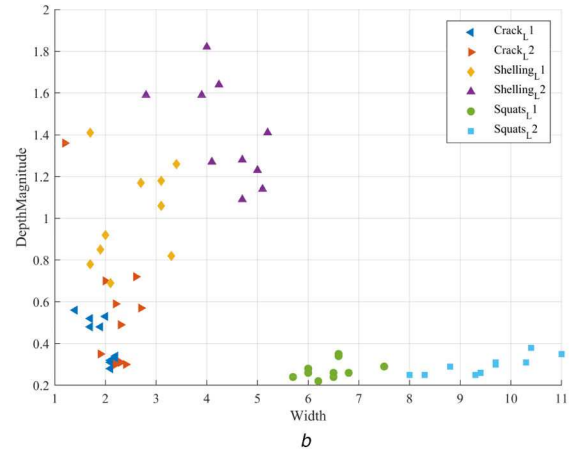
The quantified geometrical features from the 3D model provide a lower dimensional description for each defect. For these, eligible learning models are selected to recognise the patterns associated

with each category of RCF defect automatically. This thereby has the potential to remove the requirement for human involvement in RCF defect classification and evaluation. To select appropriate learning models, one commonly used method is to visualise the features from all classes to analyse their spatial distribution characteristics (linearly separable or not), which is a key index for automatic classification methods. Fig. 8 shows the scatter plot results of two random features from all six classes. The data points with different shapes representing six classes of RCF defect are linearly separable in some dimensions with relatively clear margins between each other as shown in Fig. 8a, but are linearly non-separable in other dimensions as shown in Fig. 8b. Accordingly, SVM-based classifiers would be good candidates because of their relatively low computational complexity and flexible definition of hyperplanes that divide different classes of a sample set while allowing the maximum margins for linearly separable data sets. Additionally, linearly non-separable data sets can be converted to be linearly separable ones by mapping them into a higher dimensional space with the introduction of suitable kernel functions.

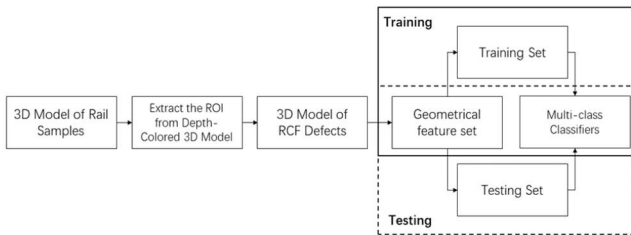
To evaluate the performance of classifiers, the test accuracy of unseen data matters. However, with a relatively small data set, a simple training and test split will give a smaller test set which is not representative of the performance of the classifier for the full data set and can be biased. The use of cross-validation offers the chance of solving this problem. It randomly splits the data sample into  $k$  subsets, each of which will be retained as a test data set, and the remaining  $k - 1$  subsets will be taken as a training data set. This process will be iterated throughout the whole  $k$  folds. In each fold, the testing data set will be hidden from the model during the training phase. This means that each sample has the opportunity to



**Fig. 8** Space distribution characteristics of the full feature set



**Fig. 9** *k*-fold cross-validation



**Fig. 10** 3D model-based RCF defect classification block processing

be used as the test data set once, from which  $k$  evaluation scores ( $E_i$ ) can be obtained and averaged to assess the performance of the classifier, giving a less biased or less optimistic estimate of the model's skill [32]. Fig. 9 shows the process of  $k$ -fold cross-validation used, where  $k$  is set to 5 here to make sure that each group is large enough to be statistically representative of the full data set.

## 4 Experimental results

### 4.1 Laboratory tests

After confirming the candidate classifiers and the model evaluation method, the full feature set, which is a  $60 \times 6$  matrix, is ready to be fitted into the classifier. Multi-class SVMs include a linear SVM and non-linear SVMs with a quadratic kernel, cubic kernel, and a Gaussian kernel, for which the regularisation parameter ( $C$ ) is set to 1. For the quadratic kernel and the cubic kernel, the kernel coefficient ( $\gamma$ ) is set to  $1/F$ , where  $F$  is the number of predictors (features) and is equal to 6 herein. As the Gaussian kernel is typically used for small numbers of feature size (e.g. 1–1000) and a moderate sample set (e.g. from 10–10,000), it is considered first. The Gaussian kernel is tested on three different kernel scales:  $\sigma = \sqrt{F}/4$ ,  $\sigma = \sqrt{F}$ , and  $\sigma = 4 * \sqrt{F}$ . One-vs-one configuration is used to give a balanced data set for each classification. Meanwhile, due to the different scales of geometrical features, data

**Table 2** Comparison of the performance of multi-class classifiers on the 3D-model based feature set

Kernel	Full feature set		
	Cross-validation accuracy, %	Score variance, %	Prediction speed, obs./s
linear SVM	91.7	1.34	~590
quadratic SVM	91.7	0.00	~910
cubic SVM	88.3	0.66	~1200
Gaussian SVM ( $\sigma = \sqrt{F}/4$ )	71.7	1.34	~1100
Gaussian SVM ( $\sigma = \sqrt{F}$ )	95.0	0.00	~1200
Gaussian SVM ( $\sigma = 4 * \sqrt{F}$ )	85.0	0.68	~1200
Euclidean KNN ( $k = 10$ )	90.0	0.68	~900
FNN (6-10-6)	93.3	2.02	~850
result	95.0	0.00	~1200

normalisation is considered in the form of  $x_j = x_j - \mu_j$ , where  $x_j$  is the original value of each feature and  $\mu_j$  is the mean value of the specific type of feature, to improve the fit. For crosswise comparison of the performance of the SVM classifiers, multi-class nearest-neighbour classifiers (KNNs) and a two-layer feed-forward neural network (FNN) with a sigmoid transfer function are tested under the same conditions. Different hidden layer sizes for the neural network (e.g. 6 input features  $\times$  5 hidden neurons  $\times$  6 outputs) are applied. Fig. 10 summarises the basic process of the proposed 3D model-based RCF defect classification system.

Table 2 shows the performance of each type of SVM classifier and the best results of KNNs and FNNs based on three indices including the average classification accuracy, the variance of five-fold scores as an evaluation of the stability, and the prediction speed in the form of observations (samples) per second. It can be seen that the classifier with a Gaussian kernel ( $\sigma = \sqrt{F}$ ) outperforms other classifiers, giving a classification accuracy up to 95% with no variance, and a prediction speed around 1200 obs./s. To further demonstrate the value of the proposed 3D model-based method, a comparison with the existing 2D profile-based methods is also considered by testing the aforementioned classifiers with the 2D feature set ( $W$  and  $D$ ) under the same conditions. The corresponding classification accuracy for the 2D solution is in the range 73.3–78.3%, which is significantly inferior to that of the proposed 3D model-based method and thus is not described in detail here.



## 4.2 Feature evaluation

A 5% error rate is acceptable for explorative research. However, this value could fluctuate with an increase of data volume in future work. To ensure the performance of the learning model to avoid risks such as high computational complexity, the feature set is expected to be non-redundant and able to classify the samples accurately with as few features as possible. However, for a feature set with six features, the distribution of samples in high dimensional space is inaccessible and thus it is difficult to verify the contribution that each feature makes to the classification or clustering of these samples. PCA gives a solution for this issue; by projecting the scaled feature set onto the fitting line, high dimensional feature variables can be converted to a one-dimensional principal component (PC), which is actually a linear combination of all the feature variables and produces a simpler description of the feature set [33]. In statistics, the variance of the original feature set is 100%, and the percentage variance that a PC accounts for is proportional to the sum of the squared projection line for all the samples, known as the eigenvalue in PCA. By selecting the PCs that account for the vast majority of the variance, we can give a good approximation of the high dimensional feature set with a lower dimensional PC set. The scree plot in Fig. 11 shows PCA results and the percentage of variance that each PC accounts for based on its eigenvalue. It can be seen that the first PC accounts for 49.6% of the total variance, the second PC for 26.7%, and the third for 19.0%; the cumulative percentage variance accounted for by the first three PCs is 95.3%.

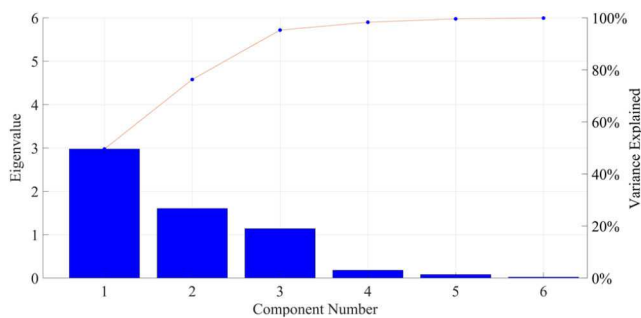


Fig. 11 Scree plot of the PCs and the variance explained

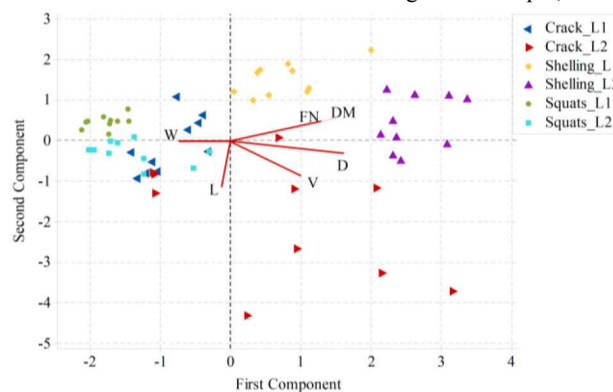


Fig. 12 Bi-plot of the full feature set and score plot of the first two PCs

According to the 'proportion of variance accounted for' criterion introduced by Kim and Mueller [34], further investigation can be made by visualising the distribution samples on the basis that the first two components account for over 70% of the cumulative variance in Fig. 12. The red vectors in Fig. 12, representing the bivariate correlation between the original features and the first two PCs, show that FN and DM have overlapped projection vectors for the first two PCs with approximately the same direction, angle, and vector size, which means that when the samples are projected onto the first two PCs, FN and DM will give approximately the same distribution properties. The results from the PCA suggest that a reduced feature set with FN or DM eliminated can be considered to reduce the computational complexity of the classification if the dimensionality of the problem is to be reduced with an increase of data volume in future work.

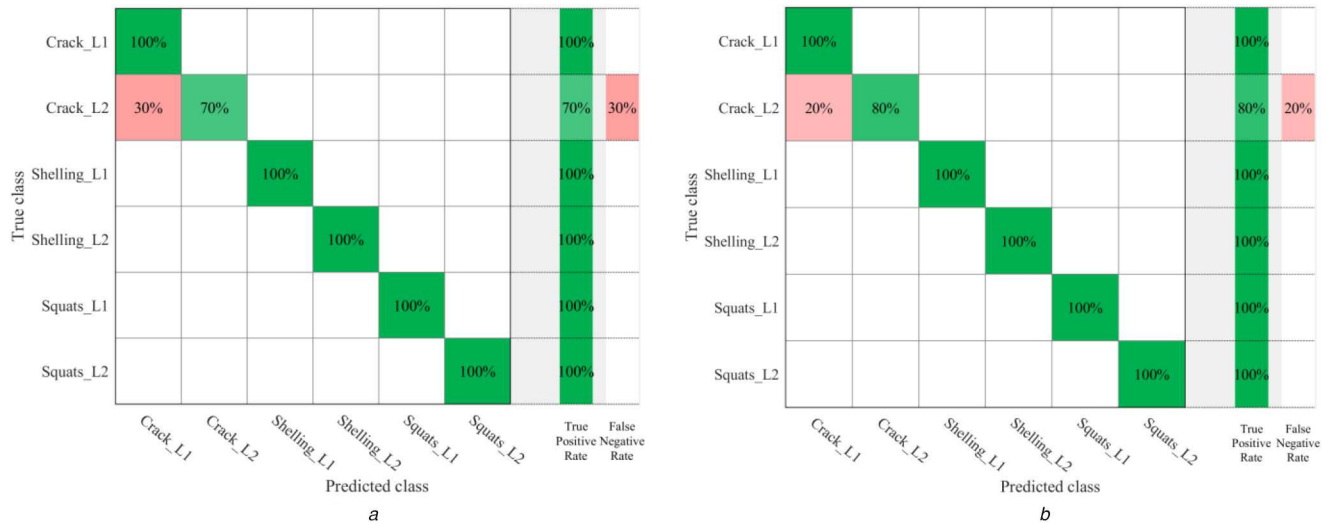
In order to evaluate the effect of the PCA, the eliminated feature sets are tested under the same conditions. Table 3 shows that the best classification accuracy of the Gaussian SVM is enhanced from 95 to 96.7% with no influence on the stability when the FN is eliminated. Moreover, the corresponding speed increases from around 1200 to 1300 obs./s. Conversely, the elimination of DM generally reduces classification accuracy and even reduces prediction speed, suggesting that FN should be the first feature candidate to be eliminated if the dimensionality of the problem is to be reduced or when faced with an increasing amount of data in future work.

## 4.3 Error analysis

To further analyse the performance of the 3D model-based defect classification and find incorrectly classified samples that cause errors, the confusion matrices of the best two classification works in Tables 2 and 3 are plotted in Fig. 13. The exact position of incorrectly classified samples and the positive classification rate for each class are indicated. It can be seen that only 30% (3 out of 10) of samples in the true class of Crack\_L2 are predicted as Crack\_L1 when the full feature set is fitted. Similarly, the classification results for the reduced feature set with FN eliminated have an error rate of only 20% (2 out of 10) for the class of Crack\_L2; all other classes are correctly classified with a 100% positive rate. Meanwhile, it is remarkable that there is no miscategorised sample; all the incorrectly classified samples are

Table 3 Comparison of the performance of multi-class SVMs on the eliminated feature set

Kernel	FN-eliminated feature set			DM-eliminated feature set		
	Cross-validation accuracy, %	Score variance, %	Prediction speed, obs./s	Cross-validation accuracy, %	Score variance, %	Prediction speed, obs./s
linear SVM	91.7	0.66	~1300	91.7	0.66	~1300
quadratic SVM	90.0	0.68	~1100	91.7	0.66	~960
cubic SVM	91.7	0.00	~1300	93.3	0.68	~850
Gaussian SVM ( $\sigma = \sqrt{F}$ )	96.7	0.00	~1300	90.0	0.68	~940
result	96.7	0.00	~1300	93.3	0.68	~850



**Fig. 13** Confusion matrices for Gaussian SVM classification results

(a) Full feature set-based classification result, (b) Eliminated feature set-based classification result

just predicted to be at the wrong severity level. Further investigation of the incorrectly classified samples shows that the corresponding cracks are mainly distributed at the edge of the rail. The length of these samples can be measured incorrectly due to the shadowing effects of the laser sensor, but the width and depth have relatively small measurement errors [14]. This can result in a sample measured with a length in the range of Crack\_L1, but with other features such as volume in the range of Crack\_L2, which could potentially influence the classification result. This can be solved in further works either by improving the measuring unit or eliminating the influence of measurement errors in post-processing such as by adding confidence indices to the samples based on their geometry. Crosswise comparisons show that FNN gives an average accuracy of 93.3%. This could result from the existence of a local minimum when finding the optimal parameters for the neural network. Hence it can suffer from a relatively high bias especially when faced with a small data volume. This could be improved with increased data volumes in future work.

## 5 Conclusion

The development of sensory technologies and the related processors advances the application of 3D measurement in many domains. A 3D model can represent the characteristics of objects more comprehensively compared to conventional 2D methods, and provide more informative features for further model-based processing such as defect inspection. In combination with artificial intelligence technologies, this approach has the potential to replace existing human-dominated inspection work. This paper takes rail RCF defect inspection as a case, exploring a 3D model-based automatic RCF defect classification method. Based on a high-quality 3D model of the rail surface, six geometrical features of RCF defects are extracted. These features describe the patterns of three common types of RCF defect and two severity levels. Multi-class classifiers are tested, giving an accuracy of up to 96.7%, which indicates the feasibility of automatic surface defect classification and evaluation using a laser-based 3D model. Finally, PCA is performed to further analyse the contribution of each geometrical feature to the classification. This can be considered if the dimensionality of the problem is to be reduced or when faced with an increasing amount of data in future cases. This exploratory work enhances the application of 3D model-based inspection technologies in railway systems, and also provides a reference for its application in other domains.

## 6 Acknowledgments

This work was supported in part by the Shift2Rail Joint Undertaking under the European Union's Horizon 2020 research and innovation programme as part of the S-CODE project under

Grant Agreement No. 730849, in part by the Guangzhou Science and Technology Plan (Ref. 201704030048), and in part by the China Scholarship Council.

## 7 References

- [1] Network Rail: 'Inspection and maintenance of permanent way' (NR, UK., 2018)
- [2] Cannon, D.F., Edel, K.O., Grassie, S.L., et al.: 'Rail defects: an overview', *Fatigue Fract. Eng. M.*, 2003, **26**, (10), pp. 865–887
- [3] Department of Transportation: 'Rolling contact fatigue: a comprehensive review' (NRC, USA., 2011), pp. 9–19
- [4] Alahakoon, S., Sun, Y.Q., Spiryagin, M., et al.: 'Rail flaw detection technologies for safer, reliable transportation: a review', *J. Dyn. Sys., Meas., Control*, 2017, **140**, (2), p. 020801
- [5] 'Track Measuring Systems: Full Rail Profile', <http://www.mermecgroup.com>, accessed May 2019
- [6] Rusu, M.F.: 'Automation of railway switch and crossing inspection'. PhD thesis, University of Birmingham, 2017
- [7] Xiong, Z., Li, Q., Mao, Q., et al.: 'A 3D laser profiling system for rail surface defect detection', *Sens. (Basel)*, 2017, **17**, (8), p. E1791
- [8] Zhou, P., Xu, K., Wang, D.: 'Rail profile measurement based on line-structured light vision', *IEEE Access*, 2018, **6**, pp. 16423–16431
- [9] Chondronasios, A., Popov, I., Jordanov, I.: 'Feature selection for surface defect classification of extruded aluminum profiles', *Int. J. Adv. Manuf. Technol.*, 2015, **83**, (1–4), pp. 33–41
- [10] Hani, A.F.M., Eltegeani, N.M., Arshad, L., et al.: 'Wound model reconstruction from three-dimensional skin surface imaging using the convex hull approximation method', *IET Image Process.*, 2012, **6**, (5), pp. 521–533
- [11] Zou, C., He, B., Zhang, L., et al.: 'Scene flow for 3D laser scanner and camera system', *IET Image Process.*, 2018, **12**, (4), pp. 612–618
- [12] Yang, Y., Li, B., Li, P., et al.: 'A two-stage clustering based 3D visual saliency model for dynamic scenarios', *IEEE Trans. Multimedia*, 2019, **21**, (4), pp. 809–820
- [13] Wang, Y., Yang, Y., Liu, Q.: 'Feature-aware trilateral filter with energy minimization for 3D mesh denoising', *IEEE Access*, 2020, **8**, pp. 52232–52244
- [14] Ye, J., Stewart, E., Roberts, C.: 'Use of a 3D model to improve the performance of laser-based railway track inspection', *Proc. IMechE, F: J. Rail Rapid Transit*, 2018, **233**, (3), pp. 337–355
- [15] Bishop, C.M.: 'Pattern recognition and machine learning' (Information Science and Statistics) (Springer-Verlag, USA., 2006)
- [16] Şevik, U., Karakullukcu, E., Berber, T., et al.: 'Automatic classification of skin burn colour images using texture-based feature extraction', *IET Image Process.*, 2019, **13**, (11), pp. 2018–2028
- [17] Xue-Wu, Z., Yan-Qiong, D., Yan-Yun, L., et al.: 'A vision inspection system for the surface defects of strongly reflected metal based on multi-class SVM', *Expert Syst. Appl.*, 2011, **38**, (5), pp. 5930–5939
- [18] Luo, Q., Fang, X., Sun, Y., et al.: 'Surface defect classification for hot-rolled steel strips by selectively dominant local binary patterns', *IEEE Access*, 2019, **7**, pp. 23488–23499
- [19] Kumar, A.: 'Computer-vision-based fabric defect detection: a survey', *IEEE Trans. Ind. Electron.*, 2008, **55**, (1), pp. 348–363
- [20] Trinh, H., Haas, N., Li, Y., et al.: 'Enhanced rail component detection and consolidation for rail track inspection'. 2012 IEEE Workshop on the Applications of Computer Vision (WACV), Breckenridge, CO, USA, January 2012, pp. 289–295
- [21] Townes, C.H.: 'How the laser happened: adventures of a scientist' (Oxford University Press, UK., 1999)

- [22] Forest, J., Salvi, J.: 'A review of laser scanning three-dimensional digitisers'. IEEE/RSJ Int. Conf. Intelligent Robots Systems, Lausanne, Switzerland, September–October 2002, pp. 73–78
- [23] Wu, B., Xue, T., Zhang, T., *et al.*: 'A novel method for round steel measurement with a multi-line structured light vision sensor', *Meas. Sci. Technol.*, 2010, **21**, (2), p. 025204
- [24] Zhang, G., He, J., Li, X.: '3D vision inspection for internal surface based on circle structured light', *Sensor. Actuat. A: Phys.*, 2005, **122**, (1), pp. 68–75
- [25] Papaefias, M.P., Lugg, M.C., Roberts, C., *et al.*: 'High-speed inspection of rails using ACFM techniques', *NDT E Int.*, 2009, **42**, (4), pp. 328–335
- [26] Micro-Epsilon: 'scanCONTROL // 2D/3D laser scanner (laser profile sensors)' (Micro-Epsilon, Germany, 2015)
- [27] Otsu, N.: 'A threshold selection method from gray-level histograms', *IEEE Trans. Syst. Man Cybern.*, 1979, **9**, (1), pp. 62–66
- [28] Chen, K., Kvasnicka, V., Kanen, P.C., *et al.*: 'Supervised and unsupervised pattern recognition: feature extraction and computational intelligence [book review]', *IEEE Trans. Neural Netw.*, 2001, **12**, (3), pp. 644–647
- [29] Xie, X.: 'A review of recent advances in surface defect detection using texture analysis techniques', *ELCVIA*, 2008, **7**, (3), pp. 1–22
- [30] Davies, E.R.: 'Computer and machine vision: theory, algorithms, practicalities' (Academic Press, Inc., UK., 2012, 1st edn.), p. 912
- [31] Keshet, L.: 'Math 103' (University of British Columbia, Canada, 2012)
- [32] Berrar, D.: 'Cross-validation', in Ranganathan, S., *et al.* (Eds.): 'Encyclopedia of bioinformatics and computational biology' (Academic Press, Netherlands, 2019), pp. 542–545
- [33] Hatcher, L., Stepanski, E.J.: 'A step-by-step approach to using the SAS system for univariate and multivariate statistics' (SAS Institute, USA., 1994), p. xiv, p. 552
- [34] Kim, J., Mueller, C.W.: 'Introduction to factor analysis: what it is and how to do it' (Quantitative Applications in the Social Sciences) (Sage Publications, 1978), vol. **13**