

# Real-Time Techniques for Fault Detection on Railway Door Systems

Minoru Shimizu  
Integrated Vehicle Management  
Centre, Cranfield University  
College Road, Cranfield, Bedford,  
Bedfordshire, UK, MK43 0AL  
minoru.shimizu@cranfield.ac.uk

Suresh Perinpanayagam  
Integrated Vehicle Management  
Centre, Cranfield University  
College Road, Cranfield, Bedford,  
Bedfordshire, UK, MK43 0AL  
suresh.nayagam@cranfield.ac.uk

Bernadin Namono  
Digital Engineering and Manufacturing  
Centre, Cranfield University  
College Road, Cranfield, Bedford,  
Bedfordshire, UK, MK43 0AL  
bernadin.namono@cranfield.ac.uk

**Abstract**—This paper focuses on real-time techniques for fault detection in railway assets through large real-world datasets. It aims to investigate data mining methods to detect faulty behaviour in time series data. A fault detection on railway door systems is carried out using motor current and encoder signal. The door data highlighted start-stop characteristics, with discontinuities in the data. This paper presents a successful fault detection technique, which is a feature-based machine learning method that requires several steps for time-series data processing, such as signal segmentation and the extraction of features. Principal Component Analysis (PCA) is applied to reduce the dimensionality of the extracted feature set and generate condition indicators. Then, the k-means algorithm is employed to separate normal and abnormal behaviour. This is followed by an evaluation of the proposed method and discussion about current challenges and prognosis possibility.

## TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. HEALTH MONITORING SYSTEM.....	2
3. DATA PRE-PROCESSING.....	2
4. FEATURE EXTRACTION.....	3
5. PRINCIPAL COMPONENT ANALYSIS.....	4
6. K-MEANS CLUSTERING.....	4
7. RESULTS AND DISCUSSION.....	5
8. HOW TO AUTOMATE FAULT DETECTION.....	6
9. PROGNOSIS.....	7
10. SUMMARY.....	7
ACKNOWLEDGEMENTS.....	7
REFERENCES.....	8
BIOGRAPHY.....	9

## 1. INTRODUCTION

Maintenance tasks are fundamental to meet the growth of the number of customers and economic demand in the railway industry. With the improvement of infrastructure and connectivity within and between cities, the railway sector is witnessing an increase in its customer numbers over the years. From an ecological point of view, globally, the railway industry spends a large amount of money on maintenance activities. For example, the annual maintenance expenditure of British railway infrastructure was more than £1 billion in

2015; almost two-thirds of Network Rail's employees are engaged in maintenance work [1].

The train door is one of the most critical subsystems that can cause service delay or breakdown, leading to the increased cost of operation and maintenance. As reported in [2], [3], door systems are responsible for 30% to 60% of the total failures in railway vehicles. In order to prevent these failures, predictive maintenance is attracting more and more attention recently. Predictive maintenance is a predictive framework to estimate the time when a fault is likely to occur and to adopt maintenance interventions accordingly [1]. To achieve successful predictive maintenance of door systems, a recent study has been conducted on fault detection and diagnosis by utilising the sensor signals gathered from machinery [4].

In this area of study, the methodologies usually centre on model-based or data-driven approaches. Model-based approaches incorporate a physical understanding of the door systems through mathematical representations and include door system modelling. The output of the model is then compared with the actual output measurement throughout the residual analysis [5]. Several works in model-driven methods encompass differential equation-based modelling [6]. However, the train door system contains many components interconnected with various uncertainties, which makes the modelling approach of limited value in predictive maintenance. On the other hand, data-driven approaches use statistical pattern recognition and machine learning to detect changes[5]. Data-driven approaches do not require mathematical modelling of the door systems and have gained much attention with the increasing availability of data. The previously proposed methods include self-organising maps (SOM) [7], logistic regression [8], dynamic time warping [9], and convolutional neural networks [10], support vector machine (SVM) by using audio sensors [11], and fault classification and evaluation for comparison purposes between the traditional machine learning method and deep learning approaches [12].

The present work also explores data-driven approaches but differs from proposed methods in the literature because prior knowledge and labelled data of the fault behaviour is not available. In practical application with complex engineering systems, expert knowledge and fault-labelled datasets sometimes cannot be obtained in advance, which makes fault detection techniques difficult. To address the issue, the

unsupervised clustering method, which does not require labelled data, is applied in this paper. In addition, this paper focuses on the data-driven method by using a real-world large dataset of the railway. In general, it has been reported that examples of successful fault detection and prognostic applications in complex engineering systems are still scarce [5].

This paper aims to report an example of successful fault detection techniques in practical application with railway door systems. In Section 2, the fault detection workflow by the data-driven approach is proposed. Sections 3 and 4 describe the data processing method, including pre-processing, segmentation, and feature extraction. Sections 5 and 6 explain condition indicator extraction and the unsupervised clustering method applied in this research. The results are provided in Section 7 to validate the methodology, followed by Sections 8 and 9, including the way to automate fault detection and the concept of prognosis. A summary is given in Section 10 with a discussion of the future development of health monitoring systems of railway assets.

## 2. HEALTH MONITORING SYSTEM

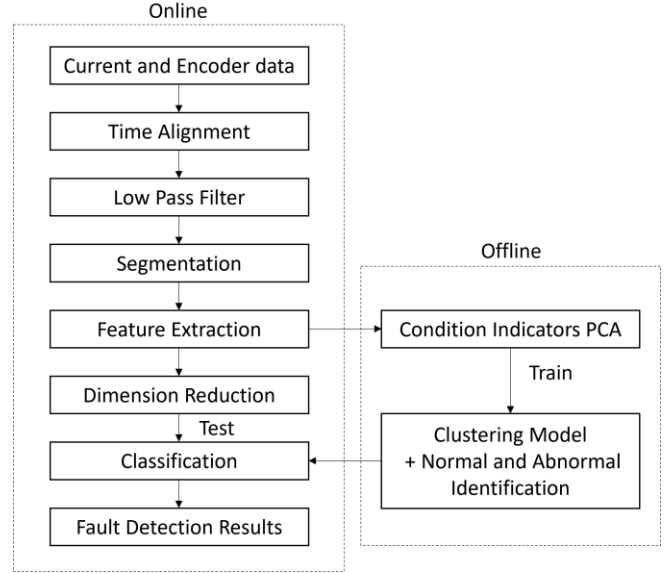
The proposed fault detection workflow for train doors is shown in Figure 1. The workflow is divided into two categories, online and offline. In the online procedure, the data-driven approach is proposed by using motor current and encoder signals that have been collected from railways. In this approach, time-series data is pre-processed to be aligned, eliminating noise by a low pass filter. Then, pre-processed data is segregated into the opening and closing operations, followed by additional segmentation into three different movement phases: acceleration, steady-state, and deceleration, which is described in Section 4. Then, various time-domain features, which characterise the dynamic behaviour of the door system, are extracted accordingly.

In the offline procedure, railway data is used as training datasets to train an unsupervised clustering model to separate normal and abnormal behaviours. First, Principal Component Analysis (PCA) is used to reduce the dimensionality of the extracted feature set and generate condition indicators. Then, the K-means algorithm is trained and employed to cluster condition indicators, followed by identification of fault and normal clusters, which is described in Section 7.

The clustering model created offline is implemented on the online procedure to detect faulty behaviour. The online fault detection workflow can be executed once one operation with opening and closing is terminated so that abnormal modes can be detected as early as possible to allow maintenance activities to be predictive.

One of the advantages of the proposed workflow is that the offline clustering model can be improved by additional operational railway data that enable fault detection to be more accurate. Furthermore, it is also possible to create more clusters of condition indicators by increased data, which means unknown fault modes can be identified by gathering

operational data, which is also beneficial for health monitoring and predictive maintenance systems.



**Figure 1: The proposed fault detection workflow for train doors**

## 3. DATA PRE-PROCESSING

In this study, an electric door is considered, which is composed of a voltage power source, a DC motor, a door control unit (DCU), a transmission and door leaves. An example of a door system is shown in Figure 2. In short, a DC motor, powered by a voltage source and controlled by DCU, can output the specified shaft angular velocity and torque, which are transmitted to transmission so that the door leaves can move in a pre-designed manner. The door data, which consists of current and encoder signals, is collected through the communication port from the DCU at a frequency of 20 Hz. The time lag is often observed between the motion profile and the current. To align the time series, the dynamic time warping (DTW) method is used for the first alignment. The DTW is one of the widely used algorithms for measuring the similarity between two temporal sequences that may vary in time [4]. The low pass filter is applied on a window of 0.25 seconds, representing five consecutive measurement time intervals to reduce noise carried by both current and encoder signals.

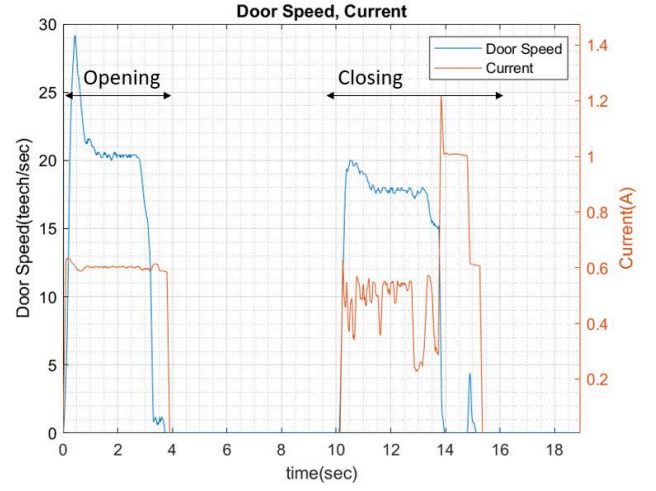


**Figure 2: Example of train doors**

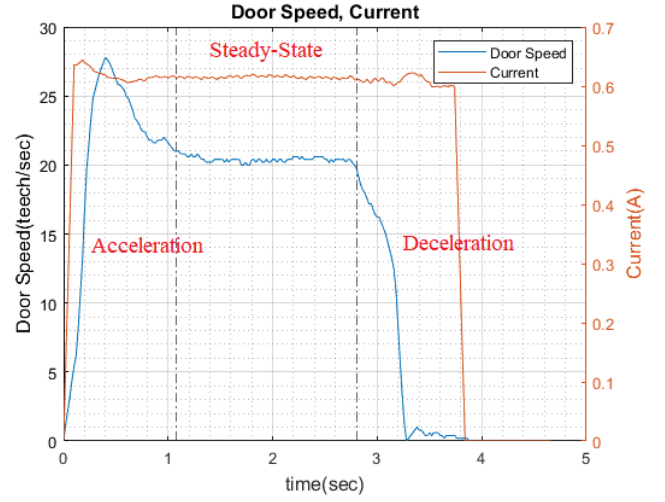
#### 4. FEATURE EXTRACTION

An example of the signal profile of the opening and closing operations is shown in Figure 3. In the opening profile, the speed and current increase steadily up to a maximum, followed by a slight curve, and then decrease to zero. The closing profile follows a similar pattern with two main differences in the current. One is that the peak in the closing profile is lower than the opening. The second is an abrupt change at the end of the closing profile, followed by a slight bump of the speed, which promotes pushing the door to its maximal reachable position where a locking process can be triggered [13].

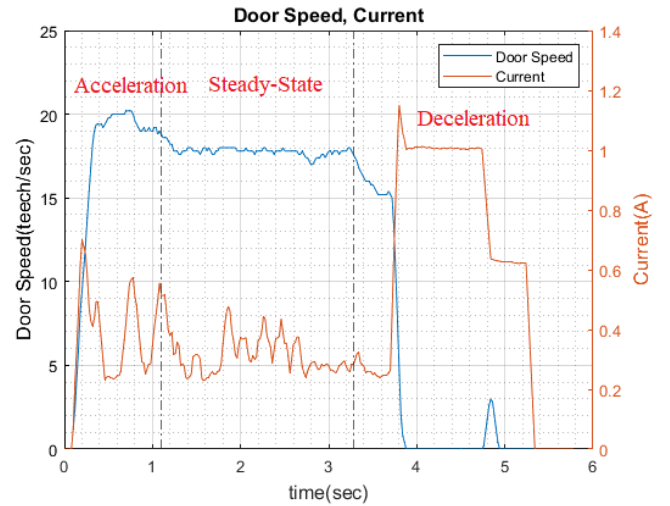
The opening and closing operations have different movements in terms of velocity regime: acceleration, steady-state, and deceleration, each of which can be identified by using the encoder signals as shown in Figure 4 and Figure 5. It is more appropriate to segregate them into each segment to extract features accurately. The bottom-up algorithm is applied to segregate the opening and closing operation automatically. The bottom-up algorithm is one of the most well-known algorithms for segmenting time series data. This algorithm begins by creating the finest possible approximation of the time series so that  $n/2$  segments are used to approximate the  $n$ -length time series. Next, the cost of merging each pair of adjacent segments is calculated, and the algorithm begins to iteratively merge the lowest-cost pair until a stopping criterion is met[14]. In this research, the minimum mean squared error of linear approximation is used for the cost, and the bottom-up algorithm keeps merging the lowest-cost pair until each signal is segregated into three segments: acceleration, steady-state, and deceleration. Indeed, the current and speed signals might be able to be segregated by a user-defined algorithm. However, the signals can be separated more appropriately by using a segmentation algorithm. Then, time-domain features, which are likely sensitive to degradation, are extracted from the current and encoder signals. The features for the opening and closing operations are given in Table 1 and Table 2, respectively [10], [12], [15].



**Figure 3: Door speed and current signals of the opening and closing operation.**



**Figure 4: Segmentation of the opening operation.**



**Figure 5: Segmentation of the closing operation.**

**Table 1: Extracted features for the opening operation**

Signal	Segments	Feature	Feature ID
Speed	Whole	Settling Time	1
		Skewness	2
		Kurtosis	3
		Standard Deviation	4
		Crest Factor	5
	Acceleration	Overshoot	6
		Settling Time	7
		Settling Minimum	8
	Steady State	Mean	9
		Standard Deviation	10
		Minimum	11
		Maximum	12
Current	Whole	Settling Time	13
		Skewness	14
		Kurtosis	15
		Standard Deviation	16
		Crest Factor	17
	Acceleration	Overshoot	18
		Settling Time	19
	Steady State	Mean	20
		Standard Deviation	21
		Minimum	22
		Maximum	23

**Table 2: Extracted features for the closing operation**

Signal	Segments	Feature	Feature ID
Speed	Whole	Settling Time	1
		Skewness	2
		Kurtosis	3
		Standard Deviation	4
		Crest factor	5
	Acceleration	Overshoot	6
		Settling Time	7
		Settling Minimum	8
	Steady State	Mean	9
		Standard Deviation	10
		Minimum	11
		Maximum	12
Current	Deceleration	Small Peak	13
		Small Peak count	14
	Whole	Settling Time	15
		Skewness	16
		Kurtosis	17
		Standard Deviation	18
		Crest factor	19
	Acceleration	Overshoot	20
		Settling Time	21
	Steady State	Mean	22
		Standard Deviation	23
		Minimum	24
		Maximum	25

Deceleration	Maximum Peak	26
	Large Peak Count	27

## 5. PRINCIPAL COMPONENT ANALYSIS

Condition indicators are generated by using PCA as they allow the projection of extracted features into a lower-dimensional space. This method has been used in many previous research works to reduce dimensionality[10] and select the most relevant features [16]. PCA can be defined as the orthogonal projection of data onto a lower dimensional linear space, known as the principal subspace, such that the variance of the projected data is maximised [17]. Given an  $n \times m$  data matrix  $X$  containing  $n$  observations and  $m$  variables, it is possible to obtain a set of loading vectors  $V$  by solving the eigenvalue decomposition of the covariance matrix  $S$ :

$$S = \frac{1}{n-1} X^T X = V D V^T \quad (1)$$

where the loading vectors  $V$  are ordered by the amount of variance expressed by the corresponding eigenvalue in the diagonal matrix  $D$ . The loading vectors attached to the *largest* singular values are retained in the loading matrix  $R \in \mathbb{R}^{m \times a}$ . This set of vectors generates a lower-dimensional representation of the extracted features. The projection of the data into this reduced space captures systematic trends of the data, reducing the amount of random noise, minimising the negative effects of measurement inaccuracies. The score matrix  $W$  contains the projection of the observed data into the lower-dimensional space, while the residual matrix  $E$  represents the difference between the observations and the projection of  $W$  back into the  $m$ -dimensional space:

$$W = X R \quad (2)$$

$$E = X - W R^T \quad (3)$$

In the research, PCA is applied on features of the opening and closing. Then, principal components retaining over 80 per cent of the proportion of the variance are selected as condition indicators in each dataset.

## 6. K-MEANS CLUSTERING

K-means clustering is an unsupervised learning and data partitioning algorithm that assigns  $n$  training feature vectors to exactly one of  $k$  clusters. This method has been used in many previous research works since industrial data usually contains both normal and abnormal data in high-dimensional space, making it difficult to segregate manually [12], [13]. The steps of K-means clustering are the following [18]:

1. Choose  $k$  centroid (initial cluster centre) and use the k-means ++ algorithm for cluster centre initialisation [19].
2. Compute distances between cluster centres and training feature vectors.

3. Assign each training feature vector to the cluster with the closest centre (this step is called Batch update).
4. Compute the average of the training feature vectors in each cluster to obtain  $k$  new cluster centres.
5. Repeat steps 2, 3 and 4 until the centres do not change their values.

In this research, the K-means clustering algorithm is used to cluster the opening and closing operations represented by condition indicators extracted by using PCA, as mentioned in Section 5.

## 7. RESULTS AND DISCUSSION

The PCA and K-means clustering results are given in Table 3. The five datasets gathered from railway assets, which are 1\_A, 1\_B, 1\_C, 2\_A, and 2\_B, have been analysed. As stated in Section 5, the number of principal components was chosen so that at least over 80 per cent of the variance is retained. The result of the closing operation is not included in Table 3 since distinguished clusters of the closing operation does not appear, which means the closing operation does not have faulty behaviour.

The silhouette score is used to evaluate the clustering algorithm by incorporating measures of compactness and separation between the proposed clusters. The silhouette score is a well-known clustering evaluation approach that introduces clustering quality scores for each individual point and calculates the final quality index as an average of the point-wise quality estimate [20]. Each point-wise estimate for a point  $x_q \in C_i$ ,  $i \in \{1 \dots k\}$ ,  $k$  which denotes the number of clusters is derived from two quantities.  $a_{i,p}$  and  $b_{i,p}$  which correspond to the average distance to other points within the same cluster and the minimal average distance to points from a different cluster, respectively.  $N$  denotes the number of observations:

$$a_{i,p} = \frac{1}{|C_i| - 1} \sum_{x_q \in C_i, q \neq p} ||x_q - x_p|| \quad (4)$$

$$b_{i,p} = \min_{j \in \{1 \dots K\}, i \neq j} \frac{1}{|C_j|} \sum_{x_q \in C_j} ||x_q - x_p|| \quad (5)$$

$$SIL(x_p) = \frac{a_{i,p} - b_{i,p}}{\max\{a_{i,p}, b_{i,p}\}} \quad (6)$$

$$SIL_i = \frac{1}{N} \sum_{p=1}^N SIL(x_p) \quad (7)$$

$$Average \ Silhouette \ Score = \frac{1}{k} \sum_{i=1}^k SIL_i \quad (8)$$

As shown in Table 3, the silhouette scores in each dataset are over 0.70, which means C1 and C2 clusters are well separated by the k-means clustering algorithm. The projection example of the original features of 1\_C into a lower-dimensional space is shown in Figure 6. The results show two distinctive clusters separated by the k-means clustering algorithm. The first conclusion that can be made is that as most of the data represent normal conditions, cluster C1, which contains the majority of the observations, represents the normal operation. In contrast, the remaining cluster, which is C2, is the fault occurring in the door systems.

What is essential in this research is to detect faulty data and separate normal and abnormal behaviour by a clustering algorithm. The examples of speed and current signals of each cluster are shown in Figure 7 and Figure 8. The encoder signal of C1 has no negative peak as normal behaviour. In contrast, that of C2 has a prominent negative peak during step response as a faulty behaviour, which means faulty data is detected and normal and abnormal data are well separated. Thus, fault and normal clusters can be identified by using an unsupervised clustering method without labelled datasets. As a result, the clustering model created and identified as stated above can be implemented on the online procedure to detect faulty behaviour in real-time as described in Sections 2.

**Table 3: PCA and K-means results**

Dataset	Segment	PCA		K-means		
		Principal Components	Proportion of variance (%)	Average Silhouette Score	Number of observations of C1	Number of observations of C2
1_A	Opening	5	81	0.865	3280	14
1_B	Opening	5	83	0.748	3507	76
1_C	Opening	5	81	0.825	4135	51
2_A	Opening	9	95	0.701	10107	151
2_B	Opening	6	85	0.719	3348	47

Certainly, the proposed health monitoring system entails some limitations if an unknown fault should be detected in real time. As stated in Section 2, it is necessary to identify abnormal and normal clusters in the offline procedure, which means only known anomalies can be included in clusters. However, the offline clustering model can be improved by additional operational data that enables fault detection to be more accurate once unrevealed abnormal data can be obtained.

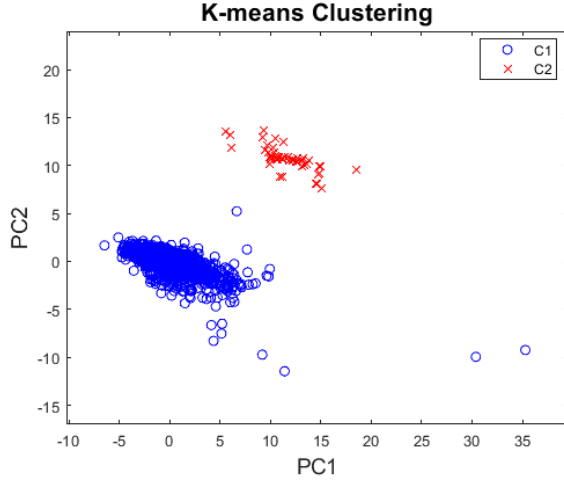


Figure 6: The example of the projection.

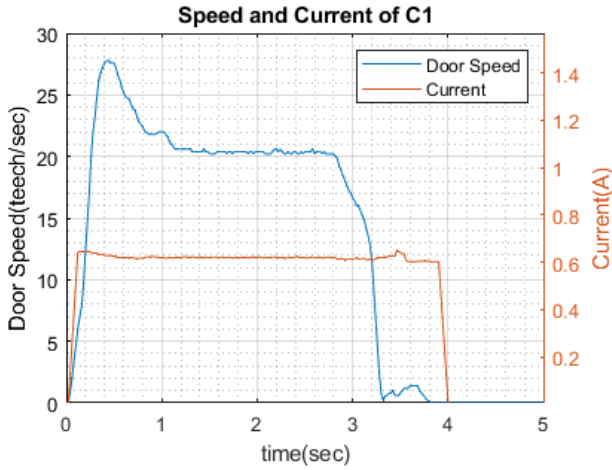


Figure 7: Speed and current signals of C1.

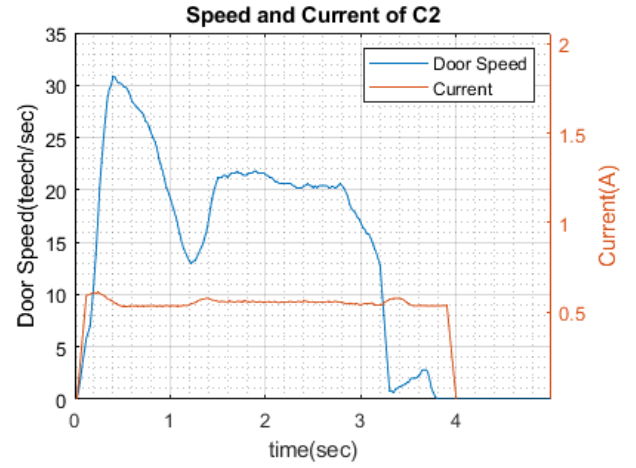


Figure 8: Speed and current signals of C1.

## 8. HOW TO AUTOMATE FAULT DETECTION

It is significant to implement an automatic fault detection procedure for practical predictive maintenance activities due to the massive amount of data acquired from railway assets in real time. The proposed automatic monitoring and fault detection system, which consists of several cloud services, is shown in Figure 9. The cloud service is being developed rapidly so that the proposed architecture is one of the examples utilising the cloud services of Microsoft. The essential point is the way how to automate the entire system from data acquisition to failure detection. First, railway sensor data is transferred to IoT Hub as streaming data. The IoT Hub is a managed service hosted in the cloud that acts as a central message hub for communications between an IoT application and its device. Then, streaming raw data is pre-processed to extract features, followed by condition indicator extraction and clustering by Spark on DataBricks. Spark is an Apache open-source software tool, distributed processing system used for big data workloads initially developed at the University of California. DataBricks is a data analytics platform optimised for the Microsoft Azure cloud services platform. Azure DataBricks offers environments for developing data-intensive applications. In this architecture, an unsupervised machine learning model trained by railway data, as shown in previous sections, is utilised in Spark to detect faulty behaviour as early as possible. Finally, the extracted features, condition indicators and fault detection results are stored in storage for a long-term purpose to utilise the data to train the machine learning model for greater accuracy.

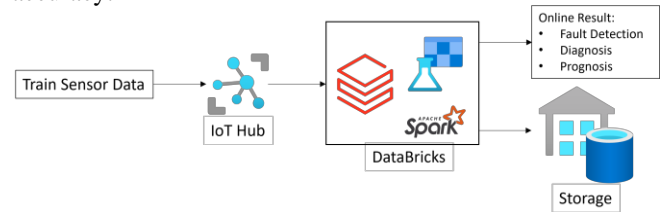


Figure 9: Proposed automation architecture.



## 9. PROGNOSIS

In this section, the concept of prognosis of railway assets is discussed. Prognosis technology aims to accurately predict and estimate the component or system remaining useful life (RUL) to enhance reliability and performance [21]. Prognosis refers to monitoring the equipment health status in real time to utilise the future state of the equipment and estimate the possible time of failure modes by computing a proper prognostic technique. During the early stages of the health monitoring technology, the traditional applicable technologies were concentrated on detecting and isolating failures. As the demand for Condition-Based Maintenance (CBM), the idea of RUL has shown up as prognostic failure prediction techniques.

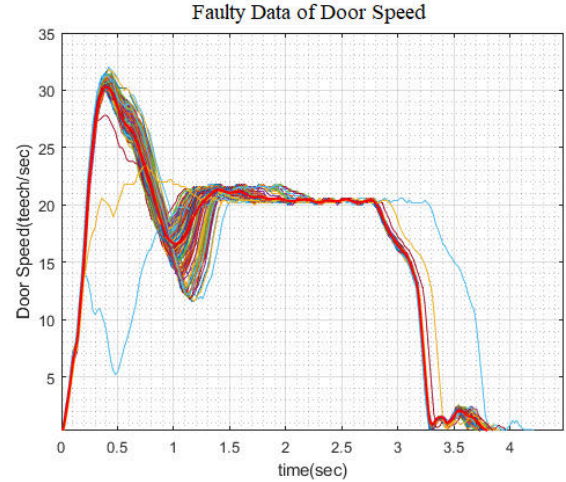
Current prognostic approaches can be categorised into three classes, namely model-based, data-driven, and hybrid prognostic approaches. A typical model-based prognostic strategy consists of dynamic models to perform the prediction function of the system future state. Model-based approaches provide technically comprehensive solutions that have been used widely to understand the failure progression [21]. The physics-based model can be used to determine the system life usage by calculating the current physics parameters for the system. Once the current physics parameters have been identified, the model can predict the future conditions based on historical conditions using stochastic techniques.

In certain situations of an engineered complex system, it might be challenging to design a model-based technique involving all the physical aspects. In such cases, a particular form of the dynamic model can be assumed, then based on the system's actual input and output, and the desired parameters can be determined to obtain accurate results [22]. For these case scenarios, it is impossible to have a prediction model without covering the physics calculations. However, the data-driven model can only be implemented if the historical failures data are available by using a nonlinear network approximator, such as Fuzzy logic, neural network, and other computational intelligence techniques to obtain the desired outputs.

A hybrid prognostic approach, which fuses the outputs from the model-based approach and data-driven approach, was proposed in [23], in which prognostics results are claimed to be more reliable and accurate [24].

The implementation of prognosis by using the railway dataset is beyond the scope of this paper. However, the concept of prognosis of railway assets by the data-driven approach is proposed. The faulty data extracted by using the proposed method mentioned in the previous sections is shown in Figure 10. As shown in Figure 10, the door speed signals have prominent negative peaks during step response as a faulty behaviour. In addition, negative peaks have different peak depths on each profile that looks fault growing. Suppose the peak depth is one of the key features to represent degradation. In that case, it can be applied to monitor the development of fault as a health indicator, followed by a data-driven approach

for prognosis, which will be future work for health monitoring systems for railway assets.



**Figure 10: Faulty data of door speed.**

## 10. SUMMARY

In this paper, the successful fault detection techniques in complex engineering systems, especially railway door systems, are proposed by using the large real-world operational dataset gathered from railway assets.

First, the time-domain features, which are likely sensitive to degradation, are extracted from the current and encoder signals. Principal Component Analysis (PCA) is applied to reduce the dimensionality of the extracted feature set and generate condition indicators. Then, the k-means algorithm is employed to separate normal and abnormal modes. Finally, the silhouette score is used to evaluate the unsupervised classification algorithm. As a result, clusters are well separated by the k-means clustering algorithm. Thus, fault and normal clusters can be identified by using an unsupervised clustering method without labelled datasets, which enable an unsupervised clustering method to be implemented on practical health monitoring systems. The way to automate fault detection and the concept of prognosis which is proposed will be future work for health monitoring systems for railway assets.

## ACKNOWLEDGEMENTS

The authors wish to thank Unipart Rail, Instrumental and the Engineering and Physical Sciences Research Council (EPSRC) for their support.

## REFERENCES

- [1] J. Xie, J. Huang, C. Zeng, S.-H. Jiang, and N. Podlich, "Systematic Literature Review on Data-Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering," *Geosciences* 2020, vol. 10, no. 11, p. 425, Oct. 2020, doi: 10.3390/GEOSCIENCES10110425.
- [2] L. Cauffriez, P. Loslever, N. Caouder, F. Turgis, and R. Copin, "Robustness study and reliability growth based on exploratory design of experiments and statistical analysis: a case study using a train door test bench," *The International Journal of Advanced Manufacturing Technology* 2012 vol. 66, no. 1, pp. 27–44, Jul. 2012, doi: 10.1007/S00170-012-4303-0.
- [3] Y. Qin, S. Yu, J. Shi, and L. Mao, "Reliability analysis based on the statistical models method for the subway door system," *IEEE ICIRT 2013 - Proceedings: IEEE International Conference on Intelligent Rail Transportation*, pp. 215–220, 2013, doi: 10.1109/ICIRT.2013.6696296.
- [4] C. S. Myers and L. R. Rabiner, "A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected-Word Recognition," *Bell System Technical Journal*, vol. 60, no. 7, pp. 1389–1409, Sep. 1981, doi: 10.1002/j.1538-7305.1981.tb00272.x.
- [5] O. F. Eker, F. Camci, and I. K. Jennions, "A New Hybrid Prognostic Methodology," *International Journal of Prognostics and Health Management*, 2019, Accessed: Aug. 27, 2021. [Online]. Available: <http://www.phmsociety.org/node/2559>.
- [6] C. Roberts, H. P. B. Dassanayake, N. Lehasab, and C. J. Goodman, "Distributed quantitative and qualitative fault diagnosis: railway junction case study," *Control Engineering Practice*, vol. 10, no. 4, pp. 419–429, Apr. 2002, doi: 10.1016/S0967-0661(01)00159-9.
- [7] J. Lee, H. Choi, D. Park, Y. Chung, H.-Y. Kim, and S. Yoon, "Fault Detection and Diagnosis of Railway Point Machines by Sound Analysis", doi: 10.3390/s16040549.
- [8] J. Yan and J. Lee, "Degradation Assessment and Fault Modes Classification Using Logistic Regression," 2005, doi: 10.1115/1.1962019.
- [9] H. Kim, J. Sa, Y. Chung, D. Park, and S. Yoon, "Fault diagnosis of railway point machines using dynamic time warping," 2016, doi: 10.1049/el.2016.0206.
- [10] P. Zhang, G. Zhang, W. Dong, X. Sun, and X. Ji, "Fault Diagnosis of High-Speed Railway Turnout Based on Convolutional Neural Network," *2018 24<sup>th</sup> International Conference on Automation and Computing (ICAC)*, Sep. 2018, pp. 1–6. doi: 10.23919/IConAC.2018.8749078.
- [11] Y. Sun, G. Xie, Y. Cao, and T. Wen, "Strategy for Fault Diagnosis on Train Plug Doors Using Audio Sensors†," pp. 24–27, 2018, doi: 10.3390/s19010003.
- [12] S. Ham, S.-Y. Han, S. Kim, H. J. Park, K.-J. Park, and J.-H. Choi, "A Comparative Study of Fault Diagnosis for Train Door System: Traditional versus Deep Learning Approaches," *Sensors*, vol. 19, no. 23, p. 5160, Nov. 2019, doi: 10.3390/s19235160.
- [13] Bernadin Namoano, "Fault Diagnosis in Time Series Data with Application to Railway Assets, PhD Thesis," Cranfield University, 2017.
- [14] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 289–296, 2001, doi: 10.1109/ICDM.2001.989531.
- [15] C. Ruiz-Carcel and A. Starr, "Data-Based Detection and Diagnosis of Faults in Linear Actuators," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 9, pp. 2035–2047, Sep. 2018, doi: 10.1109/TIM.2018.2814067.
- [16] A. Malhi and R. X. Gao, "PCA-Based Feature Selection Scheme for Machine Defect Classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004, doi: 10.1109/TIM.2004.834070.
- [17] M. Jordan, J. Kleinberg, and B. Schölkopf, "Pattern Recognition and Machine Learning".
- [18] A. Glowacz and Z. Glowacz, "Diagnosis of stator faults of the single-phase induction motor using acoustic signals," *Applied Acoustics*, vol. 117, pp. 20–27, Feb. 2017, doi: 10.1016/j.apacoust.2016.10.012.
- [19] D. Arthur and S. Vassilvitskii, "k-means++: The Advantages of Careful Seeding".
- [20] M. E. Celebi and K. Aydin, *Unsupervised Learning Algorithms*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-24211-8.
- [21] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, "Fault Prognosis," *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, pp. 284–354, Mar. 2007, doi: 10.1002/9780470117842.CH6.
- [22] Ian K Jennions, "Integrated Vehicle Health Management : The Technology," 2013.
- [23] R. J. Hansen, D. L. Hall, and S. K. Kurtz, "A New Approach to the Challenge of Machinery Prognostics," *Journal of Engineering for Gas Turbines and Power*, vol. 117, no. 2, pp. 320–325, Apr. 1995, doi: 10.1115/1.2814097.
- [24] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1–2, pp. 314–334, Jan. 2014, doi: 10.1016/J.YMSP.2013.06.004.



## BIOGRAPHY



**Minoru Shimizu** received his Bachelor in Physics and Master in Chemistry and Material Science from Tokyo Institute of Technology, Japan. He is currently pursuing his MSc by research in Transport Systems at the Integrated Vehicle Health Management (IVHM) Centre, Cranfield University. Before he joined Cranfield, he has worked for several years as a research engineer in a construction and mining machinery manufacturer. He was responsible for overseeing the research and development of IoT monitoring systems for construction and mining machinery to improve customer productivity. His research interests include fault detection, diagnosis, and prognosis by machine learning and deep learning for railway asset systems



**Suresh Perinpanayagam** has a Bachelor and Master in Aeronautical Engineering from Imperial College London. He pursued his PhD in Mechanical Engineering at the Rolls-Royce University Technology Centre at Imperial College, under the auspices of the European FP7 programme. He has worked on over 25 industry-led projects funded by industry partners, Aerospace Technology Institute (ATI), Engineering and Physical Sciences Research Council (EPSRC), UK Research and Innovation, and the European Commission. He has worked at the Ford Motor Company Development Centres in Dunton, UK and Merknick, Germany, and implemented an integrated data-centric engineering platform for new vehicle design and development. He is a Senior Lecturer in Intelligent Systems and a Chartered Engineer at Integrated Vehicle Health Management (IVHM) Centre / DARTeC. Suresh's expertise is in Data-Centric Engineering, Digital Twins & High-Fidelity Simulation, and Artificial Intelligence (AI) for Intelligent Systems.



**Bernadin Namoano** received his MSc degree in computer science in 2013 from Polytechnique University (France). Before joining Cranfield University in 2017, he worked as a software engineer in Paris, focusing on architectural design for big data, time series data analysis, and software lifecycle implementation and maintenance. He completed his PhD funded by EPSRC and Unipart-rail/Instrumentel at Cranfield University in the field of condition monitoring applied to railway assets and was awarded the EPSRC Doctoral Fellowship Prize. He is currently working on the development of novel techniques for detecting, diagnosing and predicting faults using big temporal data.