Original article

Evaluation of machine learning algorithms for detection of road induced shocks buried in vehicle vibration signals

Julien Lepine[1] , Vincent Rouillard[2] and Michael Sek[2]

Abstract

Road surface imperfections and aberrations generate shocks causing vehicles to sustain structural fatigue and functional defects, driver and passenger discomfort, injuries, and damage to freight. The harmful effect of shocks can be mitigated at different levels, for example, by improving road surfaces, vehicle suspension and protective packaging of freight. The efficiency of these methods partly depends on the identification and characterisation of the shocks. An assessment of four machine learning algorithms (Classifiers) that can be used to identify shocks produced on different roads and test tracks is presented in this paper. The algorithms were trained using synthetic signals. These were created from a model made from acceleration measurements on a test vehicle. The trained Classifiers were assessed on different measurement signals made on the same vehicle. The results show that the Support Vector Machine detection algorithm used in conjunction with a Gaussian Kernel Transform can accurately detect shocks generated on the test track with an area under the curve (AUC) of 0.89 and a Pseudo Energy Ratio Fall-Out (PERFO) of 8%.

Introduction

During operation, road vehicles are subject to shocks induced by road surface aberrations. These shocks can affect the vehicle structural and functional integrity, as well as driver and passenger comfort and health, and can also cause damage to freight. This is a significant problem as shock-causing surface aberrations are present on almost any road. Their harmful effect can be mitigated at different levels. For instance, road design and maintenance can be improved to reduce their occurrence; vehicle suspension systems can be improved to limit shock propagation throughout the vehicle chassis, and to occupants and freight. The efficiency of these preventive measures highly depends on the identification and characterisation of the shocks.

The protective packaging is another example of shock mitigation as its primary function is to protect products from distribution hazards such as the shocks and vibration produced by road vehicles. Current practice is to validate the effectiveness of protective packaging in laboratory tests involving physical simulation of

road transportation hazards as Gaussian random vibrations. However, even if this method is advocated by a number of standard organisations,[1–4] it has a fundamental limitation. The shocks and the vibration created by the randomness of a road profile are characterised concurrently, which has a significant effect on the severity and character of the simulated vibration.[5–8] Realistic simulations must consider two distinct components of Road Vehicle Vibration (RVV): (1) the nonstationary random vibration caused by variations in the road roughness and vehicle speed; and (2) the shocks caused by road surface aberrations.[9] The nonstationary random vibration results in a sustained excitation causing mostly fatigue-type damage, whereas the shocks are

[1]Department of Engineering, University of Cambridge, UK
[2]College of Engineering, Victoria University, Australia

Corresponding author:
Julien Lepine, Cambridge University Engineering Department, Trumpington St, Cambridge, CB2 1PZ, UK.
Email: J.lepine@eng.cam.ac.uk

short, high amplitude events causing more stress-driven damage or failure. To avoid the extra cost and environmental harm caused by over-packing, protective packaging must be optimised to provide just the right amount of protection to withstand the expected levels of hazards during transportation. This can only be achieved by accurately identifying and characterising the shocks that occur during transport.

A number of techniques have been used, with limited success, to detect shocks buried in RVV signals based on different analysis tool such as crest factors,[10–15] kurtosis,[12] Hilbert–Huang Transform (HHT)[16–18] and Discrete Wavelet Transform (DWT).[18–21] A thorough evaluation of these approaches[22] has revealed significant limitations especially when the underlying random vibrations are nonstationary. To improve shock detection reliability and effectiveness, the authors have proposed a method by which all these techniques are integrated into a machine learning classification algorithm.[9] This technique has been validated on artificially generated RVV signals but not yet tested on a real RVV measurement. This paper proposes and validates a complete process for training shock-detecting machine learning algorithms using an amalgamation of real and synthetic RVV signals. The best performing algorithm can then be used to efficiently identify road-generated shocks transmitted to a vehicle.

Methodology

Learning process

Machine learning algorithms are trained (taught) to recognise shocks for predictors using signals with known shock positions. These predictors are the output of different RVV analysis tools that reveal different behaviours in the signal that could indicate the presence of shocks. The minimum number of shocks required to perform supervised machine learning is substantial. Preliminary work on synthetic RVV signals showed that more than 300 shocks are needed for the training of the classifier.[22] Therefore, to teach classifiers uniquely from real RVV signals, measurement must be performed on a vehicle passing over 300 well-identified shocks. As classifiers are specifically trained for certain vehicle payloads, it is unrealistic to apply this approach to characterise RVV in real-life. Long measurements can be made on road vehicles relatively easily; however, surveying long sections of roads in order to accurately identify the position of the shocks is far more challenging.

To overcome this limitation, a learning method integrating both real and synthetic signals to train the classifier is proposed (Figure 1). This method requires two datasets of RVV measurements to be made on the same vehicle which are: (1) the training measurement dataset where the position of the shocks is not identified and (2) the validation measurement dataset where the position of the shocks is identified.

The training measurement dataset with a duration of 3500 s was collected on open roads representing normal travelling routes. This measurement dataset is used to characterise the vehicle's dynamic characteristics from which a synthetic signal is generated. The RVV synthetic signal is nonstationary (varying RMS level) and contains a sufficient amount of shocks needed for optimal classifier training. This synthetic signal provides a training dataset specific to the vehicle payload.

Once the classifier training is completed, the classifiers are validated using the validation measurement datasets. These measurement datasets are shorter than the training measurement dataset and are performed on road circuits where the position of the shocks is identified. This validation is essential to calibrate the classifier's threshold and to define the validity of the classification. The calibrated classifiers can then be used to identify shocks in other measurements made with the vehicle.

Predictors

As stated in the introduction, different techniques have been used, with limited success, to detect shocks buried in RVV signals (e.g. moving crest factors,[10–15] moving kurtosis,[12] HHT[16–18] and DWT[18–21]). The benefit of using the machine learning process to detect shocks is
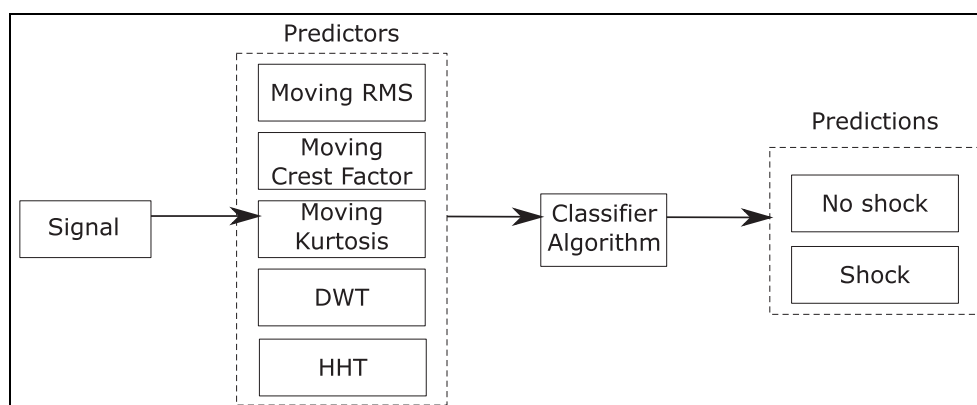

Figure 1. Learning flow chart to detect shocks buried in real RVV signal.
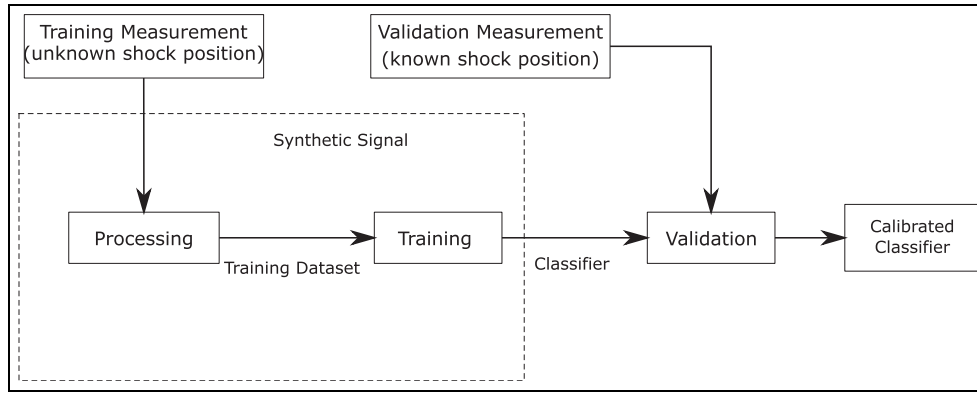
Figure 2. Machine learning process used to detect shocks.

that it can enhance shock-detection performance by combining the best performing analysis techniques found in the literature. As shown in Figure 2, the classifier algorithms based their prediction on the so-called predictors (which are the outcomes of these RVV analysis techniques). The complete development of the machine learning process has been previously published by the authors.[9,22] A brief overview of the process is presented here.

The process starts with decomposing the analysed signal into predictors. The first predictor class consists of the moving Root Mean Square (RMS) value of the signal computed with two different window lengths, T, i.e. 0.5 s and 4 s;

$$m_{RMS}(t) = \sqrt{\frac{1}{T}\int_0^T x(t+\tau)^2 d\tau}. \qquad (1)$$

The second predictor class consists of the moving crest factor computed with two window lengths, T, i.e. 8 s and 64 s;

$$m_{CF}(t) = \frac{|x(t)|}{\sqrt{\frac{1}{T}\int_0^T x(t+\tau)^2 d\tau}}. \qquad (2)$$

The third predictor class consists of the moving kurtosis which represents the fourth moment of the signal. The moving kurtosis is computed with two window lengths, T, i.e. 4 s and 8 s;

$$\kappa(t) = \frac{\frac{1}{T}\int_0^T [x(t+\tau)]^4 d\tau}{\left(\frac{1}{T}\int_0^T [x(t+\tau)]^2 d\tau\right)^2} \qquad (3)$$

Note how the moving statistics (equations 1–3) are calculated "forward"$(t+\tau)$ to consider the vehicle causal reaction from the road excitation.

The fourth predictor class comes from the DWT. This predictor is composed of the coefficients of the first 12 scales of DWT using the Daubechies 10 mother wavelet.

The fifth and last predictor class consists of the HHT which is an adaptive time-frequency transform developed by Huang et al.[23] The HHT uses a sifting process to decompose the signal into narrow-banded components called Intrinsic Mode Functions (IMF). The instantaneous frequency and envelope amplitude functions of each IMF are then computed using Hilbert Transform and used as predictors.

Dataset measurement

All the measurement datasets required to train and validate shock classifiers were measured with the same small utility vehicle, Mitsubishi Triton. The vehicle has a kerb mass of 1555 kg and can transport a maximum payload of 1165 kg. During the measurement, the only payload was the weight of two persons sitting in the cabin ($\approx$150 kg). The vehicle has a front suspension composed of coil springs with telescopic dampers and a rear suspension comprising leaf springs with telescopic dampers. Its wheelbase and track width are 3.0 m and 1.5 m respectively.

A Slam Stick X accelerometer (Midé Technology, USA) was fixed on the chassis of the vehicle near the left leaf spring rear mount. This is approximately 1.5 m away from the centre of gravity of the vehicle. As recommended by Long,[24] this position away from the centre of gravity reduces the effect of wheelbase filtering which can otherwise create an anti-resonance in the measured spectrum. The signal was sampled at 1024 Hz with the MEMs DC accelerometer presents on the Slam Stick which has a range of $\pm$160 m/s^2 and a flat response up to 1000 Hz. Only the vertical acceleration was measured, as this is the main axis affected by shocks.

Training measurement dataset

The objective of the training measurement dataset is to characterise the vehicle dynamics during normal operation. The measurement was performed on 52.8 km of

Table 1. Summary of the measurement datasets.

Measurement	Road type	Time (s)	Speed (km/h)			Distance (km)	Number of shocks ( )	RMS (m/s^2)
			Mean	Standard deviation	Max			
Training dataset	Public road	3500	55.6	29.8	103	52.8	N.A.	1.6
Validation dataset #1	Asphalt circuit	680	38.7	8.5	67.1	7 × 1	42	1.5
Validation dataset #2	Off road circuit	450	19.5	6.8	29	10 × 0.25	40	2.5
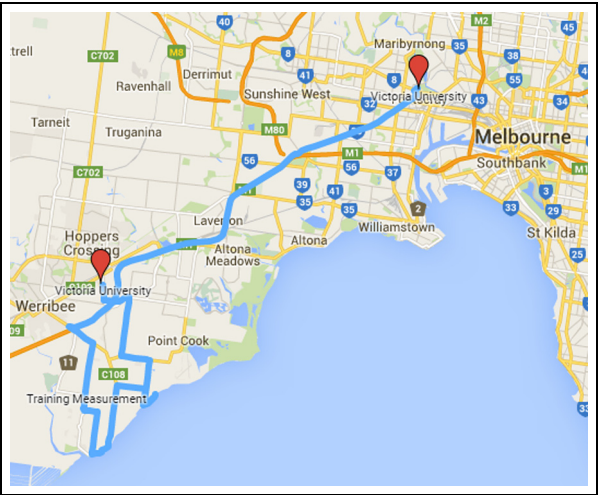
Figure 3. Geographic location of the training measurement dataset.

Figure 4. Map of the validation measurement datasets and the shock positions.

public roads in the west of Melbourne, Australia (Figure 3), which represents a mix of typical motorways, urban roads and country roads. Vehicle location and speed were recorded with a GPS (summarised in Table 1). The measurement duration was close to 1 h (3500 s) and the vehicle had a mean speed of 55 km/h with a standard deviation of 29.8 km/h. It reached a maximum of 103 km/h. The overall RMS value of the vehicle vertical acceleration was 1.6 m/s^2.

Validation measurement datasets

The objective of the validation measurement datasets is to calibrate the classifier's Optimal Operation Point (OOP) and to assess how reliably shocks are detected. Therefore, the location of the shocks encountered during the measurements needs to be clearly identified. To do so, the measurements were performed on closed circuits. As the classifier needs to work with varying RMS levels (nonstationary components), asphalted and off-road circuits were used. These circuits were located at Victoria University's Werribee campus.

The validation measurement dataset #1 was performed on a 1 km circuit of smooth asphalted road free from any significant surface aberrations (Figure 4). Six steel plates were fixed on the road pavement to gener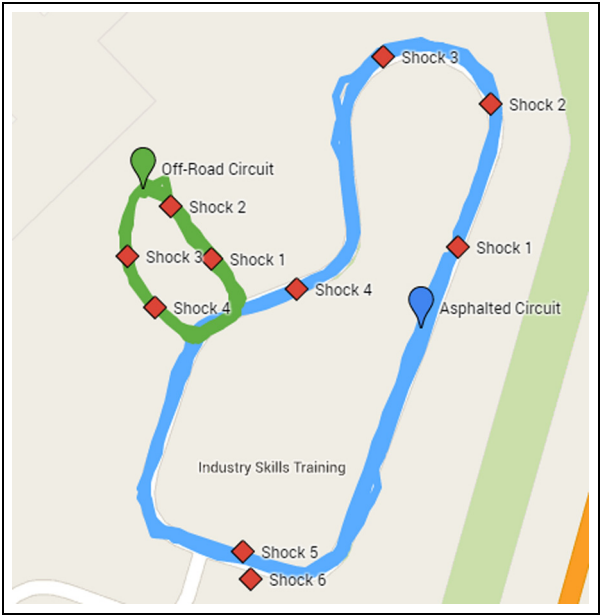ate shocks when in contact with the left wheels of the vehicle (same side as the accelerometer). Shocks #2 and #5 were composed of a 10 mm high by 80 mm long steel plate which was considered as a small shock. Shocks 3 and 4 were composed of a 15 mm high by 100 mm long steel plate which was considered as a medium shock. Shocks 1 and 6 were composed of a 20 mm high by 150 mm long steel plate which was considered as a large shock.

The vehicle was driven around the asphalted circuit seven times for a total of 42 shocks (Table 1). The measurement duration was 680 s with the shocks density of 6.2 shocks per 100 s of signal. The vehicle had an average speed of 38.7 km/h (standard deviation and a maximum of 8.5 km/h and 67.1 km/h, respectively). The RMS level of the vertical acceleration was 1.5 m/s^2, which was similar to the value measured for the training measurement dataset.

The validation measurement dataset #2 was performed on a 250 m off road circuit (Figure 4). The surface was composed of rough uneven gravel and rocks. Four artificial shocks were added to the circuit but, given the quality of the road surface, there may have been other surface aberrations causing the generation
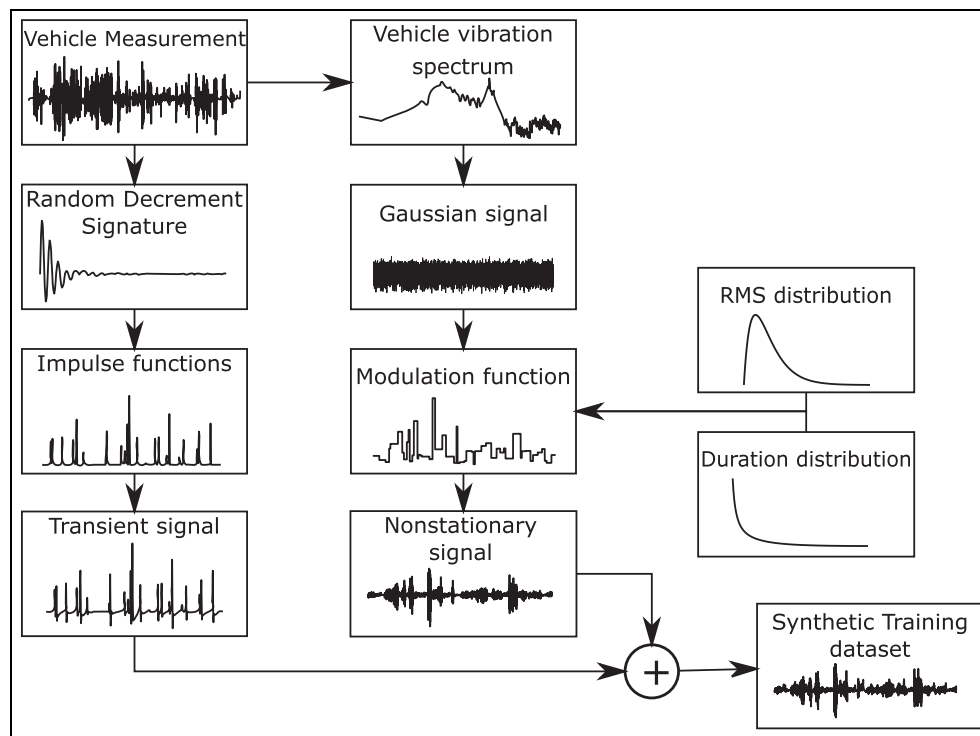
Figure 5. Flow chart of the synthetic signal creation from the learning dataset.

of unidentified shocks. For the first five laps, shocks 2 and 4 were composed of a 15 mm high by 80 mm long steel plate and for the remaining five laps they were replaced by a 48 mm high by 98 mm long lumber beam in order to increase the shocks amplitude. This change was made because these shocks were perceived by the experimenters during the measurement as not severe enough for shock detection purposes. Shocks 1 and 3 were composed of a 20 mm high by 150 mm long steel plate throughout the 10 laps of the measurement.

The vehicle was driven around the off-road circuit 10 times (2.5 km) producing a total of 40 shocks (Table 1). The measurement duration was 450 s and the artificial shocks density of 8.9 shocks per 100 s, which was similar to the validation dataset #1. The vehicle had an average speed of 19.5 km/h (standard deviation and maximum values of 6.8 km/h and 29 km/h, respectively). The RMS level of the vertical acceleration was 2.5 m/s^2, which was representative of the severity of the off-road circuit.

Synthetic training dataset

It is impractical to perform supervised classifier learning on a RVV measurement because the required number of perfectly identified shocks is too large. To overcome this issue, a synthetic RVV signal needs to be created using the vehicle's dynamic characteristics identified from the learning measurement dataset. The nonstationary component of the RVV signal was created using the vehicle's vibration spectrum and the statistical distribution of the RMS, whereas the shocks (transient
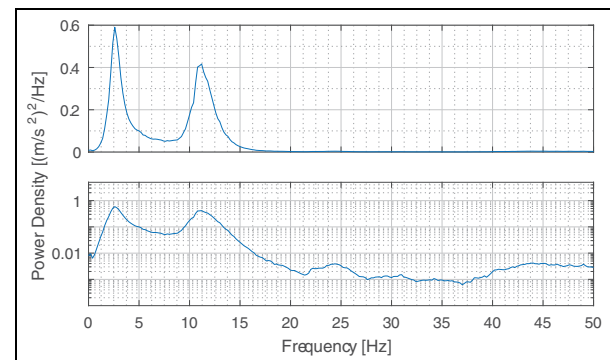

Figure 6. PDS ($\Delta f = 0.2$ Hz, Hanning window, 1405 averages) of the training measurement dataset (linear scale at the top and semi-log scales at the bottom).

signal) were created using the Random Decrement Signature of the vehicle (Figure 5).

Nonstationary signal

The nonstationary component of RVV was modelled with a series of Gaussian signals of different RMS values. To represent a realistic spectrum, the Gaussian sequences were created from the average Power Density Spectrum (PDS) calculated on the training measurement dataset (Figure 6).

To create a random Gaussian time signal with the same spectrum as the vehicle, the PDS was then scaled into an amplitude (acceleration) spectrum with a uniformly-distributed random phase. The frequency resolution of the spectrum was interpolated to match
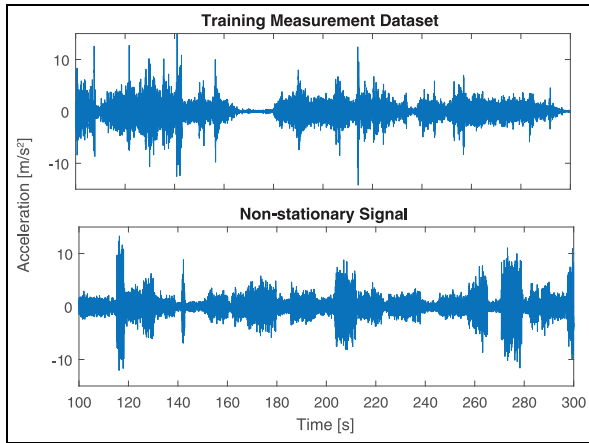
Figure 7. Typical segment of the training measurement dataset and nonstationary synthetic signal.

Figure 8. Random decrement signature of the vehicle.

Figure 9. Acceleration spectrum of the random decrement signature of the training measurement dataset.

the desired time signal duration. Once interpolated, the spectrum was transformed to the time domain with the Inverse Fast Fourier Transform (IFFT).

The RMS value of the resulting signal was then modulated in a random fashion to generate a nonstationary signal. The distribution of the duration and relative amplitude of the segments used for modulation follow the distributions of typical RVV as measured by Rouillard.[6] The modulation levels were scaled to match the overall RMS level of the measurement, which was 1.6 m/s². Samples of the training measurement dataset and the equivalent nonstationary signal are presented in Figure 7.

Transient signal

Shocks superimposed onto the RVV signals were synthesised using the impulse response function of the vehicle obtained using *in-situ* measurement of the Random Decrement Signature[25,26] such as:

$$D_k = \frac{1}{N} \sum_{k=1}^{w} x(t_{i+k-1}), \qquad (4)$$

where the *k*th point of the signature D_k is the average of N points from N segments of the signal with window width w, and t_i is the commencing data point that meets trigger criteria for the *N*th segment.

For instance, the triggering condition can be set on the zero-crossing where the Random Decrement Signature will be calculated every time the signal crosses zero. The triggering conditions can be alternatively set on a threshold value and on the signal slope polarity.

The optimal parameters to estimate the impulse response of the vehicle using Random Decrement Signature were studied by Ainalis.[25] The trigger was set on the zero-crossing with no specific slope polarity and a minimal number of points between the averaging. The triggering condition produced around 30,000 averages on the training measurement dataset which provided accurate impulse response estimation. A 2 s
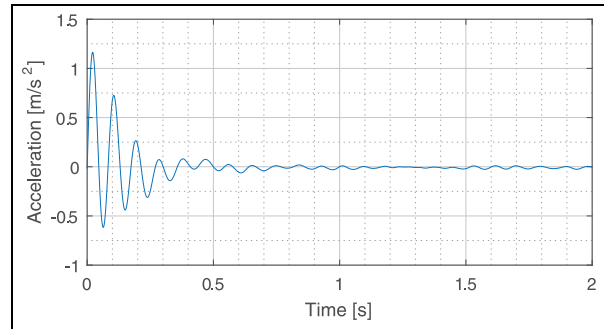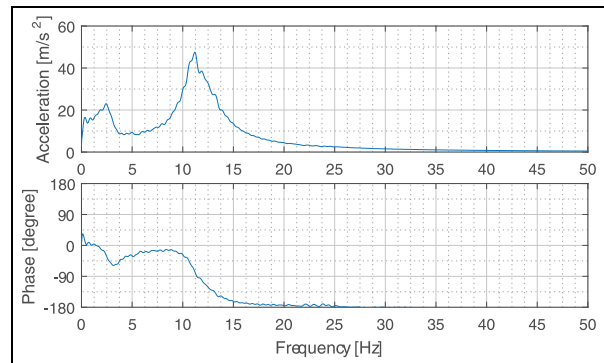
window was used for the signature length as it was found sufficient to allow complete decay of the signature. The measurement dataset was filtered with a low-pass filter cut-off at 25 Hz (5th order Butterworth) in order to estimate only the first two natural frequencies of the vehicle.

The Random Decrement Signature of the vehicle is presented in Figure 8. The frequency spectrum of the signature gives a scaled estimate of the vehicle FRF (Figure 9).

To generate shocks, the Random Decrement Signature was convolved with a sequence of randomly distributed half-sine shock (impulse) functions. These impulses had randomly distributed amplitude and duration ranging between 5 mm and 40 mm and 0.5 s and 1.4 s, respectively, which generate vehicle transient responses (i.e. shocks) with maximum absolute amplitudes ranging between 1.6 m/s² and 22 m/s². The shock density was set at 10 shocks per 100 s of signal, which is slightly more than the density of the validation datasets. The resulting transient signal was superimposed onto the nonstationary signal (Figure 10). The position of the shock functions was used to index the shocks during the learning process.

Classifier algorithms

The shock detection performance of several machine learning classifiers has been assessed using synthetic
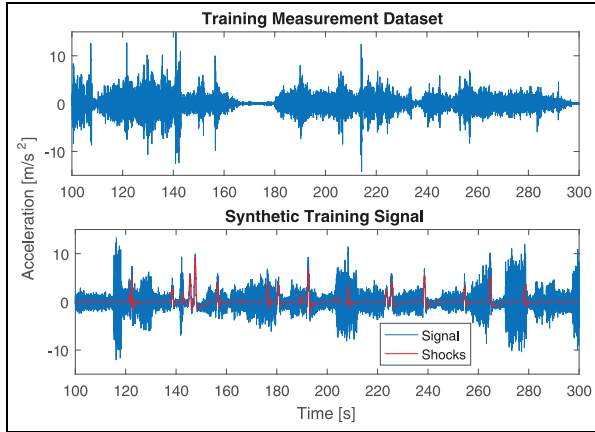
Figure 10. Typical segment of the training measurement dataset and synthetic training signal (nonstationary and shocks).

RVV signals in preliminary work undertaken by Lepine.[22] From all the classifiers evaluated, the best four were selected for detection on a real RVV signal, i.e. the Decision Tree, k-Nearest Neighbours (kNN), Bagged Ensemble and Support Vector Machine (SVM).

The Decision Tree algorithm bases its prediction on a cascade of statistical tests made on the predictors. For shock detection application, a Decision Tree composed of 20 tests was used.

The kNN algorithm groups a training dataset by class in as many spaces as there are predictors. It classifies new data points by grouping them to the most common class of their k nearest neighbours. In other words, the classification is made by association. For shock detection application, a 100NN algorithm is used. The Euclidean distance is used to define the nearest neighbours.

The Bagged Ensemble algorithm uses a combination of Decision Trees on which it bases their prediction. Bagged algorithm stands for "bootstrap aggregation". It generates Decision Trees on different random data samples. The classification is made by using the average response of each of these Decision Trees. An ensemble of 150 decision trees was used for shock detection purposes.

The SVM algorithm finds a hyperplane that maximises the distance between the two classes' data points represented in all their dimensions (predictors). A Gaussian Kernel transform is used to fit and hyperplane between the classes of the training dataset.

These algorithms were trained using their implementation in Matlab®. For more information refer to Hastie et al.,[27] Rogers and Girolami,[28] Cherkassky and Mulier,[29] Shalev-Shwartz and Ben-David.[30]

Classifier training

Decision Tree, 20NN, Bagged Ensemble and SVM classifiers were trained with the synthetic training signal composed of 300 shocks (10 shocks per 100 s). The

validity of the algorithms was initially verified using a synthetic signal. This validation was made with the same type of synthetic signal composed of 350 shocks (10 shocks per 100 s).

Results

Synthetic validation

Under ideal conditions, synthetic RVV signals would be sufficient to fully train, calibrate and validate the classifiers. However, as described in the next section this was not possible as the OOPs defined with this synthetic calibration were not sufficiently sensitive to detect real measurement shocks. The synthetic calibration approach was therefore performed to ensure the classifiers' learning was effective and enabled comparison of their performance with that of the measured signals.

One method to assess the classifier performance is to evaluate the relationship between the detection sensitivity (probability of detecting a real shock) and the fall-out (probability of detecting a shock when there is none).

In the context of detection of shocks buried in the RVV signal, the shocks are considered as a sequence of data points instead of individual data points. The sensitivity is therefore defined as the number of true detection sequences over the total number of shocks,

$$\text{Sensitivity} = \frac{\text{Number of True Detections}}{\text{Number of shocks}}, \quad (5)$$

where a detection has to overlap at least 75% of the shock duration to be considered true. This value ensures most of the shock duration is detected.

As opposed to sensitivity, the specificity is the proportion of the signal without shock that is correctly classified as such. This can be considered as the accuracy of prediction of the duration of the signal without transients. The specificity is defined as the number of true non-detections data points over the number of data points without shocks,

$$\text{Specificity} = \frac{\text{Number of Nondetection Data Points}}{\text{Number of Data Points without shocks}}. \quad (6)$$

The sensitivity and specificity are particular to an operating point, i.e. the threshold value used for the detection. As the prediction threshold decreases, the classifier predicts a shock with less confidence increasing both the number of real detections (sensitivity) and the number of false detections (fall-out). The Receiver Operating Characteristic (ROC) curve illustrates this relationship. The Area Under the Curve (AUC) is commonly used to evaluate the classifiers performance.[31,32] The ideal classifier has an AUC of 1 and a classifier based on chance has an AUC of 0.5. This means that a classifier with an AUC above 0.5 has a better prediction performance than chance.
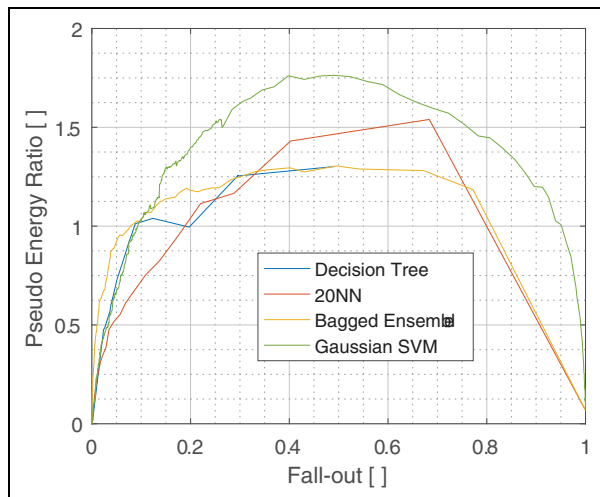
Figure 11. PERFO curves of the synthetic validation dataset.

Table 2. Classifiers' performance comparison.

Classifier	AUC	PERFO fall-out
Decision tree	0.86	0.08
Euclidean 20NN	0.86	0.17
Bagged ensemble	0.92	0.08
Gaussian SVM	0.89	0.08

Another method used to evaluate the classifier performance was based on the Pseudo Energy Ratio—Fall-Out (PERFO) curve. It was specifically developed to assess shock detection performance.[22] Since the importance of detecting shocks increases with their amplitude, the classifiers' accuracy and quality can be assessed as the ratio between the detection and the shocks' maximum absolute acceleration pseudo energy, calculated as:

$$R = \frac{\sum [\text{Detections' Max Absolute Acceleration}]^2}{\sum [\text{Shocks' Max Absolute Acceleration}]^2}.$$

(7)

As for sensitivity, the pseudo energy ratio depends on the classifier's detection threshold and directly affects the fall-out level. As for the ROC curve, this relationship can be illustrated by a curve which is called the PERFO curve (Figure 11). The fall-out at the pseudo energy ratio of one indicates the level of false detection when the distribution of the detections' maximum values is equivalent to the distribution of the real shocks, which is a good indication of the classifier's performance. This specific point on the PERFO curve can also be used to define the classifiers' prediction threshold (OOP).

The classifier validation is made with those two measurands (Table 2): i.e. the AUC and the Fall-out value at the OOP defined with the PERFO criterion. All the classifiers had an AUC value between 0.86 and 0.92 and PERFO fall-out value between 0.08 and 0.17, which indicated good detecting performance on the synthetic RVV signal.

Validation on real measurements

Validation using a synthetic validation dataset gave an appreciation of how the classifiers were successfully trained, but it did not give any insight into their application to real measurement datasets. This ultimate assessment can be performed using the same validation procedure presented in the previous section on the validation measurement datasets, which are composed of the acceleration signal record on a vehicle travelling on test tracks.

One of the first observations from the validation measurement datasets application is that finding classifiers' OOP with a synthetic dataset was inadequate. It was found that the detection threshold defined from the synthetic dataset using the PERFO criterion was not sensitive enough for real shock detection. Using this threshold, each classifier detected less than 10 shocks in all the validation measurement datasets composed of more than 80 shocks. The classifiers needed to be recalibrated using the PERFO criterion on the validation measurement datasets.

Before inspecting the PERFO curves, it is important to look at the ROC curves to evaluate the actual sensitivity of the classifiers. The PERFO curve is a good tool to assess the detection algorithm's maximum amplitude with respect to the maximum amplitude of the real shock. However, it does not indicate if the detections are true or false. For instance, a classifier can have a pseudo energy ratio of one just by guessing where the shocks are without using any significant statistics. This is where the ROC curves have an important role because they represent the relationship between the sensitivity and the fall-out, which gives a good appreciation of the classifiers' detection accuracy.

To assess the detection accuracy, ROC curves were computed with the validation measurement dataset made on the asphalted circuit (Figure 12). This circuit has potentially fewer unidentified shocks than the off-road circuit so it provides more accurate ROC curves. These ROC curves lead to very interesting results. The detection of the Decision Tree, 20NN and Bagged Ensemble classifiers are not better than a classifier based on random guessing (chance). They have AUC values of 0.49, 0.43 and 0.53 respectively. Considering the potential uncertainty associated with AUC,[33] these values are equivalent to 0.5, which represents detecting shocks by guessing. Only the Gaussian SVM provides accurate detections with an AUC of 0.94, which is a 6% increase from the synthetic validation.

The Gaussian SVM's PERFO curve was computed for both validation measurement datasets (Figure 13). The off-road curve reaches a pseudo energy ratio of one at a higher fall-out value (0.06) than the asphalted curve
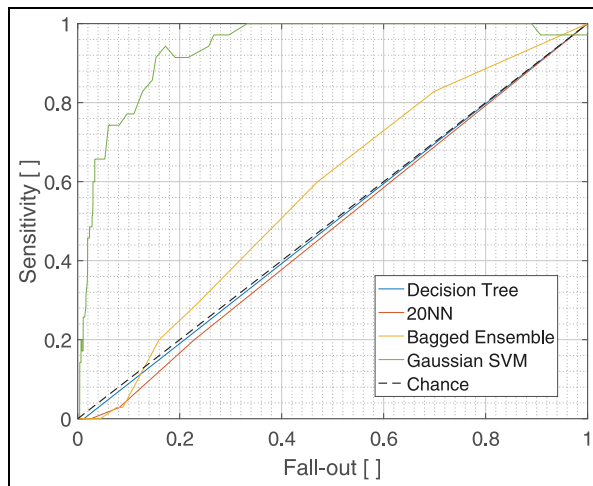
Figure 12. ROC curves of the validation measurement dataset made on the asphalted circuit.
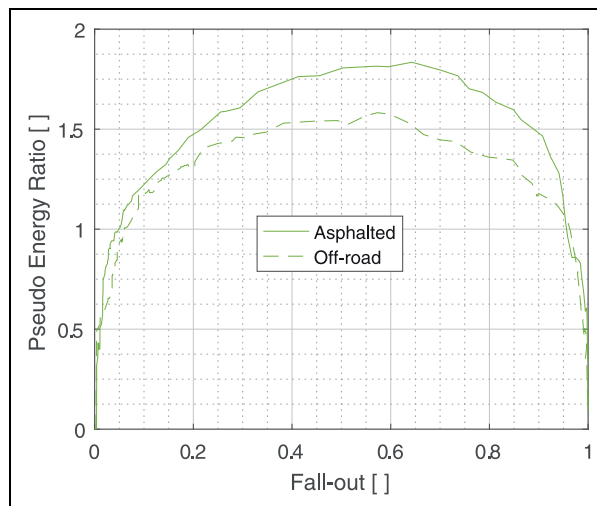

Figure 13. Gaussian SVM's PERFO curve of the validation measurement datasets.

Figure 14. Maximum absolute acceleration distributions of the detections and shocks of the asphalted circuit validation measurement dataset (#1).

(0.05). This could have been caused by the unidentified shocks present on the off-road circuit that could have been detected. The small difference between both values suggests that there may not be too many unidentified shocks and the off-road measurement dataset is reliable for validation purposes.

The OOP calculated from the PERFO criterion gives a detection threshold for the asphalted and off-road circuit of -0.96 and -0.87 respectively. The SVM represents shocks in the positive space and the 'no-shocks' in the negative space, so the smaller the detection threshold is, the more shock sensitive the classifier is. In theory, scores under -1 represent the space where the 'no shock' events of the training dataset are separated from the shocks. The detection threshold being very close to this value suggests that the shocks buried in the synthetic dataset were too easily identifiable. In other words, the artificial shocks superimposed on the
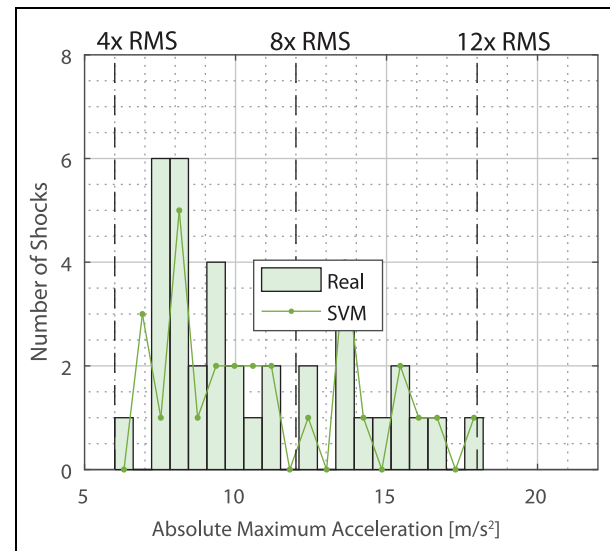
synthetic nonstationary signal were too severe. This also explains why the synthetic calibration was not sensitive enough to detect real shocks.

The difference between the threshold values from both circuits is not significant enough to affect the detection distributions. For both validation measurement datasets, the maximum amplitude distribution remains the same using either detection thresholds (Figures 14, 15). There are almost no errors on the high amplitude shocks encountered on the asphalted circuit. This high amplitude accuracy was expected as this circuit has a smooth pavement which does not generate high amplitude vibration that could be misclassified as shocks. The same level of high amplitude detection accuracy is almost achieved on the off-road measurement where only one shock above 16 m/s² (10.7 times the overall RMS value) is misdetected.

As the off-road measurement inherently contains more background vibration, its low amplitude detections are less accurate than its asphalted counterpart. Nonetheless, detection is still within the range of the synthetic validations (Figure 15). For both validation datasets, most of the non-detected shocks have low maximum acceleration (between 6 m/s² and 9 m/s²; i.e. 2.4–3.6 times the overall RMS value). There are more over-detections on the off-road validation dataset than on the asphalted dataset, especially between 10 m/s² and 15 m/s² which could be caused by unidentified shocks present on the off-road circuit.

Another way to appreciate the accuracy of shock detections is to carry out direct observation in the time domain. Figure 16 shows the location of the detected shocks in the asphalted circuit measurement dataset. The classifier missed Shock #5 four times over seven, it was a 'small' shock positioned 10 m before a big shock (Shock #6). Other than that, it seems as if there are no
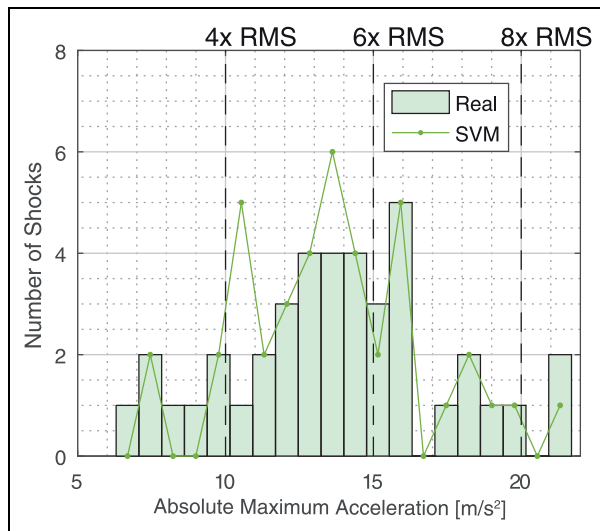
Figure 15. Maximum absolute acceleration distributions of the detections and shocks of the off-road circuit validation measurement dataset (#2).

apparent patterns in the false- and misdetection in this measurement dataset.

The interpretation of the off-road measurement dataset's detection in the time domain (Figure 17) is interesting as the signal is noisier and the shocks do not clearly stand out as on the asphalted circuit. Most of the false-detections occurred in the last segment of the lap where the circuit surface was rougher and where there were more likely to be unidentified shocks.

There are very few misdetections in the measurement dataset. Most of those occurred in the first five laps where Shocks #2 and #4 were physically smaller than for the remaining five laps. Each of those smaller shocks was missed twice during the first five laps. This is remarkable because these shocks were detected most of the time even though these two smaller shocks were perceived as not severe enough for shock detection purposes by the experimenters present in the vehicle during the measurement. This suggests that the Gaussian SVM classifier could be better than humans at detecting shocks. The two other misdetections are Shocks #3 and #4 (in their more severe state).
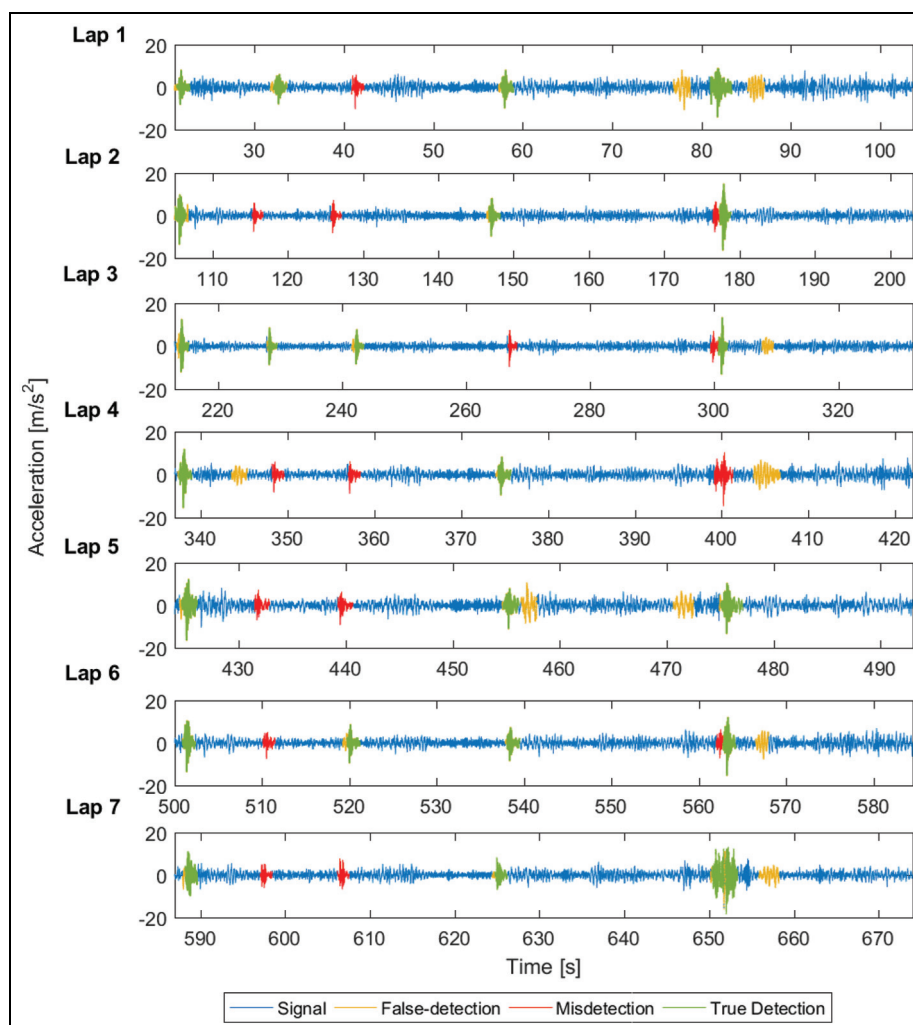

Figure 16. Position of the detections and shocks in the asphalted circuit validation measurement dataset (#1).
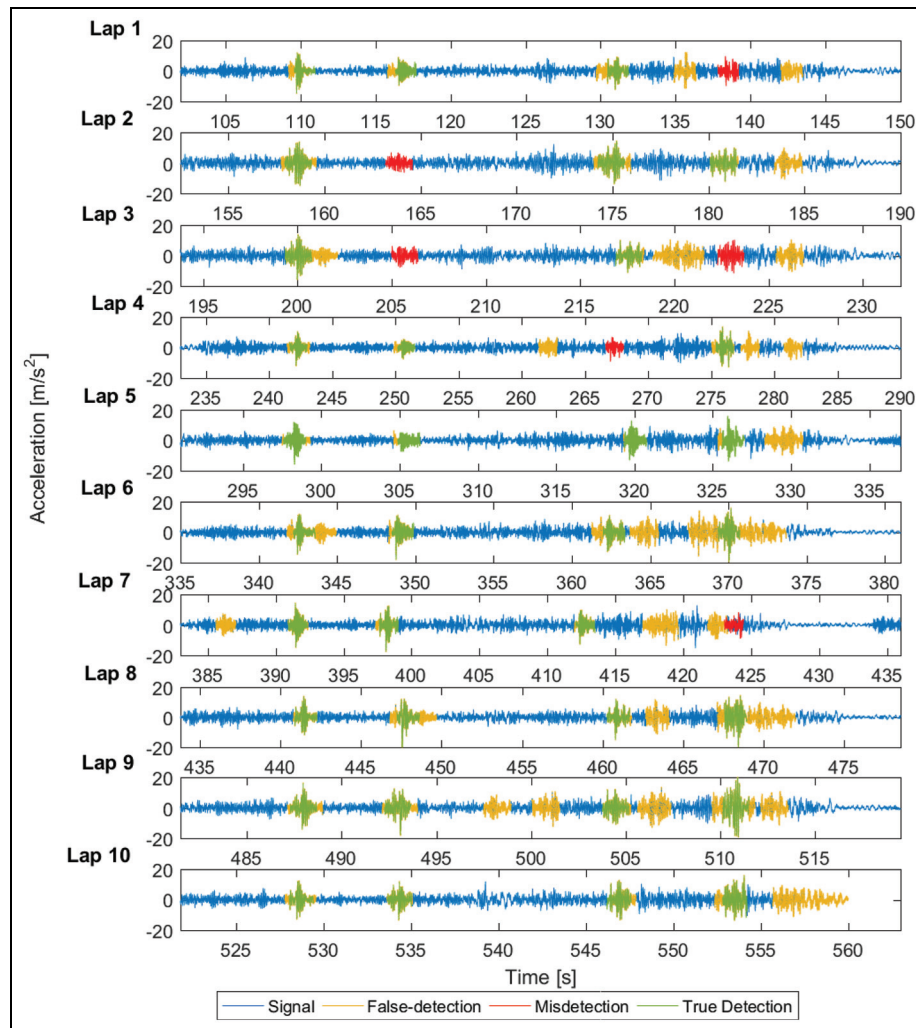
Figure 17. Position of the detections and shocks in the off-road circuit validation measurement dataset (#2).

Discussion

The Gaussian SVM is the only classifier that successfully detects shocks buried into the validation measurement datasets. This is despite all the classifiers assessed having AUC values well over 0.5 when validated with a synthetic dataset. This means that learning from a synthetic dataset cannot be adapted for a real measurement dataset for the Decision Tree, 20NN and Bagged Ensemble classifiers; only the Gaussian SVM can manage this adaptation. This could be because the Gaussian Kernel transform used with the SVM makes it the only nonlinear classifier in this evaluation. It is, therefore, possible that real shocks have behaviours that cannot be accounted for by linear classifiers.

Considering the relative importance of detecting high amplitude shocks, the Gaussian SVM classifier performs very well on real RVV measurements as most of the misdetected shocks have low maximum amplitude. Results on the asphalted road also show that there are relatively few false detections in the classifier predictions.

An interesting feature of the machine learning process presented in the paper is that a trained classifier can be used to identify the shocks buried in the training measurement dataset. The classifier being trained on a synthetic reconstruction of the training measurement dataset, the classifier has no *a priori* knowledge of the signal; there is no risk of bias of the detection made on the training measurement dataset. This is an important benefit of this approach; the classifier can be trained and applied to the same signal after a calibration from a short validation measurement dataset.

Implementation remarks

It was shown that the Gaussian SVM classifier can detect shocks on real RVV measurements, but some additional validations have to be undertaken before generalising and implementing the algorithm to any specific RVV and shocks problems. For instance, the limits of the learning have to be defined, i.e. assessing of a specific learning be applied to the same type of vehicle and to different payloads. Different vehicles may require a fully customised artificial training dataset or simply a customised validation.

The complexity of the validation depends on the application contexts such as specific route distributions,

transport vehicles and shipments. This would be the final step before the classifier implementation to model the RVV. However, such work is beyond the scope of this paper, which was to explore how the different modes of RVV can be identified.

Conclusion

This paper has demonstrated that at least one machine learning classifier can be used to detect shocks buried in real RVV signals (acceleration measurements), namely the Gaussian SVM. This classifier was trained on a synthetic dataset which was created from the dynamic behaviour of a specific vehicle. The vehicle dynamics were characterised with a training measurement dataset made during normal operation. Once trained, the classifier was calibrated and validated using measurement datasets made on closed circuits composed of artificially generated shocks.

Declaration of conflicting interests

Funding

ORCID iD

Julien Lepine [iD] https://orcid.org/0000-0003-3682-846X

References

1. ASTM-D4728. *Test method for random vibration testing of shipping containers*. West Conshohocken, PA: ASTM International, 2012.
2. MIL-STD-810F. *Environmental test methods and engineering guides*. USA: Department of Defense Test Method Standard, 2000.
3. ISO-13355. *Packaging. Complete, filled transport packages and unit loads. Vertical random vibration test*. London, UK: BSI, 2002, 16 pp.
4. ISTA. Test series, https://ista.org/test_procedures.php (accessed 6 February 2018).
5. Charles D. Derivation of environment descriptions and test severities from measured road transportation data. *J IES* 1993; 36: 37–42.
6. Rouillard V. *On the synthesis of non-Gaussian road vehicle vibrations*. PhD Thesis, Monash University, Melbourn, Australia, 2007.
7. Kipp WI. Random vibration testing of packaged-products: considerations for methodology improvement. In: *16th IAPRI world conference on packaging*, Bangkok, Thailand, 2008.
8. Sek MA. Optimisation of packaging design through an integrated approach to the measurement and laboratory simulation of transportation hazards. In: *12th international conference on packaging*, Warsaw, Poland, 2001.
9. Lepine J, Rouillard V and Sek M. On the use of machine learning to detect shocks in road vehicle vibration signals. *Packag Technol Sci* 2017; 30(8): 387–398.
10. Rouillard V, Bruscella B and Sek M. Classification of road surface profiles. *J Transp Eng ASCE* 2000; 126: 41–45.
11. Bruscella B, Rouillard V and Sek M. Analysis of road surface profiles. *J Transp Eng ASCE* 1999; 125: 55–59.
12. Bruscella B. *The analysis and simulation of the spectral and statistical properties of road roughness for package performance testing*. Doctoral Dissertation, Department of Mechanical Engineering, Victoria University of Technology, Melbourne, Australia, 1997.
13. Garcia-Romeu-Martinez MA, Rouillard V and Cloquell-Ballester VA. A model for the statistical distribution of road vehicle vibrations. In: *World congress on engineering*, London, UK, 2007, pp.1225–1230.
14. Rouillard V and Sek MA. A statistical model for longitudinal road topography. *Road Transp Res* 2002; 11: 17–23.
15. Rouillard V and Sek M. Creating transport vibration simulation profiles from vehicle and road characteristics. *Packag Technol Sci* 2013; 26: 82–95.
16. Lepine J, Sek M and Rouillard V. Mixed-mode signal detection of road vehicle vibration using Hilbert–Huang transform. In: *ASME 2015 noise control and acoustics division conference at internoise 2015*, American Society of Mechanical Engineers, August 2015, pp.V001T01A002–V001T01A002.
17. Lepine J, Rouillard V and Sek M. Using the Hilbert-Huang transform to identify harmonics and transients within random signals. In: *8th Australasian congress on applied mechanics*, ACAM 8, 2014.
18. Ayenu-Prah A and Attoh-Okine N. Comparative study of Hilbert–Huang transform, Fourier transform and wavelet transform in pavement profile analysis. *Vehicle Syst Dyn* 2009; 47: 437–456.
19. Nei D, Nakamura N, Roy P, et al. Wavelet analysis of shock and vibration on the truck bed. *Packag Technol Sci* 2008; 21: 491–499.
20. Wei L, Fwa T and Zhe Z. Wavelet analysis and interpretation of road roughness. *J Transp Eng ASCE* 2005; 131: 120–130.
21. Lepine J, Rouillard V and Sek M. Wavelet transform to index road vehicle vibration mixed mode signals. In: *ASME 2015 noise control and acoustics division conference at internoise 2015*, 2015, pp.V001T01A003–V001T01A003.
22. Lepine J. *An optimised machine learning algorithm for detecting shocks in road vehicle vibration*. PhD Thesis, College of Engineering and Science, Victoria University, Melbourne, Australia, 2016.
23. Huang NE, Shen Z, Long SR, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond Ser A* 1998; 454: 903–995.
24. Long M. *On the statistical correlation between heave, pitch and roll motion of road transport vehicles*. PhD Thesis, College of Engineering and Science, Victoria University, Melbourne Australia, 2016.
25. Ainalis DT. *Estimating the dynamic characteristics of road vehicles using vibration response data*. Doctoral

Dissertation, College of Engineering and Science, Victoria University, Melbourne, 2014.

26. Milliken P, de Pont J, Mueller T, et al. Assessing road-friendly suspensions. Transfund New Zealand Research Report, 2001.

27. Hastie T, Tibshirani R, Friedman J, et al. *The elements of statistical learning: data mining, inference and prediction*, vol. 27. New York, USA: Springer, 2005.

28. Rogers S and Girolami M. *A first course in machine learning*. Florida, USA: CRC Press, 2011.

29. Cherkassky V and Mulier FM. *Learning from data: concepts, theory, and methods*. New Jersey, USA: John Wiley & Sons, 2007.

30. Shalev-Shwartz S and Ben-David S. *Understanding machine learning: from theory to algorithms*. Cambridge: Cambridge University Press, 2014.

31. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 1997; 30: 1145–1159.

32. Ling CX, Huang J and Zhang H. AUC: a better measure than accuracy in comparing learning algorithms. In: *Advances in artificial intelligence*, 2003, pp.329–341. Springer.

33. Powers DMW. The problem of area under the curve. In: *2012 international conference on information science and technology (ICIST)*, 2012, pp.567–573.