# Monitoring of lateral and cross level track geometry irregularities from onboard vehicle dynamics measurements using machine learning classification

De Rosa A., Kulkarni R., Qazizadeh A., Berg M., Di Gialleonardo E., Facchinetti A., Bruni S.

# Monitoring of lateral and cross level track geometry irregularities from onboard vehicle dynamics measurements using machine learning classification

De Rosa A. [a]*, Kulkarni R. [b]*, Qazizadeh A. [b], Berg M. [b], Di Gialleonardo E. [a], Facchinetti A. [a], Bruni S. [a]

[a] *Politecnico di Milano, Dep. of Mechanical Engineering, Milan, Italy*

[b] *KTH Royal Institute of Technology, Dep. of Aeronautical and Vehicle Engineering, Stockholm, Sweden*

[a]* anna.derosa@polimi.it , [b]* rohank@kth.se

In recent years, significant research efforts have focused on monitoring of track geometry irregularities from measurements of vehicle dynamics acquired onboard. Most of them analyse the vertical irregularity and the vertical vehicle dynamics since the lateral direction is much more challenging due to the non-linearities caused by the contact between the wheels and the rails. In the present work, a machine learning based fault classifier for condition monitoring of the track irregularity in the lateral direction is proposed. The classifiers are trained with a dataset composed of numerical simulation results and validated with a dataset of measurements acquired by a diagnostic vehicle on straight track sections of a high speed line (300 km/h). Classifiers based on Decision Tree (DT), Linear and Gaussian Support Vector Machine (SVM) algorithms are developed and compared in terms of performance: good results are achieved with the three algorithms, especially with the Gaussian SVM. Even though classifiers are data driven, they retain the essence of lateral dynamics.

Keywords: High-speed Railway, Lateral Dynamics, On-board Diagnostics, Fault Classifier, Decision Tree (DT), Support Vector Machine (SVM)

## Introduction

Condition monitoring of track irregularities is important to assure safety and proper performance of railway network systems. Nowadays, most railway infrastructure

managers carry out the track monitoring by measuring the geometry irregularities with sophisticated devices installed on dedicated diagnostic railway vehicles [1]. Although this method provides accurate measurements of the track irregularities, it is a highly expensive and time consuming process. For this reason, in recent years, many researches were focused on the development of new methods for track condition monitoring using in-service vehicles dynamic measurements [2]. These methods can be categorised into model-based and signal-based.

Model-based approaches analyse the mathematical relationship between the input and the output of the dynamic system. Most of them are focussed on the estimation of the vertical irregularity, while a limited number of published articles can be found related to the identification of the lateral alignment. In [3] a frequency decomposition of the vertical track irregularity and of the vehicle response is performed. Then the vertical irregularity, considered as the input of the system, is identified from the output: the vehicle accelerations (carbody, bogie frame and axle-box), computing the pseudo inversion of the frequency response function matrix. In [4], a Kalman filter is proposed to solve the inverse problem: the vertical track irregularity is estimated from the carbody accelerations. Similar approaches based on pseudo inversion of the frequency response function matrix, Kalman filter and unknown input observer have been applied in [5] and [6] to estimate the lateral alignment, considering mainly numerical simulations.

Signal-based methods are aimed to develop signal processing and statistical analyses on the output data to draw conclusions on the input data, when their mathematical relationship is difficult to be formulated. Many works can be found on the monitoring of vertical track irregularity through measurement of bogie and carbody accelerations with inertial sensors and applying double integration and high-pass filters [7–10]. The main

problem of these methods is that the data have to be processed in a proper way to avoid errors in the integration. A stochastic method has been proposed in [11] and [12], in which the vertical track irregularity is found with a cross-entropy optimisation applying Monte Carlo simulations and identifying the track geometry profile that generates a response that best fits the measured one. The results are acceptable, but the method requires significant computational effort.

In recent years, signal-based approaches have incorporated machine learning (ML) based fault classifiers which can be beneficial for monitoring the condition of complex mechanical systems such as rail vehicle and infrastructure. ML classification algorithms are very effective fault classifiers, and, in the railway field, they are mainly used to detect defective conditions of railway vehicles. In [13], Random Decrement Technique (RDT) was applied to extract the signal-based features from the lateral bogie accelerations, which were then fed into k Nearest Neighbour (kNN) and Artificial Neural Network (ANN) fault classifiers to diagnose the reason behind the observed vehicle running instability. In [14] two methodologies were proposed to detect wheel defects such as wheel flat and out of roundness using Support Vector Machine (SVM) classifier and custom ANN. Both methods have outperformed conventional wheel defect detection algorithms. In [15], a predictive maintenance strategy is proposed based on Decision Tree (DT) classifier for railway switches. In another study, researchers from the German Aerospace Centre (DLR) presented a hybrid method which combines a model-based estimation method with data driven techniques [16]. In the proposed method, measured vehicle response is used for estimation of vertical track irregularity which was then classified using an SVM based fault classifier. However, this method is applied only to simulated vertical track irregularities and the applicability of a hybrid approach to monitor the lateral and cross level irregularities has not been explored.

In recent years, researchers across academia and industry have demonstrated the potential of machine learning based fault classifiers to monitor the condition of systems and components of rail vehicle and infrastructure. However, there has not been enough emphasis on the development of machine learning based fault classifier for monitoring of lateral and cross level track irregularities. The machine learning based classifiers show great potential especially for monitoring of lateral and cross level track irregularities. In fact, model-based estimation techniques are not accurate enough to address the nonlinearities present at the wheel-rail interfaces and uncertainties associated with parameters of vehicle components [17]. The machine learning based classifiers have capabilities to overcome the above-mentioned challenges and can provide a reliable solution for monitoring of lateral and cross level track irregularities. In the present work, machine learning based fault classifiers are proposed to monitor the lateral and cross level track irregularities in the D1 (3 – 25 m) wavelength range. The fault classifiers are trained to monitor these track irregularities based on standard deviation (SD) of measured lateral and roll bogie frame accelerations of a vehicle running on straight track sections of a high speed line. Some numerical simulations are performed for higher level track irregularities to enlarge the parameter space and to train the fault classifier to identify the irregularities which are safety critical and for which measurements are not allowed. Finally, the DT, Linear and Gaussian SVM classifiers are validated with data acquired by field measurements. The paper is organized as follows: in section 2, the relevant theoretical background of rail vehicle dynamics and machine learning is presented followed by preliminary analysis of the dynamic behaviour of the system in the lateral plane. In section 3, the methodology adopted to conduct this research work is described. The results for the three classifiers are given in section 4 which is followed by the conclusions of the work.

**Background**

*Track geometry*

According to EN 13848-1:2019 standard [18], track irregularity is the deviation of the track geometry profile from the ideal position. Track parameters are described in a cartesian coordinate system centred to the track with clockwise rotation. The x-axis is the running direction, y-axis is the axis parallel to the running surface, and z-axis is perpendicular to the running surface and pointing downwards. In this coordinate system, three track irregularities can be defined, as shown in Figure 1. Longitudinal level or vertical irregularity is the deviation in z-direction of consecutive running table levels on any rail. Lateral alignment or lateral irregularity is the deviation in y-direction of consecutive position of point P on any rail. Cross level irregularity is the difference in height of the adjacent running table computed from the angle between the running surface and a horizontal reference plane.
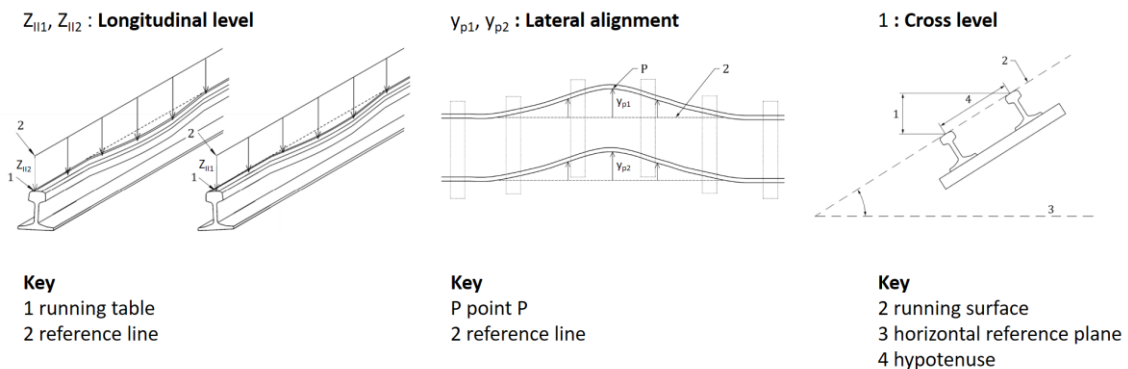


Figure 1: Track irregularities [18]

Three wavelength ranges can be considered for vertical and lateral irregularities: D1, D2, and D3 as defined in Table 1 [18].

Table 1: Wavelength ranges of the track irregularities [18]

| D1 | $3\ m < \lambda \leq 25\ m$ | |
|---|---|---|
| D2 | $25\ m < \lambda \leq 70\ m$ | |
| D3 | $70\ m < \lambda \leq 150\ m$ | vertical |
| | $70\ m < \lambda \leq 200\ m$ | lateral |

Special attention should be paid to the wavelength range D1: irregularities in this range could affect the running safety, whereas irregularities in range D2 and D3 are mainly linked with ride comfort. According to the UIC 518:2009 [19] and EN14363:2005 [20] standards, three quality levels are defined for track geometry irregularities:

- QN1: requires keeping a track section under observation or taking maintenance measures within the frame of normal operations scheduling.

- QN2: requires taking short-term maintenance measures.

- QN3: leads to the track section being excluded from the analysis, in case of assessment tests, because the track geometry quality encountered is not representative of usual quality standards. At this stage, this value is not as poor as the value reflecting the most unfavourable maintenance condition but remains acceptable.

The limits are given as a function of vehicle speed and they are normally based on SD over a defined length, 100 m is considered in the present work. SD is one of the most commonly used indexes for track irregularities of the European railway network. It represents the dispersion of the signal in a specific track section. The standards [19, 20] define the threshold levels for SD of longitudinal level and lateral alignment, but not for the cross level irregularity. Therefore, the cross level limits are chosen equal to the most

restrictive ones (lateral alignment). The threshold levels for SD of longitudinal level, lateral alignment and cross level are summarised in Table 2.

Table 2: Track irregularity quality levels limits set on SD

| 250 ≤ V ≤ 300 km/h | |
|---|---|
| QN1 | $SD_{lateral\ alignment} \leq 0.7$ mm <br><br> $SD_{longitudinal\ level} \leq 1.0$ mm <br><br> $SD_{cross\ level} \leq 0.7$ mm |
| QN2 | $0.7$ mm $< SD_{lateral\ alignment} \leq 1.0$ mm <br><br> $1.0$ mm $< SD_{longitudinal\ level} \leq 1.3$ mm <br><br> $0.7$ mm $< SD_{cross\ level} \leq 1.0$ mm |
| QN3 | $1.0$ mm $< SD_{lateral\ alignment} \leq 1.3$ mm <br><br> $1.3$ mm $< SD_{longitudinal\ level} \leq 1.7$ mm <br><br> $1.0$ mm $< SD_{cross\ level} \leq 1.3$ mm |

*Acceleration measurements*

Track geometry irregularity data used in the present work are acquired by the High Speed Diagnostic vehicle ETR500 Dia.Man.Te [21]. The measurements are performed with laser-optical and inertial systems installed onboard. The optical measurements are very accurate and reliable when the environmental conditions are favourable. However, this diagnostic technique has some drawbacks: the track recording vehicles are very expensive to be built, to run and to be maintained. The monitoring procedure is performed at defined time intervals and the distance in time between two consecutive recordings of track geometry, on the same line section, is often not optimal for diagnostics. Moreover, the inspection with diagnostic vehicles may interfere with the rail transport service. For these reasons, the present work is aimed to develop an

alternative track condition monitoring methodology based on acceleration measurements. The diagnostic train is also provided with accelerometers to acquire the vertical and lateral vehicle dynamic response. The accelerometers are located on the carbody, the bogie frames and the wheelsets. Likewise, accelerometers can be easily installed on in-service vehicles of the entire fleet allowing continuous monitoring of track irregularities from onboard measurements [22]. In case of abnormal vehicle response, root cause can be identified by comparing the acceleration measurements acquired on different trains of the fleet. In the present work, the bogie frame accelerations are used to predict the track condition in the range D1. The choice of using acceleration acquired on the bogie frames instead of on the wheelsets is made according to the sensors already installed on the in-service vehicles and used, for example, for assessing the running instability. The accelerometers installed on the wheelsets are also suitable to analyse the wavelength range D1, but they provide the same information of the bogie measurements since the primary suspension does not influence significantly the frequency range of interest. The track irregularities considered are acquired on an Italian High-Speed Line where the operating speed is 300 km/h and the diagnostic vehicle measures the track parameters at the operating speed.

*Machine learning based classifiers*

Supervised Machine Learning (ML) is the task of learning a function that maps an input to an output based on labelled input-output pairs provided in training phase [23]. In other words, the objective of the supervised machine learning algorithm is to tune the hyperparameters of the compact predefined algorithm based on the distribution of class labels with respect to the predictors. The trained model is then used to predict class labels of the test observations where the values of the predictors are known, but the value of the class label is unknown. ML algorithms are powerful in capturing the

knowledge in data and finding hidden patterns which facilitate efficient and accurate decision-making. Considering that track geometry irregularities for a specific section are either within or outside an allowed threshold, the resulting classification problem is a well posed binary classification problem, the intended outputs being discrete as opposed to being continuous or quantitative.

The two most common binary classification algorithms are DT and SVM.

*Decision Tree Classifier*

DT consists of trees which classify observations by sorting them based on the predictors' value. Each node in DT represents a predictor of an observation to be classified and each branch represents a value that the node can assume. The training phase of DT can be visualised as dividing and conquering the predictors' space by employing a recursive partitioning. In the present work, DT is trained with the Classification And Regression Tree (CART) algorithm [24]. The CART algorithm develops binary tree by formulating the optimal strategy to divide predictor space into several orthogonal subspaces with objective of maximizing the classification accuracy. DT has been successful in many problems because of easy interpretability and representation. In fact, DT more closely mirrors the human decision-making process than all other classification algorithms. However, DT is not robust and is sensitive to changes in training data: a small change in the training data can change the DT drastically. Generally, DT doesn't have good predictive accuracy as other classification algorithms, but still DT is a popular choice within the ML community.

*Support Vector Machine Classifier*

The most popular supervised ML technique is the SVM [25]. The fundamental concept of SVM is the generation of 'margin' on both sides of a hyperplane that separates two

data classes. In the training phase the algorithm tunes hyperparameters to maximize the margin, which then creates the margin with largest possible distance between the separating hyperplane and the instances. It has been proven to reduce an upper bound on the expected generalisation error [23]. Most problems in the real world are non-separable; there does not exist the optimal separating hyperplane. In this situation, the training observations are transformed into a higher dimension design space by applying a kernel trick with the aim of converting non-separable data into linearly separable data. The same kernel function is applied for converting the hyperplane in a higher dimension space into a predictor space to generate a decision boundary. There are several kernel functions available to fulfil the objective and the selection of the kernel function is very critical to obtain optimal classification accuracy. However, there are no guidelines to select the best suitable kernel function for a specific application but it is common practice to estimate a range of potential settings and use cross validation over the training set to find the best one [26]. In this work, Linear and Gaussian kernel functions are used to train different SVM classifiers.

Generally, SVM classifiers are better options for continuous and multidimensional feature datasets than DT ones. The kernel trick makes SVM a unique classification technique which gives the possibility of transformation to higher dimensions. The SVM classifiers are more robust and not very sensitive to changes in the training datasets. On the other hand, SVM classifiers are very difficult to interpret and the handling of the model parameters is difficult. SVM is also comparatively computationally expensive and slower compared to other classifiers in the training phase. However, SVM classifiers are fast in the testing phase.

*Preliminary numerical simulations*

To better understand the vehicle-track dynamic behaviour in the lateral plane and

analyse whether ML techniques could provide good results, numerical simulations have been carried out.

The simulations are performed using the software AdTreS [27,28]. This software has been developed at Dipartimento di Meccanica of Politecnico di Milano. The software considers a three-dimensional multibody model (35 degrees of freedom) of the ETR500 rail vehicle that fully accounts for the non-linear effects of the wheel-rail contact and the non-linear behaviour of suspension components (i.e. bumpstops and dampers). The model has been verified against measurements as explained in [29] . Since the interest of this work is focused only on low frequencies, flexibility of the vehicle components and flexibility of the infrastructure are neglected [30].

The simulations have been performed considering straight track, vehicle speed equal to 300 km/h, and three different cases of track irregularities (Table 3). The vertical irregularity has been set equal to zero, whereas the lateral and the cross level irregularities have been considered separately (1st and 2nd cases) and simultaneously (3rd case). For each case, 10 different track irregularity signals have been simulated. Figure 2 shows the average behaviour of the lateral bogie accelerations obtained. These simulations have shown the influence of the cross level irregularity on the response of the bogie in the lateral direction, pointing out that the lateral and roll motions are not independent, considering the ETR500 vehicle running at 300 km/h on straight track sections.

Table 3: Track irregularities considered in the preliminary numerical simulations

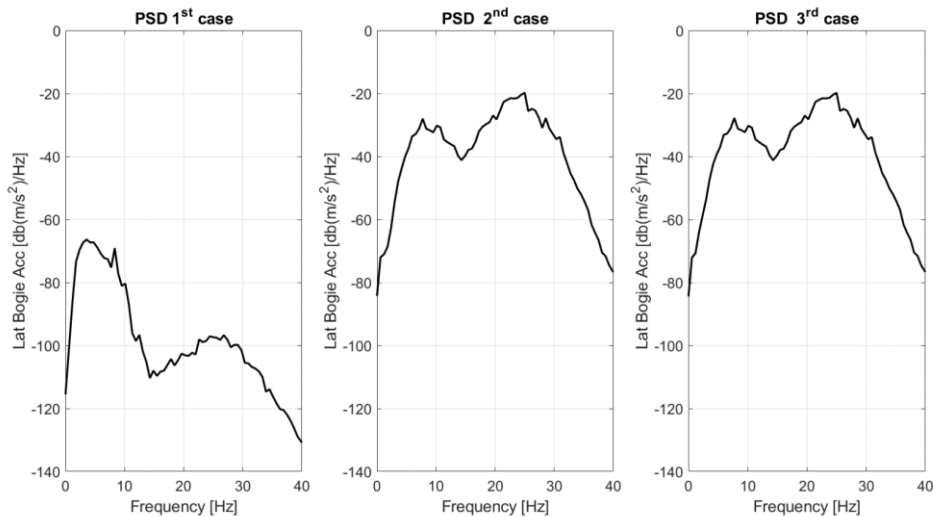|  | 1st case | 2nd case | 3rd case |
|---|---|---|---|
| Vertical irregularity | = 0 | = 0 | = 0 |
| Cross Level Irregularity | = 0 | ≠ 0 | ≠ 0 |
| Lateral Irregularity | ≠ 0 | = 0 | ≠ 0 |

Figure 2: PSD of lateral bogie frame accelerations obtained with numerical simulations (see Table 3)

Observing the vehicle response, it can be noticed that the bogie accelerations in the 2nd case with only cross level irregularity are very similar to the 3rd case with both irregularities, meaning that the effect of lateral track irregularity is significantly lower in respect to the cross level one. The cause of this behaviour is in the lateral dynamics of the system: with respect to the vertical direction, where the geometrical contact takes place, the interaction in the lateral direction is due through contact forces so the vehicle, in general, does not follow directly the alignment irregularity. For this reason, the ML classifier should consider both lateral and cross level irregularities. The focus of this application is on the lateral and cross level track irregularities with the objective of investigating the relationship between track irregularities and bogie frame lateral and roll accelerations. The aim is to define fault classifiers based on bogie frame accelerations for condition-based track maintenance.

**Methodology**

The decisions on track maintenance intervention are generally based on track

measurements performed by diagnostic vehicle and not on the basis of the dynamic response of the vehicle. The current decision-making approach is a preventive maintenance approach which is very expensive for the railway infrastructure managers. There is a need to shift from an irregularity-based threshold to a more economical approach such as a vehicle response-based threshold. Unfortunately, the relationship between the threshold on track irregularity and the recorded vehicle response has not been established yet, especially when alignment irregularity is considered.

In the present work, ML classifiers are applied to monitor the condition of irregularities. The objective is to detect the limit exceedances of the SD of the track irregularities in the D1 wavelength range from SD of the bogie acceleration measurements performed onboard the vehicle. The application is currently focused on straight track sections with no singular point that could cause a sudden change in the track stiffness, such as bridge, tunnel, viaduct, switch, etc. Moreover, only measurements performed at the operating speed of 300 km/h are considered.

In the vertical plane, the correlation between track irregularity and vehicle response is essentially linear and can be described with a simple linear regression [31,32]. In the lateral plane, the same approach is not suitable, due to the non-linear effects at the contact points between the wheels and the rails. So, in this work, ML-based classification algorithms are presented to analyse the relationship between track irregularities and vehicle response with the ultimate objective of defining the vehicle response-based threshold.

### *Configuration of databases and classification*

The configuration chosen is a binary classification: the Class1 is related to small track irregularities, in terms of cross level and alignment, that corresponds to track conditions not requiring maintenance intervention; the Class2 includes cases in which one or both

the present irregularities are higher than a set threshold level (QN1: 0.7 mm, see Table 2) and requires scheduling of maintenance interventions. The binary partition of parameter space is shown in Figure 3, where SD of lateral irregularity is on the x-axis and SD of cross level irregularity on the y-axis.
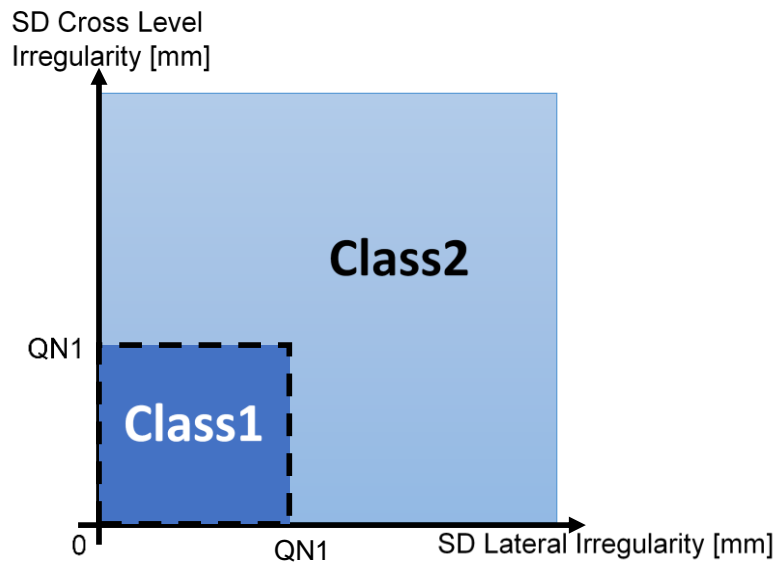


Figure 3: Track irregularity labelling strategy

The track irregularity database consists of 100 m long track sections of measured and simulated track irregularities. The measured irregularities are collected from 10 runs of the diagnostic vehicle, selecting only straight track sections where the vehicle speed is 300 km/h ± 2%. Track sections with SD higher than 1.5 mm for lateral and cross level, and 2 mm for vertical, which are greater than QN3 level, are considered as outliers and removed from the database.

The measured irregularities have low SD: most of the track sections (87.8%) are lower than QN1. This poses a risk of obtaining a highly biased classifier if trained with only measured track sections. The risk is mitigated by performing numerical simulations with higher SD of track irregularities. This approach permits to spread and extend the track irregularity database to higher level. Track sections of the database are labelled

according to the strategy presented in Figure 3. The scatter plot of the track irregularity database is presented in Figure 4 where SD of lateral and cross level track irregularities are on x and y-axis respectively. The track irregularity database contains 518 measured sections (455 in Class1 and 63 in Class2); and 1296 simulated sections (745 in Class1 and 551 in Class2).
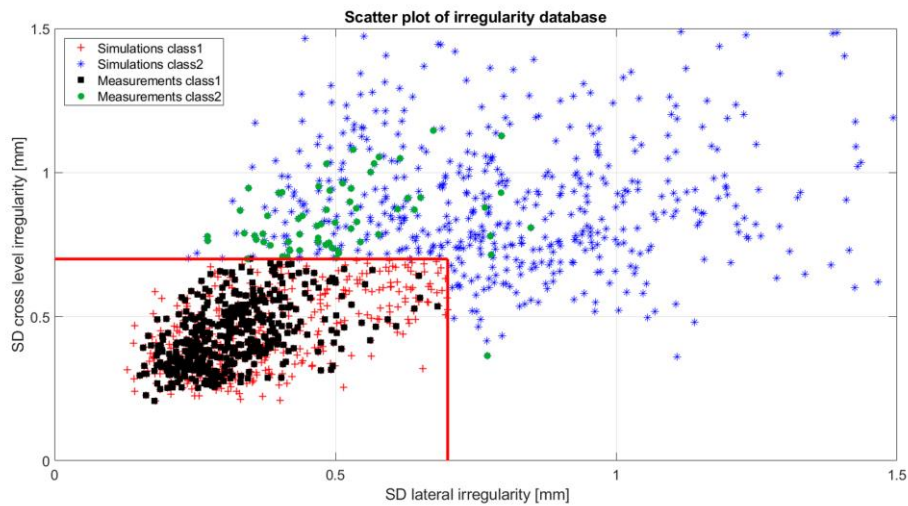


Figure 4: Scatter plot of track irregularity database

The overall database of vehicle dynamic response is composed by simulated (used for the training phase) and measured (used for the test) lateral and roll bogie frame accelerations of the vehicle running on the track irregularities of Figure 4. The vehicle accelerations are then band-pass filtered in 3 – 27 Hz to comply with the D1 range for the present speed. The scatter plot of the measured and simulated vehicle response is presented Figure 5.
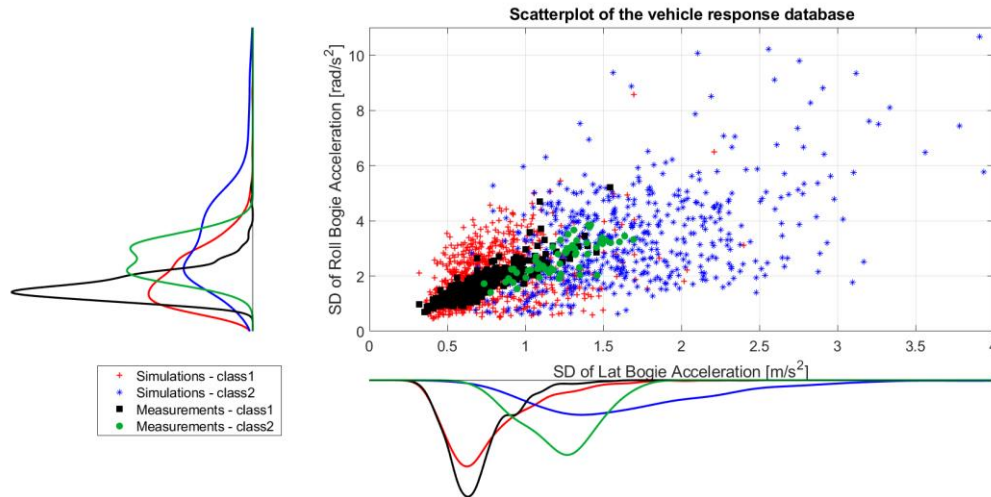
Figure 5: Scatter plot of the vehicle response database

Figure 5 represents the SD of the lateral bogie frame accelerations on the x-axis and the SD of the roll bogie frame accelerations on the y-axis. The accelerations obtained in the simulation performed considering a class1 track irregularities are represented with (red) crosses; the acceleration obtained considering class2 irregularities are represented with (blue) stars. Analogously, the acceleration acquired on field are represented with (black) squares when the irregularities are in class1 and with (green) dots for class2. The simulated and measured accelerations are also compared in terms of probability distributions of Class1 and Class2. The measured and simulated lateral accelerations show a very similar distribution in Class1, while in Class2 the simulations have a wider dispersion, as intended. The distributions of the simulated roll acceleration have a greater variance in respect to the measured ones, but they have a similar mean value. The ML model receives as input the SD of the lateral and roll bogie frame accelerations and gives as output the classification of the SD of the lateral and cross level track irregularities. So, the accelerations are selected as predictors whereas the track irregularities are outputs. The datasets are labelled based on the SD magnitude of the

track irregularities. The labelled dataset enables the application of supervised learning algorithms. In the context of selecting the best algorithm, the theorem of 'no free lunch' is popular in the ML community. The theorem states that there is no single algorithm that performs best for all problems and all datasets [26]. This means that an algorithm which performs best for a specific problem might not be the best for others. It is very common practice to try out many applicable algorithms and then shortlist the top performing algorithms for detailed investigation.  The method adopted in the current work is explained henceforth.

In the training phase, classification algorithms are trained on dataset composed of numerical simulations by adopting the 5 fold Stratified Cross Validation (SCV) process [33]. The final performance of each classifier is evaluated with test dataset in accordance with Hold-out Test (HT) procedure. In this work, the trained classifiers are tested with dataset composed of measurements. In the first step, many algorithms were trained and tested. The DT, the Linear SVM and the Gaussian SVM are the top performing algorithms among all. The hyperparameters of each algorithm are optimized by Bayesian optimization algorithm for minimization of SCV error. The optimal choice of hyperparameters is crucial to avoid overfitting of classifiers

*Evaluation Approaches*

The performance of the classification models in the SVC and HT phase is assessed with confusion matrix, performance metrics and reliability plot.

- Confusion Matrix

   In the classification problem, a Confusion Matrix (CM) is used to analyse the predicted category labels of the data against the true category labels which are already known from the laser-optical measurements acquired by the diagnostic

vehicle. The CM for a binary classification model is shown in Table 4: the True Class1 (TC1) are instances belonging to Class1 which are correctly predicted as Class1. Analogously, the True Class2 (TC2) are instances belonging to Class2 which are correctly predicted as Class2. The False Class1 (FC1) are instances belonging to Class2 which are incorrectly predicted as Class1. Vice versa, the False Class2 (FC2) are instances belonging to Class1 which are incorrectly predicted as Class2. The CM is presented in the percentages form in which correct/incorrect observation classification rate is evaluated by considering the correct/incorrect predictions of the specific class over the total instances belonging to the same class.

Table 4: CM for a binary classification problem.

| | | Predicted Class | |
|---|---|---|---|
| | | Class1 | Class2 |
| **True Class** | Class1 | True Class1 (TC1) | False Class2 (FC2) |
| | Class2 | False Class1 (FC1) | True Class2 (TC2) |

- Performance metrics

The CM is used to compute several performance measures to compare the performance of trained classifiers on test dataset [34]. Performances can be interpreted from operational or economical conservative perspectives. The classifier trained for operational conservativeness should have high performance for correctly predicting TC2 observations. On the other hand, the classifier trained for economical conservativeness should not misclassify observations belonging to Class1, in order to avoid unnecessary interventions. In the present work, the

performances are evaluated with focus on safety. Thus, the performance metrics introduced in the following are defined on Class2.

(1) **Accuracy** is a measure of correct predictions over all instances of the dataset for the classification model [34] and it is calculated with the following formulation:

$$Accuracy = \frac{TC1 + TC2}{TC1 + FC1 + FC2 + TC2} \tag{1}$$

(2) **Precision** is a measure of correctness of Class2 predictions among all Class2 predictions of the model over dataset [34] and it is calculated as:

$$Precision = \frac{TC2}{TC2 + FC2} \tag{2}$$

(3) **Recall** measures the ability of the model to accurately predict Class2 instances among all Class2 instances of the dataset [34] and it is calculated as:

$$Recall = \frac{TC2}{TC2 + FC1} \tag{3}$$

(4) **F$_1$ score** is the harmonic mean of precision and recall measures given by Equations (2) and (3); it can be calculated as:

$$F_1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

(5) **Kappa** measures an inter-rater agreement between qualitative items, which assess the relative observed agreement ($p_o$) with the hypothetical probability of chance agreement ($p_e$) [35]. In simple terms, the kappa is a measure of how well a classifier performed ($p_o$) as compared to how well it could have performed by random chance ($p_e$). In the case of imbalanced test dataset, the kappa is a more

robust performance measure in comparison to F1 score and accuracy. It is

calculated as:

$$kappa = 1 - \frac{1 - p_0}{1 - p_e}$$

(5)

**Results of ML classification**

The ML-based classifiers are developed using MATLAB and three independent

classifiers based on DT, Linear SVM and Gaussian SVM are trained with 5-

fold Stratified Cross Validation (SCV) method. The hyperparameters of each model are

optimised with Bayesian optimization approach to minimize the SCV error.

It is important to note that the three classification models are independently trained on

simulation results only, with the purpose of training the classifiers to identify instances

which may not be available from field measurements. The performance evaluation of

classifiers is made in two steps: the performance is evaluated with the SCV approach in

the training phase and classifiers are tested on the field measurements in accordance

with the Hold-out Test (HT) in the testing phase. The results of these two steps are

presented in the following subsections which show that the developed fault classifiers

are able to classify track irregularities based on vehicle response.  The results section is

concluded with an interpretation of fault classification boundaries in view of the

underlying vehicle-track dynamic interaction.

*Training of Classification Models*

During the training phase, the classification models generates a decision

boundary function  which separates Class1 and Class2 observations based on their

respective vehicle response parameters (roll and lateral accelerations of the bogie

frame). The decision boundary along with all observations of the training dataset is

mapped on a plot between the roll and the lateral accelerations. The observations on the left side of the decision boundary are identified as Class1 (no need of maintenance) and observations on the right side are assigned to Class2 (need of maintenance). In Figure 6 , the decision boundaries of DT, Linear SVM and Gaussian SVM are represented for the training dataset, where x-axis is SD of lateral acceleration and y-axis is SD of roll acceleration of the bogie. Observations belonging to Class1 and Class2 are shown in (red) squares and (blue) dots respectively.
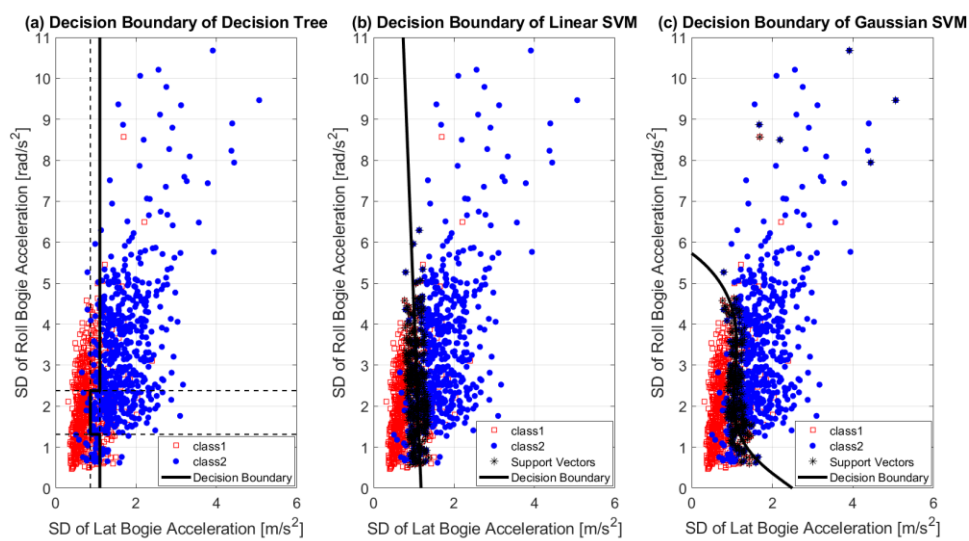


Figure 6: Graphical representation of the decision boundary of trained classifiers

The DT algorithm divides the parameter space with orthogonal boundaries parallel to either x-axis or y-axis which subdivides the parameter space into several rectangles and each rectangle is a leaf node. The graphical representation of DT classifier is in Figure 6a, where the solid line is the decision boundary separating both classes, whereas dashed lines are secondary boundaries, which link different rectangles of the same class. The DT classifier designates observation as Class1 or Class2 based on if it is on the left or right side of the decision boundary. The DT correctly classifies 90.7% of

Class1 and 84.8% of Class2 observations, whereas 9.3% of Class1 and 15.2% of Class2

observations incorrectly (Figure 7a).

The SVM classifiers generate the decision boundary as a linear combination of x and y

coordinates of support vectors. Linear SVM classifier uses a linear kernel function

which results in a linear decision boundary, the decision boundary of the Linear SVM

is shown in Figure 6b along with the corresponding support vectors in (black) stars. The

Linear SVM correctly classifies 86.6% of Class1 and 86.8% of Class2 observations,

whereas 13.4% of Class1 and 13.2% of Class2 observations incorrectly, as shown in

confusion matrix in Figure 7b. Gaussian SVM classifier is trained with Gaussian kernel

function which transforms observations into higher dimensional space and generally it

results in a nonlinear decision boundary. The decision boundary of Gaussian SVM

classifier is shown in Figure 6c. Gaussian SVM classifies 88.2% of Class1 and 85.8%

of Class2 observations correctly, whereas 11.8% of Class1 and 14.2% of Class2

observations incorrectly. These results are summarised in Figure 7c.
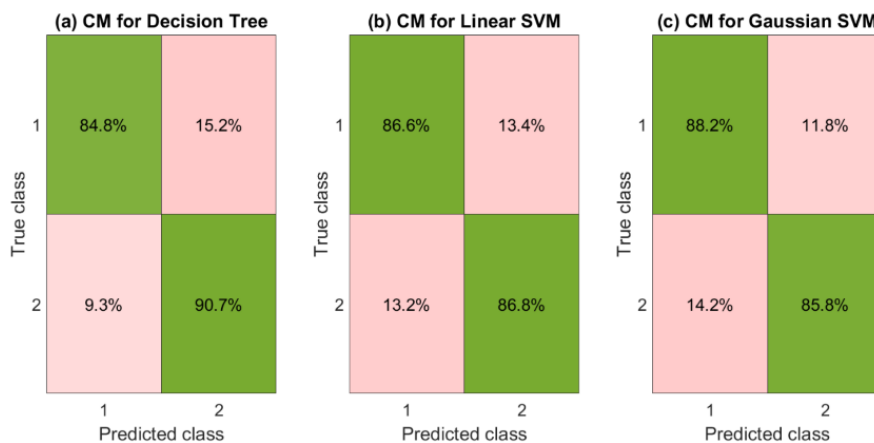


Figure 7: CM of trained classifiers in the training phase

In Table 5, the performance of the three classifiers in the training phase is compared

with each other and it is evident that the performance of the three algorithms is similar

to each other. The higher accuracy is for the DT, equal to 87.3%, while the Gaussian and Linear SVM are slightly less accurate with rates 87.2% and 86.7% respectively. The precision and recall rates related to DT are 81.6% and 90.7% respectively. Precision and recall are evaluated also for Linear SVM: 82.7% and 86.8%, which shows a better balance between the two percentages compared to DT. However, the best trade-off between precision and recall is achieved with Gaussian SVM: 84.3% and 85.8%. The $F_1$ score, which is a harmonic mean of precision and recall rate, shows the same trend as accuracy: DT presented the higher value, closely followed by Gaussian and Linear SVMs. The performance similarity of the three classifiers is confirmed evaluating kappa score. Kappa scores are 0.74, 0.73 and 0.74 for DT, Linear SVM and Gaussian SVM respectively. These percentages confidently prove that each classifier is trained to produce the best possible classification performance on the given training data.

Table 5: Summary of classifier performances in the training phase

|  | Accuracy | Precision | Recall | $F_1$ Score | kappa |
|---|---|---|---|---|---|
| **Decision Tree** | 87.3 % | 81.6 % | 90.7 % | 85.9 % | 0.74 |
| **Linear SVM** | 86.7 % | 82.7 % | 86.8 % | 84.7 % | 0.73 |
| **Gaussian SVM** | 87.2 % | 84.3 % | 85.8 % | 85.1 % | 0.74 |

*Testing of Classification Models*

Figure 8 and Figure 9 show the performances of the classifiers in the testing phase, using measured accelerations. DT correctly predicts 87.0% of Class1 and 92.1% of

Class2 observations, as shown in Figure 8a, in which most of the Class2 observations are on the right side of the decision boundary.
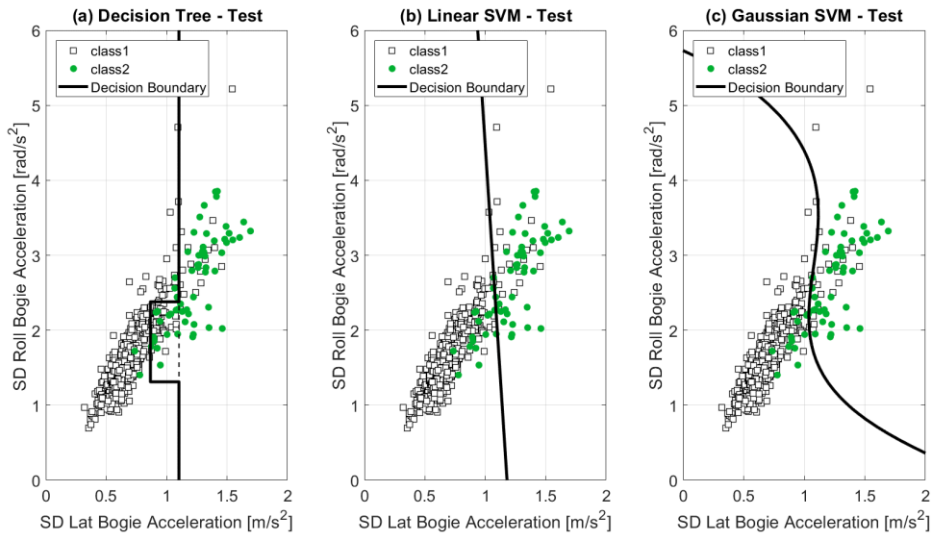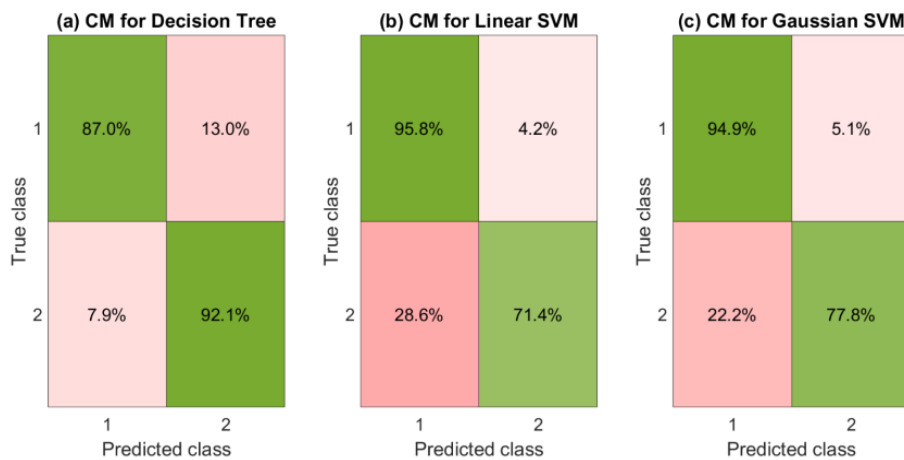


Figure 8: Testing of classifiers with test data



Figure 9: CM of trained classifiers in the testing phase

The Linear and Gaussian SVM predicts Class1 observations more correctly than DT with prediction rate of 95.8% and 94.9% respectively. These can be observed in Figure 8b and Figure 8c in which most of the Class1 observations are on the left side of the

decision boundary. However, the Linear and Gaussian SVM show only 71.4% and 77.8% correct prediction rate for Class2 which is lesser than DT.

DT has very high accuracy and recall rates related to Class2, that means that the algorithm is more conservative from running safety point of view. DT is better for correctly predicting TC2 as Class2. However, the precision is significantly lower and that means that a lot of points belonging to Class1 are predicted as Class2. This leads to schedule unnecessary maintenance intervention that increase the operation costs.

The performance of the three classifiers in the testing phase is summarised in Table 6.

Table 6: Summary of classifier performances in the testing phase

|  | Accuracy | Precision | Recall | F$_1$ Score | kappa |
|---|---|---|---|---|---|
| **Decision Tree** | 87.6 % | 49.6 % | 92.1 % | 64.4 % | 0.58 |
| **Linear SVM** | 92.9 % | 70.3 % | 71.4 % | 70.9 % | 0.67 |
| **Gaussian SVM** | 92.9 % | 68.1 % | 77.8 % | 72.6 % | 0.69 |

The SVMs are 5% more accurate and 20% more precise than DT, which means that SVMs are capable of predicting observations with better accuracy and precision. However, both the SVMs show less recall rate than DT, but F$_1$ score is better for both SVMs, and specifically, the Gaussian SVM has the highest F$_1$ score among all. The difference between the three classifiers is evident after considering the kappa score. DT's kappa score is 0.58 which is significantly lower compared to the values related to Linear and Gaussian SVM: 0.67 and 0.69. The higher score of kappa complies with other parameters and confirms that the Linear and Gaussian SVMs are better performing classification algorithms in the present application.
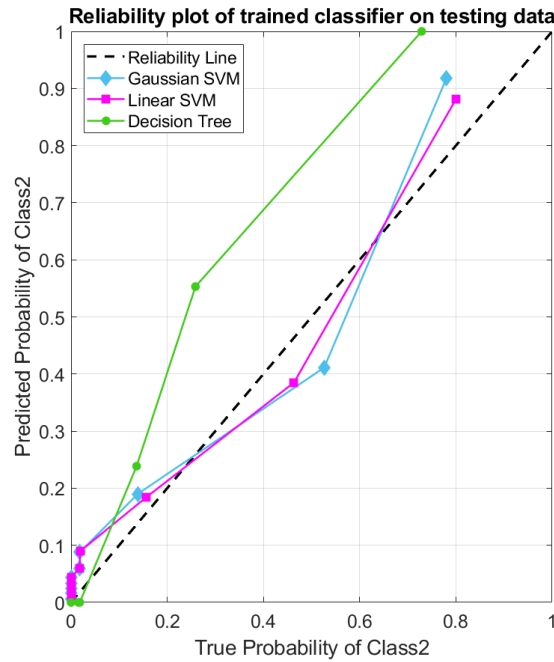
Figure 10: Reliability plot of the three classifiers in the testing phase

In Figure 10, the algorithms are compared in terms of reliability plots [36] against the

reliability line shown by the (black) dashed line. The probability of classification

of test observation as Class2 is plotted against the true probability of the Class2.

The reliability plot of the Linear and Gaussian SVMs displayed with diamond marked

(blue) and square marked (magenta) lines follows each other very closely, pointing out

the similarity in the performance of both classifiers. Most datapoints for these two

classifiers are at the left bottom corner which signifies

the reliability of the classifying true Class1 datapoints as Class1. The prediction

probability of Class2 increases uniformly as the true probability of Class2 increases and

reliability lines of both SVM classifiers follow the reliability line closely. This

behaviour shows that the Linear and Gaussian SVMs are unbiased classifiers as

observed in Table 6 that reports the similar recall and precision rates for SVM

classifiers. The reliability plot of the DT classifier represented with dot marked (green)

line, the prediction probability for test data as Class2 increases very rapidly as the true probability of Class2 increases, and the green line moves away from the reliability line. The green line lies above the reliability line which leads to the conclusion that the DT is biased towards Class2 and the same behaviour is reflected as a high recall rate and poor precision in Table 6. The parameters of the classifiers can be tuned to achieve 100 % of recall rate, however this would lead to a very low precision. Since the classification procedure can be performed recursively using the acceleration measurements acquired onboard the in-service vehicles, the undetected track sections will certainly be detected once track section crosses the classification threshold.

**Conclusions**

In this paper, machine learning based classifiers are proposed to monitor the amplitude of the lateral and cross level track irregularities from lateral and roll bogie frame accelerations. The dataset is composed with field measurements acquired onboard the Italian diagnostic vehicle Dia.Man.Te operating at 300 km/h, and by numerical simulations performed using a 3D multibody software developed at Dipartimento di Meccanica of Politecnico di Milano. The training phase is carried out using only simulation data, then the trained classifier is tested with the measurements. In this way, the classifier is trained with magnitudes of track irregularities and accelerations greater than the level found in the measurements. Therefore, the classifier is capable to detect the unusual condition of track irregularities.

Three different algorithms are applied: Decision Tree, Linear SVM and Gaussian SVM. DT is a simple algorithm which is easy to implement and interpret. On the other hand, it is not robust: a small variation in the training data can lead to a large change in the structure of the classifier. The accuracy and precision rates for DT obtained in testing phase are also lower than other classifiers. The Linear and Gaussian SVMs provide

similar results in terms of accuracy, precision and the other parameters. However, the Gaussian SVM has more degrees of freedom while defining the decision boundary function and can be considered more capable to formulate correct decision boundary in the dynamic perspective. The relative accuracy rate, with respect to maximum possible accuracy, achieved with Gaussian SVM classifier is equal to 0.69. The promising results highlight the potential of data driven classifiers to monitor the track irregularities based on dynamics in the lateral plane. However, future works could improve the results achieved. Moreover, the same application should be performed considering curve sections of the track, since in curve the relationship between track irregularities and vehicle accelerations, as well as between lateral and cross level irregularities may be different compared to straight track. Further investigations are therefore required to enhance the research results on condition monitoring of track irregularities using ML techniques.

**Appendix: List of Abbreviations**

| Abbreviation | Stands for | Defined in |
|---|---|---|
| ANN | Artificial Neural Network | Introduction |
| CART | Classification And Regression Tree | Background |
| CM | Confusion Matrix | Methodology |
| DT | Decision Tree | Abstract |
| FC1 | False Class1 | Methodology |
| FC2 | False Class2 | Methodology |
| HT | Hold-out Test | Methodology |
| kNN | k-Nearest Neighbour | Introduction |

| | | |
|---|---|---|
| ML | Machine Learning | Introduction |
| SCV | Stratified Cross Validation | Methodology |
| SD | Standard Deviation | Introduction |
| SVM | Support Vector Machine | Abstract |
| TC1 | True Class1 | Methodology |
| TC2 | True Class2 | Methodology |

## References

[1] Weston PF, Ling CS, Goodman CJ, et al. Monitoring lateral track irregularity from in-service railway vehicles. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. 2007;221:89–100.

[2] Weston PF, Roberts C, Yeo G, et al. Perspectives on railway track geometry condition monitoring from in-service railway vehicles. Veh. Syst. Dyn. 2015;53:1063–1091.

[3] Alfi S, Bruni S. Estimation of long wavelength track irregularities from on board measurement. 4th IET Int. Conf. Railw. Cond. Monit. (RCM 2008). Derby, UK: IEEE; 2008. p. 1–6.

[4] Odashima M, Azami S, Naganuma Y, et al. Track geometry estimation of a conventional railway from car-body acceleration measurement. Mech. Eng. J. 2017;4:1–12.

[5] De Rosa A, Alfi S, Bruni S. Estimation of lateral and cross alignment in a railway track based on vehicle dynamics measurements. Mech. Syst. Signal Process. 2019;116:606–623.

[6] Alfi S, De Rosa A, Bruni S. Estimation of lateral track irregularities from on-board measurement: effect of wheel-rail contact model. 7th IET Conf. Railw. Cond. Monit. 2016 (RCM 2016). Birmingham, UK: Institution of Engineering and Technology; 2017. p. 1–7.

[7] Weston PF, Ling CS, Roberts C, et al. Monitoring vertical track irregularity from

in-service railway vehicles. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. 2007;221:75–88.

[8]     Lee JS, Choi S, Kim SS, et al. A mixed filtering approach for track condition monitoring using accelerometers on the axle box and bogie. IEEE Trans. Instrum. Meas. 2012;61:749–758.

[9]     Li M, Persson I, Spännar J, et al. On the use of second-order derivatives of track irregularity for assessing vertical track geometry quality. Veh. Syst. Dyn. 2012;50:389–401.

[10]    Tsunashima H, Naganuma Y, Matsumoto A, et al. Condition monitoring of railway track using in-service vehicle. Reliab. Saf. Railw. InTech; 2012. p. 333–356.

[11]    OBrien EJ, Bowe C, Quirke P, et al. Determination of longitudinal profile of railway track using vehicle-based inertial readings. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. 2017;231:518–534.

[12]    OBrien EJ, Quirke P, Bowe C, et al. Determination of railway track longitudinal profile using measured inertial response of an in-service railway vehicle. Struct. Heal. Monit. 2018;17:1425–1440.

[13]    Gasparetto L, Alfi S, Bruni S. Data-driven condition-based monitoring of high-speed railway bogies. Int. J. Rail Transp. 2013;1:42–56.

[14]    Krummenacher G, Ong CS, Koller S, et al. Wheel defect detection with machine learning. IEEE Trans. Intell. Transp. Syst. 2018;19:1176–1187.

[15]    Allah Bukhsh Z, Saeed A, Stipanovic I, et al. Predictive maintenance using tree-based classification techniques: A case of railway switches. Transp. Res. Part C Emerg. Technol. 2019;101:35–54.

[16]    Schenkendorf R, Dutschk B. Improved Railway Track Irregularities Classification by a Model Inversion Approach. Third Eur. Conf. Progn. Heal. Manag. Soc. 2016. Bilbao, Spain; 2016. p. 358–365.

[17]    Li C, Luo S, Cole C, et al. An overview: modern techniques for railway vehicle on-board health monitoring systems. Veh. Syst. Dyn. 2017;55:1045–1070.

[18]    EN 13848-1:2019 Railway applications – Track – Track geometry quality – Part

1: Characterization of track geometry.

[19]     UIC 518 - Testing and approval of railway vehicles from the point of view pf their dynamic behaviour - safety - track fatigue - ride quality.

[20]     EN 14363:2005 Railway applications – Testing for the acceptance of running characteristics of railway vehicles – Testing of running behaviour.

[21]     Treno Diagnostico - Dia.Man.Te [Internet]. Milano, Italy; 2014. Available from: http://www.rfi.it/cms-file/allegati/rfi_2014/TrenoDiamente.pdf.

[22]     Diana G, Cangioli F, Carnevale M, et al. Condition monitoring of the rolling stock and infrastructure : results of a pilot project. :1–8.

[23]     Kotsiantis SB. Supervised machine learning: A review of classification techniques. Informatica. 2007;31:3--24.

[24]     Breiman L. Classification and regression trees. Chapman & Hall; 1993.

[25]     Cortes C, Vapnik V. Support-Vector networks. Mach. Learn. 1995;20:273–297.

[26]     Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd ed. Springer Ser. Stat. New York, NY: Springer New York; 2009.

[27]     Bruni S, Collina A, Diana G, et al. Lateral dynamics of a railway vehicle in tangent track and curve: Tests and simulation. 16th IAVSD Symp. "The Dyn. Veh. Roads Tracks", August 30, 1999 - Sept. 3, 1999. 2000;33:464–477.

[28]     Alfi S, Bruni S. Mathematical modelling of train-turnout interaction. Veh. Syst. Dyn. 2009;47:551–574.

[29]     Mazzola L, Alfi S, Bruni S. Evaluation of the hunting behaviour of a railway vehicle in a curve. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. 2015;229:530–541.

[30]     Di Gialleonardo E, Braghin F, Bruni S. The influence of track modelling options on the simulation of rail vehicle dynamics. J. Sound Vib. 2012;331:4246–4258.

[31]     Karis T, Berg M, Stichel S, et al. Correlation of track irregularities and vehicle responses based on measured data. Veh. Syst. Dyn. 2018;56:967–981.

[32]     Karis T, Berg M, Stichel S. Analysing the correlation between vehicle responses and track irregularities using dynamic simulations and measurements. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. 2019;1–13.

[33]    Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques. Data Min. Pract. Mach. Learn. Tools Tech. Elsevier; 2016.

[34]    Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 2009;45:427–437.

[35]    Cohen J. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 1960;20:37–46.

[36]    Vuk M, Curk T. ROC curve , Lift chart and Calibration plot. Metod. Zv. 2006;3:89–108.