

AUTOMATIC PREDICTION OF MAINTENANCE INTERVENTION TYPES IN ROADS USING MACHINE LEARNING AND HISTORICAL RECORDS

By

F.J. Morales¹, PhD, Research Associate
A. Reyes², PhD, Research Associate
N. Caceres³, PhD, Research Associate
L. Romero⁴, PhD, Assistant Professor
F.G. Benitez⁵, Professor (*)

Transportation Engineering, Faculty of Engineering
University of Seville
Camino de los Descubrimientos, s/n, Seville 41092, Spain
Tel.: +34 954 487315

Email: fmorales2@us.es¹, areyes@us.es², ncaceres@us.es³, lromero1@us.es⁴, benitez@us.es⁵

(*) Corresponding Author

Word count: 6219 + 5 Figures = 7469

Manuscript submitted to TRB, 1st August 2017

ABSTRACT

A methodology to support and automate the prediction of maintenance intervention alerts in transport linear infrastructures is a very useful tool for maintenance planning and managing. This piece of work goes along this track combining the current and predicted state condition of the assets, unit components of the infrastructure, with operational and historical maintenance data, to derive information about the needed maintenance operations to avoid later severe degradation. By means of data analytics and machine learning techniques, the proposed methodology generates a prioritized listing, ranked on severity levels, corresponding to the pre-alerts and alerts generated by all assets of the transport infrastructure. The methodology is applied and tested to a real case consisting of a road network with different section classes. The analysis of the results shows that the algorithms and tools developed have good predicting capabilities.

KEYWORDS: roads; predictive maintenance; machine learning

INTRODUCTION

Research on linear-asset management has attracted considerable attention in the last two decades; seminal publications have focused on specific types of linear-assets, with important advances achieved both in railways (1,2) and roads (3). Earlier contributions are also reported on other infrastructures such as cables (4) and pipelines (5,6,7,8). Currently, several either specific commercial codes (9,10,11) or adapted general purpose codes (12,13) are on market to address linear-asset information management issues; however, they have neither predictive or very basic decision analysis capabilities on basic stages; though these are strategic objectives for most of them.

It is worth remarking that most transport infrastructure Administrative Bodies in developed countries are aware of the importance of pushing forward reliable predicting tools in order to maximise the availability of theses constructions and optimise the resources devoted to maintaining. The Federal Highway Administration has demonstrated that an accurate knowledge of asset condition and a proper management and analysis of gathered data could lead to a more effective maintenance planning able to extend the lifecycle of assets (e.g. pavements) from 3 to 10 years, which means a range between 12% and 20% (14,15). In the case of railway sector the life span of ageing infrastructures can also be extended up to 25% by asset condition monitoring, prediction and decision-making support (16).

The triggering of maintenance alerts, regarding the state condition of transport infrastructure assets, has been customarily based on surpassing deterministic fixed thresholds defined by technical standards prescribed by the corresponding infrastructure administration/regulator. These thresholds are grounded on the accumulated knowledge acquired during a prolonged period of time in relation to the adequacy of the condition of the analysed assets, and they respond to a conservative envelope which guarantees the safety, integrity and right performance of the asset as a part of the system it works for. Alerts may be triggered by the appearance of corrective failures, faults or malfunctioning, which have not been detected in advance, in many cases, due to the lack of awareness concerning the explicative features which rule the failure behaviour as a result of the absence of supervising/monitoring actions. Being aware of these cases, new and additional explicative single and combined features have been considered and are incorporated to the listing of indexes to be monitored and measured. However, the large diversity of asset typologies make unlikely to envisage an ideal working case where any failure can be comprehensively explained using a fixed set of measurable features. At present, building systems and procedures based on identifying and using explanatory features, using data mining and analytics are becoming broadly investigated. A good structured database of historical maintenance interventions, founded on quantifying the information available on the said repository of knowledge, paves the way for using techniques in order to infer rules of combined explanatory features to correlate failures, enabling the creation of tools for decision making, be either fully or semi-automatic.

Diverse methodologies have been used to find a way to infer those trends, most of them in the orbit of Artificial Intelligence (AI) and Machine Learning (ML), a field of computer science and engineering concerned with the computational understanding of what is commonly called intelligent behaviour (17). An elaborate explanation concerning AI technique and discussion regarding other techniques can be found in the literature (18,19); specific works achieved in the area of linear infrastructures (e.g. railways) can be found in (20,21,22); and in the road inspection field, a considerable amount of works, essentially based on pattern recognition techniques and AI methods, are also available (20,23,24), showing the capability of performing

1 fast outcomes with minimal computing requirements.

2 The topic of *maintenance alert prediction* (25,26) has evolved going hand in hand with the
3 concept of reliability. Nowadays in this area, diagnosis is one of the main practical applications
4 of AI systems, focussed on predicting faults and condition evolving in specific asset components.
5 This is also the way maintenance on transport infrastructures has travelled along (e.g. road
6 pavement conditions, geometry integrity, etc). During the last decade, an escalating number of
7 literature pieces were published on applying ML to linear infrastructures (27,28,29,30). In all
8 cases, the common factor for this progression lays on the increasing availability of data captured
9 from auscultation/monitoring activities and campaigns; thus, ML techniques have promoted the
10 concept of learning from data, facilitating the extraction of patterns and trends by “let the data
11 speak by themselves”. During the last years, many efforts focused on improving the predicting
12 capability of asset management systems (31). At present, all data inferring-based disciplines (i.e.
13 machine learning, data mining, statistics, big data) are working in the same direction; it is
14 remarkable the numerous commercial software packages which evolved from a statistical origin
15 (32). But even though, the issue is still in an accelerated state of evolution.

16 This communication presents the methodologies, approaches and models for triggering
17 alerts associated to assets of road linear infrastructures needed of maintenance interventions, be
18 corrective, preventive or predictive. The estimated alerts are assessed according to the
19 information provided by a decision making tool based on the forecast evolving state reflected by
20 physical explanatory features, relevant to the state condition of the assets of interest, and the
21 historical interventions database. The output of the said tool will tag each estimated alert with a
22 level of severity and will rank all alerts in a hierarchical listing of interventions and their
23 associated probabilities of occurrence. The final purpose is to provide a procedure for managing
24 all active and predicted alerts, optimizing maintenance operations.

25 **METHODOLOGY FRAMEWORK**

26 The framework for predicting alerts in a road linear infrastructure, proposed in this research, is
27 sketched in Figure 1 where different modules (embodying techniques, methodologies, algorithms
28 and models) and their interactions, inputs and outputs are shown; the general objective pursued
29 focuses in four main goals:

- 30 i) Detecting pre-alerts and their reliabilities, based on the estimated values reached by the
31 explanatory features of any specific asset of the infrastructure, in pre-defined further
32 scenarios.
- 33 ii) Determining and triggering alerts, based on the predictions of a set of supervised machine
34 learning techniques and models, previously trained with the historical interventions
35 repository and the historical nowcast state condition assessment conducted prior to carrying
36 out any intervention.
- 37 iii) Evaluating technical severity levels (TSL) of pre-alerts and estimating a global technical
38 severity level (GTSL) for each estimated alert.
- 39 iv) Estimating the most probable intervention type required by the estimated triggered alert,
40 along with a ranking of those intervention types sorted in descending order of probabilities
41 of occurrence.

42 The developed methodologies implemented in each module are described below.

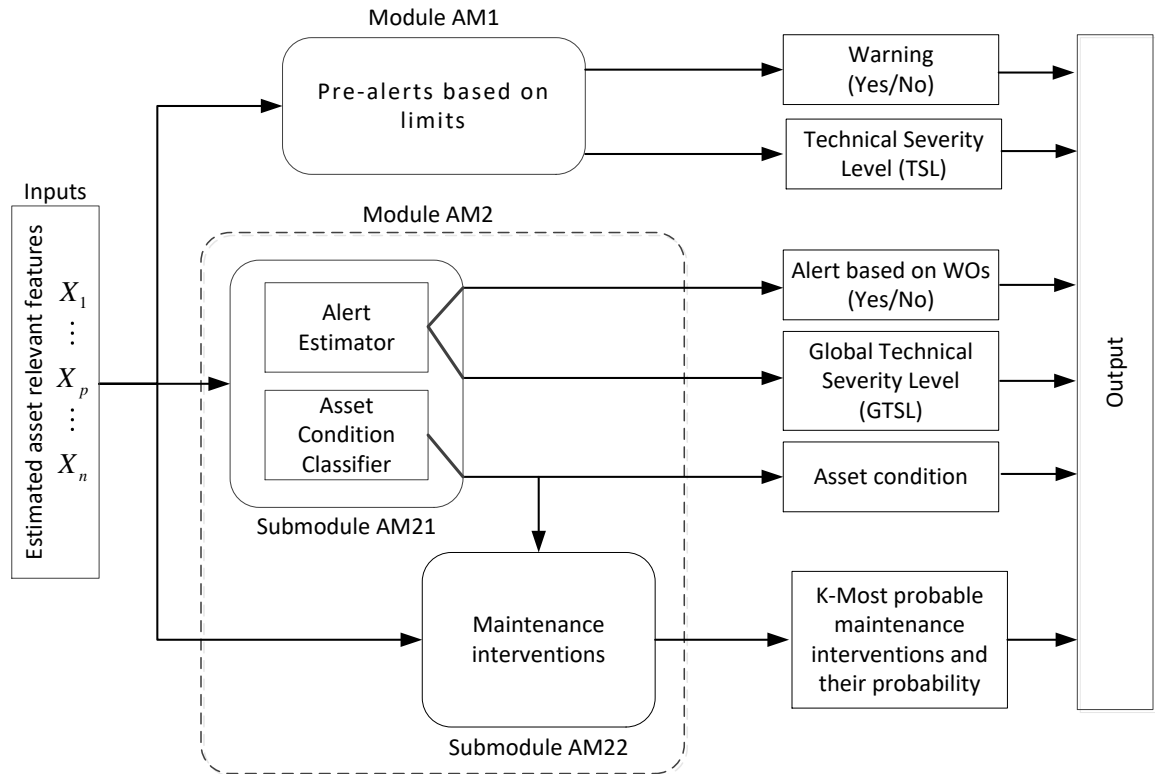


FIGURE 1 Methodology framework.

A Framework for Estimating Pre-alerts Based on Feature Limits

The feature-based procedure (module AM1) is built on deterministic and statistical techniques of the asset directly-related relevant explanatory features and their estimated evolving patterns. These patterns characterize not just the actual state condition of the asset itself but also the forecast state conditions at further scenarios, with their associated uncertainties. A description of the course the inputs are processed follows.

Figure 2 depicts the estimated evolving asset condition characterized by a generic explicative feature p -th, $X_p|_{a-i}$, of a particular road-segment asset, a -th, function of one or several other independent features of the asset X_t (e.g. time, accumulated load), showing a sample of the evolution of the feature in a hypothetical scenario; the most probable value is identified by the bold solid line of a fan of five evolving patterns corresponding to different probabilities according to some statistical reliability (e.g. $1-\gamma, 0.5$). The rightmost vertical cross-section line stands for the value of the independent variable X_t at one assumed further scenario X_{m+k} ; the nowcast scenario X_{tm} is pinpointed by the square dot (note that the dot might not lay on the bold solid line due to the fact that a divergence might exist between the evolving pattern model and the real value of the feature). The two upper horizontal lines identify two reference thresholds limits (RT), denoted by RT_i and RT_{i+1} , of the asset state condition criteria set by the relevant standards on design/quality/safety parameters (e.g. Normal Limit L_N , and Exceptional Limit L_E). The lower horizontal broken line identifies the expected value (i.e. with probability 0.5) at scenario X_{m+k} . According to the defined applicable criteria (e.g. European Standard, Road Administration Standard, Infrastructure Maintenance Managerial Body-MMB), the asset condition will be quantified regarding to the proximity the forecast values are from the threshold

limits. Besides, the previous cited probabilities of the estimated condition values are also very valuable pieces of information to assess the severity of the asset in each forecasted scenario.

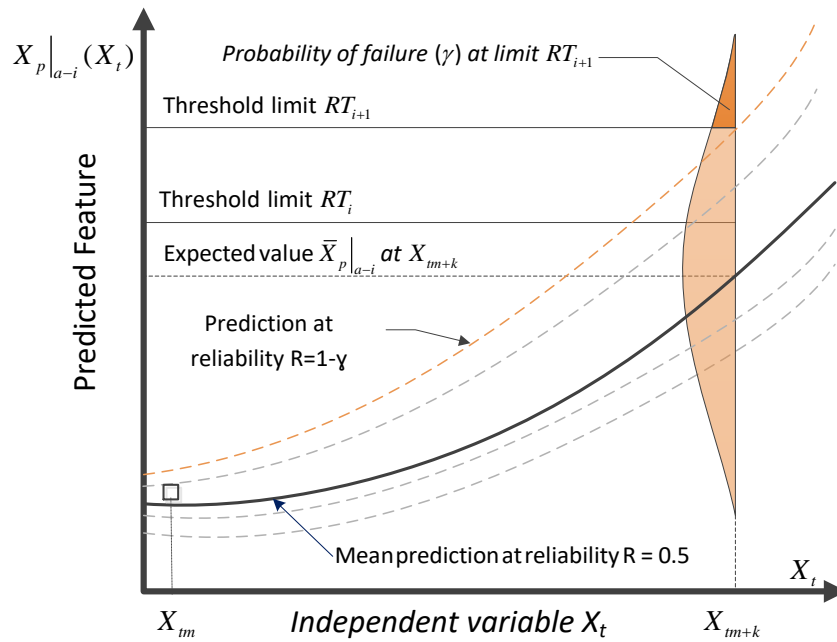


FIGURE 2 Asset condition prediction based on feature X_p , function of independent feature X_t .

Figure 2 depicts the representation of the estimated value of a feature with a known probability distribution. A pre-alert is generated when the condition of the asset, identified by any of its explanatory features, surpasses a threshold value defined by a pre-set limit under a specific probability, in a particular forecast scenario. Any triggered pre-alert is quantified by a technical severity level (TSL) associated to each explanatory feature of the asset state condition. The pre-set limits, used to trigger a pre-alert, serve to assess the state condition of the asset based on the value of the predicted feature. According to the distance between the forecast asset state condition to those limits, a TSL and the associated degree/level of uncertainty can be defined; the straightforward evaluation of that distance can follow a multiplicity of criteria, previously pre-defined by the MMB.

Taking as an example the IRI of a road segment and assuming that the pre-alert will be triggered according to equation (1),

$$P(IRI > Limit_{prescribed}) \geq \gamma \rightarrow \text{Pre-alert} \quad (1)$$

depending on the value of γ , an pre-alert may be either considered or discarded even when the mean value of the feature is the same. The value γ is a means to demand a higher reliability for those features considered more relevant from the maintenance activity point of view. The stating, by the MMB, of a low value of γ to a particular feature means that the feature is very relevant and a low possibility of failure is allowed. In this case $1-\gamma$ quantifies the reliability, thus, looking at Figure 2, equation (1) is rewritten as:

$$\text{Prediction (at reliability } 1-\gamma) \geq Limit_{prescribed} \rightarrow \text{Pre-Alert} \quad (2)$$

So far, the pre-alert is triggered when the probability of failure exceeds the limit γ . This definition of pre-alert can be generalized by the value of the asset condition technical severity

level (TSL). So when the TSL is greater than a predefined parameter α , a pre-alert is triggered. A case, for instance, of definition for the TSL_γ is presented in expression (3):

$$TSL_\gamma = \text{Prediction (at reliability } 1-\gamma) - \text{Limit}_{\text{prescribed}}$$

$$\text{If } TSL_\gamma \geq \alpha \rightarrow \text{Alert} \quad (3)$$

wherein, setting $\alpha=0$, the pre-alert recovers the previous equation (1).

This example shows that, with generic values of TLS and α , it is possible to define a general pre-alert criterion; module AM1 embodies different strategies.

A Framework for Estimating Alerts and Interventions Based on Historical Work Orders

When the existing variability of presumed-similar assets and endogenous and exogenous conditions make the knowledge insufficient to establishing clear rules for identifying explanatory features and their indicator thresholds, ML models, based on inferring knowledge from existing historical data, may pave the way to extracting the hidden know-how. This context is dealt with by module AM2.

With the developed approach, alerts are inferred by correlating the estimated values of the explanatory features of the asset state condition, in a requested further scenario of interest for the MMB, with the recorded information stored in the historical maintenance work-order repository. The relevant data stored in the repository contains, at least: a) the intervention type, b) the corresponding values of the asset explanatory features prior to the intervention, and c) the nowcast general state condition assessed by the maintenance team just before the intervention. This makes the triggering of alerts is based not on comparing their estimated TSLs with pre-set thresholds, but using the hidden non-explicit information which may explain the needed intervention carried out in past cases. By a training process in a prior step, the hidden information (i.e. non-explicit thresholds) is extracted to be used in the following step of estimating alerts. This procedure is implemented in two different functional submodules, based on distinct approaches. The first one (AM21) is devoted to detecting whether an alert (using, as inputs, the forecast values of the explanatory features of the asset under a pre-set reliability γ) will take place, its estimate Global Technical Severity Level (GTSL) and the estimated asset state condition (AC). The methodology is built on a set of supervised ML algorithms. The second one (AM22) estimates the most probable intervention type to be conducted, also providing the probability of the estimate.

The GTSL of an asset is derived from the TSLs of the corresponding explanatory features; this global index may not be referred to any threshold, as this limit is a hidden non-explicit information, rated in absolute value. The basic steps to obtain this GTSL follows herein below:

- i. Evaluate the TSLs of each individual explanatory feature involved (TSL_1, \dots, TSL_n). These can be calculated in terms of absolute values, implying that each TSL is referred to the value of the corresponding feature X_p or in terms of relative values to prescribed threshold limits.
- ii. Normalize the previous TSLs to refer all values to the same scale, in terms of their relative values: $TSLN_1, \dots, TSLN_n$. This normalization can be different depending on the nature of the feature, and subject to an external criterion defined by the MMB; for instance, in case of crocodile cracking, where the maximum value is 100, a normalizing criterion could be referred to this value.

1 iii. Pre-set different weights to each individual feature: $X_1 : \alpha_1, \dots, X_p : \alpha_p, \dots, X_n : \alpha_n$, subject to

2
$$\sum_{i=1}^n \alpha_i = 1.$$

3 iv. Compute the normalized GTSL (*GTSLN*): $GTSLN = \sum_{i=1}^n \alpha_i \cdot TSLN_i.$

4 At this point it is noteworthy to underline that, in most cases, intervention types are not
 5 directly associated to the asset state condition that resulted in triggering a pre-alert, but
 6 influenced by other factors; as an instance, two assets under the same state condition could be
 7 fixed with different intervention types depending on external factors (e.g. budget, available
 8 machinery, merging criteria to fix different sections belonging to the same road segment,
 9 maintenance policies, time opportunity). Due to this fact, the relevant information to be used in
 10 module AM22 should be the proposed intervention to be conducted based on technical criteria,
 11 and not the actual intervention carried out based on other considerations. Fail to follow this rule
 12 will have negative effects regarding the capabilities of training the ML algorithms and
 13 consequently their estimating capabilities.

15 *Supervised ML Approaches for Estimating Maintenance Alerts and Asset' Condition*

16 In order to set on the whole process of estimating alerts, the first stage of the ML model
 17 (submodule AM21) implies a training process with the adequate information. The data repository
 18 contains the historical interventions regarding the monitored infrastructure case study. This
 19 repository may also contain recorded information regarding the subjective assessment of the
 20 asset condition inspected by the MMB team just before the intervention was carried out; in
 21 particular, the assessment of each individual explanatory feature (SX_p), each combined
 22 explanatory feature (CX), and a global valuation of the state condition of the whole asset (G).
 23 Three other set of pieces of information are also relevant: i) the measurements carried out in the
 24 analysed linear asset section/segment (previous to the corresponding maintenance work-order),
 25 associated to physical explanatory features; ii) endogenous and exogenous
 26 characteristics/variables related to the asset state condition evolution.

27 The developed methodology is split in two parallel predicting blocks. The *Alert Estimator*
 28 block is used to trigger alerts using a classifier, which just correlates the values of the
 29 explanatory features with the requirement for maintenance. The target is a binary variable
 30 informing whether a requirement for maintenance was needed or not. The learning methodology
 31 consists of an automatic classification in a binary variable (1-0: Yes/No) and a set of four
 32 automatic binary classification models (i.e. DT-Decision Tree, ANN-Artificial Neural Network,
 33 KNN-K Nearest Neighborhood, SVM-Support Vector Machine). The second block, *Asset*
 34 *Condition Classifier* correlates feature measurements with different subjective evaluations of the
 35 asset condition provided and recorded by the MMB team. By this way, the system “learns” from
 36 the MMB know-how and, when a new measurement is introduced in the system, the asset
 37 condition is predicted. According to this, the inputs for training are the value of the features, and
 38 the target variables are the subjective evaluations.

39 Once the system is initialized and the machine learning models trained, the objective of
 40 detecting road segments where maintenance will be required in a specific queried further
 41 scenario is achieved by the “Alert Estimator” (upper box of AM21 in Figure 1); the needed
 42 information are the forecasted values of all explanatory features, $\hat{X} \equiv [\hat{X}_1, \dots, \hat{X}_p, \dots, \hat{X}_n]^T$.

These are the only inputs to the system, which yields the requirement for maintenance as the outcome; and in case an intervention is forecasted, an alert will be triggered.

The prediction on the asset state condition is materialized by estimating a subjective assessment. The global asset condition estimate can be seen as the predicted overall state condition of the asset regarding the simultaneous contribution of all feature effects as a whole.

Supervised and Un-supervised ML Approaches for Estimating Maintenance Types

The repository of historical interventions is the main source of information on the actual interventions carried out to get the transport infrastructure back into service. In most cases the reported information does correspond to the precise operations needed to recover the assets' functionalities from detected faults/failures by either corrective/preventive actuations or predictive policies; in other cases the conducted interventions followed other strategic considerations. In order to estimate the most probable intervention type on a specific asset, according to the know-how contained in the historical repository, a process has to be launched by correlating the estimated values of the relevant explanatory features \hat{X} and the forecast subjective state condition $(\hat{S\hat{X}}, \hat{C\hat{X}}, \hat{G})$ provided by the second block of submodule AM21, versus similar asset samples reported in the historical repository.

The intervention type prediction is based on a two-phase un-supervised ML scheme, carried out by the second submodule (AM22). This block determines a hierarchical listing of the most probable interventions to be conducted, and their probabilities of occurrence, via a learning procedure based on information contained in the historical database. The first phase is based in a k-clustering technique, the second phase in a k-neighborhood technique (KNN) from the outcome inferred by the *Asset Condition Classifier* block from a supervised ML technique (SML) submodule AM21. A description of the techniques used and methodologies implemented follows.

Un-supervised Clustering of Historical Data. The available historical data is grouped into different more or less well-defined clusters (e.g. L), embodying samples with similar condition (i.e. similar values for all features) in the same cluster. The clustering is carried out taking into account just the asset condition (i.e. values of features X) without using the actual conducted intervention. Therefore, the same cluster could contain different intervention types as shown in Figure 3a, where only two features, X_r and X_s , are considered, but several maintenance types are involved, identified by different plotting shapes (i.e. squares, triangles and circles).

Once the clustering is obtained, in order to determine the most probable type of intervention associated to a triggered alert, the estimated values of its features (\hat{X}) are the inputs to the model, and the corresponding location P of the asset state condition can be inferred; this location will belong to a specific k -cluster based on the distance to its centroid. The information corresponding to the historical interventions of the samples implied in this cluster defines the different maintenance types involved; from a simple rating analysis a probability for each type can be easily assigned to the estimated alert.

To take into account MMB's know-how, related to the subjective assessment of all historical samples implied in the k -cluster, the empirical occurrence of each maintenance type can be weighted by scores giving more relevance to those samples whose assessment assimilate to the predicted subjective evaluation of the features $[S\hat{X}, C\hat{X}, \hat{G}]$ of the asset implied in the triggered alert, derived by submodule AM21. Regarding the example, Figure 3a shows the k -

cluster the sample P belongs to; each sample contained in the cluster carries information regarding the historical features subjective evaluation $[SX_r, SX_s]$. Those samples with similar subjective evaluation to that predicted asset will be scored with a higher weight than the rest. In this way, an unsupervised machine learning approach is complemented with the information provided by the historical know-how. Using this technique, a huge number of assets under similar state condition can be used for estimating maintenance strategies, circumventing recurrent problems such as bias.

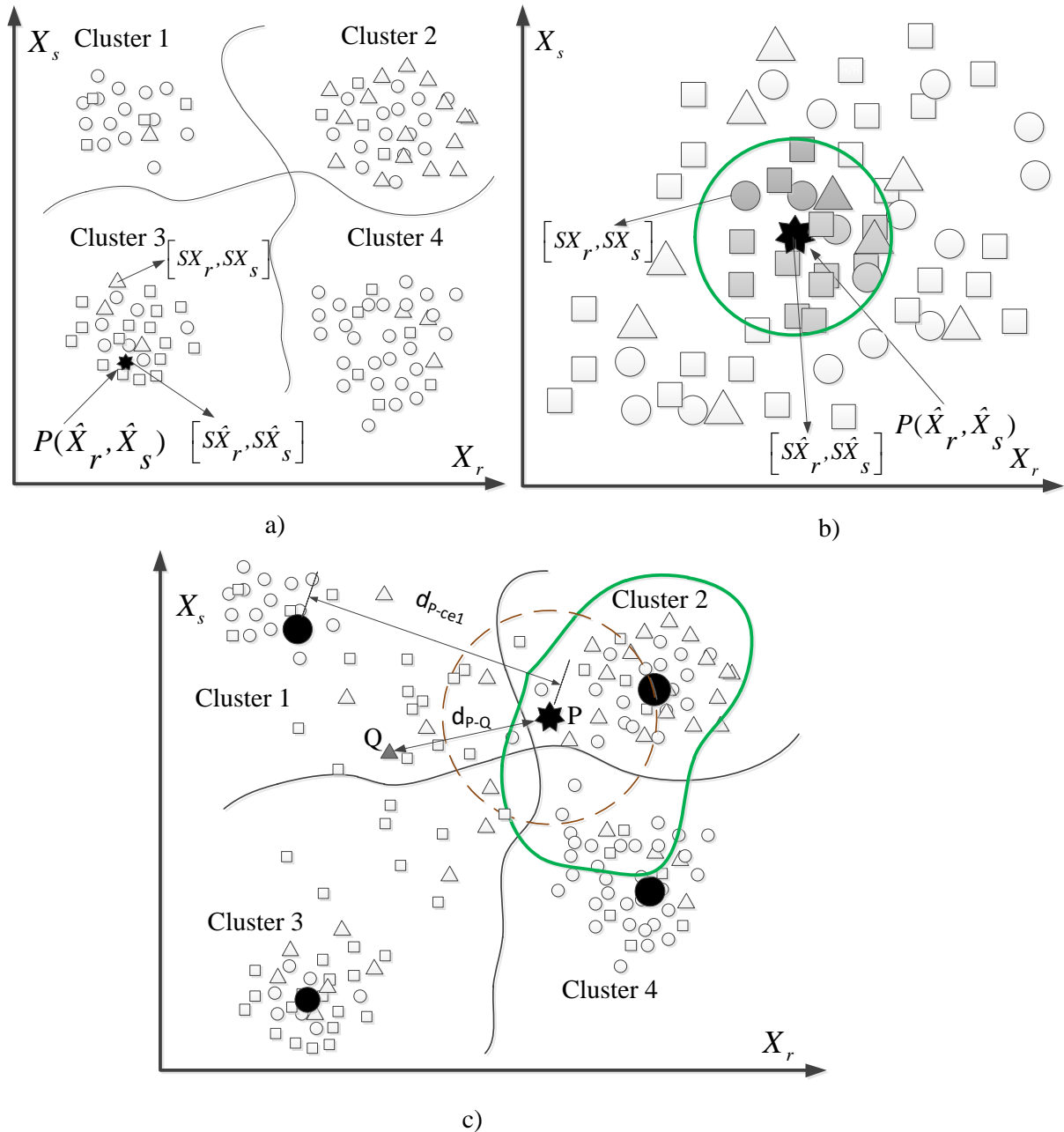


FIGURE 3 Prediction of intervention type and probability: a) clustering from historical state condition, b) KNN plus know-how approach, c) Fusion model.

The above-described methodology has serious limitations when data do not show a clear grouping, as shown in Figure 3a; this drawback can be circumvented by using a second-phase based on the K-Nearest Neighborhood algorithm (KNN). In this case, the k nearest historical samples to a given triggered alert (samples inside the circle in Figure 3b) are chosen to define the most probable types of maintenance associated to an alert. The methodology is quasi-similar to the previous one considering just k samples. This means that the selected samples including P belong to the same unique “cluster” in order to apply the technique described in phase-one. Following the same approach, the underlined know-how can be taken into account by giving more importance to those samples with similar evaluation than the predicted one for sample P , $[\hat{S}\hat{X}, \hat{C}\hat{X}, \hat{G}]$, provided by submodule AM21. In the example, each sample occurrence (from the k -nearest set) is weighted by a score which depends on its subjective assessment, higher for those samples with similar evaluation to P than for the rest. This methodology is less sensitive to the data distribution than the clustering but, in general, it uses fewer samples to define the most probable intervention types and some pieces of information can be screened out depending on the value of parameter k .

Fusion model. To take into account the advantages of the two single methodologies developed, clustering and KNN, for estimating the type of intervention, a merged one has been constructed, summarized in the following steps:

- i. Generate a proposed cluster model, as outlined in previous paragraph. Figure 3a depicts a recreation where clusters are not very clear identified, due to samples close to borders.
- ii. Compute the Euclidean distance between sample P and each of the L cluster centroids:

$$(d_{P-ce1}, d_{P-ce2}, \dots, d_{P-ceL}).$$

- iii. Choose the k nearest samples to triggered alert P using a criterion distance z between sample P and a generic sample Q belonging to cluster J -th, $z_{P-Q} = d_{P-Q} \cdot d_{P-ceJ}$, where d_{P-Q} stands for the Euclidian distance between P and Q . In this way, samples belonging to different clusters are penalized, as it is the case of sample Q in Figure 3c.

Figure 3c shows the difference of applying the KNN model (broken circle) and the fusion model (solid closed line). As it can be seen, this last model is not so dependent on the existence of a clear data grouping (i.e. cluster definition) and enables to increment the value of k in a guided way, regarding those samples belonging to the same cluster than P .

As in those previous single models, MMB’s know-how is taken into account by giving more importance to those samples with the same evaluation than the predicted one for sample P , $[\hat{S}\hat{X}, \hat{C}\hat{X}, \hat{G}]$ inferred by submodule AM21.

EMPIRICAL STUDY CASE

The pilot case selected is a meshed road network in the central region of Portugal, managed by Infraestruturas de Portugal, totalling 539 km; it includes several road categories, as principal itineraries, supplementary itineraries, national roads, regional roads and other roads; the road categories are classified on the basis of features such as travel speed, traffic volume, traffic mix and strategic importance. Regarding traffic levels, the chosen demo case presents heterogeneity among the chosen sub-networks (i.e. itineraries), between 2,500 and 10,000 vehicles per day, with an average of 9% of heavy vehicles.

Relevant Pieces of Information and Datasets

The conducted maintenance strategy followed categorizes interventions in major or routine maintenance. Major maintenance includes relevant works in terms of cost, length and complexity, it is planned in a medium-term basis (5 year periods) and follows a prioritization process, annually reviewed. Routine maintenance includes smaller scale and lower complexity works, such as pavement repairs, drainage system cleaning, shoulder treatment, minor works performed in bridges and any urgent repair.

Needless to say that a filtering pre-process has to be conducted to the data recorded to extract the relevant information free of inconsistencies (e.g. unidentified/undetermined geometrical location of measurements, measurements captured by different auscultation vehicles, unclear description of the intervention carried out, assets affected, etc), which may reduce the final number of valid records a non-negligible percentage. The expected outcome is a set of alerts each one linked to a certain infrastructure component where maintenance is necessary in order to keep a certain level of infrastructure condition, indicating the type of maintenance to be applied.

The number of filtered records in the historical interventions database is limited according to the network extension of the pilot case and the time period. In order to circumvent this drawback, a multiplicity of simulated data-sets derived from the statistics distributions of the available empirical real data was generated in order to select the most appropriate machine learning model.

Model selection for estimating maintenance interventions

This section presents the results after applying a set of machine learning techniques: Decision Trees (DT), K-Nearest Neighborhood (KNN), Support Vector Machines (SVM) and Artificial Neural Networks (ANN), based on the proposed methodology, to the repeated random sampling explained in the previous paragraphs. The main objective is to build models with good generalization capabilities to perform well on new data (i.e. test data for which the model has not been trained). Another objective is to determine the sample size required for each model to be able of generalizing. The evaluation of the performance of these classification models has been done by the confusion matrix.

The relationships between training set size, model complexity, and prediction error, have been analyzed and identified. Two separate goals are addressed: (i) Model selection: estimating the performance of different models and their parameters in order to choose the best one; and once a final model is chosen (ii) Model assessment: estimating its prediction error on new data.

The dataset, with a sample of 2,000 bootstrapping interventions is randomly divided into three sets: training to fitting the models, validating to estimating prediction errors for model selecting, testing for assessing the error of the chosen model.

The main parameter to be calibrated in DT techniques is the number of splits or branches, ranging from 2 to 40. The final best model has 28 splits.

In KNN classification technique, a total of 9 models that ranges from 2 to 20 neighbors have been analyzed. The final calibration of this method includes the choice of additional parameters such as distance metric (Euclidean) and distance (square inverse) weighting function.

The SVM technique is mainly characterized by its kernel function. Testing different ones (linear, quadratic, 3-degree polynomial, gaussian, ...) it is concluded that a 3-degree polynomial kernel function is the best option.

The ANN technique is a more complex algorithm from the calibrating point of view due to the number of parameters to be configured. The design of the number of hidden layers and neurons is just addressed. A pattern neural model, with Levenberg-Marquardt backpropagation training function and mean squared normalized error performance function, has finally been selected. Figure 4a shows seven series of different ANN designs, whose description follows: (i) 1 h-layer: only a hidden layer, where the number of neurons ranges from 2 to 20; (ii) 2 h-layers [2 X]: two hidden layers, where first hidden layer has 2 neurons and second hidden layer a number of neurons ranging from 2 to 10; (iii) 2 h-layers [3 X]: two hidden layers, where first hidden layer has 3 neurons and second hidden layer a number of neurons from 2 to 10; and so forth. In order to compare the different models, the complexity parameter is defined as the sum of the weights to be trained in each model. Model complexity increases as the number of neurons and hidden layers increase. Figure 4a displays that models with only one hidden layer (1 h. layer) achieve better performance. For greater complexity there are several models that reach high accuracy. The optimal model is one that achieves high accuracy with less complexity. The final chosen model has one hidden layer with only 6 neurons (black diamond sample in Figure 4a); the complexity is 78, unlike the model with 8 neurons reaches a similar accuracy but its complexity increases up to 104.

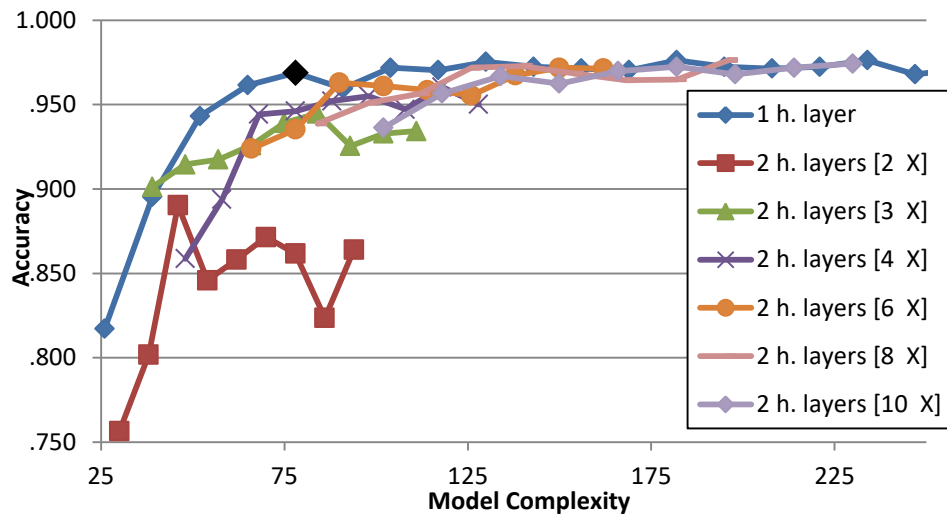
The final models selected of each technique achieve the following accuracies: (DT, 0.989; KNN, 0.960; SVM, 0.969; ANN, 0.969), resulting DT the model that reaches the higher accuracy.

The expected test error of each estimated model is calculated using learning curves (Figure 4b). These plots illustrate the important issue in assessing the ability of a learning technique to generalize.

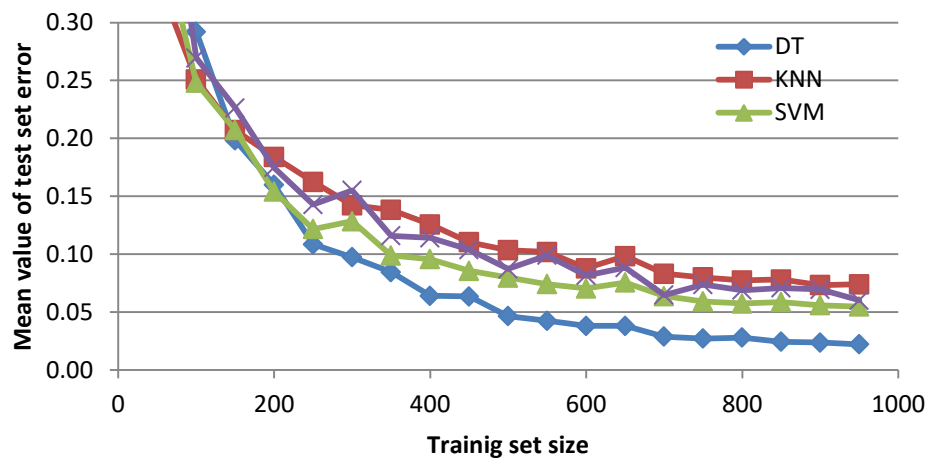
As the training set sizes get larger, the curves converge toward a threshold representing the amount of irreducible error in the data.

The DT model achieves the smallest error. SVM and ANN models yielded similar results and none of them reduces the test set average error below 5%. Only the DT model has reached an average test set error of less than 5% using a training set size of 500 records. So, this model is the one used to infer the result of the use case.

In order to know more deeply the selected model the confusion matrix is calculated (Figure 5a). In this matrix it can be seen that the prediction fails appear when the model estimates T3 or T4 maintenance types instead of T3.1 and when T2 is predicted instead of T1.



(a)



(b)

FIGURE 4 Model selection: (a) Analysis of ANN method, (b) Learning curves.

Results Inferred from Empirical data

In order to check the accuracy of the final selected model, the 2014 campaign data-set is used as a testing sample. Figure 5b shows a comparison between the real interventions carried out and the predicted ones. According to this figure, results are reliable and the errors occur between T4 and T3.1 maintenances.

A total of 1241 sections of 500 meters were used as input of the DT model, in 1159 sections the predicted value coincide with the real one, obtaining a final accuracy of 93.4%. Comparing the predicted work orders (WO) and the real ones in all sections, 843 sections out of 859 are correctly predicted as T0 maintenance type, 39 out of 41 as T1, 148 out of 187 as T2, 28 out of 30 as T3, 80 out of 93 as T3.1 and 21 out of 31 as T4. This implies that the main error is obtained when the model has to predict T2 (only 79.1% of accuracy) and T3.1 (acc. 86%) as suggested by the confusion matrix (Figure 5a).

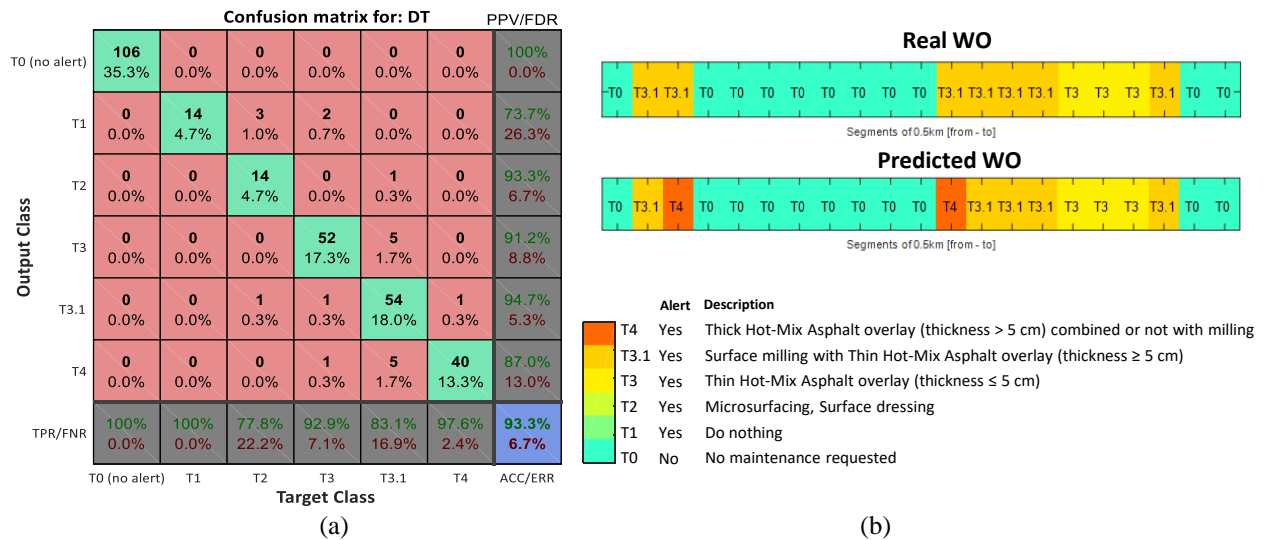


FIGURE 5 Results of DT model. (a) Confusion matrix, (b) Real vs. Predicted WO.

CONCLUSIONS AND PROSPECTIVE HORIZONS

In this paper, a road network was studied from the maintenance interventions predictive approach. Four machine learning techniques have substantiated the optimum choice of the best predictive models which have framed the automatic learning methodology from historical intervention work-orders, asset features and measurement auscultations. The main predicted outcomes are: a) the estimated intervention type for each road section and the probability of occurrence, b) a sorted out listing of estimated alerts according to the technical severity level. Each prediction set is referred to a future scenario identified by its time-stamp.

The results evidence that, even with a limited historical database, using a bootstrapping replicating sampling according to the data statistics, an optimum model choice, parameter fitting and proper training of the supervised models, the framework methodology provide good predictive capabilities.

The methodologies and results presented herein are far from being exhaustive and conclusive, and several parallel lines of research are open: a) sensitivity to the quality of intervention description in the historical repository regarding the intervention timestamp, b) importance of the detailed/undetailed description of the asset state condition previous to the intervention, c) self-learning rules from automatic learning from false positive/negatives, among others.

The methodology presented constitutes a step forward in generating a smart decision support tool to derive intervention plans based on alert forecasting generation and the optimal selection of activities regarding the most critical interventions to be carried out, which are the single bricks to build the final purpose of covering the full range of planning maintenance at operational, tactical and strategic level, under an expert intelligent framework.

ACKNOWLEDGEMENTS

The research has received funding from European Union's Horizon 2020 Research and Innovation Programme (grant agreement n° 636496). Some of the authors express their gratitude to the Spanish Ministry of Economy and Competitiveness for the partial subsidy granted under the national R&D program (TRA2015-65503) and the Torres Quevedo Programme (PTQ-13-06428). The authors acknowledge Infraestructuras de Portugal (IP) for making available the

database used in this research. The content reflects only the authors' view and it is stated that the EU and IP are not liable for any use that may be made of the information contained therein.

REFERENCES

- [1] Ahern A., and G. Anandarajah (2007). Railway projects prioritisation for investment: application of goal programming. *Transp Policy* 14(1):70–80.
- [2] Li M.Y., K.Q. Han, X.Y. Zhang, X.I. Li, Y.J. Zhai, and W.H. Yu (2008). A management information system for mine railway transportation equipment. *J China Univ Min Tech* 18(3):373–376.
- [3] Link H. (2006). An econometric analysis of motorway renewal costs in Germany. *Transp Res Part A: Policy Pract* 40(1):19–34.
- [4] Coffen-Smout S., and G.J. Herbert (2000). Submarine cables: a challenge for ocean management. *Marine Policy* 24(6):441–448.
- [5] Cagno E., F. Caron, M. Mancini, and F. Ruggeri (2000). Using AHP in determining the prior distributions on gas pipeline failures in a robust Bayesian approach. *Reliab Eng Syst Saf* 67(3):275–284.
- [6] Creig-Smith S. (2001). To refurbish or replace steel water pipelines, that is the question. *Eng Fail Anal* 8(2):107–112.
- [7] Castanier B., and M. Rausand (2006). Maintenance optimization for subsea oil pipelines. *Int J Pres Ves Pip* (The 16th European safety and reliability conference), 83(4):236–243.
- [8] Brito A.J., and A.T. de Almeida (2009) Multi-attribute risk assessment for risk ranking of natural gas pipelines. *Reliab Eng Syst Saf* 94(2):187–198.
- [9] IBM Maximo, <http://www-03.ibm.com/software/products/en/maximoassetmanagement>. Accessed Jun. 6, 2017.
- [10] Infor Corporation. USA, <http://www.infor.com/solutions/eam/>. Accessed Jun. 15, 2017.
- [11] Optram. Bentley Systems Inc. USA, <http://www.bentley.com/esES/Products/Bentley+Optram>. Accessed Jun. 15, 2017.
- [12] Oracle EAM. <http://www.oracle.com/us/products/applications/060286.html>. Accessed Jul. 06, 2015.
- [13] SAP (2014). <https://www.sap.com/products/linear-asset-management.html>. Accessed Jun. 15, 2017.
- [14] FHWA (2006)., *Pavement preservation compendium II*, Federal Highway Administration.
- [15] FHWA (2008). *Life-Cycle Cost Analysis in Pavement Design (In Search of Better Investment Decisions)*. Federal Highway Administration. Office of Asset Management. U.S. Department of Transportation.
- [16] SUST (2008). *Sustainable Bridges - Assessment for Future Traffic Demands and Longer Lives*. FP6 collaborative project.
- [17] Shapiro S. (1991). *Foundation Without Foundationalism. A case for Second-Order Logic*. Clarendon Press, Oxford University Press.
- [18] Peter, J. (1999). *Introduction to expert systems*, Third edition, Harlow, England: Addison Wesley Longman.

- 1 [19] Negnevitsky, M. (2005). *Artificial intelligence: A guide to intelligent systems*, Second
2 Edition, London: Addison-Wesley.
- 3 [20] Yella, S., M.S. Dougherty, and N.K. Gupta (2006). Artificial intelligence techniques for the
4 automatic interpretation of data from non-destructive testing, *Insight Journal of the British*
5 *Institute of Non-Destructive Testing*, vol 48, no 1, 2006.
- 6 [21] Podder, T. (2010). *Analysis & study of AI techniques for automatic condition monitoring of*
7 *railway track infrastructure*. Master Thesis. Reg. No: E3845D. Höskolan Dalarna.
- 8 [22] Iqbal, Z. (2010). *A Study of AI techniques for Railheads, Vegetation, Switches & Crossings*.
9 Master Thesis. Reg.No. E38440. Höskolan Dalarna.
- 10 [23] Sambath, S., P. Nagaraj, and N. Selvakumar (2011). Automatic Defect Classification in
11 Ultrasonic
- 12 [24] Martin, O., M. Lopez, and F. Martin (2007). Artificial neural networks for quality control by
13 ultrasonic testing in resistance spot welding. *J. Mater. Process. Technol.* 183, 226–233.
- 14 [25] Kirkman M.M. (1886). *Maintenance of Railways*, C.N. Trivess Printers, Chicago.
- 15 [26] Morrow L.C. (1957) Ed., *Maintenance Engineering Handbook*, McGraw-Hill, New York.
- 16 [27] Witten I.H., E. Frank, and M.A Hall (2011). *Data Mining: Practical Machine Learning*
17 *Tools and Techniques* . Morgan Kaufmann Publishers.
- 18 [28] Dick C.T., C.P.L. Barkan, E.R. Chapman, and M.P. Stehly (2003). *Multivariate statistical*
19 *model for predicting occurrence and location of broken rails*.
- 20 [29] Podofillini L., E. Zio, and J. Vatn (2006). Risk-informed optimisation of railway tracks
21 inspection and maintenance procedures. *Reliability Engineering and System Safety* 91(1), 20–35.
- 22 [30] Adarkwa O.A., and N. Attoh-Okine (2013). Pavement crack classification based on tensor
23 factorization. *Construction and Building Materials*, 48 (2013) 853–857.
- 24 [31] Faiz R.B., and E.A. Edirisinghe (2009). Decision Making for Predictive Maintenance in
25 Asset Information Management. *Interdisciplinary Journal of Information, Knowledge, and*
26 *Management Volume 4*.
- 27 [32] SAS (2014). *Predictive Asset Maintenance* . SAS Institute Inc. World Headquarters.