Francisco J. Morales[a], Antonio Reyes[a], Noelia Caceres[a], Luis M. Romero[a], Francisco G. Benitez[a*], João Morgado[b] and Emanuel Duarte[b]

[a]*Transportation Engineering, Faculty of Engineering, Universidad de Sevilla, Camino de los Descubrimientos, 41092-Sevilla, Spain*

[b]*Infraestruturas de Portugal, Almada, Portugal*

* corresponding author: benitez@us.es

SHORT RESUME OF AUTHORS (alphabetical order)

Francisco G. Benitez (benitez@us.es) was born in Sevilla, Spain, 1956. He received his diploma (M.Eng) degree in Industrial/Mechanical Engineering from the University of Sevilla (Spain 1979) and his PhD from the Technical University of Madrid (1981). He was a Research Fellow at the Department of Engineering Science, University of Oxford, Oxford (U.K. 1981-1983), and a Fulbright Scholar and Visiting Associate at the California Institute of Technology, Pasadena (USA 1984-1986). He is currently a Professor with the Department of Transportation Engineering, University of Seville, and the Head of the Transportation Engineering and Infrastructure Division. His actual fields of research are transport infrastructure, transportation modelling and numerical methods.

Noelia Caceres (noelia.caceres@gmail.com) was born in Don Benito, Spain, in 1980. Ph.D. degree in telecommunication engineering from the University of Seville, Spain, in 2010. Senior Researcher with the Department of Transportation Engineering, University of Seville. Her research interests include transport demand modeling, traffic flow, software design, and intelligent transportation systems.

Emanuel Duarte (emanuel.duarte@infraestruturasdeportugal.pt) was born in Porto Moniz, Portugal, in 1970. BSc. Applied Mathematics University of Coimbra in 1994. Joined the Portuguese Road Administration Company in 1993, currently called Infraestruturas de Portugal, S.A., with both road and rail administration. His area of work and expertise is in Pavement Management System, including road condition assessment, pavement performance indicators, data processing and analysis, quality control of road inspection technology, and measurements.

F. J. Morales (fmorales2@us.es) He received the Ph.D. degree ("Doctor Europeus") in industrial engineering from the University of Seville, Spain, in 2012. Senior Researcher with the Department of Transportation Engineering, University of Seville. His research interests include vehicle dynamics, transmission system and transport infrastructure.

João Morgado (joao.gmorgado@infraestruturasdeportugal.pt) was born in Lisbon, Portugal, in 1981. He graduated in Civil Engineering in 2004 from IST (University of Lisbon), starting his career as a researcher at the same university in the field of Pavement Management. Later (2006-2009) he was a lecturer in transportation and transport infrastructures. In 2010, he joined Estradas de Portugal, S.A. and worked in the development of its Pavement Management System. Since 2012, holds a Ph.D. in Civil Engineering from IST and currently works for Infraestruturas de Portugal, S.A., being responsible for the road and rail assets performance reporting.

Antonio Reyes (areyes@us.es) received a Ph.D. degree in industrial engineering from the University of Seville, Spain, in 2014. He is currently a Senior Researcher with the Department of Transportation Engineering, University of Seville. His fields of research include transport infrastructure, transportation modelling, numerical methods and software design.

Luis M. Romero (lromero1@us.es) was born in Badajoz, Spain, in 1971. He received the Ph.D. degree in industrial engineering from the University of Seville, Seville, Spain, in 2007. He is currently a Senior Researcher and Lecturer with the Department of Transportation Engineering, University of Seville. His research topics are transport demand modeling, traffic flow, numerical methods, intelligent transportation systems, and software design.

A machine learning methodology to predict alerts and maintenance interventions in roads

This contribution is about predicting maintenance alerts in roads and selecting the most appropriate type of interventions recommended for preventing the occurrence of future failures. The objective is aligned with that covered by pavement maintenance decision support systems (PMDSS), though the methodology presented can be applied to other non-pavement road linear assets.

The purpose is to summarise the main findings in the development of an approach based on testing the four most extended machine learning techniques (ML), namely Decision Trees (DT), K-Nearest Neighbourhood (KNN), Support Vector Machines (SVM) and Artificial Neural Networks (ANN), using data from the historical inventory of inspections and maintenance interventions of a case study to illustrate the potential that such approach can offer to road maintenance managers. The correlation process embodies supervised and unsupervised training of models. Despite the fact that ML techniques have already proved to be a serious counterpart to traditional modelling, their practical use in estimating maintenance alerts and ways to ascertain their reliability are being thoroughly investigated. The maintenance predictions of the proposed approach, alerts, interventions, and their severity, are presented and compared over various segments corresponding to the real maintenance interventions conducted on an existing road network of a geographical zone. This comparison showed a high reliability level.

Keywords: road; predictive maintenance; machine learning; linear infrastructure

1. Introduction

Research on linear-asset management has been a point of broad attention from engineering practitioners and academics in the last two decades; seminal publications have focused on specific types of linear assets, with important advances achieved in roads.

To push forward the development of reliable prediction tools, in order to foresee and maximise the availability of transport networks and to optimise the resources devoted to keep them serviceable, is of paramount importance for society. The forecasting of maintenance interventions and operations, either in the short term or for strategic further scenarios, is a current hot subject for most transport infrastructure administrative bodies in developed countries, whether public or private. The Federal Highway Administration, a world leader in road preservation programmes, has demonstrated that an accurate knowledge of asset condition and a proper management and analysis of gathered data could lead to more effective maintenance planning that is able to extend the lifecycle of assets (e.g. pavements) from 3 to 10 years, which means by a range between 12% and 20% (FHWA, 2006, 2008).

Currently, from the managerial aspect, several linear-asset management professional codes (e.g. Maximo from IBM; Infor; Optram from Bentley) or adapted general purpose codes (e.g. EAM from Oracle, SAP) are on the market to address linear-asset information management issues; however, they either do not have any predictive and/or decision analysis functionalities or are just at the stage of reinforcing in this vein, even though these are prime strategic objectives for most of them. On roads, a set of codes (Exor from Bentley, Optram from Bentley, HDM-4 from HDMGlobal, HIMS, Infor, Loc8, Paver from CSU, PMS Core from Heller, SAP-EAM from SAP) includes applications providing some predictive capabilities on maintenance alerts and operations, which facilitate the forecasting tasks; however, they have a long way to go before reaching the desired goal and much effort is expected to be invested in the foreseeable future.

The triggering of maintenance alerts/alarms regarding the 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. In a non-negligible percentage of cases, 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 govern the failure behaviour, as a result of insufficient supervisory/ monitoring actions. This is, for example, the case of a stretch of road in which some deterioration cannot be explained just by analysing the longitudinal International Roughness Index (IRI), transversal unevenness depth (RUT) or crocodile cracking (CT) indices in an independent manner, but by using a combination of those features, and in most cases incorporating other specific physical explanatory characteristics of the stretch. Nevertheless, the large diversity of asset typologies due to the appearance of new designs, construction processes, materials, and notwithstanding the externalities to the asset itself (e.g. service loadings, environmental conditions, other features associated with either the directly involved or other parent-child assets and influencing neighbours) make it unlikely to envisage an ideal working scenario where any failure can be comprehensively explained using a fixed set of measurable features and, as a further consequence, to get a reliable estimate of the evolution of the condition of the implied asset. Building systems and procedures based on identifying and using as many sufficient explanatory features as possible is being broadly investigated. Still, due to the difficulties of characterising and mapping all explanatory features, a multiplicity of procedures have emerged, most of them based on more or less sophisticated data analytics techniques which make use of past faults/failures in order to predict eventual ones and their probabilities of occurrence. These techniques are based on the existing knowledge of historical maintenance works and their corresponding asset condition encountered during the repair. In the last decade, the potential of methods in the area of artificial intelligence, machine learning, expert systems and other soft computing techniques has enriched the studies in this field.

Early work on this issue has focused on quantifying the capabilities of several machine learning techniques in predicting maintenance interventions using a reduced set of historical records (Morales et al., 2017). A lateral study was conducted on defining the relevant information that needs to be recorded during maintenance works, regarding the asset's condition just before and after being maintained, in order to provide self-learning capabilities to facilitate maintenance predictions (Reyes et al., 2018).

This communication reflects a step further in this research, presenting a detailed description of the automated prediction scheme, the methodologies, approaches and models for triggering alerts/alarms associated with linear assets of road infrastructures that are in need of maintenance interventions. The estimated alerts are assessed according to i) the evolving state reflected by physical explanatory features relevant to the condition of the assets of interest (not presented herein), and ii) the information contained in the historical maintenance intervention database. The predictions (outputs) will tag each estimated alert with a level of severity and will rank all predicted alerts in a hierarchical listing of maintenance interventions and their associated probabilities of occurrence. The final purpose is to provide a procedure for managing all active alerts in order to optimise the management of maintenance operations.

2. State of the art

In the predictive maintenance sphere, there was a major leap forward in the years 1940–60 with the introduction of the concept of reliability and the use of statistics and optimization techniques (see Dekker, 1996), which gave rise to Reliability-centered Maintenance (see Moubray, 1998), one of the most widely used tools nowadays. A second step came later on with the use of computing data analytics for inferring models, based on empirical quantitative information data, with the purpose of projecting the evolving condition of particular assets and components. Since the appearance of automated data-based techniques (e.g. artificial intelligence, expert systems, statistical learning) in the '80s, the organization in a scientific corpus (e.g. machine learning) in the '90s and 2000s, and the recent popularization of massive data analysis (e.g. data mining), there has been huge interest in predicting the condition of transport infrastructure, with the purpose of envisaging the appropriate interventions and setting maintenance planning programs, according to the available resources and with the minimum impact on the infrastructure serviceability.

In the road maintenance issue, since the initial development of pavement management systems (PMS) in the late 1960s, major efforts have been made to incorporate more prediction capabilities to assist agencies in selecting the procedures to maintain this type of infrastructure in functional conditions using different rule-based expert systems (Hajek et al., 1987; Ritchie, 1987; Ritchie et al., 1987; Tandon & Sinha, 1987; Ross et al. 1992). Since the '90s this tendency has diversified in modelling the diagnosis and evolution of the state of specific road components (e.g. pavement, layers, sub-base, base, platform) using several mathematical tools, such as ML techniques and data analytics. These have become very much explored in an indirect manner through explanatory features and indices (e.g. roughness, rutting, cracking, deflection) as a function of the multiplicity of factors that affect the condition (e.g. traffic, climate, construction materials and processes, soil platform properties). Herein some works can be referenced as applying ML to estimate the forecast assessment of several pavement indices such as roughness, rutting and cracking (Attoh-Okine, 1994; Huang & Moore, 1997; Roberts & Attoh-Okine, 1998; Bosurgi & Trifiro, 2005; Thube, 2012; Kumar et al., 2013; Karlaftis & Badr, 2015; Ziairi et al., 2016; Gong et al., 2018), pavement condition assessment (Eldin & Senouci, 1995; Lin et al., 2003; Suman & Sinha, 2012; Salini et al., 2015; Hamdi et al., 2017), or to determine the nowcast structural condition through deflections and strains (Meier & Rix, 1995; Ceylan et al., 2004b; Saltan & Terzi, 2008; Gopalakrishnan, 2010; Plati et al., 2016). However, as stated by Ziari et al. (2016): "*finding an efficient model with all effective variables has hardly ever been achieved. Most of the existing models include only few variables and, to the best of our knowledge, none of them analyse all variables*". This motivates a more holistic perspective, tackled almost in parallel, to predict the overall condition of pavements by predicting measurable condition indices and maintenance interventions (Taha & Hanna, 1995; Alsugair & Al-Qudrah, 1998; Attoh-Okine, 2001; Yang et al., 2003; Mosa, 2017; Domitrovic et al., 2018), or even inferring the best combined index to estimate and prioritise maintenance actions (Fwa & Chan, 1993; Goh, 1997; Bianchini, 2012; Yang et al., 2015). For an exhaustive historical review, several publications are available in the literature (TRB, 1999; Sundin & Braban-Ledoux, 2001; Flintsch & Chen, 2004; Ismail et al., 2009; Ceylan et al., 2014a; Sharma & Gupta, 2016). All these pillars constitute the foundation for addressing the optimal planning of the maintenance management of road networks.

The critical literature highlights the need for expert systems that integrate reliable information with effective multiple ML approaches motored by decision support architectures. In the maintenance management of assets, expert systems combined with fuzzy logic reasoning, using simple reports tools, were one of the first approaches tested. There were also prior attempts to use more sophisticated artificial intelligence systems using Bayesian statistics, genetic algorithms or neural networks (Fu, 1995). During the last years many attempts were made to improve the prediction potentiality of asset alerts and the efficiency of the asset management systems, and several recommendations and directions were indicated (Faiz & Edirisinghe, 2009; Marsh et al., 2016). At present, in the topic of asset alert prediction management and asset diagnosis prediction, all the data inferring-based disciplines (i.e. ML, data mining, statistics, big data) are working in the same direction; it is remarkable the numerous commercial software packages devoted to this, which evolved (some of them) from a statistical origin (e.g. SAS, 2014). Even so, the issue is still in an accelerated state of evolution.

The scope of this paper is focused on the prediction of maintenance alerts that might appear in further scenarios, the most probable type of interventions required for preventing the forecast failures, and other similar interventions that might also restore the serviceability of an affected road asset. The goal does not include predicting the evolving nature of the explanatory variables affecting the road infrastructure, which is taken in this communication as external information to the developed methodological scheme.

3. Aim, problem definition and methodology scheme

The aim of the proposed methodology is to predict maintenance alerts and the required interventions for all linear assets of a road network, and to rank the interventions according to the forecast severity of degradation/failure of the assets (i.e. road segments) themselves in specific pre-selected scenarios. This methodology is founded on using the know-how derived from the information recorded in the historical maintenance work-orders repository in order to define the N-most probable candidate interventions to be conducted.

To achieve this goal, several intermediate calculations have to be done. The methodology scheme is outlined in Figure 1, which shows the inputs and outputs. First of all, it is necessary to detect the assets with a degradation that implies the need for a maintenance intervention. For this, a specific approach that estimates whether a maintenance intervention is needed (alert) or not (no-alert) has been developed. These alerts will be associated with a severity level to rank them according to the expected gravity of their condition. This approach incorporates recorded subjective assessments of the Maintenance Managerial Body (MMB) surveyed just before the maintenance interventions took place; this data is processed to be used in order to improve the accuracy of predictions.

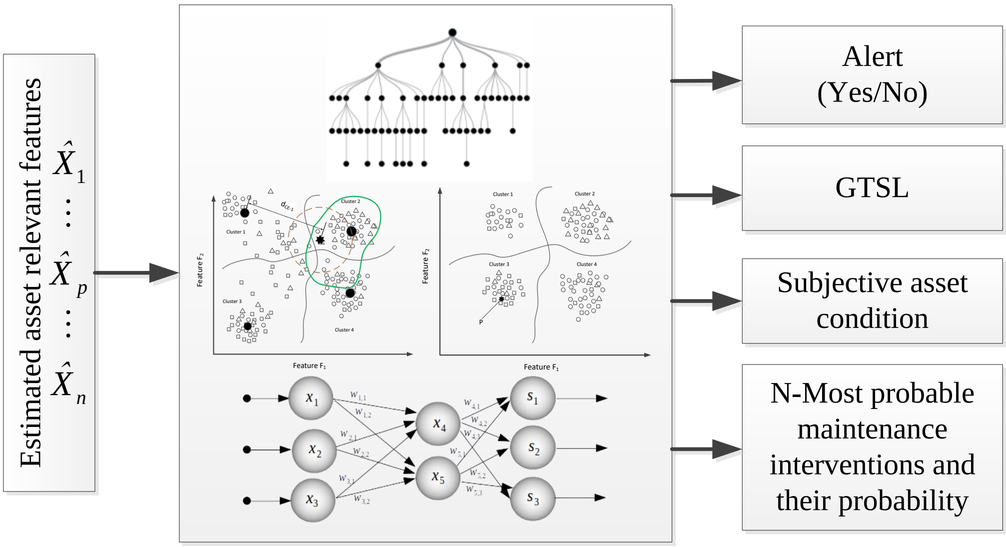

Figure 1. Methodology scheme

The maintenance predictions are inferred by correlating the estimated values of explanatory features (X_p) of the asset's condition, in the required scenario of interest to the user (i.e. the MMB), with the information stored in the historical maintenance repository. This information refers to all historical records corresponding to any asset of the infrastructure, interventions performed and timestamps stored in the repository, in which several data sources are relevant: (a) geometrical and structural measurements, (b) endogenous and exogenous characteristics (i.e. traffic flow, road category, average and distribution of vehicle load, freezing index), (c) maintenance intervention types, (d) assessed condition just before and after the intervention is performed, among other relevant pieces of information.

3.1 Prediction of alerts

In a standard predictive maintenance plan, interventions can be programmed according to the knowledge gathered over time on the basis of historical failures identified in the past (i.e. from similar cases). These may have either a deterministic nature (e.g. one, several or a combination of explanatory features of the asset failure surpassing some standard threshold values), stochastic type (e.g. the features show certain probabilities of surpassing threshold values), or a hybrid deterministic-stochastic character. These thresholds are set by relevant standards for either design, quality or safety parameters (e.g. Normal Limit L_N, Exceptional Limit L_E). Other thresholds affecting the asset condition parameter can also be considered, as in the case of the Alert Limit (AL), Intervention Limit (IL) and the Immediate Action Limit (IAL) indicators. As defined by applicable criteria (e.g. European Standard, Road Administration Standard, Infrastructure Maintenance Managerial Body), the condition of the asset will be quantified in regard to the proximity of the forecast values to the threshold limits, a distance criterion named the technical severity level (TSL). Within this frame, the triggering of a pre-alert is through the technical severity level (TSL) associated with each explanatory feature. When the mentioned thresholds are consolidated on a solid experience (i.e. time), the predictive plan reaches an acceptable confidence level. Nonetheless, there are often cases in which the existing variability of presumed similar cases makes the knowledge insufficient to establish clear rules for identifying explanatory features and their thresholds. In these cases, data-driven approaches based on ML procedures can pave the way to extracting the hidden know-how and make the estimating of alerts, asset conditions and interventions more reliable. This context is dealt with by the proposed methodology, which identifies whether an intervention is needed (alert) or not (no-alert). Additionally, the alerts will be ranked by their global technical severity level (GTSL) in terms of all the forecast features considered as a whole.

The cited GTSL, calculated for each asset, is derived from the technical severity levels (TSL) of each explanatory feature $X \equiv [X_1, \ldots, X_p, \ldots, X_n]^T$. However, the global index (GTSL) may not be referred to any threshold, as it is hidden non-explicit information, but is rated as an absolute value. To proceed, all individual TSLs are normalised ($TSLN_i$) in order to refer all individual features X_i to the same scale and, later, weighted by pre-set values αi (according to the external criteria set by the MMB to give more importance to some features versus others), subject to a constraint $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 1$. Then, a generic expression for the GTSL follows:

$$GTSL = \alpha_1 \cdot TSLN_1 + \ldots + \alpha_n \cdot TSLN_n$$

The prediction of alerts is made using supervised ML techniques. The first phase of the ML model implies a training–testing process based on the information referred to in section 3.3. The data repository, containing the historical maintenance regarding the monitored infrastructure case study, provides all interventions (i.e. raw work orders) which, after being retrieved, are filtered to keep only those records that are valid according to the intervention types relevant for the forecasting. Moreover, measurements carried out on the asset (e.g. auscultation, far or close, prior to the corresponding intervention was initiated and completed), associated with the explanatory features, have to be available.

A concise listing of basic relevant data for training the alert-prediction model is presented in Table 1.

Table 1:Data inputs for training phases (prediction of alerts).

Data input
Asset identification
Values of the assumed explanatory features ($X \equiv [X_1, \ldots, X_p, \ldots, X_n]^T$)
Maintenance intervention executed (1-0: Yes-No)

The ML methodology estimates alerts using a classifier that correlates the values of the explanatory features with the need for maintenance. It consists of an automatic classification technique in a binary variable (1-0: Yes-No). A set of four automatic binary classification models are launched and tested: decision tree (DT), artificial neural network (ANN), k-nearest neighbours (KNN) and support vector machine (SVM), in order to choose the most suitable according to the information contained in the historical repository. The model that yields the best testing results is chosen as the estimator model. The required inputs for the training phase are indicated in the previous paragraph and shown in Figure 2a.

Once the ML models are trained, the only inputs needed to detect assets where maintenance will be required, in a specific further scenario, are the forecast values of all the explanatory features, $\hat{X} \equiv [\hat{X}_1, \ldots, \hat{X}_p, \ldots, \hat{X}_n]^T$ (Figure 2b).

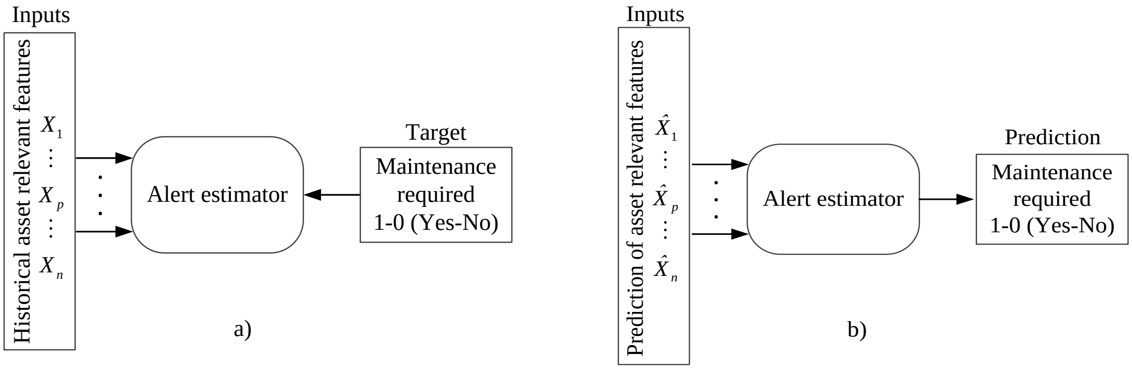

Figure 2. Classifier schemes of alert estimator: (a) training process and (b) prediction process.

3.2 Prediction of subjective asset condition

A very important piece of information to infer the know-how of the MMB is the subjective assessment of the condition of any asset inspected by the maintenance team before the intervention was carried out; in particular, the assessment of each individual explanatory feature (SX_p), each combined explanatory feature (SCX), and a global evaluation (SG) of the condition of the whole asset. This know-how, to be extracted using again a supervised ML, will be used to increase the accuracy of the prediction of the most probable intervention recommended to be carried out on the asset. Table 2 summarises the relevant data for training the prediction model.

Table 2:Data inputs for training phases (prediction of subjective asset condition).

Data input
Asset identification
Values of the assumed explanatory features ($X \equiv [X_1,\ldots,X_p,\ldots,X_n]^T$)
Subjective assessment associated with any of the individual features ($SX \equiv [SX_1,\ldots,SX_p,\ldots,SX_n]^T$)
Subjective evaluations associated with a combination of features (SCX)
Global subjective evaluation associated with the condition of the asset as a whole (*SG*)

By correlating feature measures with the subjective evaluations of the asset condition recorded in the repository, the system "learns" from the MMB know-how and, when a new prediction of explanatory features is introduced in the system, it predicts the asset condition without any extra monitoring/ assessment from the MMB. According to this, the inputs for training are the historical values of the features, and the target variables are the historical subjective evaluations (Figure 3a). Depending on the information provided by the historical maintenance database, it is possible to train the model just with either the subjective evaluations associated with the individual features ($SX \equiv [SX_1,\ldots,SX_p,\ldots,SX_n]^T$), the subjective evaluations associated with a combination of features (SCX) or with the global subjective evaluation associated with the global asset condition (*SG*).

As in the previous case, once the ML models are trained, the inputs needed to predict the subjective asset condition, in a specific further scenario, are the forecast values of all the explanatory features, $\hat{X} \equiv [\hat{X}_1,\ldots,\hat{X}_p,\ldots,\hat{X}_n]^T$ (Figure 3b). The prediction is materialised by estimating a subjective assessment by the triplet ($S\hat{X}, SC\hat{X}, S\hat{G}$), where the first element is a multi-dimensional vector. The global asset condition estimate ($S\hat{G}$) can be seen as the predicted overall state of the asset regarding the simultaneous contribution of all feature effects as a whole.

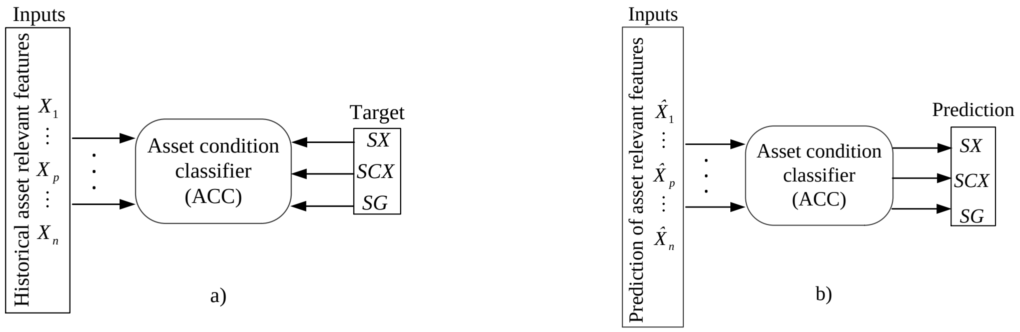

Figure 3. Classifier schemes of asset condition: (a) training process and (b) prediction process.

3.3 Prediction of N-most probable interventions

In order to estimate the most probable intervention type for a specific asset, a process is launched by correlating the estimated condition, together with the estimated subjective condition $(S\hat{X}, SC\hat{X}, S\hat{G})$, versus similar asset samples reported in the historical repository. This is the reason that records associated with samples contained in the historical repository should have assigned information relative to: (a) the nowcast subjective condition (SX, SCX, SG) of the asset just before the historical intervention took place and (b) the recommended technical intervention to be conducted to restore its serviceability and functional condition. This last piece of information is very important due to the model provides predictions regarding the just-technical maintenance interventions required to be conducted in order to fix the alerts, but it does not provide any hint regarding the maintenance finally suggested/performed by the MMB considering other non-technical variables (budget, commercial/exploitation strategy, etc).

The prediction of the intervention type is based on a two-phase ML scheme. The result of the model approach is a hierarchical listing of the most probable interventions that should be conducted, and their corresponding probabilities of occurrence. The first phase uses a clustering technique, the second one a k-nearest neighbours technique (KNN). A description of the techniques and implemented methodologies follows.

<u>*Unsupervised clustering of historical data*</u>

The data available from the historical repository are grouped into different clusters. The application of this approach depends on the quality of the data regarding the existence of more or less well-defined clusters. Assuming this hypothesis, *L* clusters can be inferred from the historical data so that samples with a similar condition (i.e. similar values for all features) are grouped in the same cluster. The clustering is carried out taking into account just the asset condition (i.e. values of features *X*) without the recommended technical intervention. Thus, the same cluster could contain different intervention types as shown in Figure 4, where only two features, X_r and X_s, are considered for the sake of clarity, although the set of explanatory features configures a multi-dimensional space according to the number of variables. For the sake of illustration, the intervention types involved are identified by different plotting shapes (i.e. squares, triangles and circles). Once the clustering is defined, the most probable type of intervention associated with a triggered alert is determined using the estimated values of the explanatory features ($\hat{X}$) as inputs to the model; the location *P* of the asset condition can determine the belonging to a specific *i*-cluster based on the distance to the centroid. From the historical intervention types of the samples implied in the specific cluster, a simple rating analysis among those types yields a probability for each type to be assigned to the estimated alert.

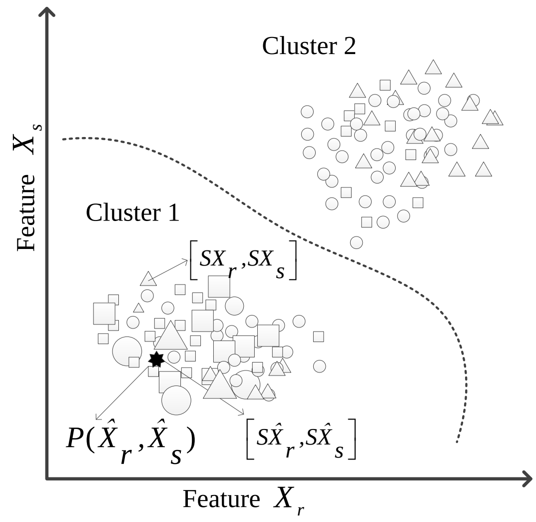

Figure 4. Unsupervised clustering.

To take into account the MMB's know-how related to the subjective assessment of all historical samples implied in the *i-th* cluster, the empirical occurrence of each intervention type can be weighted by scores (represented by different sizes of the shapes in Figure 4). These weights assign more relevance to those samples whose assessment is similar to the predicted subjective evaluation of features ($S\hat{X}, SC\hat{X}, S\hat{G}$) of the asset implied in the triggered alert, as derived by the asset condition classifier (ACC). Regarding the example, Figure 4 shows the cluster that 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 a similar evaluation to that predicted asset will be scored with a higher weight than the rest.

The above methodology has serious limitations when the data do not show a clear grouping; this drawback can be circumvented by using a second phase based on the k-nearest neighbours algorithm (KNN). In this case, the k-nearest historical samples to a given triggered alert are chosen to define the most probable types of intervention associated with an alert. The methodology is quite similar to the previous one, but restricted to k samples. This means that the selected samples, including *P*, belong to the same unique cluster in order to apply the technique described in the first phase. Following the same approach, the underlined know-how can be taken into account by giving more importance to samples with a similar assessment to the one predicted for sample *P*,($S\hat{X}, SC\hat{X}, S\hat{G}$). In the example, each sample occurrence (from the k-nearest cluster) is weighted by a score which depends on its subjective assessment, higher for samples with a 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 technique has been built, summarised in the following steps:

i. Use a clustering technique to generate the proposed cluster model, as outlined in the previous section.
ii. Compute the Euclidean distance between sample *P* and each of the L cluster centroids (ceL): $(d_{P-ce1}, d_{P-ce2}, \ldots, d_{P-ceL})$
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 the *J-th* cluster, $z_{P-Q} = d_{P-Q} \cdot d_{P-ceJ}$, where d_{P-Q} stands for the Euclidean distance between *P* and *Q*. In this way, samples belonging to different clusters are penalised, as is the case of sample *Q* in Figure 5.

The example shows the difference between applying the KNN model (broken circle) and the fusion model (solid closed line). As shown, the fusion model is not as dependent on the existence of a clear data grouping (i.e. cluster definition) and enables the value of *k* to be incremented in a guided way, in regard to samples belonging to the same cluster as *P*. With this single modelling, the MMB's know-how can also be taken into account by giving more importance to samples with the same assessment as the one predicted for sample *P*, $(S\hat{X}, SC\hat{X}, S\hat{G})$.

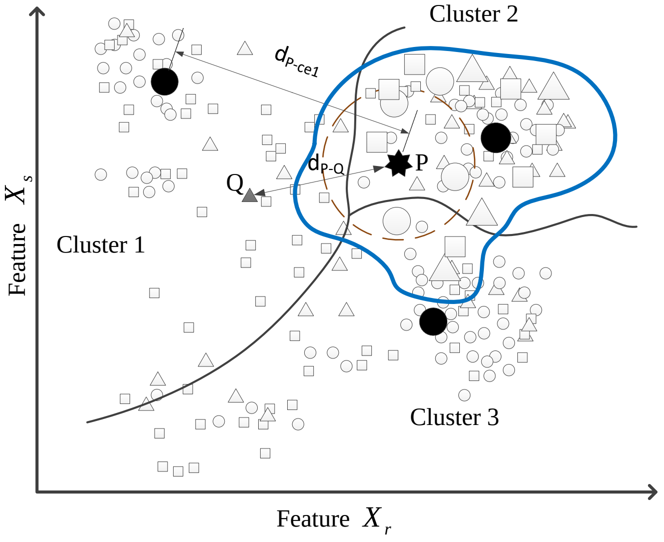

Figure 5. Fusion model.

4. Case study

The empirical case selected is a meshed road network in the central region of Portugal, managed by *Infraestruturas de Portugal*, totalling 620 km; it includes several types of road categories, such as principal itineraries, complementary itineraries, national roads, regional roads and other road classes, which are under its jurisdiction; the pilot case does not include motorways or stretches with major bridges and/or tunnels. Principal and complementary itineraries often consist of a single carriageway with one lane for each direction but with additional lanes for heavy vehicles in locations with higher gradients; national and regional roads can show narrower cross-sections. The studied road system is classified into network classes, or road categories, on the basis of features such as travel speed, traffic volume, traffic mix and strategic relevance; each road is subdivided into road sections according to a multiplicity of characteristics (i.e. geometry, layout). Regarding traffic levels, the chosen demonstration case presents heterogeneity among the selected sub-networks (i.e. itineraries), with between 2,500 and 10,000 vehicles per day, and an average of 9% of heavy vehicles; it also includes locations with high maintenance costs or very frequent interventions.

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

The available historical information consists of two different data type sources. The first one corresponds to the results of different measurement campaigns. The auscultation strategy achieved a network coverage of almost 50% annual reach during the period 2007–2012 and 90% reach since 2012, collecting parameters related to cracking, surface defects (including ravelling, bleeding, polished aggregates and deformation), potholes, patching and rutting, macrotexture and front/rear imaging. The relevant features selected were the International Roughness Index (IRI), the rut depth (RUT) and the surface area with crocodile cracking (CT) expressed by the percentage of affected pavement surface area. Data was captured every 10 m of road and 100% of the network was monitored. The second data type source compiles the interventions carried out from 1933 (oldest construction record) to 2016, from which the period 2012–2016 has been chosen to keep the same level of consistency on maintenance management criteria (e.g. description of operations and interventions performed, laser-based auscultation capturing system).

4.1 Model selection for estimating interventions

This section presents the results after training a set of four ML algorithms. The techniques used are decision trees (DT), k-nearest neighbours (KNN), support vector machines (SVM) and artificial neural networks (ANN). 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). The model performance should be robust to over-fitting, to guarantee reliable predictions. Another objective is to determine the required sample size for each model to be capable of generalizing. The performance of these classification models is assessed by the learning curves and the confusion matrix (Kohavi & Provost, 1998), derived from counting the test records correctly and incorrectly predicted by each classification model. Within this scheme, the relationships between the training set size, model complexity, and prediction error have been analysed and identified. Two separate goals are addressed: (i) model selection by estimating the performance of different models in order to choose the best one; this action includes the calibrating of parameters until the optimal model is chosen; (ii) model assessment by estimating the prediction (generalised) error of the final model on new data. Assessment of this performance is crucial in practice, since it guides the choice of the ML technique, and gives a measure of the quality of the model ultimately chosen.

To perform the optimization of the models (adjusting their complexity), a sample of 2000 intervention records is used. The selection of this sample size is built on experience based on the number of model inputs and possible alerts (outputs). The sample size allowed the inference of models without under-fitting behaviour and which generalised well. The figures describing the optimization for each model are shown in Figure 6, Figure 7 and Figure 8.

In decision tree (DT) methods, the main parameter to calibrate is the number of splits or branches. Figure 6 shows twelve series of cases where the number of splits ranges from 2 to 40. The horizontal axis represents simulations of cross-validation tests with different fold sizes. The first and last simulations divide the sample into 5 and 10 folds, each fold consisting of either 20% or 10% of the sample set, respectively. Based on the results, it is concluded that the size of the folds does not affect the accuracy. As shown, a DT model with 40 splits achieves the same accuracy as one with 28 splits. Other parameters have been calibrated but their influence is not very noticeable.

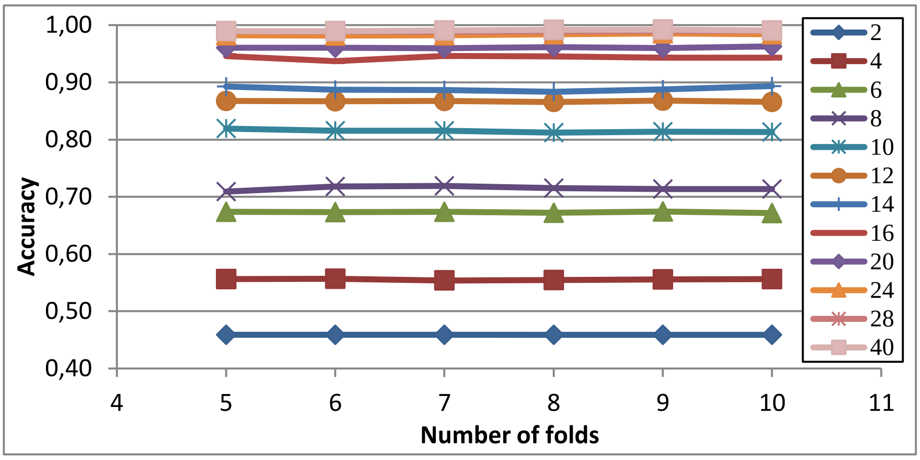

Figure 6. Accuracy of decision tree method versus the number of splits (cases 40 and 28 overlap).

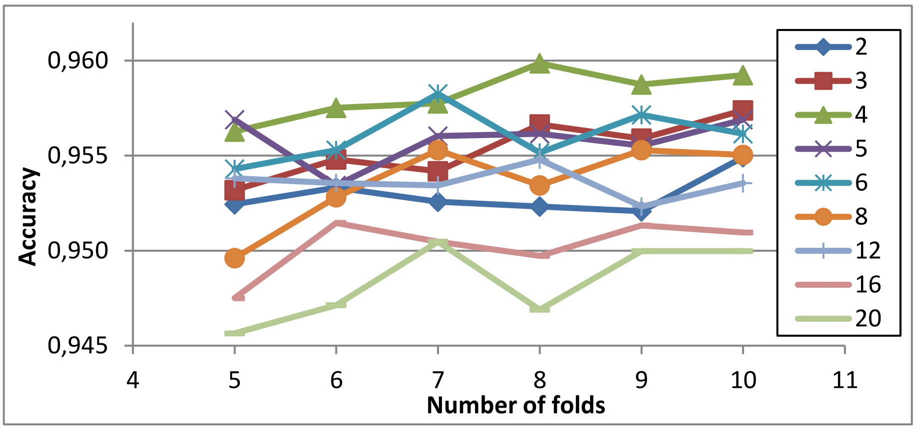

Figure 7. Accuracy of k-nearest neighbours method versus the number of neighbours.

For the k-nearest neighbours method, an object is classified according to the majority contribution of its neighbours. Figure 7 shows a total of nine cases ranging from 2 to 20 neighbours. As in the previous figure, the horizontal axis represents different simulations of cross-validation with different fold sizes. It is noticeable that the structure with four neighbours gives the best performance in spite of the small difference of 1.5% between the best and worst cases. It is observed that the choice of the number of folds into which the sample is divided may improve the final accuracy of the method although up to a maximum of only 0.4%. A comparison with the DT approach shows that in both methods, DT and KNN, the selection of the number of folds is not determinant; there are no significant bias or variance issues. This is because the size of the sample is large enough and all possible sample diversity is collected in both the test and training sets. Therefore, any of the nine cases can be a good candidate.

It has been concluded that the number of folds into which the sample is divided is not an influential parameter; therefore, in the next two methods only a cross-validation testing partitioning the sample into 7 folds is performed.

The SVM method is mainly characterised by its kernel function. Five functions were tested; the accuracy is reported in Table 3, from which the 3-degree polynomial kernel provided the highest value, resulting in the best option (case kfold=7).

Table 3: Accuracy of support vector machine method versus the kernel function.

Kernel function	Accuracy
linear	0.889
quadratic	0.960
polynomial of degree 3	0.969
polynomial of degree 4	0.965
Gaussian	0.948

The artificial neural network (ANN) is a more complex algorithm from the calibrating point of view due to the number of parameters to be configured. Only the design of the number of hidden layers and neurons is addressed.

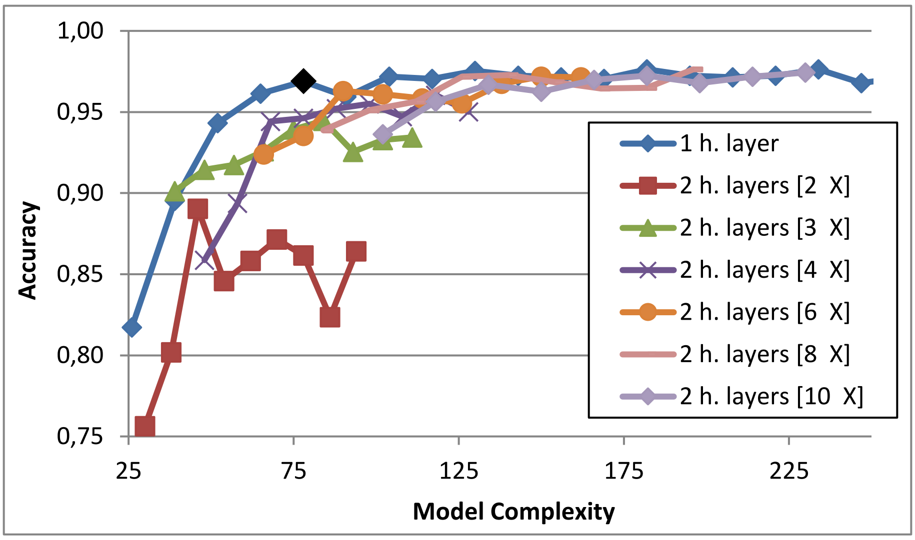

Figure 8. Accuracy of artificial neural network method versus the number of neurons.

Figure 8 depicts seven series of different ANN designs, whose description follows: (a) 1 h-layer, with only one hidden layer and where the number of neurons ranges from 2 to 20; (b) 2 h-layers [2 X], with two hidden layers, where the first hidden layer has 2 neurons and the second one ranges from 2 to 10 neurons; (c) 2 h-layers [3 X], with two hidden layers, where the first layer contains 3 neurons and the second a number of neurons from 2 to 10; and so forth.

In order to compare the different ANN designs, the model complexity parameter is defined as the sum of the weights to be trained in each model. The model complexity increases as the number of neurons and/or hidden layers increases.

As depicted in Figure 8, models with only one hidden layer (1 h-layer) achieve better performance. For complexity greater than 110 there are several models that reach high accuracy. The optimal model is one that achieves high accuracy with less complexity. The lower the complexity, the smaller the sample needed to train the model. The ANN model chosen has one hidden layer with only 6 neurons (black diamond sample in Figure 8) and a complexity of 78.

Table 4 presents the accuracies achieved for all the selected models, wherein the DT model reaches the highest value.

The expected testing error of each estimated model is calculated using learning curves. These plots illustrate the importance of assessing the ability of a learning technique to generalise; thus they reflect both the training and testing errors.

Table 4: Accuracy of selected models.

Model	Accuracy
DT	0.989
KNN	0.960
SVM	0.969
ANN	0.969

The conducted testing procedure of each selected model follows the sequence: (i) a structured set of 1000 samples is generated; (ii) for a fixed training set size (e.g. 50, 100, 150,…,700), several random extractions from the sample set are made; (iii) by following a Monte Carlo method, each extracted set is used for training and testing, and the training and testing errors for this set are obtained. In addition, for each extraction, the 20^{th} and 80^{th} percentiles are computed.

The learning curves of the selected models are shown in Figure 9, where a total of six plots are depicted; the x-axis represents different training set sizes and the y-axis the error obtained. The two thick lines represent the average value of the training and test sets; the thin lines represent the 20th and 80th percentiles of the training and test sets and provide an insight on the variance of the predictions.

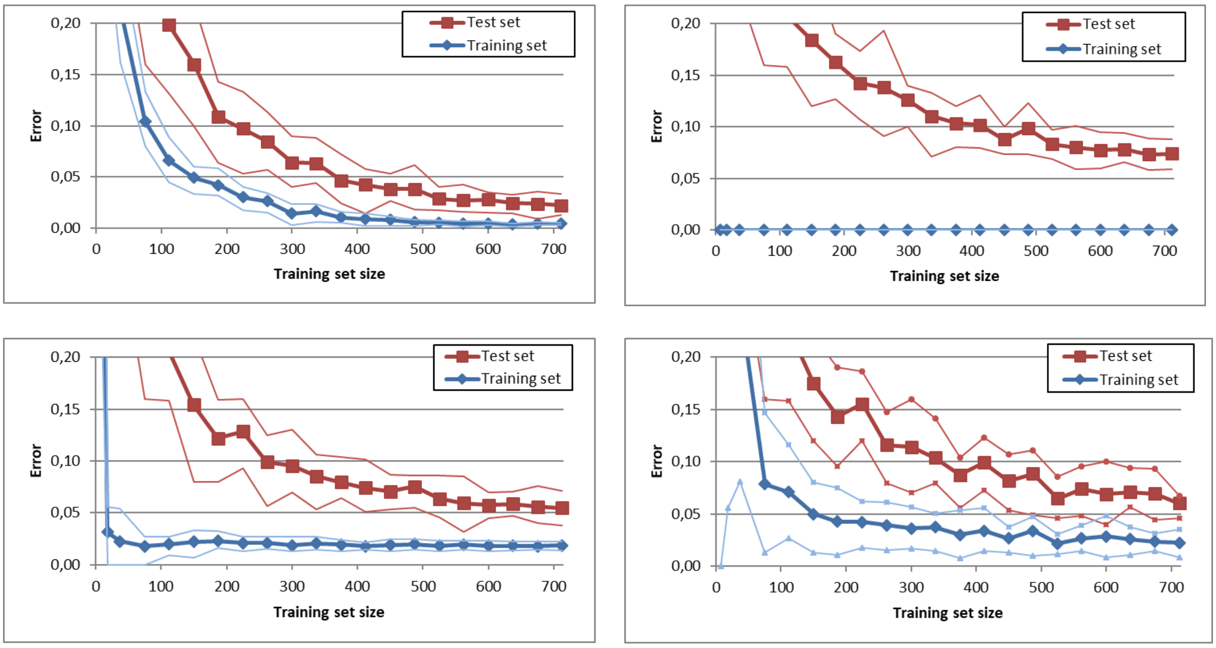

Figure 9. Learning curves: (a) decision tree, (b) k-nearest neighbour models, (c) support vector machine and (d) artificial neural network models.

As the size of the training set gets larger, the lines converge toward an asymptote representing the amount of irreducible error in the data. The error of the training-set prediction in the k-nearest neighbours model is always 0%. This is due to the fact that the prediction of a record is calculated from a neighbourhood wherein the record is included. A summary of the most relevant conclusions deduced from the learning curves shown is tabulated in Table 5.

Table 5: Summary of results.

	DT	KNN	SVM	ANN
Final test set average error	2.21%	7.39%	5.48%	6.01%
Final difference between testing and training average error	1.8%	7.39%	3.62%	3.76%
Size of training set to reach a test error < 5%	375	-	-	-

The DT model achieves the smallest error. The SVM, ANN and KNN models yielded similar results; none of them reduced the test set average error below 5%. Only the DT model reached a test set average error below 5% using a training set size of 375 records.

Attending to the evolution of the test-set error curves, the error values are still decreasing over a training set size of 700. Therefore, smaller errors would be expected for larger training sets. Although it is clearly noticeable, the flatness of the curve trends is almost horizontal; pursuing an additional error decrease might need the availability of a larger set of records.

It is worth underlining that the results presented correspond to a medium-rich data case. In real practice, once the models are designed, other issues arise. As a general rule, the historical maintenance database provides a limited valid sample set due to various causes: (a) few data on interventions are recorded with sufficient rigour; (b) the information stored lacks chronological consistency (i.e. actual interventions may not be described with the same rigour along time, the condition before and after interventions is not consistently assessed by maintenance personnel); (c) insufficient physical explanatory features are quantified. Therefore, it is important that the selected models achieve a low average test-set error with a limited number of records. In this case, the DT model is the best option, as it needs a smaller number of simulations to generalise the problem.

4.2 Statistical treatment of the case study

Once the ML models are trained, the only inputs needed for the correct functioning of the methodology are the forecast values of all the explanatory features, $\hat{X} \equiv [\hat{X}_1,\ldots,\hat{X}_p,\ldots,\hat{X}_n]^T$, which are IRI, RUT and CT in our case study. So far, these predicted features have been treated in a deterministic way; however, as predictions, they actually have a stochastic nature and will have associated uncertainties that quantify and assess the probabilities of the estimated values.

In the methodology tested, alerts and interventions are predicted together; however, the inputs are not exactly the same for both if the stochastic nature of the features is taken into account. This is because alerts have to be identified taking into account the probability of failure, stating the specification of the desired level of alert reliability. However, in order to specify the intervention, the most probable asset condition should be handled, instead of the condition provided by the desired level of reliability; this is because, in general, the forecast asset conditions will be a more realistic and reliable piece of information. In this way, the alert is triggered under the condition of only permitting a pre-specified probability of failure γ, but the forecast interventions associated with this alert will be calculated using the most probable asset condition (e.g. the expected value, said mean, of the forecast feature density function). To clarify this issue, Figure 10 shows an example involving just one feature with an evolving behaviour according to two different probability density functions ("pdf 1" and "pdf 2"). The solid line, named "Expected value", stands for the feature value with a 0.5 probability (in both pdfs); the middle broken line exemplifies the feature's evolving values under a hypothesis of $R=1-\gamma$ reliability under the probability density function "pdf 1" (stretched bell curve); the upper broken line represents the presumed feature value under the hypothesis of $R=1-\gamma$ reliability and the "pdf 2" probability density. As shown, the probability density "pdf 2" (upper broken line) implies that the alert is triggered before the case of probability density "pdf 1" (short bell curve), for identical reliability level R, because the value used to define the alert is higher. This can be explained by the fact that the dispersion of "pdf 2" is larger and it is more probable that higher feature values come into play.

On the other hand, the MMB may design maintenance plans based on the most probable interventions taking place, instead of the most severe ones and it is for that reason that the expected value is used regardless of the probability density function.

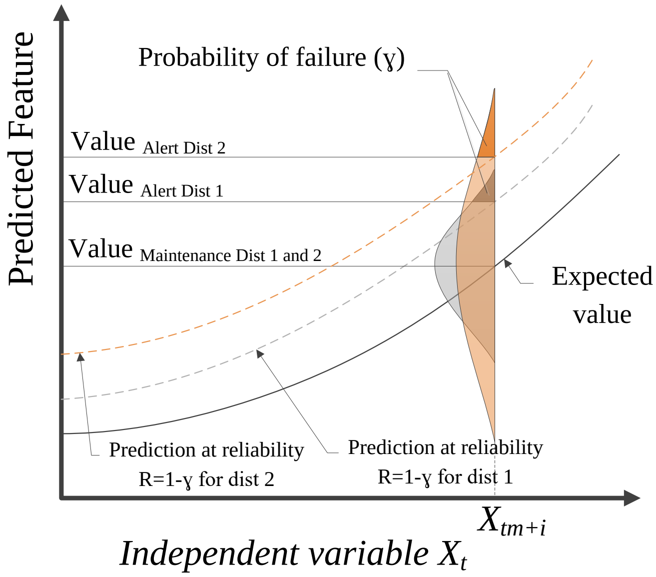

Figure 10. Example of dealing with one feature under the hypothesis of two different probability density functions.

5. Results and estimations

5.1 ML model comparison

Once all the models were calibrated, a comparison between them was performed. The scenario base to compare with is a technical state (technical-based ideal maintenance plan). It is not possible to define a specific baseline to compare with since the final maintenance depends on a multiplicity of external variables taken into account by the MMB in different situations (e.g. budget, availability of maintenance teams, intervention time, and infrastructure availability time window). Currently, although any actual alert requires a specific type of intervention (technical-based intervention), the planner may not select such intervention on the basis of the external variables. Because of this, in the case where a current maintenance plan is used as a baseline, many types of errors may arise in the comparison with the predictions provided by any predicting model, as the models provide the maintenance required to fix the alerts, but they do not provide any hint regarding the maintenance finally performed considering the external variables.

So, the technical state used as baseline is based on the maintenance plan generated by the MMB relaxing some of the constraints that exist in any maintenance exploitation management of all transport infrastructures:

- Limited budget constraint relaxation: Each asset will be maintained with the maintenance intervention type that best fixes the alert without the constraint of limiting/reducing costs (by aggregating several assets, for instance).
- Intervention time constraint relaxation: Maintenance interventions can be performed without limitations on intervention times.
- Infrastructure time-window constraint relaxation: Availability of the infrastructure is full. Any maintenance intervention can be carried out on any asset in a frame free of constraints affecting the availability of the infrastructure network.

- Limited resources constraint relaxation: The maintenance interventions can be proposed without taking into account the actual limitation in resources (equipment, material, staff…).

The results obtained for the free-of-constraint case correspond to the optimal predictions from a purely technical point of view. That is, all assets will be repaired in accordance with the recommended predicted interventions in order to reach a health condition of the network.

The results provided by the different models are shown below. The graphs are generated considering 2016 as the scenario of reference. For this year, the DT, KNN, SVM and ANN models have been applied and the predictions are compared with the target ideal scenario (for the same year). This ideal scenario is always considered as a bound for the models (the historical records have been labelled with the technical maintenance intervention by the MMB according to the previously described relaxed constraints).

The comparison is carried out based on the false positive; in aggregate (total number of alerts, Figure 11) and disaggregate terms (by maintenance intervention type, Figure 12). A total of 1257 assets were involved.

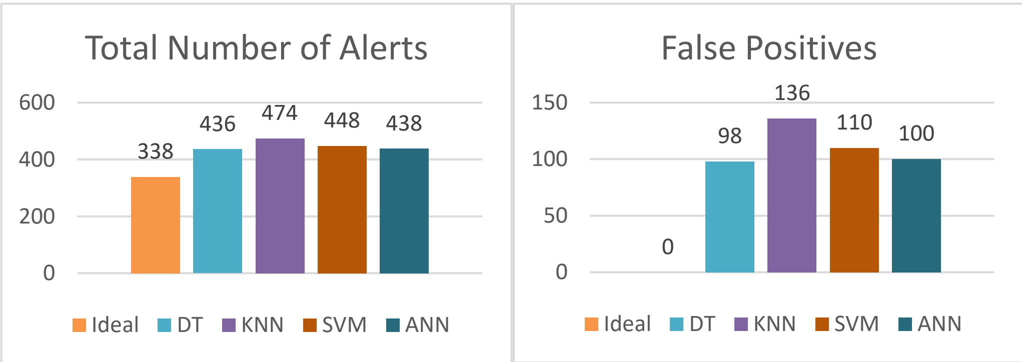

Figure 11. Model assessment: aggregate method.

As shown in Figure 11, the DT model has a lower false-positive ratio and, because of this, the total number of predicted alerts is the closest to the ideal state (as the false-negative ratio is the same for all models and equal to 0). The ANN model, next best after DT, is the most accurate approach, followed by SVM and KNN. This ranking does not coincide with the conclusion derived by looking at the mean value of the test set error presented in the previous section (Figure 9). This is because the results shown in section 4.1 were obtained in the training process, and the results in this section were derived from applying the trained models to a real pilot case and they are compared with the ideal maintenance plan (technical-based) previously described.

In order to check those cases where the models present lower accuracy, the total number of predicted alerts has been disaggregated by maintenance intervention types (Figure 11). According to the values in the figures, the models overestimate almost all maintenance interventions; this behaviour was unintentionally sought because false positives are preferred to false negatives. The most severe cases affect T2 and T3 types because, for all models, the amount of predicted interventions is larger than in the ideal case (almost all false-positives predictions are concentrated in these two types of interventions). This situation will be reduced as the system automatically learns from new data in which this type of maintenance intervention might be better represented. For example, for year 2016, there were only 28 maintenance interventions of type T3 (the same proportion is presented in the whole historical maintenance database used for the training process). As the size of the database increases, if more T3-type maintenance interventions are carried out, the system will be self-adapted by learning.

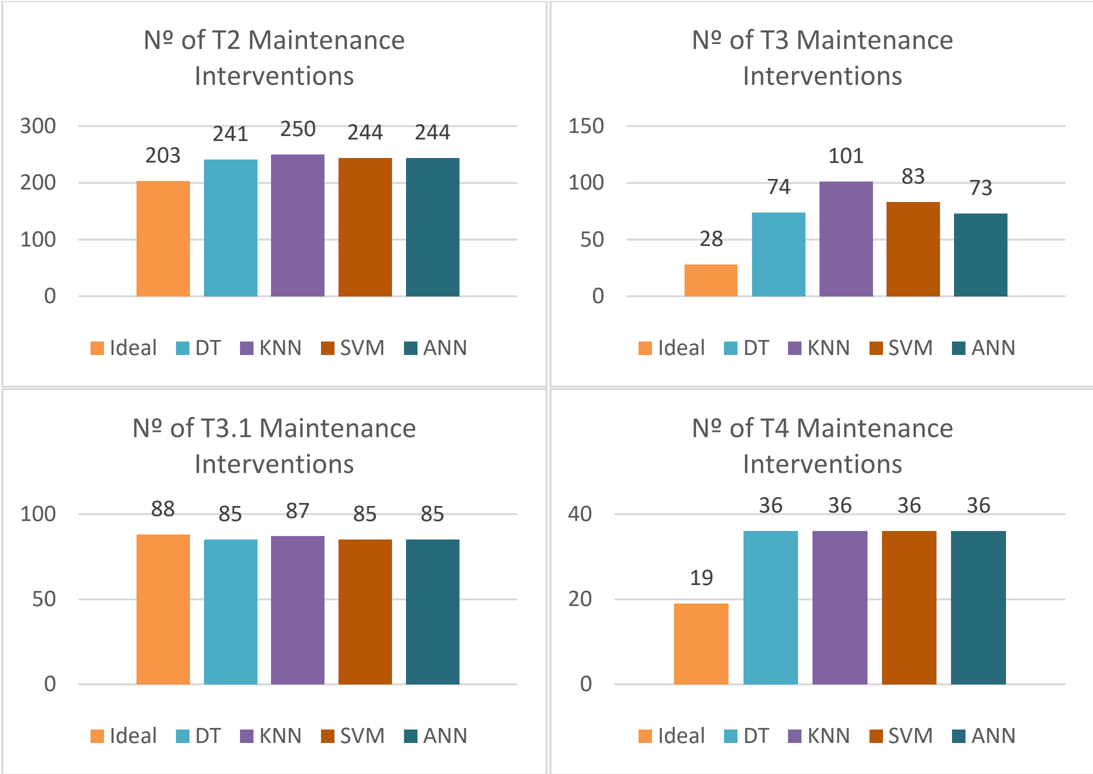

Figure 12. Model assessment: disaggregate method.

In order to characterise more deeply the selected model (DT), the methodology is applied to a real case of a road network comprising 1241 segments of 500 meters. In 1159 segments the predicted value coincides with the real one, obtaining a final accuracy of 93.4%. Comparing the predicted work orders (WO) and the real ones in all segments, 843 of 859 segments 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 predicts T2 (only 79.1% of accuracy) and T3.1 (acc. 86%).

5.2 Results inferred from empirical data

Figure 13 shows a combined colour and scale-symbol theme map that exposes the results obtained in the case study. The conditions, indicated with the GTSL value by the line thickness, of all road segments of the pilot case in 2015 and the most recommended intervention proposed by the methodologies (line colour) can be seen. The box contains a description of the different interventions managed by the MMB; case T0 indicates that no intervention is needed as the segment is in good condition with a low severity level; case T1 indicates that the predicted condition of the segment presents a not null severity level, an alert is triggered, but the alert-prediction model recommends no intervention to be conducted; the rest of cases, T2 to T4, describe the interventions recommended by the model to be carried out.

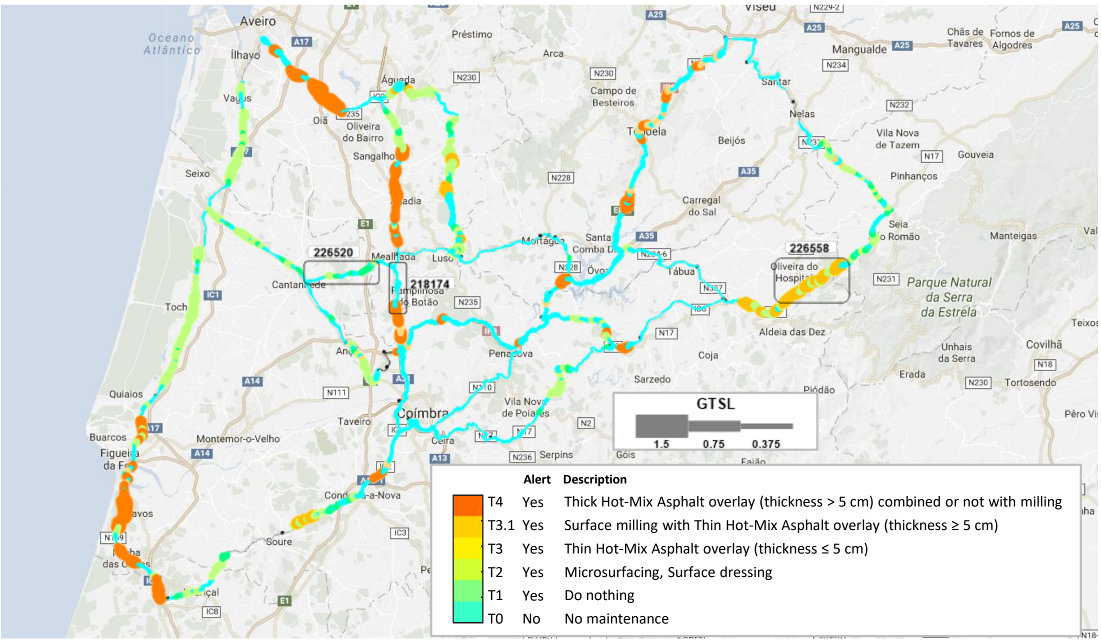

Figure 13. Results obtained in the pilot case in 2015.

¡Error! No se encuentra el origen de la referencia. Table 6 shows the predictions of GTSL and alerts (red cells) for road section 218174, partitioned into segments of 500 meters ; each segment is identified by the initial and final local kilometre points. As previously explained, the methodology provides GTSL values based on the values of IRI, RUT and CT. In parallel, the methodology generates the associated alerts for each segment of the road, revealing those where maintenance is needed by means of triggering alerts. As can be seen for the year 2012, focusing on the values of GTSL, some segments have reached an inadequate asset condition (triggering the corresponding alerts). As no maintenance is executed, these inadequate levels remain in a similar (or even worse) order of magnitude in the next year. Based on the new measurements for the following years (2014–2016), the methodology detects a maintenance intervention between years 2013 and 2014, for segments 0 to 7, since there are no alerts and the GTSL reflects that the asset conditions have been restored to adequate levels. This fact has been contrasted with the historical maintenance database and the detection is correct.

Table 6: Results of GTSL and alerts for road section 218174, in time period 2012–2016.

	Global technical severity level (GTSL)				
Segment	2012	2013	2014	2015	2016
0	0.36	0.35	0.29	0.31	0.40
0.5	0.37	0.65	0.29	0.32	0.34
1	0.60	0.69	0.32	0.44	0.43
1.5	0.74	0.65	0.45	0.66	0.60
2	0.48	0.48	0.28	0.37	0.45
2.5	1.14	1.05	0.19	0.21	0.23
3	1.08	1.25	0.20	0.23	0.23
3.5	0.80	1.05	0.27	0.35	0.41
4	0.56	0.76	0.19	0.21	0.23
4.5	0.34	0.59	0.16	0.17	0.19
5	0.50	0.53	0.20	0.22	0.29
5.5	0.97	0.56	0.17	0.19	0.20
6	0.51	0.33	0.22	0.23	0.33
6.5	0.38	0.29	0.21	0.22	0.23
7	0.78	0.64	0.25	0.26	0.28
7.5	0.48	0.93	0.72	1.04	1.06
8	0.31	0.98	0.80	1.02	0.86

A last issue worth highlighting focuses on unattended alerts. As can be seen in the last two segments on the analysed road, alerts are predicted; however, they were not attended to and the methodology continues to generate alerts there.

Another case derived from the results is analysed in Table 7. In this case, it is straightforward to see that the GTSL has significantly decreased since 2015, but no maintenance intervention has been reported in the historical data. After checking this point with the MMB, a non-reported intervention was actually detected. This was discovered by looking at the videos of inspections over the road. In this regard, it is necessary to clarify that sometimes missing/erroneous pieces of information are present in the historical maintenance repository and not all the maintenance/work done on each piece of road is properly reflected. Even though it is mandatory for major interventions to be registered, these are less numerous than small corrective operations as major maintenance plans are conducted in intervals of 5–10 years; minor maintenance is also executed on roads, normally in a few meters, and frequently only on one of the lanes. In general, minor maintenance is executed by contractors under 3-year contracts for all aspects of road maintenance (not only pavement). This type of intervention is seldom registered in the historical data, except if there is a more significant extent.

Table 7: Results of GTSL and alerts for road section 226520 in time period 2012–2016.

	Global technical severity level (GTSL)				
Segment	2012	2013	2014	2015	2016
0	0.49	0.48	0.67	0.40	0.49
0.5	0.55	0.52	0.29	0.27	0.26
1	1.10	1.10	0.98	0.28	0.28
1.5	1.10	1.11	0.92	0.28	0.30
2	1.09	1.09	1.01	0.26	0.26
2.5	0.65	0.71	0.61	0.30	0.28
3	0.57	0.42	0.51	0.69	0.62
3.5	1.17	1.06	1.19	0.59	0.63
4	0.76	0.61	0.82	0.53	0.71
4.5	0.26	0.25	0.45	0.33	0.60
5	0.33	0.31	0.45	0.39	0.41

There may be other situations (Table 8) where the GTSL estimates a bad road condition, such as the case of roads with low traffic flow (less important road), the situation when the IRI level is rather high but accompanied by low cracking (CT), or when the IRI is low and CT high. In all these cases the segments are under low-priority maintenance based on diverse factors (e.g. low traffic, budget restrictions, other considerations).

Table 8: Results of GTSL and alerts for road section 226558 in time period 2012–2016.

	Global technical severity level (GTSL)				
Segment	2012	2013	2014	2015	2016
0	1.23	0.99	0.97	0.65	1.10
0.5	1.15	1.16	1.07	0.95	0.86
1	1.19	1.30	1.23	1.17	1.36
1.5	1.22	1.27	1.25	1.22	1.28
2	0.63	1.11	1.07	1.10	1.15
2.5	0.99	1.26	1.15	1.18	1.29

6. Conclusions and prospective research lines

This paper presents an ML scheme to predict maintenance alerts and the required interventions in road linear asset infrastructures in further scenarios. A road network pilot case was used to contrast the techniques' capabilities by cross-checking actual conducted maintenance with predictions. The methodology, based on several supervised (decision tree, nearest neighbours, neural network and support vector) and unsupervised ML techniques, has substantiated the optimum choice of the best predictive models based on historical intervention work-orders, asset features and measurement auscultations.

The main predicted outcomes are: (a) the estimated intervention type for each road segment and the probability of occurrence, (b) a sorted listing of estimated alerts according to the global technical severity level. Each prediction set is referred to scenarios identified by its time-stamp. A cross-checking analysis is presented between actual conducted maintenance versus predictions.

The developed ML scheme constitutes one of the components of a smart decision support tool to derive intervention plans. This step is of the utmost interest for generalising a full range of planning maintenance at operational, tactical and strategic level, within an expert intelligent scheme.

The methodologies and results presented herein are far from being exhaustive and conclusive, but they pave the way towards several open research lines: (a) sensitivity to the quality of intervention description in the historical repository, (b) importance of the detailed/undetailed description of the asset condition prior to the intervention, among other open issues.

Acknowledgements

This study has received funding from the 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 (MINECO) for the partial subsidy granted under the national R&D program (TRA2015-65503) and the Torres Quevedo Programme (PTQ-13-06428). The authors acknowledge Infraestruturas de Portugal (IP) for making available the database used in this research. The content reflects only the authors' view and it is stated that neither the EU nor MINECO is liable for any use that may be made of the information contained therein.

Disclosure statement

No conflict of interest was reported by the authors.

References

Alsugair A.M., Al-Qudrah A.A., 1998. Artificial neural network approach for pavement maintenance. Journal of Computing in Civil Engineering, 12(4), 249-255.

Attoh-Okine N.O., 1994. Predicting roughness progression in flexible pavements using artificial neural networks. In Proceedings of the Third International Conference on Managing Pavements. San Antonio, Texas. Conference Proceedings, 1, 55–62.

Attoh-Okine N.O., 2001. Grouping pavement condition variables for performance modelling using self-organizing maps. Computer-Aided Civil and Infrastructure Engineering, 16 (2), 112-125.

Bentley Exor. https://communities.bentley.com/products/assetwise/exor, (last accessed on 05-February-2020).

Bentley Optram. Bentley Systems Inc. USA, https://www.bentley.com/en/capabilities/apm/analyze-rail-track, (last accessed on 05-February-2020).

Bianchini A., 2012. Fuzzy representation of pavement condition for efficient pavement management. Computer-Aided Civil and Infrastructure Engineering, 27, 608–619.

Bosurgi G., Trifiro F., 2005. A model based on artificial neural networks and genetic algorithms for pavement maintenance management. International Journal of Pavement Engineering, 6(3), 201–209.

Ceylan H., Bayrak M.B., Gopalakrishnan K., 2014a. Neural networks applications in pavement engineering: A recent survey. International Journal of Pavement Research and Technology, 7(6), 434–444.

Ceylan H., Guclu A., Tutumluer E., Thompson M.R., 2004b. Use of Artificial Neural Networks for analyzing full depth asphalt pavements. 2004 FWD Users Group Meeting, University Inn, West Lafayette, Indiana.

CSU. Paver. Colorado State University. http://www.paver.colostate.edu/, (last accessed on 05-February-2020).

Dekker R., 1996. Applications of maintenance optimization models: a review and analysis. Reliability Engineering & System Safety, 51(3), 229-240.

Domitrovic J., Dragovan H., Rukavina T., Dimter S., 2018. Application of an artificial neural network in pavement management system. Technical Gazette 25, Suppl. 2, 466-473.

Eldin N.N., Senouci A.B., 1995. A pavement condition-rating model using backpropagation neural networks. Microcomputers in Civil Engineering, 10(6), 433-441.

Faiz R.B., Edirisinghe E.A., 2009. Decision making for predictive maintenance in asset information management. Interdisciplinary Journal of Information, Knowledge, and Management, 4, 23-36.

FHWA, 2006. Pavement Preservation Compendium II, Federal Highway Administration.

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.

Flintsch G.W., Chen C., 2004. Soft computing applications in infrastructure management. Journal of Infrastructure Systems, 10(4), 157-166.

Fu L.M., 1995. Knowledge-based neural networks. Knowledge-Based Systems: Special Issue, 8(6).

Fwa T.F., Chan W.T., 1993. Priority rating of highway maintenance needs by neural networks. Journal of Transportation Engineering, 119(3), 419–32.

Goh A.T.C., 1997. A hybrid neural network based pavement management system. Road and Transport Research, 6(4), 62-71.

Gopalakrishnan K., 2010. Effect of training algorithms on neural networks aided pavement diagnosis. International Journal of Engineering, Science and Technology, 2(2), 83-92.

Gong H., Sun Y., Mei Z., Huang B., 2018. Improving accuracy of rutting prediction for mechanistic-empirical pavement design guide with deep neural networks. Construction and Building Materials, 190, 710–718.

Hajek J.J., Chong G.J., Haas R.C., Phang W.A., 1987. Knowledge-based expert system technology can benefit pavement maintenance. In Transportation Research Record 1145, 37-47.

Hamdi, Hadiwardoyo S.P., Correia A.G., Pereira P., Cortez P., 2017. Prediction of surface distress using neural networks. AIP Conference Proceedings, 1855, 040006.

HDMGlobal. HDM-4, http://www.hdmglobal.com/ (last accessed on 05-February-2020).

Heller. https://www.heller-ig.de/asset-management/asset-management-im-strassenwesen/. (last accessed on 05-February-2020).

HIMS. http://himsltd.com/products.html (last accessed on 05-February-2020).

Huang Y., Moore R.K., 1997. Roughness level probability prediction using artificial neural networks. Transportation Research Record, 1592, 89–97.

IBM. IBM Maximo, https://www.ibm.com/es-es/marketplace/maximo (last accessed on 05-February-2020).

Infor. Infor Corporation., USA, https://www.infor.com/products/eam (last accessed on 05-February-2020).

Ismail N., Ismail A., Atiq R., 2009. An overview of expert systems in pavement management. European Journal of Scientific Research, 30(1), 99-111.

Karlaftis A.G., Badr A., 2015. Predicting asphalt pavement crack initiation following rehabilitation treatments. Transportation Research C, 55, 510-517.

Kohavi R., Provost F., 1998. Glossary of terms. Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. Machine Learning, 30, 271-271.

Kumar S. A., Rao V.R.V., Kumar S.A., 2013. Road roughness modelling with clustered data using ANN approach. International Journal of Civil and Structural Engineering, 4(1), 21-35.

Lin J.D.L., Yau J.T., Hsiao L.H., 2003. Correlation analysis between international roughness index (IRI) and pavement distress by neural network. In: 82nd annual meeting of the Transportation Research Board, Washington, D.C.

Loc8. https://www.loc8.com/, (last accessed on 02-February-2020).

Marsh W., Nur K., Yet B., Majumdar A., 2016. Using operational data for decision making: a feasibility study in rail maintenance. Safety and Reliability, 36(1), 35-47.

Meier R.W., Rix G.J., 1995. Backcalculation of flexible pavement moduli from dynamic deflection basins using artificial neural networks. In Transportation Research Record No. 1473, TRB, National Research Council, Washington, D.C., pp. 72-81.

Morales F.J., Reyes A., Caceres N., Romero L. M., Benitez F.G., Morgado J., Duarte E., Martins T., 2017. Historical maintenance relevant information roadmap for a self-learning maintenance prediction procedural approach. IOP Conf. Series: Materials Science and Engineering. 236, 012107, doi:10.1088/1757-899X/236/1/012107.

Mosa, A.M., 2017. Neural network for flexible pavement maintenance and rehabilitation. Applied Research Journal, 3(4), 114-129.

Moubray J., 1998. Introduction to Reliability-centered Maintenance. USA: TWI Press Inc.

Oracle EAM. http://www.oracle.com/us/products/applications/sb-oracle-eam-executive-2530019.pdf (last accessed on 05-February-2020).

Plati C., Georgiou P, Papavasiliou V., 2016. Simulating pavement structural condition using artificial neural networks. Structure and Infrastructure Engineering, 12(9), 1127-1136.

Reyes A., Morales F.J., Caceres N., Romero L. M., Benitez F.G., Morgado J.; Duarte E., Martins T., 2018. Automatic prediction of maintenance intervention types in transport linear infrastructures using machine learning.Transport Research Arena TRA2018, Paper ID 10392,Vienna, Austria, 16-19 April 2018.

Ritchie S.G., 1987. Expert systems in pavement management. Transportation Research, 21 A, 145-152.

Ritchie S.G., Yeh C., Mahoney J. P., Jackson N.C., 1987. Surface condition expert system for pavement rehabilitation planning. Journal of Transportation Engineering, 113(2), 155-167.

Roberts C.A., Attoh-Okine N.O., 1998. A comparative analysis of two artificial neural networks using pavement performance prediction. Computer-Aided Civil and Infrastructure Engineering, 13, 339–348.

Ross T., Verzi S., Shuler S., McKeen G., Schaefer V., 1992. PARES-An expert system for preliminary flexible pavement rehabilitation design. Transportation Research Record, 1374, 81-89.

Salini R., Xu B, Lenngre C.A., 2015. Application of artificial intelligence for optimization in pavement management. International Journal of Engineering and Technology Innovation, 5(3), 189-197.

Saltan M., Terzi S., 2008. Modelling deflection basin using artificial neural networks with cross-validation technique in backcalculating flexible pavement layer moduli. Advances in Engineering Software, 39, 588–592.

SAP. https://www.sap.com/products/linear-asset-management.html (last accessed on 05/02/2020).

SAS. 2014. Predictive Asset Maintenance. SAS Institute Inc. World Headquarters.

Sharma S., Gupta R., 2016. Soft computing applications in infrastructure management. Imperial Journal of Interdisciplinary Research, 2(7), 188-195.

Suman S.K., Sinha, S., 2012. Pavement condition forecasting through artificial neural network modelling. International Journal of Emerging Technology and Advanced Engineering, 2(11), 474–478.

Sundin S., Braban-Ledoux C., 2001. Artificial intelligence-based decision support technologies in pavement management. Computer-Aided Civil and Infrastructure Engineering, 16, 143-157.

Taha M.A., Hanna A.S., 1995. Evolutionary neural network model for the selection of pavement maintenance strategy. Transportation Research Record, 1497, 70-76.

Tandon R.P., Sinha K.C., 1987. An expert system to estimate highway pavement routine maintenance work load. Civil Engineering Systems, 4, 206-213.

Thube D.T., 2012. Artificial neural network (ANN) based pavement deterioration models for low volume roads in India. International Journal of Pavement Engineering, 5, 115–120.

TRB A2K05(3), 1999. Use of artificial neural networks in geomechanical and pavement systems. Transportation Research Circular E-C012, Transportation Research Board, NRC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec012.pdf (last accessed on 05/02/2020.

Yang J., Lu J., Gunaratne M., Xiang Q., 2003: Forecasting overall pavement condition with neural networks: Application on Florida Highway Network. Transportation Research Record: Journal of the Transportation Research Board, 1853. 3–12.

Yang C., Remenyte-Prescott R., Andrews J.D., 2015. Pavement maintenance scheduling using genetic algorithms. International Journal of Performability Engineering, 11(2), 135-152.

Ziari H., Sobhani J., Ayoubinejad J., Hartmann T., 2016. Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods. International Journal of Pavement Engineering, 17(9), 776-788.