International Journal of Pavement Engineering

ISSN: 1029-8436 (Print) 1477-268X (Online) Journal homepage: https://www.tandfonline.com/loi/gpav20

Machine learning approach for pavement performance prediction

Pedro Marcelino, Maria de Lurdes Antunes, Eduardo Fortunato & Marta Castilho Gomes

To cite this article: Pedro Marcelino, Maria de Lurdes Antunes, Eduardo Fortunato & Marta Castilho Gomes (2019): Machine learning approach for pavement performance prediction, International Journal of Pavement Engineering, DOI: 10.1080/10298436.2019.1609673

To link to this article: https://doi.org/10.1080/10298436.2019.1609673

Published online: 10 May 2019.

INTERNATIONAL JOURNAL OF PAVEMENT ENGINEERING
https://doi.org/10.1080/10298436.2019.1609673

Machine learning approach for pavement performance prediction

Pedro Marcelino [a], Maria de Lurdes Antunes [a], Eduardo Fortunato [a] and Marta Castilho Gomes [b]

[a]National Laboratory for Civil Engineering, Lisbon, Portugal; [b]CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

ABSTRACT
In recent years, there has been an increasing interest in the application of machine learning for the prediction of pavement performance. Prediction models are used to predict the future pavement condition, helping to optimally allocate maintenance and rehabilitation funds. However, few studies have proposed a systematic approach to the development of machine learning models for pavement performance prediction. Most of the studies focus on artificial neural networks models that are trained for high accuracy, disregarding other suitable machine learning algorithms and neglecting the importance of models' generalisation capability for Pavement Engineering applications. This paper proposes a general machine learning approach for the development of pavement performance prediction models in pavement management systems (PMS). The proposed approach supports different machine learning algorithms and emphasizes generalisation performance. A case study for prediction of International Roughness Index (IRI) for 5 and 10-years, using the Long-Term Pavement Performance, is presented. The proposed models were based on a random forest algorithm, using datasets comprising previous IRI measurements, structural, climatic, and traffic data.

ARTICLE HISTORY
Received 19 November 2018
Accepted 15 April 2019

KEYWORDS
Machine learning; pavement performance models; pavement management systems (PMS); time series forecasts; international roughness index (IRI); predictive maintenance

Introduction

Performance prediction models are a critical element of network-level pavement management systems (PMS) (Uddin et al. 2013). These models relate a selected performance indicator to a set of causal variables (e.g. age, traffic load, and material properties), predicting pavements future condition. Inaccurate predictions are a problem because they can compromise budget allocation. Accordingly, effective and efficient maintenance practices require accurate models for pavement performance prediction (Haas et al. 2015).

There is a growing body of literature exploring machine learning techniques for pavement performance prediction. In this paper, a literature survey was conducted to investigate the applications of machine learning techniques in current research on pavement performance prediction. The following conditions were defined to limit the collection of articles:

- Most of the scholarly articles came from the Google Scholar database, which index research articles and abstracts from most major academic publishers and repositories worldwide. Due to its importance, a set of articles selected by the TRB ABJ70 Committee (Artificial Intelligence and Advanced Computing Applications) were also included in the survey.
- Two types of keywords were combined. The first type refers to search term related to machine learning ('machine learning', 'artificial intelligence', 'soft computing', 'data mining', 'knowledge discovery') and specific techniques ('artificial neural networks', 'classification and regression trees', 'support vector machines', 'random forests', 'k-nearest neighbors', 'naïve bayes'). The second type refers to terms related to pavement prediction models ('pavement performance models', 'road pavement performance models', 'pavement performance prediction models'). Only the literature that had been published over the last two decades was taken into account.
- To achieve the highest level of relevance, only articles with at least 50 (if between 2000 and 2009), 25 (if between 2010 and 2015), or 5 (if between 2016 and 2019) citations were considered.

Based on these considerations, six articles were collected. Table 1 presents these articles, mentioning the techniques used, the indicators adopted to describe pavement performance, and the data sources.

As can be seen, the most relevant literature on the application of machine learning technique to pavement performance prediction has three main characteristics: (1) it uses the IRI as an indicator of pavement performance; (2) it has a predominance of artificial neural networks applications; (3) it shows a growing tendency to use data from the Long-Term Pavement Performance (LTPP) database.

These characteristics may be explained by the context of application of the machine learning techniques to pavement performance prediction. Machine learning is a data-driven approach, which makes IRI a suitable indicator: it is widely available in pavement databases (e.g. LTPP database), it is measured by objective means (e.g. laser profilometer), and it is known as one of the most common indicators for pavement performance evaluation (Marcelino et al. 2018). In what regards the predominance of artificial neural networks applications, historical reasons explain this situation. Artificial neural networks are not only one of the first machine learning techniques, but also one of the most used machine learning techniques in Civil and

CONTACT Pedro Marcelino pmarcelino@lnec.pt

Table 1. Summary of studies applying machine learning techniques to pavement performance prediction.

Reference	Pavement indicator	Techniques used	Data source
Lin et al. (2003)	IRI	Artificial neural networks	Data from provincial highways and county roads in Taiwan
Bianchini and Bandini (2010)	PSI (based on IRI measurements)	Artificial neural networks	Data from the Minnesota Department of Transportation
Kargah-Ostadi et al. (2010)	IRI	Artificial neural networks	Long-Term Pavement Performance (LTPP) database
Kargah-Ostadi and Stoffels (2015)	IRI	Artificial neural networks, support vector machines, radial basis function network	Long-Term Pavement Performance (LTPP) database
Ziari *et al.* (2016c)	IRI	Group method of data handling, Adaptive neuro fuzzy inference system	Long-Term Pavement Performance (LTPP) database
Ziari *et al.* (2016b)	IRI	Artificial neural networks, group method of data handling	Long-Term Pavement Performance (LTPP) database
Ziari *et al.* (2016a)	IRI	Support vector machine	Long-Term Pavement Performance (LTPP) database
Mazari and Rodriguez (2016)	IRI	Artificial neural networks, gene expression programming	Long-Term Pavement Performance (LTPP) database

Legend: Present serviceability index (PSI), International roughness index (IRI).

Pavement Engineering (Adeli 2001, Ceylan et al. 2014). Finally, the use of the LTPP database is justified by its free access, amount of data and standard data format.

In the current context, the use of the IRI and the LTPP database is more likely to remain as the standard in studies related to the application of machine learning to pavement performance prediction than the use of artificial neural networks. While few alternatives are available to replace (with evident benefit) the use of the IRI and the LTPP database, the field of machine learning is rapidly expanding and new techniques have been emerging (Jordan and Mitchell 2015). Thus, it is important to discuss if artificial neural networks are still the most adequate machine learning technique for pavement performance prediction.

One of the main limitations of artificial neural networks is overfitting (Geman et al. 1992, Srivastava et al. 2014). Overfitting happens when the error of the model on the training set is small, but it is large on the testing set (new data). This means that the artificial neural network has memorised the training examples, but it has not learned to generalise to new observations. Since the central challenge in machine learning is to build models that perform well on new/unseen inputs – and not just on those used to train the model – overfitting is an important limitation of artificial neural networks.

Several methods have been developed to further improve the generalisation capabilities of artificial neural networks, such as different variants of cross-validation (Haykin 1994), noise injection (An 1996), regularisation (Wan et al. 2013), and optimised approximation algorithms (Liu et al. 2008). These methods can be dispensed if it is easy to collect more data and increase the size of the training set. However, that is hardly the case in pavement performance problems. Pavement performance data relies on historical information and expensive data collection methods, making it difficult to acquire more data. Accordingly, there is a danger of overfitting when applying machine learning techniques to pavement performance prediction.

Most of the studies on the application of artificial neural networks to pavement performance prediction apply this technique without addressing overfitting properly and disregarding its impact on model performance. For example, artificial neural networks models are often trained on datasets whose (small) size hardly avoids the occurrence of overfitting, but even so no methods are applied to take that issue into account (e.g. Lin et al. 2003). Also, previous studies (e.g. Ziari *et al.* 2016c) evaluated model's performance through a validation set approach instead of a k-fold cross-validation, which is known to provide a better estimation of model's generalisation capability (James et al. 2013). Finally, published works (e.g. Bianchini and Bandini 2010) often seem to ignore that the training data must be independently and identically distributed, which is an assumption underlying most machine learning techniques (Darrell et al. 2015). In the case of pavement performance prediction, this issue is particularly relevant because the data exhibits temporal dependence. Accordingly, the data used in these studies to train the artificial neural networks may not be suitable for the development of pavement performance models with generalisation capability. In summary, it is debatable if the models based on artificial neural networks have generalisation capability and, consequently, its application to network-level PMS is questionable (Kargah-Ostad and Stoffels 2015).

This paper aims to extend the application of machine learning techniques in pavement performance prediction, by providing a structured approach to the development of machine learning models that is applicable to network-level PMS and allows for the use of other techniques than artificial neural networks. A case-study design is followed to show how to use the proposed approach, as well as to demonstrate how it can be a useful tool to predict pavement performance in network-level PMS.

The first sections of this paper provide a brief overview of modelling approaches for pavement performance prediction and machine learning strategies for time-series, as well as the description of a machine learning framework for pavement performance prediction. Then, the research methodology is delineated, giving details about the case study and the analysis carried out in this work. Finally, the results are presented, discussed, and conclusions of this study are drawn.

Pavement performance prediction

According to Yang et al. (2003), there are two types of modelling approaches for pavement performance prediction: static and dynamic. The former can be described by:

$$P_t = f(X_t, t) \quad (1)$$

where, P_t = pavement performance at age t; X_t = auxiliary variables (e.g. structural characteristics, climatic conditions, traffic) at age t.

Static models do not take into account the lagged values of the output as inputs. This is a limitation for pavement performance prediction since it is known that pavements deteriorate in an incremental manner due to several factors (Haas et al. 2015). Thus, by understanding the dynamics of the deterioration process over time, future condition predictions can be improved.

By contrast, dynamic models are described in a different way:

$$P_t = f(P_{t-1}, P_{t-2}, \ldots, P_{t-n}, X_t, X_{t-1}, X_{t-2}, \ldots, X_{t-n}) \quad (2)$$

where, P_t = pavement performance at age t, X_t = auxiliary variables value at age t, n = number of past observations to consider.

As can be seen, dynamic models forecast pavement performance using historical pavement performance data. Taking into account what is known about the dynamic process of pavement deterioration, this should provide a more accurate prediction of future pavement conditions.

Dynamic models are time-series models because they describe time-dependent processes (Harvey 1993). One crucial characteristic of time-series models is that its data points are indexed in order by time, representing a sequence of discrete-time data.

In this work, dynamic models are used as time-series models have proven themselves to be a successful way to represent infrastructure deterioration (Durango-Cohen 2007).

Machine learning for time-series

For a long time, forecasting future values of an observed time series has been dominated by linear statistical models such as ARIMA models (Box and Jenkins 1970). In the 1980s, linear models started to be called into question (De Gooijer and Hyndman 2006), and nonlinear approaches started to became more and more popular (Poskitt and Tremayne 1986, Tjøstheim 1986). Nevertheless, it must be said that nonlinear approaches are still far from the level of development shown by linear approaches (Clements et al. 2004).

More recently, machine learning models established themselves as a valid alternative to classical models (Ahmed et al. 2010). The performance of machine learning models has been studied in several works (De Oliveira et al. 2000, Kim 2003, Kane et al. 2014) and forecasting competitions under different data conditions (Balkin and Ord 2000, Crone et al. 2011, Taieb et al. 2012). According to the results, machine learning models can outperform traditional methods.

Bontempi et al. (2012) present an approach to the application of machine learning in time-series problems. In particular, the authors show how to tackle time-series forecasts from a supervised learning perspective. The goal in supervised learning is to learn a mapping from input variables x to output variables y, given a labelled set of input/output examples (Murphy 2012).

The approach presented by Bontempi et al. (2012) is based on the idea that the input variables of the machine learning model are the lagged time series values, while the output variables are the values to be predicted. Accordingly, the forecasting task uses a sequence $<y_1, \ldots, y_t>$ to predict the next $t+1$ value. If other input variables $<x_1, \ldots, x_t, x_{t+1}>$ are available, it is possible to extend this case. In those situations, the problem is no longer a univariate time-series problem (set of values from a single variable over time) and starts to be a multivariate time-series problem (set of values from several variables over time).

As noted by Bontempi et al. (2012), machine learning can be used for multi-step forecasts (predict the next $t+N$ value). The main forecasting strategies applied in these cases are the recursive and the direct strategy. Figure 1 illustrates each of these strategies.

In this work, multi-step forecasts are made applying a direct strategy. The direct strategy builds a separate time-series model for each forecasting horizon. The forecasts are computed only on the observed data. For example, if it is a five-year forecast ($t+5$), only data previous to those five years will be used by the learning model.

The direct strategy is not prone to error accumulations because it does not use approximated values to compute the forecasts. However, it has some flaws, such as a higher demand of computational time, and the need to apply higher functional complexity to model the dependencies between two series values at two distant instants (Guo et al. 1999, Tong 2012). Since none of these limitations represent a significant constraint for the development of machine learning models in this work, the direct strategy was considered appropriate for this study.

Machine learning framework for pavement performance prediction

This section proposes a general framework for the development of pavement performance prediction models, based on machine learning algorithms, to use in network-level PMS. As shown in Figure 2 this framework is composed of five steps.

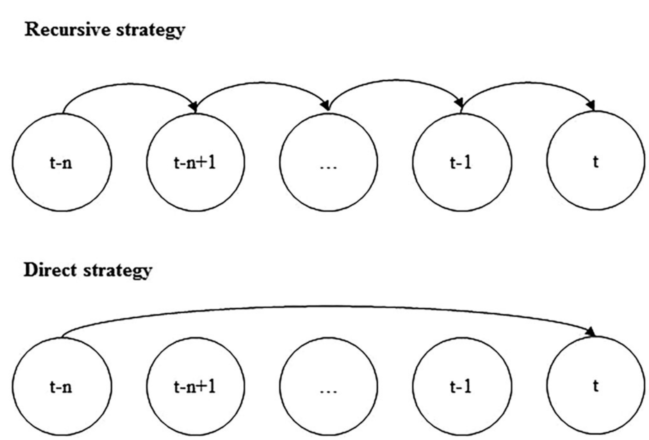

Figure 1. Recursive and direct strategies.

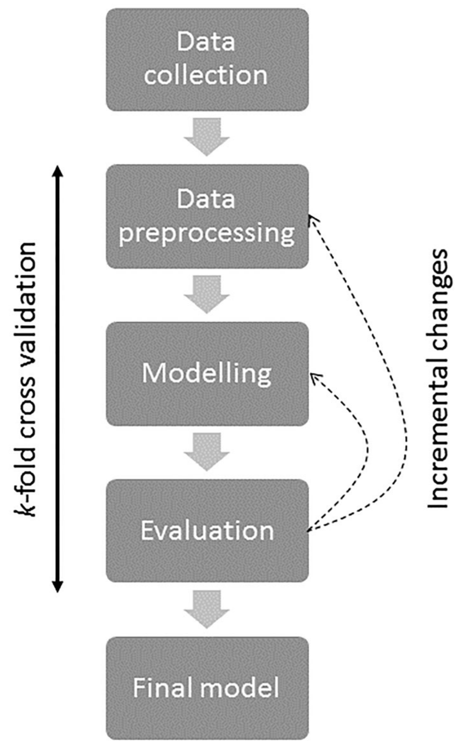

Figure 2. Framework for the development of pavement performance prediction models with machine learning.

Data collection

The first step is to collect data. In general, data for pavement performance studies result from two different sources: transportation agencies and public sources available.

Data preprocessing

After data collection, it is necessary to pre-process the data. The pre-processing step should result in a data set that is complete, clean and suited for the learning process. Several techniques can be applied during data pre-processing. Its application depends on the problem and no optimal sequence of data pre-processing techniques exists. Kotsiantis et al. (2006) present a comprehensive list of data pre-processing techniques. Due to their impact on the generalisation performance, the following set of preprocessing techniques should be considered:

- **Time series preprocessing**. As mentioned before, the approach presented by Bontempi et al. (2012) is recommended for the application of machine learning in time series problems. In these problems, there is a temporal dependence between the values. Consequently, data should be preprocessed so that the input variables of the machine learning model are the lagged time series values (variables values at time *t-1*, *t-2*, … , and *t-n*), and the output variables are the desired variables values at time *t*.
- **Sampling**. Sampling is a common practice in machine learning applications and it helps to ensure the generalisation of models (Raschka 2015). The idea behind sampling consists of randomly splitting the dataset into training and test datasets. The training dataset is used to train the model, while the test dataset evaluates its final performance. Since the goal of a prediction model is to make accurate predictions on new data, its performance should be evaluated on unseen data. Thus, sampling is necessary because it allows us to avoid a typical methodological mistake, which is learning and testing the model on the same data. In the case of pavement performance prediction, caution is necessary when choosing the observations to include in the dataset because data are independently and identically distributed. Preferably, no observations from identical road sections should be repeated.
- **Missing data**. This is a recurrent problem in real-world datasets. One of the practical impacts of missing data is the reduction of the dataset size. Thus, proper methods to deal with missing data may need to complete the dataset. When dealing with missing data, it is essential to identify the type, the extent, and the randomness of the missing data processes, to define which treatment method should be used. There are several options available to the researcher (Hair et al. 1998), and no universal method will deliver the best performance in every situation (Schmitt et al. 2015).
- **Data cleaning.** The goal of data cleaning is to detect and remove errors and/or inconsistencies from data so that its quality is improved. Data cleaning problems can result from several sources (e.g. data entry errors) and the way to approach them is strongly context dependent. For pavement performance problems, it makes sense to perform sanity checks based on engineering judgment. Also, more complex data analysis approaches (e.g. outlier detection) may be applied. Hodge and Austin (2004) provide a structured and comprehensive overview of the research on outlier detection.

Modelling

Once the data set is preprocessed, the learning algorithm can be applied. Usually, several machine learning algorithms can be applied to the same problem. For regression problems, some of the most popular algorithms are random forests (Breiman 2001), support vector regression (Smola and Schölkopf 2004), and nearest neighbours regression (Altman 1992).

Evaluation

Following the choice of the machine learning algorithm, performance metrics should be evaluated and, if necessary, incremental changes should be made to improve the performance of the model. Based on the evaluation of the performance metrics, it is possible to identify improvement opportunities and start an incremental process that seeks to enhance the model's performance. A comprehensive framework for the evaluation of pavement performance models is provided by Kargah-Ostadi and Stoffels (2015).

To improve the performance of machine learning models, several strategies can be considered:

- **Gather more data**. It is often better to gather more data than to improve the machine learning model (Goodfellow et al. 2016). However, gathering more data can be costly or unfeasible.

- **Get better data**. This option is especially relevant when the performance of the model on the training set is poor. To improve the quality of the data, it may be necessary to remove noisy data, which can be done through outlier detection. It may also be necessary to remove noisy features or to include additional ones. Features are the input variables/attributes of the model. Accordingly, its optimisation plays an essential role in performance improvement. The problem of finding the most compact and informative set of features is addressed by 'feature extraction'. A broad introduction to the various aspects of feature extraction can be found in Guyon and Elisseeff (2006).
- **Adjust hyperparameters**. Most machine learning algorithms have a set of parameters that control many aspects of the algorithm's behaviour and express higher-level properties of the model. These parameters are called 'hyperparameters'. Hyperparameters are fixed before the learning process begins. Adjusting these hyperparameters, it is possible to improve the quality of the model and its ability to predict correct estimates when applied on new inputs (unseen or future data). To adjust hyperparameters, approaches such as grid search and random search are common practice (Goodfellow et al. 2016). These approaches iteratively consider different combinations of hyperparameters so that the performance of the model can be optimised. In the case of grid search, all possible combinations are considered, while in random search only a number of combinations is considered.

One important and transversal aspect in the proposed framework is the application of the *k*-fold cross-validation (cf James et al. 2013). *K*-fold cross-validation is a technique to evaluate learning algorithms by splitting the training set into two segments: one to learn the model, and another to validate it. Proper use of cross-validation should be done in all the steps of the algorithm to ensure generalisation and provide an almost unbiased estimate of the model's true error (Varma and Simon 2006), which supports its preference over other techniques such as the validation set approach or the leave-one-out cross-validation (James et al. 2013). The framework definition implies that the k-fold cross-validation is applied to the training set.

Final model

When all the previous steps are completed, a final evaluation using the test dataset is performed. Assuming that the results are consistent and satisfactory, the model should then be trained with the entire dataset (training and test datasets), respecting all the considerations previously defined. Finally, the model complete and ready to make predictions.

Research methodology

A case study design was used to investigate the suitability of the approach proposed in this work. This case study shows how to use the machine learning framework to develop prediction models and demonstrates that the resulting models improve the predictive capacity of PMS.

The case study was based on data from the LTPP database (2018). These data were preprocessed in accordance with the proposed framework, in order to make the data suitable for the application of machine learning algorithms.

Random forests (cf. Breiman 2001) was the machine learning algorithm chosen for this application. Despite its popularity, few studies have explored the application of random forests in pavement performance prediction. Note that the approach proposed in this study is generic and can be applied to different machine learning algorithms (e.g. Support Vector Machines).

The International Roughness Index (IRI) (Sayers and Karamihas 1998) was used to describe pavement performance. Since the IRI tends to remain relatively stable from one year to the next (Perera et al. 1998), only long-term predictions (5 and 10 years) were considered in this case study. However, as was the case with the machine learning algorithms, the approach can be applied to other performance indicators (e.g. cracking).

To evaluate the performance of the models developed in this study, two types of evaluation were considered. Firstly, models' performance was characterised as suggested by Kargah-Ostadi and Stoffels (2015), which proposed a framework for the evaluation and comparison of models developed through machine learning techniques. Secondly, the machine learning models were compared with other existing prediction models, which were built for specific states/provinces.

Data collection and preprocessing

The data used in this study was collected from the Long-Term Pavement Performance (LTPP) database (LTPP 2018). This database has information about a large number of pavement test sections in the United States (US) and Canada, namely on materials, performance, traffic, weather, and maintenance and rehabilitation interventions. Data from the Standard Data Release 31 July 2017, was used.

Although several factors contribute to pavement deterioration (Haas et al. 2015), only those with less than 30% of missing values were included. Since this study intends to produce network-level models that are applicable in the daily practice of PMS, it was considered that the variables representing these factors should be easily accessible. Nevertheless, engineering judgment was applied in the choice of these variables and an effort was made to include performance, structural, climatic, and traffic data. To handle missing values, a popular imputation technique called 'missForest' (cf. Stekhoven and Bühlmann 2011) was used. Table 2 presents the resulting list of variables.

As mentioned before, a direct strategy was applied for the development of machine learning models. Different datasets were created for the 5 and 10-years predictions. These datasets must be complete (no observations with missing data) and comprehensive (include as many lagged values as possible). Since longer historical datasets tend to have a lower number of complete observations, there is a trade-off between having a complete dataset and a comprehensive dataset. Accordingly, several datasets were created to explore this trade-off and find the most accurate prediction models. Table 3 and Table 4 presents these datasets, considering that:

Table 2. Dataset variables.

Structure	Climate	Traffic	Performance
• Pavement thickness (PT) • Structural number (SN)	• Annual average precipitation (AAP) • Annual average temperature (AAT) • Annual average freeze index (AAFI) • Minimum annual average humidity (MinAAH) • Maximum annual average humidity (MaxAAH)	• Cumulative annual average daily truck traffic (CAADTT)	• International roughness index (IRI)

- Training sets use 80% of the available data, while test sets use the remaining 20%.
- The input variables correspond to the lagged values of IRI, AAP, AAT, AAFI, MinAAH, MaxAAH, CAADTT, PT and SN. Each dataset considers a different set of lagged values as detailed in Table 3 and Table 4. For example, Dataset 5F has nine input variables. IRI(t-5), AAP(t-5), AAT(t-5), AAFI(t-5), MinAAH(t-5), MaxAAH(t-5), CAADTT(t-5), PT(t-5) and SN(t-5).
- The output variable is the IRI at time *t*.

All the observations included in the datasets refer to annual data. Variables' annual data was calculated using the average of the values collected over the year. In the case of the IRI, it is important to refer that the average of the left and right wheel path measurements is used to represent the IRI value in each road section. Considering that the purpose of this study is to build a simple model that can be applied in network-level PMS, no other considerations or corrections were made to the values in the database.

By definition, all the observations refer to periods in which no interventions occurred. In case of an intervention, the observation ends and a new one starts. This option intended to keep the modelling process sufficiently reliable and far from a complexity level that could compromise its application in network-level PMS.

In addition, no observations from identical road sections should be repeated. Since one of the main assumptions of most machine learning algorithms is that all the samples in the training set are independently and identically distributed, data series should be collected from different road sections. By default, the most recent data series was considered.

Modelling

The approach proposed in this study is designed for different machine learning algorithms. For this case study, a random forests algorithm (cf. Breiman 2001) was used. Despite its general popularity and effectiveness (Caruana and Niculescu-Mizil 2006, Bowles 2015), few applications of this algorithm are known in the Pavement Engineering field. It was considered that the application of this algorithm could contribute to improving the accuracy of the machine learning model.

Random forests

As an ensemble method (cf. Dietterich 2000), random forests provide a way to reduce the variance of a machine learning model by combining different models. In order to build these models, several training sets from the population should be used. The problem is that when dealing with real data, it is usually not possible to obtain multiple training sets from the population.

Random forests use bootstrap (Efron 1979) to generate multiple training data sets. This technique emulates the process of obtaining new data sets by repeatedly sampling from the original dataset (the one used to compute the model), instead of repeatedly obtaining new data sets from the population.

Once several training sets are available, the algorithm trains a base learner on each of these sets. The resulting models are averaged, and the various outputs are merged into a single prediction. In random forests, all the models receive equal weight. Figure 3 summarises the random forest process.

One distinctive feature of random forests is that it only uses a random sample of the available *m* features. Meaning that, during the sampling process, not all the available features are used. Forcing each sample to consider only a subset of the features, random forests separate the resulting trees and make their average less variable. Accordingly, random forests usually limit overfitting issues.

In this work, the application of the random forests algorithm was made using scikit-learn (Pedregosa et al. 2011). Accordingly, the hyperparameters corresponding to the random forest algorithms must be defined:

- Number of trees in the forest (*n_estimators*).
- Maximum depth of the tree (*max_depth*).
- Minimum number of samples required to split an internal node (*min_samples_split*).

Table 3. Datasets composition (5-years prediction models).

Dataset ID	Lagged values	Number of input variables	Number of observations (total)	Number of observations (training set)	Number of observations (test set)
5A	t-10, t-9, t-8, t-7, t-6, t-5	54	289	231	58
5B	t-9, t-8, t-7, t-6, t-5	45	421	336	85
5C	t-8, t-7, t-6, t-5	36	554	443	111
5D	t-7, t-6, t-5	27	793	634	159
5E	t-6, t-5	18	1079	863	216
5F	t-5	9	1370	1096	274

Table 4. Datasets composition (10-years prediction models).

Dataset ID	Lagged values	Number of input variables	Number of observations (total)	Number of observations (training set)	Number of observations (test set)
10A	t-15, t-14, t-13, t-12, t-11, t-10	54	34	28	6
10B	t-14, t-13, t-12, t-11, t-10	45	66	53	13
10C	t-13, t-12, t-11, t-10	36	92	74	18
10D	t-12, t-11, t-10	27	139	112	27
10E	t-11, t-10	18	193	155	38
10F	t-10	9	289	232	57

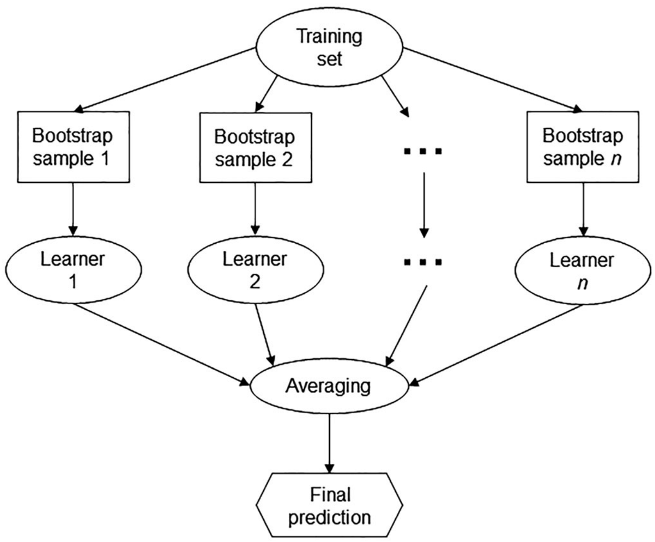

Figure 3. Random forests algorithm.

- Minimum number of samples required to be at a leaf node (*min_samples_leaf*).
- Number of features to consider when looking for the best split (*max_features*).

These hyperparameters were defined in an optimised way using random search (cf. Bergstra and Bengio 2012). Table 5 presents the hyperparameters values considered in this study.

Evaluation

This study uses the framework proposed by Kargah-Ostadi and Stoffels (2015) to characterise models' performance. According to this framework, pavement performance models should be evaluated according to three categories:

- Quantitative;
- Qualitative;
- Quantitative + qualitative.

The principles and metrics proposed by Kargah-Ostadi and Stoffels (2015) for each performance category were followed in this study. Figure 4 summarizes the framework for the characterisation of pavement performance models considering its categories, principles, and metrics.

Table 5. Values of random forests hyperparameters.

Parameter	5-years prediction models	10-years prediction models
n_estimators	800	800
max_depth	None	15
min_samples_split	2	10
min_samples_leaf	2	2
max_features	3	9

As established by the machine learning framework described in this study, a *k*-fold cross-validation (cf. James et al. 2013) approach was used to estimate the error of the model (average mean squared error and standard deviation of mean squared error). A 5-fold cross-validation was used because it works well in practice and it has computational advantages over other options (James et al. 2013).

A comparison between the machine learning models presented in this paper and a set of other existing models, which were developed for specific states/provinces in the US and Canada, was made. This comparison intended to evaluate the performance of the model in real-case scenarios and assess its generalisation capability. Table 6 details the models considered in this study for comparison purposes.

According to their states/provinces applicability, the models mentioned in Table 6 were used to predict the IRI values in a selected set of road sections. These predictions were compared with those resulting from the application of the machine learning models. All the road sections included in this comparative study were excluded from the original training dataset so that the machine learning models had to make predictions on unseen data. This way it is possible to assess its generalisation capability. The road sections considered in this part of the study are listed below, as identified in the LTPP database:

- Indiana: 18-1028, 18-5518, 18-6012.
- Saskatchewan: 90-6410.
- Texas: 48-0166, 48-0801, 48-0802.

Despite the efforts to choose road sections with complete data, it was not always possible to do so. Some models are significantly complex and demanding in what regards input data. Accordingly, when necessary, simplifying assumptions were made:

- When the IRI measurements did not match with the months when other performance data measurements (e.g. rutting) were made, the closest measurements were selected. For example, if the IRI was measured in March and rutting was measured in April and September, the April value was used and the September value was discarded.
- When different several data points were available for a specific year, the mean values were considered.
- When yearly traffic data was missing, the available traffic data were used to calculate an average annual growth rate and predict the missing traffic data values.
- Road sections with other missing data were discarded.

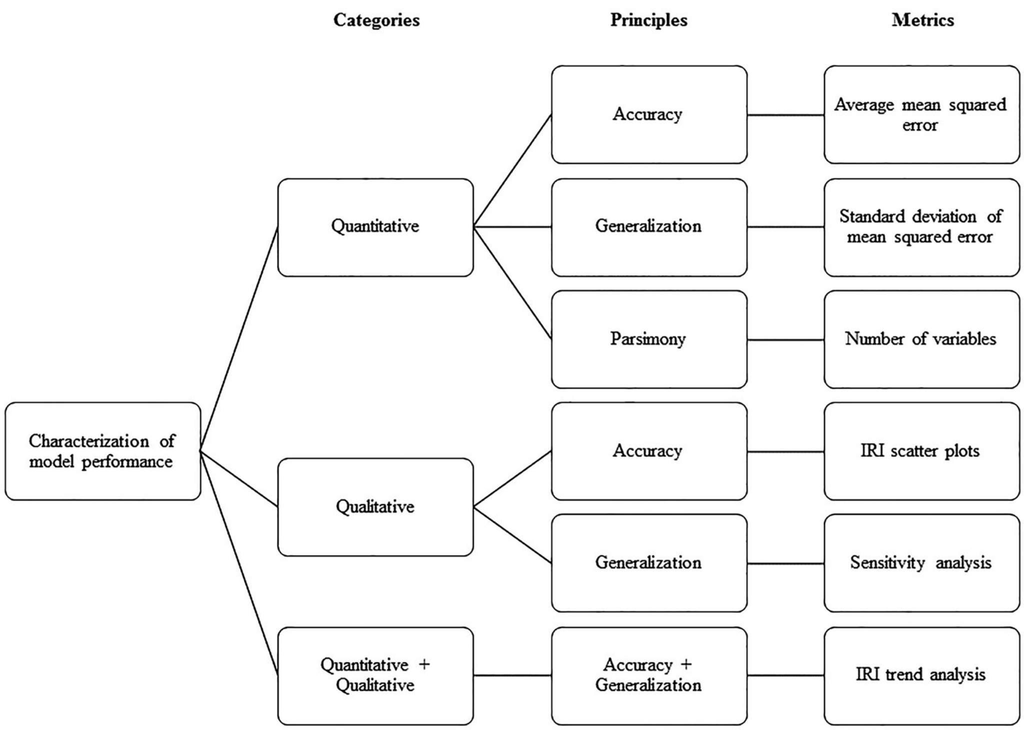

Figure 4. Characterisation of model performance.

Results and discussion

In general, the results suggest that the proposed approach is able to develop machine learning prediction models that can improve accuracy and generalisation of IRI prediction models for use in network-level PMS. The quantitative and qualitative analyses that support these findings are detailed in the following sections.

Error analysis

The average mean squared error (AVG MSE), the standard deviation of the mean squared error (STD MSE), and the number of variables were evaluated to characterise the performance of the model in quantitative terms. Table 7 and Table 8 show the results obtained for the 5 and 10-years prediction models, respectively.

It can be seen from the data in Table 7 that Model 5B is the best performing model as it has the lowest AVG MSE (0.064), an error that is about 65% less than the error resulting from the worse performing model (Model 5F). These results suggest that Model 5B benefits from the additional number of input variables it has in relation to Model 5F. In this case, more variables correspond to longer sequences of data which can be captured by the machine learning models to produce better predictions.

Taking the square root of the AVG MSE resulting from Model 5B, to express the error in IRI units, it is possible to see that the error corresponds to a value of 0.253 m/km. From an engineering perspective, given the prediction horizon and the wide range of site types covered by the model, it can be said that the results are within what is acceptable in practice. In fact, considering that in the LTPP database there are no reports of road sections with an IRI value higher than 6 m/km, it can be said that an error of 0.253 m/km represents an error close to 4% of the possible range. Moreover, recent studies (Dalla Rosa et al. 2017) reported errors in a similar order of magnitude (0.316 m/km) to models developed and applied at a state-level (Texas, in this case). Here, the models are applicable to several states/provinces included in the LTPP database, which makes the results encouraging.

Similar conclusions can be taken for the 10-years prediction models. In this case, the best performing model (Model 10D) has an AVG MSE of 0.104, which is 88% lower than the error resulting from the worse performing model (Model 10A). The poor performance of Model 10A can be explained by the size of its dataset. A total of 34 observations is not enough to develop a proper machine learning model. For example, Model 10B, which has a dataset with almost twice the size of the Model's 10A dataset, has an AVG MSE value that is 34% lower than the value resulting from Model 10A.

As expected, the 5-years prediction models are more accurate and generalise better than the 10-years prediction models. The difference in accuracy between the two best performing model is about 63% (0.064 vs. 0.104), and the standard deviation of the 5-years prediction model is considerably lower than the standard deviation resulting from the 10-years prediction model (0.014 vs. 0.078). However, taking into account the complexity of the problem and considering that the forecast accuracy decreases as the forecasting horizon increases, the results seem to be normal.

In both prediction scenarios, it is possible to see that there is a trade-off between the number of input variables and the number of observations figuring in the dataset. This study is inconclusive regarding the optimal balance between the number of input variables and the number of observations. However, it

Table 6. Reference models.

Source	Formula	Variables	Applicable states/provinces
Gulen et al. (2001)	$IRI\} = (43 + 1.8 \times AGE\} + 0.0004 \times AADT) \times \frac{0.0254}{1.60934}$	IRI = International Roughness Index (m/km) AGE = Time between the last major surface work and the year of data collection (years) AADT = Average annual daily traffic	Indiana
Nassiri et al. (2013)	$IRI\} = 0.34\} + \ 0.059 \times e^{AGE/25} \times log(AADT) + 0.006 \times P200\}$ $+ \ + 0.088 \times TransCrk\} + \ 0.013 \times MiscCrk\} + \ 0.07 \times Rutting$	IRI = International Roughness Index (m/km) AGE = Age since last major treatment (years) AADT = Average annual daily traffic P200 = Percentage of aggregate passing #200 sieve (%) TransCrk = Transverse cracking (% area) MiscCrk = Miscellaneous cracking (% area) Rutting = Rutting (mm)	Saskatchewan
Dalla Rosa et al. (2017)	$\ln\left(\frac{IRI_i}{IRI_n - IRI_i}\right) = \beta_3 + \beta_2 \times e^{AGE \times \beta_1}$	IRI_i = Initial IRI (m/km) IRI_n = IRI in year n (m/km) β_1, β_2, β_3 = Calibration coefficients AGE = Number of years since IRI_i (m/km)	Texas

Table 7. Performance of the 5-years prediction models.

Model ID	Dataset	Average of the mean squared error	Standard deviation of the mean squared error	Number of variables
5A	5A	0.073	0.026	54
5B	5B	0.064	0.014	45
5C	5C	0.067	0.015	36
5D	5D	0.078	0.014	27
5E	5E	0.095	0.024	18
5F	5F	0.106	0.011	9

Table 8. Performance of the 10-years prediction models.

Model ID	Dataset	Average of the mean squared error	Standard deviation of the mean squared error	Number of variables
10A	10A	0.196	0.270	54
10B	10B	0.129	0.142	45
10C	10C	0.149	0.162	36
10D	10D	0.104	0.078	27
10E	10E	0.128	0.073	18
10F	10F	0.119	0.055	9

is clear that a small number of input variables leads to poor performances (e.g. Model 5F), as well as a small dataset (e.g. Model 10A).

Scatter plots

The results of the scatter plot analysis for the best performing models (Model 5B and Model 10D) are shown in Figure 5. These plots refer to the application of the model to the test dataset since predictive models should be evaluated on unseen data.

The results indicate that the 5-years prediction model has an acceptable predictive capacity because the scatterplots show a coefficient of determination (R^2) of 0.98 (5-years prediction model). In what regards the 10-years prediction model, its R^2 is 0.93. This result illustrates the accuracy difference existing between the models. The visual analysis of the Model 10D scatter plot also suggests that for IRI values higher than 2 m/km, the model tends to overestimate the IRI value. A possible explanation for this is the predominance of IRI values between 0.5 and 2.0 m/km, which makes the model more suitable for prediction in that range of values.

Sensitivity analysis

To understand the consistency of the two best performing models (Model 5B and Model 10D) to changes in input variables, a sensitivity analysis was performed. This analysis consisted of removing one type of input variable at a time and then evaluating the effect that this action had on model performance. Since the previous IRI values are expected to be the most influent variable, it was also considered an additional model in which all input variables but the previous IRI values were removed. Accordingly, several prediction models were built for the 5 and 10-years prediction cases.

The Kolmogorov–Smirnov test was used to analyze the relationship between the original error distribution and the error distribution resulting from the models modified for the

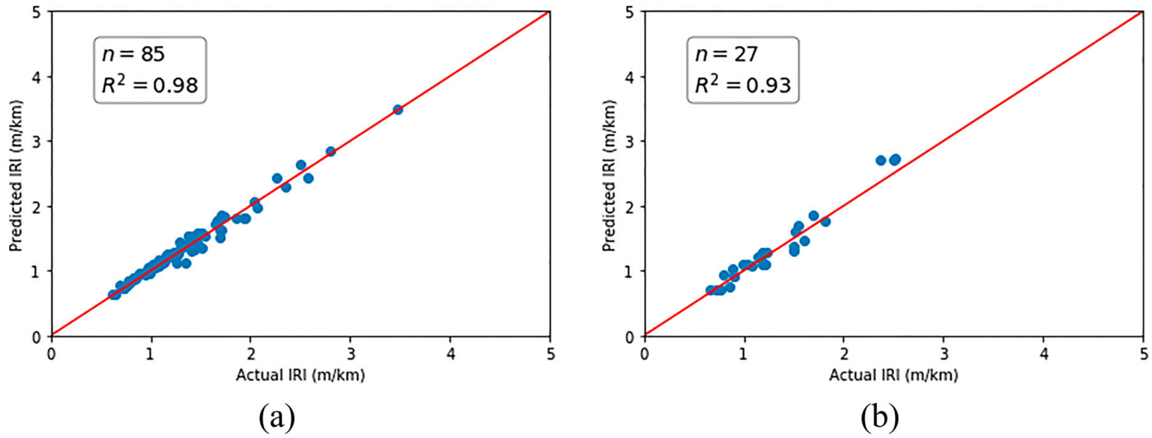

Figure 5. IRI scatter plots: (a) Model 5B (5-years prediction model); (b) Model 10D (10-years prediction model).

sensitivity analysis. This test checks if two datasets are drawn from the same distribution and has the advantage of making no assumption about the distribution of data.

In the context of this sensitivity analysis, the model is not sensitive to the removed variables if the errors resulting from the original and modified models are drawn from the same distribution. The fact that the error distribution remains the same after variable removal, shows how it does not significantly affect the predictive capacity of the model. By contrast, when the error distribution changes, it can be said that the model is sensitive to the removed variable.

Table 9 summarises the quantitative metrics of the modified models, as well as the p-values resulting from the Kolmogorov–Smirnov test. Note that this is a test for the null hypothesis that two datasets are drawn from the same distribution. Accordingly, the null hypothesis can be rejected if the p-value is less than a specific significance level. The values of the original model (with all variables) were included for reference.

It is apparent from the table that the developed machine learning models are most sensitive to previous IRI values. As it is shown for the 5-years prediction case, when the IRI values are removed, the null hypothesis can be rejected at a significance level of 3.8%. Moreover, the accuracy of the model decreases considerably. These results suggest that the model is sensitive to IRI removal. By contrast, when other variables are removed, the accuracy of the model decreases but the p-values are high, indicating that the null hypothesis cannot be rejected. Thus, the model does not seem to be sensitive to the removal of other variables than the IRI.

Similar results have been observed in the 10-years prediction case. In this case, when the IRI values are removed, the accuracy of the model decreases and the null hypothesis can be rejected at a significance level of 8.8%. The results also indicate that the model is not sensitive to the removal of other variables than the IRI.

Overall, these results indicate that the predictive capacity of the machine learning models developed in this study is primarily influenced by the previous IRI values. In general, these results are consistent with those reported in previous studies (Perera and Kohn 2001, Kargah-Ostadi and Stoffels 2015) and with what is expected from an indicator that reports a phenomenon of progressive deterioration.

Model comparison

In an attempt to show that the machine learning approach provides prediction models that are applicable in network-level PMS, a comparison between the models developed in this study and three reference models was made. These reference models resulted from a literature review that focused on models specifically developed at different times for states/provinces figuring in the LTPP database. A summary of the selected models was already presented in Table 6.

Table 9. Sensitivity analysis.

	5-years prediction model			10-years prediction model		
Variable removed	Average mean square error	Standard deviation of mean square error	p-values	Average mean square error	Standard deviation of mean square error	p-values
IRI	0.256	0.063	0.038	0.302	0.156	0.088
AAP	0.070	0.014	0.980	0.113	0.075	0.995
AAT	0.071	0.012	1.000	0.112	0.076	1.000
AAFI	0.070	0.011	1.000	0.111	0.075	1.000
MinAAH	0.072	0.012	1.000	0.112	0.075	0.917
MaxAAH	0.071	0.012	1.000	0.110	0.075	1.000
CAADTT	0.071	0.012	0.998	0.108	0.074	1.000
PT	0.071	0.012	1.000	0.112	0.076	1.000
SN	0.072	0.012	0.998	0.111	0.075	1.000
All except IRI	0.072	0.012	0.925	0.111	0.075	0.917
None (original model)	0.064	0.014	N/A	0.104	0.078	N/A

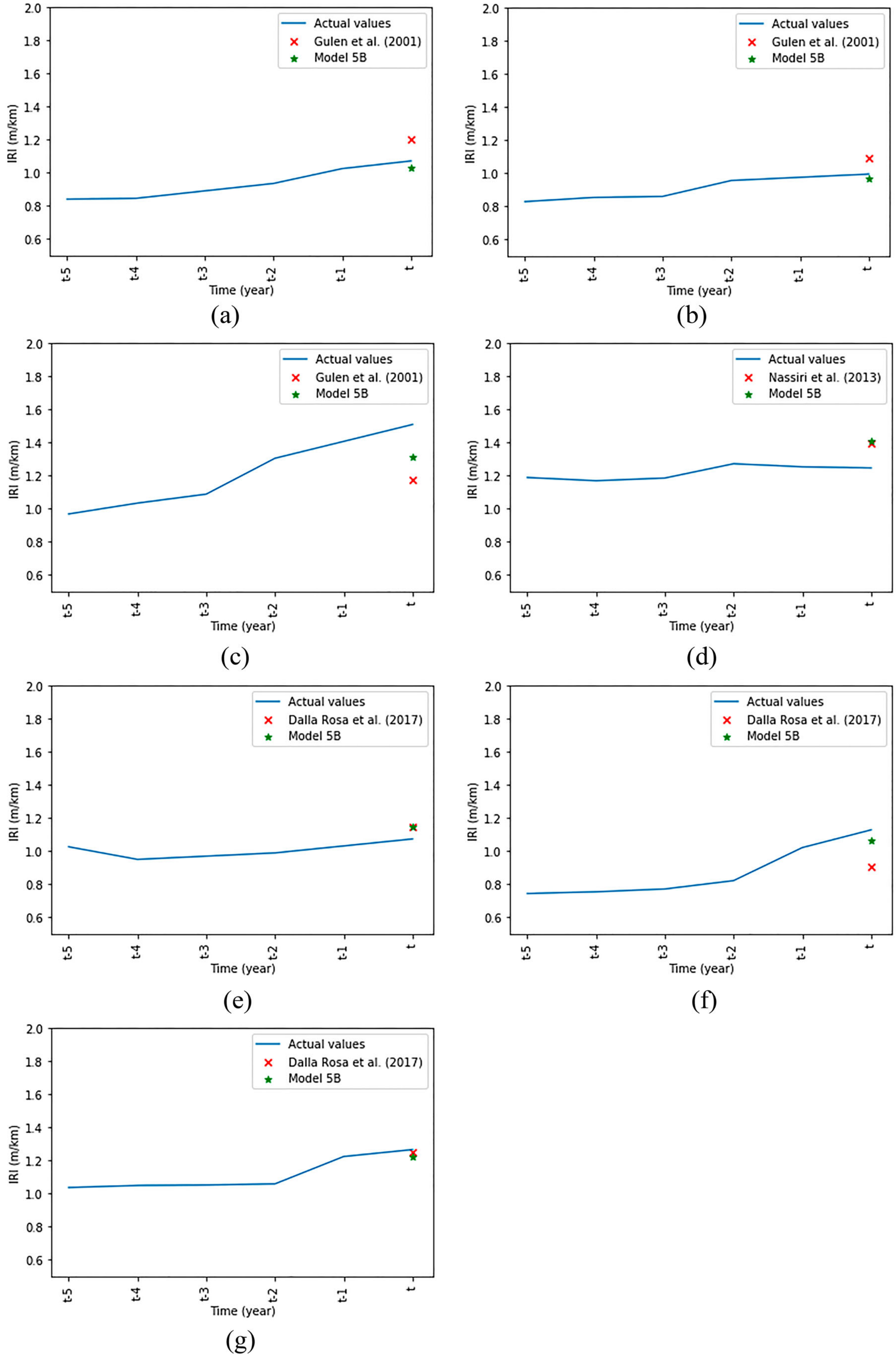

Figure 6. IRI trend analysis for 5-years predictions in the following road sections: (a) 18-1028; (b) 18-5518; (c) 18-6012; (d) 90-6410; (e) 48-0166; (f) 48-0801; (g) 48-0802.

Table 10. Error analysis.

	Error (Actual − Estimated)		Relative error	
Road section	Reference model	Model 5B	Reference model	Model 5B
18–1028	0.129	−0.042	12.01%	−3.91%
18–5518	0.096	−0.029	9.65%	−2.91%
18–6012	−0.335	−0.197	−22.19%	−13.04%
90–6410	0.145	0.161	11.64%	12.90%
48–0166	0.076	0.075	7.07%	6.98%
48–0801	−0.225	−0.062	−19.90%	−5.49%
48–0802	−0.017	−0.044	−1.36%	−3.43%

The IRI trends of specific road sections were used to compare the models. Due to data availability restrictions, it was not possible to include more than seven road sections. Also, for this reason, only the 5-years prediction model (Model 5B) was applied.

Figure 6 shows the IRI trend and the predictions resulting from the different models for the 5-years prediction scenario. Table 10 summarises the absolute and relative prediction errors.

Comparing the results obtained with the reference models with those resulting from the random forest approach, it can be seen that the reference models have an average absolute error of 0.146 (11.97%), while Model B has an average absolute error of 0.087 (6.95%). Moreover, in 5 of the 7 cases analyzed, the random forest model outperformed the reference models. These results suggest that the random forest model has a better predictive capacity, showing that the method proposed in this study is able to develop general and accurate prediction models.

It is important to bear in mind that the model comparisons reported in this section are limited to the cases where there was sufficient data to run the models. Accordingly, the findings might not be representative of all cases. However, it can be said that there is evidence that the machine learning models are able to predict the IRI accurately and using a set of variables that are commonly available in pavement management databases.

Conclusions

Pavement performance prediction plays an essential role in pavement management systems (PMS). Models with an increased predictive capacity are able to estimate future pavement conditions more accurately, improving pavement maintenance decisions.

This study aimed to develop a general approach for the development of pavement performance prediction models applicable in network-level PMS, using machine learning techniques. The proposed method intended to structure the entire modelling process, addressing issues that have a negative impact on models' generalisation performance and prevent their application in a wide range of site types.

A case-study approach was used to illustrate the proposed approach and to test the performance of the resulting prediction models. Based on the random forests algorithm, several prediction models were developed to predict the International Roughness Index (IRI) values in 5 and 10-years. A set of variables that are commonly available was used to develop the machine learning models. Since this work intended to build a model that could be easily implemented in network-level PMS, it is important to mention that all the variables considered are usually available in pavement management databases.

The research has shown that, through proper modelling, machine learning algorithms can be used to develop accurate prediction. A reasonable predictive performance was achieved by the developed models, as stated by the average mean squared error of the best-performing ones: 0.064 (5-years prediction model) and 0.104 (10-years prediction model). As expected, the 5-years prediction model performed better than the 10-years prediction model. However, considering their prediction horizon, both models have an acceptable accuracy for long-term IRI predictions at the network-level.

In what regards the models' generalisation capability, it can be said that the results are promising. The standard deviation of the mean squared error was 0.014 (5-years prediction model) and 0.078 (10-years prediction model), which are low values if it is taken into account that the models are applied on a wide range of site types. This generalisation capability is evident in the illustrative examples explored in this study, as the machine learning models outperform the reference models in almost every case.

A sensitivity analysis showed that the machine learning models are particularly sensitive to previous IRI values, which was expected due to the physical process of IRI deterioration. Variations in the remaining variables showed little effect on the machine learning models.

The findings of this research will be of interest to road agencies with road maintenance responsibilities. By improving predictions accuracy, machine learning algorithms can optimise maintenance and rehabilitation interventions and reduce maintenance costs. Moreover, the approach presented in this study is absolutely generic. It can be applied to different machine learning models and to different pavement performance indicators, according to the needs and goals of road agencies, leveraging the potential of machine learning applications in this research field.

Further research might explore other machine learning algorithms, such as the nearest neighbours regression, as well as focus on the prediction of different performance indicators. The approach presented in this study can also be tested in different prediction horizons.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Fundação para a Ciência e a Tecnologia: [Grant Number SFRH/BD/129907/2017].

ORCID

Pedro Marcelino https://orcid.org/0000-0002-5521-6066
Maria de Lurdes Antunes http://orcid.org/0000-0002-1911-517X
Eduardo Fortunato https://orcid.org/0000-0002-9968-0821
Marta Castilho Gomes https://orcid.org/0000-0002-6662-7397

References

Adeli, H., 2001. Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, 16 (2), 126–142. doi:10.4203/ccp.16.

Ahmed, N.K., *et al.*, 2010. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29 (5-6), 594–621. doi:10.1080/07474938.2010.481556.

Altman, N.S, 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46 (3), 175–185. doi:10.2307/2685209.

An, G, 1996. The effects of adding noise during backpropagation training on a generalization performance. *Neural Computation*, 8 (3), 643–674. doi:10.1162/neco.1996.8.3.643.

Balkin, S.D., and Ord, J.K, 2000. Automatic neural network modeling for univariate time series. *International Journal of Forecasting*, 16 (4), 509–515. doi:10.1016/s0169-2070(00)00072-8.

Bergstra, J., and Bengio, Y, 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13 (Feb), 281–305.

Bianchini, A., and Bandini, P, 2010. Prediction of pavement performance through neuro-fuzzy Reasoning. *Computer-Aided Civil and Infrastructure Engineering*, 25 (1), 39–54. doi:10.1111/j.1467-8667.2009.00615.x.

Bontempi, G., Taieb, S.B., and Le Borgne, Y.A., 2012. Machine learning strategies for time series forecasting. *In*: *European business intelligence summer school*. Berlin: Springer, 62–77.

Bowles, M., 2015. *Machine learning in Python: essential techniques for predictive analysis*. Indianopolis, IN: Wiley.

Box, G.E.P., and Jenkins, G., 1970. *Time series analysis, forecasting and control*. San Francisco: Holden-Day.

Breiman, L, 2001. Random forests. *Machine Learning*, 45 (1), 5–32.

Caruana, R., and Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. *In*: Proceedings of the 23rd international conference on machine learning, 161–168.

Ceylan, H., Bayrak, M., and Gopalakrishnan, K, 2014. Neural networks applications in pavement engineering: A recent survey. *International Journal of Pavement Research and Technology*, 7 (6), 434–444. doi:10.6135/ijprt.org.tw/2014.7(6).434.

Clements, M.P., Franses, P.H., and Swanson, N.R, 2004. Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20 (2), 169–183. doi:10.1016/j.ijforecast.2003.10.004.

Crone, S.F., Hibon, M., and Nikolopoulos, K, 2011. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27 (3), 635–660. doi:10.1016/j.ijforecast.2011.04.001.

Dalla Rosa, F., Liu, L., and Gharaibeh, N.G., 2017. IRI prediction model for use in network-level pavement management systems. *Journal of Transportation Engineering, Part B: Pavements*, 143 (1), 04017001. doi:10.1061/jpeodx.0000003.

Darrell, T., *et al.*, 2015. *Machine learning with interdependent and non-identically distributed data*. Dagstuhl Reports, 5 (4). Schloss Dagstuhl.

De Gooijer, J.G., and Hyndman, R.J., 2006. 25 years of time series forecasting. *International Journal of Forecasting*, 22 (3), 443–473. doi:10.1016/j.ijforecast.2006.01.001.

De Oliveira, K.A., Vannucci, Á., and da Silva, E.C., 2000. Using artificial neural networks to forecast chaotic time series. *Physica A: Statistical Mechanics and its Applications*, 284 (1-4), 393–404. doi:10.1016/S0378-4371(00)00215-6.

Dietterich, T.G., 2000. Ensemble methods in machine learning. *In*: *International workshop on multiple Classifier systems*. Berlin: Springer, 1–15.

Durango-Cohen, P.L, 2007. A time series analysis framework for transportation infrastructure management. *Transportation Research Part B: Methodological*, 41 (5), 493–505. doi:10.1016/j.trb.2006.08.002.

Efron, B, 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7 (1), 1–26. doi:10.1007/978-0-387-75692-9_9.

Geman, S., Bienenstock, E., and Doursat, R, 1992. Neural networks and the bias/variance dilemma. *Neural Computation*, 4 (1), 1–58. doi:10.1162/neco.1992.4.1.1.

Goodfellow, I., Bengio, T., and Courville, A., 2016. *Deep learning*. Cambridge, MA: MIT Press.

Gulen, S., *et al.*, 2001. *Development of improved pavement performance prediction models for the Indiana pavement management system*. Final Report. Indiana Department of Transportation.

Guo, M., Bai, Z., and An, H.Z, 1999. Multi-step prediction for nonlinear autoregressive models based on empirical distributions. *Statistica Sinica*, 9, 559–570.

Guyon, I., and Elisseeff, A, 2006. An introduction to feature extraction. *In*: *Feature extraction*. Berlin: Springer, 1-25.

Haas, R., Hudson, R., and Falls, L.C., 2015. *Pavement asset management*. Upper Saddle River, NJ: John Wiley & Sons.

Hair, J.F., *et al.*, 1998. *Multivariate data analysis*. Upper Saddle River: Prentice Hall.

Harvey, A.C, 1993. *Time series models*. New York: Harvester Wheatsheaf.

Haykin, S., 1994. *Neural networks: a comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall.

Hodge, V., and Austin, J, 2004. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22 (2), 85–126. doi:10.1023/b:aire.0000045502.10941.a9.

James, G., *et al.*, 2013. *An introduction to statistical learning*. New York: Springer.

Jordan, M., and Mitchell, T, 2015. Machine learning: trends, perspectives, and prospects. *Science*, 349 (6245), 255–260. doi:10.1126/science.aaa8415.

Kane, M.J., *et al.*, 2014. Comparison of ARIMA and random forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics*, 15 (1), 276–284. doi:10.1186/1471-2105-15-276.

Kargah-Ostadi, N., and Stoffels, S.M, 2015. Framework for development and comprehensive comparison of empirical pavement performance models. *Journal of Transportation Engineering*, 141 (8), 04015012. doi:10.1061/(asce)te.1943-5436.0000779.

Kargah-Ostadi, N., Stoffels, S., and Tabatabaee, N, 2010. Network-level pavement roughness prediction model for rehabilitation recommendations. *Transportation Research Record: Journal of the Transportation Research Board*, 2155, 124–133. doi:10.3141/2155-14.

Kim, K.J, 2003. Financial time series forecasting using support vector machines. *Neurocomputing*, 55 (1-2), 307–319. doi:10.1016/S0925-2312(03)00372-2.

Kotsiantis, S.B., Kanellopoulos, D., and Pintelas, P.E, 2006. Data preprocessing for supervised leaning. *International Journal of Computer Science*, 1 (2), 111–117.

Lin, J.D., Yau, J.T., and Hsiao, L.H., 2003. Correlation analysis between international roughness index (IRI) and pavement distress by neural network. *In*: *Proceedings of the 82nd annual meeting of the transportation research board*. Washington, DC: Transportation Research Board, 101–121.

Liu, Y., Starzyk, J.A., and Zhu, Z, 2008. Optimized approximation algorithm in neural networks without overfitting. *IEEE Transactions on Neural Networks*, 19 (6), 983–995. doi:10.1109/tnn.2007.915114.

Long-Term Pavement Performance (LTPP), 2018. *LTPP DataPave online* [dataset]. Available from: https://infopave.fhwa.dot.gov/ [Accessed 20 Mar 2018].

Marcelino, P., Antunes, M. L., and Fortunato, E, 2018. Comprehensive performance indicators for road pavement condition assessment. *Structure and Infrastructure Engineering*, 1–13. doi:10.1080/15732479.2018.1446179.

Mazari, M., and Rodriguez, D, 2016. Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. *Journal of Traffic and Transportation Engineering*, 3 (5), 448–455. doi:10.1016/j.jtte.2016.09.007.

Murphy, K., 2012. *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Nassiri, S., Shafiee, M.H., and Bayat, A, 2013. Development of roughness prediction models using alberta transportation. *International Journal of Pavement Research and Technology*, 6 (6), 714–720.

Pedregosa, F., *et al.*, 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12 (Oct), 2825–2830.

Perera, R.W., Byrum, C., and Kohn, S.D, 1998. *Investigation of development of pavement roughness*. Technical Report FHWA-RD-97-417. Federal Highway Administration.

Perera, R.W., and Kohn, S.D, 2001. *LTPP data analysis: factors affecting pavement smoothness*. NCHRP Project 20-50 (8/13). Transportation Research Board.

Poskitt, D.S., and Tremayne, A.R, 1986. The selection and use of linear and bilinear time series models. *International Journal of Forecasting*, 2 (1), 101–114. doi:10.1016/0169-2070(86)90033-6.

Raschka, S., 2015. *Python machine learning*. Birmingham: Packt Publishing.

Sayers, M.W. and Karamihas, S.M., 1998. *The little book of profiling*. Ann Arbor, MI: University of Michigan.

Schmitt, P., Mandel, J., and Guedj, M., 2015. A comparison of six methods for missing data imputation. *Journal of Biometrics and Biostatistics*, 6 (1), 224–229.

Smola, A.J., and Schölkopf, B, 2004. A tutorial on support vector regression. *Statistics and Computing*, 14 (3), 199–222. doi:10.1023/b:stco.0000035301.49549.88.

Srivastava, N., *et al.*, 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15 (1), 1929–1958.

Stekhoven, D.J., and Bühlmann, P, 2011. MissForest - non-parametric missing value imputation for mixed-type data. *Bioinformatics (oxford, England)*, 28 (1), 112–118. doi:10.1093/bioinformatics/btr597.

Taieb, S.B., *et al.*, 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, 39 (8), 7067–7083. doi:10.1016/j.eswa.2012.01.039.

Tjøstheim, D, 1986. Estimation in nonlinear time series models. *Stochastic Processes and Their Applications*, 21 (2), 251–273. doi:10.1016/0304-4149(86)90099-2.

Tong, H., 2012. *Threshold models in non-linear time series analysis*. New York: Springer Science & Business Media.

Uddin, W., Hudson, W.R., and Haas, R., 2013. *Public infrastructure asset management*. New York: McGraw Hill.

Varma, S. and Simon, R., 2006. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7 (1), 91–98.

Wan, L., *et al.*, 2013. Regularization of neural networks using dropconnect. *In*: *Proceedings of the 30th international conference on machine learning*. Atlanta: PMLR, 1058–1066.

Yang, J., Lu, J.J., and Gunaratne, M, 2003. *Application of neural network models for forecasting pavement crack index and pavement condition rating*. Final Report. University of South Florida.

Ziari, H., *et al.*, 2016a. Prediction of pavement performance: application of support vector regression with different kernels. *Transportation Research Record: Journal of the Transportation Research Board*, 2589, 135–145. doi:10.3141/2589-15.

Ziari, H., *et al.*, 2016b. Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods. *International Journal of Pavement Engineering*, 17 (9), 776–788. doi:10.1080/10298436.2015.1019498.

Ziari, H., *et al.*, 2016c. Analysing the accuracy of pavement performance models in the short and long terms: GMDH and ANFIS methods. *Road Materials and Pavement Design*, 17 (3), 619–637. doi:10.1080/14680629.2015.1108218.