Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition

Article *in* Journal of Infrastructure Systems · January 2020

DOI: 10.1061/(ASCE)IS.1943-555X.0000512

Tamer El-Diraby
University of Toronto

Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index

S. Madeh Piryonesi[1] and Tamer E. El-Diraby[2]

Abstract: Understanding the deterioration of roads is an important part of road asset management. In this study, the long-term pavement performance (LTPP) data and machine learning algorithms were used to predict the deterioration in the pavement condition index (PCI) over 2, 3, 5, and 6 years. In selecting the attributes for conducting the analysis, we targeted ones that are freely available. This approach can help smaller municipalities, which could be short on money or required expertise. For larger ones and transportation agencies, this can save the increasingly significant costs for collecting field data and any associated safety or traffic implications. In addition, we used this category of attributes to better examine the role of data analytics in asset management. Without considering a causal model, can trends in data help assess deterioration in the PCI? Several models using combinations of 15 attributes were learned and tested. The algorithms used in this study were two types of decision trees and their boosted models based on gradient boosted trees. The accuracy of the ensemble of boosted classifiers was considerably higher than their base learners, with some reaching over 80% in predicting unseen data. We also found that dividing data into different climatic zones can change the relative importance of attributes and the overall accuracy of the models. Increasing the prediction span reduces accuracy, while reducing the number of prediction classes (levels of deterioration) increases the accuracy. In addition to automating the calculation and prediction of PCI, this study presented informative or important attributes for prediction. Such analyses could help municipalities and departments of transportations with forming a more effective policy for data collection and management. **DOI: 10.1061/(ASCE)IS.1943-555X.0000512.**

Author keywords: Data analysis; Asset management; Deterioration modeling; Pavement condition index; Gradient boosted trees.

Introduction

Evidence-based decision making in infrastructure asset management (AM) is a challenging problem. The domain is multidisciplinary with diversified stakeholders. This complexity is matched with an increased subjectivity of objectives, including a consideration of levels of services, sustainability, and resilience. This is where the use of data analytics can be very valuable. In addition to statistical analysis, data analytics incorporate the use of data mining and machine learning algorithms and can handle structured and unstructured data sources. One of the main advantages of the expanded scope of data analytics is its ability to detect trends and relations as well as to learn without the need for a causal model (Provost and Fawcett 2013).

In this research work, we used data analytics (specifically, classification algorithms) to help predict deterioration in the pavement condition index (PCI) over a short time span. One of the key features of our analysis is limiting the data used by algorithms to that which can be acquired without field work—in other words, data that is free. The rationale behind this type of data is twofold. First, being able to work with free data helps municipalities in doing cost-effective assessments. The current approach for calculating and predicting PCI values requires extensive field work, and conducting this work frequently and with adequate quality may be expensive for smaller municipalities. For larger municipalities, such work may have negative implications on staff safety and traffic flows. Second, testing the power of data analytics can determine whether obtaining patterns in (not necessarily engineering) data can yield quality predictions to causal models of analysis (based on technical and engineering definitions and equations). In so doing, on the technical dimension, we can help municipalities conduct the perceptions anytime they need to do so. On the data analytics dimension, we can explore the role, advantages, and limitations of data analytics in the domain of AM. In addition, we focused on testing data that are specific to local contexts such as weather and traffic data.

The process used in this work can serve as benchmarks for similar investigations of other AM analyses. On a larger scope, lessons learned in this research work can help municipalities develop more effective data collection policies: Which data is more informative, and how to collect it? This is an important question. A study by Pantelias et al. (2008) on 50 US agencies and state departments of transportation (DOTs) revealed that, in most cases, decisions and practices in data collection are still highly subjective and conventional. Data are collected according to staff experience and past practices rather than objective analyses of relevance and value-added.

Literature Review

Proper management of performance data is essential to AM practices. Municipalities and transportation agencies use different physical performance indicators (PIs) to evaluate the condition and remaining life of their roads. Some of the most popular PIs include PCI, international roughness index (IRI), structural condition index

[1]Ph.D. Candidate, Dept. of Civil and Mineral Engineering, Univ. of Toronto, 35 St. George St., Toronto, ON, Canada M5S 1A4 (corresponding author). Email: madeh.piryonesi@mail.utoronto.ca

[2]Associate Professor, Dept. of Civil and Mineral Engineering, Univ. of Toronto, 35 St. George St. Toronto, ON, Canada M5S 1A4. Email: tamer@ecf.utoronto.ca

Note. This manuscript was submitted on August 2, 2018; approved on June 3, 2019No Epub Date. Discussion period open until 0, 0; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342.

(SCI) (Chi et al. 2014; Elbagalati et al. 2017), and present serviceability index (PSI). In this study, we chose to use PCI because it is commonly used by DOTs and municipalities in North America. For example, by surveying 58 municipalities in Ontario, we learned that more than 95% of them use the PCI as the major performance indicator of their roads. Of course, the approach developed here can be used in analyzing other PIs such as IRI, SCI and PSI. The value of PCI varies between 0 and 100, respectively, representing the worst and the best possible values of PCI. ASTM D6433-07 explains the process of calculating PCI in detail (Way et al. 2015). The Ontario Ministry of Transportation (MTO) has developed another set of guidelines in its manual #SP-024 (Chong et al. 1982). In both cases, calculating the PCI requires collecting distress data such as fatigue cracking, bleeding, edge cracking, and block cracking.

Although deterioration modeling (prediction) is an essential part of AM, many municipalities do not pay careful attention to it, and some municipalities disregard this part altogether. For instance, according to a recent study in Ontario (El-Diraby et al. 2017), most small municipalities in Ontario failed to include a deterioration model in their AM analyses. The same study revealed that those municipalities that use deterioration modeling depend on standardized curves. These curves are deterministic—users have no guidelines on how to add variability to their values when conducting probabilistic risk analysis. Moreover, these curves are, to some extent, context insensitive. Future PCI values are estimated based only on the passage of time. These curves overlook the impact of other road-specific attributes such as maintenance history, pavement type/composition, and traffic (Fig. 20) (Wu 2015). Some of the weaknesses of deterioration curves are addressed by probabilistic deterioration models, which are discussed in the following paragraphs.

One category of commonly used infrastructure deterioration models comprises Markov models (Ford et al. 2012). A Markov chain is a memoryless stochastic process, and these probabilistic models consider the deterioration of roads as a series of discrete events. Unlike the deterministic nature of standardized deterioration curves, they can incorporate probabilities. However, crude Markov models overlook the history of deterioration and remedial actions (i.e., previous maintenance and rehabilitation actions) (Anyala et al. 2014; Ens 2012; Neves and Frangopol 2005; Piryonesi and Tavakolan 2017). As a result, these models have limited ability to realistically describing the effects of previous maintenance or rehabilitation on the condition of assets. Although some researchers have used Markov processes in modeling the climate change effect on infrastructure (Pozzi et al. 2017; Špačková and Straub 2017), and more specifically in the domain of pavement, efforts have been made to consider the effects of climate and traffic by creating multiple models and comparing them (Osorio-Lird et al. 2018), but the pavement performance models developed based on Markov chain cannot use attributes such as climatic and traffic-related features as explicit inputs to quantify the impacts of such stressors on pavement health (Anyala et al. 2014; Piryonesi and El-Diraby 2018). Future states (i.e., future conditions of roads) are dependent only on the present state; hence, they are time-independent. This characteristic of crude Markov models is called time-homogeneity. From an AM perspective, this means that regardless of the age of the road, the probability of transitioning to another level of deterioration is always the same (Ens 2012). Consequently, semi-Markov models are used to address this issue (Black et al. 2005). Unlike time-homogeneous Markov models, a semi-Markov model uses independently distributed random variables to model the time between the states. Therefore, these models can consider the road's age. Semi-Markov models usually require more longitudinal data in comparison to time-homogeneous models. Such data are not easily available to many municipalities (Ens 2012). An example of applying Markov processes with a non-homogeneous transition matrix is available in (Li et al. 1996). It is important to indicate that a few studies in the literature of using Markov models for infrastructure performance modeling have addressed some of these issues (Guillaumot et al. 2003; Memarzadeh and Pozzi 2016). As mentioned, the deterioration models generated based on Markov models cannot explicitly represent attributes such as climate or traffic. Li et al. (1996) developed multiple Markov models based on different traffic volumes to test the sensitivity of their models to this attribute. The results of their work revealed that traffic has a significant role in the deterioration of roads. Therefore, roads with different traffic volumes should not be assessed by the same model.

Calculating the transition probabilities are the most significant part of developing a Markov model, and several methods have been used in the literature for this purpose. Transition probabilities can be found by expert opinion when sufficient historical data is not available (Ford et al. 2012; Kleiner 2001). Other more objective approaches for calculating the transition probabilities are optimization, Weibull distribution (Kleiner 2001), statistical modeling based on observed frequencies (Ford et al. 2012), and approximations using pairs of inspections (Ens 2012; Ford et al. 2012). To correct the effect of time on deterioration, multiple transition matrices could be established for several age ranges. Lastly, an important point is that by using Markov models, one can go directly from models to decisions (using Markov decision processes), while using machine learning, it may not be easy to quantify the decision-making part.

Based on the presented discussion, sufficient research has been performed about correlation-based and Markov models. Another approach for understanding the deterioration of roads is using data analytics, which is relatively new, and harnessing its power requires more research. Data analytics algorithms can learn from large and heterogeneous datasets. National Cooperative Research Program (NCHRP) reports machine learning techniques in several cases have been applied to the problem of pavement deterioration. However, these techniques have been used to a lesser extent, likely because such estimates are perceived as coming from a black-box (Ford et al. 2012). The NCHRP's report limits machine learning to neural networks. Neural networks, which are among the most commonly used algorithms in the domain of pavement management, are criticized for their lack of interpretability (Shahin et al. 2009).

Several examples of using neural networks in the domain of pavement management are as follows. Alsugair and Al-Qudrah (1998) trained a neural network to predict the appropriate type of maintenance and rehabilitation for asphalt pavement. They used several road distresses (e.g., bump, edge cracking, slippage cracking, and swelling) as predictive attributes. The developed models were trained to predict the most suitable maintenance action out of 13 options. Studies that are more specifically devoted to predicting the PIs or the remaining life of roads are also available. Ferregut et al. (1999) used an artificial neural network to predict the remaining life of pavement. Their input attributes included an input vector with nine elements that represent the thickness of the asphalt concrete (AC) and base layers and seven falling weight deflectometer (FWD) readings. The output is any of the two critical strains at the interfaces of the layers or the remaining life of the pavement when it experiences either fatigue cracking or rutting. The Asphalt Institute models were used to link the remaining life with rutting and fatigue cracking (Ferregut et al. 1999).

Another study that used neural networks in predicting pavement performance was performed by Lou et al. (2001). They used attributes such as three consecutive values of the crack index (CI) and the age of road as input to predict the CI in the near future. They prepared their training set based on the data from the Florida Department of Transportation (Lou et al. 2001; Yang et al. 2003). Similar studies that have used neural networks in predicting the performance or life of pavement are available in the literature (Ford et al. 2012; Karlaftis and Badr 2015; Miradi and Molenaar 2006; Terzi 2007). Furthermore, other probabilistic prediction studies on pavement that mostly rely on Bayesian methods are available. For instance, Ramia and Ali (1997) used Bayesian methods to predict the rutting in flexible pavements. Another example of using Bayesian methods was performed by Anyala et al. (2014), in which they assessed the impact of different climatic scenarios on rut depth progression for asphalt pavement. Their study is an endeavor to combine mechanistic-empirical models with probabilistic distributions and simulation to create deterioration models for different types of asphalt pavements. They used Bayesian regressions to estimate the distribution of coefficients of mechanistic-empirical models under different climatic scenarios.

Neural networks are not easy to interpret, and their black-box nature does not help in understanding the relative importance of the attributes used (Ford et al. 2012; Shahin et al. 2009). Therefore, other researchers have tried using other algorithms. An example of the application of other machine learning algorithms in predicting pavement performance is given by Chi et al. (2014). They trained decision trees based on the Texas Department of Transportation data. They reported that DOTs can use the results of their models for parts of their network when the data of FWD is not available. The accuracy of their models for predicting five levels of SCI was around 60%, which seems to be satisfactory. The size of their training set was 354, which is pretty small, potentially adversely affecting the accuracy. They trained their models using attributes such as the amount of distress and ride score. Furthermore, the attributes were averaged over a period of 5 years (e.g., 5-years average of distress). Kargah-Ostadi and Stoffels (2015) used the LTPP database to train artificial neural networks, radial basis function networks, and support vector machines (SVM) to predict the progression of roughness. Some of their predictive attributes included previous IRI, ESAL, the plasticity index of the base, and percent of clay in subgrade soil. Generally, most studies conducted on the LTPP data tend to use the IRI as their main performance indicator instead of the PCI (Haider et al. 2007; Kargah-Ostadi et al. 2010; Tighe 2002; Ziari et al. 2016) probably because the PCI is not included in the LTPP database and because its calculation requires some data preparation.

Previous studies mostly focus on predicting a single engineering index such as roughness or rutting. First, these indices by themselves cannot comprehensively represent pavement health. Second, understanding these operational PIs is not convenient for high-level management or the public, who are other important parties in the practice of AM. Another issue of previous models is that their input data is not free or easy to collect. For instance, most small municipalities have no access to updated FWD or distress data. Furthermore, previous researchers tend to use longitudinal data for training their models. Thus, more than 1 year of data is required as input data. Augmenting multiple-year data into one attribute may not be a suitable approach as many municipalities and DOTs do not have (updated) data for several consecutive years (El-Diraby et al. 2017).

These limitations open the larger topic of management strategies of data in AM. Several researchers have studied the challenges of road data collection and the techniques for correcting missing or faulty data (Al-Zou'bi et al. 2015; Farhan and Fwa 2014; Shalaby and Reggin 2007). However, despite the extensive need and the investments in data collection and management, limited research is available regarding how to find informative data or their significance/contribution to decision making in AM. Pantelias et al. (2008) proposed a framework for data collection that aimed to support project selection for maintenance and rehabilitation. Within this scope, the study provided general guidelines for assessing data collection needs based on literature review and survey results. Another study by Woldesenbet et al. (2015) used network analysis to model the interaction between data and decision-making processes in road management. By using surveys and interviews to realize networks, they assessed how frequently a specific piece of data is used in decision making.

Decision Trees and Gradient Boosted Trees

Decision trees are among the most popular classification algorithms. Part of their popularity stems from their intelligibility and ease of interpretation (Hastie et al. 2009). As their name suggests, these algorithms generate a classifier tree based on the trends in the dataset. Primitive versions of decision trees, such as ID3 and CLS, were only able to be learned from discrete data (Wu et al. 2008), whereas their descendants (e.g., C4.5) are capable of learning from both continuous and discrete variables.

Decision trees create segmentations of the dataset. They start with the most informative attribute and split the data based on a test. Therefore, at least two branches grow out of each node. Then, the nodes in each branch split based on their informativeness. The terminal node of a tree is called a leaf, which is ideally pure and belongs to a particular class. Different trees may use different measures for defining informativeness. Most trees rely on entropy or the Gini index (Hastie et al. 2009; Provost and Fawcett 2013; Wu et al. 2008). The trees developed in this paper are based on entropy. The information gain and the homogeneity of each leaf are determined by the parameters of the tree (Lin and Chen 2012).

Despite their popularity and intelligibility, sometimes decision trees cannot reach the expected accuracy in classifying unseen data (Chi et al. 2014; Hastie et al. 2009; Piryonesi and El-Diraby 2018). A recent technique for improving their accuracy is the use of an ensemble of trees instead of a single tree. This technique, which is called ensemble learning, is not limited to decision trees and could also be used for other classifiers. Two popular approaches of ensemble learning are bagging and boosting. Both approaches have two general stages. The first stage is generating a population of base (or weak) learners from the training set, and the second stage is then combining them to create a stronger predictive model. Their main difference lies in the way they combine the base learners. While bagging relies on simple committee voting, boosting gives a larger weight to the votes of more accurate base learners (Hastie et al. 2009). For decision trees, random forest (RF) and gradient boosted trees (GBT) are the most popular realizations of ensemble learning (Hastie et al. 2009; Provost and Fawcett 2013; Wu et al. 2008). In this paper, the latter is used because it usually outperforms its counterpart.

Objectives: Predictions Using Cost-Effective Data

The first objective of this study is to train a machine learning algorithm that could adequately predict short-term PCI deterioration through easy-to-collect and less-expensive data. The rationale behind focusing on short-term deterioration is that most municipalities in Ontario conduct a comprehensive survey on their road network every five years [see, for the example, Beckwith Township (Beckwith 2013)]. Consequently, having intermediate estimates at 2, 3, 4, or 5 years can give them early warnings about areas that

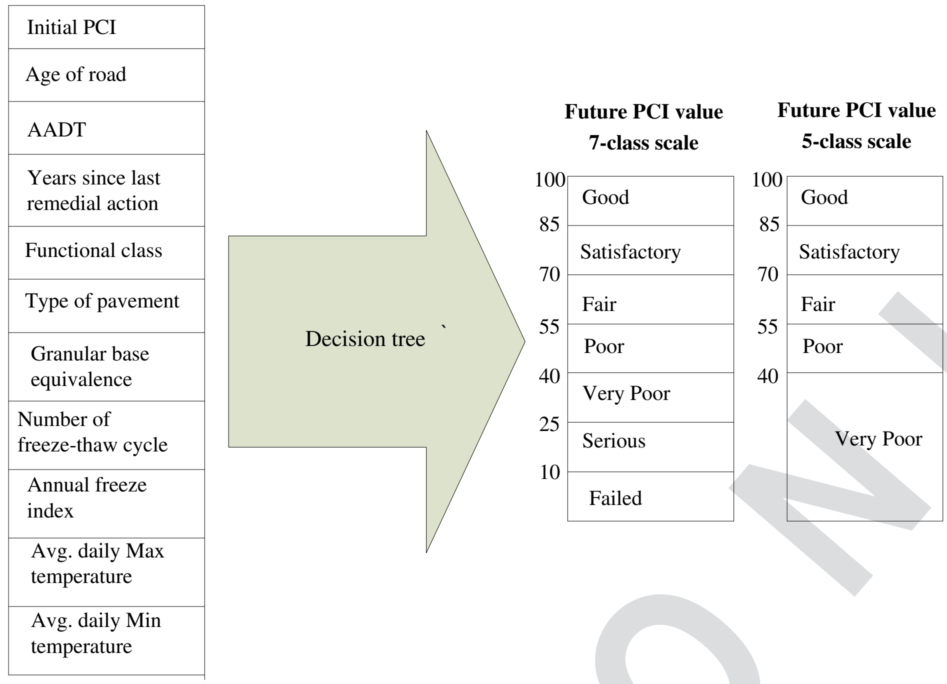


Fig. 1. A conceptual representation of the models; several predictive attributes are used to predict the class of PCI in the future; PCI could be categorized into seven or five classes.

may need more detailed assessment. Fig. 1 depicts the scope of our work: given a set of affordable attributes, LTPP data are used to train a model to predict the level of PCI in the near future using either a 5- or 7-level scale. ASTM divides the PCI into seven categories. However, models were also tested with five levels of PCI. Merging the lowest classes of PCI increased the model's accuracy. We tested models containing different numbers of attributes. Moreover, different prediction time horizons of 2, 3, 5, and 6 years were tested and compared.

Choosing the type of algorithm to use is not a trivial problem. We decided to mostly rely on decision trees, as they provide an open-box approach where decision-makers can see/test the role/impact of every attribute at different stages of the analysis. There are other reasons for choosing decision trees. First, training a decision tree (for example, the frequently used type called C4.5) requires almost no pre-requisites or assumptions about predictive attributes. There is almost no limitation on the type of attributes used in training decision trees, and this is not the case for some other algorithms. For instance, the attributes to train a naive Bayes classifier must be independent. In other words, a naive Bayes classifier assumes that given a specific label, the predictive attributes are independent of each other. Therefore, if this assumption is not satisfied, the prediction error is increased. Second, decision trees are intuitive and easy to interpret given that they result in a graphic model. Finally, they can be easily implemented and reused for classifying data points out of a training set. Unlike some other classifying algorithms such as k-nearest neighbor (k-NN), decision trees result in an explicit model, which can easily classify examples from new data. With k-NN, for example, every new data point must be assessed based on previously classified data points (Hastie et al. 2009; Najibi et al. 2017; Provost and Fawcett 2013; Wu et al. 2008).

Decision trees have a few weaknesses that could be addressed by their ensembles. Examples of such weaknesses are the lack of robustness and relatively low accuracy (Hastie et al. 2009; Wu et al. 2008). RF and GBT are examples of ensembles of trees that considerably perform better than a single learner. In this paper, the GBT is used to accommodate for the relatively low accuracy of decision trees.

Methodology

The methodology used is summarized in this section. First, since PCI values are not included in the online LTPP database (InfoPave 2017), a tool was developed in Python to calculate the PCI from available distress data in the LTPP database. The ASTM methodology was adopted (Way et al. 2015), and it relies on extracting several interim data values from existing curves. To help automate the calculation of PCI for road sections, we first digitized all ASTM curves. A mathematical formula (polynomial) was fitted to each curve. Then, a Python program was developed to calculate PCI values from the distresses data in LTPP and the digitized curves. The generated PCIs and other attributes of road were used for training the machine learning models. In the next step, to select which attributes to consider, we conducted a literature review and a few interviews with experts. Furthermore, we analyzed the AM plans and databases of ten small municipalities to check which attributes are easily available to them. A provisional list of 15 attributes was prepared. Data were then retrieved from the online LTPP database using SQL-based queries. After completing data preparation and cleansing, we used seven ranking algorithms and a heuristic feature-selection algorithm to select an initial set of attributes that could be significant in predicting PCI values. Finally, two types of decision trees were used to predict the PCI in the near future. To improve the accuracy, two GBT models developed based on these trees were trained and tested as well. Fig. 2 concisely shows the explained procedure.

After identifying the initial list of attributes, two types of decision trees and their boosted versions were trained to predict PCI values for road segments in the upcoming years. Learned decision

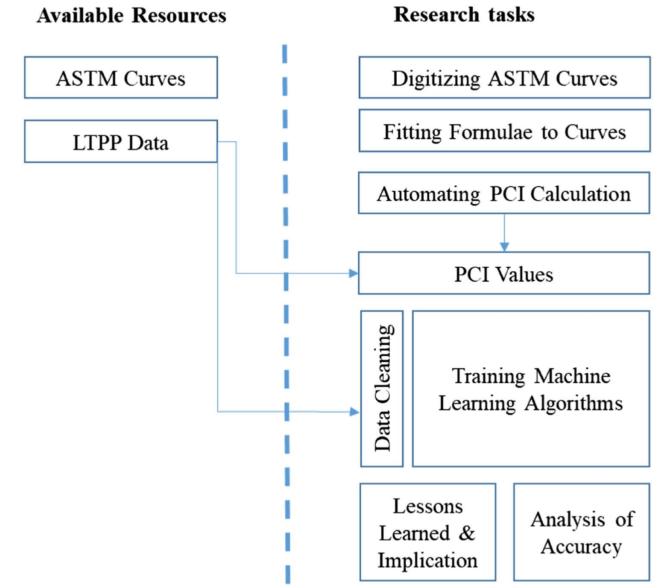


Fig. 2. Summary of adopted methodology.

trees successfully predicted the class of PCI in both short and long intervals of time. The ensemble of trees (i.e., the GBT) increased the accuracy of prediction significantly. Initially, our scale included seven classes/levels. These were later reduced to five to enhance accuracy. The horizon of prediction was initially 3 years. The prediction interval was then changed to study 2, 5, and 6 years. The accuracy of models was tested for unseen data using cross-validation.

Furthermore, the impact of the size of the training set and the quality of data was tested. To test the effect of the size of the training set on the model's accuracy, models were trained using different data sizes—namely, 110, 256, 550, and 942 road sections. Increasing the size of the training set significantly boosted accuracy. Furthermore, the accuracy of models learned from cleaned data was significantly higher. Moreover, the results show how separating the data can result in a clearer understanding of the results. Different models were learned for different subsets of data, for instance, different climatic regions or roads with different rehabilitation histories. Finally, the confusion matrices of trained models were developed. A confusion matrix tracks correct and incorrect predictions and compares them to actual PCI values. By knowing the correct class/level of PCI for different sections, a confusion matrix indicates the number the model predicted correctly. For the wrongly predicted ones, the confusion matrix determines which were an overestimation and which were an underestimation. Therefore, by studying a confusion matrix, model evaluation is not limited to a one-number assessment such as the correlation factor. This model evaluation exposes the variety of mistakes made, which is a major advantage of machine learning classifiers over descriptive statistics or correlation-based analyses.

Data Preparation and Data Cleansing

Data were retrieved from the LTPP database, which is available online (InfoPave 2017). With several hundred tables, the LTPP database is the world's largest and most comprehensive pavement performance database. It is updated every 6 months and includes both asphalt and concrete pavements. This study concentrated on asphalt pavement only. Data were retrieved using SQL-based queries. In addition, sections denoted by specific pavement studies (SPS) in the dataset were not used in the training because of replication: SPS sections that are co-located have identical traffic and climatic data (Elkins et al. 2003). After downloading the data, a major step was calculating PCI values from distress data, as the LTPP database does not include the PCI. Furthermore, since LTPP data is stored in different tables, different fields had to be collated by SQL join queries. Moreover, some attributes were created by combining two or more attributes. For instance, the attribute *last remedial action* used in this model is a combination of LTPP attributes of major rehabilitation and maintenance actions, which are stored in different tables. Data cleansing included removing erroneous records. An example of an erroneous record was when the PCI value increased after a few years with no maintenance.

Per ASTMD 6433-07, PCI was calculated based on the distresses and the geometry of the road sections (Way et al. 2015). Therefore, the distress data and the dimensions of road sections were retrieved from the online platform of LTPP. To automate the calculation of PCI, two sets of curves (i.e., deduct value and corrected deduct-value curves) were digitized and mathematically represented through polynomials. Altogether, 31 deduct values for distress density curves and 8 correction curves were digitized. They were embedded into a spreadsheet and a Python script to calculate PCI values automatically. For further information on the process of digitizing the curves and generating PCI values, see Piryonesi and El-Diraby (2018). In addition, this process is explained in the Appendix (Fig. 18).

Choosing an Initial Set of Attributes

An initial set of 15 relevant attributes were chosen as candidates for use by the models. Unlike most related previous works, these attributes were not chosen based on mere engineering (physical) reasoning; rather, they were based on the costs for managing the acquisition and maintenance of their data. These attributes were selected through reviewing related literature, interviews with experts, and studying the AM plans and the databases developed by ten small municipalities in Ontario. Table 1 shows the initial 15 attributes selected. Most of the attributes are related to weather and basic information about the road segment.

Using all 15 attributes to train a model may not be convenient for all municipalities given that they have different data collection programs (El-Diraby et al. 2017). Thus, a set of algorithms were used to rank the relative importance of the selected attributes. The algorithms used were information gain, information gain ratio, correlation-based feature selection, Chi-squared, Gini index, weighting by rule, and symmetrical uncertainty. These algorithms assign a weight to an attribute based on its contribution to the prediction of the target variable (i.e., future PCI). For instance, the correlation-based feature-selection algorithm ranks attributes according to their correlations with the PCI, while the information gain algorithm ranks them based on the reduction they bring to entropy (Hall 1999; Hastie et al. 2009; Kira and Rendell 1992; Provost and Fawcett 2013). Moreover, the importance of attributes was ranked based on the outputs of the GBT. Notice that the criteria used by a GBT are not different from the one used by its base decision tree. The only difference is that the GBT averages out the importance of attributes over all base learners. As a result of this aggregation, the results of ranking attributes using ensemble learning algorithms are more reliable (Hastie et al. 2009). To this end, it is important to indicate a significant difference between traditional regression analysis and data analytics. We did not rely only on obvious ranking algorithms, such as correlation, but we also used the ensemble learning algorithms presented to ensure that we consider the nature of data-based analysis. This difference is not only a

Table 1. The initial set of 15 attributes

Field name	Description	Number of NULL examples (out of 943)
PCI0	The (initial) value of PCI at the time of analysis	0
AGE	Age of road (since construction)	0
PAVEMENT_TYPE	Type of pavement (as defined by the Federal Highway Administration in LTPP)	0
FREEZE_INDEX_YR	Calculated freeze index for year (in Celsius days)	0
MAX_ANN_TEMP_AVG	Average of daily maximum air temperatures for year	0
MIN_ANN_TEMP_AVG	Average of daily minimum air temperatures for year	0
TOTAL_ANN_PRECIP	Total precipitation for year (in mm)	0
FUNC_CLASS	Functional class of road (as defined by the Federal Highway Administration in LTPP)	0
FREEZE_THAW_YR	Number of freeze–thaw cycles per year	0
OVERLAY_THICKNESS	Thickness of the placed layer in rehabilitation	432
AADT_ALL_VEHIC_2WAY	Average annual daily traffic	382 (reduced to 32 after correcting missing data)
REMED_TYPE	Type of last remedial action (as defined by FHWA in LTPP)	200
REMED_ YEARS	Number of years since the last remedial action	210
CONSTRUCTION_NO	Number of conducted remedial actions	0
GBE	Granular Base Equivalence	5
PCI (target variable)	The class of PCI in the near future (as categorized by the ASTM)	0

fitting problem. Rather, it also includes, at least indirectly, a judgment about the variation in data quality (in the given set) and variations in future sets. This is why it is important to study the relative importance of attributes beyond the simplified fitting to the current sample.

It is important to indicate that LTPP attributes have different levels of quality. For instance, the attributes related to distress data or climatic data within the training set were complete and consistent, whereas attributes related to traffic and maintenance history were missing many values. This lack of values could be because the latter attributes are partially collected and reported by local highway agencies prior to the inclusion of the test sections in the LTPP study (Elkins et al. 2003). The last column of Table 1 shows the number of examples that are missing for each attribute. As shown in Table 1, traffic and maintenance history data are missing for a considerable part of the training set. Therefore, it is expected that lower quality would (unfairly) impact the importance of these attributes in predicting the target variable. Missing data is a common problem in the domain of pavement management. Some of the most common data imputation approaches to address the missing data are substitution by mean, linear interpolation, and using regression (Farhan and Fwa 2014).

Table 1 shows that the annual average daily traffic (AADT) data are missing for 40% of the examples of the training set. As a part of data cleansing, the missing traffic data were estimated by interpolating from the historical trends within the LTPP data. Therefore, the number of missing data points for the AADT was reduced from 382 to 31. Such data cleansing efforts can increase the accuracy of the models and analysis validity.

The ranks and the weights of different attributes in predicting the future PCI are summarized in Fig. 3. These graphs were developed from 943 examples. The initial PCI values and their corresponding deterioration within 3 years were calculated by applying our automated system to LTPP data at the start and end of a 3-year span. We then examined the possible co-variation (not necessarily correlation) between each attribute and PCI deterioration level to determine the sensitivity of these changes to any of the used attributes. Vertical axes represent the normalized relative importance of each attribute in predicting the future PCI. Obviously, all algorithms suggest that the current PCI (PCI0) is the most influential attribute in predicting the level of PCI after 3 years. This observation is the basis for using standardized deterioration curves—the assumption is that initial PCI and time are the only two factors to consider. However, this simplified initial screening shows that the effects of other attributes are not negligible. Specifically, attributes such as the last type of maintenance (REMED_TYPE), the number of years since last maintenance (YEARS_REMED), AADT, the age of road (AGE), and several climatic attributes should be considered. Note that the graphs of Fig. 3 are created after correcting for the missing AADT values. AADT was ranked as the seventh most informative attribute (average of the results of all ranking algorithms), and it was the eleventh prior to replacing the missing values.

Further cleaning of data was significant in the initial analysis of relative importance. We removed LTPP entries that were missing data about YEARS_REMED, and the revised rankings by each algorithm are shown in Fig. 4. Excluding the missing data added consistency in all entries and resulted in two changes. First, on average, the contributions of non-PCI attributes increased. Second, as expected, the removal of missing entries for a major maintenance attribute increased the overall importance of attributes related to maintenance history. This is another feature for data analytics; the results are sensitive to the quality and completeness of data.

In parallel, the *Optimize Selection* operator of the software Rapidminer was also applied to the dataset. This operator selects the most relevant attributes in a dataset. It has a heuristic approach and applies two deterministic greedy feature-selection algorithms named *forward selection* and *backward elimination* (Hastie et al. 2009; Rapidminer 2017). In a dataset with n attributes, this operator recursively selects m features ($m < n$) that maximize the accuracy of learned models (i.e., decision trees in this case). The algorithm tries different combinations of attributes. Therefore, applying it to a large training set with many attributes requires large computational power and hence a long processing time. Solving such a problem with exact algorithms would be impractical given its large size and computational complexity, but metaheuristic algorithms (e.g., genetic or Big Bang–Big Crunch algorithms) could be used to avoid some pitfalls of heuristic approaches. However, a discussion on solving the optimization problem of feature selection by different algorithms is beyond the scope of this paper and could be studied separately in the future.

The result of applying this operator to the attributes of Table 1 is presented in Table 2. A weight of 1 indicates that an attribute has a role in increasing the accuracy of a decision tree, while a weight of

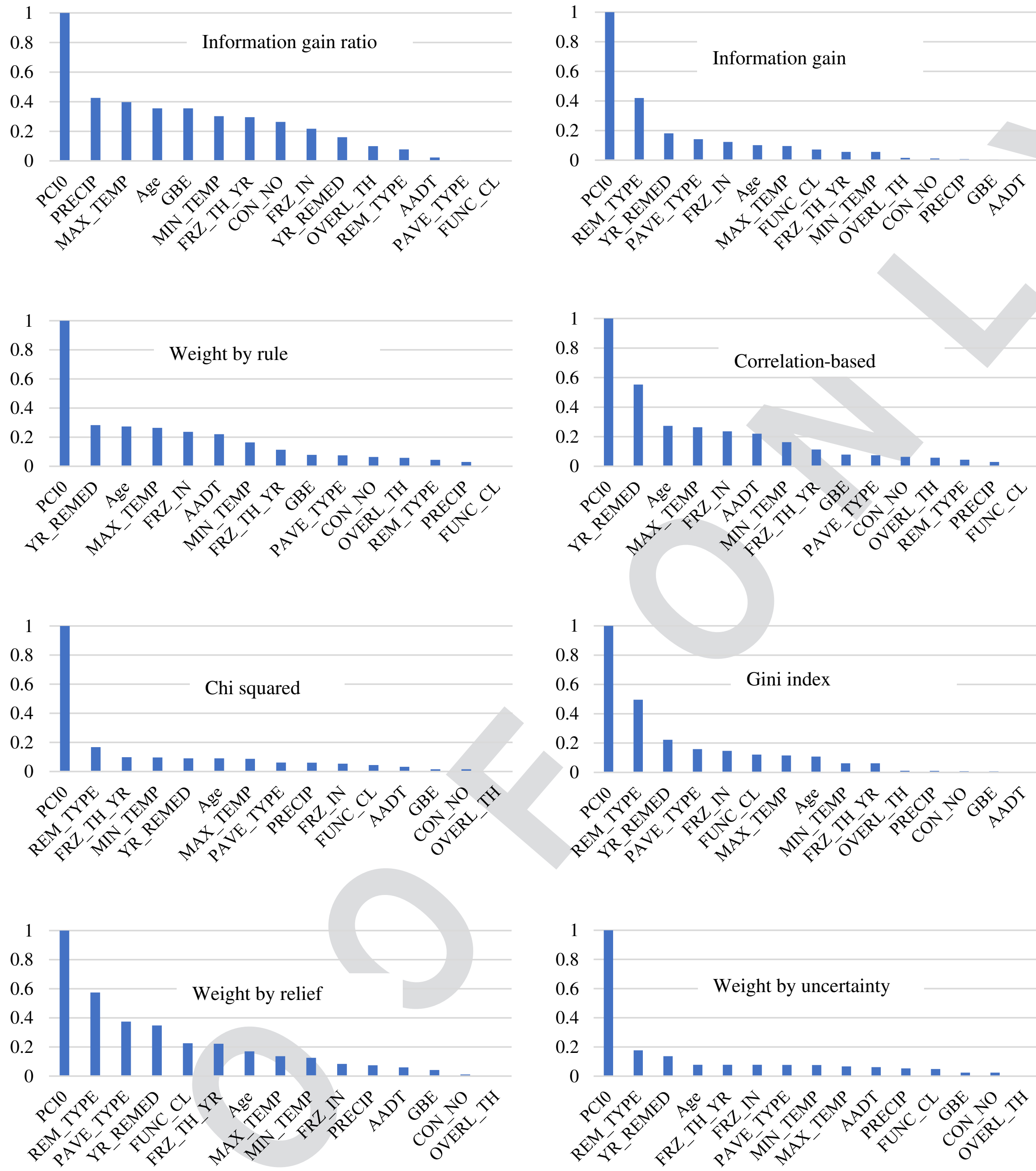


Fig. 3. Ranks and normalized weights assigned to each attribute by different ranking algorithms before removing missing examples.

0 means that it is possible to train a tree with the same accuracy without using that attribute. Table 2 confirms the collective results of the algorithmic analysis of importance (Figs. 3 and 4). Most attributes with a high rank in Figs. 3 and 4 have a weight of 1 in Table 2. It is worth noting that the *Optimize Selection* operator does not guarantee finding a global optimal because of its heuristic nature (Rapidminer 2017). By combining the results of all algorithms and the heuristic, the following attributes were identified as possibly the most informative and were used for training: initial PCI value, number of years since the last remedial action, number of freeze–thaw cycles, the age of road, AADT, total precipitation, granular base equivalence, annual freeze index, and average of daily maximum temperature.

Finally, the importance of the attributes was ranked using the GBT. As mentioned, this algorithm calculates the importance of each feature by averaging it over all base learners. The results of the GBT ranking are shown in Fig. 5. These results are similar to the results shown in Figs. 3 and 4 and in Table 2.

Models for Predicting the PCI

Using all attributes does not necessarily guarantee a more accurate model. Consequently, we used combinations of the ten attributes. Furthermore, simple models developed based on subsets of initial attributes can be reproduced and used by different transportation agencies with different data availabilities. We present here five

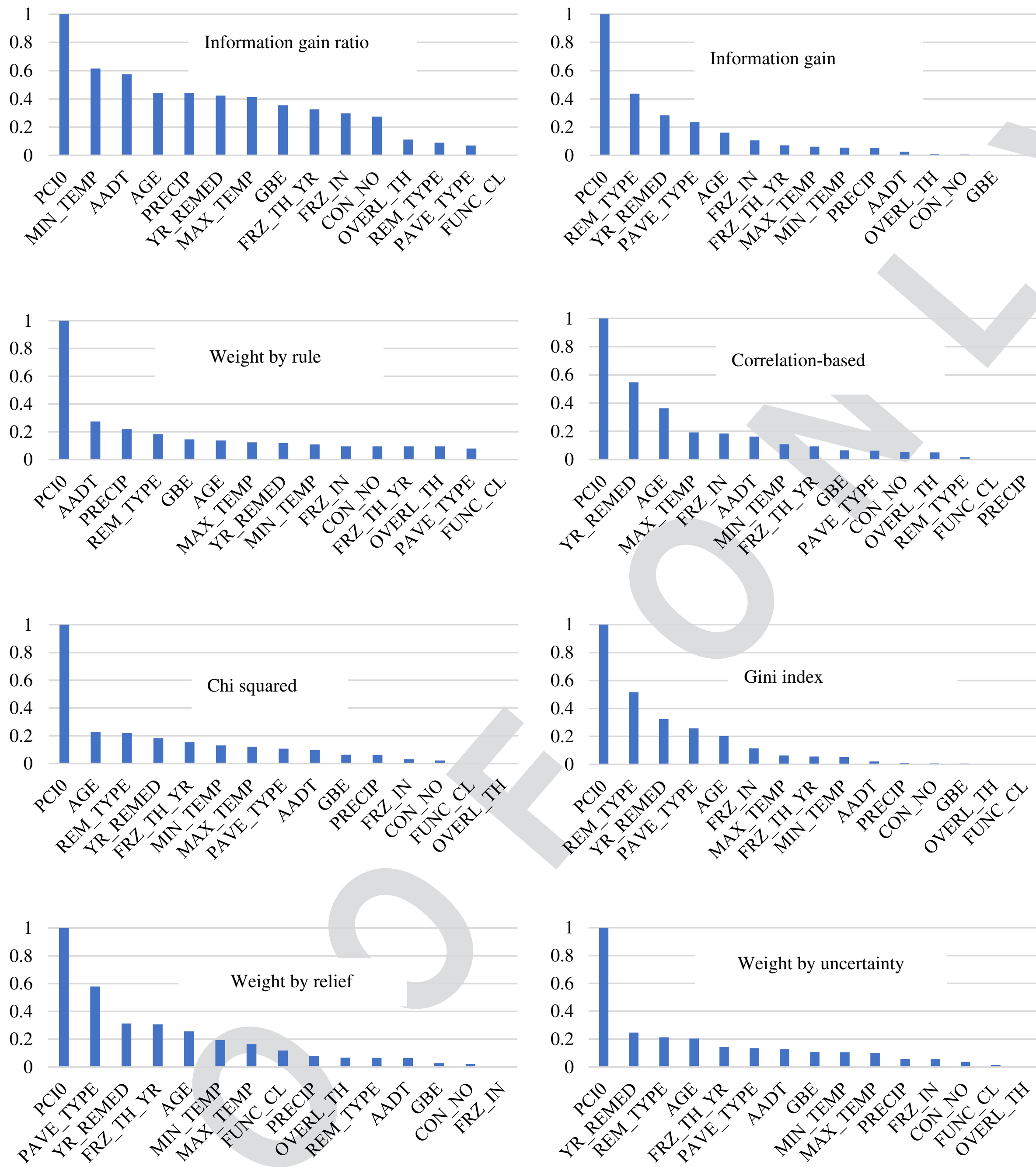


Fig. 4. Ranks and normalized weights of attributes by different ranking algorithms after removing missing examples.

models, and two of them were the most stable and accurate. Models 1, 4, and 5 each have two versions: one with five labels and one with seven. Using various combinations of the attributes provided insights into our analysis of the informativeness of attributes and helped different municipalities with different levels of access to data. Furthermore, variations in models can be beneficial to municipalities that may have only a subset of the attributes available.

The training set contained 942 examples of road sections. The target of prediction is PCI values after 3 years. For training a decision tree and a GBT, the target value must be discrete. Therefore, PCI values were discretized according to the ASTM rating scale (Fig. 1). As shown on the right side of Fig. 1, ASTM D6433-07 divides PCI values into seven classes (Way et al. 2015).

Rapidminer software includes a set of tested and frequently used decision trees. Two were specifically used: the default decision tree of Rapidminer (decision tree I) and a tree called C4.5 (decision tree II), recognized for its high performance (Chi et al. 2014; Wu et al. 2008). Unlike old decision trees, such as CLS and ID3, C4.5 is capable of learning from both continuous and categorical data. To illustrate this, we trained two models with both trees using all the attributes in Table 1. Fig. 6 shows the learning performance of both of these when using different data sizes. Note that Fig. 6 simply compares the learning capability of the two mentioned algorithms when applied to the training set. However, comparing the accuracy of the two models requires further evaluation.

Table 3 summarizes the information of a few different models learned from different attributes. Decision trees provide an open

Table 2. Assessing the initial 15 attributes using a heuristic process

Attribute	Weight
PCI0	1
AADT_ALL_VEHIC_2WAY	1
AGE	1
PAVEMENT_TYPE	0
FREEZE_INDEX_YR	0
MAX_ANN_TEMP_AVG	0
MIN_ANN_TEMP_AVG	0
GBE	1
TOTAL_ANN_PRECIP	0
FUNC_CLASS	0
CONSTRUCTION_NO	0
FREEZE_THAW_YR	1
OVERLAY_THICKNESS	0
Remed_YEARS	1
Remed_TYPE	1

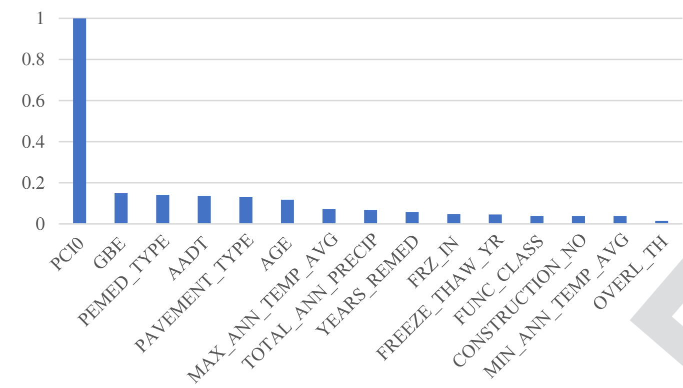


Fig. 5. Importance of attributes based on the GBT algorithm.

model for understanding the role of all attributes. To illustrate, Fig. 7 shows a snapshot of decision tree I for Model #2, and Fig. 8 shows tree II (learned from a C4.5 algorithm) for the same model. Fig. 8 shows that when the value of current PCI is smaller than 91.6 and larger than 85.1 (i.e., $85.1 < \text{PCI0} < 91.6$) and the road section has been maintained within the last 3 years (REMED_YEARS <= 3), PCI will be in a *Good* condition 3 years later. However, if the last remedial action occurred more than 3 years ago and the road has experienced more than 24 freeze–thaw cycles per year (FREEZE_THAW_YR > 24), its condition will fall to *Satisfactory*. Note that the entire tree is too large to be graphically represented here. Therefore, the rules of trees I and II (for model #3) were included in the Appendix.

Both types of decision trees and all models of Table 3 were tested multiple times with a similar number of examples and similar parameters: leaf size of 2 and confidence factor of 0.25. Decision tree II (i.e., C4.5) showed a higher accuracy on both training and test data (the next section provides the details of the model evaluation). Moreover, Table 3 contains the values of accuracy resulted from the GBT, and these values are substantially higher than those of the base learners. The GBT was generated using 20 base learners.

The hierarchy of the attributes within a decision tree is an approximate reflection of how informative the attributes are. Therefore, decision-makers can simply identify the most informative attributes upon testing a decision tree. This ease of interpretation is a great advantage of decision trees. The GBT is not as intuitive as decision trees (Hastie et al. 2009; Wu et al. 2008). Studying graphs such as Fig. 5 could help with understanding the results of the GBT (Hastie et al. 2009). For the convenience of users, the decision trees of all models were programmed into a web-portal using Python. Users can input the attributes of their road segments and climatic data and receive assessments about their future deterioration.

Model Evaluation

The accuracy of developed models (including all models of Table 3) was tested using cross-validation. Cross-validation is the best method to test the accuracy of models in predicting the label of unseen data, especially when the size of the training set is small. By using cross-validation, one can make sure that the model is not fitted to the noise of data. Usually, the accuracy resulted from cross-validation is the most critical accuracy (Hastie et al. 2009).

The training data was divided into 10 subsets. Recursively, nine subsets were used to train the decision trees, and the 10th was used to test accuracy. This process was repeated 10 times. The accuracy of the model is the average of the accuracies of the 10 iterations. Using cross-validation a standard deviation is calculable as well, making the analysis more reliable (these are reported in Table 3). Fig. 9 compares the cross-validation accuracy of the two decision trees for Model #5 (i.e., all attributes) with four sizes of training sets. The first training set included 110 LTPP sections, the second included 256, the third included 550, and the last included 942. Fig. 9 demonstrates that decision tree II outperformed the other in all four cases. The higher accuracy of the C4.5 decision tree agrees

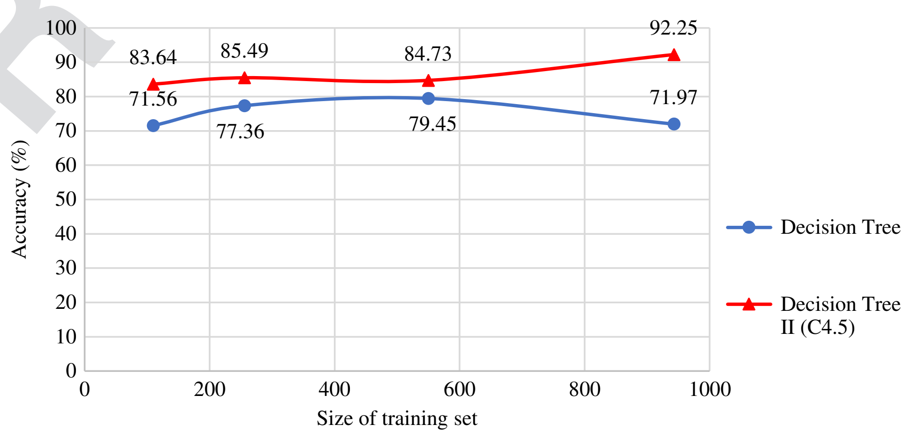


Fig. 6. Comparing the performance of decision trees: C4.5 outperforms decision tree I in predicting the class of PCI (5 levels) in 3 years.

Table 3. Different models trained by different combinations of attributes (numbers are rounded)

Attributes	Model #1 (5 levels)	Model #1 (7 levels)	Model #2 (7 level)	Model #3 (7 levels)	Model #4 (5 levels)	Model #4 (7 levels)	Model #5 (5 levels)	Model #5 (7 levels)
Initial PCI value	X	X	X	X	X	X	X	X
Age of road	X	X	—	—	X	X	X	X
Annual average daily traffic	—	—	—	X	X	X	X	X
Years since last remedial action	—	—	X	X	X	X	X	X
Type of remedial action	—	—	X	—	X	X	X	X
Functional class	X	X	—	—	—	—	X	X
Type of pavement	X	X	—	—	X	X	X	X
GBE	—	—	—	—	X	X	X	X
Number of freeze–thaw cycles	—	—	X	X	X	X	X	X
Annual freeze index	X	X	—	—	X	X	X	X
Av. daily max. temperature	X	X	—	—	—	—	X	X
Av. daily min. temperature	X	X	—	—	—	—	X	X
Annual precipitation	X	X	—	—	—	—	X	X
Overlay thickness	—	—	—	—	—	—	X	X
Number of remedial actions	—	—	—	—	—	—	X	X
Cross-validation Accuracy of C4.5	72 ± 5%	67 ± 5%	67 ± 3%	69 ± 4%	**78** ± **4**%	76 ± 4%	73 ± 4%	70 ± 4%
Cross-validation Accuracy of GBT	77 ± 3%	75 ± 3%	72 ± 3%	77 ± 6%	**82** ± **4**%	79 ± 3%	80 ± 4%	79 ± 3%

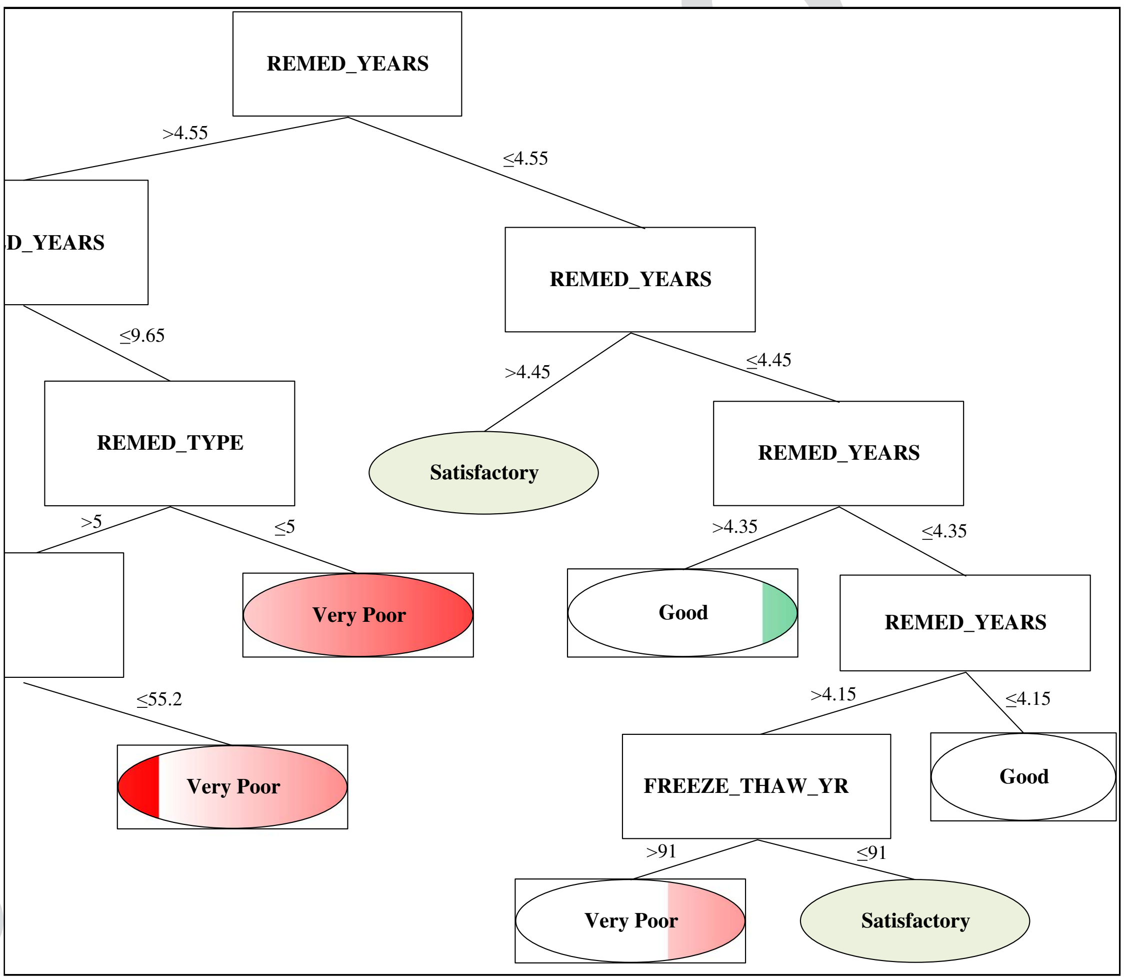
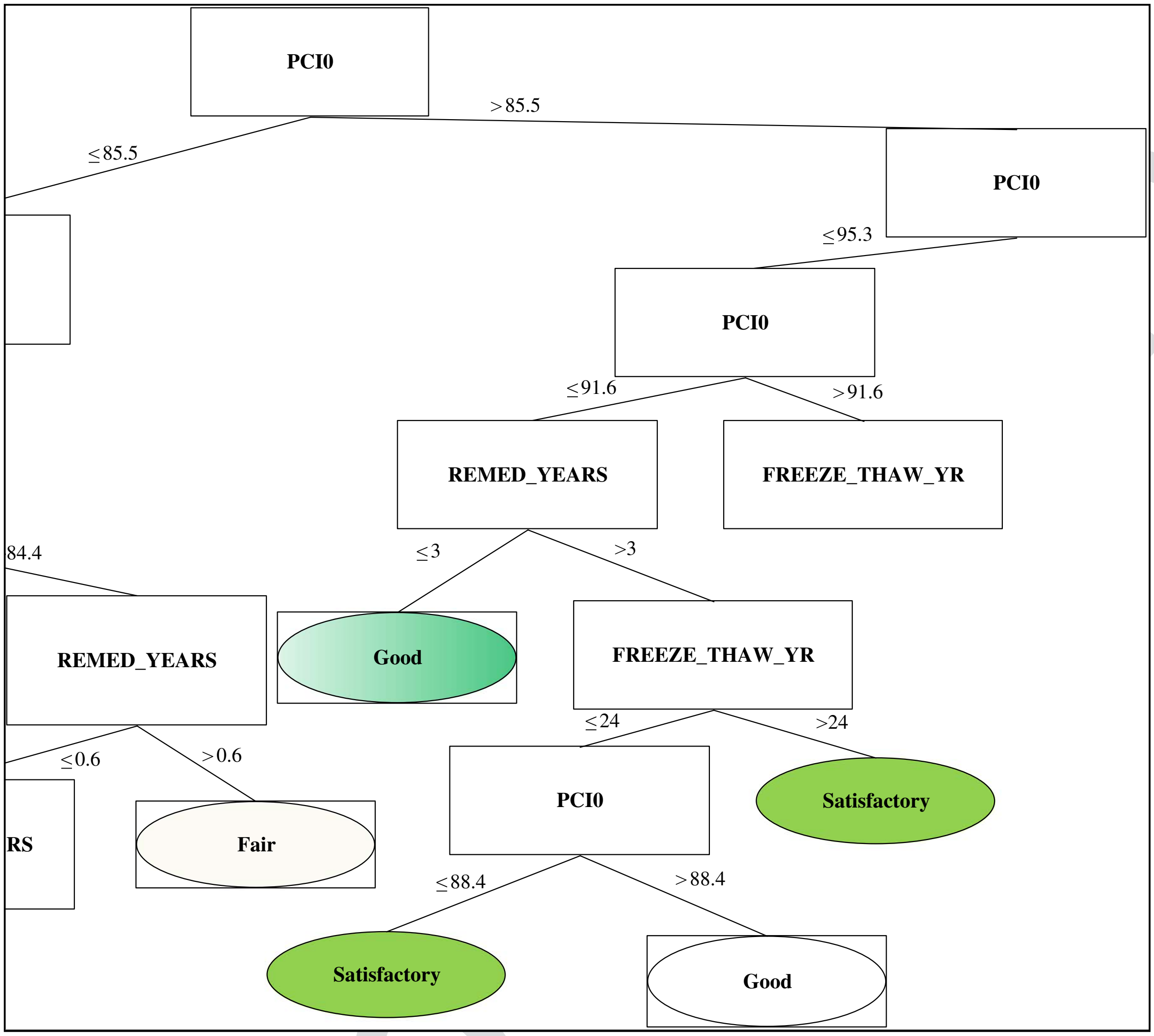


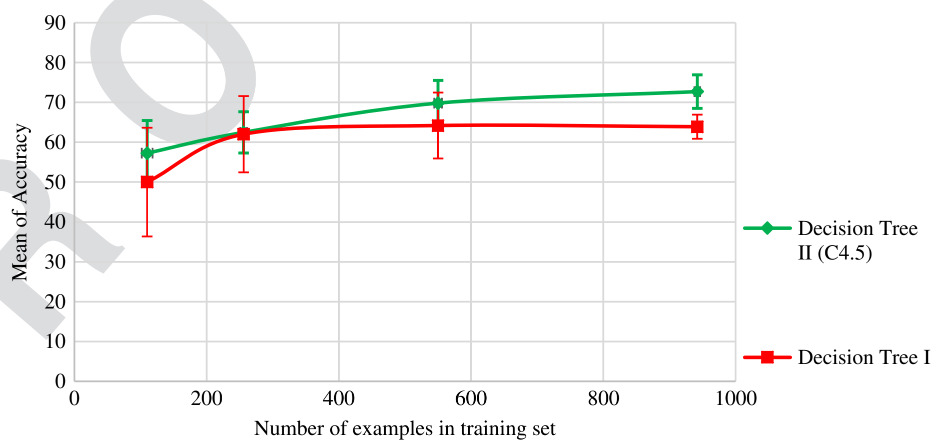
Fig. 7. Illustration of decision tree I for Model #2.

with the results of previous research such as the study by Chi et al. (2014). Furthermore, Fig. 9 shows that by increasing the number of examples, the mean accuracy increases and its standard deviation decreases. The standard deviation is shown by vertical lines over the curves. The curves prove that the models trained by larger datasets are more accurate and robust.

The graph in Fig. 9 shows the impact of data size on model performance. The type of algorithm and its parameters are important as



F8:1 **Fig. 8.** The illustration of a C4.5 decision tree (decision tree II) for Model #2.



F9:1 **Fig. 9.** Comparing the accuracy of two different decision trees; all attributes of Table 1 are used. (Model #5, 5-level.)

well. As indicated, in Table 3, the GBT outperforms a single decision tree. The GBT models of Table 3 were trained using 20 base learners. To better explore the performance of GBT, different numbers of trees were used. GBT models using 5, 10, 20, 50, and 100 trees were developed, and the results are presented in Figs. 10 and 11. These two figures clearly demonstrate that the ensemble of boosted trees reaches a higher accuracy in comparison with their base learner. Furthermore, increasing the number of base learners

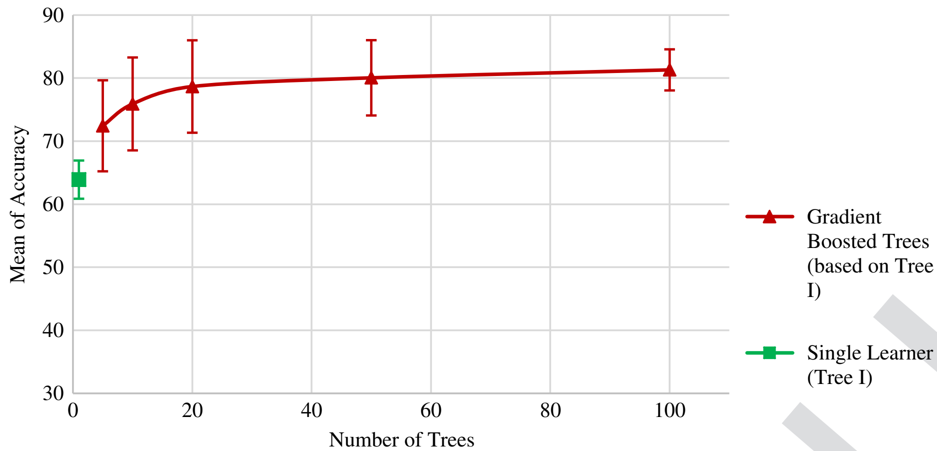


Fig. 10. The performance of GBT versus a single learner (decision tree I); increasing the number of base learners increases the accuracy.

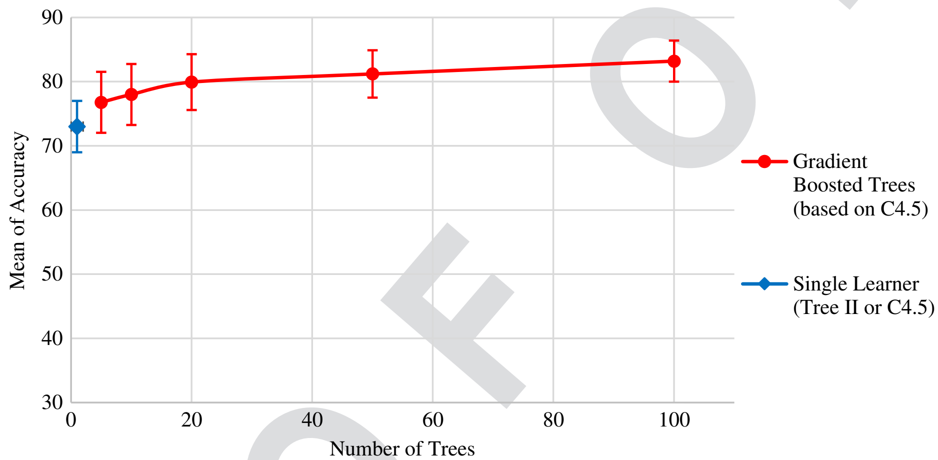


Fig. 11. Performance of GBT versus a single learner (decision tree II); increasing the number of base learners increases the accuracy.

Table 4. The confusion matrix for Model #1 (using the C4.5 algorithm trained by 942 examples)

	Actual (True) levels of PCI (calculated by ASTM method)							
Predicted PCI	Good	Satisfactory	Fair	Poor	Very poor	Serious	Failed	Class precision (%)
Levels	205	165	241	129	135	59	8	—
Good	180	38	18	1	1	0	0	75.6
Satisfactory	16	98	26	7	2	1	0	65.3
Fair	7	23	161	14	10	6	0	72.9
Poor	2	3	22	91	19	6	0	63.6
Very Poor	0	2	10	14	92	11	3	69.9
Serious	0	1	4	2	11	30	4	57.7
Failed	0	0	0	0	0	5	1	16.6
Class Recall	87.8%	59.4%	66.8%	70.5%	67.6%	50.8%	20.0%	—

(which is called the number of rounds in some packages) could enhance the accuracy. Finally, a more accurate base learner could result in a more accurate ensemble of classifiers.

Confusion matrices are very valuable in evaluating prediction models. Table 4 shows a confusion matrix that resulted from testing the accuracy of Model #1 with seven classes. The columns of a confusion matrix show how many examples belong to each class in reality, while the rows show the predictions of the model for each class. For example, as shown in Table 4, there were 59 sections with an actual rating (class) of "*Serious*." The model predicted 30 of these correctly, but the others were incorrectly predicted. For example, the model predicted six sections to be "*Fair*" and one to be "*Satisfactory*". Thus, the model *recalled* (correctly predicted) only 50.8%. The *class recall* for each of the seven classes used is shown in the bottom row of Table 4.

The developed model can detect the class "*Good*" with an accuracy of 87.8%, while the performance of the model in predicting "*Failed*" and "*Serious*" classes is lower. From a data analytics perspective, the reason for the high-class recall for "*Good"* and the low recall for "*Failed"* is the large number of existing *Good* segments and the small number of recorded *Failed* segments in the training set. This is because the percentage of roads in the real world with a *Failed* condition is very low. A practical solution for increasing the accuracy of this model is to merge the three lower classes (i.e., *Very Poor, Serious and Failed*) into one class (Fig. 1). Decreasing the number of classes is a common approach for enhancing the accuracy of classification models (Hastie et al. 2009). In our case, it is quite practical for two reasons. First, these three classes constitute only 21% of available sections in LTPP. This limited data subset will always result in poor training. More importantly, these three classes encompass PCI values that are below 40. Thus, second, in the real world, a finer assessment of PCI at this level is not needed given that roads with a PCI lower than 40 are in need of major rehabilitation or maintenance, i.e., finer classification does not add significant value.

Unlike the correlation-based and descriptive statistical measures, a confusion matrix can reveal the severity of the incorrect predictions (Hastie et al. 2009; Provost and Fawcett 2013). For example, in the class "*Fair*," 161 out of 241 were predicted correctly (class recall $= 66.8\%$). Among the incorrectly predicted 80 sections, none was categorized as "*Failed*," which would have been a major mistake; while four were categorized as "*Serious*," which is a major deviation. In contrast, 26 were categorized as "*Satisfactory*" and 22 were categorized as "*Poor*," which are the proceeding and succeeding classes, which is an acceptable outcome because this is not a gross deviation. Thus, the confusion matrix shows the distribution of mistakes and not just the overall percentage of successful predictions (which could conceal acceptable and very bad mistakes). Exploring the spectrum and nature of accuracy is one advantage of data analytics over the traditional one-indicator nature of traditional statistical analysis.

The overall accuracy of the model $69.32 \pm 4.7\%$. The results show that the model was learned and tested with a leaf size of 3. Consequently, its overall accuracy is slightly different from the Model #1 of Table 3. The mean accuracy (i.e., 69.32%) can be calculated by dividing the sum of diagonal elements of the confusion matrix (the correctly predicted ones) by the number of all predictions (942). This accuracy is better than a random guess given that because the target classes are seven, the basic odds of making a correct prediction (by random guess) is $1/7 = 14.3\%$. The $\pm 4.7\%$ value represents one standard deviation of the accuracy. However, comparing this accuracy to the results of the previous literature is not easy because most predictive models in the literature of performance modeling are defined based on regression (such as linear and nonlinear regression and neural networks) models rather than classifiers. Furthermore, using machine learning classifiers for predicting PCI values is somewhat rare in the literature. For example, this accuracy is higher than the numbers reported by Chi et al. (2014).

Defining Significant Attributes

Within each decision tree, several indicators can help determine the most significant attributes (the attributes with the largest impact). First, the indicators can help determine which attributes maximize the accuracy of trained models. Second, the indicators can help determine which attributes appear in a higher position in the tree hierarchy. For a detailed discussion on the importance of attributes within a decision tree that are assessed based on information gain, see Hastie et al. (2009) or Provost and Fawcett (2013). Third, the attributes that result in a more effective confusion matrix can be determined. The answers to the first and second questions from the results can be obtained by using the ranking algorithms and *Optimize Selection* operator, as discussed previously. For the third indicator, we must study overestimation and underestimation as the direction of deviation matters. Overestimating is misleading, as the actual PCI will be worse than what the decision-maker is expecting. In the case of PCI prediction, it can be interpreted as a *false-negative* prediction. Overestimation should then be avoided as much as possible. Thus, when comparing two models with (generally) the same accuracy, the one with a smaller number of false-negative predictions is more efficient.

For the models in Table 3, Fig. 10 shows the resultant accuracy (on the y-axis), standard deviation (the size of bubbles), and percentage of overestimation (x-axis) for each model. The combinations of attributes used in each model are available in Table 3. An additional model presented in Fig. 10, which is not in Table 3, is Model #3M. It is the same as Model #3, except for one difference: the data of traffic is still not corrected— i.e., missing values are not replaced with estimates of traffic. As a result, unlike Model #3, which is learned based on corrected AADT values, Model #3M is trained based on missing traffic data. As shown in Fig. 10, Models #1, Model #2, and Model #3M all have very similar overall accuracies. However, the number of false-negative predictions in Model #2 is considerably lower than its counterparts. Fig. 10 also shows that Model #3 has higher accuracy and a smaller percent of overestimations than Model #3M. This improvement is a result of correcting traffic data. Lastly, the results show that Model #4 has the highest accuracy and the smallest percent of overestimation and standard deviation; hence, it is the best model. Furthermore, all solid bubbles of Fig. 10 represent the models predicting the level of PCI after 3 years out of seven classes, while bubbles with a texture represent models learned by five classes. Both Table 3 and Fig. 10 show that the models learned to predict a smaller number of classes are more accurate.

Before discussing the relative importance of attributes across all models, it is important to note that the aim of this study, much like most data analytics work, is not necessarily to discover or validate a causal relationship. Rather, the objective is to develop a prediction based on observing trends in data. Thus, our analysis here relates to which attributes enhance the prediction of deterioration rather than cause it. In other words, when considering the relationship between attributes and accuracy, we are looking for informativeness and not necessarily causation or even correlation. In general, data analytics is a mechanism that helps to discover what is true when deterioration happens without asserting that it caused it. For example, data analytics predicts that it is raining from the number of umbrellas in the streets or from delayed traffic. Neither caused the rain. The number of umbrellas is definitely more informative than traffic delays, as the latter could be related (not necessarily caused) to many other events.

Considering the results of ranking algorithms, the hierarchy of developed decision trees, and the results of the analysis of the confusion matrices, the following observations are made. It is obvious that the functional class of road, the thickness of the last-placed overlay were the least informative attributes. They were followed by the average daily minimum temperature and the number of maintenance interventions throughout the life of a road. In the case of overlay thickness, this small informativeness could be a result of missing data for 45% of the examples rather than a natural irrelevance of the attribute. Thus far, the initial PCI is the most informative attribute in all models, and this result agrees with the available literature. Attributes related to maintenance history (i.e., type of maintenance and number of years since the last remedial action) could be the second most informative set of attributes. They had a high rank according to ranking algorithms and contributed to

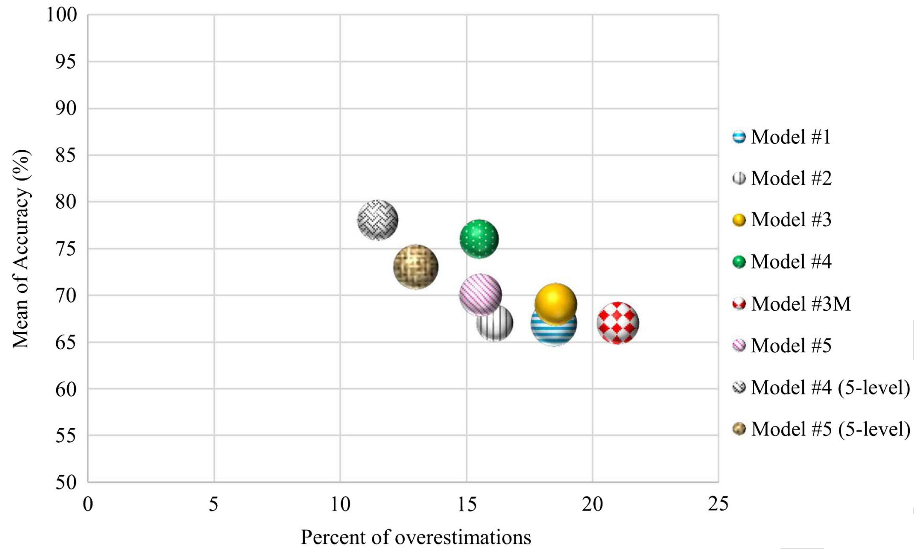


Fig. 12. Comparing the performance of models based on accuracy and percent of overestimation. The size of bubbles represents the standard deviation of accuracy. Solid bubbles are learned based on 7-classes.

increasing the accuracy of the models. This observation confirms the importance of maintenance history in predicting the deterioration of roads. Average annual traffic is significant. In addition, correcting the traffic data resulted in increased accuracy and a decreased number of false-negative predictions. Corrected AADT attributes appeared in Model #4, which had the highest accuracy. The pavement type proved to be an important attribute according to most performed analyses. This observation agrees with the literature of pavement deterioration modeling, and total precipitation and annual freeze index were informative attributes in several models. However, the number of freeze–thaw cycles was more informative. Perhaps the alternation in temperatures (near the freezing level) was more important than the temperatures or the freezing itself.

Role of Age in Predicting the Future PCI

Age (since construction) was identified as one of the most informative attributes. However, logically, it should be irrelevant once significant rehabilitation occurs. "Years since last remedial action" is more significant and should have trumped absolute age. This situation is related to an interesting and very important issue: what is the data measuring. The data available are not laboratory data of aging pavements but data about actual aging pavements, which are highly affected by the maintenance regime. In other words, the appearance of absolute age as an important attribute is more of a reflection of the maintenance culture than technical attributes. In other words, the data we have correspond to physical deterioration and maintenance patterns and adequacy. The results show that older roads deteriorate faster than more recent ones because older roads never receive a perfect maintenance regime. This does not mean no road goes through major rehabilitation, which would re-establish PCI to 100. Instead, in older roads, the number of roads that do not get complete overhauls or equivalent maintenance is, as expected, relatively large. In essence, there are two types of age: the overall age, which indicates the overall upkeep of an average road; and "years since last remedial action", which if small is purely a reflection of natural (physical) deterioration. Thus, there are two relevant issues about data analytics: a deep contemplation is required to understand what the data are corresponding to and the need to avoid causal inferences.

To better explore the role of maintenance in highlighting absolute age as a major attribute, a subset of the dataset containing only recently rehabilitated roads was studied. This subset included 256 road sections that were rehabilitated within the last 5 years. The previous ranking algorithms were applied to the recently rehabilitated roads. The results are presented in Fig. 13, with the attribute AGE highlighted. Fig. 13 clearly suggests that the attribute AGE is not among the most important attributes anymore. The average rank of attribute AGE dropped from three or four in Fig. 4 to eleven in Fig. 13. Another attribute that has a significantly different weight is OVERLAY_THICKNESS. Not only its weight increased, but also its rank improved from twelve in Fig. 4 to eight in Fig. 13. This difference is because of the smaller number of missing examples in this subset (around 3%). Another observation is that the weight of most predictive attributes in Fig. 13 is larger than Figs. 3 and 4. Thus, the contribution of attributes to predicting future PCI increased on average. Similar changes occurred when ranking the importance of attributes using the GBT. The rank of attribute AGE dropped from the third (in Fig. 5) to seventh. The cross-validation accuracy of applying Model #5 to recently rehabilitated roads was $75 \pm 7\%$, which is considerably larger than its corresponding accuracy in Fig. 9. This difference could be because of the larger weights of predictive attributes in this subset of the training set.

Different Climatic Zones

The training data was retrieved from the LTPP dataset, which contains roads in the United States and Canada, with drastic differences in climatic conditions. Therefore, it is expected that conducting similar analyses on the data of a particular road network or limited geographic spread would result in models with better accuracy (Lou et al. 2001; Wu 2015). For example, the reduction in the variations in weather indicators may provide more insights into the role of other attributes. To address this issue, the training set into four subsets based on the climatic zones defined by the LTPP. Table 5 summarizes the information about the number of road sections in each climatic zone, and the average deterioration of PCI within 3 years is presented in Fig. 11. The figure also presents the median and the standard deviation of the change in the PCI for

F13:1 **Fig. 13.** Results of ranking algorithms on the roads rehabilitated within five years.

different climatic zones. The wet-freeze climatic zone has the largest average deterioration. The deterioration of PCI also has the largest variance in this climatic zone.

Model #5 was trained and tested for different climatic regions. The cross-validation accuracy of the model for all climatic zones is given in Table 6. The numbers within Table 6 clearly indicate the accuracy of Model #5 trained based on different climatic regions is better than the results of the aggregate model shown in Fig. 9. According to Fig. 9, when the size of the training set is smaller than 400 examples, the mean of the accuracy is smaller than 65%. Considering the climatic zones with similar or less size, the accuracy of each model is higher than 65%, with zone #2 reaching an accuracy of 83%. A relatively high accuracy with these smaller small sizes was observed when we studied the subset of recently rehabilitated roads.

The higher cross-validation accuracies of Table 6 are the result of eliminating some of the noise from the dataset. Furthermore, the reduction in randomness helped shed some light on the relative importance of attributes. These points are illustrated in the Appendix (Fig. 19).

In addition to enhancing the accuracy, separating the dataset was insightful in understanding the relative importance of attributes. In general, there are three main sets of attributes that have bearing on deterioration prediction (not necessarily causing it). The first set is age and maintenance, including "years since last remedial action" and "type of last remedial action." We combine these into one set because age is implicitly an indicator of maintenance history or adequacy in the average road. The second set comprises climate attributes. Obviously, freezing (and freeze–thaw cycles) affects the accuracy of prediction. Alternatively, in areas with no freezing, the difference between the maximum and minimum temperatures can affect the deterioration predictions. Third, AADT is the third set.

Our work helps explain deterioration prediction. Typically, it was overwhelmingly based on a physical model: aging (and lack of maintenance) make roads deteriorate. Our work, without

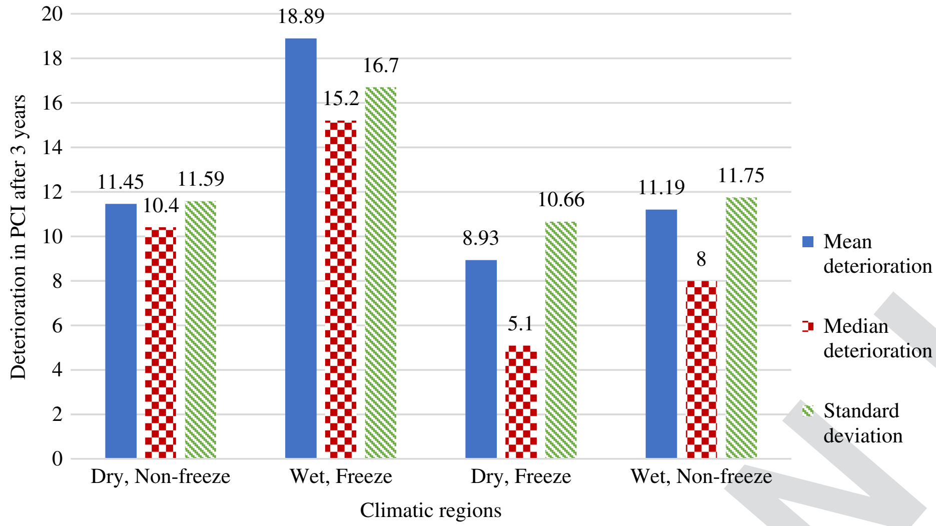


Fig. 14. Aggregate statistical measures for deterioration of the PCI in different climatic regions within 3 years.

Table 5. Number of roads in each LTPP climatic zone

Code	Climatic zone	Number of examples
1	Dry, freeze	151
2	Dry, non-freeze	213
3	Wet, freeze	228
4	Wet, non-freeze	350
Total		942

Table 6. The cross-validation accuracy of Model #5 for different climatic zones

Climatic zone	Climatic region	Cross-validation accuracy	Number of examples
1	Dry, freeze	70% ± 8%	151
2	Dry, non-freeze	83% ± 5%	213
3	Wet, freeze	72% ± 9%	228
4	Wet, non-freeze	67% ± 7%	350
Total			942

claiming causation, may help explain the role of weather and traffic on the deterioration. It is interesting to note that studies that considered only physical attributes using statistical approaches may have corroborated the results (Fig. 20).

Predicting PCI for Other Intervals

The same approach was used to predict the PCI deterioration in 2, 5, and 6 years. Such analyses could be helpful to municipalities in medium-range decision making. The analyses also help us explore the sensitivities of models and could also help researchers understand the differences between the long-term and the short-term deterioration of roads. Fig. 15 shows the number of examples that we could extract for training for each time horizon. Obviously, for a longer horizon, a smaller size dataset was available. This is mainly because as time increases, there is a good chance that the road section will be maintained (not too many sections stay without maintenance for a long time).

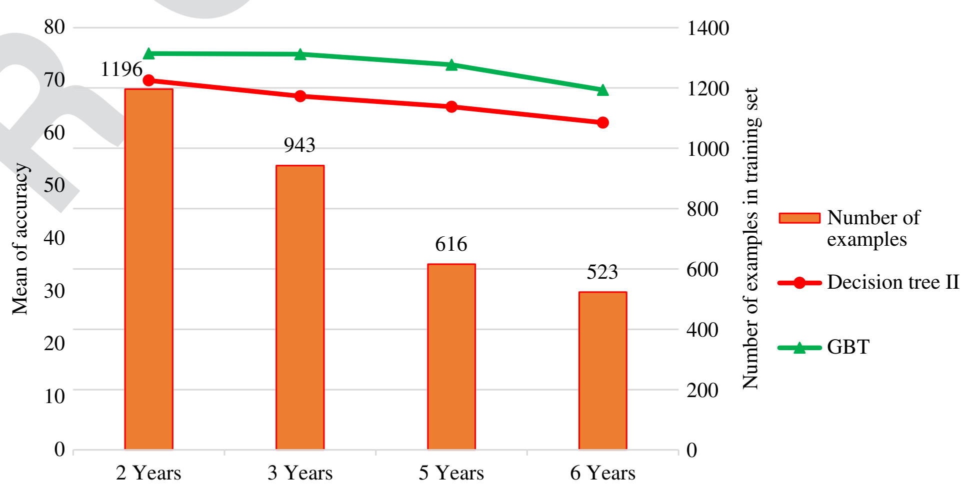


Fig. 15. The number of examples in the training set for predicting PCI at different intervals and the associated accuracy of decision tree II and its boosted ensemble.

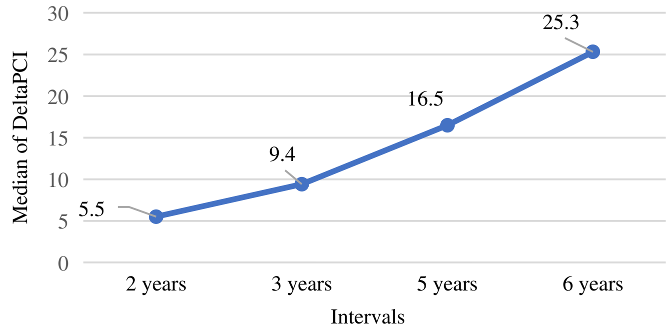


Fig. 16. Median of change in PCI value (DeltaPCI) within different prediction horizons.

After preparing and cleaning data, models were trained for each of the presented intervals. A major pattern that was observed was the higher accuracy for smaller intervals of prediction. This pattern is clearly visible in Fig. 15. The vertical axis on the left shows the mean of accuracy, and the one on the right shows the number of examples in the training set. The bars show the number of examples in the training set for each interval, and the curves show the mean of cross-validation accuracy. The red curve dotted with circles shows Model #1 learned from decision tree II for predicting seven classes of PCI, and the green curve dotted with triangles represents its boosted version. The results show that the GBT performs better than the single decision tree. The lower accuracy for longer intervals could be explained as follows. First, from a data analysis perspective, the size of the training set for longer intervals is smaller, impacting the accuracy of the model. Second, from an engineering perspective, in a nonlinear phenomenon such as deterioration, more variations can occur in longer time periods. It is expected that a larger number and variety of distresses could occur and their extent could take more divergent values. Fig. 16 shows the median change in actual PCI (i.e., DeltaPCI) for different intervals. As shown in Fig. 16, generally, by increasing the interval of prediction, the actual deterioration in PCI tends to increase. In other words, the PCI deterioration has a larger dispersion in longer intervals.

Other models showed larger accuracies in predicting the PCI over larger time intervals. For instance, Fig. 17 shows the mean of cross-validation accuracy of Model #5 in predicting PCI after 5 years. The figure clearly suggests that the accuracy of the model in predicting five levels of PCI is better than when predicting seven levels. However, Fig. 17 includes another interesting insight with respect to data quality and quantity. With regard to quantity, the literature suggests that a larger training set usually results in more accurate models (Hastie et al. 2009; Provost and Fawcett 2013). This notion was confirmed by Fig. 9. However, less research is available about assessing the impact of data quality on the accuracy of models. The quality issue is specifically serious in the domain of road infrastructure management. This result is because, as reported by a recent study in Ontario, a significant number of municipalities either do not assess the quality of their data or have no well-defined procedure to do so (El-Diraby et al. 2017).

We examined the impact of removing Null data—LTPP fields with no data reported in them. Fig. 17 is an attempt to visualize this. The x-axis is the number of examples in the training set, and the left vertical axis is the accuracy. The right vertical axis is a counter for the number of NULL data fields. Within the chart, the top curve shows testing with 5-level PCI, while the lower curve is with 7. The inclined dashed line tracks the number of missing data, and we tested five cases. Each is shown by three dots (one on each curve). The training set originally included 616 examples (case 1), and we removed the examples holding missing records for AADT, granular base equivalence (GBE), "years since remedial action" and "overlay thickness" one after the other. Every time that a cluster of missing data points was removed, the accuracy of the model increased. Therefore, the trend in Fig. 17 can help assess the

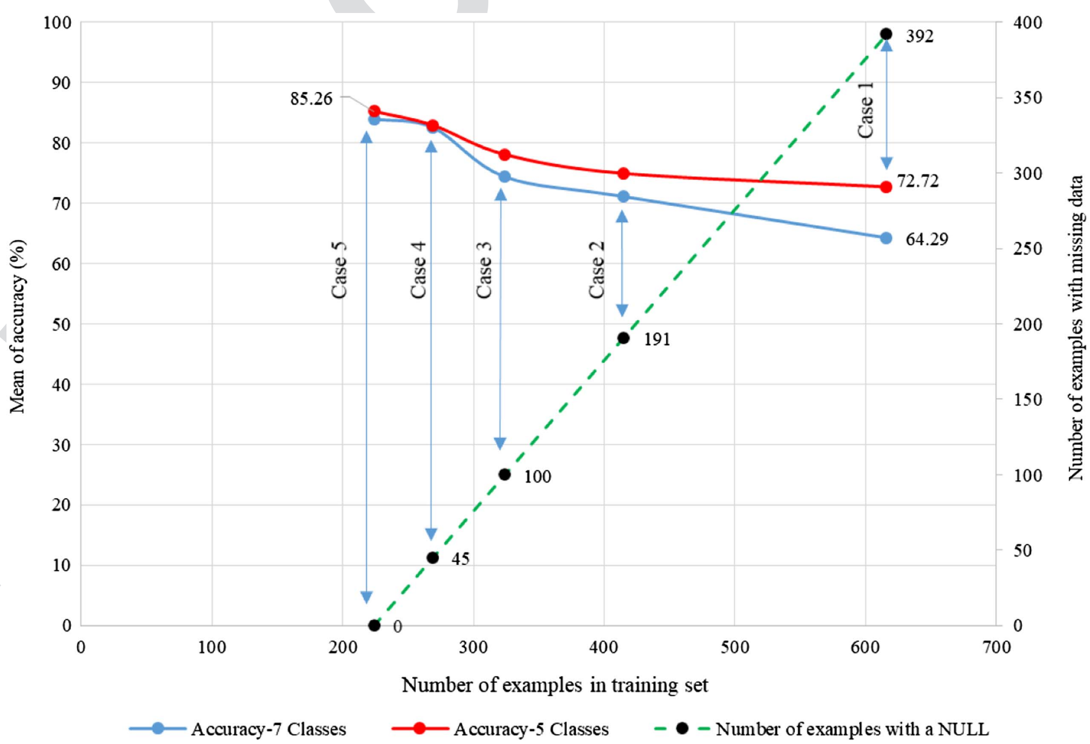


Fig. 17. Cross-validation of Model #5 in predicting the PCI after 5 years; having more examples with missing data decreases the accuracy.

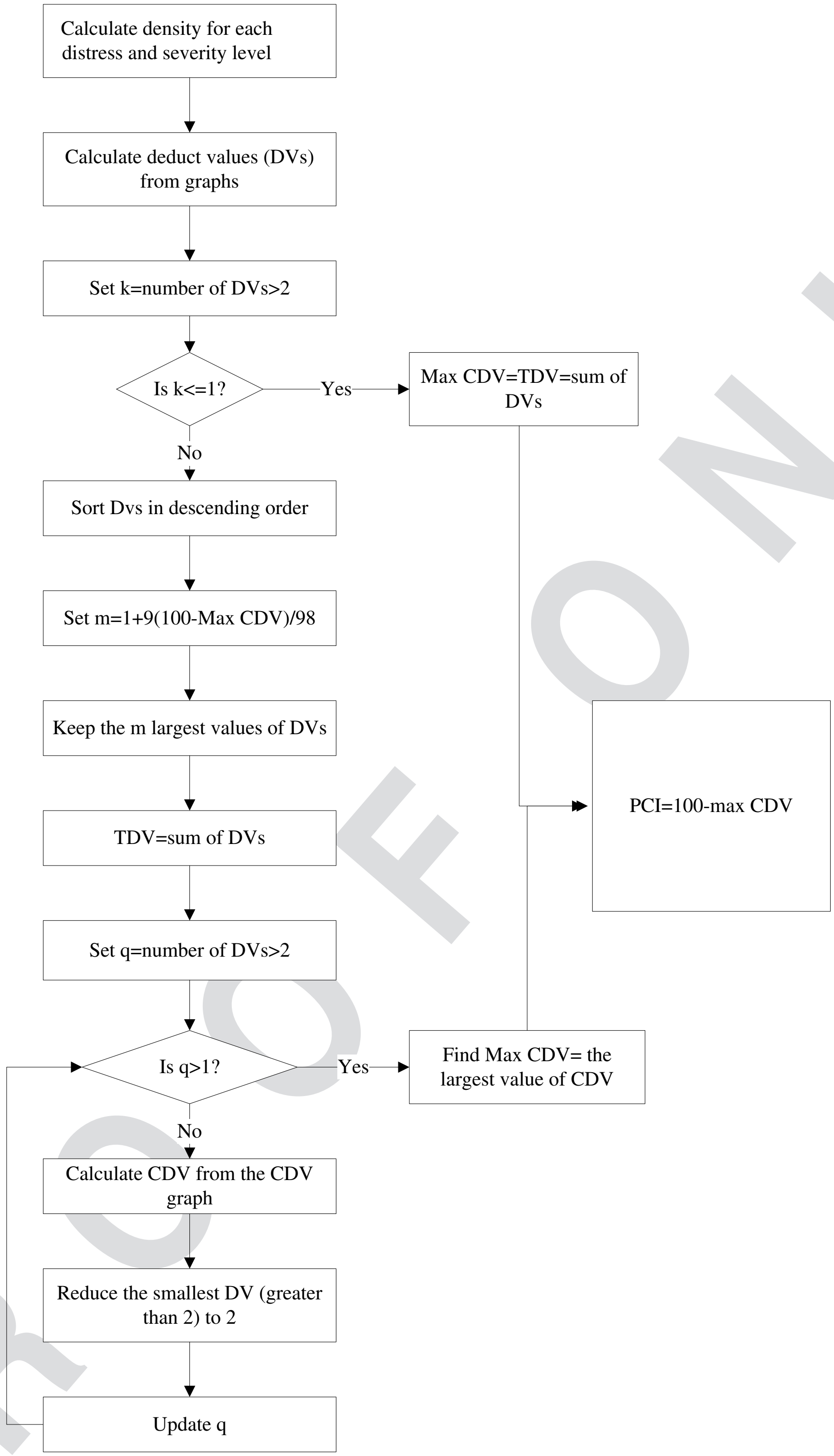


Fig. 18. Process of calculating the PCI from distress values.

magnitude of the impact of the lack of data completeness (our reference to quality) on the accuracy of models.

Conclusion and Recommendations

Combinations of 15 different attributes were used for training decision trees and GBT to help predict the PCI over 3 years. We also trained and tested the models for 2, 5, and 6 years. Shorter time horizons were associated with higher prediction accuracy. Several models were trained to predict the PCI value on either a 5-level and 7-level scale. The 5-level scale was always more accurate. All the developed models are probabilistic, making them easy to integrate into a risk assessment framework.

The cross-validation accuracy of some of the models for predicting future PCI values reached around 85%. The best accuracy of a

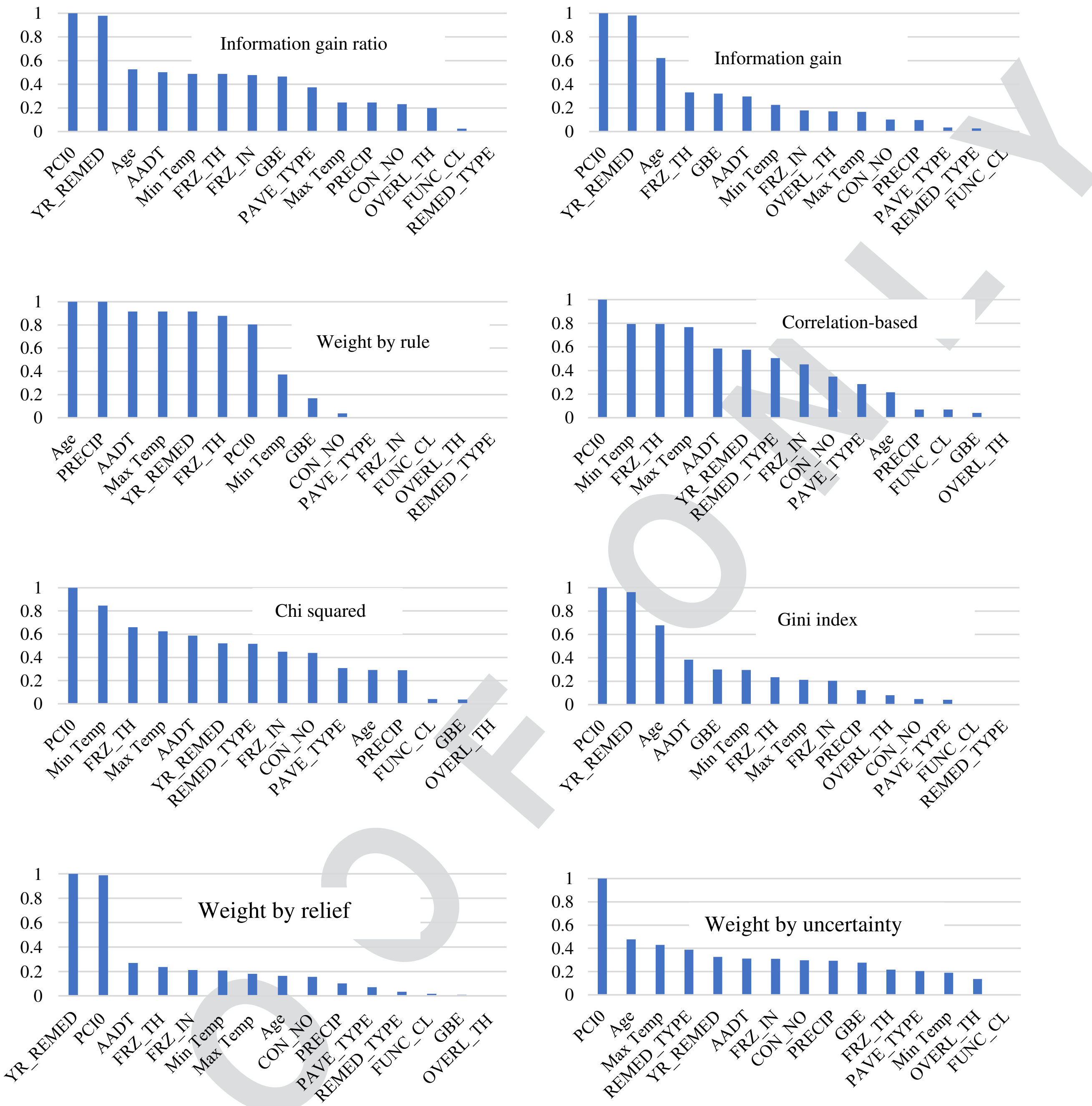


Fig. 19. Ranks and normalized weights assigned to each attribute by different ranking algorithms for Climatic zone 2 (dry and nonfreeze); the weights of most attributes are substantially higher than Figs. 3 and 4.

model predicting PCI within 3 years reached 78 ± 4% with 5 classes and 76 ± 4% with 7 classes. This accuracy was higher when dividing the data into climatic regions. As expected, the accuracy of the ensemble of classifiers was considerably higher when compared to single learners.

The methods and results of this paper can help municipalities determine their priorities/plans in regard to data collection and maintenance. The results of our analysis suggest that the maintenance/rehabilitation history data has an important role in predicting the condition of roads. Therefore, other than condition data, one of the pieces of data that municipalities should track regularly is the time and type of remedial actions.

This study was conducted based on the LTPP data, which are collected within a very large geographic zone. This collection may have increased the disparity of data because of large variations in weather conditions and different data and management practices. Therefore, data were divided into smaller geographical zones for further investigation. Notwithstanding the smaller size of training sets for each created subset, the accuracy of developed models was considerably higher.

Data analytics is a process as well as a tool. As municipalities start to experiment with the use of data analytics in their decision making, we designed the scope of this work to include several steps and features to expose some of the advantages and limitations of data analytics. The quantity and quality of data had obvious effects on the accuracy. Specifically, missing or inaccurate data play a role in qualifying the importance of analysis attributes for overall model accuracy. Data cohesion was also a factor in

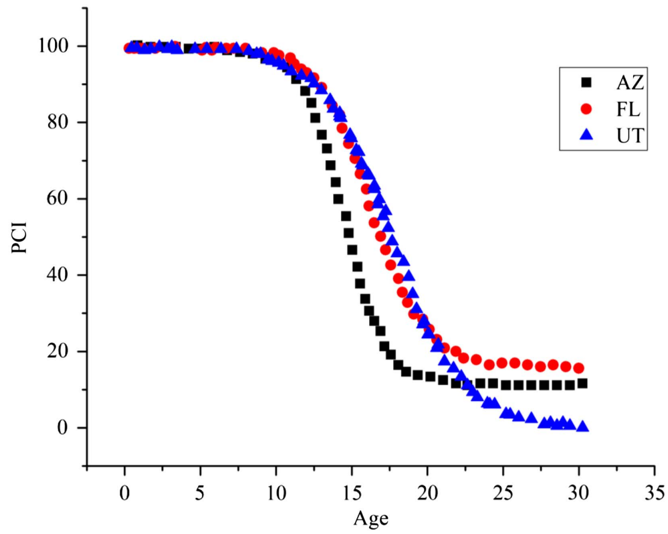


Fig. 20. Three PCI master curves developed based on the LTPP dataset; adapted from Wu (2015).

enhancing the quality of models and in assessing the relative importance of attributes. In the case of PCI, this was manifested through separating the data based on climatic zones. This separation reduced the noise in data (the variance or spread). With a smaller dataset, this more consistent data produced very good levels of accuracy. To this end, it is important to recognize that data analytics is about discovering trends. Thus, it is important to consider several sub-datasets and variations of modeling objectives and attribute combinations.

Of greater importance is to understand the nature of data. In the case of PCI, the "age since construction" attribute was frequently important, while many sections included in the analysis were maintained and attributes such as "years since last remedial action" were also important. We explored that the data are actually data about the physical features of roads and, implicitly, about the pattern, frequency, and adequacy of maintenance. The age of the newly built road did not matter. So, should the age of a newly maintained road. The fact that the attribute age was an important attribute is more related to maintenance patterns than the number of years. Finally, data analytics is an attempt to find patterns of co-variation: what things happen together. It would be a mistake to infer causation every time co-variation is observed. Cases should be excessive in transferring data analytics from the sphere of prediction to the sphere of the causal model.

Finally, this study included several climate stressors such as precipitation, temperature, and freeze–thaw cycles. These characteristics could help with a better understanding of the impact of climate on the pavement. This type of analysis is especially important for areas such as Canada, which will be highly affected by climate change. In the course of future research, this study could be expanded to consider the impact of climate change.

Appendix. Automating PCI Calculation

PCI Calculation

This appendix includes the information that was not included in the main body of the paper for brevity. First, the process of calculating the PCI was not explained in detail. The process of digitizing the curves and finding formulas is available at Piryonesi and El-Diraby (2018), and the general repetitive process of translating distresses into PCI is shown in Fig. 18.

Weight of Attributes for Climatic Region 2

Fig. 19 shows the weight of predictive attributes for climatic zone 2. By comparing Fig. 19 with Fig. 3, there are substantially higher weights of attributes in Fig. 19. These higher weights mean that most attributes, specifically the number of years since the last remedial action and the AADT, have larger roles in predicting the future PCI and, thus, a higher accuracy ($83 \pm 5\%$).

Fig. 20 includes three PCI master curves developed by Wu (2015) based on the LTPP data. The curves show the deterioration of roads in three different states. Generally, the roads within Arizona (AZ), which is a dry and nonfreeze climatic zone, are expected to have slower deterioration. However, Fig. 20 shows that they deteriorate faster than roads in Florida (FL) and Utah (UT). Wu (2015) argues that the faster deterioration of the AZ curve is because the road sections used to develop it have more traffic. This figure reveals one of the weaknesses of PCI master curves, which is the lack of flexibility in considering attributes such as traffic or the formal maintenance history (instead of being implicitly represented by age) for each road section.

Decision Tree Rules

Since showing the entire decision tree graphically is not possible, the rules of decision tree I and II for Model #3 are presented here.

Decision trees II (Model #3)
PCI0 <= 85.1
| PCI0 <= 57.6
| | PCI0 <= 46.4
| | | PCI0 <= 34.3
| | | | PCI0 <= 18.2
| | | | | PCI0 <= 15.4: Serious (4.0/1.0)
| | | | | PCI0 > 15.4: Failed (2.0)
| | | | PCI0 > 18.2
| | | | | AADT <= 6917
| | | | | | FREEZE_THAW_YR <= 113: Serious (16.92/2.0)
| | | | | | FREEZE_THAW_YR > 113: Very Poor (2.0)
| | | | | AADT > 6917
| | | | | | AADT <= 18200
| | | | | | | YEARS_REMED <= 2.5
| | | | | | | | PCI0 <= 24.7: Serious (3.65/0.38)
| | | | | | | | PCI0 > 24.7: Very Poor (2.65/0.53)
| | | | | | | YEARS_REMED > 2.5: Very Poor (10.51/3.01)
| | | | | | AADT > 18200: Serious (5.26)
| | | PCI0 > 34.3
| | | | AADT <= 4290
| | | | | PCI0 <= 44.2: Poor (5.28/1.28)
| | | | | PCI0 > 44.2
| | | | | | AADT <= 1450: Serious (2.14)
| | | | | | AADT > 1450: Very Poor (2.14/0.14)
| | | | AADT > 4290
| | | | | AADT <= 34500: Very Poor (52.06/3.53)
| | | | | AADT > 34500: Serious (6.38/1.19)
| | PCI0 > 46.4
| | | AADT <= 32760
| | | | FREEZE_THAW_YR <= 107
| | | | | FREEZE_THAW_YR <= 51
| | | | | | FREEZE_THAW_YR <= 16
| | | | | | | FREEZE_THAW_YR <= 0
| | | | | | | | PCI0 <= 52.6: Poor (5.0)

```
| | | | | | | | PCI0 > 52.6: Fair (2.0/1.0)
| | | | | | | FREEZE_THAW_YR > 0: Very Poor (2.0)
| | | | | | FREEZE_THAW_YR > 16: Poor (5.0)
| | | | | FREEZE_THAW_YR > 51
| | | | | | YEARS_REMED <= 3.9
| | | | | | | YEARS_REMED <= 0.7: Very Poor (4.49/1.82)
| | | | | | | YEARS_REMED > 0.7: Poor (7.49/1.39)
| | | | | | YEARS_REMED > 3.9: Very Poor (14.97/4.75)
| | | | FREEZE_THAW_YR > 107: Poor (31.95/5.0)
| | | AADT > 32760
| | | | PCI0 <= 54.2: Serious (2.11/0.11)
| | | | PCI0 > 54.2: Failed (2.0)
| PCI0 > 57.6
| | PCI0 <= 70.6
| | | PCI0 <= 64.6
| | | | FREEZE_THAW_YR <= 130
| | | | | PCI0 <= 63.1
| | | | | | FREEZE_THAW_YR <= 56
| | | | | | | FREEZE_THAW_YR <= 31: Poor (11.0/3.0)
| | | | | | | FREEZE_THAW_YR > 31
| | | | | | | | PCI0 <= 60.7: Fair (6.0)
| | | | | | | | PCI0 > 60.7: Satisfactory (3.0/1.0)
| | | | | | FREEZE_THAW_YR > 56
| | | | | | | FREEZE_THAW_YR <= 115
| | | | | | | | AADT <= 7800: Poor (11.0/2.0)
| | | | | | | | AADT > 7800
| | | | | | | | | PCI0 <= 62.3: Very Poor (13.0/6.0)
| | | | | | | | | PCI0 > 62.3
| | | | | | | | | | AADT <= 19870: Very Poor (2.0/1.0)
| | | | | | | | | | AADT > 19870: Fair (4.0)
| | | | | | | FREEZE_THAW_YR > 115
| | | | | | | | PCI0 <= 59.7: Fair (3.0/1.0)
| | | | | | | | PCI0 > 59.7: Very Poor (8.0)
| | | | | PCI0 > 63.1: Poor (14.0/1.0)
| | | | FREEZE_THAW_YR > 130
| | | | | FREEZE_THAW_YR <= 154
| | | | | | FREEZE_THAW_YR <= 144: Fair (11.0)
| | | | | | FREEZE_THAW_YR > 144
| | | | | | | PCI0 <= 60.7: Fair (3.0)
| | | | | | | PCI0 > 60.7
| | | | | | | | FREEZE_THAW_YR <= 147: Poor (3.0)
| | | | | | | | FREEZE_THAW_YR > 147: Fair (3.0/1.0)
| | | | | FREEZE_THAW_YR > 154: Very Poor (3.0/1.0)
| | | PCI0 > 64.6
| | | | PCI0 <= 68.2
| | | | | FREEZE_THAW_YR <= 43: Fair (21.0)
| | | | | FREEZE_THAW_YR > 43
| | | | | | FREEZE_THAW_YR <= 57: Very Poor (5.0)
| | | | | | FREEZE_THAW_YR > 57: Fair (46.0/10.0)
| | | | PCI0 > 68.2
| | | | | PCI0 <= 70.2
| | | | | | AADT <= 1410: Poor (6.22)
| | | | | | AADT > 1410
| | | | | | | FREEZE_THAW_YR <= 102
| | | | | | | | PCI0 <= 69.4: Poor (7.0)
| | | | | | | | PCI0 > 69.4: Fair (2.78/0.78)
| | | | | | | FREEZE_THAW_YR > 102: Fair (12.0/2.0)
| | | | | PCI0 > 70.2: Fair (5.0/1.0)
| | PCI0 > 70.6
| | | PCI0 <= 84.4
| | | | FREEZE_THAW_YR <= 5
| | | | | PCI0 <= 80.7: Fair (24.0/1.0)
| | | | | PCI0 > 80.7: Satisfactory (3.0/1.0)
| | | | FREEZE_THAW_YR > 5
| | | | | AADT <= 9950
| | | | | | AADT <= 1804: Fair (9.28/2.07)
| | | | | | AADT > 1804
| | | | | | | FREEZE_THAW_YR <= 28: Satisfactory (19.0/2.0)
| | | | | | | FREEZE_THAW_YR > 28
| | | | | | | | PCI0 <= 78.2
| | | | | | | | | AADT <= 6800
| | | | | | | | | | YEARS_REMED <= 0.6: Very Poor (3.76/2.51)
| | | | | | | | | | YEARS_REMED > 0.6: Satisfactory (26.33/8.26)
| | | | | | | | | AADT > 6800: Fair (11.42/3.14)
| | | | | | | | PCI0 > 78.2: Fair (6.5)
| | | | | AADT > 9950
| | | | | | FREEZE_THAW_YR <= 77
| | | | | | | YEARS_REMED <= 10.2
| | | | | | | | FREEZE_THAW_YR <= 18: Satisfactory (5.0/1.0)
| | | | | | | | FREEZE_THAW_YR > 18
| | | | | | | | | YEARS_REMED <= 4.8
| | | | | | | | | | PCI0 <= 81.2: Fair (5.35)
| | | | | | | | | | PCI0 > 81.2: Satisfactory (2.27/0.27)
| | | | | | | | | YEARS_REMED > 4.8: Fair (15.23/0.54)
| | | | | | | YEARS_REMED > 10.2
| | | | | | | | YEARS_REMED <= 12.3: Poor (3.21/0.12)
| | | | | | | | YEARS_REMED > 12.3: Fair (3.21/0.09)
| | | | | | FREEZE_THAW_YR > 77
| | | | | | | AADT <= 89000
| | | | | | | | YEARS_REMED <= 5.9: Satisfactory (6.8/0.67)
| | | | | | | | YEARS_REMED > 5.9
| | | | | | | | | YEARS_REMED <= 14.2
| | | | | | | | | | YEARS_REMED <= 7.5: Fair (3.4/0.06)
| | | | | | | | | | YEARS_REMED > 7.5: Serious (5.66/1.66)
| | | | | | | | | YEARS_REMED > 14.2: Satisfactory (4.53/0.44)
| | | | | | | AADT > 89000: Poor (2.04/0.04)
| | | PCI0 > 84.4
| | | | PCI0 <= 84.9: Very Poor (13.0/1.0)
| | | | PCI0 > 84.9: Fair (3.0/1.0)
PCI0 > 85.1
| PCI0 <= 95.3
| | AADT <= 21500
| | | PCI0 <= 86.4: Satisfactory (15.85/3.0)
| | | PCI0 > 86.4
| | | | AADT <= 1940
| | | | | YEARS_REMED <= 0.5: Fair (2.62/0.62)
| | | | | YEARS_REMED > 0.5: Good (21.91/4.2)
| | | | AADT > 1940
| | | | | YEARS_REMED <= 3.9
| | | | | | AADT <= 4408
| | | | | | | AADT <= 3668: Fair (3.65/1.7)
| | | | | | | AADT > 3668: Poor (2.14/0.14)
| | | | | | AADT > 4408: Good (14.83/7.81)
| | | | | YEARS_REMED > 3.9
| | | | | | PCI0 <= 90.4: Satisfactory (12.6/2.0)
| | | | | | PCI0 > 90.4
| | | | | | | PCI0 <= 91.1: Fair (3.58/1.58)
| | | | | | | PCI0 > 91.1
| | | | | | | | FREEZE_THAW_YR <= 95: Fair (3.93/1.87)
| | | | | | | | FREEZE_THAW_YR > 95: Satisfactory (3.0)
| | AADT > 21500
| | | PCI0 <= 92.3: Satisfactory (8.75/0.3)
| | | PCI0 > 92.3
| | | | AADT <= 34500: Very Poor (4.1/0.1)
| | | | AADT > 34500: Satisfactory (2.05/0.05)
| PCI0 > 95.3
| | FREEZE_THAW_YR <= 88
| | | YEARS_REMED <= 1.3
```

| | | | YEARS_REMED <= 1.2
| | | | | FREEZE_THAW_YR <= 85
| | | | | | AADT <= 26000
| | | | | | | FREEZE_THAW_YR <= 73: Good (54.58/6.34)
| | | | | | | FREEZE_THAW_YR > 73
| | | | | | | | AADT <= 7500: Good (3.88/0.94)
| | | | | | | | AADT > 7500: Satisfactory (5.94/1.47)
| | | | | | AADT > 26000: Satisfactory (13.43/5.43)
| | | | | FREEZE_THAW_YR > 85
| | | | | | YEARS_REMED <= 0.3: Good (4.7/0.23)
| | | | | | YEARS_REMED > 0.3: Poor (4.7/0.7)
| | | | YEARS_REMED > 1.2
| | | | | AADT <= 13120: Good (4.24/0.94)
| | | | | AADT > 13120: Satisfactory (17.56/0.45)
| | | YEARS_REMED > 1.3
| | | | YEARS_REMED <= 6.8: Good (71.83/9.86)
| | | | YEARS_REMED > 6.8
| | | | | AADT <= 19228: Good (2.06/0.25)
| | | | | AADT > 19228: Satisfactory (3.07/0.04)
| | FREEZE_THAW_YR > 88
| | | FREEZE_THAW_YR <= 142
| | | | FREEZE_THAW_YR <= 103
| | | | | PCI0 <= 97.4: Satisfactory (5.0/1.0)
| | | | | PCI0 > 97.4
| | | | | | YEARS_REMED <= 0.2: Satisfactory (3.13/1.13)
| | | | | | YEARS_REMED > 0.2
| | | | | | | AADT <= 6017: Fair (12.87/1.0)
| | | | | | | AADT > 6017
| | | | | | | | AADT <= 36842: Good (6.0)
| | | | | | | | AADT > 36842: Fair (2.0)
| | | | FREEZE_THAW_YR > 103
| | | | | AADT <= 11500: Good (20.0/3.0)
| | | | | AADT > 11500
| | | | | | AADT <= 33224: Satisfactory (10.0/3.0)
| | | | | | AADT > 33224: Fair (2.0)
| | | FREEZE_THAW_YR > 142: Good (25.0)
Number of Leaves : 100
Size of the tree : 199
Decision tree I (Model #3)
YEARS_REMED > 26.200: Serious {Serious=1, Very Poor=1, Poor, Good, Failed, Fair, Satisfactory=0}
YEARS_REMED ≤ 26.200
| YEARS_REMED > 23.250
| | PCI0 > 58.100: Poor {Serious=0, Fair=0, Satisfactory=1, Very Poor=0, Poor=2, Good=0, Failed=0}
| | PCI0 ≤ 58.100: Failed {Serious=0, Fair=0, Satisfactory=0, Very Poor=0, Poor=0, Good=0, Failed=2}
| YEARS_REMED ≤ 23.250
| | YEARS_REMED > 22.900: Poor {Serious=0, Fair=0, Satisfactory=0, Very Poor=0, Poor=2, Good=0, Failed=0}
| | YEARS_REMED ≤ 22.900
| | | FREEZE_THAW_YR > 189
| | | | PCI0 > 59.650: Fair {Serious=0, Fair=2, Satisfactory=0, Very Poor=0, Poor=0, Good=0, Failed=0}
| | | | PCI0 ≤ 59.650: Poor {Serious=0, Fair=0, Satisfactory=0, Very Poor=0, Poor=3, Good=0, Failed=0}
| | | FREEZE_THAW_YR ≤ 189
| | | | YEARS_REMED > 22.100: Serious {Serious=1, Fair=0, Satisfactory=1, Very Poor=0, Poor=0, Good=0, Failed=0}
| | | | YEARS_REMED ≤ 22.100
| | | | | YEARS_REMED > 6.700
| | | | | |FREEZE_THAW_YR > 179.500: Satisfactory {Satisfactory=3, Very Poor, Poor, Good, Failed Serious, Fair=0}
| | | | | | FREEZE_THAW_YR ≤ 179.500
| | | | | | | YEARS_REMED > 21.200: Fair {Fair=1, Good=1, Satisfactory, Very Poor, Poor, Failed, Serious=0}
| | | | | | | YEARS_REMED ≤ 21.200
| | | | | | | | | FREEZE_THAW_YR > 0.500: Fair {Fair=88, Satisfactory=25, Very Poor=49, Poor=28, Serious=16, Good=1, Failed=1}
| | | | | | | | | FREEZE_THAW_YR ≤ 0.500: Poor {Serious=0, Fair=0, Satisfactory=0, Very Poor=0, Poor=3, Good=0, Failed=0}
| | | | | YEARS_REMED ≤ 6.700
| | | | | | PCI0 > 22.050
| | | | | | | PCI0 > 28.250
| | | | | | | | | PCI0 > 46.300: Good {Serious=7, Fair=112, Satisfactory=113, Very Poor=36, Poor=43, Good=164, Failed=0}
| | | | | | | | | PCI0 ≤ 46.300: Very Poor {Serious=3, Fair=0, Satisfactory=0, Very Poor=13, Poor=0, Good=0, Failed=0}
| | | | | | | | PCI0 ≤ 28.250: Serious {Serious=6, Fair=0, Satisfactory=0, Very Poor=0, Poor=0, Good=0, Failed=0}
| | | | | | | PCI0 ≤ 22.050: Failed {Serious=0, Fair=0, Satisfactory=0, Very Poor=0, Poor=0, Good=0, Failed=3}

Acknowledgments

Input and evaluation as well as support in suggesting work steps provided by Dr. James Smith, Manager, Member/Technical Services, Ontario Good Roads Association (OGRA) is highly appreciated and is recognized here.

References

Alsugair, A. M., and A. A. Al-Qudrah. 1998. "Artificial neural network approach for pavement maintenance." *J. Comput. Civ. Eng.* 12 (4): 249–255. https://doi.org/10.1061/(ASCE)0887-3801(1998)12:4(249).

Al-Zou'bi, M. M., C. M. Chang, S. Nazarian, and V. Kreinovich. 2015. "Systematic statistical approach to populate missing performance data in pavement management systems." *J. Infrastruct. Syst.* 21 (4): 04015002. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000247.

Anyala, M., J. B. Odoki, and C. J. Baker. 2014. "Hierarchical asphalt pavement deterioration model for climate impact studies." *Int. J. Pavement Eng.* 15 (3): 251–266. https://doi.org/10.1080/10298436.2012.687105.

Beckwith. 2013. *Asset management plan*. Washington, DC: US Dept. of Veterans Affairs.

Black, M., A. T. Brint, and J. R. Brailsford. 2005. "A semi-Markov approach for modelling asset deterioration." *J. Oper. Res. Soc.* 56 (11): 1241–1249. https://doi.org/10.1057/palgrave.jors.2601967.

Chi, S., M. Murphy, and Z. Zhang. 2014. "Sustainable road management in Texas: Network-level flexible pavement structural condition analysis using data-mining techniques." *J. Comput. Civ. Eng.* 28 (1): 156–165. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000252.

Chong, G. J., W. A. Phang, and G. Wrong. 1982. *SP-024, Manual for Condition Rating of Flexible Pavements*. Ministry of Transportation. 30

Elbagalati, O., M. A. Elseifi, K. Gaspard, and Z. Zhang. 2017. "Implementation of the structural condition index into the Louisiana pavement management system based on rolling wheel deflectometer testing." *Transp. Res. Rec.* 2641 (1): 39–47. https://doi.org/10.3141/2641-06.

El-Diraby, T. E., S. Kinawy, and S. M. Piryonesi. 2017. "A comprehensive review of approaches used by Ontario municipalities to develop road asset management plans." In *Proc., Transportation Research Board 96th Annual Meeting*, Washington, DC. 31

Elkins, G. E., P. Schmalzer, T. Thompson, and A. Simpson. 2003. *Long-term pavement performance information management system pavement performance database user reference guide*. McLean, VA: Federal Highway Administration.

Ens, A. 2012. *Development of a flexible framework for deterioration modelling in infrastructure asset management.* Toronto: Univ. of Toronto.

Farhan, J., and T. Fwa. 2014. "Augmented stochastic multiple imputation model for airport pavement missing data imputation." *Transp. Res. Rec.* 2449 (1): 96–104.

Ferregut, C., I. Abdallah, O. Melchor-Lucero, and S. Nazarian. 1999. *Artificial neural network-based methodologies for rational assessment of remaining life of existing pavements*. El Paso, TX: Center for Highway Materials Research.

Ford, K., M. Arman, S. Labi, K. C. Sinha, P. D. Thompson, A. M. Shirole, and Z. Li. 2012. *Estimating life expectancies of highway assets*. NCHRP Rep. No. 713. Washington, DC: Transportation Research Board.

Guillaumot, V. M., P. L. Durango-Cohen, and S. M. Madanat. 2003. "Adaptive optimization of infrastructure maintenance and inspection decisions under performance model uncertainty." *J. Infrastruct. Syst.* 9 (4): 133–139. https://doi.org/10.1061/(ASCE)1076-0342(2003)9:4(133).

Haider, S. W., K. Chatti, N. Buch, R. W. Lyles, A. S. Pulipaka, and D. Gilliland. 2007. "Effect of design and site factors on the long-term performance of flexible pavements." *J. Perform. Constr. Facil.* 21 (4): 283–292. https://doi.org/10.1061/(ASCE)0887-3828(2007)21:4(283).

Hall, M. 1999. *Correlation-based feature selection for machine learning.* Hamilton, New Zealand: Univ. of Waikato.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. "The elements of statistical learning." *Bayesian Forecasting Dyn. Models* 1: 1–694.

InfoPave. 2017. "LTPP Infopave–Home." Accessed July 27, 2017. https://infopave.fhwa.dot.gov/.

Kargah-Ostadi, N., S. Stoffels, and N. Tabatabaee. 2010. "Network-level pavement roughness prediction model for rehabilitation recommendations." *Transp. Res. Rec.* 2155 (1): 124–133.

Kargah-Ostadi, N., and S. M. Stoffels. 2015. "Framework for development and comprehensive comparison of empirical pavement performance models." *J. Transp. Eng.* 141 (8): 04015012. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000779.

Karlaftis, A. G., and A. Badr. 2015. "Predicting asphalt pavement crack initiation following rehabilitation treatments." *Transp. Res. Part C: Emerging Technologies* 55 (Jun): 510–517. https://doi.org/10.1016/j.trc.2015.03.031.

Kira, K., and L. Rendell. 1992. *The feature selection problem: Traditional methods and a new algorithm.*, 129–134. San Jose, CA: AAAI Press.

Kleiner, Y. 2001. "Scheduling inspection and renewal of large infrastructure assets." *J. Infrastruct. Syst.* 7 (4): 136–143. https://doi.org/10.1061/(ASCE)1076-0342(2001)7:4(136).

Li, N., W.-C. Xie, and R. Haas. 1996. "Reliability-based processing of Markov chains for modeling pavement network deterioration." *Transp. Res. Rec.* 1524 (1): 203–213. https://doi.org/10.3141/1524-24.

Lin, S.-W., and S.-C. Chen. 2012. "Parameter determination and feature selection for C4.5 algorithm using scatter search approach." *Soft Comput.* 16 (1): 63–75. https://doi.org/10.1007/s00500-011-0734-z.

Lou, Z., M. Gunaratne, J. J. Lu, and B. Dietrich. 2001. "Application of neural network model to forecast short-term pavement crack condition: Florida case study." *J. Infrastruct. Syst.* 7 (4): 166–171. https://doi.org/10.1061/(ASCE)1076-0342(2001)7:4(166).

Memarzadeh, M., and M. Pozzi. 2016. "Value of information in sequential decision making: Component inspection, permanent monitoring and system-level scheduling." *Reliab. Eng. Syst. Saf.* 154 (Oct): 137–151.

Miradi, M., and A. A. A. Molenaar. 2006. "Application of artificial neural network (ANN) to PA lifespan: Forecasting models." In *Proc., 2006 IEEE Int. Joint Conf. on Neural Network*, 3679–3685. New York: IEEE.

Najibi, N., N. Devineni, and M. Lu. 2017. "Hydroclimate drivers and atmospheric teleconnections of long duration floods: An application to large reservoirs in the Missouri River Basin." *Adv. Water Resour.* 100 (Feb): 153–167. https://doi.org/10.1016/j.advwatres.2016.12.004.

Neves, L. C., and D. M. Frangopol. 2005. "Condition, safety and cost profiles for deteriorating structures with emphasis on bridges." *Reliab. Eng. Syst. Saf.* 89 (2): 185–198. https://doi.org/10.1016/j.ress.2004.08.018.

Osorio-Lird, A., A. Chamorro, C. Videla, S. Tighe, and C. Torres-Machi. 2018. "Application of Markov chains and Monte Carlo simulations for developing pavement performance models for urban network management." *Struct. Infrastruct. Eng.* 14 (9): 1169–1181. https://doi.org/10.1080/15732479.2017.1402064.

Pantelias, A., G. W. Flintsch, J. W. Bryant, and Chen Chen. 2008. "Asset management data practices for supporting project selection decisions." *Public Works Manage. Policy* 13 (3): 239–252. https://doi.org/10.1177/1087724X08327574.

Piryonesi, S. M., and T. E. El-Diraby. 2018. *Using data analytics for cost-effective prediction of road conditions: Case of the pavement condition index*. Washington, DC

Piryonesi, S. M., and M. Tavakolan. 2017. "A mathematical programming model for solving cost-safety optimization (CSO) problems in the maintenance of structures." *KSCE J. Civ. Eng.* 21 (6): 2226. https://doi.org/10.1007/s12205-017-0531-z.

Pozzi, M., M. Memarzadeh, and K. Klima. 2017. "Hidden-model processes for adaptive management under uncertain climate change." *J. Infrastruct. Syst.* 23 (4): 04017022. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000376.

Provost, F., and T. Fawcett. 2013. *Data science for business*. Newton, MA: O'Reilly Media, Inc.

Ramia, A. P., and N. Ali. 1997. "Bayesian methodologies for evaluating rutting in Nova Scotia's Special B asphalt concrete overlays." *Can. J. Civ. Eng.* 24 (1): 1–11. https://doi.org/10.1139/l96-082.

Rapidminer. 2017. "Optimize selection–RapidMiner documentation." Accessed July 27, 2017. https://docs.rapidminer.com/studio/operators/modeling/optimization/feature_selection/optimize_selection.html.

Shahin, M. A., M. B. Jaksa, and H. R. Maier. 2009. "Recent advances and future challenges for artificial neural systems in geotechnical engineering applications." *Adv. Artif. Neural Syst.* 2009 (1): 1–9. https://doi.org/.

Shalaby, A., and A. Reggin. 2007. "Optimization of data collection needs for manual and automated network-level pavement condition ratings based on transverse variability and neural networks." *Can. J. Civ. Eng.* 34 (2): 139–146. https://doi.org/10.1139/l06-126.

Špačková, O., and D. Straub. 2017. "Long-term adaption decisions via fully and partially observable Markov decision processes." *Sustainable Resilient Infrastruct.* 2 (1): 37–58. https://doi.org/10.1080/23789689.2017.1278995.

Terzi, S. 2007. "Modeling the pavement serviceability ratio of flexible highway pavements by artificial neural networks." *Constr. Build. Mater.* 21 (3): 590–593. https://doi.org/10.1016/j.conbuildmat.2005.11.001.

Tighe, S. 2002. "Evaluation of subgrade and climatic zone influences on pavement performance in the Canadian Strategic Highway Program's (C-SHRP) long-term pavement performance (LTPP) study." *Can. Geotech. J. Can.* 39 (2): 377–387. https://doi.org/10.1139/t01-111.

Way, N. C., P. Beach, and P. Materials. 2015. *Standard practice for roads and parking lots pavement condition index surveys*. ASTM D6433-07. West Conshohocken, PA: ASTM.

Woldesenbet, A., H. D. Jeong, and H. Park. 2015. "Framework for integrating and assessing highway infrastructure data." *J. Manage. Eng.* 32 (1): 4015028. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000389.

Wu, K. 2015. *Development of PCI-based pavement performance model for management of road infrastructure system*. Tempe, AZ: Arizona State Univ.

Wu, X., et al. 2008. "Top 10 algorithms in data mining." *Knowl. Inf. Syst.* 14 (1): 1–37. https://doi.org/10.1007/s10115-007-0114-2.

Yang, J., J. Lu, and Q. Xiang. 2003. "Forecasting overall pavement condition with neural networks: Application on Florida highway network." *Transp. Res. Rec.* 1853 (1): 3–12. https://doi.org/10.3141/1853-01.

Ziari, H., J. Sobhani, J. Ayoubinejad, and T. Hartmann. 2016. "Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods." *Int. J. Pavement Eng.* 17 (9): 776–788. https://doi.org/10.1080/10298436.2015.1019498.

Queries

1. Please provide the ASCE Membership Grades for the authors who are members.

2. Please confirm that the “In addition to automating” sentence conveys your intended meaning.

3. Please check the hierarchy of section heading levels.

4. [ASCE Open Access: Authors may choose to publish their papers through ASCE Open Access, making the paper freely available to all readers via the ASCE Library website. ASCE Open Access papers will be published under the Creative Commons-Attribution Only (CC-BY) License. The fee for this service is $1750, and must be paid prior to publication. If you indicate Yes, you will receive a follow-up message with payment instructions. If you indicate No, your paper will be published in the typical subscribed-access section of the Journal.]

5. Please confirm that the “This complexity” sentence conveys your desired meaning.

6. Please confirm the “Second, testing” sentence conveys your intended meaning.

7. The citations (Figs. a1–a3) have been changed to (Figs. 18–20). Please check and confirm.

8. Please confirm “discussed in the following paragraphs” presents your intended meaning.

9. Please confirm that “and approximations using pairs of inspections” conveys your intended meaning.

10. Please check all figures, figure citations, and figure captions to ensure they match and are in the correct order.

11. Please check that the “For the wrongly predicted ones” sentence conveys your intended meaning.

12. Please confirm that FHWA should be defined as “Federal Highway Administration”.

13. Please confirm that the “and used the ensemble learning algorithms presented” phrase conveys your desired meaning.

14. Please check that the “Some of the most common data imputation” sentence conveys your desired meaning.

15. Please confirm that the “We then examined” sentence conveys your desired meaning.

16. Please use letters (e.g., a and b) to differentiate different sections of the figure instead of “left” and “right.”

17. Please provide significance of Bold values in Table 3.

18. Please confirm that the “The accuracy” sentence conveys your intended meaning.

19. Please confirm this is the correct figure.

20. Please consider providing a reference for the poor training conclusion.

21. Please consider further clarifying “with a texture”.

22. Please clarify the “It is obvious” sentence. Do you mean “The results show that the functional class of road and the thickness of the last placed overlay were the least informative attributes.”

23. Please confirm that “what is the data measuring” conveys your intended meaning.

24. Fig. 14 was not cited in the text of the paper. Please provide in-text citations for Fig. 14 after Fig. 13 and before Fig. 15 citation, as per style table citations have to be in sequential order in the text.

25. Please confirm “separating the dataset” conveys your intended meaning.

26. Please confirm “the difference between the maximum” sentence conveys your desired meaning.

27. Please confirm “corroborated the results.”

28. Please confirm the “In the case of PCI” sentence conveys your intended meaning.

29. Please clarify the meaning of the “So, should the” sentence or consider removing the sentence.

30. Please provide publisher location for Chong et al. (1982).

31. Please provide the publisher or sponsor name and location (not the conference location) for Ref. El-Diraby et al. (2017).

32. This query was generated by an automatic reference checking system. This reference could not be located in the databases used by the system. While the reference may be correct, we ask that you check it so we can provide as many links to the referenced articles as possible.

33. This query was generated by an automatic reference checking system. This reference could not be located in the databases used by the system. While the reference may be correct, we ask that you check it so we can provide as many links to the referenced articles as possible.

34. This query was generated by an automatic reference checking system. This reference could not be located in the databases used by the system. While the reference may be correct, we ask that you check it so we can provide as many links to the referenced articles as possible.

35. Please provide expansion for AAAI in Kira and Rendell (1992).

36. Please provide publisher name for Piryonesi and El-Diraby (2018).

37. Please check and confirm the edit made in the Ref. Way et al. (2015) is okay.