Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems

S. Madeh Piryonesi, Ph.D.[1]; and Tamer E. El-Diraby, Ph.D.[2]

Abstract: This study explores the performance regime of different classification algorithms as they are applied to the analysis of asphalt pavement deterioration data. The aim is to examine how different algorithms deal with the typically limited and low-quality data sets in the infrastructure asset management domain, and whether better configurations of relevant algorithms help overcome these limitations. Furthermore, the emphasis on choosing the most affordable attributes (e.g., temperature and precipitation levels) makes the results reproducible to smaller municipalities. This analysis used the data of more than 3,000 examples of road sections, which were retrieved from the Long-Term Pavement Performance (LTPP) database. The algorithms examined in this study include two types of decision trees, naïve Bayes classifier, naïve Bayes coupled with kernels, logistic regression, k-nearest neighbors (k-NN), random forest, and gradient boosted trees. The performance of these algorithms is compared, and their weaknesses and strengths are discussed. They were all applied to predict the deterioration of pavement condition index (PCI). Of specific importance is the positive role of ensemble learning. It is shown how using higher efficiencies by using ensemble learning can compensate for data shortcomings. The accuracy of some of the models in predicting the PCI after 3 years exceeded 90%. Suggestions are made to improve the performance of some algorithms. For instance, the naïve Bayes classifier was coupled with kernel estimates to achieve a better accuracy. It is demonstrated that using kernel estimates can increase the accuracy of the naïve Bayes classifier dramatically. Further, the study examines the impact of data segmentation. Data were divided into four different climatic regions. The accuracy of prediction was sufficiently high after segmentation, with the highest accuracy in the dry and nonfreeze zone and the lowest performance in the region with a wet and freezing climate. **DOI: [10.1061/JPEODX.0000175](https://doi.org/10.1061/JPEODX.0000175).** © *2020 American Society of Civil Engineers.*

Author keywords: Machine learning; Ensemble learning; Transportation asset management; Pavement condition index; Highway maintenance; Data preparation.

Introduction

Transportation asset management (TAM) planning helps municipalities keep their road networks in an acceptable level of service with a limited budget. One of the basic, yet very necessary, steps in preparing a TAM plan is assessment of current and prediction of future road conditions (El-Diraby et al. 2017). However, the accuracy of such assessments is hampered by the typically limited data size and, equally important, the lower quality of the data themselves [i.e., missing (El-Diraby et al. 2017; Piryonesi and El-Diraby 2020), inconsistent (Gharaibeh et al. 2010), and inaccurate data (Abdelaty et al. 2018)]. This study aims to examine the role of analytics algorithms not only in predicting conditions/deterioration in roads, but also doing so with limited data sets. So, an important question is if some algorithms can address this typical problem in asset management.

The selected target variable for deterioration modeling in this study is the pavement condition index (PCI). The PCI has a significant role in all TAM decisions. It is one of the main indicators for measuring asphalt road deterioration and assessing levels of service. In fact, 58 municipalities in Ontario, Canada, were surveyed about the type of performance indicators they use to represent the level of service of their roads. As shown in Fig. 1, the PCI was the most popular. Yet, there is limited research on the application of machine learning to PCI prediction. With an exception of two studies that targeted PCI (e.g., Piryonesi and El-Diraby 2018; Kırbaş and Karaşahin 2016), most researchers have focused on predicting the International Roughness Index (IRI) or the analysis of a single distress. This could be because preparing PCI values from the Long-Term Pavement Performance (LTPP) database requires some data preparation effort (Piryonesi and El-Diraby 2018; Shahnazari et al. 2012; Wu 2015). Table 1 reviews related work in this domain (43 studies). Fig. 2 also offers a breakdown of studies that addressed pavement performance modeling.

Although the PCI is important in itself, because this study's analysis is targeting the effects of data size and algorithms, PCI predictions serve as a case for such a line of analysis. Future studies can benchmark this work to conduct similar analyses to address data limitations for other performance measures of TAM.

The deterioration of roads is a complicated nonlinear process that depends on many factors that relate to design characteristic, climatic, and operational attributes. Such a complicated and context-sensitive phenomenon cannot be easily formulated using a single logical syllogism or a mathematical formula. It requires analyses that are sophisticated enough to consider different facets of the problem. However, currently, most DOTs and municipalities

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

© ASCE   04020022-1   J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022
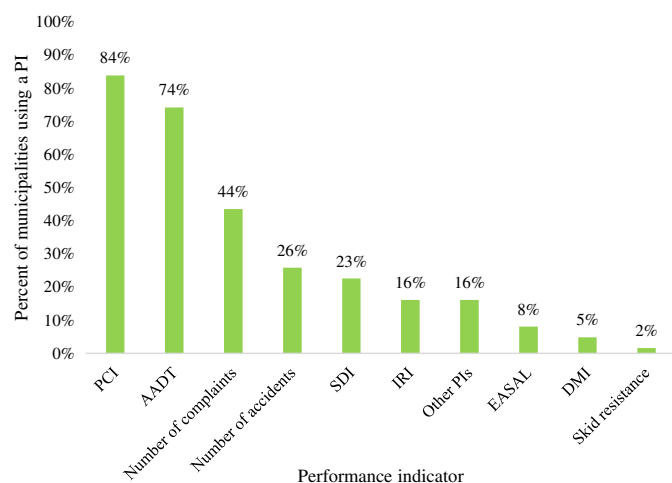
Fig. 1. Performance indicators used by municipalities in Ontario, Canada.

depend on deterioration curves for predicting road condition (El-Diraby et al. 2017; Wu 2015). Despite their simplicity and popularity, these curves have a few drawbacks. First, they are deterministic and cannot take into account uncertainties. Furthermore, they only rely on age and overlook important attributes such as the maintenance history and traffic of roads (Ens 2012; Piryonesi and El-Diraby 2018; Wu 2015). Other deterioration models, such as Markov models, can address some of these drawbacks (Black et al. 2005; Ens 2012; Li et al. 1996). For instance, Markov models are probabilistic. However, they have been criticized for disregarding the maintenance history of roads (Ens 2012; Piryonesi and Tavakolan 2017).

Data analytics is an alternative tool for predicting the deterioration of roads. However, it is not as popular as the aforementioned techniques. This lack of popularity rooted in two possible reasons: (1) the lack of reliable data for training models (Abdelaty et al. 2018; El-Diraby et al. 2017; Ens 2012); and (2) the lack of interpretability. The latter results from most previous studies tending to use neural networks, which are not easy to interpret (Ford et al. 2012). A handful of researchers have used more intelligible classifiers, such as decision trees, but they reported accuracy numbers that may not be high enough (Chi et al. 2014; Piryonesi and El-Diraby 2017, 2018).

In this paper, ensemble learning is deployed as a solution to the low accuracy of classifiers in predicting the future condition of roads. Particularly, boosting and bagging as two examples of ensemble learning are showcased, and their results are compared and interpreted. The main advantage of these approaches is that a more accurate model can be trained by merging weak learners such as decision trees. The training data were retrieved from the LTPP database.

Furthermore, in this paper, different types of classifiers are examined, and their weaknesses and strengths are explained. Some solutions for improving the accuracy of classifiers are discussed and implemented. The following algorithms were considered in this comparison.

Naïve Bayes Classifier

As the name suggests, naïve Bayes classifier works based on the Bayes rule. Prior and conditional probabilities are calculated from the data set. For continuous variables, a univariate Gaussian distribution is used to estimate their class-conditional marginal densities. A histogram is used for discrete attributes. The underlying

assumption of this classifier is that predictor attributes are independent; hence, it is called naïve. This simple classifier is popular when the number of features is large given its small computational complexity (Hastie et al. 2009; Provost and Fawcett 2013).

Naïve Bayes Classifier with Kernel

Kernel density estimation could help increase the accuracy of a naïve Bayes classifier. Kernel, in this context, is a localization technique and should not be confused with the kernel method. The latter is used for regularized nonlinear modeling coupled with algorithms such as support vector machine (Hastie et al. 2009). When coupled with naïve Bayes classifiers, kernel density estimates are used to estimate conditional marginal densities of each class. Different functions, such as Gaussian or Epanechnikov, could be used for local smoothing (Hastie et al. 2009). In this paper, Gaussian kernels are used. Naïve Bayes per se is a nonparametric algorithm. However, when using a kernel-based naïve Bayes classifier, the user should set two parameters: the bandwidth and the number of kernels.

Decision Trees

Decision trees are among the most popular classification algorithms. Part of their popularity stems from their intelligibility and ease of interpretation. As their name suggests, these algorithms generate a classifier tree based on the trends in the data set. Primitive versions of decision trees, such as iterative dichotomiser 3 (ID3) and concept learning system (CLS), were only able to learn 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 segmentation of the data set. 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 will 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 information gain and level of homogeneity of each leaf will be determined through the parameters of the tree (Lin and Chen 2012).

Ensemble Learning

Ensemble learning refers to a group of technique that include combining multiple base learners to diminish their weaknesses and superpose their strengths. The general process of ensemble learning is twofold, first generating a population of base learners from the training set and then combining them to create the strong predictor. Different ensemble techniques have different approaches for combining the base learners. Two important categories of ensemble learning include bagging and boosting. These, respectively, rely on committee voting and weighted voting (Hastie et al. 2009). In this paper, both approaches were used, and it was shown that although both methods increase the accuracy of a classifier, boosting is more effective given its more sophisticated learning process.

Literature Review

Condition assessment and deterioration modeling are not only the basis for maintenance and rehabilitation, but also help decision makers proactively develop capital planning and schedule future maintenance interventions. Some of the most well-known measures

© ASCE

Table 1. Summary of studies on pavement performance modeling

Study No.	Size of training data	Performance indicator	Data source	Type of modeling	References
1	—	IRI	LTPP	DC, nonlinear regression	Tighe (2002)
2	148	IRI	LTPP	Nonlinear regression	Smith and Tighe (2004)
3	117	IRI	LTPP	Neural networks	Choi et al. (2004)
4	212	IRI, single distress	LTPP	ANOVA and logistic regression	Haider et al. (2007)
5	162	IRI	LTPP and Utah DOT	M-E, linear regression	Darter et al. (2009)
6	178	IRI	LTPP	M-E, linear regression	Souliman et al. (2010)
7	858	IRI	LTPP	Neural network	Kargah-Ostadi et al. (2010)
8	135	IRI	LTPP and New Mexico DOT	Nonlinear regression, survival curves	Chen and Zhang (2011)
9	1,206	Single distress	LTPP	Logistic regression	Wang (2013)
10	—	Single distress	LTPP	Weibull distribution	Dong and Huang (2014)
11	163	IRI	LTPP	Weibull distribution	Meegoda and Gao (2014)
12	148	IRI	LTPP	Bayesian regression with neural networks	Liu and Gharaibeh (2014)
13	143	IRI	LTPP	M-E, linear regression	Guo and Timm (2015)
14	1,143	Single distress	LTPP	Neural network	Karlaftis and Badr (2015)
15	96	PCI	LTPP	DC, nonlinear regression	Wu (2015)
16	205	IRI	LTPP	Neural network and GMDH	Ziari et al. (2016)
17	<100	IRI	LTPP	Neural networks	Hossain et al. (2017)
18	342	IRI	LTPP	Neural networks	Sollazzo et al. (2017)
19	460	IRI, single distress	LTPP	Random forest	Gong et al. (2019)
20	382	IRI	Local data (Brazil)	E, linear, and nonlinear regression	Paterson (1989)
21	—	PCS	—	Markov chain	Li et al. (1996)
22	<100	Single distress	Nova Scotia, Canada, DOT	Bayesian models	Ramia and Ali (1997)
23	—	Critical strain	Texas DOT	Neural network	Ferregut et al. (1999)
24	—	Single distress	AASHO Road Test and experimental data	M-E, linear regression	Archilla and Madanat (2000)
25	114	IRI	Indiana DOT	E, linear regression	Gulen et al. (2001)
26	—	CI	Florida DOT	Neural network	Lou et al. (2001)
27	400	IRI	Local data (UAE)	DC, linear, and nonlinear regression	Al-Suleiman and Shiyab (2003)
28	125	IRI	Local data (Taiwan)	Neural network	Lin et al. (2003)
29	—	PCR	Florida DOT	Neural networks	Yang et al. (2003)
30	—	Single distress	Own experiment data	Linear regression	Fwa et al. (2004)
31	102	Single distress	Strategic Highway Research Program Netherlands (SHRP-NL)	Neural network	Miradi and Molenaar (2006)
32	133	PSI	AASHO	Neural network	Terzi (2007)
33	—	PCR	Ohio DOT	Markov chain	Pulugurta et al. (2009)
34	—	PCI	Local data (Iran)	Neural networks and genetic programming	Shahnazari et al. (2012)
35	580	IRI	Local data (India)	Linear and nonlinear regression, neural networks	Chandra et al. (2013)
36	94	IRI	Local data (India)	Linear regression	Prasad et al. (2013)
37	160	Single distress	UK Highways Agency (Area 6)	M-E, nonlinear regression	Anyala et al. (2014)
38	354	SCI	Texas DOT	Decision tree	Chi et al. (2014)
40	—	RCI and SDI	Local data (Canada)	M-E, linear regression	Ayed (2016)
41	88	PCI	Local data (Turkey)	Neural network and nonlinear regression	Kırbaş and Karaşahin (2016)
42	720	PCI	Local data (Iran)	DC, nonlinear regression	Sadeghi et al. (2017)
43	150	UPCI	Local data (Chile)	Markov chain	Osorio-Lird et al. (2018)

Note: DC = deterministic curve; E = empirical; GMDH = group method of data handling; M-E = mechanical-empirical; and UPCI = urban pavement condition index.

are distress manifestation index (DMI), international roughness index (IRI), PCI, surface distress index (SDI), and present service-ability index (PSI) (Ford et al. 2012).

Collecting data about these performance indicators needs both human and financial resources. Small municipalities are usually short of both (El-Diraby et al. 2017). In larger jurisdictions, on the other hand, the annual cost of data collection could be high, given the large size of their networks. Furthermore, sometimes data collection could cause limitations to passing traffic (Chi et al. 2014; McCabe et al. 2017). In such situations, a predictive model can be highly useful in estimating performance indicators, hence avoiding some of the laborious and expensive data collection work. It can help smaller municipalities by saving the costs of interim data collection. Also, larger municipalities could benefit from such model by avoiding data collection for their entire road network.

Pavement performance or deterioration modeling for asphalt roads has been of interest for more than 3 decades. Table 1 summarizes some of the main studies conducted on this topic. The table
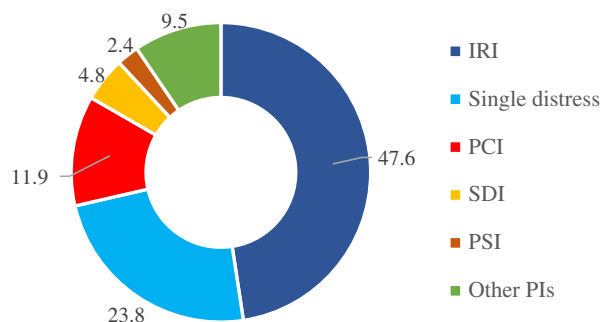
Fig. 2. Percent of studies conducted on pavement performance modeling categorized based on their main PI.

considers the data sources, the size of the training set, the target variable, and the type of models. Studies that used the same data source are listed in a chronological order. It is clear that most researchers developed their models for the IRI or a single distress. This pattern is specifically more common in the studies conducted on the LTPP data. Second, most studies that are based on machine learning deployed artificial neural networks (ANNs). Finally, the size of most training sets prepared and used by these researchers are smaller than 1,000 examples.

In practice, standardized deterioration curves are widely used by different municipalities and DOTs (El-Diraby et al. 2017; Ford et al. 2012). Most of these deterioration curves model the changes in a performance indicator (e.g., PCI) over time using a sigmoidal function (Witczak et al. 2002). For example, Wu (2015) used LTPP data to develop PCI master curves. Because of the high computational demands of the adopted method, no more than 96 examples sections could be used to develop a curve. Wu (2015) also reported that sometimes PCI curves could be misleading given that they do not take into account attributes such as traffic and maintenance history. In another example, Sadeghi et al. (2017) extracted multiple deterioration curves for roads under low, average, and heavy traffic within 4 years. Such an approach could be a solution to incorporating one or more of the additional attributes to problem of modeling deterioration curves.

Researchers have tried to address the weaknesses of master curves via mechanistic-empirical (M-E) methods. These methods usually rely on using statistical methods such as correlation analyses for finding relations in historical data (Walubita et al. 2017). They are widely used to model the performance of roads. For example, Archilla and Madanat (2000) developed a mechanistic-empirical model to predict the incremental changes in rut depth. They used the data of equivalent single-axle load (ESAL), structural number (SN), and thawing index, which was collected from the AASHO Road Test. Similar studies on rutting progression are available. Fwa et al. (2004) used similar regression analyses to estimate the effects of traffic load, loading speed, and temperature on the depth of ruts in the asphalt pavement. The literature also includes similar studies on other performance indicators, such as IRI (Ens 2012; Wu 2015).

Over the last 2 decades, such M-E models gained momentum, and several DOTs have used this type of analysis for predicting the condition of their road network (AASHTO 2008; Darter et al. 2009; Souliman et al. 2010). Ayed (2016) reported that M-E models result in different outcomes when compared with traditional network-level models such as master deterioration curves. This difference is understandable not only due to the different nature of these models but also the incorporation of other attributes into the M-E models (Ayed 2016; Ens 2012; Sakhaeifar et al. 2015). Despite these

strengths, it is not easy to assess the reliability of M-E models when they are used out of the original data range for which they were developed (Ens 2012; Wu 2015). Furthermore, calibrating such models requires engineering and mechanistic data, such as deflection data collected by the falling weight deflectometer (FWD), which is not easily available to small municipalities (Ayed 2016; Chi et al. 2014; El-Diraby et al. 2017; Li and Wang 2018; Walubita et al. 2017).

Markov models are another category of probabilistic methods frequently used in the literature of deterioration modeling. These models are based on a memoryless stochastic process called a Markov chain. Unlike regression-based techniques, Markov models define the deterioration of roads as a series of discrete events. Therefore, their output will be a class of performance indicators in lieu of a real number. They are being used recently to develop more sophisticated deterioration models for roads and other assets (Ens 2012; Ford et al. 2012; Pulugurta et al. 2009). They, however, cannot include/consider other factors such as traffic or climatic attributes.

More importantly, Markov chains are time-independent: future states depend only on the current state with disregard to the pattern of previous variation. Therefore, they neglect the history of deterioration and previous maintenance interventions due to their memoryless nature (Anyala et al. 2014; Ens 2012; Piryonesi and Tavakolan 2017). To address this problem (called time homogeneity), semi-Markov models have been developed. They use independent random variables to model the interstate transitions (Black et al. 2005). However, semi-Markov models require more longitudinal data than time-homogeneous models. Examples of using homogenous and nonhomogeneous Markov models for deterioration modeling are, respectively, available from Pulugurta et al. (2009) and Li et al. (1996).

The most recent method for modeling road performance is using data analytics or machine learning techniques. Although some data analytics algorithms are similar to statistical analyses, they are different. Briefly, unlike statistical tools, data analytics algorithms can learn from large and heterogeneous data sets. Furthermore, they can handle high dimensional data, which could be in different formats (Hastie et al. 2009). Finally, unlike statistical tools, the result of a machine learning analysis is not necessarily an equation or a formal model—simply, they just capture patterns in data.

According to the National Cooperative Research Program (NCHRP) (Ford et al. 2012), data analytics or machine learning techniques have been used to a lesser extent, possibly because the results of such analyses are not easy to interpret (Ford et al. 2012). Interestingly, the report limits machine learning to neural networks (Ford et al. 2012). Neural networks have been criticized for their lack of interpretability despite their popularity and high learning capability (Shahin et al. 2009). However, ANNs are the most frequently used machine learning algorithms in the domain of pavement management. Shahnazari et al. (2012) used ANNs and genetic programming to determine the PCI of asphalt roads from distresses without referring to deduct value curves. The attributes they used were mainly the same pavement distresses recommended by ASTM D6433-07 (ASTM 2015): alligator cracking, block cracking, edge cracking, longitude cracking, pothole, bleeding, patching, and transverse cracking. A limitation of this study is that it relies on the same parameters that are used for PCI calculation. Therefore, such a model only saves engineers the time needed to extract the data from ASTM curves. Alsugair and Al-Qudrah (1998) trained an ANN to predict the most suitable type of remedial action for asphalt pavement. They used different road distresses, such as edge cracking, bumping, swelling, and slippage cracking, as inputs of their model. Their model was meant to select the most relevant maintenance intervention out of 13 different options.

© ASCE   04020022-4   J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

Furthermore, ANNs have been specifically used to study the change in performance indicators or the remaining life of roads. For instance, Ferregut et al. (1999) used ANNs in predicting the remaining life of asphalt pavement. They used an array of nine attributes as the input of their model: the thickness of the asphalt concrete; the thickness of base layers; and seven readings of a FWD. The model used these attributes to predict two critical strains at the interfaces of the layers or the remaining life of the pavement when it experiences either fatigue cracking or rutting. Lou et al. (2001) used predictive attributes such as the age of the road and three consecutive values of CI to predict the CI in the near future. Their training set was prepared based on the data set of Florida DOT (Lou et al. 2001). Kırbaş and Karaşahin (2016) developed three models using deterministic regression analysis, multivariate adaptive regression splines (MARS), and ANNs to predict the deterioration in the PCI based on the age of asphalt pavement. They reported R^2 of 0.7, 0.71, and 0.74 for each model, respectively (i.e., the ANN had the best performance).

Other probabilistic prediction studies on pavement were conducted—mainly based on Bayesian models. For instance, Ramia and Ali (1997) used Bayesian models to predict rutting in asphalt pavements. Anyala et al. (2014) investigated the impact of different climatic scenarios on asphalt pavement rut progression. They combined M-E models with probabilistic distributions and simulation to create more realistic deterioration models. Bayesian regressions was used to estimate the distribution of coefficients of M-E models.

ANNs have been criticized for being hard to interpret. Researchers have argued that the black-box nature of these algorithms has created issues such as difficulty in knowledge extraction (Ens 2012; Ford et al. 2012; Hameed et al. 2016; Shahin et al. 2009). As a consequence, other algorithms have been explored. Chi et al. (2014), for example, trained different decision trees for predicting structural condition index (SCI) based on data from Texas DOT. Their models could be used in SCI calculation when the data of the FWD are missing. The accuracy of all of their models in predicting five classes of SCI was less than 62% ["below the authors' expectation" according to Chi et al. (2014)]. The small size of their training set, which included 354 examples, may have adversely affected the accuracy of their models. Furthermore, they only relied on decision trees, which are not as powerful as random forest or gradient boosted trees (GBT) algorithms. Among the attributes used by Chi et al. (2014) were the amount of distress and ride score over 5 years (i.e., the model needed data from 5 consecutive years).

In another study, Wang (2013) used ordinal logistic regression to predict the percent of alligator cracking in asphalt overlays using the LTPP data. Although the number of training examples used by Wang (2013) was larger than that used by Chi et al. (2014), relying on logistic regression, which is a linear classifier, hindered the accuracy of the model. The accuracy in detecting four labels was 70%. However, the probability of a correct prediction by guessing is 25% when there are only four classes. The accuracy could have been improved by using nonlinear classifiers (as showcased by this study).

Kargah-Ostadi and Stoffels (2015) used LTPP data to learn ANNs, radial basis function networks, and support vector machines (SVM) to predict the roughness progression of asphalt pavement. Some of their inputs included previous IRI values, ESAL, the plasticity index of base and the percent of clay in subgrade soil. It is a common trend that studies based on LTPP data tend to use the IRI or roughness as their performance indicator instead of the PCI (Chen and Zhang 2011; Haider et al. 2007; Hossain et al. 2017, 2019; Kargah-Ostadi et al. 2010; Kargah-Ostadi and Stoffels 2015; Liu and Gharaibeh 2014; Ziari et al. 2016). Most probably, this is because the PCI data are not given in this database.

The value of ensemble learning seems to be acknowledged by researchers in the domain of road management. For example, two of studies in this domain (Nitsche et al. 2014; Tsai and Wang 2015) used random forests, which is an example of ensemble learning. Nitsche et al. (2014) used random forest regression and support vector regression (SVR) to estimate the weighted longitudinal profile (wLP) of roads. The predictive attributes they used included the angular speed as well as the longitudinal, lateral, and vertical acceleration of the four wheels. Also, Tsai and Wang (2015) used random forests for classifying types of raveling based on the three-dimensional (3D) laser-scanning data.
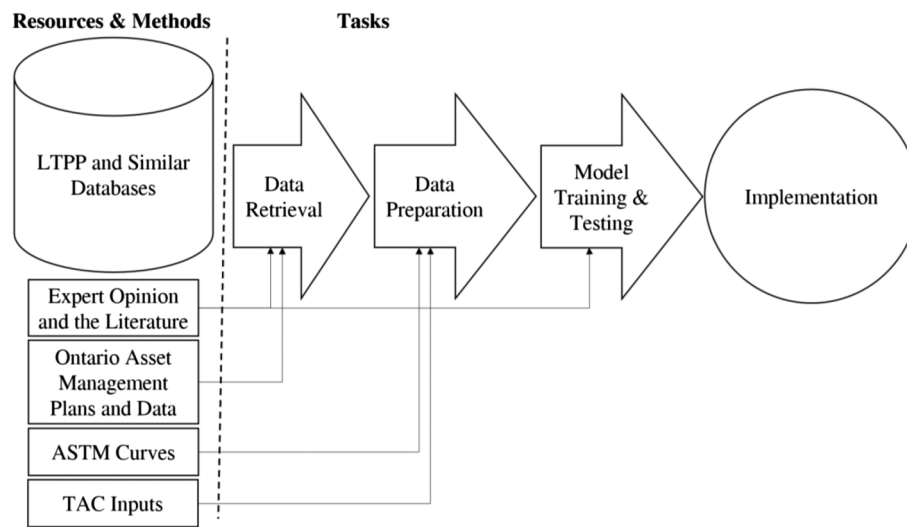
Scope and Contribution

Most researchers who deployed machine learning in deterioration modeling have relied on neural networks or a type of regression. Notwithstanding the availability of different classification algorithms, most researchers relied on regression algorithms. This could be partially a result of the tendency to see the performance indicators as real numbers rather than discrete variables. However, in most practical cases roads within a specific interval of a performance indicator (e.g., PCI or IRI) receive the same type of maintenance and rehabilitation. Furthermore, three studies that used classifiers such as decision trees have expressed concern about the accuracy levels (Chi et al. 2014; Piryonesi and El-Diraby 2018; Wang 2013). In this paper, different classifiers are trained and tested for predicting the PCI deterioration in the short term. The paper will address some of these mentioned methodological and theoretical barriers; particularly, the capability of ensemble learning in improving the accuracy of simple classifiers will be investigated.

To the best knowledge of the authors, the only study that has used an ensemble learning method for performance modeling is recent research conducted by Gong et al. (2018). They used a random forest regression (RFR) in predicting the IRI of roads based on the LTPP data. Their research suggested that their RFR outperformed the regularized linear regression significantly. Their study, however, did not include other ensemble learning methods such as boosting. In this study, it will be also shown how by using boosting, a classifier could outperform even powerful algorithms such as random forest. Furthermore, like the mainstream literature, Gong et al. (2018) conducted their study based on the IRI. The current study will investigate PCI as an understudied pavement performance indicator.

Furthermore, a key differentiator for the scope of this study is the selected attributes. The predictive attributes of previous studies are mostly causal and, more importantly, expensive to acquire. They use machine learning or statistics to recalculate the equation of the performance indicator. This study is noncausal. It uses circumstantial data that are free or easy to collect. Finally, most studies rely on longitudinal data as input of models. i.e., sequences of data of multiple years are required as the input of models. This may be difficult given that many municipalities do not have reliable data of consecutive years (Abdelaty et al. 2018; El-Diraby et al. 2017). In this study, thanks to the power of machine learning in discovering patterns in data, there is no need for longitudinal data.

Methodology

This research aimed to examine the application of machine learning algorithms—other than neural networks—to pavement performance prediction. A key focus is to compare the performance of different algorithms when tackling different sizes of data and different intervals of prediction (i.e., short-term and long-term behavior).

© ASCE   04020022-5   J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

Fig. 3. Schematic representation of the main tasks and the methods and resources adopted to perform them.

Furthermore, the attributes used in this study are either free or affordable. So, it is hoped that this will make it possible for municipalities to move to more data-driven analyses. Fig. 3 summarizes the methodology, the main tasks, and resources of this research. Utilized resources are shown on the left and the main performed tasks are on the right. As shown in Fig. 3, first, the needs of municipalities and DOTs were assessed through literature review and soliciting expert opinions via surveys and interviews. Then, the training data were retrieved from LTPP. The selection of features (attributes of the pavement/roadway) was performed based on expert opinion and literature review, with emphasis on the availability and affordability of data. To this end, 40 asset management plans developed by Ontario, Canada, municipalities were studied. The databases of 10 municipalities who volunteered their data were examined to learn about their available data fields. After identifying a list of affordable attributes, the data were retrieved from the LTPP online database using SQL-based queries (Elkins et al. 2003; InfoPave 2018).

In the next step, the retrieved data were prepared and cleaned. Two attributes that required considerable data preparation effort were the PCI and granular base equivalence (GBE). The first was prepared based on ASTM D6433-07 (ASTM 2015), and the latter using the guideline of Transportation Association of Canada (TAC) (Haas and Kazmierowski 1996). PCI values are not listed in LTPP. Several tables and fields from LTPP had to be combined.

After the data preparation task was finished, different classification models were trained based on the generated training set. These models include discriminative algorithms such as decision trees, generative algorithms such as naïve Bayes classifier, and ensemble learning algorithms such as random forests and gradient boosted trees. The inputs of experts were considered in the process of model training. Examples of these inputs include the following: (1) the most available and affordable predictive features; (2) the type of performance indicator as target variable; (3) the number of classes for PCI; and (4) acceptable accuracy levels. Trained models were tested using unseen data, and their effectiveness in prediction was evaluated. Training and cross-validation were performed based on a part of the LTPP data called general pavement studies (GPS). Another partition of the data set, named specific pavement studies (SPS), was used for testing. Eventually, the results of different

models were implemented in a computer program for easier use of practitioners.

Data Preparation

The LTPP database includes data collected by the Federal Highway Administration (FHWA) since the 1980s. The database is not limited to road performance measures but includes climatic data and operational conditions of roads as well. The LTPP database includes two major types of road sections: GPS and SPS (Elkins et al. 2003). The SPS, in brief, is concerned with the study of specific variables involving new construction, maintenance actions, and rehabilitation, but the GPS is more focused on the impact of environmental and/or exterior conditions on the in-service pavement. Furthermore, unlike GPS sections, SPS road sections are collocated.

Data were retrieved from the LTPP's online platform (InfoPave 2018) using SQL-based queries. The LTPP database only entails the data for distresses and does not include PCI values. Thus, a major part of data preparation was devoted to calculating PCI values from available distresses data. Another main task in data preparation was collating different attributes and transforming data because the LTPP data are stored in different tables and in a format that is not necessarily the most suitable for analytic purposes. Data cleansing entailed removing erroneous records, for example, if PCI (calculated) values were to increase drastically without any recorded maintenance—a type of inconsistency frequently found in road performance databases (Abdelaty et al. 2018).

Because the training set consisted of more than 3,000 examples, calculating the PCI manually was not feasible. So, a computer program was developed to calculate the PCI from distresses data provided in LTPP. To this end, both deduct value and corrected deduct value curves were digitized, and their mathematical formulas were found. These equations were imbedded into a spreadsheet and a computer program to calculate PCI values automatically. The process of finding the formulas and automating the PCI calculation is explained in the Appendix. For each of the 3,277 samples, PCI values were calculated for 6 consecutive years. The algorithms were trained to get the first as input to predict the others.

© ASCE 04020022-6 J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

Description of Predictive Attributes

There are challenges in collecting relevant PCI data and in assuring the quality of such data. So, unlike the majority of previous studies, the attributes used in this study were not selected based on mere mechanistic reasoning, but rather based on the required effort for their acquisition. In other words, the algorithms were trained to use only a set of attributes that are free to collect and/or easy to find reliably. The aim of the analysis is to find patterns in LTPP data in a noncausal manner. In other words, can the patterns of variation in these (easy to find) attributes and those of PCI be similar.

The selected attributes were chosen based on literature review, expert opinion, and studying the asset management plans of 40 and databases of 10 municipalities in Ontario. Table 1 includes the name of these attributes and their brief description. For each of the 3,000 samples, PCI values were calculated for 6 consecutive years. The algorithms were trained to get the first as input and then study patterns in these attributes to predict the other values of PCI (target feature). As can be seen, the attributes selected are easy to collect reliably. More interesting, they are not necessarily causal. For example, the age of road is calculated from its construction even if the road was rehabilitated. Again, the aim was to see if the patterns in attribute values and PCI deterioration match without any claim on causation.

Of course, these attributes are not equally informative. Further, the same subset of attributes will not be as informative to all algorithms. In other words, what is informative to one algorithm may not be as such for another. A detailed screening of the informativeness was conducted by Piryonesi and El-Diraby (2018, 2020). However, because the emphasis of this study is on exploring the sensitivities of different algorithms, all attributes of Table 2 are used in testing all algorithms, i.e., the same (extended) set of attributes were used to train all algorithms to establish a common ground for the comparative analysis.

**Table 2.** Predictive attributes and their description

Field name	Description
PCI0	(Initial) value of PCI at time of analysis
AGE	Age of road (since construction)
PAVEMENT_TYPE	Type of pavement (as defined by FHWA in LTPP)
FREEZE_INDEX_YR	Calculated freeze index for year (in Celsius days)
MAX_ANN_TEMP_AVG	Average of daily maximum air temperatures for year
MIN_ANN_TEMP_AVG	Average of daily minimum air temperatures for year
TOTAL_ANN_PRECIP	Total precipitation for year (mm)
FUNC_CLASS	Functional class of road (as defined by FHWA in LTPP)
FREEZE_THAW_YR	Number of freeze-thaw cycles per year
OVERLAY_THICKNESS	Thickness of the placed layer in rehabilitation
AADT_ALL_VEHIC_2WAY	Average annual daily traffic
REMED_TYPE	Type of last remedial action (as defined by FHWA in LTPP)
REMED_ YEARS	Number of years since the last remedial action
CONSTRUCTION_NO	Number of conducted remedial actions
GBE	Granular base equivalence
PCI (target variable)	Class of PCI after 3 years (as categorized by ASTM)

## Model Learning and Evaluation

Different models were learned from the training set and tested using 10-fold cross-validation. Fig. 4 shows the schematic process of how the models are designed. Using the attributes of Table 2 as input and the initial PCI, the trained models are to predict the value of PCI after 2, 3, 5, and 6 years. The reason behind limiting the horizon of prediction to 6 years is that most municipalities tend to do a comprehensive condition assessment every 5 years. Therefore, the prediction models will be mainly needed in-between. As shown in Fig. 4, the values of PCI are discretized according to ASTM's suggestions (ASTM 2015), except for that PCI values lower than 40 are labeled as Very Poor. Roads in this class must go through a major rehabilitation. Therefore, five classes provide sufficient granularity.

The models were first tested using 10-fold cross-validation. Fig. 5 shows the mean of cross-validation accuracy for different algorithms trained by different numbers of examples. As expected, increasing the number of training examples boosts the mean of accuracy. Also, both decision trees (i.e., RapidMiner decision tree and C4.5) outperformed the naïve Bayes classifier and the logistic regression. This is understandable considering the nonlinear nature of the classification problem and having multiple classes. Unlike decision trees, logistic regression is a linear classifier.

Moreover, naïve Bayes classifier, which is considered a generative algorithm, had a lower accuracy than both decision trees. Generally, discriminative algorithms usually have a better asymptotic
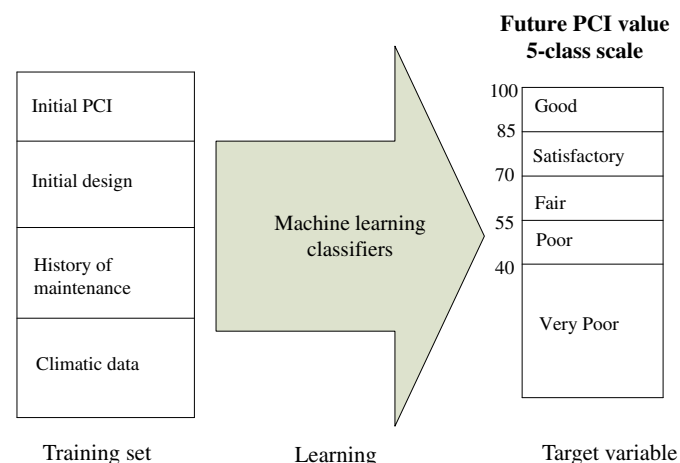
Fig. 4. Schematic representation of models for predicting future PCI using a number of easy-to-collect attributes.
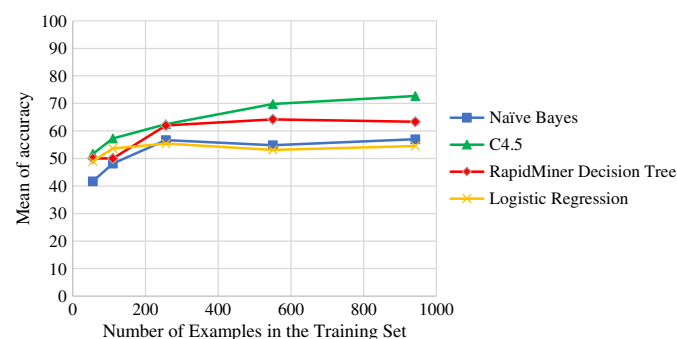
Fig. 5. Performance of different classification algorithms versus different sizes of the training set.

performance (Kaveh et al. 2013; Wu et al. 2008). However, in this example, the logistic regression did not outperform its generative counterpart (i.e., naïve Bayes). The reason is that logistic regression is originally a binary classifier and may not have the same performance when tackling a multiclass problem. Furthermore, the data set contained a considerable number of missing values, which affected the performance of this classifier.

Naïve Bayes Classifier with Kernel

Naïve Bayes classifiers use Gaussian distributions to estimate the marginal probability distributions of classes. However, sometimes the distributions of the actual data are far from normal. Therefore, better density estimations can help in finding better values for $P(x|Y = y_i)$, and hence more accurate predictions. For this, kernel density estimates were used to enhance the accuracy of naïve Bayes classifier. Fig. 6 shows the change in cross-validation accuracy of a naïve Bayes coupled with kernel estimation. As shown in Fig. 6, the accuracy increases as the number of kernel estimators increases.
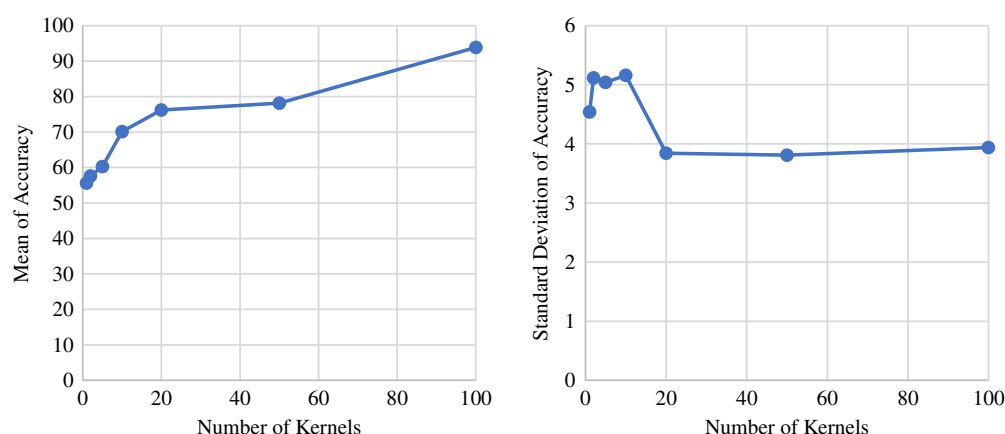
Comparing Figs. 7 and 8 can cast some light on the reason behind the enhancement of the accuracy. Fig. 7 shows the conditional marginal distribution of the number of freeze-thaw cycles per year for each class calculated by naïve Bayes classifier. According to Fig. 7, for a road with a target variable (future condition) of Good or Satisfactory (i.e., y = Good or y = Satisfactory), the probability of experiencing −25 freeze-thaw cycles is nonzero. Obviously, this is not valid (the value of this attribute must be larger than zero). However, in Fig. 8, which is based on using three kernel estimators, this probability is near zero. This makes the latter estimation more reasonable. In other words, by increasing the number of kernel functions, the probability densities became more realistic. Moving from a generic (smoothed) distribution of Fig. 7 to more realistic (and variable) distributions in Fig. 8 resulted in a higher accuracy.

The distributions presented in Fig. 8 are only estimations of class-conditional marginal densities of a specific attribute (i.e., annual freeze-thaw cycles) for a specific database (i.e., the LTPP). Therefore, they should not be generalized to other data sets. Furthermore,

Fig. 6. Using kernel density estimation increases the accuracy of naïve Bayes classifier.
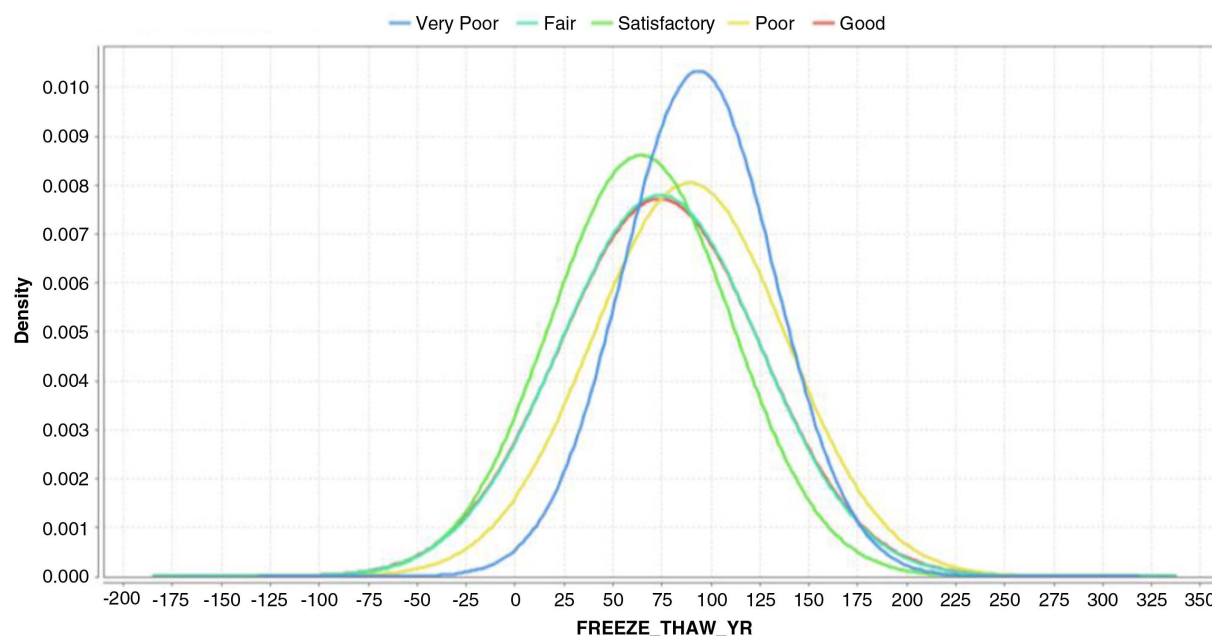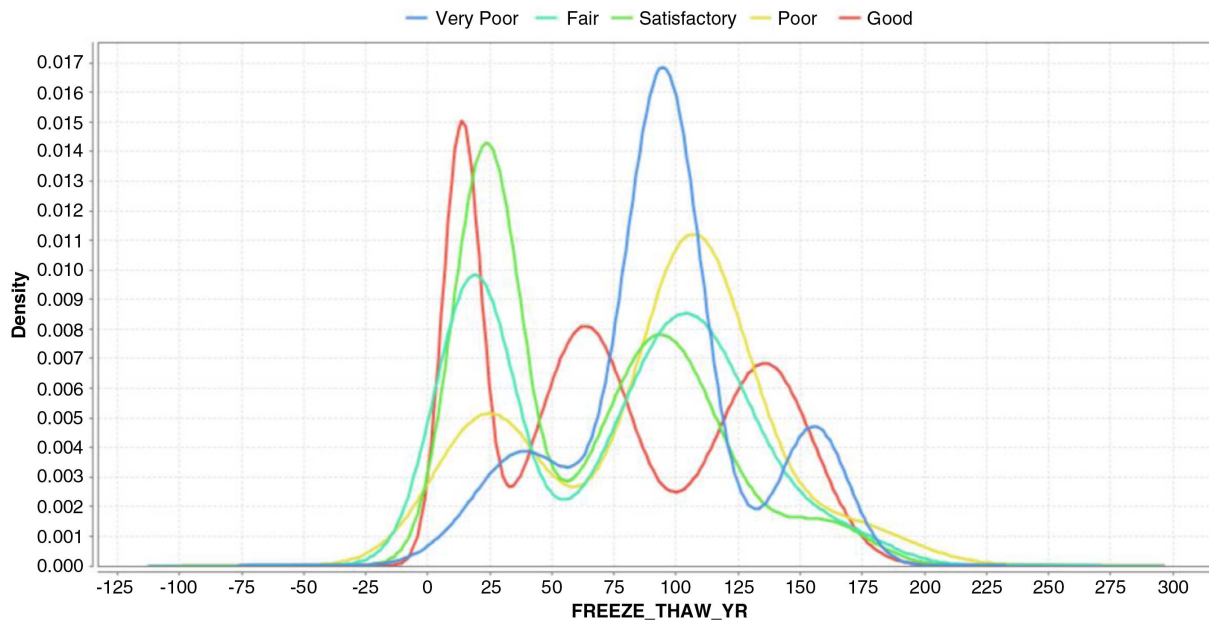

Fig. 7. Class-conditional marginal densities of number of freeze-thaw cycles per year in naïve Bayes classifier. The probability of negative freeze-thaws is estimated not to be zero.

© ASCE 04020022-8 J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

Fig. 8. Using three kernel estimation functions made the estimation of marginal densities more accurate.

because the LTPP data are collected over a large geographic spread with four different climatic regions, the multimodal distribution of this climatic attribute is no surprise. In fact, it represents the distribution of that specific attribute more realistically.

k-Nearest Neighbors Algorithm

The only parameter of this algorithm is the number of neighbors, which is an integer denoted by k. Fig. 9 shows the change in the mean of cross-validation accuracy resulted from applying a k-NN with different values of parameter k. As shown in Fig. 9, in general, the accuracy decreased by increasing the size of the neighborhood.

The algorithm was applied to both normalized and unnormalized attributes. The triangles in Fig. 9 show the results after the data were normalized, and the circles represent the results prior to normalization. Fig. 9 clearly shows that normalizing data enhances the performance of k-NN. In almost all cases, the accuracy of the k-NN on normalized data is at least 10% better. This is understandable knowing that similarity in the k-NN algorithm is measured using Euclidean distance. This constructive impact could
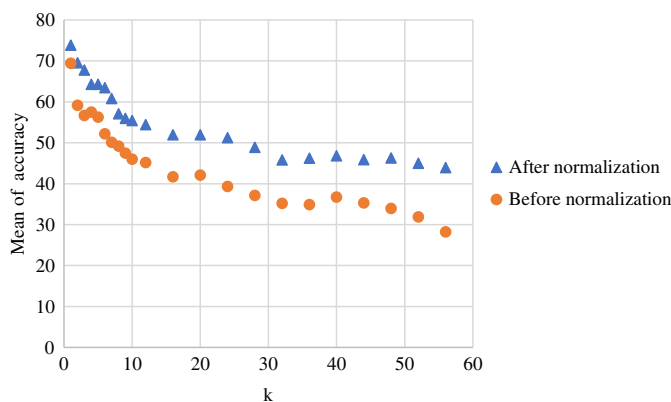

Fig. 9. Cross-validation accuracy of applying k-NN. Normalizing data increased the accuracy.

also be explained by the heterogeneous features used in the training set (and their span of values). For instance, although the annual average daily traffic (AADT) is an integer varying between 0 and 120,000, the years since the last remedial action is a real number between 0 and 26. Therefore, using these numbers in a Euclidean distance formula without any normalization does not result in the best solution. Although algorithms such as k-NN or clustering algorithms, which rely on a distance measure, are sensitive to data normalization, the variation in dimensions of data does not affect other algorithms such as decision tree or random forest (Hastie et al. 2009; Provost and Fawcett 2013).

Ensemble Learning

A similar analysis was conducted for ensemble learning algorithms. Fig. 10 compares the performance of two ensemble algorithms with their corresponding base learner. The curve with triangles shows the accuracy of GBT versus the number of trees, and the curve with circles represents a random forest. The accuracy of the base learner (i.e., the decision tree in this case) is shown by a square. These numbers are the result of 10-fold cross-validation. The accuracy of algorithms has two components: the mean and the standard deviation of accuracy.

Fig. 10 shows that both ensemble algorithms perform better than a single base learner. Also, it demonstrates that increasing the number of weak leaners boosts the performance of the ensemble of classifiers (i.e., higher mean and lower standard deviation of accuracy). Finally, when it comes to comparing the two ensemble learning methods, one can easily discern that GBT is continuously outperforming its counterpart. Fig. 10 also indicates that the asymptotic performance of the two ensemble learners is different. Whereas the mean of accuracy for the random forest almost stabilizes after 20 trees, the GBT's average accuracy continues to improve by including more base learners. Also, the standard deviation of accuracy decreases by increasing the number of trees. It means that the ensemble models learned from more base learners are usually more stable.

Better performance of GBT could be understood based on two reasons: (1) the classification approach of the algorithm itself; and
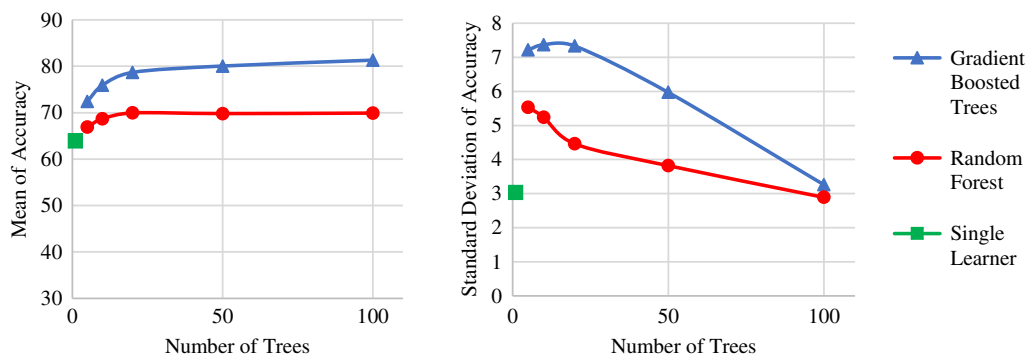
Fig. 10. Mean and standard deviation of accuracy versus the number of base learners for gradient boosted trees and random forests. Both algorithms do considerably better than a single learner.
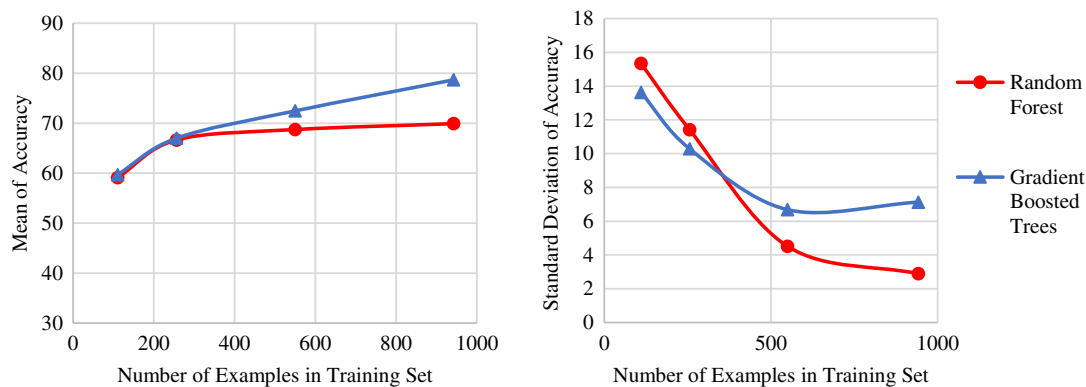
Fig. 11. Impact of the size of training set on accuracy. Increasing the number of examples makes models more accurate and stable.
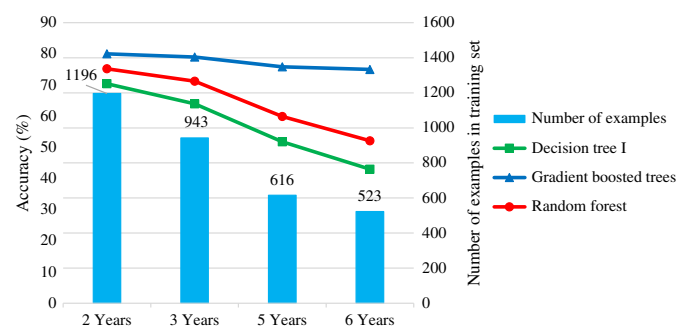
(2) the informativeness of the training data. As explained, GBT, and boosting algorithms in general, is one level more sophisticated than random forest. Whereas in the random forest, each base learner casts a vote and the majority vote predicts the class, in a GBT, a weight is assigned to each base learner. Therefore, the prediction of more accurate learners (i.e., better trees) is emphasized in a GBT algorithm.

The informativeness and the relative role of each attribute play a major role in accuracy. Random forests algorithm dices the data randomly. The resulting data subsets are used to train decision trees. At the end, each tree votes for the final classification. In a data set where all attributes have the same informativeness, random forest should perform well. If not, as is the case in this study, there will be more noise in each subset, effectively reducing the accuracy. Moreover, the training data have a role in the superiority of an algorithm. In this particular problem, the information gain of attributes is highly skewed, and a few attributes have a negligible impact on perdition. In such circumstances, the random forest does not have its best performance. The reason is that random forest algorithm dices the data randomly to learn random trees. A tree that is learned based on the least informative attributes has the same vote as a tree that is learned based on the best attributes. However, obviously, the latter has a higher accuracy. Therefore, random forest is more accurate in the cases in which the attributes are more or less equally informative.

The results of Fig. 10 are based on 942 examples of GPS sections. It was shown previously that increasing the number of examples in the training set could boost the accuracy of the models. To show the impact of the size of the training set, the GBT and the

random forest were trained and tested using different numbers of examples. Fig. 11 shows the result for 110, 256, 550, and 942 examples. Increasing the number of examples enhances the accuracy, with GBT gaining more as the sample size increases. However, the standard deviation of random forest drops more with increased sample size.

Also, limited research is available on how the accuracy of prediction changes with a change in the prediction horizon. So, the accuracy of the selected algorithms was examined across different periods: two, three, five and six. The results are shown in Fig. 12 (the aforementioned numbers presented were for predicting the PCI after 3 years). Such a comparison is beneficial because roads are known to behave differently in the short and long terms

Fig. 12. Comparing the accuracy of three algorithms over four prediction intervals.

© ASCE 04020022-10 J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

(Ziari et al. 2016). The aforementioned results were for predicting the PCI after 3 years. The mean of accuracy of the algorithms is shown on the left, and the number of examples used in training the models for each interval is presented on the right axis. Of course, smaller numbers of roads have enough data (in LTPP) for longer intervals. The number of examples in the training set for each interval is shown by four bars.

In Fig. 12, the curve with squares represents the average accuracy of Decision tree I or RapidMiner decision tree. The other two curves represent ensemble algorithms trained based on this base learner. The gradient boosted trees algorithm is shown in by the curve with triangles and the random forest by the curve with circles. Both the random forest and GBT of Fig. 12 are trained using 20 base learners. As shown previously, using more base learners can increase the accuracy dramatically. Clearly, the accuracy of the GBT is superior followed by the random forest. All three algorithms have lower accuracy in predicting the PCI over longer intervals. This lower accuracy for longer horizons could be explained from two perspectives: data analytics and pavement engineering. First, from a data analytic viewpoint, the smaller size of the training set for longer intervals reduces the accuracy of the models. Second, from a pavement engineering perspective, pavement deterioration is a stochastic and nonlinear phenomenon. So, more variations can occur in longer intervals. This increase in stochasticity makes the prediction more difficult.

One final observation is that although the accuracy of all three algorithms slopes down, this decrease is less substantial for the GBT. Although the difference might be negligible in the 2-year prediction interval, it certainly is considerable when conducting the prediction for 5 and 6 years. This means that the GBT is a better choice when predicting using smaller number of data points and longer intervals of predictions.

Analysis on Different Climatic Regions

LTPP data span large variations in climatic conditions, maintenance and management regimes, and design attributes. Therefore, it is expected that conducting the analysis using more coherent data (of roads in similar climatic conditions) would result in more accurate models. By localizing data, one can cancel out some of the noises or aleatory uncertainties that are not easy to quantify (Lou et al. 2001; Wu 2015). Also, reducing the variations due to climate-related attributes may provide more insights on the role of other features. The training set was divided into four different subsets based on LTPP's definition of climatic regions. The name of each region and the number of roads in that partition is given in Table 3. Table 3 also presents the results of a 10-fold cross-validation of the GBT and the random forest learned from and tested on the data of each climatic zone.

Table 3 clearly indicates that the accuracy of the models trained based on separate climatic zones is higher than the results of the aggregate data. In Fig. 11, when the size of the training set is smaller than 350, the mean of accuracy is below 70%, but the accuracy of the model in Table 3 is above that even with smaller training examples. Considering the climatic zones with similar or smaller data size, the accuracy of each model is higher than 70%, with Region 2 reaching above 90%. This higher accuracy could be the result of eliminating some of the uncertainties that are not captured by the models. In other words, narrowing down the training data to a more coherent subset cancels out some of the aleatory and epistemological uncertainties.

The accuracy of the models for climatic Regions 3 and 4 is lower than Region 2 despite the larger size of their training set.

Table 3. Results of cross-validation for GBT and random forest in each climatic region

Climatic zone code	Climatic region	Random forest	Gradient boosted trees	Number of examples
1	Dry, freeze	71 ± 8.89%	70% ± 9.05%	151
2	Dry, nonfreeze	92% ± 6.86%	91% ± 7.82%	213
3	Wet, freeze	75 ± 7.02%	79% ± 7.97%	228
4	Wet, nonfreeze	76 ± 9.75%	79% ± 7.74%	350
Total	—	—	—	942

Table 4. Statistical measures of DeltaPCI over 3 years for different climatic regions

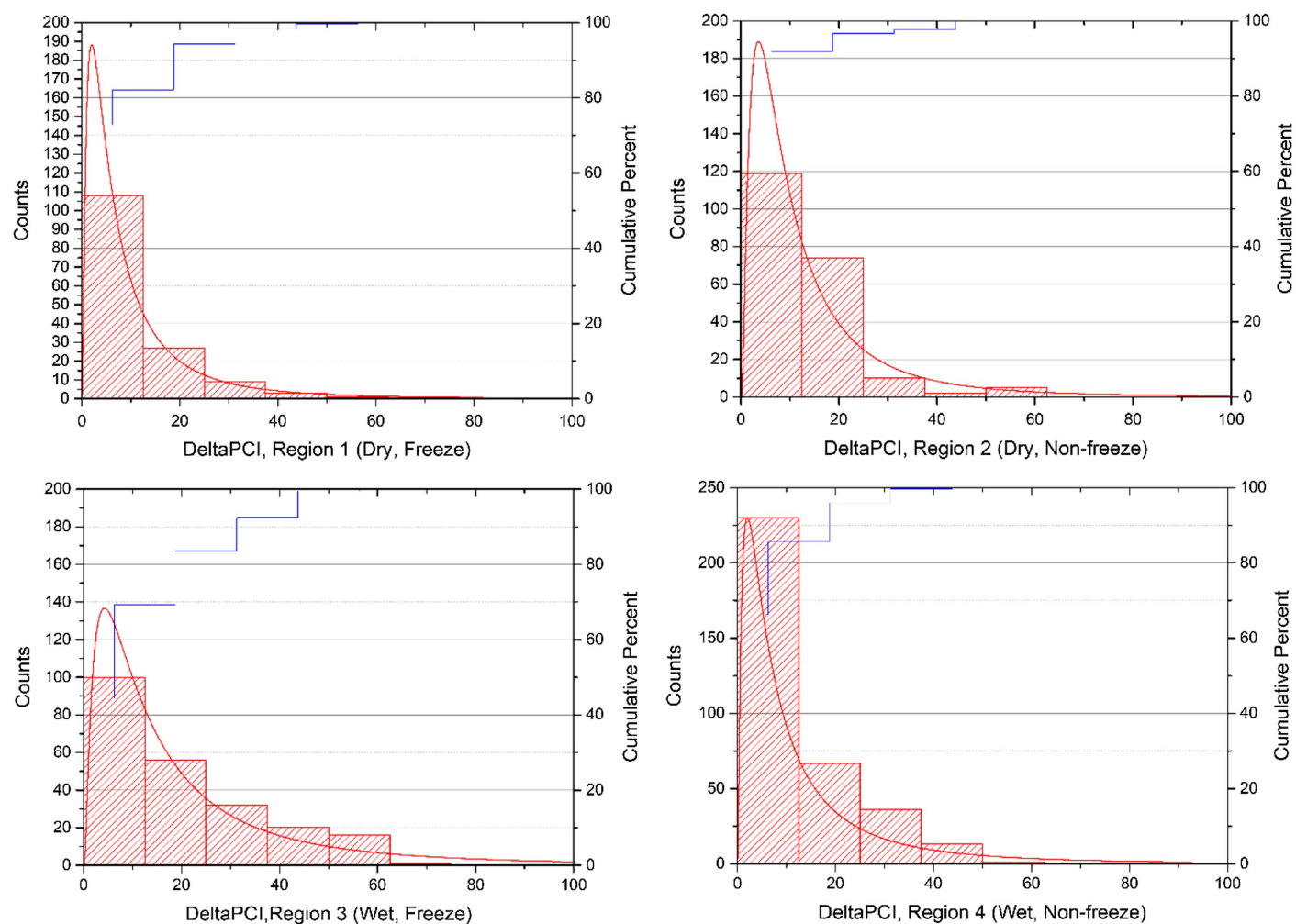| Climatic zone code	Climatic region	Mean of DeltaPCI	Standard deviation of DeltaPCI	Interquartile range
1	Dry, freeze	9.17	10.41	12.5
2	Dry, nonfreeze	11.14	10.56	11.7
3	Wet, freeze	19.05	16.5	20.4
4	Wet, nonfreeze	11.34	11.60	15.1

To investigate this point, the deterioration of the PCI (i.e., DeltaPCI) over 3 years in different regions was examined. The statistical measures of PCI deterioration in 3 years is given in Table 4. This table indicates that the roads in Regions 3 and 4 not only experience a more extreme deterioration but also a more stochastic one. This could be inferred from the higher mean, standard deviation, and interquartile range of DeltaPCI in these regions. The histograms of DeltaPCI of all regions are drawn in Fig. 13. The percent of road sections that experience a DeltaPCI larger than 25 (i.e., the percent that do not fall into the first two bins) is 31% and 14% for Regions 3 and 4. This is around 8% for other regions. Furthermore, Fig. 13 clearly suggests that the dispersion in PCI is the largest respectively in Regions 3 and 4.

The larger mean and dispersion in Region 3 suggest that the process of deterioration in wet-freeze areas is associated with more uncertainty. Such stochastic deterioration, especially in the wet-freeze region, could be a result of other factors that are specific to this type of climate. Examples of such factors studied in the literature are deicing salt (Hassan et al. 2002; Suraneni et al. 2018) and relative humidity (Caro et al. 2008; Jeong and Zollinger 2005). The impact of such factors could be captured by affordable data such as percent of relative humidity and the type and amount of deicing salt. Such features could be integrated into the current models in future research, and their impact and informativeness could be assessed. Incorporating such attributes would be more meaningful knowing that climate change is already affecting the freeze-thaw cycles, the pattern of precipitation, humidity, and temperature.

Testing the Models Using SPS Data

A benefit of machine learning algorithms is that can easily be tested on any unseen data set. Although the cross-validation technique (which was used in the preceding section) is a reliable method of accuracy evaluation, the trained models were tested on a separate data set. The data chosen as the test set are the SPS data. Table 5 summarizes the accuracy of testing models discussed previously on 137 SPS examples. So, in this case, the training set included 942 examples and Test set 137. Comparing the results of Table 5 with the cross-validation accuracy, it was found that the accuracy of models in classifying the SPS sections is slightly lower. This lower accuracy is explainable given the differences between SPS

© ASCE 04020022-11 J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

Fig. 13. Deterioration of the PCI in different climatic regions. DeltaPCI is on average larger in Region 3 and has a larger dispersion in Regions 3 and 4.

Table 5. Accuracy of models tested on SPS sections

Model	Test accuracy (%)
Gradient boosted trees	75
Random forest (based on C4.5)	73
Random forest	63
Naïve Bayes coupled with kernels	65
Decision tree	59
C4.5 decision tree	70
Naïve Bayes	55

and GPS sections. As mentioned, the SPS sections are specifically built for testing the impact of design and maintenance attributes on pavement health. Therefore, these road sections are deliberately built, operated, and maintained with many variations. A model that does not include these design and maintenance attributes will not capture all these variations—especially the simplified (and circumstantial) attributes used in this study.

Conclusions

Using an example of a data set containing roads in the US and Canada, this study showcased the value of data analytics in infrastructure asset management and decision making. Predicting the deterioration in PCI was used as the case example in this study because of its popularity among municipalities in Ontario. The study compared GBT, random forest classifier, and naïve Bayes coupled with kernels. It was shown that only with affordable (and circumstantial and noncausal) data, the algorithms were able to predict deterioration with high accuracy. The data were segmented multiple times and according to different rationales. The algorithm parameters were also altered several times. This helped in analyzing the reasons for accuracy variations. Consequently, several tips were discussed for enhancing accuracy in practice and in future studies. Of particular interest is the performance of two examples of ensemble learning algorithms: random forest and GBT. The accuracy of GBT and random forest was boosted by 25% and 10%, respectively, compared with their base learner.

Other methods of learning, such as k-NN (a lazy learner), were investigated as well. The impact of data normalization was tested on k-NN, and it was observed that normalization increased the cross-validation accuracy by at least 10%. Furthermore, kernel density estimates were used to increase the accuracy of the naïve Bayes classifier (instead of a generic Gaussian distribution). It was observed that the accuracy of naïve Bayes classifier was boosted dramatically by using more realistic density estimates.

Three different classifiers were trained to predict the PCI in 2, 3, 5, and 6 years. The accuracy of all algorithms declined as the

interval increased. However, the decrease in the accuracy of the decision tree and random forest was drastically higher than that of the GBT. This means that the GBT not only had a better performance over different data sizes but also over different prediction intervals. This observation could be highly important and beneficial to municipalities with limited data availability. The models were trained based on the GPS data and then were tested on the SPS data. The accuracy of models was slightly lower than cross-validation, but still relatively high compared with other methods (used in similar research work). This shows that models learned based on a large enough training set with relevant attributes could be used even in other jurisdictions. Also, predictive models were developed for specific climatic regions. Focusing the analysis on a particular climatic region results in a higher accuracy. The accuracy of some of the ensemble learners exceeded 90% in some climatic regions.

This study was focused on the prediction of PCI of asphalt roads within 2, 3, 5, and 6 years. It could be, however, generalized to other performance indicators and other horizons of prediction easily. Furthermore, including meaningful climate stressors such as precipitation, freeze-thaw, and temperature in the model could open the door to a study on the impact of climate change on roads. Finally, it should be pointed out that the current study was developed in close consultation with experts and municipalities in Ontario. It only included attributes available to smaller municipalities. Adding more attributes, especially more mechanistic and technical features, could enhance the accuracy of the models even more. Assessing the cost of acquiring those additional attributes and conducting a cost-accuracy trade-off analysis or a value of information analysis could be considered to check if having those additional attributes is worthwhile. Furthermore, although prediction accuracy can improve with proper analysis, no modeling technique can avoid an inevitable issue: trying to predict deterioration using data that embed design, construction, and maintenance methods geared at delaying/counteracting deterioration.

Appendix. PCI Calculation

This appendix is devoted to explaining the tasks conducted for automating PCI calculation. The ASTM (2015) guideline calculates the PCI based on distress values and their severity levels using a set of graphs and an iterative process. After the density for each defect is calculated at all three severity levels, density values are translated into deduct values (DVs) via ASTM graphs. The impact of different distresses is combined by adjusting the number of DVs and translating them to corrected deduct values (CDVs). This translation is performed via another set of ASTM curves. Finally, the maximum CDV is calculated through an iterative process. When the maximum CDV is determined, the PCI is calculated by simply deducting the maximum CDV from 100.

Using ASTM curves tends to be very time-consuming and error-prone. Following the ASTM approach manually is almost impractical when the size of data is large. For the current study the formulas for all ASTM curves were extracted via curve-fitting and embedded in a computer program to avoid the pitfalls of manual PCI calculation. Using the developed computer program, makes PCI calculation possible for large data like the training set of the current study. Details of digitizing ASTM curves and finding mathematical functions have been given by Piryonesi and El-Diraby (2018) and details on the workflow of PCI calculation have been provided by Piryonesi and El-Diraby (2020).

All relevant equations were extracted for all distresses at all severity levels and for both sets of curves. Altogether, eight correction

curves and 31 deduct value versus distress density curves were digitized, mathematically expressed, and embedded in a Python program. Using this program, the PCI was calculated for the retrieved data. The values calculated by the program were compared with the example calculated by ASTM D6433-07 and several other examples to assure the validity of developed program.

Data Availability Statement

Some or all data, models, or code generated or used during the study are available from the corresponding author by request. These items include parts of the prepared PCI data, traffic, climatic attributes, and maintenance data.

References

AASHTO. 2008. *Mechanistic-empirical pavement design guide: A manual of practice*. Washington, DC: AASHTO.

Abdelaty, A., H. D. Jeong, and O. Smadi. 2018. "Barriers to implementing data-driven pavement treatment performance evaluation process." *J. Transp. Eng. Part B: Pavements* 144 (1): 04017022. https://doi.org/10.1061/JPEODX.0000023.

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-Suleiman, T. I., and A. M. S. Shiyab. 2003. "Prediction of pavement remaining service life using roughness data—Case study in Dubai." *Int. J. Pavement Eng.* 4 (2): 121–129.

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.

Archilla, A. R., and S. Madanat. 2000. "Development of a pavement rutting model from experimental data." *J. Transp. Eng.* 126 (4): 291–299. https://doi.org/10.1061/(ASCE)0733-947X(2000)126:4(291).

ASTM. 2015. *Standard practice for roads and parking lots pavement condition index surveys*. ASTM D6433-07. West Conshohocken, PA: ASTM.

Ayed, A. 2016. *Development of empirical and mechanistic empirical performance models at project and network levels*. Waterloo, Canada: Univ. of Waterloo.

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.

Caro, S., E. Masad, A. Bhasin, and D. N. Little. 2008. "Moisture susceptibility of asphalt mixtures, Part 1: Mechanisms." *Int. J. Pavement Eng.* 9 (2): 81–98. https://doi.org/10.1080/10298430701792128.

Chandra, S., C. R. Sekhar, A. K. Bharti, and B. Kangadurai. 2013. "Relationship between pavement roughness and distress parameters for Indian highways." *J. Transp. Eng.* 139 (5): 467–475. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000512.

Chen, C., and J. Zhang. 2011. "Comparisons of IRI-based pavement deterioration prediction models using New Mexico pavement data." In *Proc., Geo-Frontiers 2011*, 4594–4603. Reston, VA: ASCE.

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.

Choi, J., T. M. Adams, and H. U. Bahia. 2004. "Pavement roughness modeling using back-propagation neural networks." *Comput.-Aided Civ. Infrastruct. Eng.* 19 (4): 295–303.

Darter, M. I., L. Titus-Glover, and H. L. Von Quintus. 2009. *Implementation of the mechanistic-empirical pavement design guide in Utah: Validation, calibration, and development of the UDOT MEPDG user's guide*. Salt Lake City: Research Div., Utah DOT.

Dong, Q., and B. Huang. 2014. "Evaluation of influence factors on crack initiation of LTPP resurfaced-asphalt pavements using parametric

© ASCE   04020022-13   J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

survival analysis." *J. Perform. Constr. Facil.* 28 (2): 412–421. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000409.

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: Transportation Research Board.

Elkins, G. E., P. Schmalzer, T. Thompson, and A. Simpson. 2003. *Long-term pavement performance information management system pavement performance database user reference guide*. Rep. No. FHWA-RD-03-088. McLean, VA: Turner-Fairbank Highway Research Center.

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

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, Univ. of Texas at El Paso.

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, National Academy of Sciences.

Fwa, T. F., S. A. Tan, and L. Y. Zhu. 2004. "Rutting prediction of asphalt pavement layer using C-ϕ model." *J. Transp. Eng.* 130 (5): 675–683. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:5(675).

Gharaibeh, N. G., Y. Zou, and S. Saliminejad. 2010. "Assessing the agreement among pavement condition indexes." *J. Transp. Eng.* 136 (8): 765–772. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000141.

Gong, H., Y. Sun, W. Hu, P. A. Polaczyk, and B. Huang. 2019. "Investigating impacts of asphalt mixture properties on pavement performance using LTPP data through random forests." *Constr. Build. Mater.* 204: 203–212.

Gong, H., Y. Sun, X. Shu, and B. Huang. 2018. "Use of random forests regression for predicting IRI of asphalt pavements." *Constr. Build. Mater.* 189 (Nov): 890–897. https://doi.org/10.1016/j.conbuildmat.2018.09.017.

Gulen, S., K. Zhu, and J. Weaver. 2001. *Development of improved pavement performance prediction models for the Indiana pavement management system*. FHWA/IN/JTRP-2001/17. Indianapolis: Indiana Department of Transportation.

Guo, X., and D. H. Timm. 2015. "Local calibration of MEPDG using national center for asphalt technology test track data." In *Proc., TRB 94th Annual Meeting*. Washington, DC: Transportation Research Board.

Haas, R., and T. Kazmierowski. 1996. *TAC's new management design and management guide*. Ottawa: Transportation Association of Canada.

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).

Hameed, A., R. Dai, and B. Balas. 2016. "A decision-tree-based perceptual video quality prediction model and its application in FEC for wireless multimedia communications." *IEEE Trans. Multimedia* 18 (4): 764–774. https://doi.org/10.1109/TMM.2016.2525862.

Hassan, Y., A. O. Abd El Halim, A. G. Razaqpur, W. Bekheet, and M. H. Farha. 2002. "Effects of runway deicers on pavement materials and mixes: Comparison with road salt." *J. Transp. Eng.* 128 (4): 385–391. https://doi.org/10.1061/(ASCE)0733-947X(2002)128:4(385).

Hastie, T., Tibshirani, R., and Friedman, J. 2009. "The elements of statistical learning." In Vol. 1 of *Bayesian forecasting and dynamic models*, 1–694. New York: Springer.

Hossain, M. I., L. S. P. Gopisetti, and M. S. Miah. 2017. "Prediction of international roughness index of flexible pavements from climate and traffic data using artificial neural network modeling." In *Airfield and highway pavements 2017*, 256–267. Reston, VA: ASCE.

Hossain, M. I., L. S. P. Gopisetti, and M. S. Miah. 2019. "International roughness index prediction of flexible pavements using neural networks." *J. Transp. Eng. Part B: Pavements* 145 (1): 04018058. https://doi.org/10.1061/JPEODX.0000088.

InfoPave. 2018. "LTPP InfoPave—Home." Accessed July 27, 2018. https://infopave.fhwa.dot.gov/.

Jeong, J.-H., and D. G. Zollinger. 2005. "Environmental effects on the behavior of jointed plain concrete pavements." *J. Transp. Eng.* 131 (2): 140–148. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:2(140).

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. https://doi.org/10.3141/2155-14.

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 Technol. Pergamon* 55: 510–517.

Kaveh, A., M. A. Motie Share, and M. Moslehi. 2013. "Magnetic charged system search: A new meta-heuristic algorithm for optimization." *Acta Mech.* 224 (1): 85–107. https://doi.org/10.1007/s00707-012-0745-6.

Kırbaş, U., and M. Karaşahin. 2016. "Performance models for hot mix asphalt pavements in urban roads." *Constr. Build. Mater.* 116 (Jul): 281–288. https://doi.org/10.1016/j.conbuildmat.2016.04.118.

Li, M., and H. Wang. 2018. "Prediction of asphalt pavement responses from FWD surface deflections using soft computing methods." *J. Transp. Eng. Part B: Pavements* 144 (2): 04018014. https://doi.org/10.1061/JPEODX.0000044.

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/10.1177/0361198196152400124.

Lin, J.-D., J.-T. Yau, and L.-H. Hsiao. 2003. *Correlation analysis between international roughness index (IRI) and pavement distress by neural network*. Washington, DC: Transportation Research Board.

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

Liu, L., and N. G. Gharaibeh. 2014. "Bayesian model for predicting the performance of pavements treated with thin hot-mix asphalt overlays." *Transp. Res. Rec.* 2431 (1): 33–41. https://doi.org/10.3141/2431-05.

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).

McCabe, B. Y., H. Hamledari, A. Shahi, P. Zangeneh, and E. R. Azar. 2017. "Roles, benefits, and challenges of using UAVs for indoor smart construction applications." In *Proc., Congress on Computing in Civil Engineering*, 349–357. Reston, VA: ASCE.

Meegoda, J. N., and S. Gao. 2014. "Roughness progression model for asphalt pavements using long-term pavement performance data." *J. Transp. Eng.* 140 (8): 1–7. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000682.

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. Vancouver, BC, Canada: IEEE.

Nitsche, P., R. Stütz, M. Kammer, and P. Maurer. 2014. "Comparison of machine learning methods for evaluating pavement roughness based on vehicle response." *J. Comput. Civ. Eng.* 28 (4): 04014015. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000285.

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.

Paterson, W. D. O. 1989. "A transferable causal model for predicting roughness progression in flexible pavements." *Transp. Res. Rec.* 1215: 70–84.

Piryonesi, S. M., and T. El-Diraby. 2020. "Data analytics in asset management: Cost-effective prediction of pavement condition index." *J. Infrastruct. Syst.* 26 (1): 04019036. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000512.

Piryonesi, S. M., and T. E. El-Diraby. 2017. "A data analytics solution for predicting the pavement condition index of roads using the most

© ASCE   04020022-14   J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022

affordable attributes." In *CSCE: Leadership in sustainable infrastructure*. Vancouver, Canada: Canadian Society for Civil Engineering.

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*. Rep. No. FHWA-HRT-18-065. Washington, DC: Federal Highway Administration.

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.

Prasad, J. R., S. Kanuganti, P. N. Bhanegaonkar, A. K. Sarkar, and S. Arkatkar. 2013. "Development of relationship between roughness (IRI) and visible surface distresses: A study on PMGSY roads." *Procedia—Social Behav. Sci.* 104: 322–331.

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

Pulugurta, H., Q. Shao, and Y. J. Chou. 2009. "Pavement condition prediction using Markov process." *J. Stat. Manage. Syst.* 12 (5): 853–871. https://doi.org/10.1080/09720510.2009.10701426.

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.

Sadeghi, J., E. R. Najafabadi, and M. E. Kaboli. 2017. "Development of degradation model for urban asphalt pavement." *Int. J. Pavement Eng.* 18 (8): 659–667. https://doi.org/10.1080/10298436.2015.1095912.

Sakhaeifar, M. S., Y. Richard Kim, and P. Kabir. 2015. "New predictive models for the dynamic modulus of hot mix asphalt." *Constr. Build. Mater.* 76 (Feb): 221–231. https://doi.org/10.1016/j.conbuildmat.2014.11.011.

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. Artificial Neural Syst.* 2009: 1–9. https://doi.org/10.1155/2009/308239.

Shahnazari, H., M. A. Tutunchian, M. Mashayekhi, and A. A. Amini. 2012. "Application of soft computing for prediction of pavement condition index." *J. Transp. Eng.* 138 (12): 1495–1506. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000454.

Smith, J. T., and S. L. Tighe. 2004. "Assessment of overlay roughness in long-term pavement performance test sites: Canadian case study." *Transp. Res. Rec.* 1869 (1): 126–135.

Sollazzo, G., T. F. Fwa, and G. Bosurgi. 2017. "An ANN model to correlate roughness and structural performance in asphalt pavements." *Constr. Build. Mater.* 134: 684–693.

Souliman, M. I., M. S. Mamlouk, M. M. El-Basyouny, and C. E. Zapata. 2010. "Calibration of the AASHTO MEPDG for flexible pavement for Arizona conditions." In *Proc., Transportation Research Board 89th Annual Meeting*. Washington, DC: Transportation Research Board.

Suraneni, P., V. J. Azad, O. B. Isgor, and J. Weiss. 2018. "Role of supplementary cementitious material type in the mitigation of calcium oxychloride formation in cementitious pastes." *J. Mater. Civ. Eng.* 30 (10): 04018248. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002425.

Terzi, S. 2007. "Modeling the pavement serviceability ratio of flexible highway pavements by artificial neural networks." *Constr. Build. Mater.* 21 (3): 590–593.

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.* 39 (2): 377–387.

Tsai, Y., and Z. Wang. 2015. *Development of an asphalt pavement raveling detection algorithm using emerging 3D laser technology and macro-texture analysis*. Washington, DC: Transportation Research Board.

Walubita, L. F., S. I. Lee, A. N. M. Faruk, T. Scullion, S. Nazarian, and I. Abdallah. 2017. *Texas flexible pavements and overlays: Year 5 report—Complete data documentation*. Rep. No. FHWA/TX-15/0-6658-3. College Station, TX: Texas A&M Transportation Institute.

Wang, Y. 2013. "Ordinal logistic regression model for predicting AC overlay cracking." *J. Perform. Constr. Facil.* 27 (3): 346–353. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000327.

Witczak, M. W., K. Kaloush, T. Pellinen, M. El-Basyouny, and H. Von Quintus. 2002. *Simple performance test for Superpave mix design*. Washington, DC: Transportation Research Board.

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. J. Lu, M. Gunaratne, and Q. Xiang. 2003. "Forecasting overall pavement condition with neural networks: Application on Florida Highway Network." *Transp. Res. Rec.* 1853 (1): 3–12.

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.

© ASCE 04020022-15 J. Transp. Eng., Part B: Pavements

J. Transp. Eng., Part B: Pavements, 2020, 146(2): 04020022