

# Time-series forecasting of road distress parameters using dynamic Bayesian belief networks

Time-series  
forecasting

Babitha Philip and Hamad AlJassmi

*Department of Civil and Environmental Engineering, United Arab Emirates University, Al Ain, United Arab Emirates and Emirates Center for Mobility Research (ECMR), United Arab Emirates University, Al Ain, United Arab Emirates*

317

Received 8 September 2022  
Revised 27 April 2023  
10 July 2023  
10 August 2023  
Accepted 14 August 2023

## Abstract

**Purpose** – To proactively draw efficient maintenance plans, road agencies should be able to forecast main road distress parameters, such as cracking, rutting, deflection and International Roughness Index (IRI). Nonetheless, the behavior of those parameters throughout pavement life cycles is associated with high uncertainty, resulting from various interrelated factors that fluctuate over time. This study aims to propose the use of dynamic Bayesian belief networks for the development of time-series prediction models to probabilistically forecast road distress parameters.

**Design/methodology/approach** – While Bayesian belief network (BBN) has the merit of capturing uncertainty associated with variables in a domain, dynamic BBNs, in particular, are deemed ideal for forecasting road distress over time due to its Markovian and invariant transition probability properties. Four dynamic BBN models are developed to represent rutting, deflection, cracking and IRI, using pavement data collected from 32 major road sections in the United Arab Emirates between 2013 and 2019. Those models are based on several factors affecting pavement deterioration, which are classified into three categories traffic factors, environmental factors and road-specific factors.

**Findings** – The four developed performance prediction models achieved an overall precision and reliability rate of over 80%.

**Originality/value** – The proposed approach provides flexibility to illustrate road conditions under various scenarios, which is beneficial for pavement maintainers in obtaining a realistic representation of expected future road conditions, where maintenance efforts could be prioritized and optimized.

**Keywords** Road distress parameters, Dynamic Bayesian belief networks, Time-series data, Performance prediction modeling

**Paper type** Research paper

© Babitha Philip and Hamad AlJassmi. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors thank the Ministry of Energy and Infrastructure (MoEI) in UAE for their immense support and for providing with the data related to this research. This research is financially supported by the Office of Associate Provost for Research at UAE University under the fund code 31R291.



Construction Innovation  
Vol. 24 No. 1, 2024  
pp. 317-340  
Emerald Publishing Limited  
1471-4175  
DOI 10.1108/CI-09-2022-0233

## 1. Introduction

The fundamental goal of a pavement management system (PMS) is the development of maintenance and rehabilitation (M&R) programs to enhance pavement performance and lengthen its service life (Sholevar *et al.*, 2022). The key performance indicators for assessing the pavement condition are the intensity of several road distress parameters, namely, International roughness index (IRI), rutting, bleeding, deflection, cracking, potholes and others (Mcdaniel and Shah, 2003). The combined effect of several pavement deterioration factors such as design features, traffic and environmental conditions results in the deterioration of the pavements over time (Visintine *et al.*, 2015). Thus, it is important to understand what are the most significant factors contributing to road deterioration, and how to quantify them to explore the current and future pavement conditions and to define optimal maintenance decisions. In addition, how to incorporate the historical road deterioration data with traffic and climatic factors to improve future predictions requires more attention. Furthermore, validating the accuracy of the road deterioration prediction models will aid in improving the wide range of its applicability.

Various performance prediction models have been reported in the literature; the majority of them fall into two categories: deterministic models and probabilistic models (George *et al.*, 1989). Different approaches were used over the years for evaluating pavement performance, including regression analysis (Pan *et al.*, 2011; Kim and Kim, 2006; Mills *et al.*, 2012), artificial neural networks (ANN) (Yang *et al.*, 2003; Roberts and Attoh-Okine, 1998; Vyas *et al.*, 2022), Bayesian belief networks (BBN) (Pantuso *et al.*, 2021; Xiao *et al.*, 2022), Markov chain models (Kobayashi *et al.*, 2010; Kobayashi *et al.*, 2012; Salman and Gursoy, 2022), decision trees (Piryonesi and El-Diraby, 2020) and others (Hu *et al.*, 2022). According to Osorio-Lird *et al.* (2018), the advantages of probabilistic models, specifically Bayesian approaches and Markov chain models over regression and ANN methods in predicting the pavement performance are their capability to capture uncertainty and employment of a transition probability matrix (TPM) for predicting future behavior based on the current state. Furthermore, the capability of the BBN structure to probabilistically estimate the unknown values in a domain can overcome the challenge faced by many statistical methods which demand a huge amount of inspection data to obtain an accurate deterioration model (Han *et al.*, 2014). Thus, it is important to further explore the capabilities of BBN in developing a comprehensive road deterioration prediction model.

Dynamic BBNs are an extension of static BBNs with the inclusion of a time factor (Neapolitan and Jiang, 2010). Considering the complexity associated with the existing time-series models, dynamic BBN models are developed based on certain hypotheses to simplify the forecasting process. Dynamic BBN models have the Markov property and follow the invariant transition probability hypothesis (Li *et al.*, 2019). This research aims to investigate the efficiency of dynamic BBNs to forecast the behavior of road distress parameters, namely, IRI, rutting, deflection and cracking based on historical data collected from 32 major road sections in the United Arab Emirates (UAE) over the years 2013 to 2019.

The major objectives of this study are:

- To develop dynamic BBN models to forecast the behavior of major road distress parameters taking into consideration the influence of various road distress parameters, traffic factors, environmental factors and road factors.
- To make inferences based on the threshold values adopted in the study region.
- To examine the accuracy and validity of the developed models based on statistical tests.

- To perform scenario analysis to demonstrate the practical application of the developed models, thereby, to guide pavement maintainers in obtaining a realistic representation of expected future road conditions, in which maintenance efforts could be prioritized, and proactive cost-effective pavement management plans could be drawn.

In the following, a brief outline of several Bayesian methods in pavement performance prediction modeling and the motivation to adopt dynamic BBN in this study is mentioned in Section 2. Further explanation of dynamic BBN is presented in Section 3. Section 4 explains the study's methodology. The steps followed in model development including data collection, data preprocessing, development of dynamic BBN models and the validation of the developed models are discussed. Finally, Section 5 presents the study's conclusion, emphasizing the implications of the developed models in better pavement management solutions and future directions.

## 2. Bayesian methods in pavement performance prediction modeling

Pavement performance prediction models play a crucial role in road infrastructure management. The service life of pavements is a function of several factors including structural characteristics, climatic conditions, traffic loading and many more. In general, the allocation of highway budget depends on the performance forecast of pavements. To date, numerous models based on various approaches have been developed to address the pavement deterioration process. Among them, the Bayesian approach has efficiently captured unobserved heterogeneity (arising from the unobserved influence of various factors) of pavement sections, expert knowledge and/or prior information and most importantly uncertainty associated with the variables involved (Hong and Prozzi, 2006).

Xiao *et al.* (2022) developed a probabilistic Bayesian neural network (BNN) model by combining Bayesian theory and neural networks based on pavement data collected from Shanxi province, China. The major steps in the development of the BNN model were the imputation of missing pavement data, the selection of significant deterioration factors based on correlation coefficient analysis, the determination of the prior probability distribution of weights using a feed-forward neural network technique, the establishment of BNN-based probabilistic prediction model and the analysis of uncertainty value associated with each road section. The prediction accuracy of the BNN model was high and comparable to that of the neural network-based deterministic model.

Inkoom *et al.* (2020) developed a Bayesian survival model (BSM) for predicting highway pavement performance. BSM was developed using the Bayes theorem and Markov Chain Monte Carlo methods. The cracking and riding conditions of the pavements collected by the Florida Department of Transportation from the years 1976 to 2011 were used to develop the model. It was observed that the BSM model had fewer prediction errors with no overfitting problems when compared to other predictive models based on classical product limit survival (Kaplan–Meier) estimates and parametric univariate survival models. The posterior probabilities and the survival curves associated with the three models were compared. The Kaplan–Meier estimate (a nonparametric model) predicted pavement deterioration with minimum evidence (observations) and the univariate parametric model had a good fit for the observed data. However, BSM described the survival probabilities efficiently based on prior information while capturing the associated uncertainty.

Pantuso *et al.* (2021) developed a pavement deterioration model to predict the Critical condition index (CCI), the deterioration index used by the Virginia Department of Transportation (VDOT) for representing the pavement condition. Initially, the measured

pavement deterioration (DI) collected from interstate, primary and secondary pavement road sections were fitted with different statistical distributions [normal, Poisson and negative binomial (NB) distribution] and it was found the NB distribution, which is a combination of Poisson distribution and Gamma distribution, provided a good representation of the observed pavement data. A linear empirical Bayesian (LEB) approach was further adopted to obtain better estimates of pavement conditions based on the fitted model and the observed data. The prediction error was very less in this method since the LEB estimator corrected inappropriate pavement condition projections for specific pavement portions (pavement sections with little deterioration at late ages and/or significantly deteriorated parts at early ages), bringing them closer to the mean value.

Based on previous studies, BBN is deemed ideal for capturing the complex time-dependent pavement deterioration process due to the following:

- Prior information is included in the analysis.
- Results are expressed as probability distributions to address uncertainty.
- Causal relationships among the variables are obtained.
- Expert opinions are incorporated when historical data are not available.

The major contribution of this study is the development of time-series models for the road distress parameters considering external factors, namely, traffic, environmental and road factors. Dynamic BBNs are selected for the analysis because it follows Markovian property and invariant transition probability. Further explanation of the mathematical background of the dynamic BBN is provided in the below section.

### 3. Dynamic Bayesian belief networks

BBNs are the directed acyclic graphs (DAG) used in probabilistic models to express uncertain relationships between random variables. Variables are represented graphically as nodes, while arcs or arrows represent relationships (both causal and noncausal). Based on relationships with other variables in the domain, a probability distribution function is assigned to each variable in the DAG. The fundamental rule involved in the development of BBN is Bayes' theorem, expressed as in [equation \(1\)](#), for events X and Y:

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} \quad (1)$$

where  $P(X|Y)$  is the conditional probability of X given Y. A BBN represents probabilistic relationships among random variables by combining our belief (prior knowledge expressed as prior distribution) and the data collected (expressed as likelihood) to make inferences based on posterior distribution about a complex problem at some point in time. The standard BBN are extended to model the temporal relationships by capturing the dynamic aspect of a process, known as dynamic BBN ([Neapolitan and Jiang, 2010](#)). The design of dynamic BBN is initiated by the simple assumption that an event in the present time can cause another event in the future and not vice-versa. Thus, in time series modeling based on dynamic BBN, directed arcs flow forward in time. In other words, dynamic BBN has Markov property as it follows first-order Markovian assumption, i.e. the value of a variable X at time t ( $X_t$ ) depends only on the value of X at a previous time  $t-1$  ( $X_{t-1}$ ). In addition, this process is stationary, i.e.  $P(X_t | X_{t-1})$  does not change with time.

Let  $X_i$  ( $i = 1, \dots, n$ ) be a discrete random variable in a DAG. If there is a direct link from  $X_i$  to  $X_j$ , then  $X_i$  is a parent of  $X_j$  represented by  $\text{Pa}(X_j)$ . To represent the time series data of

the network for time  $t$  ( $t = 1, \dots, T$ ), the dynamic BBN is used. According to the Markovian assumption,  $X$  at time  $t$  ( $X_t$ ) depends only on the value of  $X$  at a previous time  $t-1$  ( $X_{t-1}$ ). Thus, for time  $t$  ( $t = 1, \dots, T$ ), the joint probability for variable  $X$  is represented as in equation (2) (Imoto et al., 2013):

$$P(X_1, \dots, X_T) = P(X_1)P(X_2 | X_1) \dots P(X_T | X_{T-1}) \quad (2)$$

Where,

$$P(X_t | X_{t-1}) = \prod_{j=1}^n P(X_{tj} | \text{Pa}(X_j)_{t-1}) \quad (3)$$

Combining equations (2) and (3):

$$P(X_1, \dots, X_T) = P(X_1) \prod_{t=2}^T \prod_{j=1}^n P(X_{tj} | \text{Pa}(X_j)_{t-1}) \quad (4)$$

In comparison to deterministic models, probabilistic models can account for uncertainties in pavement behavior (Tosun et al., 2017). A probabilistic model, as opposed to a deterministic model, provides the probability of each possible result, allowing decision-makers to make better judgments in pavement management (Rose et al., 2018).

#### 4. Methodology

Four time-series models are developed in this study corresponding to road distress parameters: IRI, rutting, deflection and cracking. To achieve this goal, the following methodology was adopted:

- **Data collection:** The data used for this study were collected from 32 road networks lying in the northern region of the UAE. Those roads are maintained by the Road Department of the Ministry of Energy and Infrastructure Development (MoEI), UAE. The time frame of data collection is seven years, from 2013 to 2019. The data collected can be classified into the road, traffic and environmental data. Among these, road data and traffic data were collected from the database of the Ministry. Data related to environmental conditions were collected from web resources.
- **Data preprocessing:** The raw data collected was complex mainly due to the heterogeneous nature of the variables, missing data and differences in road distress measurement intervals. Hence, several data preprocessing techniques were adopted to transform the raw data into a useful and efficient format before performing the analysis.
- **Development of time-series models:** The data collected were used to develop performance prediction models for IRI, rutting, deflection and cracking. The models developed represent the influence of road distress parameters and external factors on the development and progression of the pavement deterioration process. The dynamic BBN approach is adopted for representing the pavement deterioration over time.
- **Validation and analysis of the models:** The developed models were validated based on the K-fold cross-validation method and receiver operation characteristic (ROC) curves. Furthermore, the developed time-series models were tested for various conditions, and the future behavior of road distress parameters was observed.

4.1 Data collection

The data collected in this study majorly falls into three categories: road data, traffic data and environmental data.

4.1.1 Road data. The road data can be grouped into data related to road distress parameters and the road-specific data. The road distress data involved the details of road distress parameters; IRI, rutting, deflection and cracking. These parameters were measured by laser profilometer, laser rutting measurement system, falling weight deflectometer and laser crack measurement system respectively by the MoEI and the measurements were stored within their database. The road-specific data which is also mentioned as road factors includes the starting (initial) and ending (final) points of the road section where these road distress parameters were measured which represents the exact location of the road distress. It also includes the year of construction and the type of road network. The road types involved in this study are arterial, collector, expressway and freeway.

4.1.2 Traffic data. The number of heavy and light vehicles passing through the road networks in a year reported as traffic count (heavy vehicle) and Traffic count (light vehicle), respectively, are used in the analysis. The traffic flow in the directions South-North and West-East are recorded as “Forward” and in the directions North-South and East-West are designated as “Backward.”

4.1.3 Environmental data. Temperature (°C), humidity (percentage) and air pressure (millibar) are the environmental elements addressed in this study. These climatic data were gathered from weather records available on the online Web platform (*timeanddate, n.d.*).

Additional data related to the road networks under study is provided in [Appendix](#).

4.2 Data preprocessing

In this step, the following preprocessing techniques were applied to the collected data. In addition, the assumptions related to maintenance type and age of the road sections adopted to prepare the data for further analysis are mentioned in the following sub sections.

4.2.1 Uniform measurement intervals. The details of the road distress parameters considered in this study are provided in [Table 1](#).

Although the measuring interval for IRI, rutting and cracking is 10 m as shown in [Table 1](#), the starting and ending points of the measurement sections were different. In other words, the 10-m segment in which IRI was observed was different from the 10-m segment in which the other distresses were observed. Since we had to examine the characteristics of a specific road distress parameter with respect to other distress parameters, it was necessary to uniformize the measurement segments, so that a data point in the data set would represent the road distresses corresponding to the same road segments. To perform this, spreadsheet formulas to round the starting and ending points of the measurement sections to the nearest possible multiples of 10 were used (An example of the spreadsheet formulas

Distress factor	Unit	Measuring intervals in meters	Range of values
IRI	mm/m	10	0.00–52.99
Rutting	mm	10	0.00–52.65
Deflection	mm/100	100	0.00–374.00
Cracking	Percentage (%)	10	0.00–99.87

**Table 1.**  
Details of road  
distress parameters

**Source:** Created by the authors

used in given in [Appendix 2](#). Consequently, the road networks in the study were divided into a total of 132,999 road sections, each of a uniform 10-m length.

**4.2.2 Maintenance treatments.** The road authorities perform three types of maintenance activities on the road networks selected in this study: major treatment, surface treatment and partial treatment. However, data related to M&R activities were not available. Thus, the occurrence of maintenance activities during the study period was assumed based on the knowledge of maintenance types as described below:

- *Major treatment:* A road section is considered to have undergone a major treatment if all the road distress parameters (IRI, rutting, cracking and deflection) are corrected in consecutive years. This requires a top to bottom repair of the road structure and, hence, the most expensive method.
- *Surface treatment:* A road section is considered to have undergone surface treatment if all the road distress parameters except deflection are corrected in consecutive years.
- *Partial treatment:* A road section is considered to have undergone partial treatment if the shallow cracks, IRI and rutting are corrected and the deeper distresses (deep cracks and deflection) are not repaired. This is usually the least expensive method.

The type of maintenance performed was determined by comparing the values of road distress parameters across successive years. While comparing it was observed that some road sections did not undergo maintenance, evident by a higher current year value than the previous year, while other road sections had missing data either in the current year or the preceding year. Since maintenance of the road section is important in improving the service life of pavements ([Al-Saadi et al., 2021](#)), it was decided to keep only the treated roads (major, surface or partial) for further investigation. As a result of removing data related to road sections for which maintenance was not performed during the study period and road sections for which maintenance type was unknown due to missing values of the distress parameters in consecutive years, a total of 3,272 road sections were retained for further analysis.

**4.2.3 Age of the road section.** The year that the selected road networks were built was known. However, since the service life of the road network improves with maintenance treatments, the age of the road section is determined in this study based on previous maintenance time ([Al-Saadi et al., 2021](#)). The decision to measure road age based on previous maintenance time was backed up by the fact that considerable initiatives were launched by road authorities in the research region to maintain high-performance roads.

A data point from the final data set after performing various preprocessing steps as mentioned above is given in [Table 2](#).

#### *4.3 Development of dynamic Bayesian belief networks models for road distress parameters (prediction models)*

Four models are developed in this study to examine the intensity of road distress parameters, namely, IRI, rutting, deflection and cracking, based on data related to pavement deterioration factors. A software tool based on the BBN method, BayesiaLab 10.2 ([Bayesia S.A.S., 2002](#)), is selected for model development. Each model is developed based on 17 variables which includes the 16 variables listed in [Table 2](#) and an additional variable corresponding to the road distress parameter in the previous year represented by the “road distress parameter at  $t - 1$ ” according to the model. For example, while developing the forecasting model for cracking (Cracking model), in addition to the 16 variables given in [Table 2](#), an extra variable “cracking at  $t - 1$ ” is included in the analysis. The following subsections outline the steps

**Table 2.**  
Representation of a  
data point in the data  
set for analysis

Variables under study	Value
<i>Road factors</i>	
Road	E11
Road type	Arterial
Initial distance	8,540
Final distance	8,550
Maintenance type	Partial
Age of the road section	1
<i>Road distress parameters</i>	
Cracking at $t$	0
Deflection at $t$	51
IRI at $t$	1.016
Rutting at $t$	1.75375
<i>Environmental factors</i>	
Temperature	26
Humidity	49
Atmospheric pressure	1,019
<i>Traffic factors</i>	
Traffic count (heavy vehicles)	4,203,429
Traffic count (light vehicles)	487,421
Direction of traffic	Forward

**Source:** Created by the authors

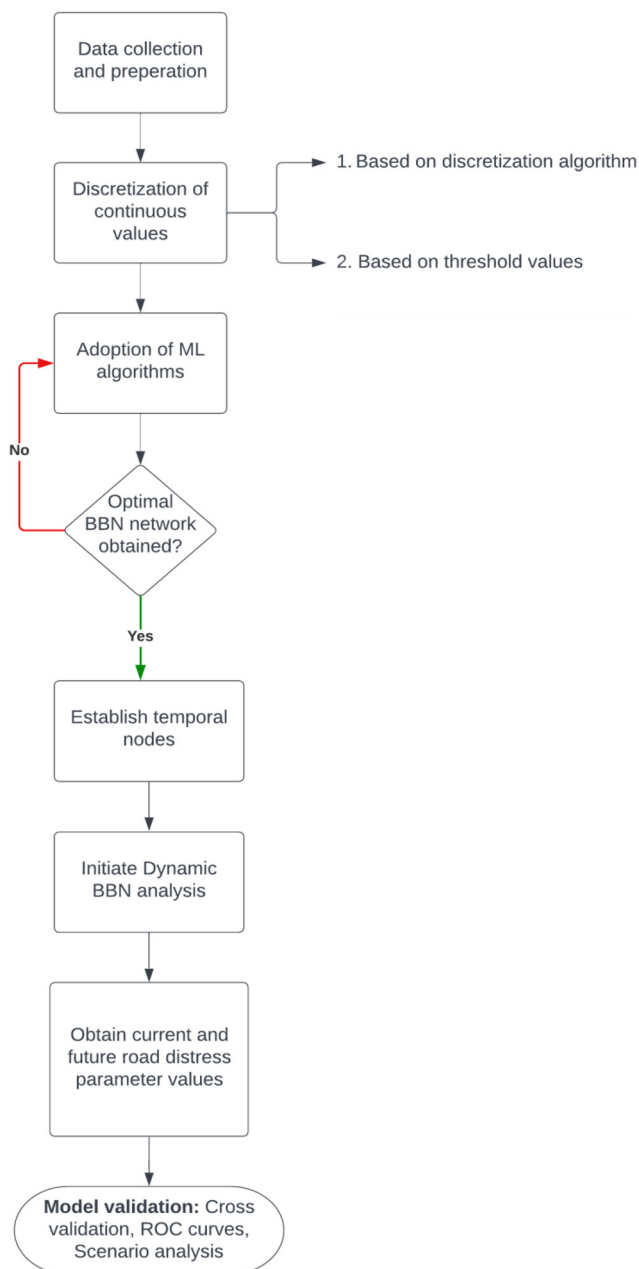
involved in the construction of the BBN model. The framework adopted in this study is given in [Figure 1](#) and the steps involved are explained in the following sections.

*4.3.1 Selection of an optimum Bayesian belief networks structure.* The data set obtained after the completion of the data preparation stage involved data related to 3,272 road sections of 10-m length. This data set was further used in the development of prediction models. Knowing the level of risk associated with the road segment is beneficial before deciding on the schedule of pavement maintenance activities. For instance, if the road has significant cracking (or any other road distress parameters), it can be deemed to be at high danger, and quick repair work is therefore recommended. On the other hand, if the cracking values are extremely low, the road stretch can be categorized as being in the low-risk group. Therefore, in this study, classification algorithms that predict discrete outcomes are advised over regression algorithms which predict a continuous value based on the input variables in a decision-making scenario for pavement management.

The data set used in this study includes both discrete and continuous values as seen in [Table 2](#). It is required to discretize the continuous values to classify the continuous values under different class labels. The discretization process involves two components:

- (1) the number of classes in which the continuous variables are divided; and
- (2) the threshold values for each class.

The discretization process can be carried out either automatically based on various discretization algorithms (univariate, bivariate and multivariate) or manually. In this study, initially, the values are discretized based on algorithms, and thereafter the discretization process is further improved manually based on the known threshold values adopted in the country. The different discretization algorithms tested in this study include Tree, Perturbed



Source: Created by the authors using LucidChart tool

**Figure 1.** Framework of dynamic BBN model development

Tree, Supervised Multivariate,  $R^2$ -GenOpt, K-means, density approximation, normalized equal distance, equal distance, equal frequency and unsupervised multivariate. Once the continuous variables are discretized, different machine learning (ML) algorithms are applied to the whole data set to obtain different BBN structures.

ML algorithms majorly belong to two categories supervised and unsupervised learning algorithms. When there is a “target node” (final output), supervised ML algorithms are applied to analyze the effect of each study component on the target. On the other hand, unsupervised ML techniques are used when there is no “target node,” to examine all direct probabilistic correlations between variables. In this work, supervised algorithms such as Naive Bayes, Augmented Naive Bayes, Tree Augmented Naive Bayes, Sons and Spouses, Markov Blanket, Tree Augmented Markov Blanket and Minimal Augmented Markov Blanket were examined, to develop the forecasting model for the road distress parameters, with road distress parameter as “target node.” For instance, the node “cracking at  $t$ ” is the target node while developing a forecasting model for studying the behavior of cracking in roads.

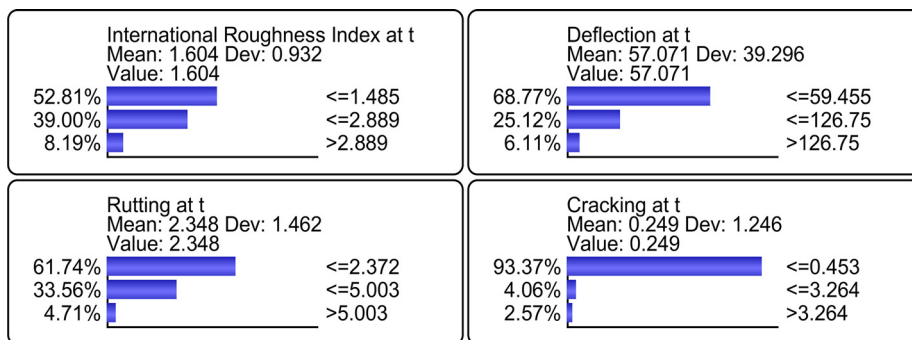
The optimum BBN model is chosen using the minimum description length (MDL) score. The intricacy of the structure and the information it yields are combined to get the MDL score. According to the MDL score’s guiding principles, every data set can be learned using a model based on how well it can compress the data and the model’s generalizability is indicated by the score. The best model is the one with the lowest MDL score, which is calculated as in equation (5), where,  $L(\text{Data}|\text{Model})$  is the data description length given the model, while  $L(\text{Model})$  is the model description length (Myung, 2001):

$$\text{MDL} = L(\text{Data} | \text{Model}) + L(\text{Model}) \quad (5)$$

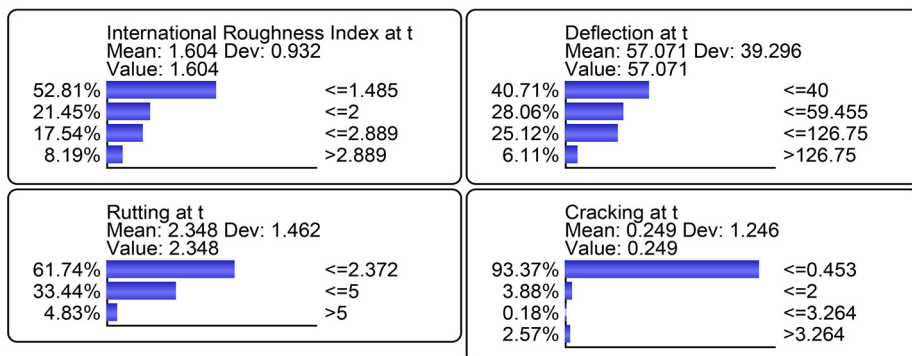
The discrete classes for the road distress parameters obtained according to the optimum network structure are shown in Figure 2(a). It can be noted that more than 50% of each road distress parameter lies in the lowest class. This might imply that the road networks under study are in comparatively good condition. However, the quality of the road network structures can be inferred based on the comparison with the threshold (safe limit) value accepted in the region, as the value should ideally fall within the predetermined safe limit. Thus, the road distress nodes were further improved based on the threshold values adopted by MoEI (the values were provided by the officials at the MoEI), which are  $\text{IRI} < 2 \text{ mm/m}$ , rutting  $< 5 \text{ mm}$ , deflection  $< 40 \text{ mm/100}$  and cracking  $< 2\%$ . The modified discretized classes for road distress parameters are given in Figure 2(b). The classes were modified by splitting the original class where the threshold value belongs into two separate classes. For instance, the threshold value for IRI is  $2 \text{ mm/m}$  and it falls into the class  $\leq 2.889$ . Thus, the class  $\leq 2.889$  were divided into two classes, one up to  $2 \text{ mm/m}$  and the other up to  $2.889 \text{ mm/m}$ . In addition, it should be noted that the splitting of the classes does not affect the values in other classes as evident from the probability distribution of the classes. Before splitting, the prior probability of class  $\leq 2.889$  was 39%, which was divided into 21.45% for class  $\leq 2$  and 17.54% for class  $\leq 2.889$ , which sums up to the original 39%.

The road distress parameter nodes were modified to include the threshold values accepted by the highway agency, to aid in better decision-making purposes for the MoEI. It can be seen that the nodes were updated by adding a class label corresponding to the threshold value, which created an extra class for all nodes except for rutting. For rutting, the safe limit is  $5 \text{ mm}$  and since one of the class labels in the previous discretization step was  $5.003 \text{ mm}$ , which was very close to  $5 \text{ mm}$ , it was decided to modify  $5.003 \text{ mm}$  as  $5 \text{ mm}$ .

*4.3.2 Establishing temporal relations.* The development of an optimal BBN structure was followed by the establishment of temporal nodes among the road distress parameters for



(a)



(b)

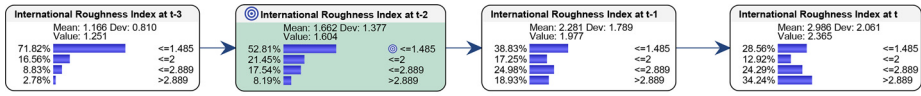
**Figure 2.** Discretization interval for road distress parameters

**Notes:** (a) Discretization interval based on optimization; (b) improved discretization intervals based on threshold values adopted at MoEI

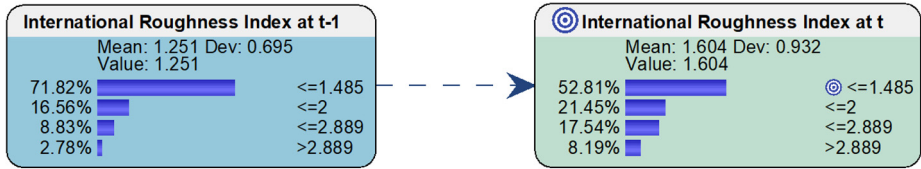
**Source:** Created by the authors using BayesiaLab software <https://www.bayesia.com/bayesia/>

implementing dynamic aspects into the model. To develop a time-series model, the relation between the nodes “road distress at  $t-1$ ” and “road distress at  $t$ ” was set to be temporal for performing dynamic BBN analysis. An example of the representation of time-series data for IRI is presented below. Initially, the arrow connecting “IRI at  $t-1$ ” and “IRI at  $t$ ” will be set to temporal, thus, establishing a temporal relation between them. The time-series data can then be represented either by duplicating the network for each time step as shown in Figure 3(a) or by a compact representation by establishing temporal relation between the nodes as in Figure 3(b). Both provide the time-series data; however, the initial figure is an extended static BBN network where a limited number of time steps can be represented and the latter one is the dynamic network with the inclusion of temporal arcs. In other words, the structure in Figure 3(a) is an unrolled version of the structure in Figure 3(b).

The time-series data, thus, obtained can be used to draw the progression curve for different classes of road distress values, which aids in understanding the deterioration rate of the roads. Figure 4 represents an illustration of the curves corresponding to different classes of IRI values for five years based on the data collected from the whole data set. The curves given here are not subjected to any other conditions. However, based on the evidence



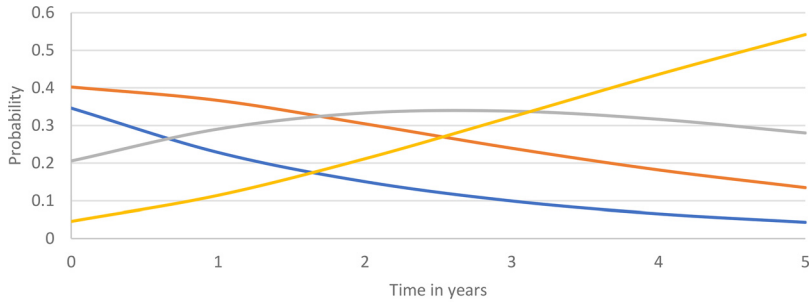
(a)



(b)

**Figure 3.**  
Representation of  
time-series data

**Notes:** (a) BBN developed for three-time steps; (b) dynamic BBN model with temporal arcs  
**Source:** Created by the authors using BayesiaLab software <https://www.bayesia.com/bayesia/>



**Figure 4.**  
Progression curve for  
different classes of  
IRI values

**Source:** Created by the authors using Microsoft Excel

(data collected) related to other factors involved in the study (traffic, environmental and road-specific factors), the curves can be modified. Such curves are highly beneficial to evaluating various scenarios, thus, generating custom solutions based on road conditions.

The progression curves for different classes of IRI values in Figure 4 imply that the probability of IRI to be in a low-risk condition (i.e. classes  $\leq 1.485$  and  $\leq 2$ ) without the need for any maintenance action is high in the initial years (till Year 2). From Year 2 to Year 3, the highest probability of occurrence is for class IRI  $\leq 2.889$ , which is slightly above the threshold value. After Year 3, a significant and rapid increase in the probability of IRI values above 2.889 mm/m is observed, recommending the initiation of maintenance activities in the study region.

*4.4 Model validation and discussion*

Model validation is the process of ensuring that the model fulfills its intended purpose. There are numerous validation methods available. The most common method for model validation is cross-validation, often known as leave-one-out or K-fold. In this study, the

developed BBN models are validated based on the K-fold cross-validation method, receiver operating characteristic (ROC) curves and scenario analysis. The following subsections discuss the applicability of the developed dynamic models.

**4.4.1 Cross-validation method.** The data set was divided into training/testing sets using a K-fold cross-validation method. This strategy is commonly used when dealing with unbalanced data sets. The data set is randomly divided into “K” groups for this purpose ( $K = 10$  is chosen for this study). Each group serves as the testing set, while the other groups serve as the training set, and this procedure is repeated “K” times. As a result, each group is used as a test set once and as a training set  $K-1$  times. The total number of sections screened for analysis in this study is 3,272, hence, the size of each “K” group is about 327. The performance metrics reported in this study are the precision, recall and area under the curve (AUC) index. These performance metrics validate the classification efficiency of the developed models. Table 3 lists the performance matrices of each model.

Precision and Recall are among the most widely adopted matrices to evaluate classification models, which are calculated below in equations (6) and (7) respectively.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (6)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (7)$$

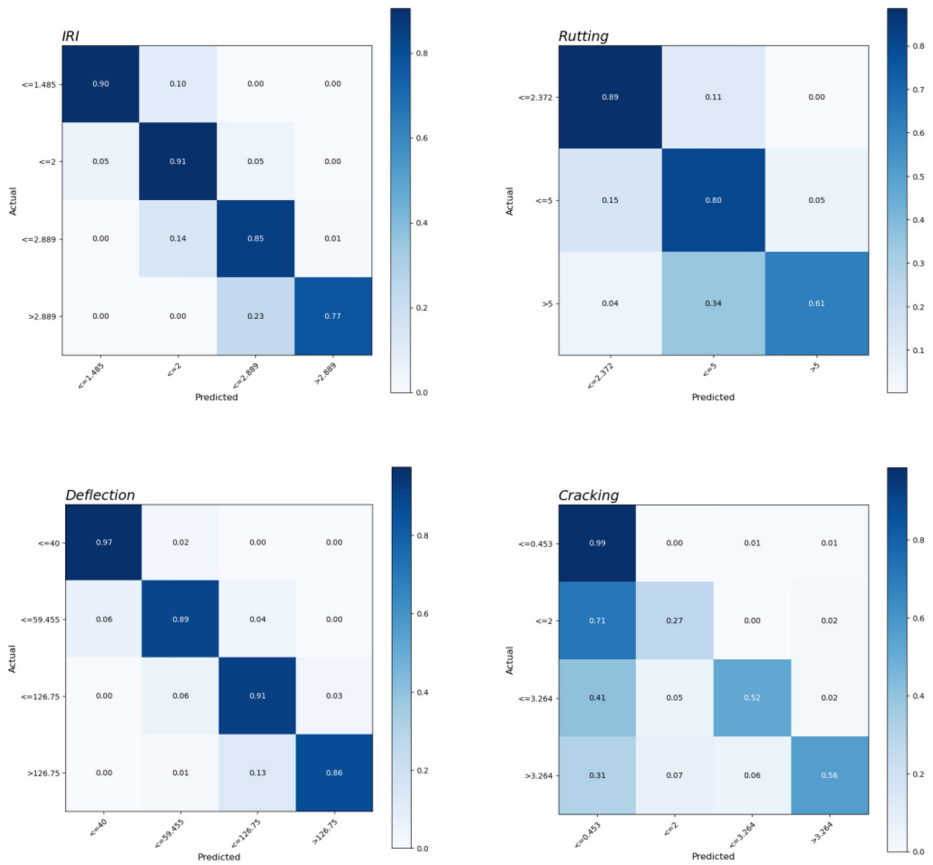
The AUC index is the two-dimensional area beneath the ROC curve. It is an aggregate measure of the model’s performance at various classification thresholds. From Table 3, it is evident that values of precision, recall and AUC are satisfactory for all the models. Figure 5 presents the confusion matrices for the road distress parameters based on the precision of the model. The distribution of the total 3,272 datapoints in each class of the road distress parameter is provided in Table 4.

The confusion matrices imply that the performance prediction models are generally performing well for all classes of the road distress parameters with some exceptions. It should be noted that the performance level is less for classes with a small sample size. For instance, the least value obtained is 27% for class  $\leq 2$  in the cracking model, which had a comparatively smaller sample size as can be seen in the Table 4. In addition, according to the confusion matrix obtained for the cracking model, for class  $\leq 2$ , the probability of the model to predict class  $\leq 2$  as  $\leq 0.453$  is 71%. To understand this observation more clearly, the distribution of the data for the class  $\leq 2$  was investigated. The probability distribution of the class  $\leq 2$  is given in Figure 6. It was noted that among the 63 values in class  $\leq 2$ , above 90% of the values fall into the interval of  $\leq 0.765$ . This means that

Model	Precision	Performance metrics (%)	
		Recall	AUC
IRI	85.69	89.64	94.34
Rutting	84.53	84.81	94.44
Deflection	91.02	90.81	97.44
Cracking	94.22	93.46	87.61

**Table 3.** Performance metrics for the performance prediction models

**Source:** Created by the authors



**Figure 5.** Confusion matrices for the road distress parameters

**Source:** Created by the authors using Python programming

Class	IRI		Rutting		Deflection		Cracking	
	Class	Sample size	Class	Sample size	Class	Sample size	Class	Sample size
$\leq 1.485$	$\leq 1.485$	1,133	$\leq 2.372$	2,020	$\leq 40$	1,747	$\leq 0.453$	2,992
$\leq 2$	$\leq 2$	1,317	$\leq 5$	1,094	$\leq 59.455$	1,080	$\leq 2$	63
$\leq 2.889$	$\leq 2.889$	673	$> 5$	158	$\leq 126.75$	328	$\leq 3.264$	133
$> 2.889$	$> 2.889$	149	-	-	$> 126.75$	117	$> 3.264$	84
Total		3,272		3,272		3,272		3,272

**Table 4.** Sample size of each class of the road distress parameters

**Source:** Created by the authors

above 90% of the data in the class  $\leq 2$  are very close to the previous class  $\leq 0.453$ . This distribution of the data can be the reason for the cracking model to highly interpret class  $\leq 2$  as  $\leq 0.453$  (71%). Further discretization of the  $\leq 2$  would improve the model's accuracy. Nonetheless, this class was intentionally left as-is to comply with the thresholds used by the MoEI, as discussed in the previous section.

Furthermore, in this study, the prediction errors were noted mainly for cracking because most of the values for cracking were close to zero, which showed a lower impact of cracking on UAE roads which could be attributed to the UAE's well-maintained road network. Another possibility is that the road sections investigated in this study were recently repaired for cracking (Colombier, 2004).

**4.4.2 Receiver operating characteristic curve.** The model evaluation is further carried out through ROC curve. A ROC curve represents the performance of a classification model at different classification thresholds. The curve is plotted based on true positive rate versus false positive rate as computed in equations (8) and (9) below:

$$TPR = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (8)$$

$$FPR = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}} \quad (9)$$

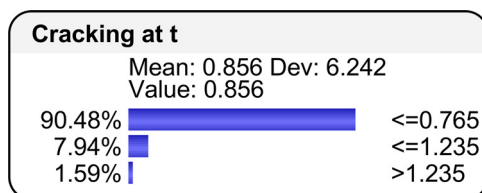
Figure 7 depicts the ROC curve associated with each class for IRI. Similar curves were obtained for other road distress parameters. The predictive performance of the model improves with the ROC curve escalating toward the top of the y-axis (Al Jassmi *et al.*, 2021).

The ability of the BBN models created to make precise predictions is demonstrated by the fact that the ROC curves corresponding to each road distress parameter in this study are found to be lying toward the top-left of the graph for all classes of the road distress parameter.

**4.4.3 Scenario analysis for the dynamic Bayesian belief networks models.** Pavement performance prediction models based on dynamic BBN can be used to track the behavior of road distress parameters under various scenarios. The Bayesian inferences based on the performance prediction models help pavement maintainers to make better decisions. To demonstrate the applicability of the models developed in this study, time-series curve corresponding to three scenarios given in the Figure 8 (selected from the data set of this study) are analyzed to generate IRI prediction curves. The influence of the age of the road section and quantity of heavy vehicles are taken into account. The scenarios considered are as below.

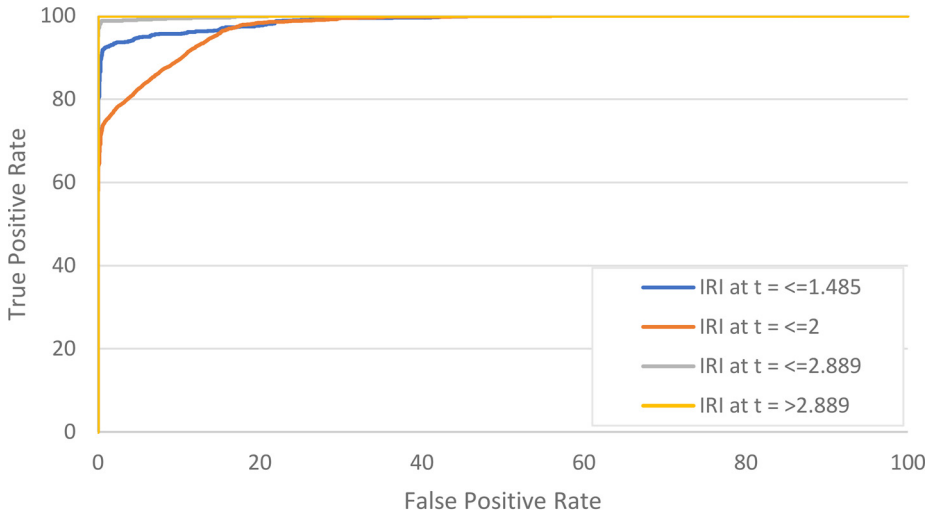
*Case 1:* Age of the road section is minimum, i.e. the roads are maintained recently (before one year). No evidence is set for other nodes, and are estimated based on their probability distributions.

*Case 2:* Age of the road is maximum, i.e. the roads have not been maintained recently (before five years). No evidence is set for other nodes, and are estimated based on their probability distributions.



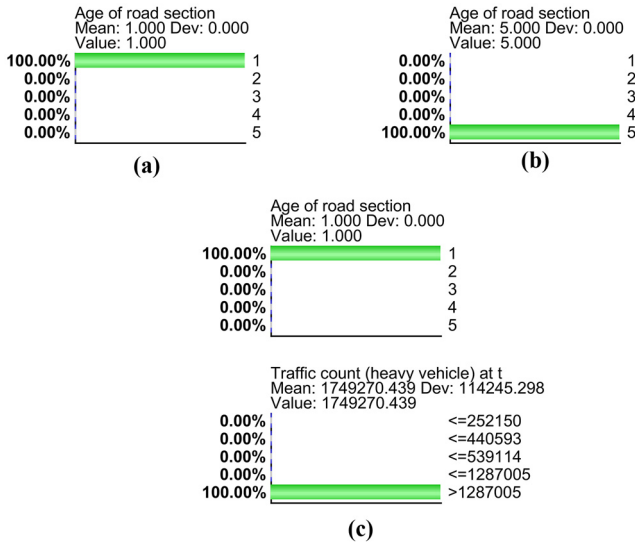
**Source:** Created by the authors using BayesiaLab software <https://www.bayesia.com/bayesia/>

**Figure 6.**  
Probability distribution for class <= 2



**Figure 7.**  
ROC curve  
corresponding to  
different classes  
of IRI

**Source:** Created by the authors using Microsoft Excel



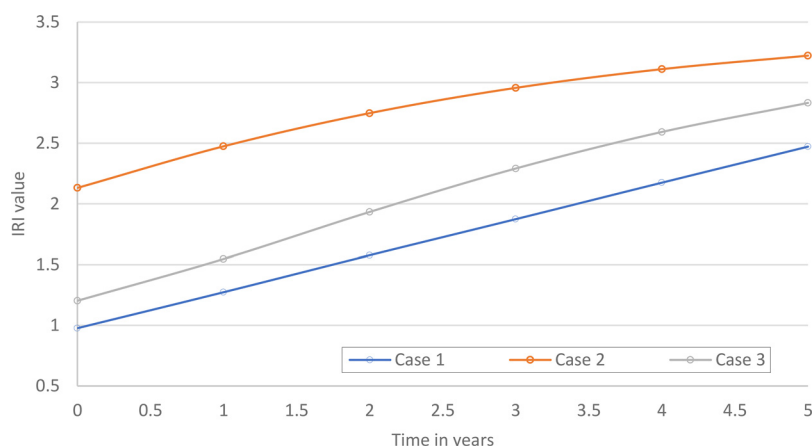
**Figure 8.**  
Different scenarios  
considered to  
generate time-series  
curves

**Notes:** (a) Case 1; (b) Case 2; (c) Case 3

**Source:** Created by the authors using BayesiaLab software <https://www.bayesia.com/bayesia/>

*Case 3:* Age of the road section is minimum (maintained before one year) and the traffic count of heavy vehicles is maximum (>1,287,005 per year). No evidence is set for other nodes, and are estimated based on their probability distributions.

The curves obtained corresponding to three different scenarios are given in [Figure 9](#). According to the curves obtained, IRI is progressing over the years, however, the rate of



**Source:** Created by the authors using Microsoft Excel

**Figure 9.**  
Progression of IRI for cases 1, 2 and 3

progress is different for different conditions. The curve corresponding to Case 1 shows that the rate of progress of IRI is slow in the initial years after maintenance. Here, the value of IRI corresponding to Year 0 is the IRI value observed for the road sections maintained before one year. The curve corresponding to Case 2 represents the effect of roads being left unmaintained. For this case, Year 0 represents the IRI value of the road sections maintained before five years. Thus, Year 0 for Case 2 marks the beginning of the sixth year of the road section. Case 3 demonstrates the combined effect of the age of the road and traffic count, which shows the progress of IRI values are higher in recently maintained roads if the traffic is extreme, compared to Case 1 where the effect of traffic is not taken into consideration.

This example shows how the models developed in this work can depict the severity of pavement deterioration under various scenarios. Similarly, various situations can be examined for different road distress parameters. Such analyses based on the BBN approach are critical in making sound pavement management decisions since pavement deterioration is a complicated phenomenon due to its unpredictable and variable character. Additional aspects, such as construction quality and drainage facility, must be considered in future research to accurately depict the pavement deterioration process (Rose *et al.*, 2018).

Several techniques are adopted for pavement condition assessment and maintenance. Yepes *et al.* (2016) developed a hybrid algorithm combining Greedy randomized adaptive search procedure (GRASP) and a series of constructive and local search heuristics to optimize maintenance programs with technical and budgetary restrictions. According to this study, preventive activities are better for the long-term effectiveness of maintenance programs compared to M&R activities. Torres-Machi *et al.* (2017) also adopt the hybrid GRASP method to optimize maintenance programs by integrating environmental impacts in addition to the technical and budgetary restrictions, where the importance of proactive actions is highlighted compared to reactive actions. Although the financial aspect is beyond the scope of the current study, the methodology adopted in the current study enables the pavement maintainers to assess the current and future pavement condition even when the data is partially available making it more favorable to adopt preventive actions at an early stage. Singh *et al.* (2018) used fuzzy analytical hierarchy process and fuzzy weighted average method to prioritize maintenance activities based on values of road distress parameters, namely, IRI, surface modulus, rut depth and friction coefficient. The

contribution of each distress parameter is analyzed through a pair-wise comparison matrix, represented as a weight coefficient. The values of these parameters serve as the criteria for prioritizing maintenance activities especially when adequate funds are not available. Similarly, the current study also evaluates the values of road distress parameters to analyze the pavement condition, however, the interrelationships among different deterioration factors are also taken into consideration which improves the practical application of the proposed approach.

## 5. Conclusion

In this study, dynamic BBN is used to develop four pavement performance prediction models based on the behavior of road distress parameters: IRI, rutting, deflection and cracking. The data used to develop the models were collected from major road sections in UAE over a time frame of seven years. The conditions under which the pavement behavior is observed include traffic conditions, environmental conditions, maintenance status and certain road features. These models can enhance the service life and the quality of the road networks by allowing the decision-makers to foresee the progression of road degradation under a variety of scenarios.

The developed probabilistic time-series models were validated based on K-fold cross-validation techniques. The confusion matrices obtained for each class of the road distress parameters showed the prediction performance of the models developed (Table 4). It was noted that the probability of correct predictions was above 70% for all cases except certain classes:  $>5$  for rutting and  $\leq 2$ ,  $\leq 3.264$ ,  $> 3.264$  classes for cracking. The reason behind these exceptions is the sample size of these discretization classes. The classes which had less precision in prediction had comparatively smaller sample sizes. Thus, lack of a sufficient quantity of data in a class will affect the prediction performance of the corresponding class. In addition, it was observed that there is a high chance of misclassification when values in a class are skewed toward the lower range of a class. This issue can be resolved by adopting better discretization intervals. In this study, initially, several discretization algorithms were used followed by manually choosing the discretization intervals to customize the model based on the threshold values adopted in the MoEI. This method of discretization was adopted to demonstrate the practicability of the model, which aids in creating alerts when the road distress values exceed the safe limit accepted for the region.

The developed time-series models represent the influence of external factors provided as evidence on the occurrence of road distress parameters. Variations in the probability distribution of road distress parameters observed based on evidence provided by the user are beneficial. Testing various scenarios based on the combination of various levels of contributing factors can yield useful inferences. For example, in the UAE, the maintenance cycle is five years, and to maintain high-quality roads, highway agencies implement "major" treatment plans. This is an expensive task, and transportation authorities frequently postpone repair tasks owing to funding restrictions. In this case, the dynamic BBN models developed in this study will allow pavement maintainers to test various cost-effective alternatives, such as performing a partial treatment every three years rather than a major treatment every five years, while taking into account the interaction of the explanatory variables. Such analysis can produce recommendations for the best use of pavement management resources.

The potential of dynamic BBN to observe a process over time and forecast its future characteristics is highlighted in this work. The behavior of road distress parameters as a result of various explanatory variables, such as traffic and environmental factors, is efficiently captured. Inferences generated by dynamic BBN models with Markovian

property and invariant transition probability hypothesis aid in the deep understanding of the complex pavement deterioration process. Prior knowledge of the future behavior of road distress parameters under the influence of various factors will aid in the adoption of cost-effective management practices without jeopardizing the quality of the road networks. In future studies, the incorporation of additional factors, particularly construction quality and drainage facility, will improve the applicability of the developed models. The findings of this paper advocate the use of dynamic BBN for structural health monitoring applications in infrastructures such as bridges, tunnels and other structures, which are deteriorating gradually.

## References

- Al Jassmi, H., Al Ahmad, M. and Ahmed, S. (2021), "Automatic recognition of labor activity: a machine learning approach to capture activity physiological patterns using wearable sensors", *Construction Innovation*, Vol. 21 No. 4, pp. 555-575, available at: [www.emerald.com/insight/content/doi/10.1108/CI-02-2020-0018/full/html](http://www.emerald.com/insight/content/doi/10.1108/CI-02-2020-0018/full/html)
- Al-Saadi, I., Wang, H., Chen, X., Lu, P. and Jasim, A. (2021), "Multi-objective optimization of pavement preservation strategy considering agency cost and environmental impact", *International Journal of Sustainable Transportation*, Vol. 15 No. 11, pp. 826-836, doi: [10.1080/15568318.2020.1821413](https://doi.org/10.1080/15568318.2020.1821413).
- Bayesia S.A.S (2002), "BayesiaLab", available at: [www.bayesiab.com/](http://www.bayesiab.com/)
- Colombier, G. (2004), "Cracking in pavements: nature and origin of cracks", *Prevention of Reflective Cracking in Pavements*, CRC Press, London, pp. 14-29, available at: [www.taylorfrancis.com/chapters/edit/10.1201/9781482294996-7/cracking-pavements-nature-origin-cracks-colombier](http://www.taylorfrancis.com/chapters/edit/10.1201/9781482294996-7/cracking-pavements-nature-origin-cracks-colombier)
- George, K.P., Rajagopal, A.S. and Lim, L.K. (1989), "Models for predicting pavement deterioration", *Transportation Research Record*, No. 1215, pp. 1-7.
- Han, D., Kaito, K. and Kobayashi, K. (2014), "Application of Bayesian estimation method with Markov hazard model to improve deterioration forecasts for infrastructure asset management", *KSCE Journal of Civil Engineering*, Vol. 18 No. 7, pp. 2107-2119, doi: [10.1007/s12205-012-0070-6](https://doi.org/10.1007/s12205-012-0070-6).
- Hong, F. and Prozzi, J.A. (2006), "Estimation of pavement performance deterioration using Bayesian approach", *Journal of Infrastructure Systems*, Vol. 12 No. 2, pp. 77-86.
- Hu, A., Bai, Q., Chen, L., Meng, S., Li, Q. and Xu, Z. (2022), "A review on empirical methods of pavement performance modeling", *Construction and Building Materials*, Vol. 342 No. PB, p. 127968, doi: [10.1016/j.conbuildmat.2022.127968](https://doi.org/10.1016/j.conbuildmat.2022.127968).
- Imoto, S., Matsuno, H. and Miyano, S. (2013), *Gene Networks: Estimation, Modeling, and Simulation, Computational Systems Biology: From Molecular Mechanisms to Disease*, 2nd ed., Elsevier, doi: [10.1016/B978-0-12-405926-9.00006-X](https://doi.org/10.1016/B978-0-12-405926-9.00006-X).
- Inkoom, S., Sobanjo, J.O., Chicken, E., Sinha, D. and Niu, X. (2020), "Assessment of deterioration of highway pavement using Bayesian survival model", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2674 No. 6, pp. 310-325, doi: [10.1177/0361198120919112](https://doi.org/10.1177/0361198120919112).
- Kim, S.-H. and Kim, N. (2006), "Development of performance prediction models in flexible pavement using regression analysis method", *KSCE Journal of Civil Engineering*, Vol. 10 No. 2, pp. 91-96, doi: [10.1007/bf02823926](https://doi.org/10.1007/bf02823926).
- Kobayashi, K., Do, M. and Han, D. (2010), "Estimation of Markovian transition probabilities for pavement deterioration forecasting", *KSCE Journal of Civil Engineering*, Vol. 14 No. 3, pp. 343-351, doi: [10.1007/s12205-010-0343-x](https://doi.org/10.1007/s12205-010-0343-x).
- Kobayashi, K., Kaito, K. and Lethanh, N. (2012), "A statistical deterioration forecasting method using hidden Markov model for infrastructure management", *Transportation Research Part B: Methodological*, Vol. 46 No. 4, pp. 544-561, doi: [10.1016/j.trb.2011.11.008](https://doi.org/10.1016/j.trb.2011.11.008).

- Li, Z., Wang, T., Ge, W., Wei, D. and Li, H. (2019), "Risk analysis of earth-rock dam breach based on dynamic Bayesian network", *Water (Switzerland)*, Vol. 11 No. 11, doi: [10.3390/w11112305](https://doi.org/10.3390/w11112305).
- Mcdaniel, R. and Shah, A. (2003), "Asphalt additives to control rutting and cracking", *Joint Transportation Research Program, In cooperation with Federal Highway Administration and the Indiana Department of Transportation*, available at: <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1495&context=jtrp>
- Mills, L.N.O., Attoh-Okine, N.O. and McNeil, S. (2012), "Developing pavement performance models for Delaware", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2304 No. 1, pp. 97-103, doi: [10.3141/2304-11](https://doi.org/10.3141/2304-11).
- Myung, I.J. (2001), "Computational approaches to model evaluation", *International Encyclopedia of the Social and Behavioral Sciences*, pp. 2453-2457, doi: [10.1016/B0-08-043076-7/00589-1](https://doi.org/10.1016/B0-08-043076-7/00589-1).
- Neapolitan, R. and Jiang, X. (2010), "Decision analysis fundamentals", *Probabilistic Methods for Financial and Marketing Informatics*, Elsevier, United States of America, pp. 177-228, available at: [https://books.google.ae/books?hl=en&lr=&id=MyU8-Jw8tZUC&oi=fnd&pg=PP1&dq=Neapolitan,+R.+and+Jiang,+X.+\(2007\),+%E2%80%9CDecision+analysis+fundamentals%E2%80%9D,+Probabilistic+Methods+for+Financial+and+Marketing+Informatics&ots=NUO5TQLFLU&sig=3xHhZUETD2jV9zof5fnczGk-k7A&redir\\_esc=y#v=onepage&q&f=false](https://books.google.ae/books?hl=en&lr=&id=MyU8-Jw8tZUC&oi=fnd&pg=PP1&dq=Neapolitan,+R.+and+Jiang,+X.+(2007),+%E2%80%9CDecision+analysis+fundamentals%E2%80%9D,+Probabilistic+Methods+for+Financial+and+Marketing+Informatics&ots=NUO5TQLFLU&sig=3xHhZUETD2jV9zof5fnczGk-k7A&redir_esc=y#v=onepage&q&f=false)
- Osorio-Lird, A., Chamorro, A., Videla, C., Tighe, S. and Torres-Machi, C. (2018), "Application of Markov chains and Monte Carlo simulations for developing pavement performance models for urban network management", *Structure and Infrastructure Engineering*, Vol. 14 No. 9, pp. 1169-1181, doi: [10.1080/15732479.2017.1402064](https://doi.org/10.1080/15732479.2017.1402064).
- Pan, N.F., Ko, C.H., Yang, M.D. and Hsu, K.C. (2011), "Pavement performance prediction through fuzzy regression", *Expert Systems with Applications*, Vol. 38 No. 8, pp. 10010-10017, doi: [10.1016/j.eswa.2011.02.007](https://doi.org/10.1016/j.eswa.2011.02.007).
- Pantuso, A., Flintsch, G.W., Katicha, S.W. and Loprencipe, G. (2021), "Development of network-level pavement deterioration curves using the linear empirical Bayes approach", *International Journal of Pavement Engineering*, Vol. 22 No. 6, pp. 780-793, doi: [10.1080/10298436.2019.1646912](https://doi.org/10.1080/10298436.2019.1646912).
- Piryonesi, S.M. and El-Diraby, T.E. (2020), "Data analytics in asset management: cost-effective prediction of the pavement condition index", *Journal of Infrastructure Systems*, Vol. 26 No. 1, available at: <https://ascelibrary.org/doi/10.1061/%28ASCE%29IS.1943-555X.0000512>.
- Roberts, C.A. and Attoh-Okine, N.O. (1998), "A comparative analysis of two artificial neural networks using pavement performance prediction", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 13 No. 5, pp. 339-348, doi: [10.1111/0885-9507.00112](https://doi.org/10.1111/0885-9507.00112).
- Rose, S., Mathew, B.S., Isaac, K.P. and Abhaya, A.S. (2018), "Risk based probabilistic pavement deterioration prediction models for low volume roads", *International Journal of Pavement Engineering*, Vol. 19 No. 1, pp. 88-97, doi: [10.1080/10298436.2016.1162308](https://doi.org/10.1080/10298436.2016.1162308).
- Salman, B. and Gursoy, B. (2022), "Markov chain pavement deterioration prediction models for local street networks", *Built Environment Project and Asset Management*, Vol. 12 No. 6, doi: [10.1108/BEPAM-09-2021-0117](https://doi.org/10.1108/BEPAM-09-2021-0117).
- Sholevar, N., Golroo, A. and Esfahani, S.R. (2022), "Machine learning techniques for pavement condition evaluation", *Automation in Construction*, Vol. 136, p. 104190, doi: [10.1016/j.autcon.2022.104190](https://doi.org/10.1016/j.autcon.2022.104190).
- Singh, A.P., Sharma, A., Mishra, R., Wagle, M. and Sarkar, A.K. (2018), "Pavement condition assessment using soft computing techniques", *International Journal of Pavement Research and Technology*, Vol. 11 No. 6, pp. 564-581, doi: [10.1016/j.ijprt.2017.12.006](https://doi.org/10.1016/j.ijprt.2017.12.006).
- Timeanddate (n.d.), available at: [www.timeanddate.com/](http://www.timeanddate.com/) (accessed 20 June 2022).

- 
- Torres-Machi, C., Pellicer, E., Yepes, V. and Chamorro, A. (2017), "Towards a sustainable optimization of pavement maintenance programs under budgetary restrictions", *Journal of Cleaner Production*, Vol. 148, pp. 90-102, doi: [10.1016/j.jclepro.2017.01.100](https://doi.org/10.1016/j.jclepro.2017.01.100).
- Tosun, A., Bener, A.B. and Akbarinasaji, S. (2017), "A systematic literature review on the applications of Bayesian networks to predict software quality", *Software Quality Journal*, Vol. 25 No. 1, pp. 273-305, doi: [10.1007/s11219-015-9297-z](https://doi.org/10.1007/s11219-015-9297-z).
- Visintine, B.A., Hicks, R.G., Cheng, D. and Elkins, G.E. (2015), "Factors affecting the performance of pavement preservation treatments", *9th International Conference on Managing Pavement Assets*, p. 23.
- Vyas, V., Pratap Singh, A. and Anshuman (2022), "Modeling asphalt pavement condition using artificial neural networks", *Materials Today: Proceedings*, Vol. 62, doi: [10.1016/j.matpr.2022.05.050](https://doi.org/10.1016/j.matpr.2022.05.050).
- Xiao, F., Chen, X., Cheng, J., Yang, S. and Ma, Y. (2022), "Establishment of probabilistic prediction models for pavement deterioration based on Bayesian neural network", *International Journal of Pavement Engineering*, pp. 1-16, doi: [10.1080/10298436.2022.2076854](https://doi.org/10.1080/10298436.2022.2076854).
- Yang, J., Lu, J.J., Gunaratne, M. and Xiang, Q. (2003), "Forecasting overall pavement condition with neural networks: application on Florida highway network", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1853 No. 1, pp. 3-12, doi: [10.3141/1853-01](https://doi.org/10.3141/1853-01).
- Yepes, V., Torres-Machi, C., Chamorro, A. and Pellicer, E. (2016), "Optimal pavement maintenance programs based on a hybrid greedy randomized adaptive search procedure algorithm", *Journal of Civil Engineering and Management*, Vol. 22 No. 4, pp. 540-550.

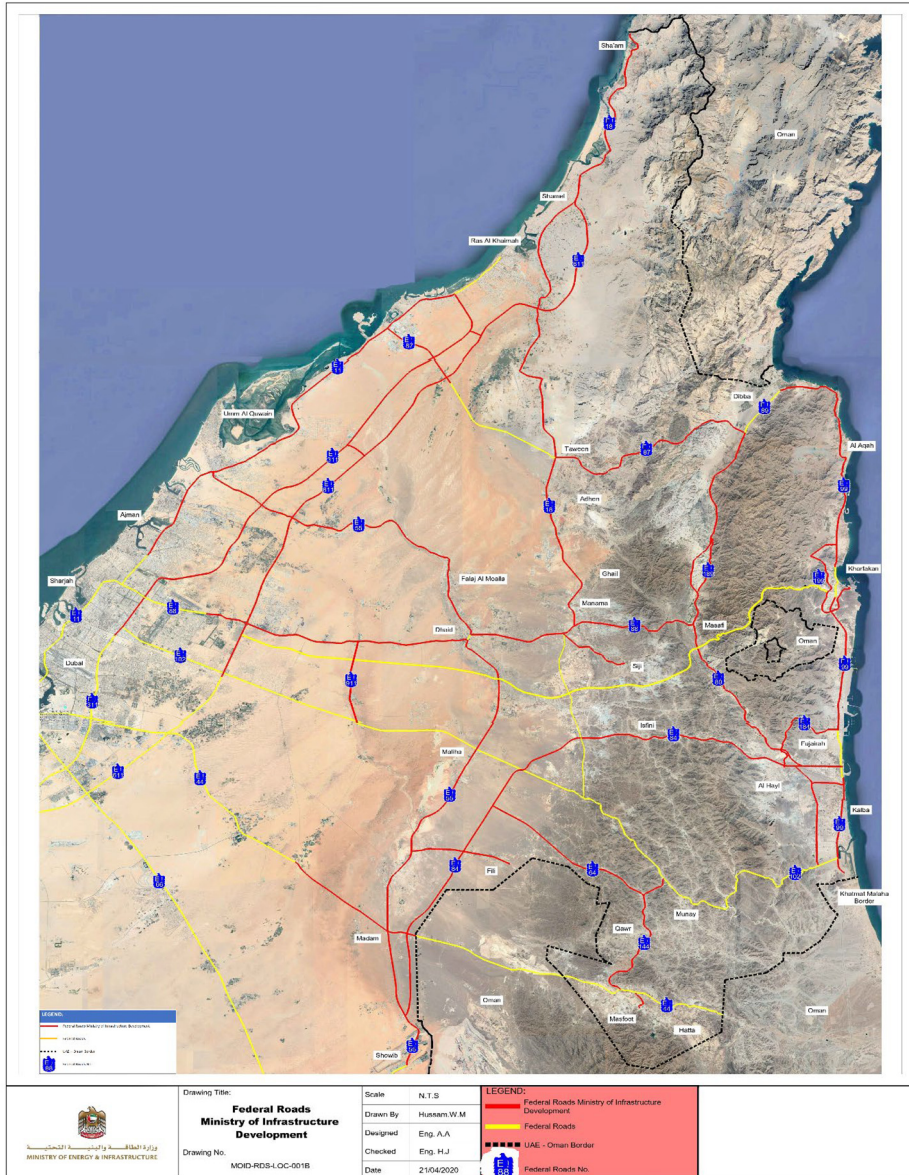


Figure A1.  
Satellite view of the  
study region

Source: Provided by the Ministry of energy and Infrastructure, MoEI and is reproduced from our previous work: <https://www.mdpi.com/2075-5309/12/7/1039>

Sl. No.	Road name	Stretch	Road type
1	E11	Ras Al Khaimah (RAK) to Sharjah	Arterial
2	E18_1	RAK Airport to Manama	Arterial
3	E18_2	Sha'am to Rak airport	Arterial
4	E18_3	Sha'am to Oman	Arterial
5	E55_1	Madam to Dhaid	Arterial
6	E55_2	Madam to Shiweb	Arterial
7	E55_3	Madam to Shiweb	Arterial
8	E84	Maliha to Fujairah	Freeway
9	E87	Tawyeen to Dibba	Arterial
10	E88_1	Sharjah to Dhaid	Expressway
11	E88_2	Dhaid to Masafi	Arterial
12	E89_1	Dibba to Ghona Bridge	Arterial
13	E89_2	Ghona Bridge to Masafi	Arterial
14	E89_3	Fujairah to Masafi	Arterial
15	E99_0	E87 to Dibba	Arterial
16	E99_1	Khor Fakkan to Dibba	Arterial
17	E99_2	Fujairah to Khor Fakkan	Arterial
18	E99_3	Oman to Fujairah	Arterial
19	E184	Al Hayl to Fujairah	Collector
20	E184_1	Al Hayl to E89	Collector
21	E199	Khor Fakkan Port to E99	Collector
22	E299	E99 to End Point	Collector
23	E311	National Paints to RAK	Freeway
24	E611_1	OB3 to Kalba Road	Freeway
25	E611_2	E611 (exit 95) to OB3	Freeway
26	E611_3	E611 (exit 95) to Al Shohada	Freeway
27	E611_B	E311 to E611 (exit 95)	Freeway
28	FA	E99 to Fujairah Airport	Collector
29	Fili link	Fili link	Local
30	Masfut link	Masfut link	Local
31	Ring road	Ring road	Local
32	Siji link	Siji link	Collector

**Source:** Created by the authors

**Table A1.**  
List of road networks  
selected for analysis

**Appendix 2. An example of data-preprocessing step**

The data pre-processing step adopted in the initial stage of data analysis is given below:

- (1) The initial point of measurement and the final point of measurement were decided to be rounded to the nearest possible multiples of 10. Spreadsheet functions “ROUND UP” and “ROUND DOWN” were used to convert the measurements in the multiples of 10. [Table A2](#) shows the starting point and final point of a road segment “A” before and after preprocessing:
  - Although majority of the points were measured in 10-meter intervals, there were many outliers:
  - Points not measured in 10-meter interval (e.g.: 100 – 120)
  - Points not measured in multiples of 10 (e.g.: 115 – 125)
  - A combination of both (e.g.: 112 – 147)

These outlier points were initially rounded into multiples of 10, if they were not in the multiples of 10 and then divided into 10-meter segments:

- (1) Then the next challenge was overlapping of the initial and final points. For example, see point A and B before and after pre-processing in [Table A3](#). When the points were split into 10-meter interval, both point A and B will have segments corresponding to 500–510. Hence, in this case the average values of 500–510 segments obtained corresponding to point A and B are taken.

**Table A2.**  
Representation of values converted into multiples of 10

Datapoint	Before pre-processing		After pre-processing	
	Initial distance	Final distance	Initial distance	Final distance
X	427	437	430	440
Y	512	522	510	520

**Table A3.**  
Overlapping of road segments

Datapoint	Before preprocessing		After preprocessing	
	Initial distance	Final distance	Initial distance	Final distance
A	399	510	400	510
B	500	601	500	600

**Corresponding author**

Hamad AlJassmi can be contacted at: [h.aljassmi@uaeu.ac.ae](mailto:h.aljassmi@uaeu.ac.ae)

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)