www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

HYBRID FEATURE SELECTION FRAMEWORK FOR PREDICTING BRIDGE DECK CONDITIONS

SUBMITTED: September 2022
REVISED: November 2022
PUBLISHED: November 2022
EDITOR: Bimal Kumar
DOI: 10.36680/j.itcon.2022.050

Abdelhady Omar, Ph.D. Student,
Concordia University, Department of Building, Civil and Environmental Engineering, Montreal, QC, Canada
Structural Engineering Department, Faculty of Engineering, Alexandria University, Alexandria, Egypt
abdelhady.omar@mail.concordia.ca

Osama Moselhi, Professor,
Director of Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM),
Concordia University, Department of Building, Civil and Environmental Engineering Montreal, QC, Canada
moselhi@encs.concordia.ca

SUMMARY: *Bridge decks' maintenance funding requirements are influenced by bridge decks' current and predicted future conditions. Additionally, the serviceability of bridges may be negatively impacted by the degradation of bridge decks. Bridge inspections require considerable effort, time, cost, and resources; besides, such inspections may introduce hazards and safety concerns. This paper introduces a data-driven hybrid feature selection framework for predicting bridge deck deterioration conditions and applying it to a bridge deck in Iowa State, USA. Firstly, the Boruta algorithm, stepwise regression, and multi-layer perceptron are employed to find the best subset of features that contribute to bridge deck deterioration. Then, four classification models were developed using the best feature subset of features, namely k-nearest neighbours, random forest, artificial neural networks, and deep neural networks. The hyperparameters of the models were optimized to get their best performance. The developed models showed comparable performance, and the random forest model outperformed the other models in prediction accuracy with fewer misclassifications. The developed models are thought to reduce field inspections and give insights into the most influential factors in bridge deck deterioration conditions.*

KEYWORDS: *Bridge Deck Deterioration, Feature Selection, Classification Models, Random Forest, Stepwise Regression, Multi-Layer perceptron.*

REFERENCE: *Abdelhady Omar, Osama Moselhi (2022). Hybrid feature selection framework for predicting bridge deck conditions. Journal of Information Technology in Construction (ITcon), Vol. 27, pg. 1028-1041, DOI: 10.36680/j.itcon.2022.050*

1. INTRODUCTION

Bridges are essential transportation systems that enable movement between different geographical locations. The deterioration of bridges is a critical issue that may result in a partial or whole collapse of such vital infrastructure components. Their failure may be catastrophic regarding human life as well as social, environmental, and economic impacts. The deterioration of bridge decks is the most frequent cause of bridges failure or being classified as structurally deficient/poor. Bridge decks provide a driving surface that facilitates the movement of people and cargo. Environmental factors, increasing traffic volumes, accidents, deferred maintenance, and aging are just a few contributing factors to this element deterioration (Omar and Moselhi 2022). Improving bridge deck maintenance, rehabilitation, and replacement work can effectively lower overall bridge expenditures as well as the bridge's life cycle costs, since maintenance actions of bridge decks account for 50% to 85% of total bridge costs (Gucunski et al. 2014). In this regard, timely condition monitoring of bridge decks is required to efficiently determine appropriate intervention actions and avoid unnecessary or expensive maintenance by taking preventative maintenance measures.

Visual examinations and non-destructive evaluation methods (e.g., digital imaging, ground-penetrating radar, and infrared thermography) are utilized to assess the current condition of bridge decks at uniform inspection intervals, e.g., 24 months (FHWA 2004; Omar and Moselhi 2022). However, due to the fact that there are a large number of bridges that require regular inspection, this process requires considerable effort, time, cost, and resources, besides being a source of substantial danger. In this context, developing data-driven methods to accurately predict bridge deck deterioration conditions is highly important to transportation agencies. A deterioration model of a bridge deck is a relationship between a bridge deck's condition and a vector of explanatory features, i.e., variables; features and variables will be used interchangeably hereinafter. These features reflect a collection of variables that impact the bridge deck's performance, such as age, load capacity, material properties, and deck geometry. The ability to anticipate deck conditions and the probability of failure is a major concern for transportation authorities (Liu and El-Gohary 2017).

This paper aims to develop a data-driven method to predict the conditions of reinforced concrete bridge decks, and apply it to a bridge deck in Iowa State, USA. The main goal can be broken down into the following objectives: (i) identifying the most influential features of the deterioration of bridge decks; and (ii) investigating the performance of four machine learning models to predict the condition of bridge decks. The developed methodology would help transportation agencies allocate resources and reduce field inspections.

2. LITERATURE REVIEW

Bridge decks' maintenance funding requirements are influenced by bridge decks' current and future conditions (Abed-Al-Rahim and Johnston 1995). Additionally, the serviceability of bridges may be negatively impacted by the degradation of bridge decks (Scott et al. 2003). Consequently, keeping an eye on the state of bridge decks is crucial to minimize any harm that can result from poor inspections and evaluations (Assaad and El-adaway 2020). Deterioration models should complement (rather than replace) field inspections to better plan for such inspections. On the other hand, data collected from the field inspections should be used to continuously improve these models (Omar et al. 2022).

Deterioration models can be classified into two main categories: deterministic models and stochastic models. Deterministic models are based on a mathematical and statistical formula for the relationship between the features influencing bridge deck deterioration and the bridge deck's condition. These models can be developed utilizing linear and non-linear regression, straight-line extrapolation, curve fitting, support vector machines, and Artificial Neural Networks (ANNs). On the other hand, stochastic models define the deterioration process as consisting of one or more random variables representing this process's uncertainty and randomness. Such random variables can be modelled using probability distribution functions (Agrawal et al. 2010). Regression analysis is considered the most common technique for deterministic models and Markov chain for stochastic models (Muñoz et al. 2016). Stochastic models can be divided into state- and time-based models (Mauch and Madanat 2001). In state-based models, such as Markov chains, the deterioration process is modeled through a probability of transition from one condition state to another in a discrete time. Whereas, in time-based models, the duration that a bridge element stays at a specific condition state is represented as a random variable utilizing Weibull-based probability density functions to define the deterioration process (Agrawal et al. 2010). Most recent research on deterioration models

of different bridge components (i.e., deck, superstructure, substructure, or the whole bridge) is highlighted below, and a summary is provided in Table 1.

TABLE 1: Summary of literature on deterioration modelling of bridges

Study	Focus & Methods	Data	Considered features
(Assaad and El-adaway 2020)	Predicting bridge deck conditions using ANNs and *k*-NNs & using feature selection and hyperparameters optimization to improve the models' performance.	NBI, Missouri	Age, ADT, maximum span length, bridge length, deck width, superstructure and substructure conditions, operating and inventory ratings, and structural evaluation
(Martinez et al. 2020)	Predicting Bridge Condition Index using *k*-NNs, DT, LR, ANNs, and DNNs.	Ontario, Canada	Age, number of lanes, material, structural type, number of spans, width, and length
(Nguyen and Dinh 2019)	Predicting bridge deck conditions using ANNs & developing deterioration curves.	NBI, Alabama	Age, ADT, design load, main structure design, approach span design, number of spans, truck ADT, and ADT growth rate
(Abdelkader et al. 2019)	Developing a defect-based deterioration model to forecast bridge deck conditions using a Markovian model.	Quebec, Canada	Age
(Ali et al. 2019)	Predicting conditions of different bridge elements using ANNs.	NBI, Missouri	Age, ADT, service type, material, structure type, length of maximum span, bridge length, deck width, year reconstructed, deck type, surface type, membrane type, deck protection, and truck ADT
(Zambon et al. 2017)	Predicting bridge deck conditions using different Markov chain models.	Portugal Infrastructures	Age
(Shim and Lee 2017)	Predicting bridge deck conditions using a Markovian deterioration model.	NBI, Wyoming	Age
(Muñoz et al. 2016)	Estimating the deterioration of different bridge components using Markov-chain models and regression analysis.	NBI, Nevada	Age and material
(Bu et al. 2015)	Developing bridge deterioration models incorporating time- and state-based models with Elman neural networks.	NBI, New York	Age, bridge component, material, ADT, and construction era
(Le and Andrews 2015)	Using Weibull distribution to model the deterioration of bridge components.	UK railways	Age, material, and intervention type
(Mašović and Hajdin 2014)	Predicting the deterioration of different elements using the Markov chain.	Serbia	Age
(Ranjith et al. 2013)	Predicting conditions of timber bridge elements using Markov chain models.	VicRoads, Australia	Age

Although the objectives of the highlighted studies in this area are quietly similar, i.e., predicting deterioration conditions of different bridge components, the methods, the data used, and the included features are different. Age and Average Daily Traffic (ADT) are among the most used features in models. It is worth noting that ADT calculations differ from those of Annual Average Daily Traffic (AADT); for more information about these

calculations, the reader may refer to FHWA (2018). The National Bridge Inventory (NBI) constitutes the primary source of data (FHWA 2022a). The NBI was compiled by Federal Highway Administration (FHWA) to serve as a database with information about all bridges and tunnels in the USA that have roads passing above or below them. Regarding employed methods, machine learning algorithms, including ANNs, Deep Neural Networks (DNNs), Decision Trees (DTs), *k*-Nearest Neighbours (*k*-NNs), and Linear Regression (LR), are thriving in recent studies by Ali et al. (2019); Assaad and El-adaway (2020); Martinez et al. (2020); Nguyen and Dinh (2019). In addition, different Markov-chain models with different calibration methods and Weibull distribution models were also applied by Abdelkader et al. (2019); Bu et al. (2015); Le and Andrews (2015); Mašović and Hajdin (2014); Muñoz et al. (2016); Ranjith et al. (2013); Shim and Lee (2017); Zambon et al. (2017). However, a few studies have considered feature selection or clustering to investigate the most influential factors on bridge deterioration, such as Assaad and El-adaway (2020).

3. METHODOLOGY

An overview of the developed 6-step methodology, along with the algorithms utilized is depicted in Fig. 1. The details of each step will be described subsequently.

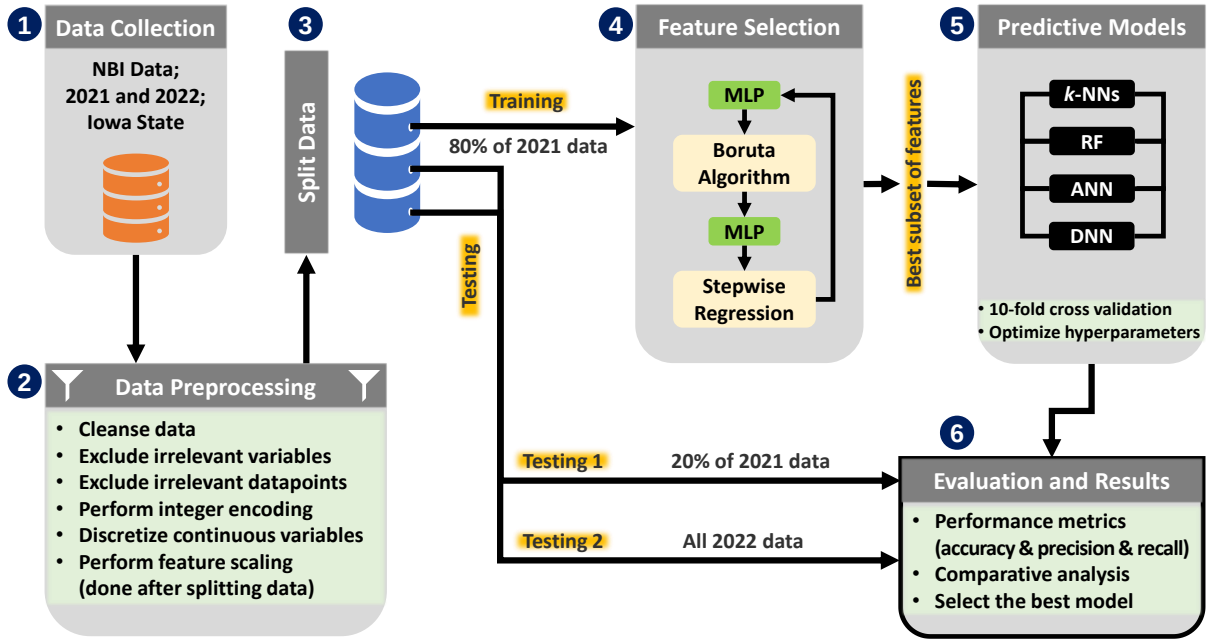

FIG. 1: Research Methodology

3.1 Data Collection

The two datasets used in this study were retrieved from NBI for Iowa State. Iowa is ranked the seventh in the number of bridges in the USA; however, it has the highest number of structurally deficient/poor bridges. The average bridge age of Iowa's primary highway system is 41 years approaching the intended lifespan of 50 years (Iowa DOT, 2022). Iowa Department of Transportation (DOT) uses NBI data to develop deterioration models to secure future funding needs for their bridges' replacement, rehabilitation, and repair (Iowa DOT, 2022). The retrieved datasets include bridge condition information for the years 2021 (FHWA 2021) and 2022 (FHWA 2022b). The Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges (FHWA 1995) describes in detail all data items (variables) in these datasets. In order to maintain consistency among bridge inspectors, FHWA provided guidelines for rating and coding conditions of different bridge components, i.e., deck, superstructure, and substructure (FHWA 1995). These condition ratings are utilized to describe the existing, in-place bridge as compared to the as-built condition using integer values from "0" (Failed Condition) to "9" (Excellent Condition), and "N" (Not Applicable) (FHWA 1995). As such, the deterioration of bridge deck conditions is a classification problem since conditions are provided as integers, i.e., categories.

3.2 Data Preprocessing

The following preprocessing steps were performed on the 2021 and 2022 datasets. The 2021 dataset consists mainly of 123 features, including 23,870 data points. Irrelevant variables to this study, such as County Code, Owner, and Culverts Condition Ratings, were excluded. For deck condition ratings, the code "N" means that this data point belongs to culverts or other structures without decks; so, their related data points (4,735 data points: 19.84%) were removed. In addition, the deck conditions rated as: "0" (Failed Condition; 253 data points: 1.06%), "1" (Imminent Failure Condition; two data points: 0.01%), or "2" (Critical Condition; two data points: 0.01%) means that these bridges are in a severe condition and need reconstruction rather than inspection. So, their related data points were dropped from the dataset. Data points with any of the following criteria were also filtered out: (i) missing values; (ii) bridges that had been reconstructed; and (iii) extreme values, i.e., "199": for detours of length 199 kilometers or more; "99": to indicate a significant variation in skews of substructure units; "99.9": when the restriction is 100 meters or greater for total horizontal clearance for the inventory route; "99.9": for operating rating and inventory rating for a structure under sufficient fill; and "99.9": for truck ADT greater than or equal 99.9% (FHWA 1995). Eventually, only reinforced concrete bridges with cast-in-place reinforced concrete decks were considered for the scope of this study. The same steps were also applied to the 2022 dataset. Accordingly, the datasets ended up with: 5,138 data points for 2021; 5,212 data points for 2022; and 42 variables. These variables are classified into eight main categories, as depicted in Fig. 2. It must be noted that the variable ID is the same as the data item number in the Recording and Coding Guide (FHWA 1995) to facilitate the tracking of variables.

The original variables have three data types; however, in this study, they are expressed in numeric-integer codes according to the Recording and Coding Guide (FHWA 1995). Categoric data were transformed into numeric-integer/ordinal values since most machine learning algorithms cannot handle such data representation (Garg 2022). Although dummy coding or one-hot-encoding can be used to transform categoric data, it considerably increases the dimensionality of the problem. Due to the fact that different numeric-real variables are measured on different scales with different ranges, they were discretized into numeric-integer values. Discretization (i.e., binning or bucketing) helps to minimize the effect of errors due to minor observations and outliers, and reduce the risk of overfitting. It can be unsupervised or supervised. Unsupervised discretization does not account for the label, e.g., discretizing a continuous variable into bins with equal widths or frequencies. In contrast, supervised discretization considers the label, e.g., discretizing a variable based on the highest information gain (Datacadamia 2022; Kotsiantis and Kanellopoulos 2006). This paper uses discretizing into bins with equal frequencies, i.e., an equal % of data points in each interval; the advantage is that this method takes the distribution of the variable into account and can be generalized to a new dataset regardless of the distribution of the label itself.

3.3 Split Data

The 2021 dataset was divided randomly into two sets: a training set with 4,110 records representing 80% of the dataset and a testing set (Testing 1) of 1,028 records representing 20% of the dataset. However, all 2022 data were used as a testing set (Testing 2) to investigate the generalization capabilities of the developed models. Then, feature scaling was applied to each set separately using Eq. 1 for min-max normalization.

$$\boldsymbol{Z_i = \frac{X_i - x_{min}}{x_{max} - x_{min}}} \qquad \text{Eq. 1}$$

Where: Z_i is the scaled value of the ith original value of the independent variable X; x_{min} and x_{max} are the minimum and maximum values of the independent variable X, respectively. The new ranges will be [0,1]. It is worth pointing out that feature scaling helps machine learning algorithms converge much faster and improve their performance. In addition, distance-based algorithms, e.g., Euclidean distance in *k*-NNs, are much more sensitive to feature scaling (Singh et al. 2015).

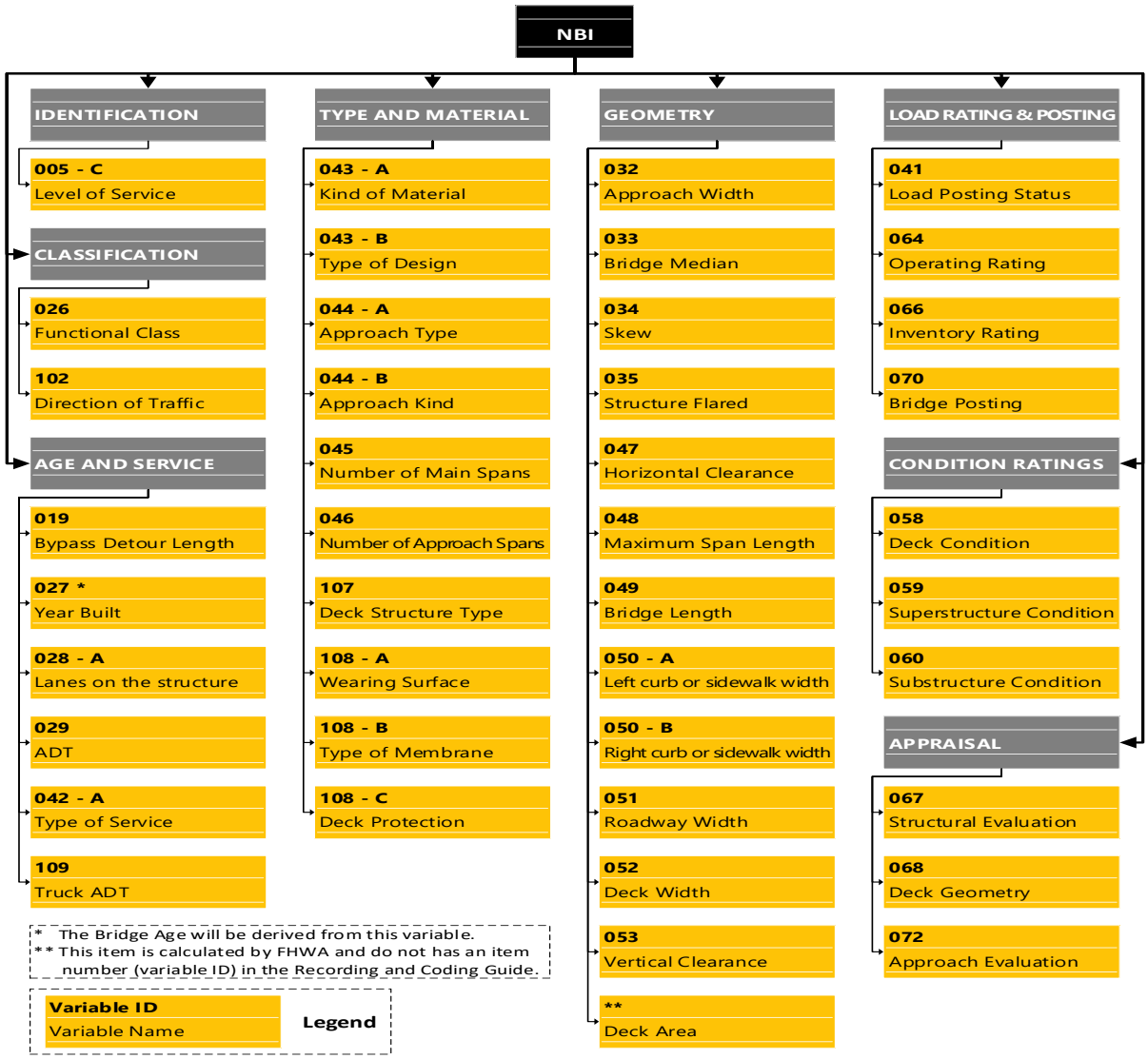

FIG. 2: Categorization of the variables included in this study

3.4 Feature Selection

Feature selection is the process of removing irrelevant and/or redundant features to find the best subset of data that contributes the most to the predicted label, i.e., deck condition in the present study. The importance of feature selection in machine learning lies in the fact that it helps: improve the model's prediction performance, reduce the dimensionality of the data; remove multi-collinearity (for regression models) and dependency (for algorithms like Naïve Bayes); reduce computational time, complexity and memory requirements; reduce the risk of overfitting; produce more interpretable predictive models; and, most importantly, provide better insights and understanding of the most influential features affecting deck deterioration conditions (Assaad and El-adaway 2020; Ebrahimi et al. 2022; Jain et al. 2014; Nik-Bakht 2021; Solorio-Fernández et al. 2020). Feature selection or selecting significant parameters was employed in different areas, including: selecting the best subset of features for predicting deterioration conditions in NBI data (Althaqafi 2021; Assaad and El-adaway 2020); finding significant parameters impacting construction labour productivity (Ebrahimi et al. 2022; Moselhi and Khan 2012); identification of significant impact factors affecting process times at workstations in modular construction (Bhatia et al. 2022); and determining the significant design parameters in modular construction workplace that contribute most to ergonomic risk scores (Zaalouk and Han 2021).

Feature selection methods can be clustered into two main categories: filter methods and wrapper methods. It is also worth pointing out that a hybrid of these methods can be utilized (Ebrahimi et al. 2022). Filter methods include: t-test, correlation analysis, chi-square test, and principal component analysis; while wrapper methods include: stepwise regression, forward selection, and backward elimination. The idea behind filter methods is to select top-ranked features using a specific criterion; a threshold is utilized to discriminate between selected and discarded features. Filter methods are applied before training predictive models relying only upon the statistical characteristics of the training set regardless of the algorithm applied. Although filter methods (i.e., ranking methods) are computationally light, they ignore predictive algorithms, and the selected subset may not be optimal for specific learning algorithms. In addition, for example, correlation analysis and chi-square test may produce a subset of features that are highly correlated with each presenting the risk of multi-collinearity in the dataset, besides needing a threshold which is, sometimes, difficult to be determined (Chandrashekar and Sahin 2014; Solorio-Fernández et al. 2020; Tsai 2009). On the other hand, wrapper methods are based on iteratively selecting those features that improve the performance of a specific algorithm. Although they are computationally intensive compared to ranking methods, they produce a subset of features that leads to the highest performance of the algorithm. The optimal subset, however, may be different from one algorithm to another. The risk of overfitting may be present in the wrapper method, necessitating careful training and testing considerations (Chandrashekar and Sahin 2014; Solorio-Fernández et al. 2020; Tsai 2009).

In the present study, three-iterative algorithms are utilized: the Boruta algorithm, stepwise regression, and Multi-Layer Perceptron (MLP). The Boruta is not a stand-alone algorithm, but rather it has an embedded Random Forest (RF) classifier. The algorithm goes through a specified number of iterations. In each iteration, features start to be classified as important or unimportant until all features are classified, or the set number of iterations is reached. This method is demonstrated to be effective in finding the best subset of features for developing high-performance forecasting models. It can perform computationally fast, even when dealing with a large number of variables. In addition, unlike filter methods, it can process features with complex and non-linear relationships. For the detailed steps involved in this method, the reader may refer to Andrew (2021); Cao et al. (2018); Kursa and Rudnicki (2010). In stepwise regression, forward selection and backward elimination criteria are used to assess which variables should be included in the regression equation. If a variable meets the statistical requirements, it is input one at a time; however, if the variable no longer meaningfully contributes to the regression model, it may be eliminated at any stage (Ho 2013). MLP was utilized as the baseline model to initially evaluate selected features against the classification problem, and find the relative weight of each feature. It is a type of feed-forward neural network. MLP may resolve complex issues that are not linearly separable. Prediction, classification, and Pattern recognition are among the main applications of MLP (Abirami and Chitra 2020; Menzies et al. 2015).

3.5 Predictive Models

Four predictive algorithms were employed in this study: *k*-NNs, RF, ANN, and DNN. The *k*-NNs algorithm is the simplest form of machine learning. It classifies a new unlabelled data point based on the majority class of its *k*-closest data points or "neighbours" in the data space (Kramer 2013). In this connection, two hyperparameters are needed for this algorithm to work: the value of *k* (number of closest neighbours), and a measure to determine those closest neighbours, i.e., a measure of proximity. Previous studies found that the best performance of *k*-NNs algorithm can be obtained at *k* values from 5 to 11 (Ashari et al. 2013; Batista and Silva 2009). In addition, among different measures of proximity (e.g., Euclidean Distance, Cosine Similarity, and Jaccard Similarity), Euclidean Distance is the most commonly used in the case of numeric predictors (Melhem and Cheng 2003). RF algorithm is an ensemble machine learning technique whose outcome is based on the majority voting of multiple decision tree models (Breiman 1996; Cutler et al. 2007; Kim et al. 2022). RF is very tolerant to outliers and noise, unlikely to overfitting, and has a good prediction accuracy (Sun et al. 2020). RF model involves many hyperparameters that should be fine-tuned to get the best performance. These hyperparameters include: the sample segmentation criterion or impurity measure, the number of decision trees, and the maximal depth of the decision tree (Koehrsen 2018; Sun et al. 2020). ANNs are inspired by biological networks consisting of an input layer, i.e., the predictors, hidden layers, and the output layer, i.e., the label (Moselhi et al. 1991). The number of neurons in the hidden layer, training cycles, and learning rate are among the hyperparameters that should be optimized for ANNs. Deep architectures like DNNs are composed of ANNs with many hidden layers. The number of hidden layers, the number of neurons in each layer, activation function (Rectifier activation function is the most popular), and deep learning epochs should be considered for its fine-tuning (Radhakrishnan 2017; Yoo 2019).

The performance of a machine learning algorithm model is significantly influenced by hyperparameters selection. The process of hyperparameter optimization or fine-tuning involves choosing the best hyperparameter values that give the highest performance (Koehrsen 2018; Sun et al. 2020). This study has utilized grid search and 10-fold cross validation to find the ideal hyperparameter values. The optimized hyperparameters, along with the search ranges and/or values, are presented in Table 2. The hidden layer size is denoted in the form $[h_1, h_2, \dots, h_m]$ to represent a network with (m) hidden layers that contain: h_1 neurons in the first hidden layer, h_2 neurons in the second layer, and h_m neurons in the last layer.

TABLE 2: Investigated hyperparameters of the models

Model	Hyperparameters and their investigated values
***k*-NNs**	k = 5−50 with step equal to 5.
RF	Number of trees = 10−100 with step equal to 10; impurity measure = Gini Index, Information Gain, Gain Ratio; maximal depth = 0−100 with step equal to10; apply/not apply pruning and pre-pruning.
ANN	Hidden layer size = 5, 8, 10, 15, 25; number of training cycles = 10−200 with step equal to 10; learning rate = 0.001−0.1 with 20 steps on logarithmic scale.
DNN	Different configurations of hidden layers = [50, 50], [25, 25], [25, 25, 25, 25], [10, 10, 10, 10]; deep learning epochs = 10−500 with 10 steps on linear scale.

3.6 Evaluation

In order to minimize the impact of a single sampling technique on model outcomes and reduce the risk of overfitting, the 10-fold cross validation with stratified random sampling was utilized to evaluate the performance of the four models and optimize their respective hyperparameters in the training dataset. The stratified random sampling was used as it allows to obtain a sample population that best represents the entire population of the dataset. In addition, in 10-fold cross validation, the dataset is divided into ten subsets or folds. The models are iteratively fitted ten times, each time trained on nine folds of the data and tested on the remaining fold. The highest average cross-validation accuracy was utilized as the objective function to optimize the models' hyperparameters. Then, the optimal hyperparameters were used to evaluate the models' performance using the two testing datasets. In addition, the performance of the models under the training and testing datasets was compared using the confusion matrix and other related performance metrics, i.e., accuracy, precision, and recall using Eq. 2, Eq. 3, and Eq. 4, respectively (Kotu and Deshpande 2019).

$$\boldsymbol{Accuracy} = (\boldsymbol{TP} + \boldsymbol{TN})/(\boldsymbol{TP} + \boldsymbol{FP} + \boldsymbol{TN} + \boldsymbol{FN}) \quad \text{Eq. 2}$$

$$\boldsymbol{Precision} = \boldsymbol{TP}/(\boldsymbol{TP} + \boldsymbol{FP}) \quad \text{Eq. 3}$$

$$\boldsymbol{Recall} = \boldsymbol{TP}/(\boldsymbol{TP} + \boldsymbol{FN}) \quad \text{Eq. 4}$$

Where, TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative. It is worth pointing out that the precision and recall were calculated as the average for all classes.

4. RESULTS AND DISCUSSIONS

The three-iterative algorithms for feature selection, referred to above, were applied to the training set. Firstly, the 39 predictors were fed into the Boruta algorithm. For the embedded ensemble RF, hyperparameters were optimized using random search and grid search cross validation. The optimized hyperparameters include: the number of trees in the forest, maximal depth of the tree, minimum number of samples required to split an internal node, minimum number of samples required to be at a leaf node, and bootstrap samples (Koehrsen 2018; Sun et al. 2020). It took the algorithm 52 iterations to converge and resulted in 10 important features. These features were then fed into a stepwise regression to check the potential for excluding any other redundant and multi-collinearity. Two features were excluded in this step as they did not contribute to the regression model. The remaining features were included

in the regression model. In addition, they have a Variance Inflation Factor (VIF) of [1.04, 9.82]. As such, small values of VIF (i.e., VIF < 10) ensure absence of multi-collinear variables (Ho 2013). The final regression model showed a high coefficient of determination, R^2, of 0.94.

The MLP was used as a baseline model to evaluate the prediction capabilities of different subsets of features. It was applied before and after each application of the Boruta algorithm and stepwise regression to test the resultant subset of features. The classification accuracy was improved from 91.60%/training and 89.10%/testing using all features to 92.20%/training and 93.60%/testing using the last eight features. The Receiver Operating Characteristic (ROC) curve (Tripepi et al. 2009), as well as the Area Under the Curve (AUC) values for each deck condition class, are depicted in Fig. 3. The AUC values are greater than or equal to 0.974 indicating improved predictive capabilities for MLP using the best subset of features. In addition, the relative importance of features derived from the MLP is shown in Fig. 4. Superstructure condition and structural evaluation are the most influential factors in predicting bridge deck deterioration conditions.

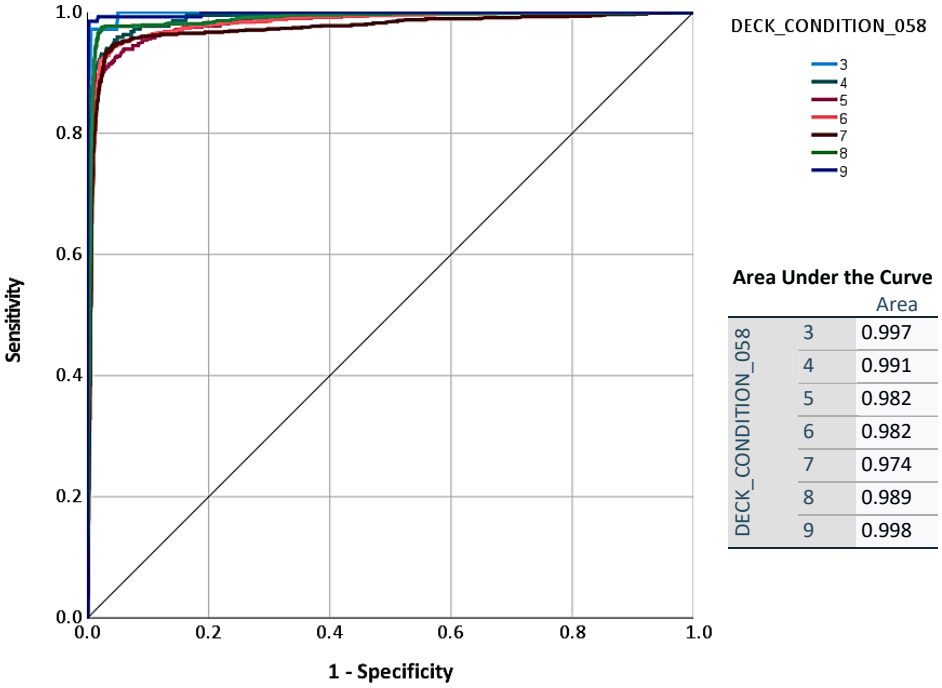

FIG. 3: ROC as well as AUC of the final iteration of MLP

The best subset of features that resulted from the previous phase was fed into the four models of *k*-NNs, RF, ANN, and DNN. In order to get the optimal (or near-optimal) hyperparameters that give the highest possible model performance, each model was then fine-tuned using the training set. These optimal hyperparameters are presented in Table 3. It is worth pointing out that the best performance, for the RF and ANN models, were obtained using different combinations of hyperparameters. For instance, for RF Model, using a combination of: number of trees = 70; impurity measure = Gain Ratio; and maximal depth = 60, produced the best performance. Likewise, changing the number of cycles for ANN to 120 or 130 with a learning rate of 0.06 yielded the best outcomes.

The results of training and testing the four models are presented in Table 4. Generally, RF, ANN, and DNN models showed comparable good performance in terms of accuracy, precision, and recall, whereas the *k*-NNs model had substantially less performance in training and testing. However, RF outperformed all other models in terms of: training accuracy and precision (92.87% and 89.96, respectively); testing accuracies (93.78% and 93.82%); testing precisions (92.15% and 91.27%); and testing recalls (94.25% and 92.93%). Concerning training recall: RF showed a little less recall than DNN by about 0.51%. In addition, as shown, the RF model showed better performance metrics in testing than in training which ensures that the model is not overfitted and possesses good generalization capabilities. Better precisions and recalls for RF also assure that the model provides better predictions with fewer misclassifications. Another point worth mentioning is that the standard deviations for training cross validation are: $\pm$1.60%, $\pm$0.94%, $\pm$1.09%, and $\pm$1.71% for *k*-NNs, RF, ANN, and DNN, respectively, indicating that the RF model is the most stable.

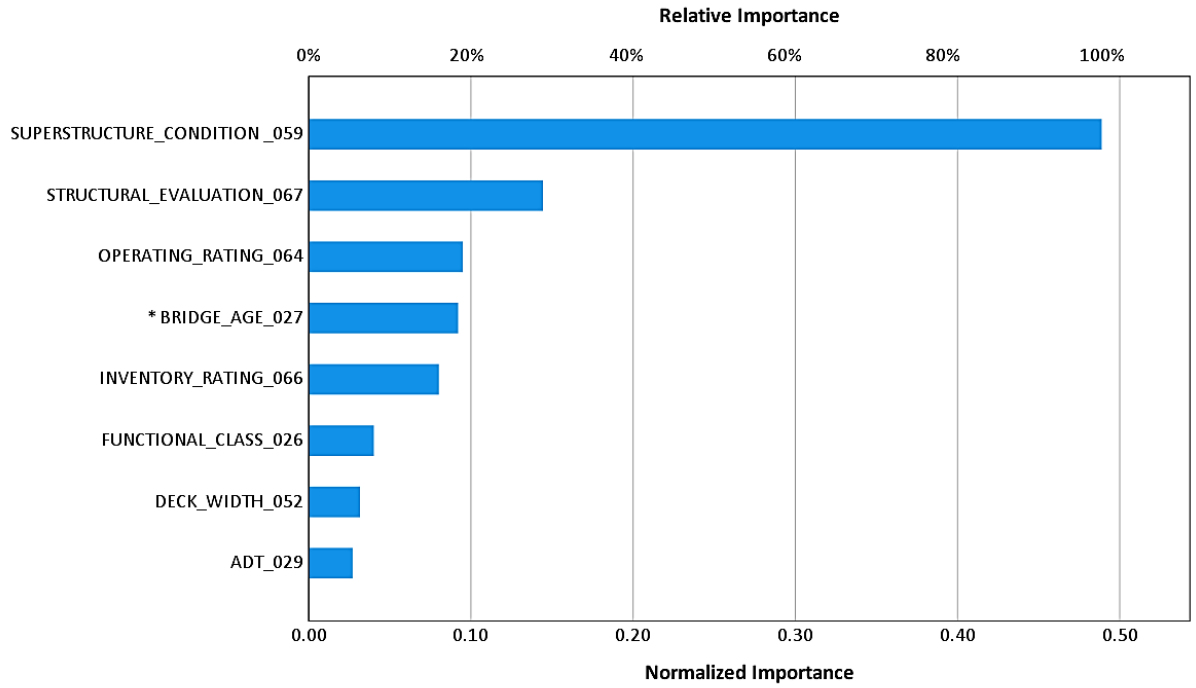

FIG. 4: Feature importance derived from MLP

TABLE 3: Optimal fine-tuned hyperparameters of the models

Model	**Optimal fine-tuned hyperparameters**
***k*-NNs**	$k = 5$; proximity measure = Euclidean Distance
RF	Number of trees = 40; impurity measure = Gini Index; maximal depth = 10; apply pruning and pre-pruning
ANN	Hidden layer size = [8]; number of training cycles = 200; learning rate = 0.05
DNN	Activation function = Rectifier; hidden layers size = [10, 10, 10, 10]; deep learning epochs = 157

TABLE 4: Cross-validation training results as well as testing results of the models

Model	**Training** (using 80% of 2021 data)			**Testing 1** (using 20% of 2021 data)			**Testing 2** (using all 2022 data)		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
***k*-NNs**	80.88%	78.62%	72.69%	80.54%	75.82%	69.05%	84.61%	82.90%	77.22%
RF	***92.87%***	***89.96%***	91.13%	***93.87%***	***92.15%***	***94.25%***	***93.82%***	***91.27%***	***92.93%***
ANN	92.19%	89.02%	87.82%	93.68%	91.97%	93.58%	93.48%	90.95%	92.56%
DNN	92.68%	89.15%	***91.64%***	91.54%	85.23%	88.39%	93.40%	90.43%	92.24%

Many studies have introduced diverse machine learning models in this domain, i.e., predicting bridge deck deterioration conditions. Unlike those studies, this paper contributes to the body of knowledge on three fronts. First, it introduces a hybrid framework to systematically find the best subset of features, associated with their relative weights, for predicting bridge deck deterioration conditions. The framework incorporates three-iterative algorithms, selected based on several iterations and trials of other algorithms (e.g., correlation analysis, Gini index, entropy theory, forward selection, backward elimination, and principal component analysis), which ensures removal of redundancy and multi-collinearity in the final subset. In addition, it can be generalized for NBI datasets to give better insights into the most influential factors on the probability of failure and deterioration of bridge decks. Second, it presents four newly developed classification models with comparable good performance. The developed models are a combination of simple models (e.g., *k*-NNs), ensemble machine learning models (e.g., RF), and computationally challenging models (e.g., DNN). The results showed the superiority of the RF model, which is rarely used before in this domain despite its ease of application compared to neural networks. Third, it provides a detailed road map for preparing NBI datasets and optimizing the models' hyperparameters, which are crucial steps in machine learning applications. The used dataset, along with the developed models and code are available upon request to facilitate future application and/or evaluation of the proposed methodology.

5. CONCLUSIONS AND FUTURE WORK

This paper introduced a hybrid feature selection framework for predicting bridge deck deterioration conditions. The Boruta algorithm, stepwise regression, and multi-layer perceptron were employed to find the best subset of features that contribute to bridge deck deterioration. When applied to Iowa State, USA, 2021 and 2022 datasets, only eight out of 38 features were selected. Superstructure condition, structural evaluation, operating rating, and bridge age were the most influential factors, followed by inventory rating, functional classification, deck width, and average daily traffic. Upon using these eight features as input to the four classification models of *k*-nearest neighbours, random forest, artificial neural network, and deep neural networks for prediction of the bridge deck conditions, the random forest model yielded the best prediction. It provided an accuracy of about 92.87% in training as well as 93.87% and 93.82 in testing. Future studies should investigate using the current features and models to develop maintenance plans and budget allocation for bridge decks.

REFERENCES

Abdelkader E. M., Zayed T., and Marzouk M. (2019). A computerized hybrid Bayesian-based approach for modelling the deterioration of concrete bridge decks. *Structure and Infrastructure Engineering*, Vol. 15, No. 9, 1178-1199. Doi: 10.1080/15732479.2019.1619782.

Abed-Al-Rahim I. J., and Johnston D. W. (1995). Bridge element deterioration rates. *Transportation Research Record*, No. 1490. https://trid.trb.org/view/452547.

Abirami S., and Chitra P. (2020). Chapter Fourteen - Energy-efficient edge based real-time healthcare support system. In Pethuru Raj & Preetha Evangeline (Eds.), *Advances in Computers*, Vol. 117, pp. 339-368: Elsevier. Doi: 10.1016/bs.adcom.2019.09.007.

Agrawal A. K., Kawaguchi A., and Chen Z. (2010). Deterioration Rates of Typical Bridge Elements in New York. *Journal of Bridge Engineering*, Vol. 15, No. 4, 419-429. Doi: 10.1061/(ASCE)BE.1943-5592.0000123.

Ali G., Elsayegh A., Assaad R., El-Adaway I. H., and Abotaleb I. S. (2019). *Artificial Neural Network Model for Bridge Deterioration and Assessment.* In Proceedings of the Canadian Society for Civil Engineering: Montreal, QC, Canada. https://scholarsmine.mst.edu/civarc_enveng_facwork/1812/.

Althaqafi E. (2021). *Developing Bridge Deterioration Model Using Artificial Neural Network and Markov Chain.* Ph.D. Thesis, The University of Toledo. Retrieved from ProQuest Dissertations & Theses Global database. https://www-proquest-com.lib-ezproxy.concordia.ca/docview/2645480104?accountid=10246.

Andrew D. (2021). Feature Selection with Boruta in Python. Retrieved August 02, 2022, from https://medium.com/geekculture/feature-selection-with-boruta-in-python-676e3877e596.

Ashari A., Paryudi I., and Tjoa A. M. (2013). Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 4, No. 11. Doi: 10.14569/IJACSA.2013.041105.

Assaad R., and El-adaway I. H. (2020). Bridge infrastructure asset management system: Comparative computational machine learning approach for evaluating and predicting deck deterioration conditions. *Journal of infrastructure systems*, Vol. 26, No. 3, 04020032. Doi: 10.1061/(ASCE)IS.1943-555X.0000572.

Batista G., and Silva D. F. (2009). *How k-nearest neighbor parameters affect its performance.* In Proceedings of the Argentine symposium on artificial intelligence. http://sites.labic.icmc.usp.br/pub/gbatista/BatistaASAI09.pdf.

Bhatia A. P. S., Han S., and Moselhi O. (2022). A simulation-based statistical method for planning modular construction manufacturing. *Journal of Information Technology in Construction (ITcon)*, Vol. 27, 130-144. Doi: 10.36680/j.itcon.2022.007.

Breiman L. (1996). Bagging predictors. *Machine learning*, Vol. 24, No. 2, 123-140. Doi: 10.1007/BF00058655.

Bu G. P., Lee J. H., Guan H., Loo Y. C., and Blumenstein M. (2015). Prediction of Long-Term Bridge Performance: Integrated Deterioration Approach with Case Studies. *Journal of Performance of Constructed Facilities*, Vol. 29, No. 3, 9. Doi: 10.1061/(ASCE)CF.1943-5509.0000591.

Cao Y., Ashuri B., and Baek M. (2018). Prediction of Unit Price Bids of Resurfacing Highway Projects through Ensemble Machine Learning. *Journal of Computing in Civil Engineering*, Vol. 32, No. 5, 04018043. Doi: 10.1061/(ASCE)CP.1943-5487.0000788.

Chandrashekar G., and Sahin F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, Vol. 40, No. 1, 16-28. Doi: 10.1016/j.compeleceng.2013.11.024.

Cutler D. R., Edwards Jr T. C., Beard K. H., Cutler A., Hess K. T., Gibson J., and Lawler J. J. (2007). Random forests for classification in ecology. *Ecology*, Vol. 88, No. 11, 2783-2792. Doi: 10.1890/07-0539.1.

Datacadamia. (2022). Discretizing. Retrieved August 12, 2022, from https://datacadamia.com/data_mining/discretization.

Ebrahimi S., Kazerooni M., Sumati V., and Fayek A. R. (2022). Predictive model for construction labour productivity using hybrid feature selection and principal component analysis. *Canadian Journal of Civil Engineering*. Doi: 10.1139/cjce-2021-0248.

FHWA. (1995). *Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges*. Retrieved January 17, 2022, from U.S. Department of Transportation, Bridges & Structures, National Bridge Inventory: https://www.fhwa.dot.gov/bridge/mtguide.cfm

FHWA. (2004). *National Bridge Inspection Standards*. Retrieved October 09, 2021, from U.S. Department of Transportation, Bridges & Structures, National Bridge Inventory: https://www.fhwa.dot.gov/bridge/nbis.cfm

FHWA. (2018). *Traffic Data Computation Method*. Retrieved January 21, 2022, from U.S. Department of Transportation, Bridges & Structures, National Bridge Inventory:

https://www.fhwa.dot.gov/policyinformation/pubs/pl18027_traffic_data_pocket_guide.pdf

FHWA. (2021). *NBI ASCII files 2021*. Delimited Files, Iowa. Retrieved February 26, 2022, from U.S. Department of Transportation, Bridges & Structures, National Bridge Inventory:

https://www.fhwa.dot.gov/bridge/nbi/ascii2021.cfm

FHWA. (2022a). NBI ASCII files. Retrieved March 13, 2022, from https://www.fhwa.dot.gov/bridge/nbi/ascii.cfm.

FHWA. (2022b). *NBI ASCII files 2022*. Delimited Files, Iowa. Retrieved July 10, 2022, from U.S. Department of Transportation, Bridges & Structures, National Bridge Inventory:

https://www.fhwa.dot.gov/bridge/nbi/ascii.cfm

Garg S. (2022). How to Deal with Categorical Data for Machine Learning. Retrieved August 12, 2022, from https://www.kdnuggets.com/2021/05/deal-with-categorical-data-machine-learning.html.

Gucunski N., Nazarian S., Imani A., and Azari H. (2014). *Performance of NDT technologies in detection and characterization of reinforced concrete deck deterioration.* In Proceedings of the Geo-Congress 2014: Geo-characterization and Modeling for Sustainability. https://doi.org/10.1061/9780784413272.236.

Ho R. (2013). *Handbook of univariate and multivariate data analysis with IBM SPSS* (Second ed.): CRC press, Taylor & Francis Group.

https://books.google.ca/books?id=DXtFAQAAQBAJ&lpg=PP1&ots=5q1jV2Q2MR&dq=Ho%20R.%20(2013).

Iowa DOT Bridges and Structures Bureau. (2022). *Annual Bridge Report 2022*. Retrieved July 23, 2022, from https://iowadot.gov/bridge/Annual-Bridge-Report

Jain R. K., Damoulas T., and Kontokosta C. E. (2014). Towards Data-Driven Energy Consumption Forecasting of Multi-Family Residential Buildings: Feature Selection via The Lasso *Computing in Civil and Building Engineering (2014)*, pp. 1675-1682. Doi: 10.1061/9780784413616.208.

Kim J., Hatzis J. J., Klockow K., and Campbell P. A. (2022). Building Classification Using Random Forest to Develop a Geodatabase for Probabilistic Hazard Information. *Natural Hazards Review*, Vol. 23, No. 3, 04022014. Doi: 10.1061/(ASCE)NH.1527-6996.0000561.

Koehrsen W. (2018). Hyperparameter Tuning the Random Forest in Python. Retrieved August 14, 2022, from https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74.

Kotsiantis S., and Kanellopoulos D. (2006). Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering*, Vol. 32, No. 1, 47-58. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.109.3084&rep=rep1&type=pdf.

Kotu V., and Deshpande B. (2019). Chapter 8 - Model Evaluation. In Vijay Kotu & Bala Deshpande (Eds.), *Data Science (Second Edition)*, pp. 263-279: Morgan Kaufmann. Doi: 10.1016/B978-0-12-814761-0.00008-3.

Kramer O. (2013). K-Nearest Neighbors. In Oliver Kramer (Ed.), *Dimensionality Reduction with Unsupervised Nearest Neighbors*, pp. 13-23. Berlin, Heidelberg: Springer Berlin Heidelberg. Doi: 10.1007/978-3-642-38652-7_2.

Kursa M. B., and Rudnicki W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, Vol. 36, No. 11, 1 - 13. Doi: 10.18637/jss.v036.i11.

Le B., and Andrews J. (2015). Modelling railway bridge degradation based on historical maintenance data. *Safety and Reliability*, Vol. 35, No. 2, 32-55. Doi: 10.1080/09617353.2015.11691040.

Liu K., and El-Gohary N. (2017). Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports. *Automation in Construction*, Vol. 81, 313-327. Doi: 10.1016/j.autcon.2017.02.003.

Martinez P., Mohamed E., Mohsen O., and Mohamed Y. (2020). Comparative Study of Data Mining Models for Prediction of Bridge Future Conditions. *Journal of Performance of Constructed Facilities*, Vol. 34, No. 1, 9. Doi: 10.1061/(asce)cf.1943-5509.0001395.

Mašović S., and Hajdin R. (2014). Modelling of bridge elements deterioration for Serbian bridge inventory. *Structure and Infrastructure Engineering*, Vol. 10, No. 8, 976-987. Doi: 10.1080/15732479.2013.774426.

Mauch M., and Madanat S. (2001). Semiparametric hazard rate models of reinforced concrete bridge deck deterioration. *Journal of infrastructure systems*, Vol. 7, No. 2, 49-57. Doi: 10.1061/(ASCE)1076-0342(2001)7:2(49).

Melhem H. G., and Cheng Y. (2003). Prediction of Remaining Service Life of Bridge Decks Using Machine Learning. *Journal of Computing in Civil Engineering*, Vol. 17, No. 1, 1-9. Doi: 10.1061/(ASCE)0887-3801(2003)17:1(1).

Menzies T., Kocagüneli E., Minku L., Peters F., and Turhan B. (2015). Chapter 24 - Using Goals in Model-Based Reasoning. In Tim Menzies, Ekrem Kocagüneli, Leandro Minku, Fayola Peters, & Burak Turhan (Eds.), *Sharing Data and Models in Software Engineering*, pp. 321-353. Boston: Morgan Kaufmann. Doi: 10.1016/B978-0-12-417295-1.00024-2.

Moselhi O., Hegazy T., and Fazio P. (1991). Neural networks as tools in construction. *Journal of construction Engineering and Management*, Vol. 117, No. 4, 606-625. Doi: 10.1061/(ASCE)0733-9364(1991)117:4(606).

Moselhi O., and Khan Z. (2012). Significance ranking of parameters impacting construction labour productivity. *Construction Innovation*. Doi: 10.1108/14714171211244541.

Muñoz Y., Paz A., De La Fuente-Mella H., Farina J., and Sales G. (2016). Estimating bridge deterioration for small data sets using regression and Markov models. *International Journal of Urban and Civil Engineering (World Academy of Science, Engineering and Technology)*, Vol. 10, No. 5, 663-670.

Nguyen T. T., and Dinh K. (2019). Prediction of bridge deck condition rating based on artificial neural networks. *Journal of Science and Technology in Civil Engineering (STCE)-NUCE*, Vol. 13, No. 3, 15-25. Doi: 10.31814/stce.nuce2019-13(3)-02.

Nik-Bakht M. (2021). *Lecture notes on "Big Data Analytics for Smart City Infrastructure: CIVI691"*. Concordia University, Montreal, QC, Canada.

Omar A., Delnaz A., and Nik-Bakht M. (2022). *Comparative Analysis of Machine Learning Techniques for Predicting Water Main Failures in the City of Kitchener*. Research Project. Submitted Manuscript to the Canadian Journal of Civil Engineering (CJCE).

Omar A., and Moselhi O. (2022). Condition Monitoring of Reinforced Concrete Bridge Decks: Current Practices and Future Perspectives. *Current Trends in Civil & Structural Engineering - CTCSE*, Vol. 8, No. 4. https://irispublishers.com/ctcse/pdf/CTCSE.MS.ID.000695.pdf.

Radhakrishnan P. (2017). What are Hyperparameters ? and How to tune the Hyperparameters in a Deep Neural Network? Retrieved August 14, 2022, from https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a.

Ranjith S., Setunge S., Gravina R., and Venkatesan S. (2013). Deterioration prediction of timber bridge elements using the Markov chain. *Journal of Performance of Constructed Facilities*, Vol. 27, No. 3, 319-325. Doi: 10.1061/(ASCE)CF.1943-5509.0000311.

Scott M., Rezaizadeh A., Delahaza A., Santos C. G., Moore M., Graybeal B., and Washer G. (2003). A comparison of nondestructive evaluation methods for bridge deck assessment. *NDT & e International*, Vol. 36, No. 4, 245-255. Doi: 10.1016/S0963-8695(02)00061-0.

Shim H. S., and Lee S. H. (2017). Balanced allocation of bridge deck maintenance budget through multi-objective optimization. *KSCE Journal of Civil Engineering*, Vol. 21, No. 4, 1039-1046. Doi: 10.1007/s12205-016-0591-5.

Singh B. K., Verma K., and Thoke A. (2015). Investigations on impact of feature normalization techniques on classifier's performance in breast tumor classification. *International Journal of Computer Applications*, Vol. 116, No. 19. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.1851&rep=rep1&type=pdf.

Solorio-Fernández S., Carrasco-Ochoa J. A., and Martínez-Trinidad J. F. (2020). A review of unsupervised feature selection methods. *Artificial Intelligence Review*, Vol. 53, No. 2, 907-948. Doi: 10.1007/s10462-019-09682-y.

Sun D., Wen H., Wang D., and Xu J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, Vol. 362, 107201. Doi: 10.1016/j.geomorph.2020.107201.

Tripepi G., Jager K. J., Dekker F. W., and Zoccali C. (2009). Diagnostic methods 2: receiver operating characteristic (ROC) curves. *Kidney International*, Vol. 76, No. 3, 252-256. Doi: 10.1038/ki.2009.171.

Tsai C.-F. (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems*, Vol. 22, No. 2, 120-127. Doi: 10.1016/j.knosys.2008.08.002.

Yoo Y. (2019). Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches. *Knowledge-Based Systems*, Vol. 178, 74-83. Doi: 10.1016/j.knosys.2019.04.019.

Zaalouk A., and Han S. (2021). Parameterized Design Optimization Framework for Worker-Friendly Workplaces in Modular Construction. *Journal of construction Engineering and Management*, Vol. 147, No. 5, 04021030. doi: Doi:10.1061/(ASCE)CO.1943-7862.0002029.

Zambon I., Vidovic A., Strauss A., Matos J., and Amado J. (2017). Comparison of stochastic prediction models based on visual inspections of bridge decks. *Journal of Civil Engineering and Management*, Vol. 23, No. 5, 553-561. Doi: 10.3846/13923730.2017.1323795.