

RESEARCH

Open Access



Modeling retroreflectivity degradation of pavement markings across the US with advanced machine learning algorithms

Ipshit Ibne Idris¹, Momen Mousa² and Marwa Hassan^{1*}

Abstract

Retroreflectivity is the primary metric that controls the visibility of pavement markings during nighttime and in adverse weather conditions. Maintaining the minimum level of retroreflectivity as specified by Federal Highway Administration (FHWA) is crucial to ensure safety for motorists. The key objective of this study was to develop robust retroreflectivity prediction models that can be used by transportation agencies to reliably predict the retroreflectivity of their pavement markings utilizing the initially measured retroreflectivity and other key project conditions. A total of 49,632 transverse skip retroreflectivity measurements of seven types of marking materials were retrieved from the eight most recent test decks covered under the National Transportation Product Evaluation Program (NTPEP). Decision Tree (DT) and Artificial Neural Network (ANN) algorithms were considered for developing performance prediction models to estimate retroreflectivity at different prediction horizons for up to three years. The models were trained with randomly selected 80% data points and tested with the remaining 20% data points. Sequential ANN models exhibited better performance with the testing data than the sequential DT models. The training and testing R^2 ranges of the sequential ANN models were from 0.76 to 0.96 and 0.55 to 0.94, respectively, which were significantly higher than the R^2 range (0.14 to 0.75) from the regression models proposed in past studies. Initial or predicted retroreflectivity, snowfall, and traffic were found to be the most important inputs to model predictions.

Keywords Pavement marking, FHWA final rule, Retroreflectivity degradation modeling, Decision tree, Artificial neural network

Introduction

Pavement Markings are retroreflective longitudinal and transverse lines that are installed on the pavement surfaces to delineate their profile [1, 2]. The most commonly used marking materials in the U.S. are either non-durable (waterborne paint) or durable (thermoplastic, epoxy, polyurea, methyl methacrylate [MMA], and tape) [3]. Markings are a critical component of overall traffic

signalization as they define boundaries between moving vehicles and a well-maintained marking system enhances safety for motorists during daytime, nighttime, and in poor visibility conditions [4].

As per the Federal Highway Administration (FHWA), over 50% of all traffic fatalities occur at night while the majority of travel happens during the daytime [5]. The visibility of pavement markings is primarily dependent on their retroreflectivity [6]. Retroreflectivity is the property of markings that describes the phenomenon in which light originating from vehicle headlights illuminates the visible pavement marking surface and a substantial portion of it returns to the eye of the motorists [7]. Retroreflectivity, measured by the coefficient of retroreflected luminance (R_r) in millicandelas per square meter per

*Correspondence:

Marwa Hassan
marwa@lsu.edu

¹ Department of Construction Management, Louisiana State University, Baton Rouge, LA 70803, USA

² Department of Engineering Technology, Sam Houston State University, Huntsville, TX 77341, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

lux ($\text{mcd}/\text{m}^2/\text{lux}$), is provided through a partial embedment of transparent glass beads inside the markings [6]. An adequate level of R_L can reduce 50% of crashes during nighttime and 28% of fatalities in dark, rainy, or snowy weather conditions [8].

R_L degrades over time depending on the type of marking material, glass bead properties, climatic conditions, traffic loading, road surface type, and snow-plow frequency [9]. Recently, FHWA published a new rule that established national standards for minimum R_L levels for longitudinal marking lines on all roads depending on speed limits [10]. This final rule requires all state agencies to implement a method before 2027 to maintain the specified minimum R_L level and consider retroreflectivity in future restriping activities, highlighting the importance of regular retroreflectivity monitoring for the marking lines. However, instead of continuously monitoring and restriping when R_L drops below the minimum threshold, many transportation agencies restripe markings as per a fixed schedule or visual inspection [11]. These restriping strategies are not optimum as markings are often restriped before or after the end of their service life, resulting in misspending available funds or compromising the safety of the motorists, respectively. As such, modeling retroreflectivity degradation is critical to estimate the service life of the markings and to accordingly plan for future restriping activities.

Objective and scope

The objective of this study was to develop machine learning models that can be used by U.S. transportation agencies to predict, with superior accuracy, the retroreflectivity of pavement markings, over a period of 3 years, based on the initial measured retroreflectivity and other key project conditions. Two different machine learning algorithms, Decision Tree and Artificial Neural Network algorithms, were utilized to develop performance prediction models for seven types of pavement marking materials (waterborne paint, thermoplastic, preformed thermoplastic, permanent polymeric tape, epoxy,

polyurea, and MMA) located in three different U.S. climate zones (Southeast, Northeast, and Upper Midwest). The proposed models are expected to provide a scientific basis for transportation agencies in predicting the service life of pavement markings based on local conditions and only one initial retroreflectivity measurement, thereby eliminating the need for the costly monitoring of the retroreflectivity of marking products.

Background

This section provides a brief background about the NTPEP as it is the source of data used in this study. After that, this section presents the results of previous studies that modeled the retroreflectivity degradation of pavement markings to be used as a baseline for comparison to the results of this study. Eventually, this section is concluded with the shortcomings in the current state of knowledge and how these shortcomings are addressed in the developed models presented in this study.

Overview of the National Transportation Product Evaluation Program (NTPEP)

American Association of State Highway and Transportation Officials (AASHTO) implements a consensus-based work plan every year through National Transportation Product Evaluation Program (NTPEP) to evaluate the field performance of a variety of pavement marking products [12]. NTPEP selects test decks from across the U.S., representing various traffic and geographical conditions and installs marking products on those in transverse direction [13]. In a typical test deck, four marking lines are installed side-by-side on an asphalt or concrete surface from the inner side of the “edge” line to the far side of the “skip” line [12] (see Fig. 1). For each line, retroreflectivity measurements are collected from both the “skip” area (within the first 9 in. from the “skip” line, known as transverse-skip retroreflectivity, R_S) and “wheel” area (9 in. on both sides of the left wheel path, known as transverse-wheel retroreflectivity, R_W) with a handheld retroreflectometer [11]. These readings are

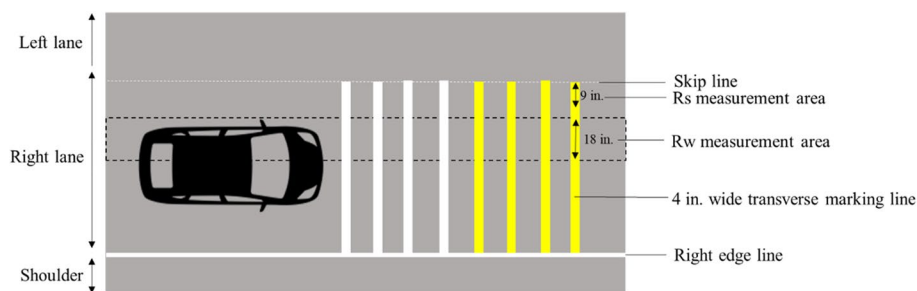


Fig. 1 Typical configuration of an NTPEP test deck and retroreflectivity measurement locations

collected at 12 different time intervals (after 0, 1, 2, 3, 11, 12, 15, 21, 24, 27, 33, and 36 months) for up to 3 years. Accelerated degradation of retroreflectivity occurs in the ‘wheel’ area due to continuous friction with vehicle tires and does not mimic the service life of an actual longitudinal marking stripe [11]. Traffic condition in “skip” area is more representative of the skip-line stripes in the field [14], and therefore, only R_S measurements were considered in this study for modeling.

Previous retroreflectivity degradation models

A literature review was conducted on past studies that proposed retroreflectivity degradation models to estimate the service life of pavement markings, as summarized in Table 1. Previous studies mostly adopted the parametric approach, more specifically, regression models to predict future retroreflectivity [9, 15–21]. Regression models make stringent assumptions about the shape of the mapping function which often significantly differ from the true shape of the relationship between the inputs and output [22]. The adopted datasets in the parametric approach typically demonstrate high dimensionality (i.e., data have many inputs) and high multicollinearity (i.e., high correlations between input variables)

[23]. This high dimensionality and high multicollinearity of the adopted datasets for the past regression models prevented the input variables to be independent of each other, violating one of the fundamental assumptions of the parametric approach. As such, an unpredictable variance was imposed on the models weakening their statistical power. High variability in retroreflectivity data made it challenging for the regression models to estimate the service life of the markings with a high level of statistical confidence, even with the collection of more data [24]. Due to these limitations, past regression models [9, 15–21] predicted retroreflectivity with relatively low accuracy (the coefficient of determination (R^2) ranging between 0.14 to 0.75).

Because of the poor reliability of the past regression models, retroreflectivity degradation models were developed using supervised machine learning algorithms [11, 25, 26]. Machine learning algorithms construct models based on a “non-parametric” approach without making explicit assumptions about the shape of the mapping function. These models can efficiently capture complex patterns from the dataset and in general, achieve higher prediction accuracy than the regression models [22]. Karwa and Donnell [25] proposed an Artificial Neural

Table 1 Summary of the literature review on retroreflectivity degradation modeling

Reference	Material(s)	Input(s)	Model(s)	R^2	Study Site(s)
Lee et al. [15]	Waterborne paint, thermoplastic, polyester, tape	Time	Simple linear regression	0.14 – 0.18	MI
Abboud and Bowman [16]	Waterborne paint, thermoplastic	Traffic	Exponential regression	0.31 – 0.58	AL
Sarasua et al. [19]	Thermoplastic, epoxy	Time	Simple linear regression	0.21 – 0.47	SC
Hollingsworth [17]	Waterborne paint, thermoplastic	Time, traffic, bead type, color, initial retroreflectivity, lateral line location, and time	Logarithmic	0.53	NC
Sitzabee et al. [18]	Thermoplastic, waterborne paint	Time, initial retroreflectivity, traffic, line lateral location, line color	Multiple linear regression	0.60 – 0.75	NC
Karwa and Donnell [25]	Thermoplastic	Initial retroreflectivity, time, traffic, marking type, marking location	Artificial Neural Network (ANN)	-	NC
Robertson et al. [20]	Waterborne paint	Time, traffic, lane width, shoulder width	Multiple linear regression	0.24 – 0.34	SC
Ozelim and Turochy [9]	Thermoplastic	Time, traffic, initial retroreflectivity	Multiple linear regression	0.49	AL
Malyuta [21]	Waterborne paint, thermoplastic	Time, traffic	Multiple linear regression	0.33 – 0.46	TN
Mousa et al. [11]	Waterborne paint	Initial retroreflectivity, manufacturer, surface type, color, thickness, bead types, time, air temperature, rainfall, snowfall, traffic, surface age	Categorical Boosting	0.83 – 0.98	FL, PA, MN, MS
Idris et al. [26]	Thermoplastic	Initial retroreflectivity, surface type, color, thickness, bead types, rainfall, traffic	Genetic Algorithm	0.64 – 0.93	FL

Network model to estimate the service life of thermoplastic markings without considering climatic conditions (i.e., rainfall and/or snowplow activities) as input variables. Mousa et al. [11] predicted retroreflectivity for up to three years with reasonable accuracy ($R^2=0.83 - 0.98$) using a Categorical Boosting model, but its applicability was limited to waterborne paint markings. More recently, Idris et al. [26] developed Genetic Algorithm models ($R^2=0.64 - 0.93$) to predict retroreflectivity. However, the models were developed for only thermoplastic markings which only predicted retroreflectivity for up to a year.

Advancements based on previous research

Based on Table 1 it can be observed that most previous retroreflectivity prediction models had limited scope as those were site-specific and marking material-specific models. These models were developed considering a few input variables (i.e., time, traffic, line lateral location, and/or initial retroreflectivity). To address the shortcomings of the past studies, retroreflectivity degradation models were developed in this study using two different machine learning algorithms, Decision Tree and Artificial Neural Network. These models were constructed considering all the significant input variables affecting retroreflectivity. Additionally, the scope of the proposed models has been further extended to seven types of commonly used marking materials along with traffic and climatic conditions of different geographical regions across the U. S.

It is worth mentioning that other common machine algorithms besides Decision Tree and Artificial Neural Network, including Support Vector Machine, LightGBM, and K-Nearest Neighbors, were initially implemented in this study to develop retroreflectivity degradation models. Decision Tree and Artificial Neural Network models yielded the highest prediction accuracy among those, and therefore, the results of Decision Tree and Artificial Neural Network models were only presented in this study.

Data collection

In this study, the measured R_S and other relevant variables were retrieved from NTPEP's online data repository and assembled into a dataset. This dataset included R_S readings from the following eight recent NTPEP test decks distributed over three different U.S. climate zones (Southeast, Northeast, and Upper Midwest):

- Minnesota: 2010, and 2013
- Pennsylvania: 2011, and 2014
- Florida: 2012, 2015, and 2019
- Wisconsin: 2017

A total of 517 marking products were considered in this study. Each product was installed as eight transverse lines, four lines on an asphalt surface and four lines on a concrete surface, resulting in a total of 4,136 marking lines (517 products \times 2 pavement surfaces \times 4 marking lines = 4,136 lines). For each line of these 4,136 lines, R_S measurements were collected at 12 time intervals (after 0, 1, 2, 3, 11, 12, 15, 21, 24, 27, 33, and 36 months) resulting in a total of 49,632 R_S values (517 products \times 8 lines per product \times 12 R_S per line = 49,632). The descriptions of retrieved variables for each marking are presented in Table 2.

Exploratory data analysis

Descriptive statistics

The general descriptive statistics (i.e., minimum, and maximum values, interquartile range, mean, median, and outliers) of the numerical variables were calculated. These descriptive statistics are presented in Fig. 2 as box-whisker plots. It is worth mentioning that the outliers in MR_S , as shown in Fig. 2d, represented true variability of R_S measurements. Therefore, these outliers were not removed to maintain the original statistical distribution of the study dataset, which is a common practice for handling outliers [27].

Correlation analysis

A correlation analysis was conducted to determine the degree of association between all the collected variables. Pearson's R, [28], eta (η) coefficient [29], and Cramer's V [30], were utilized to evaluate the association between numeric-numeric, numeric-categorical, and categorical-categorical variable pairs, respectively. Pearson's R ranges from -1.0 (a perfect, decreasing, linear association) to 1.0 (a perfect, increasing, linear association) while the values of η coefficient and Cramer's V lie between 0 (no association) to 1 (a perfect association). The developed correlation matrix is presented in Fig. 3. Among the numeric-categorical variable pairs, (TH, T) had the highest η coefficient of 0.9. (E, TR) showed the highest Pearson's R of 0.7 among the numeric-numeric variable pairs. (M, T) exhibited the highest Cramer's V of 0.6 among categorical-categorical variable pairs. These high values of η coefficient, Pearson's R, and Cramer's V indicate high multicollinearity in the compiled dataset.

Data preprocessing

Both DT and ANN algorithms cannot process categorical data [31]. Therefore, as a data preprocessing step, the categorical variables were converted into numerical forms utilizing Label Encoding technique [32]. This technique designated every unique category a number ranging

Table 2 Description of the variables in the assembled dataset

Variable types	Variables	Notations	Units	Descriptions
Numerical	Elapsed time	E	month	12 values ($E_0, E_1, E_2, E_3, E_{11}, E_{12}, E_{15}, E_{21}, E_{24}, E_{27}, E_{33},$ and E_{36}). E_0 represents the month of installation, E_1 represents 1 month since installation, E_{11} represents 11 months since installation, and similarly for other values
	Measured Rs	MR_s	$\text{mcd/m}^2/\text{lux}$	12 values (MR_{s0} through MR_{s36} , corresponding to E_0 through E_{36})
	Cumulative traffic level	TR	vehicles	12 values (TR_0 through TR_{36} , corresponding to E_0 through E_{36})
	Cumulative snowfall	SN	in	12 values (SN_0 through SN_{36} , corresponding to E_0 through E_{36})
	Marking thickness	TH	mil	172 unique values of marking thickness
Categorical	Surface type	S	-	2 unique categories (asphalt and concrete)
	Types of marking material	T	-	7 unique categories (waterborne paint, thermoplastic, preformed thermoplastic, permanent polymeric tape, epoxy, polyurea, and MMA)
	Marking color	C	-	2 unique categories (white and yellow)
	Marking manufacturer	M	-	30 unique categories (M1 through M30)
	Bead type of the first drop	b	-	7 unique categories (Type 1, Type 2, Type 3, Type 4, high-performance beads, wet reflective elements, and N/A. Here, N/A represents only one type of glass bead was applied as the second drop)
	Bead type of the second drop	B	-	7 unique categories (Type 1, Type 2, Type 3, Type 4, high-performance beads, premium optics "Utah blend", and N/A. Here, N/A indicates marking does not have dropped on glass beads)

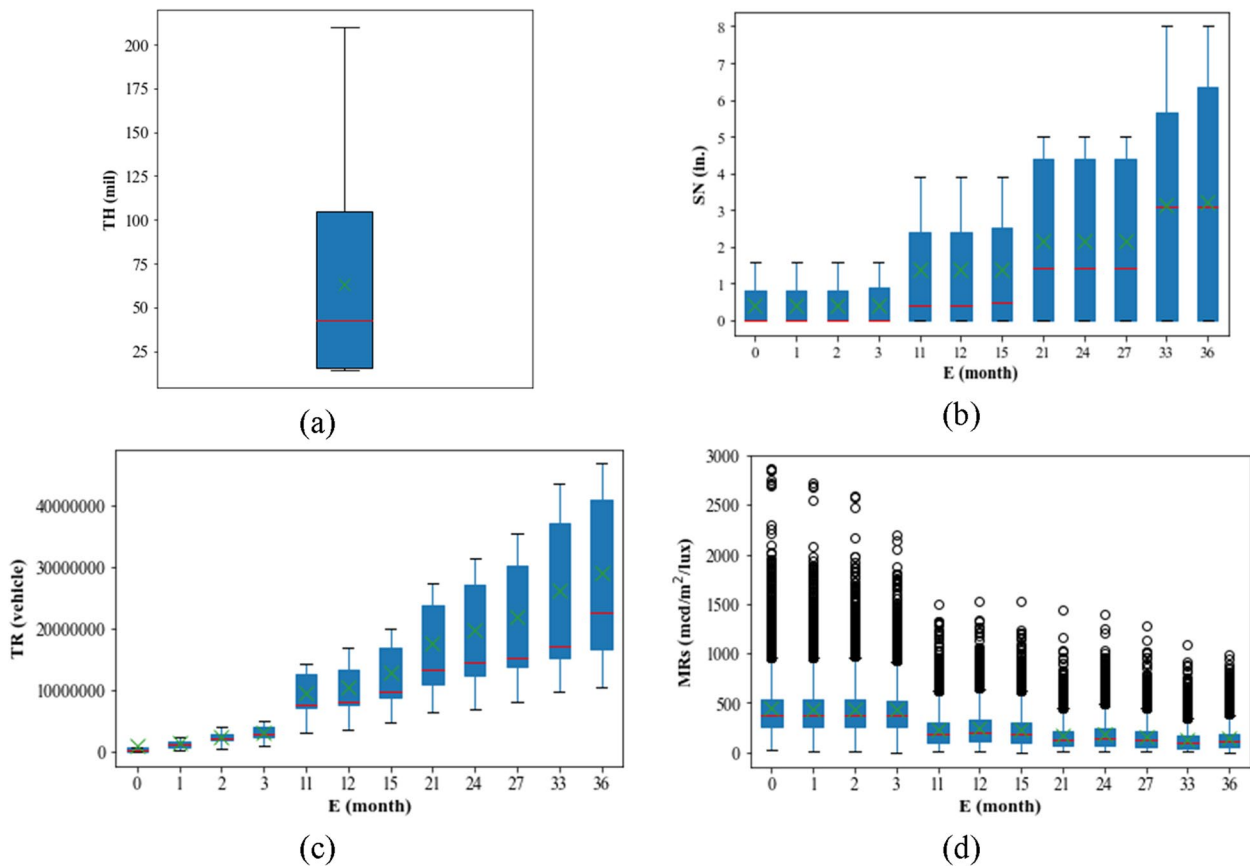


Fig. 2 Descriptive statistics of (a) TH, (b) SN, (c) TR, and (d) MR_s , respectively

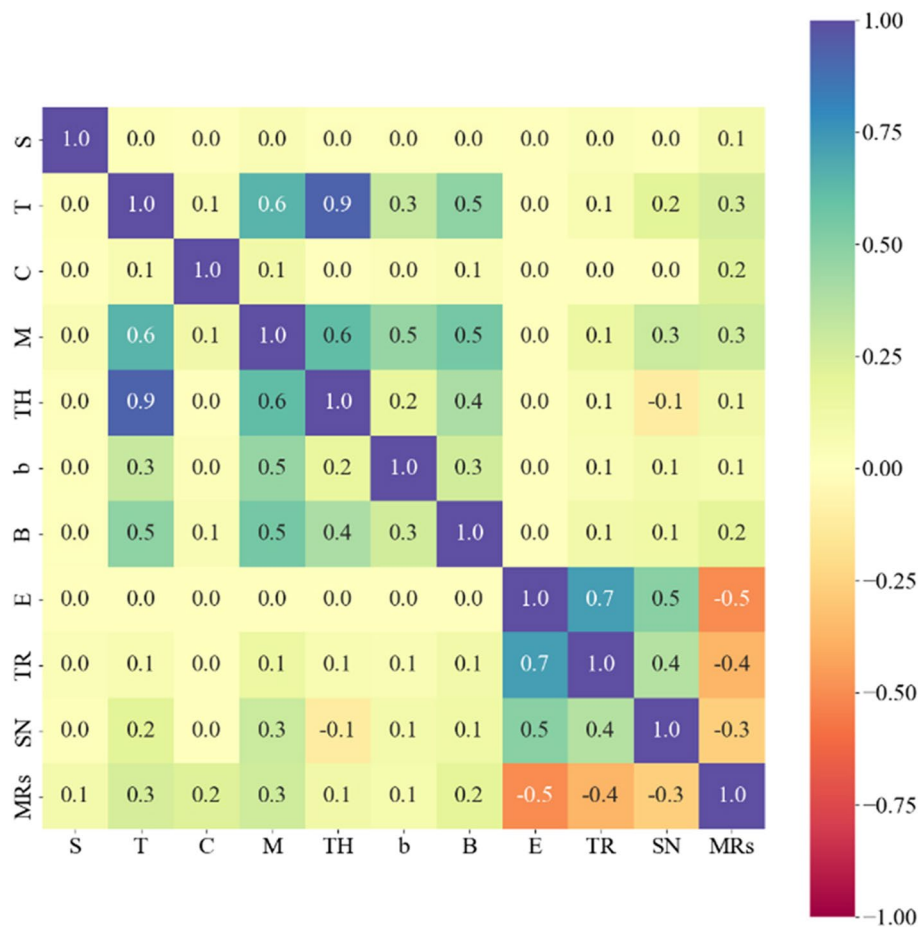


Fig. 3 Correlation matrix

between 0 and (X-1), where X represented the total number of unique categories for this categorical variable.

Model development

This section provides technical details of the machine learning algorithms implemented in this study, followed by an overview of the model development process.

Machine learning algorithms

Two machine learning algorithms were utilized in this study for model development, as described in the following sub-sections.

- a) Decision Tree (DT) is a machine learning algorithm that builds models in the form of tree-like structures consisting of a root node, internal nodes, and leaves that are connected through branches [33]. A simple DT model consisting of two predictors (X_1 and X_2) is illustrated in Fig. 4a. DT uses a series of splitting rules ($X_1 \leq s_1, X_2 \leq s_2, \dots, X_2 \leq s_4$) to divide the training observations into various regions of the input

space utilizing recursive binary splitting technique (see Fig. 4b) [22]. This process is iteratively applied until the objective function is optimized and the leaves (R_1, R_2, \dots, R_5) are established [34]. The mean response value for the observations (C_1, C_2, \dots, C_5) falling under each leaf is used as the final prediction [35]. The most important hyperparameters that control model architecture and require tuning during the training process of a DT model consist of maximum depth (D), minimum samples split (S), minimum samples leaf (L), and maximum features (F) [36]. The general mathematical form of a DT model and the objective function are presented in Eqs. 1 and 2, respectively.

$$f(x) = C_{q(x)}, (q : \mathbb{R}^m \rightarrow 1, 2, \dots, t, C \in \mathbb{R}^m) \quad (1)$$

where,
 $q(x)$ =splitting rule

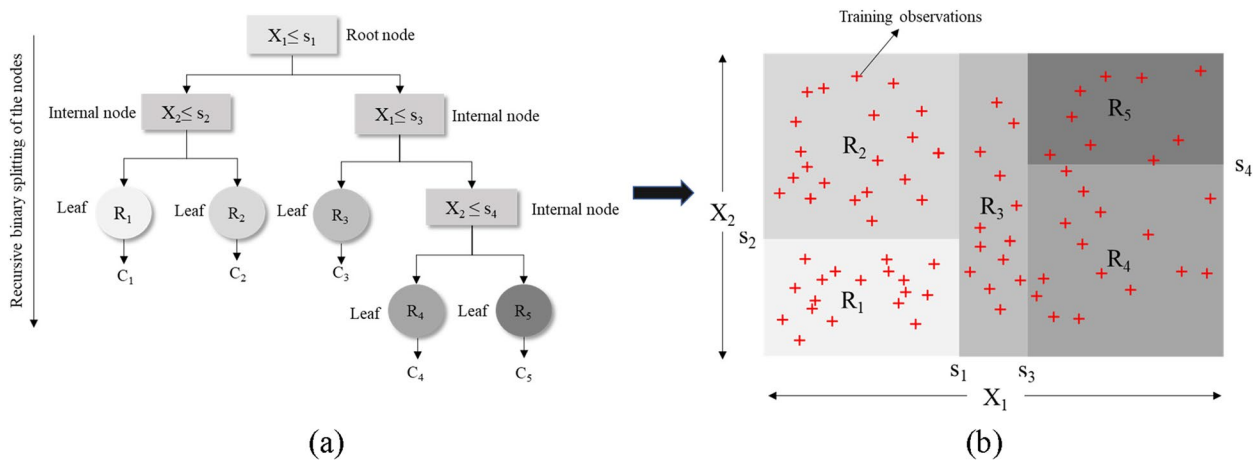


Fig. 4 The schematic representation of a Decision Tree (DT) model

m = number of input variables
 t = total number of leaves
 $C_{q(x)}$ = mean response of a leaf

$$Obj(T) = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \alpha(T) \tag{2}$$

where,

$Obj(T)$ = objective function
 $\sum_{i=1}^N (y_i - \hat{y}_i)^2$ = loss function
 $\alpha(T)$ = regularization term

b) Artificial Neural Network (ANN) is another machine learning algorithm that builds an artificial neural

network with a multilayer perceptron (MLP) architecture, consisting of an input layer, the maximum of two hidden layers, and an output layer [37, 38]. A simple ANN model consisting of a hidden layer is illustrated in Fig. 5. The primary processing elements in each layer are neurons (N_1, N_2, \dots, N_n) that are interconnected by certain weights. The network is built through a 2-stage optimization process, “forward pass” and “backpropagation” [39]. In the forward pass, the inputs (X_1, X_2, \dots, X_j) and associated weights ($W_{1,1}, W_{2,1}, \dots, W_{hh,k}$) are multiplied by the weights, summed, and added with a bias term (a_H) to produce a linear output. The linear output is

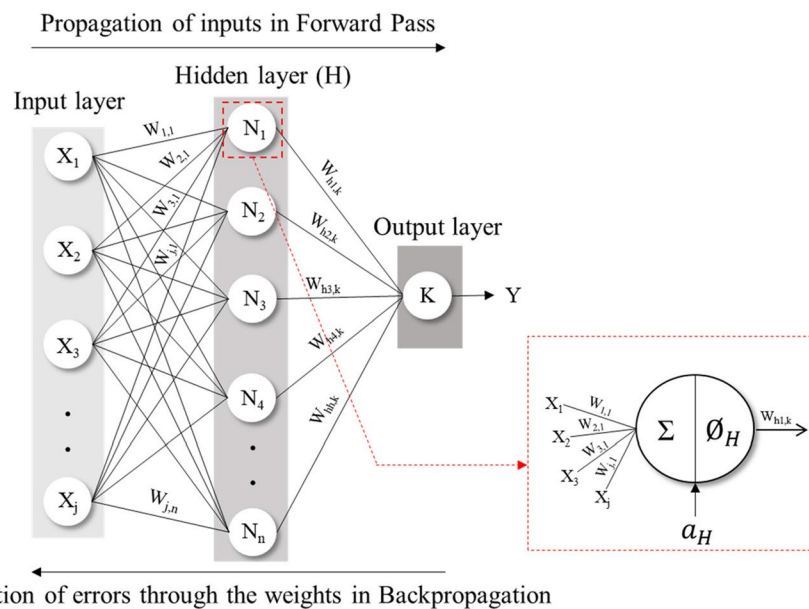


Fig. 5 The schematic representation of an Artificial Neural Network (ANN) model

passed through an activation function (\varnothing_H) (i.e., sigmoid, hyperbolic tangent, or Rectified Linear Unit Function (ReLU)) to obtain a non-linear output [22]. The output from a neuron in the hidden layer acts as an input to the neuron in output layer (K). In back-propagation, the errors made in the forward pass are distributed from the output layer to the input layer through the weights using an optimizer (i.e., adam, or Stochastic Gradient Descent (SGD)) and as a result, the weights are updated [40]. Forward pass and back-propagation are iteratively performed until the objective function is optimized [22]. The most important hyperparameters requiring tuning for an ANN model include number of hidden layers (H), number of neurons (N), batch size (B), activation function (A), optimizer (O), and learning rate (Lr) [36]. The mathematical formulations of an ANN model and objective function are presented in Eqs. 3 and 4, respectively.

$$\hat{y}_n = \varnothing_0 \left(\alpha_k + \sum_{p=1}^N w_{hk} \varnothing_h \left(\alpha_h + \sum_{j=1}^J w_{jh} x_{ij} \right) \right) \quad (3)$$

where,

\hat{y}_n = model prediction

w_{jh}, w_{hk} = weights between input and the hidden layers and between hidden and output layers

x_{ij} = inputs

a_h, a_k = bias terms in the hidden layer neuron and output neuron, respectively

N = number of neurons in the hidden layer

$\varnothing_h, \varnothing_0$ = activation functions in the hidden layer neuron, and output neuron, respectively.

$$Obj(\theta) = \frac{1}{M} \sum_{m=1}^M (y_n - \hat{y}_n)^2 + \Omega\alpha(\theta) \quad (4)$$

where,

$Obj(\theta)$ = objective function

$\frac{1}{M} \sum_{m=1}^M (y_n - \hat{y}_n)^2$ = loss function

$\Omega\alpha(\theta)$ = regularization term.

Overview of model development process

In this research, two model development strategies (Strategy A and Strategy B) and were utilized for model development and their results were compared to identify the best strategy of constructing R_S prediction models. In Strategy A, two different integral models was developed, one for DT and one for ANN, to predict R_S after one month (PR_{S1}), two months (PR_{S2}), three months (PR_{S3}), and 11 months (PR_{S11}), and similarly after 12, 15, 21, 24, 27, 33, and 36 months using $S, T, C, M, TH, b, B, E, TR,$ and SN as inputs and MR_S as the target variable. As per

this strategy, the general formulations for the DT and ANN models are presented in Eq. 5.

$$PR_S = f(S, T, C, M, TH, b, B, E, TR, SN) \quad (5)$$

where,

PR_S = Predicted R_S

S, T, C, M, TH, b, B = Time independent inputs

TR, SN = Time dependent inputs

On the other hand, utilizing Strategy B, a sequential R_S prediction models were developed for both DT and ANN and two separate sets of 11 models (DT-A through DT-K and ANN-A through ANN-K) were developed. The schematic of the model inputs and outputs for DT models is illustrated in Fig. 6, as an example. As shown in Fig. 6, Model DT-A utilizes initially measured R_S (MR_{S0}) and other key input variables at the time of installation ($S, T, C, M, TH, b, B, TR_0,$ and SN_0) to predict R_S after $E=1$ month. For Model DT-B, the output from Model DT-A (PR_{S1}) was combined with other inputs at $E=1$ month to predict R_S after $E=2$ months. This process was iteratively used for the remaining models. Based on the framework illustrated in Fig. 6, the general formulations for the DT and ANN models are presented in Eq. 6.

$$PR_{SE_i} = f(S, T, C, M, TH, b, B, TR_{SE_{i-1}}, SN_{SE_{i-1}}, X_{E_{i-1}}) \quad (6)$$

where,

E_i = Elapsed times (E_0, E_1, \dots, E_{36})

$X_{E_{i-1}} = MR_{S0}$, for $E_i = E_0$ or $PR_{SE_{i-1}}$, for $E_i > E_0$

PR_{SE_i} = Predicted R_S for month E_i

S, T, C, M, TH, b, B = Time independent inputs

$TR_{SE_{i-1}}, SN_{SE_{i-1}}$ = Time dependent inputs at previous elapsed time (E_{i-1})

MR_{S0} = Initially measured R_S

$PR_{SE_{i-1}}$ = Predicted R_S at previous elapsed time (E_{i-1})

Model training

Training DT models

DT models, developed with Strategy A or Strategy B, were trained with the training dataset (randomly selected 80% of the total data points) when the models learnt patterns from the training dataset. DT-based models are unaffected by multicollinearity [41, 42]. These models are also insensitive to the scale of the inputs as the nodes are split based on a single input and are not affected by other inputs [43]. Therefore, inputs were not scaled for training DT models.

Training ANN models

ANN models, developed with Strategy A or Strategy B, were also trained with the training dataset. ANN uses the gradient descent technique to optimize the objective

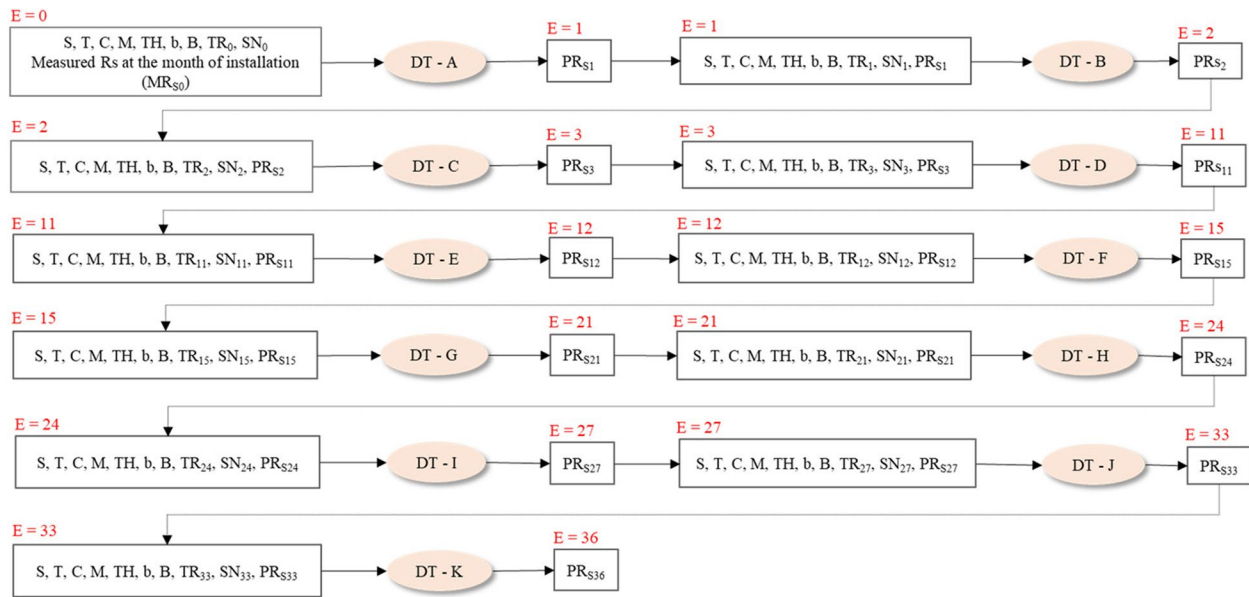


Fig. 6 Schematic representation of the inputs and output for the DT models

function and scaling the input variables enables it to reach global minima faster [44]. As such, Standard Scaler [45] was implemented to normalize the input variables by removing the mean and scaling to unit variance to ensure fast convergence. During training, each model was set to train for 1000 iterations. Early stopping was included in the training process to terminate training when validation R^2 did not improve for 100 consecutive iterations.

Hyperparameter tuning

For both DT and ANN models, model hyperparameters were tuned during the training (Strategy A and Strategy B). Maximum depth (D), minimum samples split (S), minimum samples leaf (L), and maximum features (F) hyperparameters were tuned to prevent overfitting of the DT models. On the other hand, number of hidden layers (H) and neurons (N) were tuned for the ANN models to control model complexity and prevent overfitting. Batch size (B), activation function (A), optimizer (O), and learning rate (Lr) were tuned to achieve improved predictive performance from the ANN models. Moreover, L2 regularization was utilized by tuning alpha (α) hyperparameter to mitigate the effect of high multicollinearity of the assembled dataset.

The tuning of the model hyperparameters for both Strategy A and Strategy B were achieved through the combined implementation of grid search and 10-fold cross-validation techniques [11]. Grid search evaluated all possible combinations of values within the defined hyperparameter space to identify their optimal combination with maximum accuracy. Grid search was

accompanied by 10-fold cross-validations, segmenting the training dataset into ten subsets. Training was performed with nine subsets, and validation was done with the remaining subset. This was repeated ten times by changing the validation subset. The average R^2 value for the ten trials was used to evaluate the performance of the models. The developed hyperparameter spaces and optimum hyperparameter combinations for the models developed with Strategy A and Strategy B are presented in Tables 3, 4, 5 and 6.

The training performance of the models developed with Strategy A and Strategy B was evaluated using coefficient of determination (R^2), mean absolute percentage error (MAPE), and root mean square error (RMSE), as illustrated in Table 7 and Fig. 7, respectively. It can be observed from Table 7 that in Strategy A, ANN model provided better training performance than DT model with higher R^2 (0.82), lower MAPE (44.6%), and lower RMSE (106.9 mcd/m²/lux).

On the other hand, in Strategy B, the training R^2 , MAPE, and RMSE values were similar for both DT and ANN models until a prediction horizon of 3 months (i.e., DT-A to DT-C and ANN-A to ANN-C) (see Fig. 7). Although, in general, a downward trend in R^2 and an upward trend in MAPE were observed when the prediction horizon was further extended from 3 to 36 months (i.e., DT-D to DT-K or ANN-D to ANN-K), no trend was observed in RMSE. The decrease in overall training accuracy with an increase in prediction horizon in Strategy B was expected because the predicted values were used as inputs to make further predictions, inducing

Table 3 Hyperparameter space for the models developed with Strategy A

DT model		ANN model	
Hyperparameters	Values	Hyperparameters	Values
Maximum depth (D)	5, 10, 15, 20	Number of hidden layers (H)	1, 2
Minimum samples split (S)	30, 40, 50, 100	Neurons (N)	25, 50, 100
Minimum samples leaf (L)	20, 30, 40, 50	Batch size (B)	8, 16, 32, 64, 200
Max features (F)	auto = considering all the inputs, sqrt = considering the square root of the total number of inputs, and log2 = Log Base 2 of the total number of inputs	Activation function (A)	ReLU, sigmoid, tanh
		Optimizer (O)	adam, SGD
		Learning rate (Lr)	0.001, 0.01, 0.1
		Alpha (α)	0.1, 0.4, 0.6, 0.8

Table 4 Optimal combination of the hyperparameters for Strategy A

Model	D	S	L	F	Model	H	N	B	A	O	Lr	α
DT	20	30	20	sqrt	ANN	2	(50, 100)	200	ReLU	adam	0.001	0.1

Table 5 Hyperparameter space for the models developed with Strategy B

DT models		ANN models	
Hyperparameters	Values	Hyperparameters	Values
Max depth (D)	4, 5, 10, 20, 30, 40, 50	Number of hidden layers (H)	1, 2
Minimum samples split (S)	2, 5, 10, 12, 16, 18, 20, 50, 100	Neurons (N)	15, 30, 60
Minimum samples leaf (L)	2, 5, 10, 20, 30, 40, 50	Batch size (B)	16, 32, 64, 200
Max features (F)	auto, sqrt, log2	Activation function (A)	ReLU, sigmoid, tanh
		Optimizer (O)	adam, SGD
		Learning rate (Lr)	0.001, 0.01, 0.1
		Alpha (α)	0.1, 0.4, 0.6, 0.8

Table 6 Optimal combination of the hyperparameters for the models developed with Strategy B

DT models					ANN models							
Models	D	S	L	F	Models	H	N	B	A	O	Lr	α
DT-A	10	18	2	auto	ANN-A	2	(30, 60)	32	ReLU	adam	0.001	0.1
DT-B	10	10	2	auto	ANN-B	2	(30, 60)	64	ReLU	adam	0.001	0.1
DT-C	10	10	2	auto	ANN-C	2	(30, 60)	64	ReLU	adam	0.001	0.1
DT-D	20	2	2	sqrt	ANN-D	2	(30, 60)	32	ReLU	adam	0.001	0.4
DT-E	10	50	20	auto	ANN-E	1	(30,)	32	ReLU	adam	0.001	0.1
DT-F	20	2	2	auto	ANN-F	1	(30,)	32	ReLU	adam	0.001	0.1
DT-G	20	50	2	auto	ANN-G	1	(60,)	64	ReLU	adam	0.001	0.6
DT-H	10	2	2	auto	ANN-H	2	(30, 60)	64	ReLU	adam	0.001	0.6
DT-I	50	2	2	sqrt	ANN-I	2	(30, 60)	64	ReLU	adam	0.001	0.1
DT-J	10	2	10	sqrt	ANN-J	2	(30, 60)	32	ReLU	adam	0.001	0.1
DT-K	4	2	2	auto	ANN-K	1	(30,)	32	ReLU	adam	0.001	0.1

Table 7 Training performance of the models developed with Strategy A

Model	R ²	MAPE (%)	RMSE (mcd/m ² /lux)
DT	0.78	45.5	116.4
ANN	0.82	44.6	106.9

errors in model predictions. Utilizing Strategy B, DT and ANN models fitted the training data with average R², MAPE, and RMSE of 0.89/27.1%/53.9 mcd/m²/lux and 0.86/38.1%/61.6 mcd/m²/lux, respectively and therefore, demonstrated better training performance than the models developed with Strategy A. The ranges of training R², MAPE, and RMSE for DT models were from 0.77 to 0.96, 9.7 to 57.6%, and 46.0 to 67.3 mcd/m²/lux, respectively while for ANN models, these ranges were from 0.76 to 0.96, 10.1 to 62.3%, and 53.5 to 73.5 mcd/m²/lux, respectively. Both DT and ANN models fitted training data well utilizing Strategy B with R² values higher than the models listed in Table 1. Overall, the DT models demonstrated a better training performance than ANN models in Strategy B.

Model testing

Testing dataset (remaining 20% of total data points) was used to evaluate the performance of the developed models with the unseen data. The testing performance of the models developed with Strategy A is presented in Fig. 8. From Fig. 8, it can be observed that, like training performance, ANN provided better testing performance than DT with higher R² (0.76), lower MAPE (50.5%), and lower RMSE (121.6 mcd/m²/lux).

The testing performance of the DT and ANN models developed with Strategy B are shown in Figs. 9 and 10, respectively. Results showed that, like training accuracy, the testing accuracy of these models decreased with an increase of prediction horizon beyond 3 months (Figs. 9 and 10). Using Strategy B, DT and ANN models provided average R², MAPE, and RMSE of 0.76/41.9%/79.7 mcd/m²/lux and 0.77/45.5%/74.8 mcd/m²/lux, respectively and therefore, demonstrated better testing performance than the models developed with Strategy A. The ranges of testing R², MAPE, and RMSE for DT models developed with Strategy B were between 0.54 to 0.93, 12.6 to 95.1%, and 67.9 to 95.2 mcd/m²/lux, respectively; while for ANN models, these ranges were between 0.55 to 0.94, 10.6 to 80.7%, and 61.2 to 92.9 mcd/m²/lux, respectively. These models provided close R_s estimates at different prediction horizons from a dataset that was not used during model training, and thereby demonstrating their robustness. Most of the ANN models exhibited better testing

performance than the corresponding DT models. The testing R² values of these models were also reasonably higher than the R² of most past regression models predicting retroreflectivity for up to 3 years. These high R² values indicated that proposed machine learning models developed with Strategy B can predict R_s with a superior level of accuracy than the traditional regression models.

It is worth mentioning that machine learning aims to build models that generalize well to unseen data, ensuring their effectiveness in real-world applications [46]. As such, testing performance is typically prioritized over training performance when selecting the best models [22]. With Strategy B, DT models exhibited better training performance than the ANN models, while the ANN models demonstrated superior testing performance and hence, a better generalization ability on unseen data than the DT models. Therefore, ANN was selected as the better algorithm than DT for the adopted dataset and ANN models developed with Strategy B were considered for further evaluation.

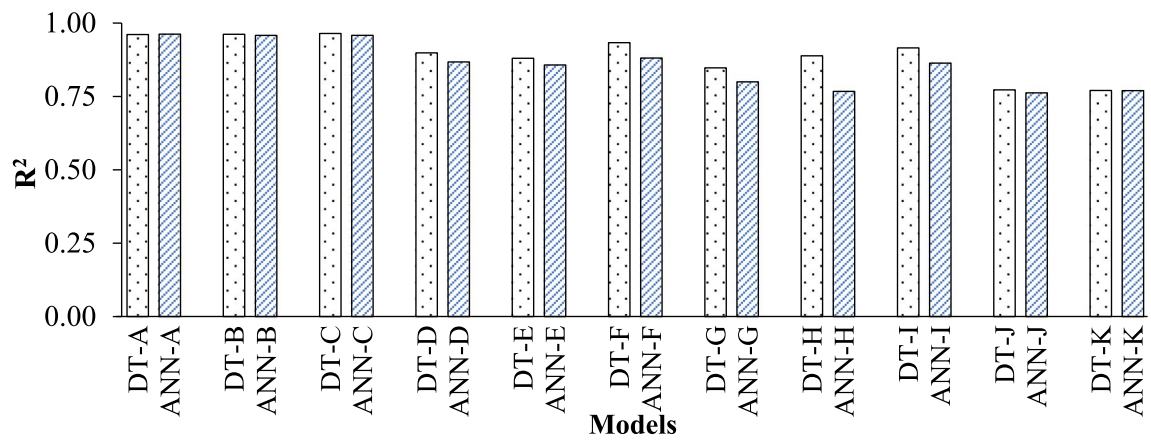
Feature importance study

The importance of each input for every ANN model was assessed using the SHapley Additive exPlanations (SHAP) values [47]. SHAP value for an input represents the contribution of that input to the difference between actual and expected prediction, averaged over all possible permutations of inputs. An input with a higher SHAP value indicates its higher impact on model prediction. For every ANN model, SHAP values for each input were calculated by taking the weighted average of the marginal contribution of the inputs. The mean absolute SHAP values of ANN-A inputs are presented in Fig. 11, as an example. The mean absolute SHAP values for the remaining models are reported in Table 8.

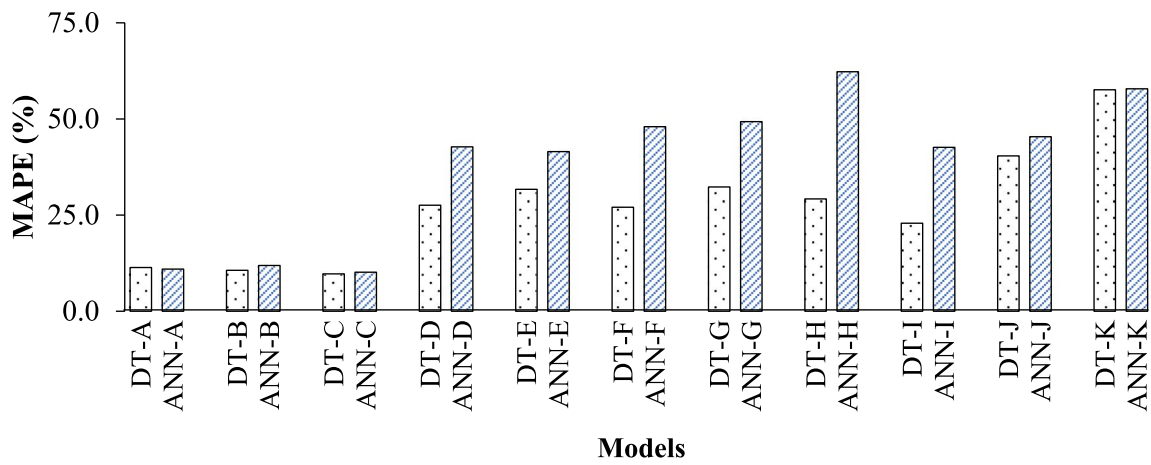
For ANN-A, initially measured retroreflectivity (MR_s at E=0) was the most important input (mean absolute SHAP value = 166.9) (in Fig. 11). For the remaining models (in Table 8), predicted skip retroreflectivity (PR_s) had the highest impact on model prediction. Additionally, SN followed by TR were the most important inputs after MR_s or PR_s for all ANN models (in Fig. 11 and Table 8). Most retroreflectivity degradation models in the literature did not consider these key inputs simultaneously which might explain their relatively low accuracy.

Illustrative implementation of the proposed models

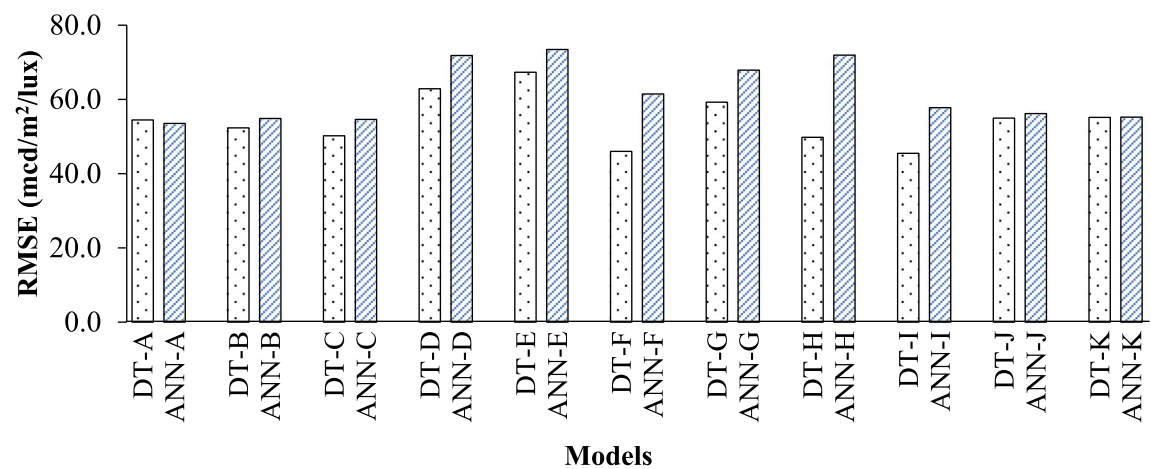
Before using a pavement marking product in a project, a transportation agency might be interested in determining the expected service life of a specific marking product based on a specified minimum retroreflectivity threshold. As an example, the process of implementing



(a)



(b)



(c)

Fig. 7 Training performance of the developed models

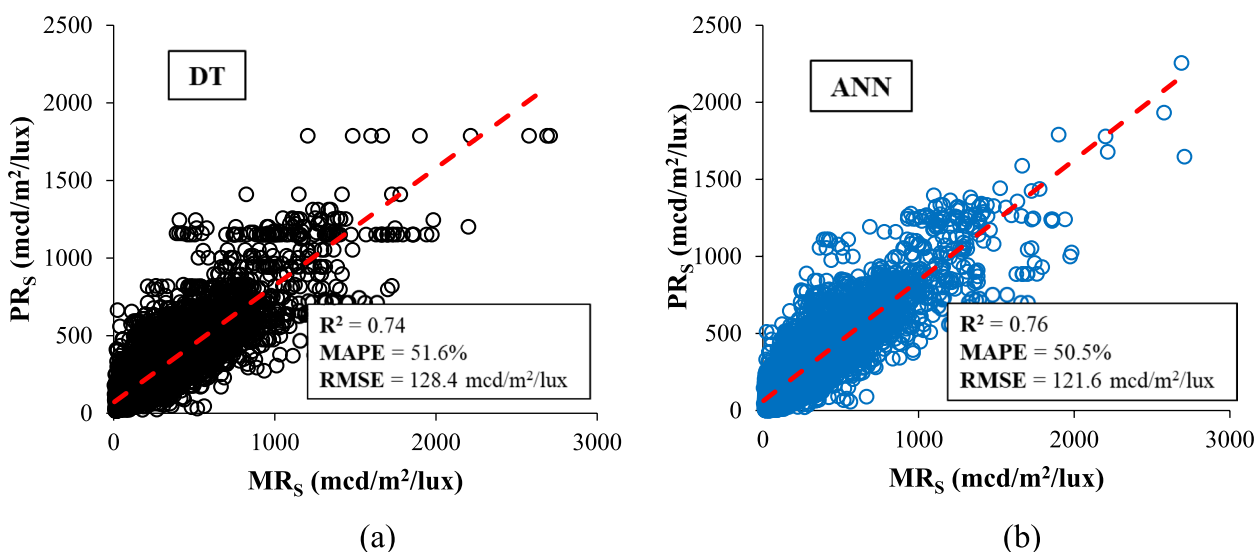


Fig. 8 Testing performance of the models developed with Strategy A

the proposed models to estimate the service life of one of the marking products (included in testing dataset) is explained in the following steps with the aid of Tables 9 and 10.

Step 1: Initial retroreflectivity measurement

Firstly, initial R_s of the marking line right after field installation needs to be determined. This could be measured in the field at the project level or assumed based on previous similar projects. As shown in Table 9, this value was 329 $\text{mcd/m}^2/\text{lux}$.

Step 2: Collection of inputs over 3 years

In Step 2, all the other inputs in Tables 9 and 10 should be collected at different elapsed times (i.e., $E=0, 1, 2, 3, 11, 12, 15, 21, 24, 27,$ and 33 months). $S, T, C, M, Th, b,$ and B are time-independent variables and could be easily determined for a specific product. TR and SN vary with E , which could be determined, predicted, or assumed based on historical data. It should be noted that for accurate prediction, all the numerical inputs applied at different E should lie within the ranges illustrated in Fig. 2.

Step 3: Sequential use of the proposed models

In this step, the user will assign all the input data at $E=0$ and employ ANN-A to calculate PR_s at month 1 (369 $\text{mcd/m}^2/\text{lux}$ in Table 9). Afterward, the user will employ all the input data at $E=1$ along with PR_s at month 1 (369 $\text{mcd/m}^2/\text{lux}$ in Table 9) and employ ANN-B to calculate PR_s at month 2 (343 $\text{mcd/m}^2/\text{lux}$ in Table 9). This process is repeated for all the remaining values of E and for all the remaining models in Tables 9 and 10 until the PR_s

at month 36 (96 $\text{mcd/m}^2/\text{lux}$ in Table 10) is obtained. A comparison between measured R_s (as collected from NTPEP) and predicted R_s from the proposed models is illustrated in Fig. 12.

Step 4: Transverse to longitudinal retroreflectivity conversion

Based on the agency’s policy, the predicted transverse skip retroreflectivity (R_s) values at different E should be transformed into longitudinal retroreflectivity (R_L), which mimics the actual field conditions. It has been widely accepted to assume that R_s correlated well to R_L [14]. Furthermore, a recent study proposed simple models to perform the conversion between R_s and R_L [48].

Step 5: Service life estimation

Based on the R_L values obtained in step 4 at different E , the service life of the marking line can be estimated until the time when a marking product’s R_L drops below 100 $\text{mcd/m}^2/\text{lux}$. As per the model predictions in Fig. 12, the expected service life of this marking product is somewhere between 27 to 33 months.

Summary and conclusions

The objective of this study was to develop machine learning models that could be utilized by U.S. local and state agencies to reliably predict the retroreflectivity of pavement markings. To fulfill this objective, transverse skip retroreflectivity data and other key variables were retrieved from NTPEP database. DT and ANN algorithms were considered to develop models for predicting R_s sequentially for up to 3 years using two different

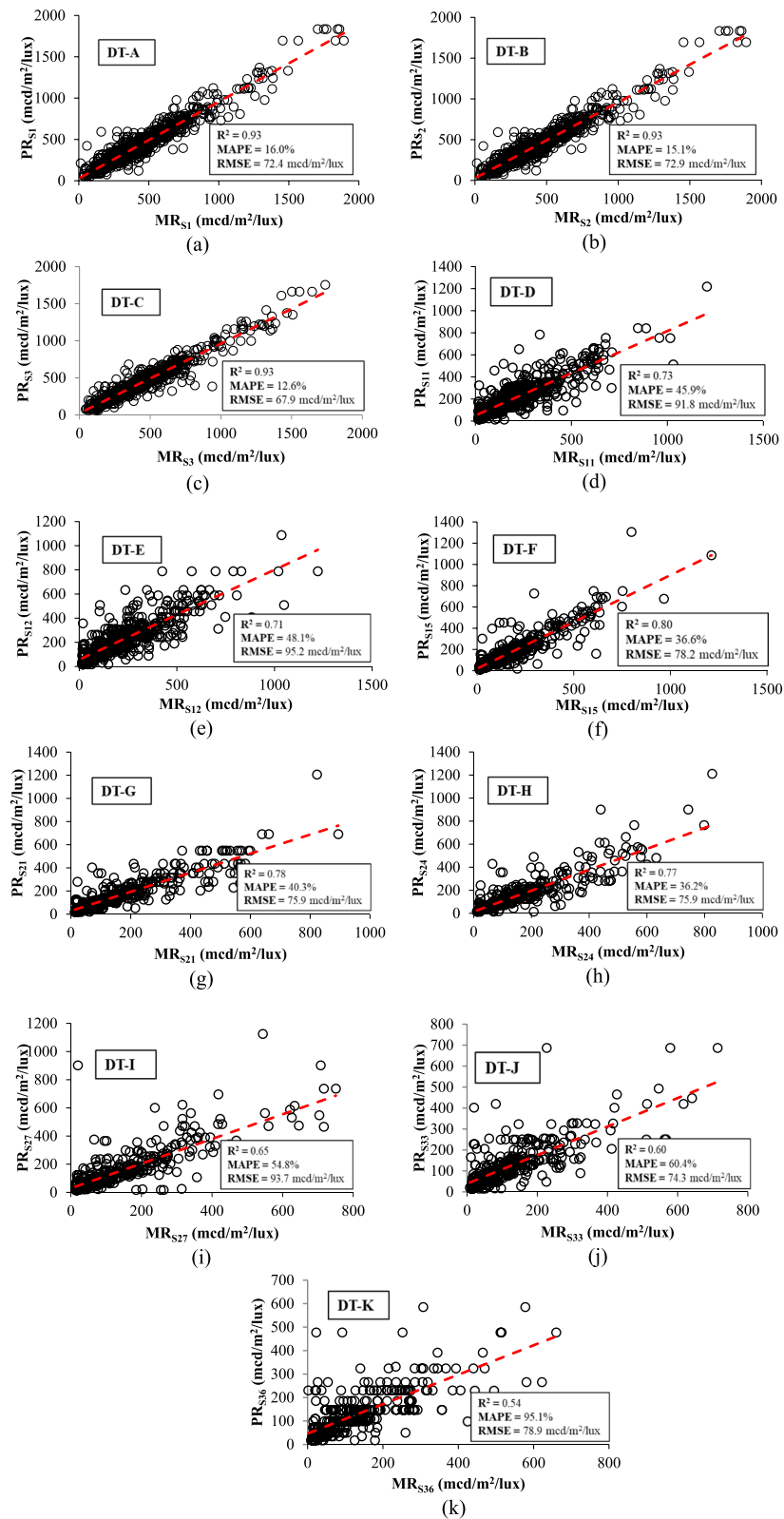


Fig. 9 Testing performances of the models developed with DT algorithm (Strategy B)

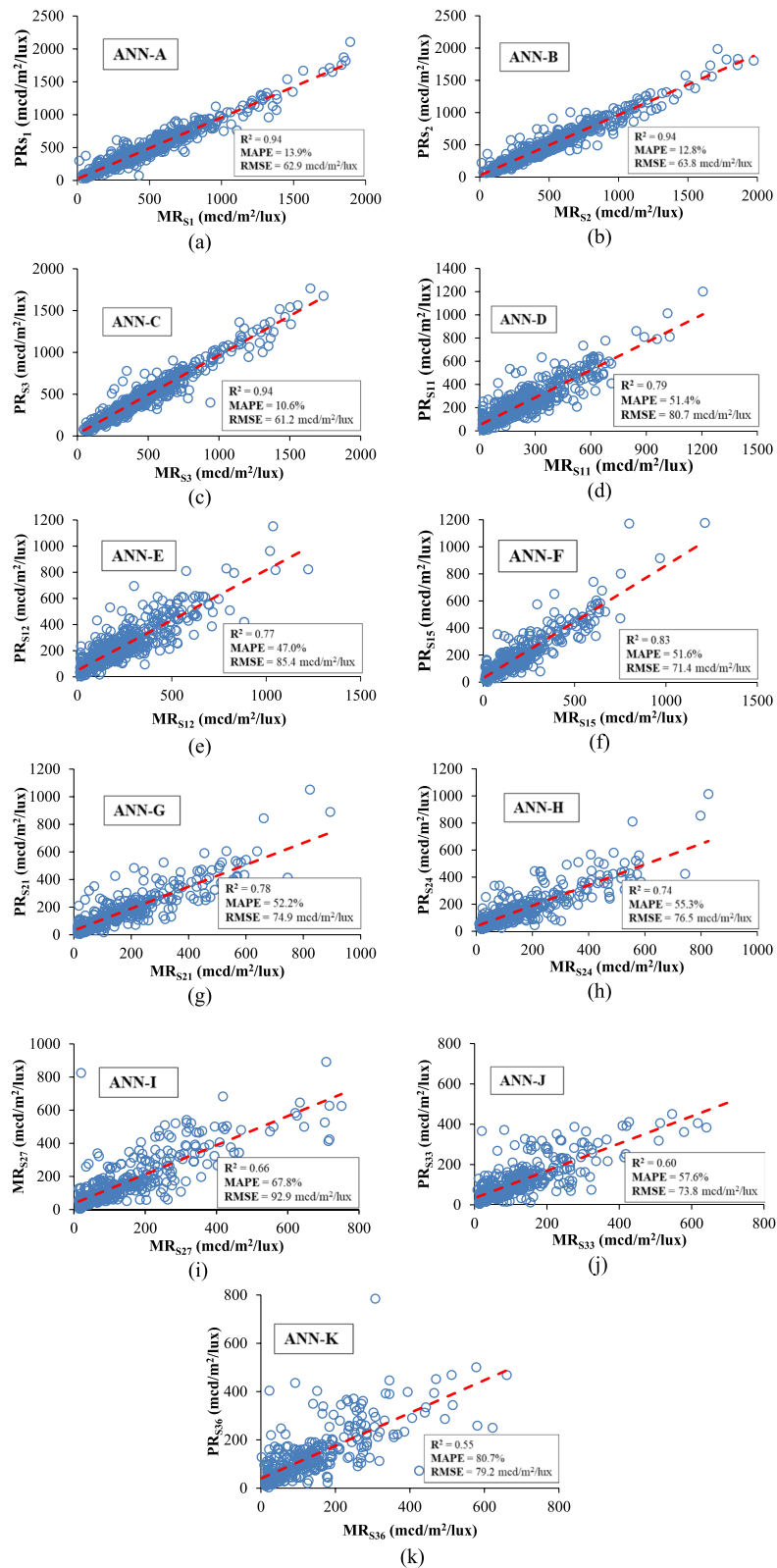


Fig. 10 Testing performances of the models developed with ANN algorithm (Strategy B)

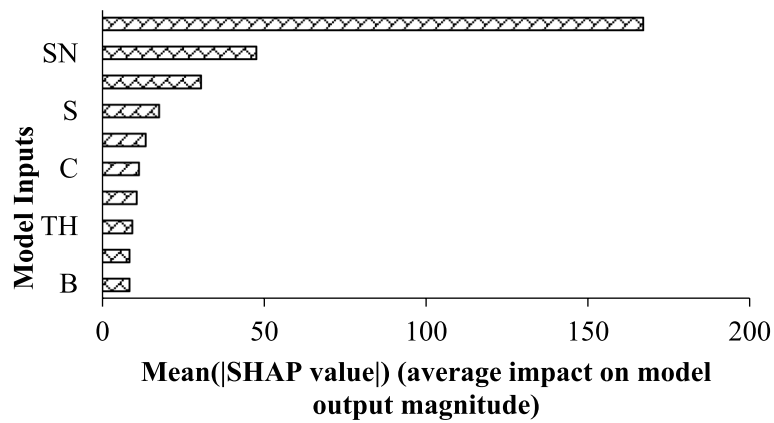


Fig. 11 Feature importance values of the input variables to ANN-A

Table 8 Feature importance of the input variables for models ANN-B through ANN-K

Mean(SHAP value) (average impact on model output magnitude)										
Models	S	T	C	M	TH	b	B	TR	SN	PR ₅
ANN-B	6.1	6.1	10.1	8.6	13	11.9	14.4	14.8	15.8	174.6
ANN-C	4.9	7.7	5.6	5.6	7.7	3.5	5.6	11.8	11.9	169.1
ANN-D	24.1	18.6	33	16.3	25.1	10	15.7	33.7	66.4	75.8
ANN-E	2.4	1.9	1.9	2.9	1.9	1.9	3.4	3.9	4.4	115.4
ANN-F	2.6	1.3	4.3	5.4	5.4	4.3	3.7	5.7	22.0	104.1
ANN-G	3.3	3.7	5.6	2.9	5.2	3.7	3.7	7.1	11.4	100.2
ANN-H	2.3	3.3	5.0	4.6	6.9	6.9	8.4	11.5	18.1	70.4
ANN-I	8.0	9.8	7.3	6.2	10.3	4.3	6.8	10.3	12.3	110.5
ANN-J	8.9	5.8	6.0	6.0	6.8	4.5	8.9	11.1	14.7	69.1
ANN-K	5.8	1.7	2.3	1.7	2.4	2.0	2.4	5.8	6.8	69.4

Table 9 Example results for models ANN-A through ANN-F

Variable type	Variables	Reported values					
		ANN-A	ANN-B	ANN-C	ANN-D	ANN-E	ANN-F
Input	E (month)	0	1	2	3	11	12
	S	Asphalt	Asphalt	Asphalt	Asphalt	Asphalt	Asphalt
	T	Polyurea	Polyurea	Polyurea	Polyurea	Polyurea	Polyurea
	C	White	White	White	White	White	White
	M	M-29	M-29	M-29	M-29	M-29	M-29
	TH (mil)	20.7	20.7	20.7	20.7	20.7	20.7
	b	N/A	N/A	N/A	N/A	N/A	N/A
	B	Premium optics "Utah blend"	Premium optics "Utah blend"	Premium optics "Utah blend"	Premium optics "Utah blend"	Premium optics "Utah blend"	Premium optics "Utah blend"
	TR (no. of vehicles)	222,000	740,000	1,776,000	2,812,000	12,617,000	13,394,000
	SN (in.)	1.6	1.6	1.6	1.6	3.9	3.9
Prediction horizon (month)	Input R ₅ (mcd/m ² /lux)	329	369	343	299	160	130
		1	2	3	11	12	15
Output	PR ₅ (mcd/m ² /lux)	369	343	299	160	130	87
Actual	Actual R ₅ (mcd/m ² /lux)	344	329	355	168	125	54

Table 10 Example results for models ANN-G through ANN-K

Variable type	Variables	Reported values				
		ANN-G	ANN-H	ANN-I	ANN-J	ANN-K
Input	E (month)	15	21	24	33	36
	S	Asphalt	Asphalt	Asphalt	Asphalt	Asphalt
	T	Polyurea	Polyurea	Polyurea	Polyurea	Polyurea
	C	White	White	White	White	White
	M	M-29	M-29	M-29	M-29	M-29
	TH (mil)	20.7	20.7	20.7	20.7	20.7
	b	N/A	N/A	N/A	N/A	N/A
	B	Premium optics "Utah blend"	Premium optics "Utah blend"	Premium optics "Utah blend"	Premium optics "Utah blend"	Premium optics "Utah blend"
	TR	16,798,000	23,791,000	27,158,000	30,229,000	37,259,000
	SN (in.)	3.9	5.0	5.0	5.0	8.0
Prediction horizon (month)	Input R_s (mcd/m ² /lux)	87	133	127	137	98
	Output PR_s (mcd/m ² /lux)	133	127	137	98	96
Actual	Actual R_s (mcd/m ² /lux)	127	58	86	89	87

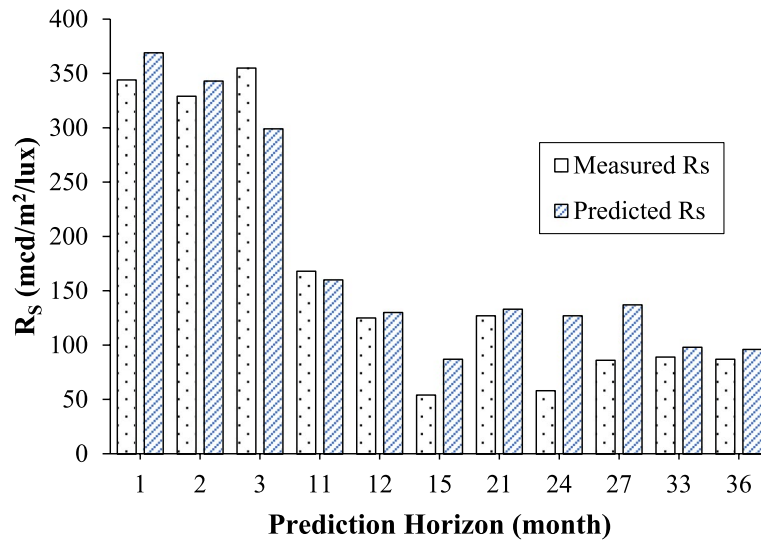


Fig. 12 Comparison of the measured and predicted R_s for the example in Tables 9 and 10

model development strategies. All the models were trained with 80% of the total data points and tested with the remaining 20% data points. The findings and conclusions of the study were as follows:

- Correlation analysis of the collected variables indicated that some variables were highly correlated which confirmed high multicollinearity in the study dataset.
- Sequential retroreflectivity prediction models demonstrated higher accuracy than integral retroreflectivity prediction models both in training and testing.
- Both sequential DT and ANN models, with overall R^2 ranging between 0.54 to 0.96 and 0.55 to 0.96, predicted retroreflectivity at different prediction horizons with a superior level of accuracy as compared with the regression models proposed in the literature.

- ANN models exhibited better testing performance than the DT models, and therefore, were selected as the better algorithm than DT for the adopted dataset.
- Study of the feature importance of the input variables to ANN models revealed that initial retroreflectivity or predicted retroreflectivity followed by snowfall and traffic were the most important inputs to model predictions.
- The proposed models are expected to assist state agencies and transportation officials in determining the service life of pavement marking products and plan for future restriping activities accordingly.

Overall, non-parametric supervised machine learning algorithms seemed to be a promising alternative to traditional parametric methods in modeling retroreflectivity degradation of pavement markings. In the future, longitudinal retroreflectivity data should be collected from northern and western states. More sophisticated DT-based ensemble algorithms (i.e., XGBoost and LightGBM) and deep learning techniques (i.e., Recurrent Neural Network) should be employed to develop marking performance prediction models. Additionally, lightweight machine learning models that require less data should be explored to increase applicability in cases with limited data availability. Models compatible with multi-functional road condition detection vehicles, which are being increasingly used for pavement surveys, should also be developed. The models with best performance might be utilized to develop a simple retroreflectivity prediction tool that can enhance the decision-making process of transportation agencies regarding future restriping activities for the marking products.

Acknowledgements

Not applicable.

Authors' contributions

MM and III developed the study concept. III collected the data, performed the analysis, developed models, and interpreted the results. III, MM, and MH prepared the draft manuscript. All authors reviewed the results and approved the final version of the manuscript.

Funding

This research was funded by the National Cooperative Highway Research Program (NCHRP) (Project Number: 20-30/IDEA 237).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Received: 18 August 2023 Revised: 27 January 2024 Accepted: 31 January 2024

Published online: 21 February 2024

References

1. Rasdorf WJ, Hummer JE, Zhang G, Sitzabee W (2009) Pavement marking performance analysis. Fin Rep, NCDOT, Raleigh
2. Mohamed MM (2019) Evaluation and modeling of pavement marking characteristics based on laboratory and field data. Dissertation, University of Idaho, Ann Arbor
3. Migletz J, Graham JL, Harwood DW, Bauer KM (2001) Service life of durable pavement markings. *Transp Res Rec* 1749(1):13–21
4. Fu H, Wilmot CG (2013) Evaluating alternative pavement marking materials. *Public Work Manag Policy* 18(3):279–297
5. Benz RJ, Pike AM, Kuchangi SP, Brackett Q (2009) Serviceable pavement marking retroreflectivity levels. Tech Rep FHWA/TX-09/0-5656-1, TxDOT, Austin
6. Sitzabee WE, White ED, Dowling AW (2013) Degradation modeling of polyurea pavement markings. *Public Work Manag Policy* 18(2):185–199
7. Xu L, Chen Z, Li X, Xiao F (2021) Performance, environmental impact and cost analysis of marking materials in pavement engineering, the-state-of-art. *J Clean Prod* 294:126302
8. Hussein M, Sayed T, El-Basyouny K, de Leur P (2020) Investigating safety effects of wider longitudinal pavement markings. *Accid Anal Prev* 142:105527
9. Ozelim L, Turochy RE (2014) Modeling retroreflectivity performance of thermoplastic pavement markings in Alabama. *J Transp Eng* 140(6):05014001
10. FHWA (2022) National standards for traffic control devices; the manual on uniform traffic control devices for streets and highways; maintaining pavement marking retroreflectivity. *Fed Regist* 87(150):47921–47931
11. Mousa MR, Mousa SR, Hassan M, Carlson P, Elnaml IA (2021) Predicting the retroreflectivity degradation of waterborne paint pavement markings using advanced machine learning techniques. *Transp Res Rec* 2675(9):483–494
12. Wang S (2010) Comparative analysis of ntpcp pavement marking performance evaluation results. Dissertation, University of Akron, Akron
13. Thomas GB, Scholz C (2001) Durable, cost-effective pavement markings phase I: synthesis of current research. Fin Rep Project No. TR-454, IOWA DOT, Ames
14. Zhang Y, Wu D (2010) Methodologies to predict service lives of pavement marking materials. *J Transp Res Forum* 45(3):5–18
15. Lee JT, Maleck TL, Taylor WC (1999) Pavement making material evaluation study in Michigan. *ITE J* 69(7):44
16. Abboud N, Bowman BL (2002) Cost-and longevity-based scheduling of paint and thermoplastic striping. *Trans Res Rec* 1794(1):55–62
17. Hollingsworth JD (2012) Understanding the impact of bead type on paint and thermoplastic pavement markings. Dissertation, Airforce Institute of Technology, Ohio
18. Sitzabee WE, Hummer JE, Rasdorf W (2009) Pavement marking degradation modeling and analysis. *J Infrastruct Syst* 15(3):190–199
19. Sarasua WA, Clarke DB, Davis WJ (2003) Evaluation of interstate pavement marking retroreflectivity. Fin Rep No. FHWA-SC-03-01, SCDOT, Columbia
20. Robertson J, Sarasua W, Johnson J, Davis W (2013) A methodology for estimating and comparing the lifecycles of high-build and conventional waterborne pavement markings on primary and secondary roads in South Carolina. *Public Work Manag Policy* 18(4):360–378
21. Malyuta DA (2015) Analysis of factors affecting pavement markings and pavement marking retroreflectivity in Tennessee highways. Dissertation, University of Tennessee at Chattanooga, Chattanooga
22. James G, Witten D, Hastie T, Tibshirani R (2021) An introduction to statistical learning. Springer, New York
23. Umali J, Barrios E (2014) Nonparametric principal components regression. *Commun Stat Comput* 43(7):1797–1810
24. Kopf J (2004) Reflectivity of pavement markings: analysis of retroreflectivity curves. Res Rep WA-RD 592.1, WSDOT, Seattle
25. Karwa V, Donnell ET (2011) Predicting pavement marking retroreflectivity using artificial neural networks: exploratory analysis. *J Transp Eng* 137(2):91–103

26. Idris II, Mousa MR, Hassan M, Dhasmana H (2022) Predicting the retroreflectivity degradation of thermoplastic pavement markings with genetic algorithm. San Antonio
27. Aguinis H, Gottfredson RK, Joo H (2013) Best-practice recommendations for defining, identifying, and handling outliers. *Organ Res Methods* 16:270–301
28. Benesty J, Chen J, Huang Y (2008) On the importance of the pearson correlation coefficient in noise reduction. *IEEE Trans Audio Speech Lang Process* 16(4):757–765
29. Mk UÇAR (2019) Eta correlation coefficient based feature selection algorithm for machine learning: e-score feature selection algorithm. *J Intell Syst Theory Appl* 2(1):7–12
30. Jorge I (2011) The influence of the e-tutor on the development of collaborative critical thinking in a student's e-forum: association levels with cramer's v. In: *Old Meets New Media Educ 61st Int Counc Educ Media XIII Int Symp Comput Educ Jt Conf*, University of Lisbon, Portugal
31. Yadav D (2019) Categorical encoding using label-encoding and one-hot-encoder. <https://towardsdatascience.com/categorical-encoding-using-label-encoding-and-one-hot-encoder-911ef77fb5bd>. Accessed 15 Jul 2022
32. Agajanian S, Oluyemi O, Verkhivker GM (2019) Integration of random forest classifiers and deep convolutional neural networks for classification and biomolecular modeling of cancer driver mutations. *Front Mol Biosci* 6:44
33. De Ville B (2013) Decision trees. *Wiley Interdiscip Rev Comput Stat* 5(6):448–455
34. Karballaezadeh N, Mohammadzadeh SD, Moazemi D, Band SS, Mosavi A, Reuter U (2020) Smart structural health monitoring of flexible pavements using machine learning methods. *Coatings* 10(11):1100
35. Loh WY (2011) Classification and regression trees. *Wiley Interdiscip Rev Data Min Knowl Discov* 1(1):14–23
36. Yang L, Shami A (2020) On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* 415:295–316
37. Walczak S, Cerpa N (2019) Artificial neural networks. *Encycl Phys Sci Technol* 631–645
38. Mostafa B, El-Attar N, Abd-Elhafeez S, Awad W (2020) Machine and deep learning approaches in genome: review article. *Alfarama J Basic Appl Sci* 2(1):105–113
39. Jin M, Liao Q, Patil S, Abdurraheem A, Al-Shehri D, Glatz G (2022) Hyperparameter tuning of artificial neural networks for well production estimation considering the uncertainty in initialized parameters. *ACS Omega* 7(28):24145–24156
40. Sharma S, Sharma S, Athaiya A (2017) Activation functions in neural networks. *Int J Eng Appl Sci Technol* 4(12):310–316
41. Kotsiantis SB (2013) Decision trees: a recent overview. *Artif Intell Rev* 39:261–283
42. Badr W (2019) Why feature correlation matters.... A Lot!. <https://towardsdatascience.com/why-feature-correlation-matters-a-lot-847e8ba439c4>. Accessed 3 Apr 2023
43. Thenraj R (2020) Do decision trees need feature scaling. <https://towardsdatascience.com/do-decision-trees-need-feature-scaling-97809eaa60c6>. Accessed 3 Apr 2023
44. Roy B (2020) All about feature scaling. <https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35>. Accessed 15 2022
45. Thara TDK, Prema PS, Xiong F (2019) Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. *Pattern Recognit Lett* 128:544–550
46. Bengio Y, Goodfellow I, Courville A (2016) *Deep learning*. MIT press, Cambridge
47. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: *31st Conf. Neural Inf Process Syst (NIPS 2017)*, Neural Information Processing Systems Foundation, Inc. (NeurIPS), Long Beach
48. Pike AM, Songchitruksa P (2015) Predicting pavement marking service life with transverse test deck data. *Transp Res Rec* 2482(1):16–22

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.