Case Studies in Construction Materials 20 (2024) e03186

Contents lists available at ScienceDirect

Case Studies in Construction Materials

journal homepage: www.elsevier.com/locate/cscm

A Machine Learning based approach to predict road rutting considering uncertainty

K. Chen [a,b,*], M. Eskandari Torbaghan [a], N. Thom [b], A. Garcia-Hernández [c], A. Faramarzi [a], D. Chapman [a]

[a] School of Engineering, University of Birmingham, Birmingham B15 2TT, United Kingdom
[b] Nottingham Transportation Engineering Centre, Department of Civil Engineering, University of Nottingham, Nottingham NG7 2RD, United Kingdom
[c] Institute of Highway Engineering, RWTH Aachen University, Mies-van-der-Rohe-Str. 1, Aachen 52074, Germany

ARTICLE INFO

Keywords:
Machine Learning
Simulation
Uncertainty Quantification
Road Condition Prediction
Digital Twins

ABSTRACT

Roads as vital public assets are the backbone for transportation systems and support constant societal development. Recently, data-driven technologies such as digital twins and especially machine learning have shown great potential to maintain the service level of the existing road infrastructure by accurate future condition modelling and optimal maintenance treatment recommendations. However, the pavement community suffers from inadequate data and errors experienced in data collection, which unavoidably limits machine learning performance. In addition, focusing solely on data without considering the underlying physical behaviour remains as a challenge for the practical implementation of machine learning. To this end, this study provides a machine learning based approach to predict road rutting taking into account the machine learning uncertainties. The US Long-Term Pavement Performance public database has been used as the main data source while supplementary synthetic data was added using Finite Element simulations based on physics. The obtained results indicate that adding extra simulation data improved the model's short-term prediction accuracy by 4.4% and reduced the long-term prediction uncertainty by 6.76%. The approach could potentially mitigate the issue of lack of data and the uncertainties around the data collected, by integrating existing understanding of pavement physical behaviour into the machine learning modelling pipeline.

1. Introduction

The road infrastructure system is critical to consistent and sustainable development in any country. Given the fact that most countries are constructing new roads quite rarely, recent focus has been shifted towards better maintenance and management of existing roads to attain a high level of sustainability [1]. However, due to the limited budgets and large cost of maintenance [2], current road maintenance strategies are mostly reactive, despite being considered not as cost-effective as proactive approaches [3].

Abbreviations: 2D, Two-dimensional; AADT, Annual Average Daily Traffic; ANN, Artificial Neural Network; DTs, Decision Trees; ESALs, Equivalent Standard Axle Loads; FE, Finite Element; GPR, Gaussian Process Regression; LTPP, Long-Term Pavement Performance; ML, Machine Learning; NNs, Neural Networks; RF, Random Forest; RMSE, Root Mean Squared Error; SHAP, SHapley Additive exPlanations; SVMs, Support Vector Machines.

* Corresponding author at: School of Engineering, University of Birmingham, Birmingham B15 2TT, United Kingdom.
E-mail address: k.chen.3@pgr.bham.ac.uk (K. Chen).

https://doi.org/10.1016/j.cscm.2024.e03186
Received 5 September 2023; Received in revised form 26 March 2024; Accepted 19 April 2024
Available online 23 April 2024

Having said that, setting a proactive asset management strategy for roads is challenging for various reasons including: 1) the complexity of the road structure, 2) various factors that influence road condition degradation, 3) lack of condition data in order to have constant monitoring and tracking of the road deterioration status as well as 4) lack of an accurate and reliable road deterioration model [4].

On the other hand, emerging into the industry 4.0 era, also called the fourth industrial revolution, where increasing interconnectivity, pervasive data, machine learning and intelligent automation have brought rapid changes across different industries, multiple recent technological developments and advances have provided the potential to enable the desired proactive approach. For example, with the growing quantity of real-time data from various sources, big data analytics are leveraged to perform improved road deterioration modelling and to enable predictions of road performance with high accuracy for optimised maintenance [5].

More importantly, to appropriately utilise and benefit from available data, various Machine Learning (ML) algorithms (e.g., neural networks (NNs) and support vector machines (SVM)) have been designed and developed to make sense of the generated data and identify patterns for future performance predictions [6,7]. In particular, ML has been seen as the enabler to support automated decision-making for assets whole lifecycle management optimisation for operation and maintenance strategies [8,9]. ML techniques have also long been adopted in pavement management domain with Artificial Neural Network (ANN) as the most common approach [10].

However, ML techniques have inherent limitations. First of all, to fully utilise ML, large amount of quality data would be required to ensure consistent model performance, while avoiding bias and overfitting [11]. However, lack of data as well as low data quality have been longstanding issues for the pavement industry [12]. To overcome these issues, researchers in various domains have endeavoured to obtain more data from different sources. For example, more recently, additional synthetic data as a supplement has been generated using Finite Element (FE) simulations in which an asset or a system is modelled to increase the total available data for ML modelling. This approach has been proved to show enhanced accuracy in various domains [13,14]. The approach has also been applied in the construction sector, where improvements have been achieved in the prediction performance [15,16].

In addition, despite ML's promising modelling capacity compared to conventional statistical approaches, in general, the lack of model's interpretability and the fact that it does not obey the underlying physics such as existing governing principles or laws, based on which the asset should behave, remain as a challenging matter for ML models. This leads to generalisation issues when a ML model is applied to scenarios which it has not been trained on, and hence the ML only finds the relationship between data while completely ignoring the physical principles in the predictions [17]. To address these pressing matters, various studies recently have attempted to develop novel approaches to incorporate physical principles into ML models to ensure a robust, stable and reliable model prediction [15, 18–20]. Therefore, the overall aim of this paper is to investigate the combination of physics and ML modelling for pavement performance prediction.

Moreover, uncertainty should not be overlooked during any ML modelling process given the range of errors that exist during the development of both models such as initial data measurement error and ML model error, as well as parameter errors in FE modelling. Hence the consideration of the handling of probability and uncertainty is also a necessary layer to ensure realistic predictions for road condition deterioration [21].

To this end, this study focuses on building a ML model to predict rutting which has been considered as a major distress mode in flexible pavements [22]. Specifically, it aims to project rutting for different road sections in the short term (1 year) and long term (2–13 years) in asphalt pavements using a ML model. It considers multiple data sources such as public field inspection data, supplemented by FE simulation data according to physics as an attempt to integrate specific domain physical knowledge into the ML development pipeline for pavement rutting modelling. This study then compares the performance of the ML model with its inherent uncertainties developed with these different data origins. The remainder of the paper is organised as follows: Section 2 provides an in-depth literature review on different approaches to combine physical domain knowledge with ML as well as existing ML models for pavement performance prediction, thereby identifying a gap for the suggested approach. In Section 3, a detailed methodology of using physics to enhance ML and all the steps taken to undertake this research are described, and Section 4 presents the results and provides the analysis and discussions. Finally, the limitations and conclusions of the research as well as potential future research directions are summarised and outlined in Section 5.

2. Literature review

Considering the limitations and challenges commonly reported in ML models such as overfitting, insufficiency of interpretability and ignorance of underlying guiding physics, multiple studies have been conducted to investigate the approaches for including existing domain knowledge to enhance ML's performance across a diverse discipline such as earth systems [23], climate science [24]. These studies have demonstrated not only the improved performance and generalizations but also the consistency and credibility of ML models. For example, Daw et al. [25] combined physics with ML in a framework for lake temperature prediction by adding output from physics-based model as one feature in an ML model, refining the prediction by 38.1%, a reduction of Root Mean Squared Error (RMSE) from 1.18 to 0.73. Similarly, adding data rich of physical knowledge as an additional input for ML has also been done to predict interlamellar spacing and mechanical properties of high carbon pearlitic steel [26], and steel connection stiffness which also resulted in an RMSE drop from 0.26 to 0.16, a 38.5% improvement [27]. Another method is using modified physics-based loss function to guide the training process of NNs to enforce physical constraints which has shown improved generalizability as well as produced physically meaningful results [18,25,28]. Other approaches include informing the ML architecture with specific physics characteristics of the problem being solved, also called physics-guided design of ML architecture where physics-constrained variables as the intermediate variables are added into the ML architecture [29] or the weights have been fixed during ML training process given the known physics [30].

As for pavement performance modelling, including rutting deterioration modelling, remarkable developments in the application of different types of ML algorithms, has been observed such as NNs, decision trees (DTs), SVM, combination of them and other advanced ML algorithms. However, despite using these encouraging ML techniques, the same drawback applies when using "black-box" ML models that are agnostic to the existing underlying physics. This issue has been acknowledged by researchers in the field, e.g., [20]. For instance, in a study conducted by Alnaqbi et al., 2023 [31], the authors collected 1584 records from the US Long-Term Pavement Performance (LTPP) database compared the ML modelling performances across ANN, SVM, DTs and Gaussian Process Regression (GPR) with the results showing GPR achieving the highest accuracy (R^2 of 0.989). Despite the promising result, the authors also acknowledged the need for combining ML approach with domain specific knowledge to yield more interpretable predictions. Similar suggestions have also been made by [32], in which a fully-connected three-layered feedforward deep neural network (128−32−8) was developed and obtained R^2 of 0.82 but the authors admit the black-box nature without any direct explanation on how NNs produced a certain output, indicating the need for model's interpretability and the integration of physical domain knowledge. Some progress has been made in this regard in pavement modelling. For example, a SHapley Additive exPlanations (SHAP) approach for the interpretation on models' rutting predictions has been adopted by [21,33] with Bayesian Neural Network and Gradient Boosting Decision Tree. While the approach has led to further enhance ML model's capacity compared to traditional ML techniques as well as the ability to understand the model's prediction output on pavement performance, the SHAP value is a model-agnostic tool to explain the individual prediction output which only happens after the model development rather than considering physical characteristics [34]. This study therefore addresses the gap on the integration of domain knowledge based on physics into ML modelling process for pavement rutting prediction.

3. Methodology

As discussed in the literature review section, this study follows the methodology adopted by Daw et al. [25] where the output from FE simulation based on known physics is added as an additional input for ML model for road rutting prediction. The methodology consists of the following steps, (Fig. 1): 1) Data collection; 2) Data preparation (e.g., Pre-process, clean the data and variable selection; 3) ML model development based on processed data considering choice of the model and hyperparameter configuration; 4) Evaluate model performance through k-fold cross validation techniques on training data, and test model performance on unseen data; 5) ML model uncertainty quantification; 6) Make multi-year predictions based on the uncertainty of the ML model quantified from Step 5.

3.1. Utilised software and packages

The end-to-end research steps have been enabled by making use of multiple open-source software, computing platform, data science libraries and packages, as well as available hardware, as presented in Table 1.

3.2. Data collection

In this study, data on pavement performance was gathered from the US LTPP database [45]. The LTPP database provides detailed information of over 2500 pavement sections in the US and Canada, with various pavement types, pavement materials, ambient environmental and climate conditions, traffic, defects as well as maintenance records [46]. This database was selected due to its open access and has been used in multiple similar studies [47–49].

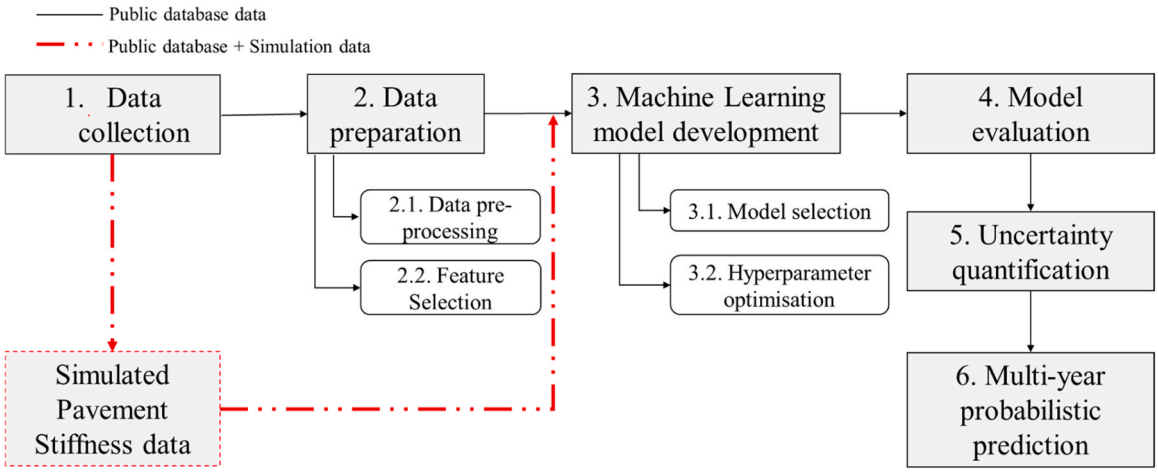


Fig. 1. Key steps taken to conduct the research.

Table 1
Utilised Software and Packages.

Software and Packages	Description	Reference
Programming language	Python 3.9.2	[35]
Computing platform	Jupyter notebook 6.1.4	[36]
Software library	Pandas 1.1.0 (Data analysis library)	[37–43]
	NumPy 1.19.5 (Library used for working with arrays)	
	SciPy 1.5.4 (Library used for scientific computing)	
	Matplotlib 3.3.2 and Seaborn 0.11.0 (Statistical data visualisation library)	
	Scikit-learn 0.23.1 (Machine Learning library)	
	Mlxtend 0.18.0 (Machine leaning extension library for data science tasks)	
Software suite	Abaqus Finite Element analysis	[44]

3.2.1. Data description

This study focuses on asphalt pavement as the most common surface type with survey condition data available for as many consecutive years as possible. Hence 99 asphalt pavement sections were selected from the LTPP database based on data availability for 13 years from 1995 to 2007. In total, 1287 pavement data records were collected in this study. After rearranging the data to obtain rutting value for the next year and removing the years with relevant maintenance records, based on the assumption that the rutting was reset to zero in the year following the relevant maintenance activity, 1152 pavement records were further analysed. The age of the pavements ranges from newly constructed to 55-year-old pavements. The geographical locations of the sections are diverse across the whole US. Fig. 2 presents details of the geographical locations of the pavement sections and the number of sections in each state.

After an initial literature review, a particular list of variables was defined to be obtained from the selected test sections in the LTPP database for this study. This list covers most factors that could influence the performance of a pavement according to existing physical understanding about road deterioration based on a review of similar studies [50]. Table 2 describes the details of the variables chosen, their associated categories and their statistical properties. Table 3 summarises the nominal variables considered in this study.

3.2.2. Data pre-processing

Once the data for selected variables was obtained, initial assessment and evaluation of the data was performed to improve its quality. Despite the completeness and comprehensiveness of the collected data, the raw data for all 99 sections from the LTPP database still suffered from several data quality issues such as missing values for certain years, duplicates, general noise, and anomalies with unreasonable data fluctuations, potentially due to measurement and human errors. For example, rutting condition sometimes slightly improves, e.g., by 1 mm, over time for some road sections without any reported relevant maintenance activities. To address these issues, multiple data pre-processing techniques, recommended by [51], were performed and a Python script was written for automatic data pre-processing and cleaning. The issues considered and corresponding techniques are described in more detail in Table 4.

Various data processing techniques have been used in this study with different purposes. Moving average and curve fitting, using the least-squares method, have been used to reduce the noise in the data and mitigate the potential measurement errors in the collected data [52]. In addition, spline interpolation method was used to ensure the completeness and the smoothness of the whole dataset [53]. Data has been stored and processed in the DataFrame which is a 2-dimensional data structure provided by Pandas data analysis library [37]. A detailed data pre-processing step flowchart, with rutting variable condition data as an example, is described in Fig. 3.

It is worth mentioning that, as shown in Fig. 3, it was assumed that road condition after a relevant maintenance treatment such as “Overlay” or “Surface Treatment” would be restored to the same service level as a newly constructed pavement in its first year in service. In other words, it has been considered that a full restoration of road defects such as rutting and longitudinal cracking occurred from the year when there is a relevant maintenance treatment.

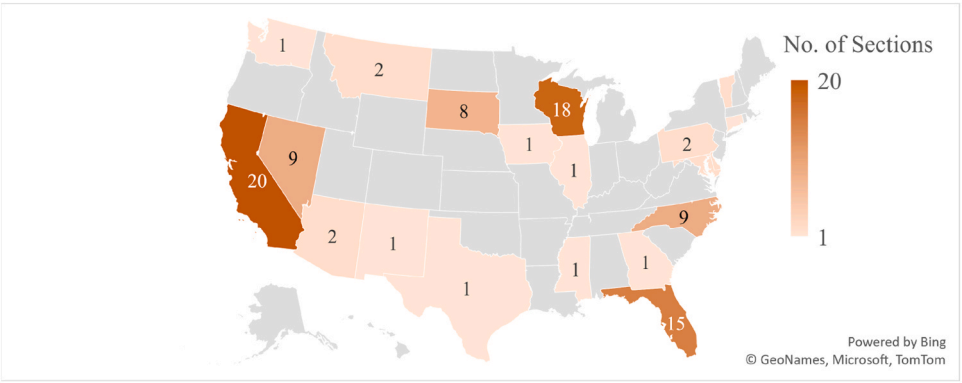


Fig. 2. Geographical locations and No. of sections.

Table 2
Descriptive statistics of the variables considered in this study.

Category	Variable	Range	Mean	Median	Standard deviation
Condition	Years from construction	0–55	27.2	30	10.07
	Rutting (t) (mm) *	0–17	4.2	3.8	2.6
	Longitudinal cracking length (m/section length)	0–123.2	5.05	0	14.5
	Transverse Cracking (Count)	0–220	21	8	32
	Fatigue Cracking (m^2)	0–816.6	26.7	0	80.6
	Rutting (t+1) (mm) *	0.5–17.2	4.5	4	2.6
Structure	Unbound foundation thickness (mm)	0–942.3	369.2	365.8	260.6
	Foundation + Asphalt thickness (mm)	185.4–1282.7	714.7	744.2	261.7
	Total asphalt thickness (mm)	0–505.5	151.2	147.3	76.3
	Dense graded asphalt thickness (mm)	0–502.9	131.6	119.4	73.9
	Open graded asphalt thickness (mm)	0–33	1.06	0	4.7
	Recycled asphalt thickness (mm)	0–167.6	15.8	0	38.1
	Emulsion-based sealing thickness (mm)	0–55.9	2.0	0	7.4
	Total foundation thickness (mm)	121.9–1135.4	563.4	602	256.8
	Bound foundation thickness (mm)	0–726.4	194.2	149.9	168.2
	Number of foundation layers	1–4	2	2	1
	Number of asphalt layers	1–11	4	4	2
Climate	Annual average ambient temperature (℃)	4.8–24	13.9	12.6	5.9
	Annual average precipitation (mm)	11.5–2070.4	797.9	845.1	500.5
	Annual average humidity (%)	51–78%	65.4	66.5	6.0
Traffic	Annual Average Daily Traffic (Equivalent Single Axle Load) - AADT (ESALs)	9–2405	585	405	531

* *t* refers to the year the data was collected; t+1 means the following year after the year the data was collected

Table 3
Description of nominal variables considered in this study.

Category	Variable	Description
Material	Contains geotextile (yes or no)	There are four pavement sections (52 records) where the material geotextile has been used with the value yes
Relevant Maintenance records	Overlay	The maintenance treatment activity - Overlay has been applied 40 times in total across 99 sections over the years between 1995 and 2007
	Surface treatment	The maintenance treatment activity – Surface treatment has been applied 13 times in total across 99 sections over the years between 1995 and 2007

Table 4
Data issues and applied pre-processing techniques.

Data issues	Impacted variables	Pre-processing technique used
Multiple entries for the same year	Rutting, All types of cracking (Longitudinal, Transverse, Fatigue)	Mean value
Missing data for some years	Rutting, All types of cracking (Longitudinal, Transverse, Fatigue), AADT (ESALs)	Moving average Curve fitting using the least-squares method Spline interpolation method
Unreasonable data	Rutting, All types of cracking (Longitudinal, Transverse, Fatigue), AADT (ESALs)	Outlier detection based on mean and standard deviation Moving average Curve fitting using the least-squares method Spline interpolation method
General data noise	Rutting, All types of cracking (Longitudinal, Transverse, Fatigue), AADT (ESALs)	Moving average Curve fitting using the least-squares method

3.2.3. Additional physics-based FE simulations data

In addition to the data collected from the LTPP database, extra data was generated by running a series of simulations based on physics-based modelling for the 99 road sections with the aim to integrate existing domain knowledge into ML. For this, all the sections have been modelled in Abaqus [44] using a two-dimensional (2D) Finite Element method, by incorporating road structure, number of layers, thicknesses, and material. The simulation produced an elastic surface deflection model under 40 kN per meter traffic loading as additional data. This was based on existing knowledge on pavement physics and the rutting severity level for each pavement section, given the established assumption that surface deflections in asphalt pavement measured in areas affected by rutting are higher than those measured in areas without rutting defect [54].

The assumptions made for the simplified 2D models are as follows:

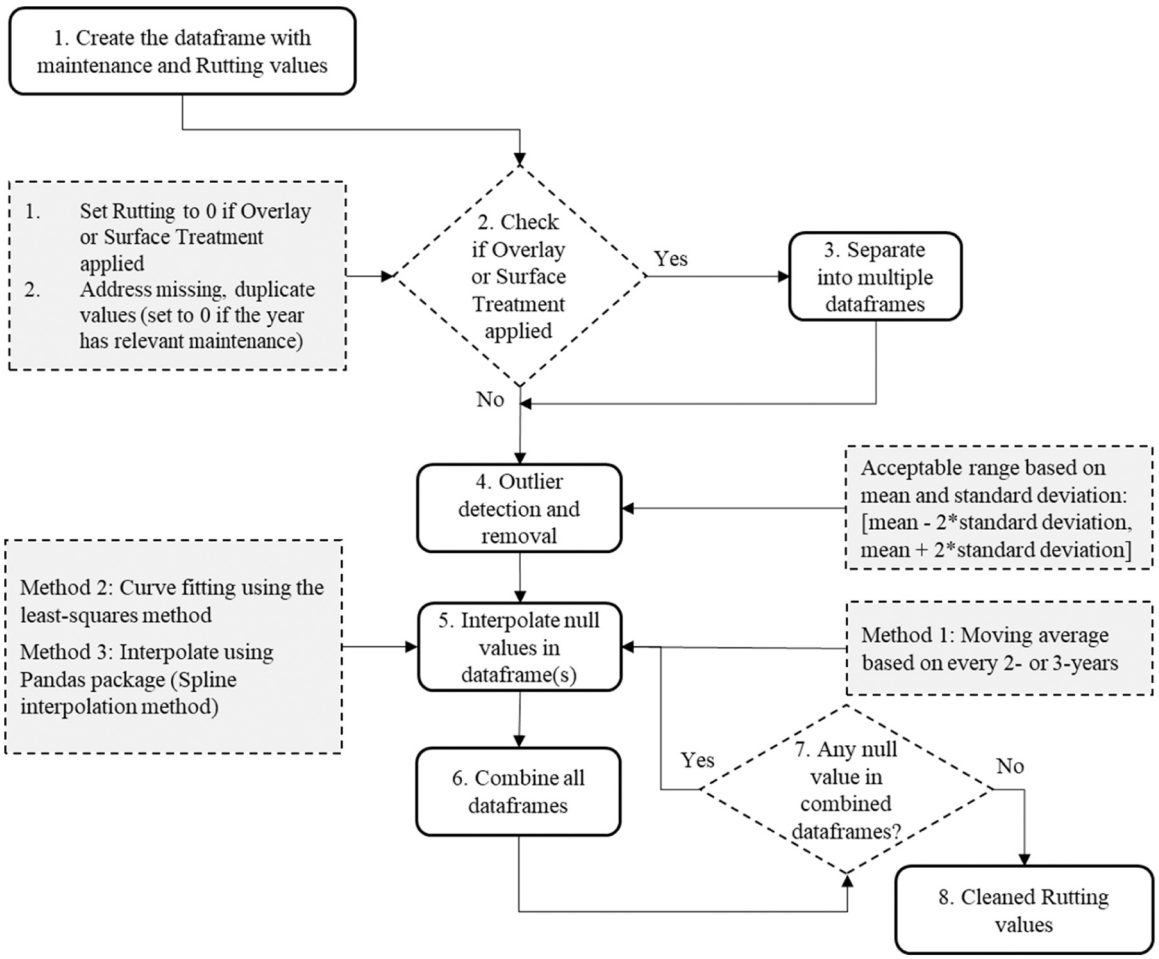


Fig. 3. Data pre-processing procedures on rutting condition data.

1) All material properties (Young's modulus and Poisson's ratio) for different layers of all pavement sections stay constant over the years.
2) Abaqus models for all the pavement sections have been created following identical procedures including the loading area, and the way it was meshed (see Appendix A for more details).
3) Subgrade layer has thickness of 200 mm

The materials used across different layers in the LTPP sections vary significantly; in total there are more than 50 types of material. More details on the Abaqus models can be found in Appendix A. Given the materials of road sections in this research vary significantly across different layers, and it is not in the scope of this research to identify accurate properties of the materials, Appendix B presents estimates of the Young's modulus and Poisson's ratio values defined for all material types used in the models based on expert judgements.

3.3. Variable selection process

3.3.1. Exhaustive variable selection process

Variables election was necessary to identify the optimal subset of input variables from the initial list of model inputs presented in Tables 2 and 3. There are three main approaches in conducting variable selection, namely: wrapper, filter, and embedded methods [55]. In this research, the method was adopted since it allows for an unambiguous understanding of the effectiveness of every single included variable. More specifically, exhaustive variable selection algorithm has been utilised because it concentrates on retrieving all possible combinations of the model inputs and gives priority to create a subset of inputs based on the performance quality of an algorithm [56], linear regression in this case which requires the least time and computing resources.

Despite being a computationally expensive method, considering the relatively small total number of variables, and the fact that there is a much stronger emphasis and need to understand the optimal variables as model inputs, the wrapper method was selected. This is to ensure better comprehension and interpretation of the factors impacting pavement rutting development.

All combinations of 21 initial variables available from Tables 2 and 3 were tested as part of the exhaustive variable selection method. Maintenance treatment was not selected as a variable, i.e., input, because roads after relevant maintenance treatment were treated as new sections. Mlxtend library [43] was implemented to perform exhaustive feature selection that included all possible combinations of the 21 model inputs interactions with input number ranging from 1 to 21, to build a linear regression model and then compare the models' performances to select the one that results in the best performance (e.g., the least mean squared error). In this study, the number of input variables is 21 which results in 2097151 ($2^{21} - 1$) numbers of combinations, which means 2097151 linear regression models have been built as part of the variable selection process.

Each of the 2097151 linear model has been evaluated through a 10-fold cross validation approach which calculated the average mean squared error of the model's performance from all iterations. Table 5 reports the best five variable combinations that yielded the least average mean squared error across the 10-folds. The negated mean squared error is simply the negated value of the mean squared error, and this is due to a convention in the Scikit-learn ML software package [42] where higher return values are better than lower return values.

The first result with the combination of variables that produced the highest negated mean squared error (i.e., the least mean squared error) from Table 5 was selected to give the variables for ML model development.

As it may be easily noticed that the results after exhaustive variable selection did not include any traffic volume or load which has been commonly considered as a direct cause of rutting, the reason could be that the traffic is heavily related to the road design with the number of foundation and asphalt layers, thicknesses, and the quality of the materials used. Regarding environmental factors, the results may indicate that the impact of ambient temperature, precipitation could be covered by humidity.

3.4. Model development

3.4.1. Random forest model

Random forest (RF) is a supervised ML algorithm that is used widely in classification and regression problems because of its high

Table 5
Top five exhaustive variable selection results.

Rank	Common Variables	Variables specific to a particular model	Average negated mean squared error score
1	Rutting (t) (mm); Unbound foundation thickness (mm); Contains geotextile (yes or no);	N/A	-0.56826
2	Longitudinal cracking length (m); Annual average humidity (%); Number of asphalt	Fatigue (m2);	-0.568703
3	layers; Emulsion-based sealing thickness (mm); Number of foundation layers.	Annual average ambient temperature (°C);	-0.568832
4		Annual average precipitation (mm);	-0.569025
5		Annual average precipitation (mm); Fatigue (m2);	-0.569451

prediction accuracy [57–59]. It is one of the decision tree algorithms where RF builds multiple decision trees and combines the results from each tree together to get a more generalised and accurate result.

In this study, RF regression was used to construct models to predict the value of rutting the next year given the defined inputs. RF uses Bootstrap and Bagging Aggregation ML techniques [60]. Its foremost advantage is the fact that it effectively deals with overfitting issues by joining multiple sub-datasets, while it requires less time for processing data when compared with other methods [61]. The following steps were taken to utilise the RF regressor algorithm:

- Step 1) The whole dataset was used to build decision trees based on the number of defined estimators;
- Step 2) Individual decision trees were constructed;
- Step 3) Each decision tree generated an output;
- Step 4) Final output was considered based on averaging for regression;

3.4.2. Model inputs for two developed scenarios

Two scenarios with different model data inputs were defined for model development and evaluation. Scenario 1 used input data from the variable selection process, whereas Scenario 2 expanded the selected input data by integrating the additional data from the corresponding physics-based FE simulation models previously mentioned.

Scenario 1 is a standard ML modelling process, also known as black-box model, given a list of inputs X and output Y (X ➡ Y). Scenario 2 differs from scenario 1 in its use of physics-based FE model simulations as input attribute, Y_{phy}, which can be represented as ($X + Y_{phy}$ ➡ Y). In Scenario 2, the simulated surface deflection was produced based on physical relationships between the load and the pavement structure using physics-based FE models. Furthermore, Y_{phy} may not offer a complete representation of the target output Y due to simplified or missing physics, resulting in model discrepancies. To avoid any systematic bias in Y_{phy} in comparison to the actual performance of pavement, the physics output as an extra input can be added, so the modelling process can learn to complement Y_{phy} by leveraging and extracting information from the space of existing inputs and thus reducing the knowledge gap.

3.4.3. Model development and evaluation of one-year prediction

In this study, the Scikit-learn ML package [62] was used for developing RF regression models. Grid Search method was used for tuning the hyperparameters used for the RF regression model in this research for both scenarios considering its exhaustive searching method that could find the optimal hyper-parameter values by checking all parameter combinations [63]. The RF hyperparameter space was defined and the final configurations are summarised in Table 6.

To assess the one-year prediction performance of each model, metrics such as Coefficient of determination (R^2) and Root Mean Squared Error (RMSE) were used, Eqs. 1–2, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}(\widehat{y_i} - y_i)^2}{\sum_{i=1}(\widehat{y_i} - \overline{y_i})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n}} \tag{2}$$

where $\widehat{y_i}$ is the predicted value from the ML model, y_i is the actual observed value, n is the total number of observations, and $\overline{y_i}$ is the average of the measured values.

To ensure the model generalisation capacity, out of the 99 road sections 15 sections (approx.15% of the total) were randomly selected and used for the final test of model performance based on suggestions made by [64,65]. The remaining 84 sections were used to train and validate the model. To avoid overfitting, which is a common issue for ML algorithms [66], a cross-validation technique was applied on the 84 sections with a configuration of 10 folds, which is the common configuration practice for cross-validation [67]. Additionally, during the random selection of training and testing sections, basic statistical analysis such as mean and standard deviation on each model variable was performed to ensure the validity of the model by checking the distribution similarity between training and testing data. The model evaluation outcome is the median value out of the results from 10 folds. Once the model was evaluated, all data from 84 sections were trained to build a model which was tested against the 15 sections to understand the performance of the model. To further reduce the bias and randomness during the selection of the sections used for model training and testing, the whole process was repeated 30 times for both scenarios, each with a different set of 15 test sections, and this was used to

Table 6
RF hyperparameter space and tuning results.

Hyperparameters	Space	Default	Tuning results (Scenario 1)	Tuning results (Scenario 2)
n_estimators	[50, 100, 150, …, 450]	100	100	50
max_depth	[3, 4, 5, …, 49]	None	5	38
min_samples_split	[2, 4, 6, 8, …, 28]	2	12	18
min_samples_leaf	[1, 2, 5, 10, 50, 100]	1	5	2

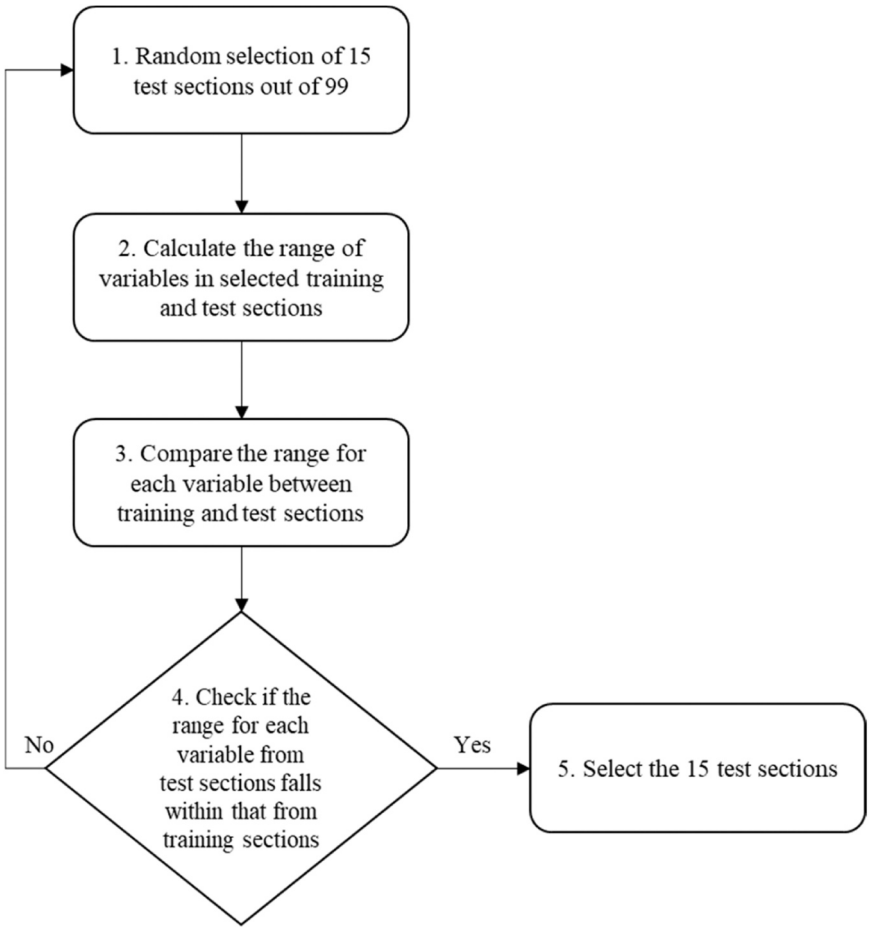


Fig. 4. Selection process of the 15 test sections.

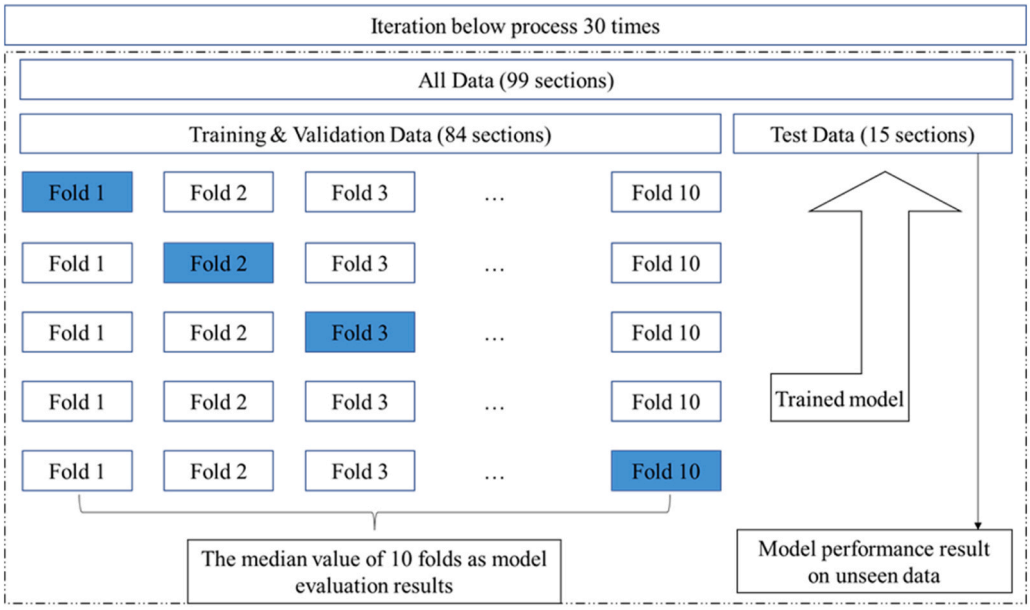


Fig. 5. RF model training, evaluation, and test iteration process.

provide an indication of the models' generalisation performance capacity. The selection of 15 test sections, was made based following the procedure described in Fig. 4 for each iteration to ensure the optimum ML learning experience. The process was developed based on the common technique when it comes to machine learning models to ensure the meaningful selection of training and testing sections so that the testing data range stays within the range of the data used to train the model to mitigate the overfitting issues [68,69].

The number of times the model was run was chosen to be 30 because this number has been used and recommended in other studies with the same approach [70–72] in order to gain enough statistical information on the variance to understand the model uncertainties. Fig. 5 shows the whole process of the RF model training, cross-validation evaluation, and test.

3.4.4. Model uncertainty quantification

Prediction intervals offer a method to quantify and communicate the probability for a single prediction point which is an estimation or approximation with some level of uncertainty. Prediction intervals can reveal the level of accuracy and confidence of model prediction. The uncertainties and potential errors in general come from two main sources: 1) ineffectiveness and error of the model itself, and 2) the noise contained in the initial measurement data before or even after the data cleaning process. In this study, the uncertainty is expressed through the variance of ML model due to different sampling of the training and testing data. Common model evaluation methods such as R^2 and RMSE fail to address the prediction confidence for individual instances [73]; In such cases the point estimation is insufficient for the forecasting of the prediction uncertainties. Therefore, prediction intervals were generated in this study to understand the precision and accuracy of the model prediction.

Prediction intervals were calculated to understand the model performance accuracy based on ML model results after 30 repetitions. Following the process described in Fig. 5, a list of predictions from the 30 RF model results was obtained and compared against the actual results to understand prediction distributions by generating prediction intervals. The generated prediction intervals quantified the uncertainty for one prediction point value by identifying range of prediction with a certain likelihood through calculating a linear regression fit of the predicted values and the actual values and the standard deviations of the residuals between linear fit line prediction and the predicted values [74].

3.4.5. Model evaluation of multi-year predictions

The multi-year predictions were generated by leveraging bootstrapped residuals based on the collection of errors already experienced in the 30 model results. This method assumes the future predicted errors would be similar to the existing ones [75,76]. By taking a uniformly distributed sample from the collection of existing sorted residuals following a cumulative distribution function and appending to the model's predicted results, different prediction outcomes were produced. Through doing this repeatedly, a collection of slightly different one-year predictions was generated. By building model inputs based on the predicted output from the previous year, multi-year predictions were then created according to the same process. A Python script was run 2000 times to generate the ranges with lower and upper bound for multi-year predictions. The 90% probability prediction intervals for the 2000 results for any single year were then computed to understand the 90% percentile prediction range for that particular year. The results for the 12th year 90% probability prediction intervals, as an example, based on this approach are presented in Section 4.

4. Results and discussion

This section provides results on model's performance from both scenarios for 1) one-year prediction 2) prediction interval considering ML uncertainty 3) multi-year probabilistic predictions, and gives insightful analysis and discussions on the overall improvement comparing the scenarios.

4.1. Model performance for one-year deterministic prediction for both scenarios

For each model generation iteration, results of model evaluation using 10-fold cross-validation on 84 randomly selected sections and the result of model performance on the randomly picked 15 unseen test sections were acquired. Without considering the uncertainty of the ML model caused by different training datasets, model's performance results are summarised in Table 7, and model's 1-year prediction performance on the testing datasets is shown in Fig. 6.

For one-year prediction, Fig. 6 shows that Scenario 2 improves the R2 performance compared to Scenario 1 from 90.3% to 94.2%,

Table 7
Model performance for 1-year prediction.

1-Year prediction	Results from 10-fold cross-validation on Training and Validation datasets (R^2)	Results from Testing dataset (R^2)
Scenario 1 (ML with LTPP data)	Median: 93.1% Max: 95.5% Min: 88.3%	90.3%
Scenario 2 (ML with LTPP + Simulation data based on physics)	Median: 93.6% Max: 94.6% Min: 83.1%	94.2%

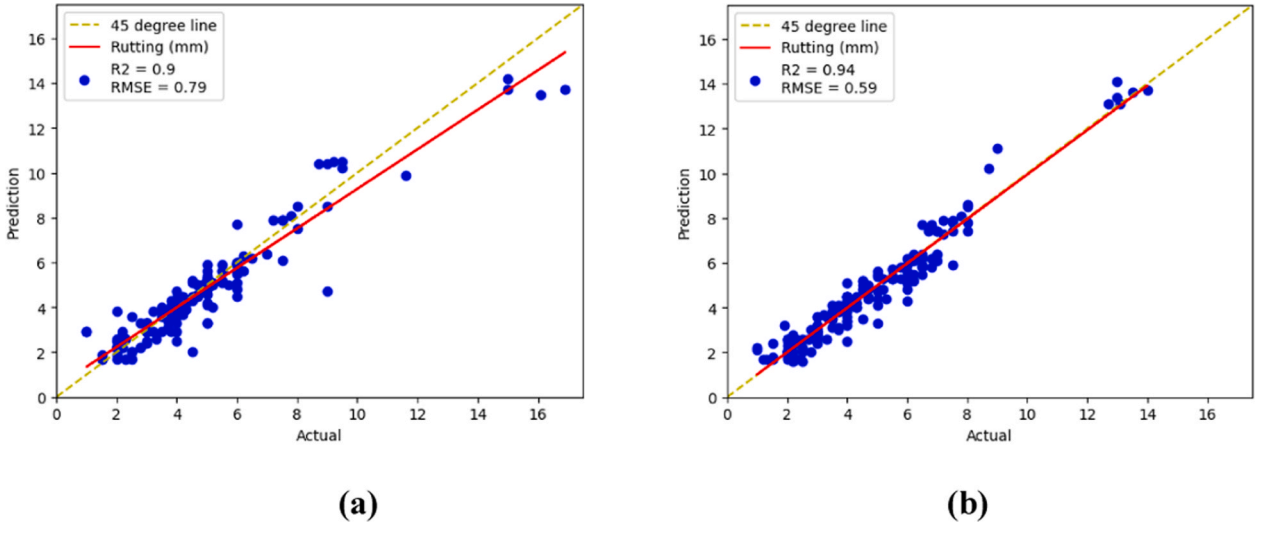


Fig. 6. One-year model deterministic prediction results on test sections using a) LTPP data only and b) LTPP + simulation data based on physics.

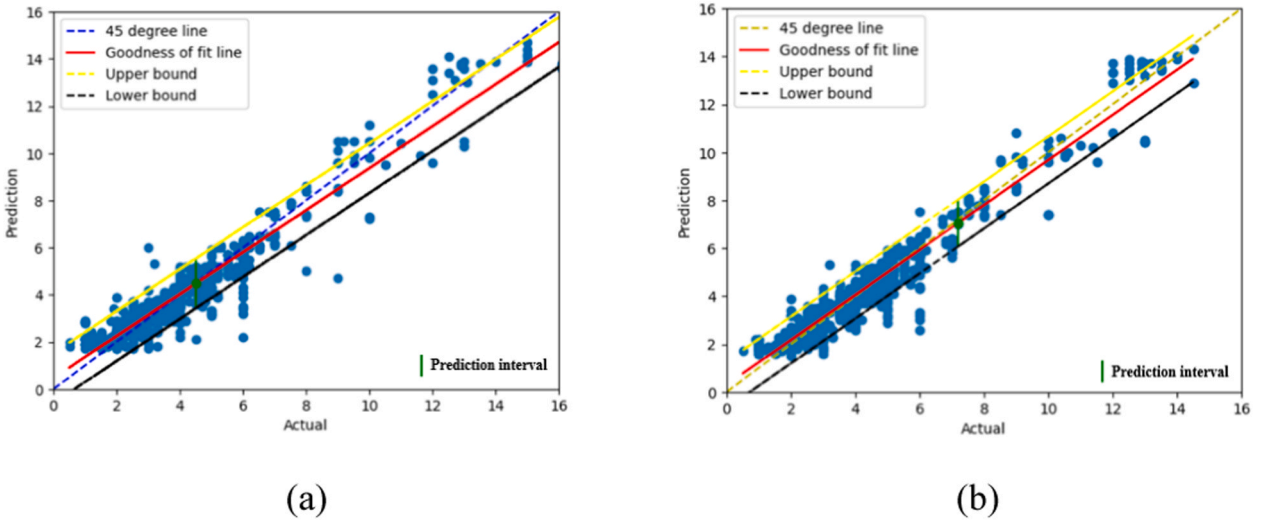


Fig. 7. Prediction intervals with 90% confidence level on test sections for a) Scenario 1 and b) Scenario 2.

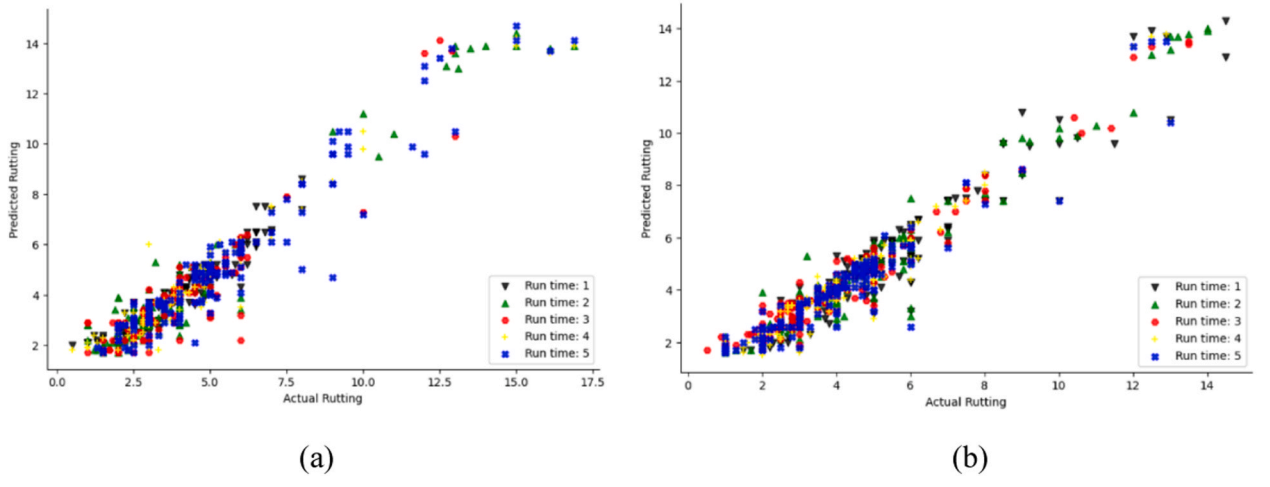


Fig. 8. Visualisation of Prediction intervals with 90% confidence level on test sections for a) Scenario 1 and b) Scenario 2 after five runs of the model development process with different training data each time. Each colour represents the result of one run.

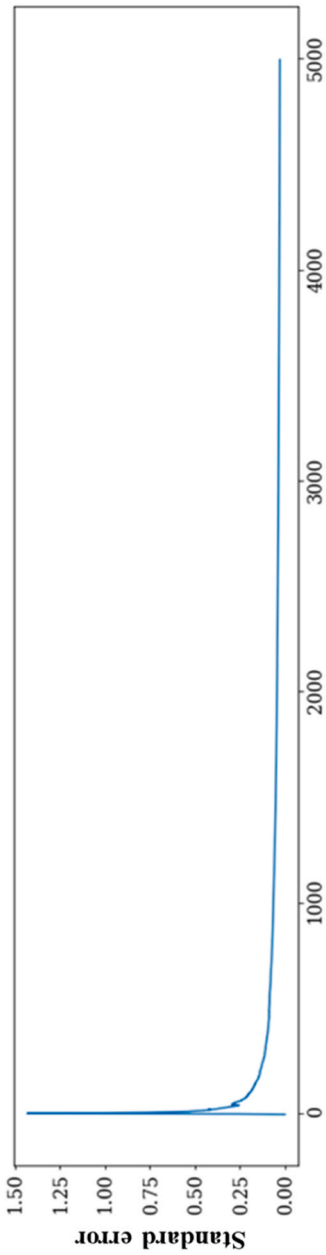
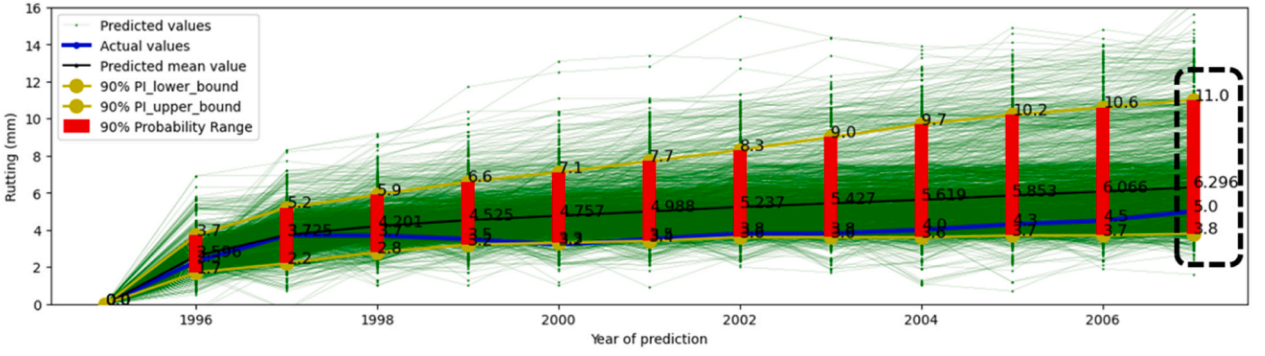
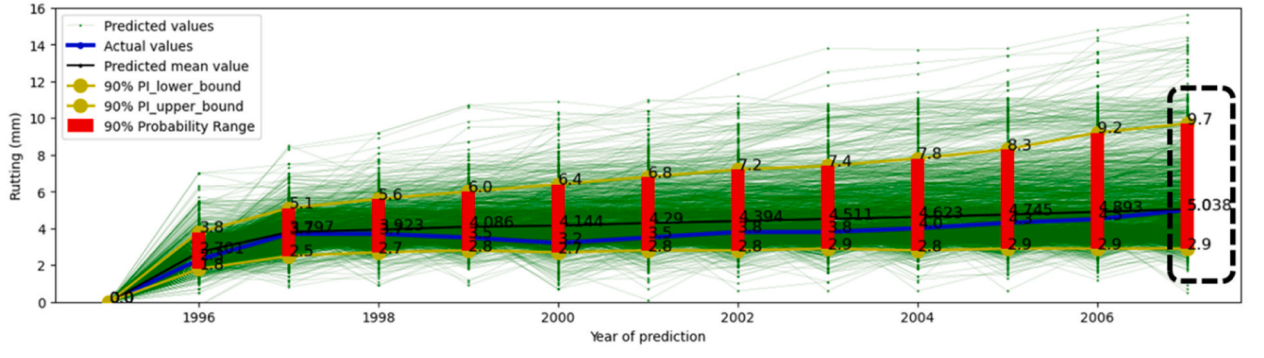


Fig. 9. Standard error vs. number of runs.



(a)



(b)

Fig. 10. Multi-year predictions of one model for Section 12–0566 a) Scenario 1 and b) Scenario 2 considering ML uncertainties.

an increase of 4.4%. Meanwhile, the RMSE decreased by 25.3%, from 0.79 to 0.59. This implies a model accuracy enhancement for one-year projection with data from physics-based FE simulations. It is worth noting the significant decrease on RMSE value despite little improvement regarding R2. This finding is in consistency with other studies in which similar approach was adopted in the prediction of steel connection stiffness [27] and lake temperature [18].

4.2. Prediction intervals

In order to quantify the uncertainty of the ML model, prediction intervals with 90% confidence level were produced. For visualisation purposes, the model development process was repeated five times to clearly show distinct model performances as a result of multiple runs with different training data. Fig. 7 shows the prediction interval for both scenarios, and Fig. 8 provides the visualisation of model performance on testing sets after five runs. The results illustrate that the one-year prediction interval decreased for Scenario 2 in comparison to Scenario 1 from the range of the fitted regression result $\pm$ 1.056 to $\pm$ 0.980. This means the predictions produced by the ML model increased their precision. This indicates that Scenario 2, with additional data generated by physics-based FE simulations, has contributed to a decrease of 7.2% in the model's one-year prediction interval range.

The visualisation shows the ML model generalisation capacity that it could yield stable results for both scenarios under varied training data. This ensures the robustness of the model performance with data from pavement sections with complete different internal and external characteristics. The reduction in the prediction interval with 90% confidence level reassures having additional FE data improves the ML model's prediction accuracy. This inherent ML model uncertainty, after being quantified, can be used to enable the model to make multi-year predictions by incorporating the underlying quantified uncertainty in the predictions year by year.

4.3. Multi-year prediction results based on the prediction intervals

To achieve multi-year predictions while taking account of the ML's underlying uncertainties, an alternative method based on residual bootstrap was used to calculate the prediction intervals, thus generating 90% percentile simulation ranges for each year. Road section 12–0566 was selected as an exemplary use case for this approach due to the fact that the data observed for this section started from 0 following a major maintenance and it also followed a pattern similar to the typical pavement deterioration curve over the years [77–79]. Two thousand sample predictions were generated for each year. This number was determined based on the standard error results (Eq. 3) from these sample predictions. Between 2000 and 5000 runs, a low and stabilised standard error was observed (Fig. 9), indicating that the number of sample predictions is enough to represent the whole population of the predictions [80].

$$SE = \frac{\boldsymbol{\sigma}}{\sqrt{n}} \tag{3}$$

where SE is the standard error of the sample predictions; $\boldsymbol{\sigma}$ is the sample predictions standard deviation; and n is the number of sample predictions.

Fig. 10 then presents the results for both scenarios with the multi-year prediction ranges and the actual values throughout the years using one of the 30 developed models for this section as an example with the bootstrapped residuals method.

Considering the collection of uncertainties made from 30 different trained models, the multi-year prediction was also run 30 times to ensure 30 different models with different training sections were used to compare between both scenarios for model generalisation purposes. Table 8 displays all the results based on all 30 models. Table 8 shows that for the majority of the 30 different models, the inclusion of the extra simulated data according to first physical principles improved the long-term (the 12th year) prediction precision

Table 8
. The 12th year prediction details with 30 models for both scenarios.

Scenario 1 (ML)	Scenario 2 (Hybrid)	% of Range Reduction from Scenario 1–2	Scenario 1 (ML)	Scenario 2 (Hybrid)	% of Range Reduction from Scenario 1–2
1 – 15 Model Results (90th Percentile Range)			16 – 30 Model Results (90th Percentile Range)		
7.2	6.8	5.56%	7.1	6.3	11.27%
7.1	6.2	12.68%	7	6.1	12.86%
6.7	6.7	0.00%	7.5	6.8	9.33%
7.5	6.6	12.00%	7	6.3	10.00%
6.8	6.8	0.00%	6.9	6.6	4.35%
7.2	6.6	8.33%	7.1	6.8	4.23%
6.9	6.9	0.00%	7.5	6.3	16.00%
7.3	6.5	10.96%	7.5	7.3	2.67%
7.1	6.2	12.68%	7.5	6.8	9.33%
7.2	6.6	8.33%	7.5	6.5	13.33%
7	6.8	2.86%	7.2	6.7	6.94%
6.6	6.5	1.52%	7	6.5	7.14%
7.3	6.4	12.33%	7.6	7.8	-2.63%
7.2	6.7	6.94%	7.4	6.8	8.11%
7.4	7.4	0.00%	6.8	7.1	-4.41%
Average % of Range Reduction from Scenario 1–2	6.76%				

by between 1.52% and 16% regarding the 90th percentile of the predicted values, while acknowledging the fact that two model results showed the potential risk of decreasing the model performance and four models' results showed no improvement. But on average, Table 8 indicates the 90th percentile range has been narrowed down by 6.76% from Scenario 1 to Scenario 2, showing an improvement in the prediction accuracy for the 12th year.

The example presented in Fig. 10 shows the 90th percentile range of the model's probabilistic predictions increases over the years for both scenarios, and looking into the predictions for the 12th year, Scenario 2 with the extra physics-based FE simulation data, improved the whole prediction compared to Scenario 1 by reducing the rutting ranges from 7.2 to 6.8.

It can also be observed that the actual values in Fig. 10 are quite distant from the centre of the confidence intervals in the prediction. This can be explained in that the actual values (the blue line) are from one particular section and the predictions (green lines) were generated based on the ML model and its associated uncertainties in the modelling process using data from all training sections. The uncertainty includes multiple sources such as measurement errors in the data collection process, as well as the ML modelling uncertainties due to the selection of training data during repetitions and the variety of road sections from which the data has been collected. Hence, Fig. 10 demonstrates the successful reduction of uncertainties in the multiple-year prediction ranges through the supplement of physics-based numerical modelling deflection data into the ML model development procedures.

Both results presented in Tables 7 and 8 demonstrated that Scenario 2 in which ML is supplemented with physics-based FE simulation data in general has made a relatively small improvement in the accuracy and reliability of the model for one year or multi-year predictions. This limited advancement could have been because of the simplistic physics-based FE models which may not fully reflect the complete realistic physics being modelled and plus various assumptions that had to be made in FE simulation modelling process, such as material properties, which do vary with different construction methods, the age of material and subgrade properties. More advanced simulation could potentially improve the outcome of this framework which is also included as one of the future works.

The key findings of this study can be summarized as follows:

- These results suggest that this combined approach of enriching the dataset from the public LTPP database with physics-based FE simulations based on the information available could result in 1) an increase in the ML model's short-term prediction accuracy; 2) a decrease in the uncertainty in the ML model's prediction ranges (both short-term and long-term).
- Especially for multi-year predictions, the addition of physics-based FE simulation data has helped the ML model to decrease the prediction ranges each year. From a roads agency perspective, this increase in prediction confidence could result in reduced maintenance spend, traffic delay and congestion and more accurate financial forecasting for multi-year investment periods without necessarily requiring additional data collection.
- While various ML algorithms have shown a decent pavement performance prediction accuracy in previous research, the lack of data and data quality has prevented the pavement community from fully unlocking ML capacity. These results are encouraging, partially overcoming this issue by generating synthetic data using physics-based FE simulations. The same approach could be used to model other meaningful pavement performance indexes or defects, giving the potential to enhance ML model prediction accuracy.
- The study has also opened up a new research direction where ML models are integrated with Simulations for pavement performance prediction. While this integration was only at the data level in this research, future studies can explore advanced combinations such as physics-informed ML training process to enhance a model's performance further as well as its interpretability.

5. Conclusion

This paper has presented a novel ML-based approach integrated with domain physical knowledge to predict road rutting considering ML inherent uncertainties. More specifically, a RF regressor algorithm was used to build the ML model with optimal hyperparameters.

A comprehensive dataset from 99 sections was collected from the US LTPP database as the main data source, and the output (pavement surface deflection) from physics-based FE simulations has been added to collected data as model inputs. The exhaustive variable selection method has been performed to identify the optimal combination of variables to use as the inputs for ML model development.

ML model uncertainties have been taken into consideration expressed in the form of variance, and this uncertainty has been quantified by executing the model development 30 times with different training and testing data in each run to guarantee the model's generalisation capacity, demonstrating the prediction ranges of the model for both short and long term.

ML model performances have been compared on two scenarios 1) without the extra FE simulation data based on physics 2) with the extra FE simulation data based on physics. The results in this study show that the ML model performance for 1-year prediction was around 90.3% R^2 and this was improved to 94.2% with the additional FE simulation data. When it comes to ML long-term prediction, the inclusion of extra FE simulation data improved the 12th year prediction accuracy by 6.76% on average considering a 90% probability prediction range compared to only data from the database.

This approach could potentially 1) mitigate the data shortage issue for pavement management, and 2) lead to further research on advanced ML model development guided by expert pavement deterioration knowledge such as ingesting specific pavement domain knowledge in the loss function of NNs, and 3) to be integrated into a pavement digital twin environment where a large amount of data with high frequency and variety can be fully utilised for better road asset management leveraging ML and physics-based simulations.

Several future research directions can be summarised below. From database point of view, it would be worthwhile to incorporate additional real-world data sources in addition to the LTPP used in this study. And from the ML perspective itself, more advanced ML techniques could be implemented to understand the impact of such a combined approach. When it comes to integrate pavement

physics into ML to overcome its limitations, different methods of utilising the physics-based model could be investigated such as customisation of the loss function of an ANN. More importantly, to further the scientific development in the area of combining domain knowledge with ML in general, exploring interdisciplinary approaches or collaborations could also lead to more robust and innovative research outcomes. Last but not least, a more complex physics-based FE model representing full pavement physics including variations of different layers over the years and its interaction with other assets could likely make a more significant enhancement to the framework.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Data availability

Data will be made available on request.

Appendix A. : Finite element (Abaqus) pavement model

Abaqus is a commonly used engineering simulation software suite based on finite element method, it possesses robust computing function and extensive simulated performance, as well as providing a huge number of multiple element models, material models and analytic processes [81].

Modelling was performed on the following modules in Abaqus: Part, Property, Assembly, Step, Interaction, Load, Mesh, Job, Visualization, Sketch. Meshing is an important process in which the Abaqus model solves the differential equations of the system models, by discretizing the model to smaller nodes and elements [44].

Considering the large number of Abaqus models developed in this study, a simplistic linear elastic 2-D pavement model simulating surface elastic deflection under 40 kN per meter in the third dimension for the corresponding pavement structure and materials for each section has been created to avoid huge computational time and resource. Young's modulus and Poisson's ratio have been defined as the elastic properties of the materials in different layers in this study.

Ninety-nine Abaqus models, corresponding to the number of considered road sections, were created according to the specific pavement structures in the relevant test section. The cross section of each pavement test section has been modelled with a 3.66 m width. The pavement model structure for Section 12–0566 is shown in Figure A.1. as an example. Boundary conditions and mesh techniques were consistent across all sections. Vertical displacements are the results produced from the models. Figure A.2. displays a) loading area and boundary conditions, b) meshing of layers and c) results of the vertical displacement from the finite element model for Section 12–0566. All material properties (i.e., Young's Modulus and Poisson's Ratio) have been defined in Table B.1. in Appendix B based on pavement engineering expert opinion.

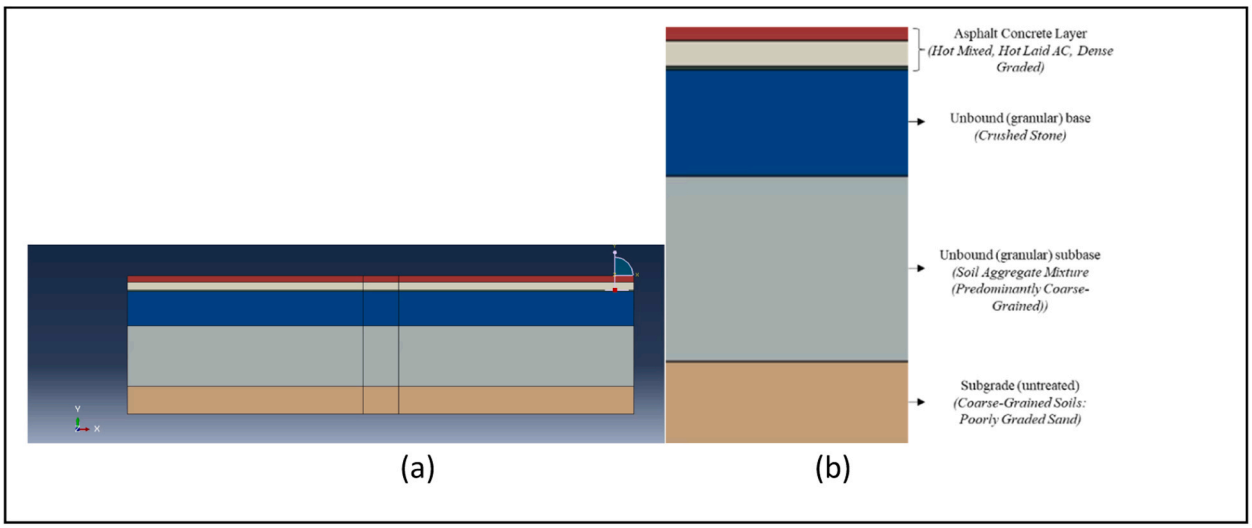


Figure A.1. a) Pavement model and b) its structural layers

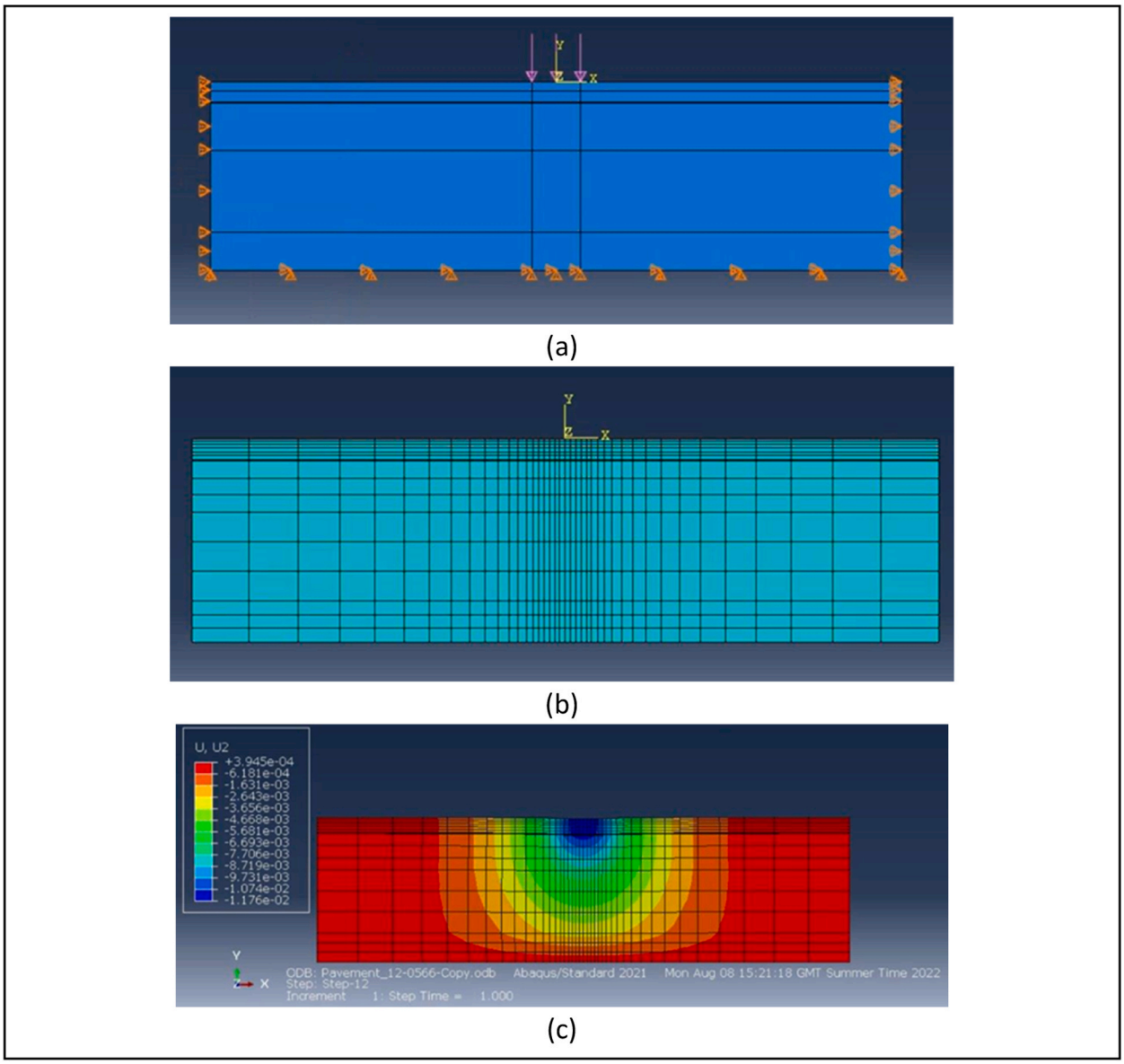


Figure A.2. Numerical model with a) its boundary conditions, b) meshing of layers, and c) results of vertical displacement.

Appendix B. : Material codes and characteristics in US LTPP Database

Table B.1. Material codes and characteristics

Material code	Material Code Description	Young's modulus (GPa)	Poisson's ratio
1	1-Hot Mixed, Hot Laid AC, Dense Graded	5	0.3
2	2-Hot Mixed, Hot Laid AC, Open Graded	3.5	0.3
4	4-Portland Cement Concrete (Jointed Plain Concrete Pavement)	41	0.2
5	5-Portland Cement Concrete (Jointed Reinforced Concrete Pavement)	41	0.2
6	6-Portland Cement Concrete (Continuously Reinforced Concrete Pavement)	41	0.2
9	9-Plant Mix (Emulsified Asphalt) Material, Cold Laid	2.5	0.3
13	13-Recycled AC, Hot Laid, Central Plant Mix	3.5	0.3
14	14-Recycled AC, Cold Laid, Central Plant Mix	3	0.3
20	20-Other	3.5	0.3
71	71-Chip Seal	Same properties as the layer below it	

(*continued on next page*)

(*continued*)

Material code	Material Code Description	Young's modulus (GPa)	Poisson's ratio
72	72-Slurry Seal		
73	73-Fog Seal		
74	74-Woven Geotextile		
75	75-Nonwoven Geotextile		
77	77-Stress Absorbing Membrane Interlayer	1	0.35
81	81-Chip Seal with Modified Binder	Same properties as the layer below it	
82	82-Sand Seal		
83	83-Asphalt-Rubber Seal Coat		
102	102-Fine-Grained Soils: Lean Inorganic Clay	0.2	0.3
108	108-Fine-Grained Soils: Lean Clay with Sand	0.1	0.2
109	109-Fine-Grained Soils: Fat Clay with Sand	0.2	0.2
111	111-Fine-Grained Soils: Gravelly Lean Clay	0.13	0.2
114	114-Fine-Grained Soils: Sandy Lean Clay	0.2	0.3
117	117-Fine-Grained Soils: Gravelly Lean Clay with Sand	0.11	0.3
131	131-Fine-Grained Soils: Silty Clay	0.1	0.3
135	135-Fine-Grained Soils: Sandy Silty Clay	0.2	0.3
141	141-Fine-Grained Soils: Silt	0.2	0.3
142	142-Fine-Grained Soils: Silt with Gravel	0.2	0.2
144	144-Fine-Grained Soils: Gravelly Silt	0.2	0.2
145	145-Fine-Grained Soils: Sandy Silt	0.2	0.2
201	201-Coarse-Grained Soils: Sand	0.13	0.3
202	202-Coarse-Grained Soils: Poorly Graded Sand	0.12	0.45
204	204-Coarse-Grained Soils: Poorly Graded Sand with Silt	0.2	0.3
211	211-Coarse-Grained Soils: Well-Graded Sand with Silt and Gravel	0.2	0.25
214	214-Coarse-Grained Soil: Silty Sand	0.15	0.42
215	215-Coarse-Grained Soil: Silty Sand with Gravel	0.2	0.4
216	216-Coarse-Grained Soil: Clayey Sand	0.2	0.3
217	217-Coarse-Grained Soil: Clayey Sand with Gravel	0.2	0.15
265	265-Coarse-Grained Soil: Silty Gravel with Sand	0.2	0.4
266	266-Coarse-Grained Soil: Clayey Gravel	0.2	0.3
267	267-Coarse-Grained Soil: Clayey Gravel with Sand	0.2	0.35
302	302-Gravel (Uncrushed)	0.25	0.35
303	303-Crushed Stone	0.25	0.35
304	304-Crushed Gravel	0.25	0.35
306	306-Sand	0.08	0.2
307	307-Soil-Aggregate Mixture (Predominantly Fine-Grained)	0.2	0.35
308	308-Soil-Aggregate Mixture (Predominantly Coarse-Grained)	0.2	0.35
309	309-Fine-Grained Soils	0.2	0.2
310	310-Other (Specify, if Possible)	0.2	0.3
319	319-HMAC	5	0.3
320	320-Sand Asphalt	2	0.3
321	321-Asphalt Treated Mixture	2	0.3
325	325-Open Graded, Hot Laid, Central Plant Mix	3.5	0.3
331	331-Cement Aggregate Mixture	20	0.1
338	338-Lime-Treated Soil	0.2	0.13
339	339-Soil Cement	1	0.13

References

[1] S. Burningham and N. Stankevich, "Why road maintenance is important and how to get it done," The World Bank, vol. Transport Note, no. 121, pp. 535–546, 2005.

[2] UK House of Commons, "Local roads funding and maintenance: filling the gap," 2019.

[3] S. Bowden, A. Dorr, T. Thorpe, C. Anumba, Mobile ICT support for construction process improvement, Autom. Constr. vol. 15 (5) (Sep. 2006) 664–676.

[4] M. Eskandari Torbaghan, et al., Robotic and autonomous systems for road asset management: a position paper, Proc. Inst. Civ. Eng. -Smart Infrastruct. Constr. vol. 172 (2) (Sep. 2020) 83–93.

[5] S.M. Piryonesi, The Application of Data Analytics to Asset Management: Deterioration and Climate Change Adaptation in Ontario Roads, University of Toronto, Canada, 2019.

[6] S. Choi, M. Do, Development of the road pavement deterioration model based on the deep learning method, Electronics vol. 9 (1) (Jan. 2020) 3.

[7] S.Amirhossein Hosseini, Data-driven framework for modeling deterioration of pavements in the state of Iowa the state of Iowa. Doctoral dissertation, Iowa State University, 2020.

[8] E. Negri, L. Fumagalli, M. Macchi, A review of the roles of digital twin in CPS-based production systems, Procedia Manuf. vol. 11 (Jan. 2017) 939–948.

[9] M. Macchi, I. Roda, E. Negri, L. Fumagalli, Exploring the role of digital twin for asset lifecycle management, IFAC-Pap. vol. 51 (11) (2018) 790–795.

[10] X. Yang, et al., Research and applications of artificial neural network in pavement engineering: a state-of-the-art review, J. Traffic Transp. Eng. (Engl. Ed.) vol. 8 (6) (Dec.2021) 1000–1021.

[11] J. Rojo, E. Moguel, C. Fonseca, M. Lopes, J. Garcia-Alonso, J. Hernandez, Time Series Forecasting to Predict the Evolution of the Functional Profile of the Elderly Persons, Gerontechnology III: Contrib. Third Int. Workshop Gerontechnology, IWoG 2020, Oct. 5-6, 2020, Évora, Port. (2021) 11–22.

[12] M. Rangelov, H. Dylla, J. Davies, N. Sivaneswaran, O. Tatari, Integration of life cycle assessment into planning and project delivery for pavements in the USA, Int J. Life Cycle Assess. vol. 25 (8) (Jun.2020) 1605–1619.

[13] Y. Du, C. McNestry, L. Wei, A.M. Antoniadi, F.M. McAuliffe, C. Mooney, Machine learning-based clinical decision support systems for pregnancy care: a systematic review, Int J. Med Inf. vol. 173 (May 2023) 105040.
[14] R. Hu *et al.*, "Imbalance multiclass Problem: A robust Feature Enhancement-based Framework for Liver Lesion Classification," 2023.
[15] M. Cabrera, J. Ninic, W. Tizani, Fusion of experimental and synthetic data for reliable prediction of steel connection behaviour using machine learning, Eng. Comput. vol. 1 (Jun. 2023) 1–19.
[16] Z.S. Liu, et al., Digital twin-based intelligent safety risks prediction of prefabricated construction hoisting, Sustainability 2022 vol. 14 (9) (Apr. 2022) 5179.
[17] K. Kashinath, et al., Physics-informed machine learning: case studies for weather and climate modelling, Philos. Trans. R. Soc. A vol. 379 (2194) (Apr. 2021).
[18] A. Daw, A. Karpatne, W.D. Watkins, J.S. Read, V. Kumar, Physics-guided neural networks (PGNN): an application in lake temperature modeling, Knowl. -Guide Mach. Learn. (Jun. 2022) 353–372.
[19] W. E, J. Han, and L. Zhang, "Integrating Machine Learning with Physics-Based Modeling," arXiv preprint arXiv, Jun. 2020, doi: ⟨10.48550/arxiv.2006.02619⟩.
[20] Y. Deng, H. Wang, X. Shi, Physics-guided neural network for predicting asphalt mixture rutting with balanced accuracy, stability and rationality, Neural Netw. vol. 172 (Apr. 2024) 106085.
[21] L. Yao, Z. Leng, J. Jiang, F. Ni, Modelling of pavement performance evolution considering uncertainty and interpretability: a machine learning based framework, Int. J. Pavement Eng. vol. 23 (14) (2021) 5211–5226.
[22] A.K. Singh, J.P. Sahoo, Rutting prediction models for flexible pavement structures: a review of historical and recent developments, J. Traffic Transp. Eng. vol. 8 (3) (Jun. 2021) 315–338.
[23] M. Reichstein, et al., Deep learning and process understanding for data-driven Earth system science, Nature vol. 566 (7743) (Feb. 2019) 195–204.
[24] J. Willard, X. Jia, S. Xu, M. Steinbach, V. Kumar, Integrating scientific knowledge with machine learning for engineering and environmental systems, ACM Comput. Surv. vol. 55 (4) (Nov. 2022), https://doi.org/10.1145/3514228.
[25] J. Willard, X. Jia, M. Steinbach, V. Kumar, and S. Xu, "Integrating Physics-Based Modeling With Machine Learning: A Survey," arXiv preprint arXiv:2003.04919, vol. 1, p. 34, 2020, doi: ⟨10.1145/1122445.1122456⟩.
[26] L. Qiao, J. Zhu, Y. Wang, Coupling physics in machine learning to predict interlamellar spacing and mechanical properties of high carbon pearlitic steel, Mater. Lett. vol. 293 (Jun. 2021) 129645.
[27] M.C. Duran, J. Ninic, W. Tizani, F. Wang, Machine learning-based fusion of experimental and synthetic data for reliable prediction of steel connection stiffness, UKACM 2022 Conf. (2022).
[28] S.J. Raymond, D.B. Camarillo, Applying physics-based loss functions to neural networks for improved generalizability in mechanics problems, arXiv Prepr. arXiv vol. 2105.00075 (2021).
[29] N. Muralidhar, et al., Phynet: Physics guided neural networks for particle drag force prediction in assembly, Proc. West Mark. Ed. Assoc. Conf. (2020) 559–567, https://doi.org/10.1137/1.9781611976236.63.
[30] J. Sun, Z. Niu, K.A. Innanen, J. Li, D.O. Trad, A theory-guided deep-learning formulation and optimization of seismic waveform inversion, Geophysics vol. 85 (2) (Mar. 2020) R87–R99.
[31] A.J. Alnaqbi, W. Zeiada, G.G. Al-Khateeb, K. Hamad, S. Barakat, Creating rutting prediction models through machine learning techniques utilizing the long-term pavement performance database, Sustainability 2023 vol. 15 (18) (Sep. 2023) 13653, https://doi.org/10.3390/SU151813653.
[32] A.J. Haddad, G.R. Chehab, G.A. Saad, The use of deep neural networks for developing generic pavement rutting predictive models, Int. J. Pavement Eng. vol. 23 (12) (2022) 4260–4276.
[33] R. Guo, D. Fu, G. Sollazzo, An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree, Int. J. Pavement Eng. vol. 23 (10) (Apr. 2022) 3633–3646.
[34] Z. Chen, F. Xiao, F. Guo, J. Yan, Interpretable machine learning for building energy management: a state-of-the-art review, Adv. Appl. Energy vol. 9 (Feb. 2023) 100123, https://doi.org/10.1016/J.ADAPEN.2023.100123.
[35] G. Van Rossum and F.L. Drake, Python 3 Reference Manual. Scotts Valley, CA: CreateSpace, 2009.
[36] T. Kluyver, et al., Jupyter Notebooks – a publishing format for reproducible computational workflows, Position. Power Acad. Publ.: Play., *Agents Agendas* (2016) 87–90.
[37] W. McKinney, others, Data structures for statistical computing in python, Proc. 9th Python Sci. Conf. vol. 445 (2010) 51–56.
[38] C.R. Harris, et al., Array programming with NumPy, Nature vol. 585 (2020) 357–362, https://doi.org/10.1038/s41586-020-2649-2.
[39] P. Virtanen, et al., SciPy 1.0: fundamental algorithms for scientific computing in python, Nat. Methods vol. 17 (2020) 261–272.
[40] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. vol. 9 (3) (2007) 90–95, https://doi.org/10.1109/MCSE.2007.55.
[41] M. Waskom *et al.*, "mwaskom/seaborn: v0.8.1 (September 2017)." Zenodo, Sep. 2017. doi: ⟨10.5281/zenodo.883859⟩.
[42] F. Pedregosa, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. vol. 12 (2011) 2825–2830.
[43] S. Raschka, MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack, J. Open Source Softw. vol. 3 (Apr. 2018) 24, https://doi.org/10.21105/joss.00638.
[44] M. Smith, ABAQUS/Standard User's Manual, Version 6.9. United States: Dassault Systèmes Simulia Corp, 2009.
[45] C.J. Churilla, National Research Council, The long-term pavement performance program roadmap: a strategic plan, U. S. Fed. Highw. Adm. (Sep. 1995), https://doi.org/10.21949/1503647.
[46] H. Ziari, M. Maghrebi, J. Ayoubinejad, S.T. Waller, Prediction of pavement performance: application of support vector regression with different kernels, Transp. Res. Rec. vol. 2589 (Jan. 2016) 135–145.
[47] P. Marcelino, M. de Lurdes Antunes, E. Fortunato, M.C. Gomes, Transfer learning for pavement performance prediction, Int. J. Pavement Res. Technol. 2019 13: 2 vol. 13 (2) (Dec. 2019) 154–167, https://doi.org/10.1007/S42947-019-0096-Z.
[48] H. Gong, Y. Sun, Z. Mei, B. Huang, Improving accuracy of rutting prediction for mechanistic-empirical pavement design guide with deep neural networks, Constr. Build. Mater. vol. 190 (Nov. 2018) 710–718, https://doi.org/10.1016/j.conbuildmat.2018.09.087.
[49] A. Fathi, M. Mazari, M. Saghafi, A. Hosseini, S. Kumar, Parametric study of pavement deterioration using machine learning algorithms, Airfield Highw. Pavements (2019) 31–41, https://doi.org/10.1061/9780784482476.004.
[50] K. Chen, M.E. Torbaghan, M. Chu, L. Zhang, A. Garcia, Identifying the most suitable machine learning approach for a road digital twin, Proc. Inst. Civ. Eng. -Smart Infrastruct. Constr. vol. 174 (3) (2022) 88–101, https://doi.org/10.1680/jsmic.22.00003.
[51] N. Kargah-Ostadi, Y. (Mina) Zhou, T. Rahman, Developing performance prediction models for pavement management systems in local governments in absence of age data, Transp. Res. Rec. vol. 2673 (3) (Mar. 2019) 334–341, https://doi.org/10.1177/0361198119833680.
[52] B. Dayananda, S. Owen, A. Kolobaric, J. Chapman, D. Cozzolino, Pre-processing applied to instrumental data in analytical chemistry: a brief review of the methods and examples, Crit. Rev. Anal. Chem. (2023) 1–9, https://doi.org/10.1080/10408347.2023.2199864.
[53] V.M. Taavitsainen, "Denoising and Signal-to-Noise Ratio Enhancement: Splines," Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Second Edition: Four Volume Set, vol. 3, pp. 165–177, Jan. 2009.
[54] Y. Huang, Pavement analysis and design. 2004.
[55] T. Thenmozhi, R. Helen, Feature selection using extreme gradient boosting bayesian optimization to upgrade the classification performance of motor imagery signals for BCI, J. Neurosci. Methods vol. 366 (Jan. 2022) 109425, https://doi.org/10.1016/j.jneumeth.2021.109425.
[56] F. Deeba, M. Islam, F.M. Bui, K.A. Wahid, Performance assessment of a bleeding detection algorithm for endoscopic video based on classifier fusion method and exhaustive feature selection, Biomed. Signal Process Control vol. 40 (Feb. 2018) 415–424, https://doi.org/10.1016/j.bspc.2017.10.011.
[57] H. Gong, Y. Sun, X. Shu, B. Huang, Use of random forests regression for predicting IRI of asphalt pavements, Constr. Build. Mater. vol. 189 (Nov. 2018) 890–897, https://doi.org/10.1016/j.conbuildmat.2018.09.017.
[58] T. Yu, L.-I. Pei, W. Li, Z. Sun, J. Huyan, Pavement surface condition index prediction based on random forest algorithm, J. Highw. Transp. Res. Dev. (Engl. Ed.) vol. 15 (4) (Dec. 2021) 1–11, https://doi.org/10.1061/jhtrcq.0000794.

[59] C. Han, T. Ma, G. Xu, S. Chen, R. Huang, Intelligent decision model of road maintenance based on improved weight random forest algorithm, Int. J. Pavement Eng. (2020) 985–997.
[60] T.H. Lee, A. Ullah, R. Wang, Bootstrap aggregating and random forest, Adv. Stud. Theor. Appl. Econ. vol. 52 (2020) 389–429, https://doi.org/10.1007/978-3-030-31150-6_13/cover/.
[61] K. Saikiran, G. Lithesh, B. Srinivas, S. Ashok, Prediction of air quality index using supervised machine learning algorithms, 2021 2nd Int. Conf. Adv. Comput., Commun., Embed. Secur. Syst. (ACCESS) (Oct. 2021) 1–4.
[62] F. Pedregosa, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. vol. 12 (2011) 2825–2830. ⟨http://scikit-learn.sourceforge.net⟩.
[63] D. Liu, W. Jiang, L. Mu, S. Wang, Streamflow prediction using deep learning neural network: case study of Yangtze River, IEEE Access vol. 8 (2020) 90069–90086, https://doi.org/10.1109/ACCESS.2020.2993874.
[64] H. Sun, C. Qiu, L. Lu, X. Gao, J. Chen, H. Yang, Wind turbine power modelling and optimization using artificial neural network with wind field experimental data, Appl. Energy vol. 280 (Dec. 2020) 115880.
[65] D. Ardila, et al., End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nat. Med. vol. 25 (6) (May 2019) 954–961, https://doi.org/10.1038/s41591-019-0447-x.
[66] S. Badillo, et al., An introduction to machine learning, Clin. Pharmacol. Ther. vol. 107 (4) (2020) 871–885, https://doi.org/10.1002/cpt.1796.
[67] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Int. Jt. Conf. Artif. Intell. vol. 14 (2) (1995) 1137–1145.
[68] S. Alzabeebee, Y.M. Alshkane, A.J. Al-Taie, K.A. Rashed, Soft computing of the recompression index of fine-grained soils, Soft Comput. vol. 25 (24) (Dec. 2021) 15297–15312, https://doi.org/10.1007/S00500-021-06123-3/FIGURES/8.
[69] S. Alzabeebee, D.N. Chapman, A. Faramarzi, Development of a novel model to estimate bedding factors to ensure the economic and robust design of rigid pipes under soil loads, Tunn. Undergr. Space Technol. vol. 71 (Jan. 2018) 567–578, https://doi.org/10.1016/J.TUST.2017.11.009.
[70] T. Menzies, D. Port, Z. Chen, J. Hihn, S. Stukes, Validation methods for calibrating software effort models, Proc. 27th Int. Conf. Softw. Eng. (2005) 587–595.
[71] W. Su, J. Qian, and L. Liu, "Communication-Efficient False Discovery Rate Control via Knockoff Aggregation," arXiv preprint arXiv:1506.05446, 2015.
[72] A. El-Gawady, M.A. Makhlouf, B.S. Tawfik, H. Nassar, Machine learning framework for the prediction of alzheimer's disease using gene expression data based on efficient gene selection, Symmetry *Vol. 14* (2022) *491*, https://doi.org/10.3390/sym14030491.
[73] F. Tavazza, B. Decost, K. Choudhary, Uncertainty prediction for machine learning models of material properties, ACS Omega vol. 6 (48) (Dec. 2021) 32431–32440.
[74] B.R. Kirkwood, J.A. Sterne, Essential medical statistics, John Wiley & Sons. Blackwell Pub, 2010.
[75] R. Tibshirani, A comparison of some error estimates for neural network models, Neural Comput. vol. 8 (1) (1996) 152–163.
[76] A. Khosravi, S. Nahavandi, D. Creighton, A.F. Atiya, Comprehensive review of neural network-based prediction intervals and new advances, IEEE Trans. Neural Netw. vol. 22 (9) (Sep. 2011) 1341–1356.
[77] K.A. Zimmerman, D.G. Peshkin, Applying pavement preservation concepts to low-volume roads, Transp. Res. Rec. vol. I (1819) (Jan. 2003) 81–87.
[78] M. Rahman, Alternatives to PCI and MicroPAVER based maintenance solutions for airport pavements, University of New Mexico, NM, USA, Jan. 2013.
[79] E.A. Amarh, Evaluating the mechanical properties and long-term performance of stabilized full-depth reclamation base materials, Dr. Diss., Va. Tech. (May 2017).
[80] D. Sisodia, D.S. Sisodia, Quad division prototype selection-based k-nearest neighbor classifier for click fraud detection from highly skewed user click dataset, Eng. Sci. Technol., Int. J. vol. 28 (101011) (Apr. 2022).
[81] J. Kong and J.Y. Yuan, "Application of linear viscoelastic differential constitutive equation in ABAQUS," 2010 International Conference on Computer Design and Applications, ICCDA 2010, vol. 5, 2010.