Construction and Building Materials 344 (2022) 128166

Contents lists available at ScienceDirect

Construction and Building Materials

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

Data-driven methods to predict track degradation: A case study

Saeed Goodarzi [a,*], Hamed F. Kashani [b], Jimi Oke [a], Carlton L. Ho [a]

[a] Department of Civil and Environmental Engineering, University of Massachusetts Amherst, 28 Marston Hall, 130 Natural Resources Way, Amherst, MA 01003, United States

[b] Loram Technologies, Inc., P.O. Box 324, Williamsburg, MA 01096, United States

ARTICLE INFO

Keywords:
Large data
Preventive maintenance
Gradient Boosting Machine (GBM)
Logistic regression
SMOTE
Geometry data
Ground Penetrating Radar (GPR)

ABSTRACT

Predicting track degradation is extremely important due to its beneficial effects on increasing track safety and scheduling preventive maintenance. Logistic regression and Gradient Boosting Machine (GBM) were employed in this paper for predicting Yellow-tag Defect in the Next Inspection (YDNI), which is an indicator of maintenance requirement in the track. The geometry data of this paper were collected from passenger tracks with a length of 155 miles from 2011 to 2020. In addition to geometry data, Ground Penetrating Radar (GPR) data were used as additional explanatory variables to improve the performance of models. As there are a limited number of YDNI observations, Synthetic Minority Oversampling Technique (SMOTE) is employed to resolve the issue of imbalance dataset. Results indicate that the GBM outperforms logistic regression and can successfully predict YDNI with a very high sensitivity of 0.94. Parametric analysis indicated the direct correlation between the Ballast Fouling Index (BFI) and the probability of defect occurrence. Finally, the model can be integrated into a decision-making framework which generates predictions based on a given probability threshold for classification. Thus, the railway companies can select a specific threshold according to their maintenance budget and desired level of track safety for scheduling preventive maintenance interventions.

1. Introduction

Railroads are the largest means of intercity freight transportation in the United States, comprising approximately 40% of total freight volume. In addition to freight transportation, passenger rail is a significant mode of intercity transportation with an average of 86,000 passengers per day [1]. The US Federal Railroad Administration Office of Safety Analysis reported that track defects are one of the major reasons for train accidents [2]. For instance, 33% of 1747 train accidents that occurred in 2012 were due to track defects [3].

Railway track maintenance can be categorized as follows: corrective maintenance and preventive maintenance [4–6]. Corrective maintenance rectifies existing defects that result in track closure or speed restriction. However, this type of maintenance has high cost due to sudden failure and system recovery. On the other hand, preventive maintenance predicts defects and maintains them before they surpass safety thresholds. Preventive maintenance has the following advantages over corrective maintenance: (i) the safety of track is increased due to less occurrence of defects; (ii) the potential of track closure or speed limit is reduced significantly [4].

Many predictive models were developed for preventive track maintenance [1,7–23]. Support Vector Machine (SVM) was applied on US Class I track data to predict deviations from Track Quality Index (TQI) threshold [13]. Track degradation rate between two maintenance interventions was modeled by linear regression algorithm for tamping prediction [9]. Data from the Swedish railway network was utilized for predicting track degradation rate using artificial neural network [12]. Mehrali et al. [24] investigated the correlation between track stiffness and geometry data using data mining techniques. They found that track alignment has the most relation with support stiffness. Kasraei et al. [25] optimized geometry maintenance limits by minimizing the total cost of maintenance. In another study, Bakhtiary et al. [26] employed genetic algorithm to find opportunistic maintenance threshold. It was shown that considering the new threshold decreases maintenance cost and increases track quality.

To develop a preventive track maintenance strategy, Sharma et al. [18] employed random forests, SVM, and logistic regression for predicting geo-defects. They considered TQI, Million Gross Tonnage (MGT), train speed limit, and data of the previous inspection as explanatory variables. Cardenas-Gallo et al. [1] predicted the evolution of previous

* Corresponding author.
E-mail addresses: sgoodarzitae@umass.edu (S. Goodarzi), hamed.f.kashani@loram.com (H.F. Kashani), jboke@umass.edu (J. Oke), ho@umass.edu (C.L. Ho).

https://doi.org/10.1016/j.conbuildmat.2022.128166
Received 30 January 2022; Received in revised form 13 June 2022; Accepted 15 June 2022
Available online 23 June 2022

defects to red tag defects using ensemble classifiers. Their independent variables were MGT, class of track, the time interval between inspections, and the missing value between previous defects and red tag threshold. In another study Soleimanmeigouni et al. [7] forecast isolated defects using logistic regression. They found that standard deviation and kurtosis of longitudinal level have correlation with the occurrence of isolated defects.

In addition to geometry independent variables (e.g., TQI and standard deviation of profile) that were investigated in the previous studies, Ballast Fouling Index (BFI) [27] has a significant impact on behavior and degradation of track [28–35]. Kashani et al. [36] investigated the effect of BFI on track behavior under cyclic loads of trains. Results indicated that highly fouled ballast settled considerably more than clean ballast at saturated conditions. Indeed, insufficient hydraulic conductivity of fouled ballast hinders the dissipation of excess pore water pressure under train loads. Kashani et al. [37] also evaluated the effect of fouling on the ballast behavior using large-scale triaxial tests. They found that fouled ballast degrades faster due to the inability for water to drain. Toloukian et al. [38] conducted large-scale direct shear test on sand contaminated ballast. It was shown that shear strength decreases by increasing contamination, especially at fouling percentages higher than 24%. Box tests with different load cycles from 100,000 to 500,000 were conducted to investigate behavior of ballast in terms of fouling condition, particle breakage, and settlement [39].

Ground Penetrating Radar (GPR) is a non-invasive method for monitoring ballast condition [40]. The GPR method determines the subsurface condition by transmitting short electromagnetic waves and recording the reflected waves. More precisely, the waves will be reflected at the interface of materials with different dielectric permittivity. There is a significant difference between dielectric permittivity of soil and water, and to a lesser extent between different soil types and densities. Therefore, the subsurface condition can be investigated by interpolating the reflected waves [41]. Important features that relate to railway problems such as water content, fouling index, and ballast thickness can be determined by GPR method [42,43]. The correlation between GPR data and ballast fouling index was validated using laboratory and field tests in the previous studies [44–46]. The collection of GPR data can be conducted by installing antennas on train that transmit and receive waves. By installing three antennas, the subsurface condition of left, center, and right sides of the track can be determined. The quality of antennas, directly correlate with the quality GPR data [47].

The ability to predict Yellow-tag Defect in the Next Inspection (YDNI) is of vital importance as it allows railway administrators to maintain the track more efficiently. It should be noted that YDNI indicates that track requires maintenance operation; the definition of YDNI and the thresholds for different classes are elucidated in Section 3.2. In this study, the combination of geometry and GPR data are utilized for training prediction models to predict the probability of YDNI. In addition to proposing models, exploratory data analysis and parametric analysis are conducted to shed light on the hidden correlation between explanatory variables and the track maintenance requirement. The importance of correctly detecting previous maintenance interventions in training an accurate model is also investigated in this paper. The remainder of this study is organized as follows. Section 2 gives brief descriptions of the methodologies that were used in the paper. Section 3 describes the case study, which includes data summary, data preprocessing, and implementation of models. Section 4 presents the results of a case study and parametric analysis. Finally, Section 5 provides the conclusions of the paper.

2. Methodology

Two classification methods namely, logistic regression and Gradient Boosting Machine (GBM) were employed for predicting the YDNI. Generally, there is a trade-off between complexity and interpretability of prediction models. Logistic regression is a model with high interpretability that provides equations for predicting the output. On the other hand, GBM is a complex and non-parametric model with high accuracy [48]. Thus, using these two methods on the same problem provides both high accuracy and insight into the relationship between inputs and the output variable. The theories of these methods are summarized in the next two subsections.

2.1. Logistic regression

Logistic regression has been widely used for track prediction models [1,7,49–53]. In this research, logistic regression is employed for predicting the probability of YDNI. The response variable of Y is defined to be 1 when the yellow-tag defect occurs in the next inspection, and 0 otherwise.

$$P(Y=1)=\frac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n}} \tag{1}$$

where $x_1, x_2, \cdots, x_n$ are explanatory variables that are used for predicting YDNI. The details of explanatory variables and their distributions are provided in Section 3.1. $\beta_0, \beta_1, \cdots, \beta_n$ are unknown coefficients of the model. The values of coefficients must be estimated based on the training dataset. The maximum likelihood method is used to fit a logistic regression on the data and estimate the values of coefficients:

$$L(\beta_0,\beta_1,\cdots,\beta_n)=\prod_{i:y_i=1}p(x_i)\prod_{i':y_{i'}=0}(1-p(x_{i'})) \tag{2}$$

The likelihood (L) equation is the joint probability of each observation, where $p(x_i)$ is the probability of yellow-tag defect in ith observation when defect has truly occurred, and $p(x_{i'})$ is the probability of yellow-tag defect in i'th observation when defect has not occurred. The values of coefficients are chosen to maximize L, or equivalently, its logarithm. In other words, the coefficients are estimated such that the predicted probability of defect occurrence is as close as possible to the observed defect status.

2.2. Gradient Boosting Machine

Gradient Boosting Machine (GBM) has many advantages compared to regression models. GBM is less vulnerable to outliers in the dataset. It can fit complex nonlinear relationships of the output with independent variables. Also it can work with missing values in the independent variables as compared with deletion, which is usually employed in regression models [54]. GBM is an ensemble model that combines many simple decision trees, called "weak learners," to achieve a "strong learner" with a high prediction accuracy [55]. The GBM algorithm starts by creating the first decision tree, which maximally decreases the loss function. Then at each step, the residuals of previous models are used for fitting a new decision tree (i.e., "weak learner"). This process continues until the maximum number of iterations is reached [56]. The employed loss function to train the models was the deviance loss:

$$L=\sum_{i=1}^{n}[-y_i\times log(p_i)\ -(1-y_i)\times log(1-p_i)] \tag{3}$$

where L is the loss function that should be minimized. y_i and p_i are the actual label and predicted probability of ith observation, respectively; n is the total number of observations. The value of y_i is {0, 1} with respect to the occurrence of yellow-tag defect.

In GBM, decision trees are grown sequentially and use the information of previous models. On the contrary to a large decision tree that is usually prone to overfitting, the GBM approach learns slowly [48]. The GBM method can be seen as an algorithm that tries to find an additive model which minimizes the loss function [56].

There are three hyperparameters in the GBM algorithm that require tuning. 1. The number of iterations (B), which indicates the number of sequential trees that were created in the model. 2. The depth of tree (d)

that represents the number of splits in each tree and controls the complexity of model. A high value of d might result in overfitting as it allows the model to learn relations that are very specific to a particular record. 3. The shrinkage parameter (λ) controls the rate of learning. A small value of λ requires a high number of iterations to achieve an accurate model as the learning rate is slow [48,57].

3. Case study

3.1. Data collection

Data for this study were collected from passenger revenue tracks, covering 155 miles of track from the year 2011 through 2020. During this ten-year period, 110 geometry inspections were conducted. The entire dataset consists of 816,836 rows and 117 columns that provide outcomes of various inspections and geo-location details including milepost, track number, and class. The entire track was divided into blocks with a length of 200 ft. Yellow-tag defects were identified using the mid-chord offset to be consistent with the Federal Railroad Administration (FRA) threshold basis. Fig. 1 shows an example of dividing the track into 200-ft long blocks.

In addition to geometry data, GPR data were collected for this study. Subsurface properties of track such as Ballast Fouling Index (BFI) can be calculated from frequency content of raw GPR data, which was discussed in previous studies [44,45,58,59]. Two GPR inspections were conducted during the study period at the dates of June 2011 and Aug 2019. As the subsurface condition of track changes by cumulative loads of trains, the two GPR inspections cannot be used for the entire 10 years period from 2011 to 2020. To solve this issue, two datasets are created with respect to the availability of GPR data. The details of these datasets and their periods are discussed in Section 3.5.

GPR provides valuable information about the BFI on three channels that represent left, center, and right sides of the track. The GPR datapoints are collected at 5 m intervals. Thus, there are 12 datapoints on each channel for a 200-ft long block. The average value of 12 datapoints is considered for each channel and then, the maximum of average values is chosen as the BFI of a block. The reason for choosing the maximum value is the fact that having high BFI on either channel results in high degradation of track. Fig. 2 shows the heatmap of BFI for a block and the average of each channel. The considered BFI for this block is 20.5 as it is the maximum of average values. Table 1 presents the list of variables that were considered in prediction models.

3.2. FRA thresholds

Geometry defects are required to be rectified by maintenance interventions such as tamping or undercutting. An increase in the value of geometry defects, such as profile, decreases the safety of track and in some extreme cases might result in derailments. Thus, FRA has set thresholds for each class to ensure track safety (Table 2). Surpassing these thresholds results in track closure or speed restrictions. To avoid these consequences, railway companies conduct maintenance operations on much lower thresholds. Taking the case study of this paper as an example, only 137 records out of 262,211 records (0.052% of total datapoints) surpassed the FRA thresholds. Thus, the term of yellow-tag defect is defined in this paper, which indicates that the block requires maintenance intervention. The ability to predict Yellow-tag Defect in the Next Inspection (YDNI) is a powerful tool that significantly enhances maintenance scheduling. The thresholds of yellow-tag defects are half of the FRA thresholds for each class. It is notable that among 262,211 records, 10,786 records were labeled as yellow-tag defects, which is roughly equal to 4% of total datapoints. Fig. 3 shows an example of applying the yellow-tag defect threshold on six inspections in 2016. The track in this figure is Class 4; thus, the yellow-tag defect threshold is equal to one inch. The track deteriorated over time and two inspections in Aug and Oct 2016 surpassed the threshold. A maintenance intervention was conducted in Nov 2016 that reduced the profile to lower values.

Fig. 4 shows the number of YDNI in each track classes. There is a lower number of defects in lower track classes. This is attributed to the fact that the thresholds for lower classes are high, and maintenance interventions are usually conducted on the track before getting to those high values. For example, in Class 4, 43,119 records are without defects and only 289 records have yellow-tag defects.

3.3. Exploratory data analysis

The data are analyzed in this section to explore patterns across each variable. The distributions of explanatory variables with respect the YDNI are plotted in Fig. 5. Statistical values such as minimum, maximum, median, and variance of each input variable is presented in Table 3. SD and defect-ratio are relevant explanatory variables for predicting YDNI as defects are densely distributed at high values of these variables. The ratio of blocks with defect to normal blocks (i.e., #YDNI = 1 to #YDNI = 0) increases significantly with increasing BFI, especially at BFI values greater than 25. This indicates the importance of ballast quality in the occurrence of YDNI. The ratio of #YDNI = 1/#YDNI = 0 increases significantly at MGT values greater than 1.5. However, given that the data are collected from passenger tracks, the values of MGT are not high enough to have a considerable effect on YDNI. Yet, MGT was retained as an explanatory variable since it increases model performance, making it more robust for future predictions. It is notable that

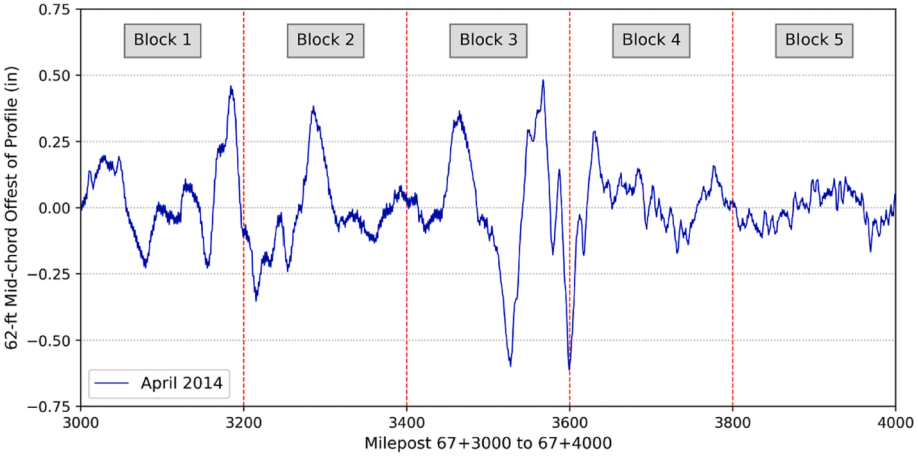

Fig. 1. Profile data of one inspection for 1000 ft that is divided into five blocks.

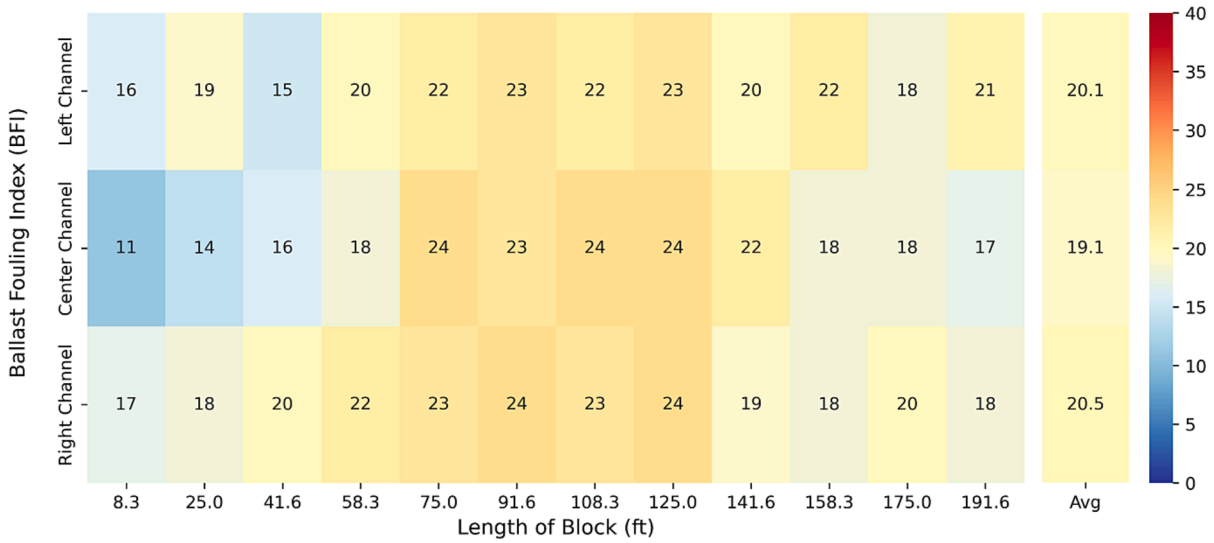

Fig. 2. Heatmap of ballast fouling index (BFI) in one block and the average value of each channel; the considered BFI in this block is 20.5.

Table 1
List of variables and their explanations.

Variable	Description
SD	The standard deviation of profile (62-ft mid-chord offset)
Defect-ratio	The ratio of maximum profile value to the yellow-tag defect threshold
SD-variation	The variations in the SD of profile from the last inspection
MGT	The loading of train between two inspections in million gross tons
Class	The class of track according to FRA standards
BFI	The ballast fouling index (BFI) of block collected from GPR data
YDNI*	The occurrence of Yellow-tag Defect in the Next Inspection. YDNI is 1 if there are defects and 0 otherwise.

* YDNI is the dependent variable that will be predicted by other variables.

when the defect-ratio is higher than 1, the profile has already passed the yellow-tag threshold. As illustrated in histogram of SD-variation, there are some negative values that erroneously represent improvement in the track. These negative values are not due to maintenance intervention as maintenance has much higher effect on the track. Although the track condition always deteriorates by train loading, observing some negative changes in SD of profile is not uncommon in the railway industry. These negative changes could be attributed to frost-heave cycles or sensor calibration of geometry cars [60].

The correlation analysis reveals how different variables positively or negatively affect each other. Fig. 6 shows the correlation plot for the entire dataset. It should be noted that Pearson equation was employed for continuous variables, while Spearman equation was used for ordinal (i.e., Class) and binary variables (i.e., YDNI). The three features that have the highest correlation with YDNI are defect-ratio, SD, and BFI, respectively. There is a high correlation between the pair of SD and defect ratio. However, neither of these two features are removed from the prediction models as they contain useful information for predicting YDNI. Removing either of these variables, significantly decreases the accuracy of models.

3.4. Data preprocessing

Data preprocessing is an essential task before training prediction models. When dealing with railway track data, appropriate preprocessing is challenging due to the following reasons:

- The lengths of track are usually high, which result in large datasets (e.g, the dataset of this study contains more than 80 million datapoints)
- The inspections from different dates are not usually aligned with each other. A considerable amount of work is required for perfectly aligning inspections
- The dates and locations of previous maintenance interventions should be determined for training accurate prediction models, which is a hard task due to complex behavior of railway tracks
- There are high amounts of noise in the data, which are attributed to different sources such as uncalibrated sensors, using different types of machines for collecting data, analyzing raw data by different operators, and highly variable weather conditions

3.4.1. Detecting previous maintenance interventions

It is important to detect the maintenance interventions prior to training the prediction models. This is because the inspections immediately before maintenance interventions have detrimental effects on training the model. This can be explained by Fig. 7 which shows the SD of profile vs time for two blocks. The inspections immediately before interventions have high SD and defect-ratio, but the next inspection will not have defect (i.e., YDNI = 0) due to maintenance intervention. If the inspections immediately before interventions are not filtered, the prediction model assumes that there are some significant improvements at random locations. Therefore, it tries to find a correlation that does not exist in reality. Consequently, the inspections immediately before interventions should be filtered to improve the performance of prediction

Table 2
FRA safety thresholds from [2].

Track profile (in)	Class of track								
	1	2	3	4	5	6	7	8	9
The maximum deviation from uniform profile on either rail at the mid-ordinate of 62-ft	3	2.75	2.25	2	1.25	1	1	1	0.75
Yellow-tag defect threshold*	1.5	1.375	1.125	1	0.625	0.5	0.5	0.5	0.375

* This row is defined in this paper and is not mentioned in FRA safety thresholds [2].

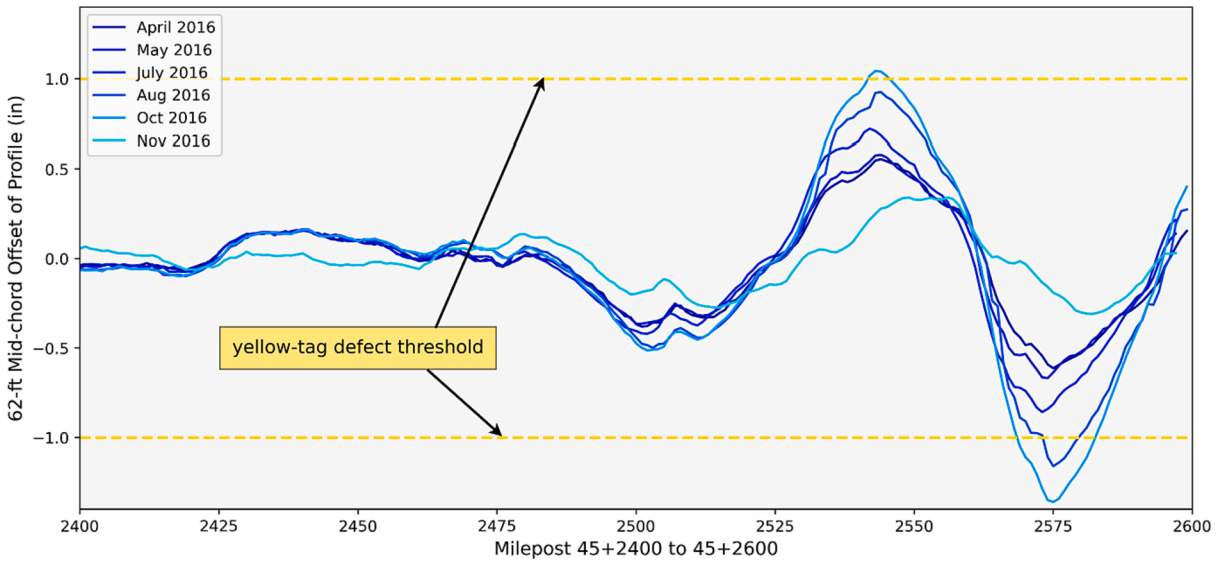

Fig. 3. An example of applying yellow-tag defect threshold on Class 4 track. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

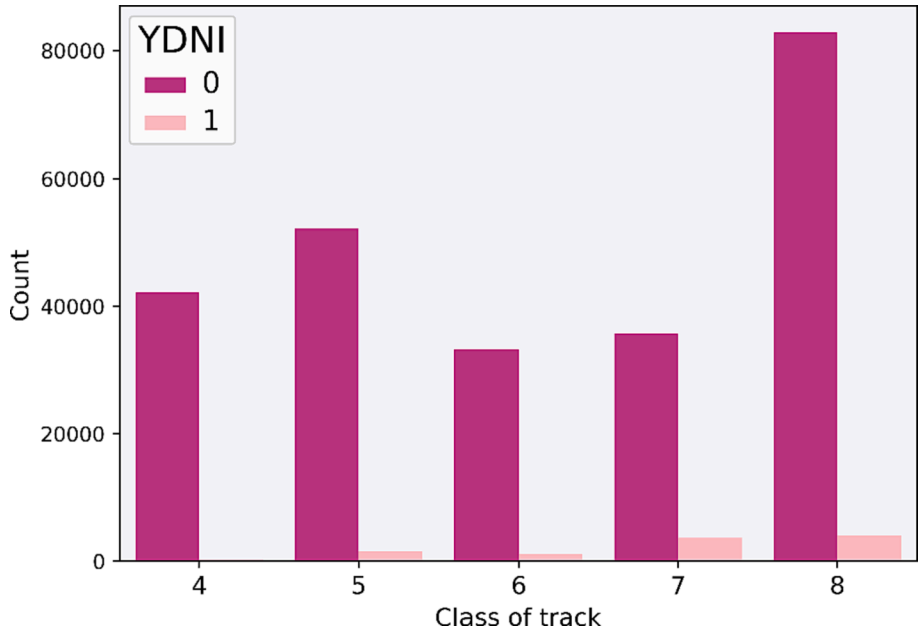

Fig. 4. The number of yellow-tag defects in different classes of track (1 represents blocks with defects). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

models. The effect of correctly determining previous maintenance interventions on the performance of predictions is discussed in detail in Section 4.2. Maintenance detection is not an easy task due to the complex behavior of track. A combination of four conditions is considered for detecting maintenance interventions. The pseudocode of the algorithm is shown in Fig. 8. SD_1 and SD_2 are the standard deviations of profile for two successive inspections.

As illustrated in block 1 of Fig. 7, Conditions 1 and 2 are satisfied and maintenance interventions are detected. However, the effects of interventions are not always significant to reduce the SD by more than 30% and satisfy condition 1. Thus, another condition should be considered to include interventions that have a lower effect on the track. Condition 3 was introduced for this purpose; it includes interventions that decrease the SD of track only by values higher than 10%. Picking the low threshold of 10% reduction is tricky since it might wrongly label noises as maintenance interventions. Taking block 2 of Fig. 7 as an example, there is a small SD reduction in Aug 2012 which is due to maintenance. However, there are multiple small SD reductions during 2018 and 2019 that are due to noises and should not be labeled as interventions. Condition 4 discriminates between these situations. In the former case, the variance of four previous inspections is low, which indicates that the data do not have noise; thus, any small reduction in SD is due to maintenance. On the other hand, the variance of four previous inspections in the latter case is high, which is the indicator of noisy data. These SD reductions during 2018 and 2019 are not labeled as intervention because condition 4 is not satisfied (see Fig. 7). Condition 2 is useful for the blocks that have SD lower than 0.1 in. In these blocks, 30% or 10% reduction is easily satisfied due to very low values of SD. Condition 2 ensures that the amount of reduction in SD is at least 0.03 in.

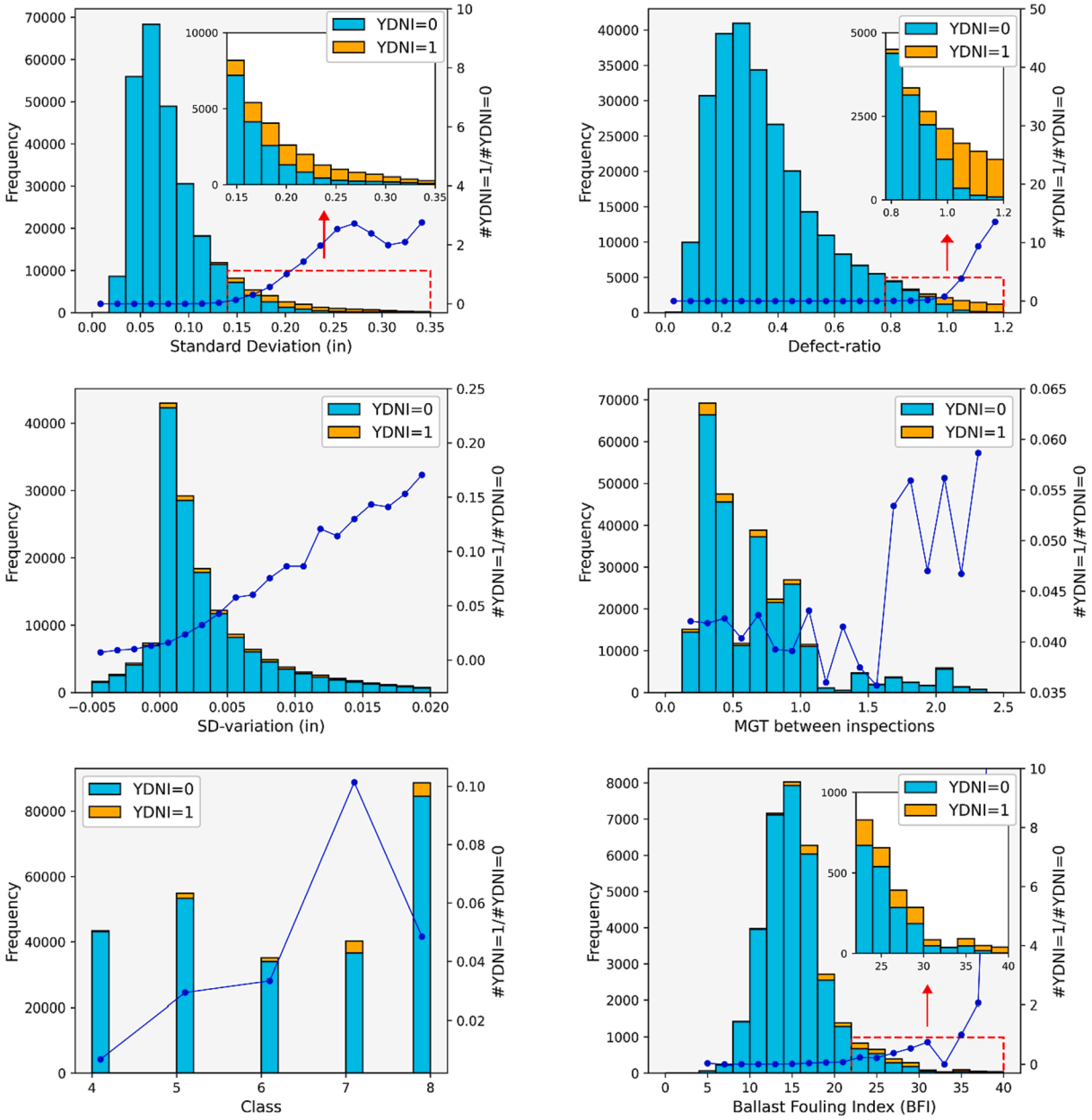

Fig. 5. The histograms of explanatory variables; the blue line shows the ratio of number of YDNI = 1 to YDNI = 0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Statistical values of explanatory variables.

	SD	Defect-ratio	SD-variation	MGT	BFI
Variance	2.421E−03	6.843E−02	1.254E−06	2.374E−01	1.562E+01
Minimum	0.0147	0.0180	−0.0048	0.2430	4.772
Median	0.0707	0.3220	0.0033	0.5670	14.90
Maximum	0.3428	1.196	0.0194	2.384	38.87

3.4.2. *Aligning inspections from different dates*

The 110 geometry inspections in 10 years of data are rarely aligned with each other. Thus, cross correlation analysis is conducted to find the amount of lag between unaligned inspections. In cross correlation, the second curve is moved along the first curve and the amount of correlation is calculated at each step. When two inspections align with each other, the value of correlation is maximized. After finding the amount of lag between curves, the un-aligned inspections are moved along other inspections. Further explanations about aligning inspections are provided in [61].

3.4.3. *Filtering assets*

The dynamic loads of train on assets, such as switches and undergrade bridges, are higher than those on normal track, which result in different behavior and faster deterioration in assets [62,63]. Thus, the blocks with assets must be filtered from the dataset. Otherwise, the

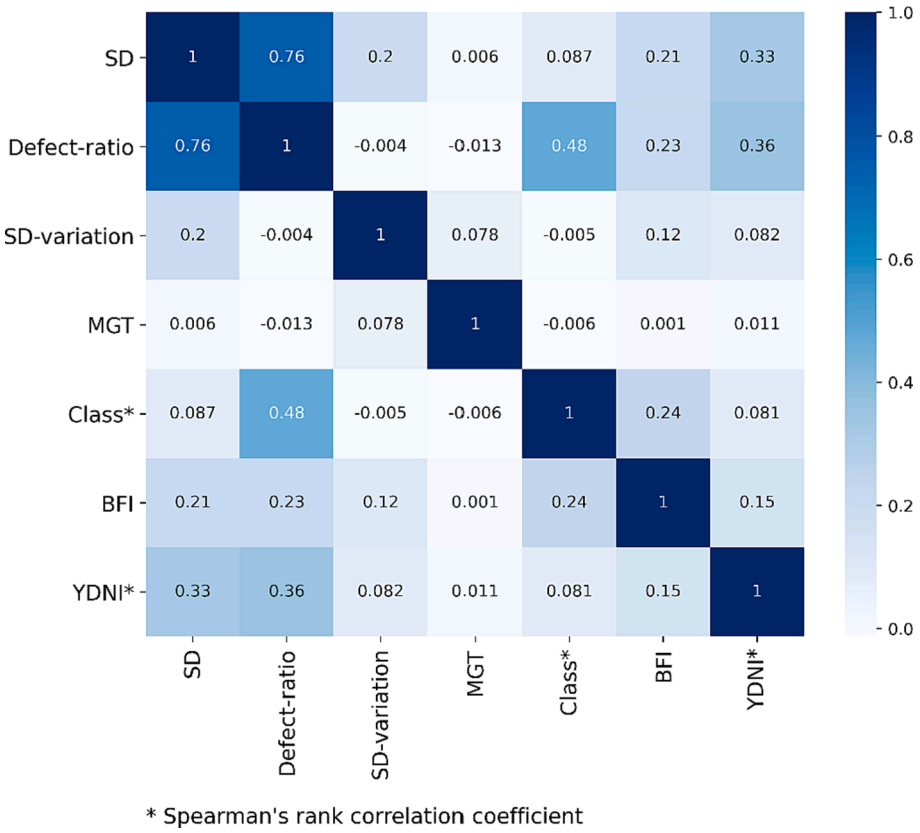

Fig. 6. The correlation plot using Pearson equation for continuous variables and Spearman equation for ordinal and binary variables.

models cannot accurately predict the behavior of normal track. It should be noted that the data regarding the start and end of assets in the tracks were provided with the geometry dataset.

3.4.4. Handling imbalanced dataset

Underrepresentation of a given class in training data can negatively impact the performance of a classification model. In these situations, prediction models tend to correctly classify only the majority class and treat the minority class as noise. To overcome this issue, two approaches can be taken: (i) cost sensitive learning that increases the weight and penalty for misclassifying the minority class; (ii) preprocessing techniques that focus on resampling the dataset. Resampling datasets that are designed to balance the imbalance dataset are divided into three categories [64]:

- Under-sampling methods: they randomly remove the samples from the majority class. The disadvantage of these methods is that the size of dataset shrinks. Thus, some valuable samples that were essential for determining the correct decision boundary might be removed.
- Over-sampling methods: they produce some synthetic samples from the minority class to overcome the issue of imbalance dataset
- Hybrid methods: these methods are a combination of under-sampling and over-sampling methods

The Synthetic Minority Oversampling Technique (SMOTE) algorithm [65] is employed in this case to resolve the issue of imbalance dataset. In this algorithm, synthetic samples are generated on euclidean distance between a minority sample and one of its nearest neighbors. More precisely, SMOTE algorithm takes three steps for generating synthetic samples. Firstly, a minority sample ($\vec{a}$) is randomly selected. Subsequently, its k-nearest neighbors form the same class are determined and one of them is chosen ($\vec{b}$). Eventually, a new sample is generated by randomly interpolating the two samples [66]:

$$\vec{x} = \vec{a} + w \times \left(\vec{b} - \vec{a}\right) \tag{4}$$

where $\vec{x}$ is the new synthetic sample; $\vec{a}$ is the randomly selected sample from the minority class; $\vec{b}$ is one of the k-nearest neighbors of $\vec{a}$; w is random weight between 0 and 1. The ratio of datapoints with YDNI = 1 to total datapoints was equal to 4.1% in the training dataset before applying the oversampling technique. To choose the best oversampling proportion, ratios from 10% to 50% of total datapoints were employed. A proportion of 20% resulted in the best model performance and was therefore chosen for further analysis.

3.4.5. k-fold cross-validation

The *k*-fold cross-validation method is employed to increase the robustness of prediction models and decrease the chance of overfitting [1,67–70]. In the *k*-fold cross-validation method, the entire dataset is shuffled randomly and then is divided into k folds with equal sizes (Fig. 9). At each iteration, a single fold is considered as a test dataset and the model is trained by the remaining folds. This process is repeated for k iterations and at each iteration, the performance of model is measured by test fold. The final performance of prediction model is equal to the average of all iterations. The value of k is chosen to be 5.

3.5. Model implementation

Two models are implemented for each method according to the availability of BFI data. As discussed in Section 3.1, two BFI datasets were collected in June 2011 and Aug 2019. The value of BFI increases over time and the amount of increment is not constant and depends on numerous factors. Thus, the collected BFI datasets cannot be utilized for the entire dataset from 2011 to 2020. Based on this limitation, two prediction models are proposed: (i) the first model considers the entire dataset from 2011 to 2020 and ignores BFI as a variable; (ii) the second

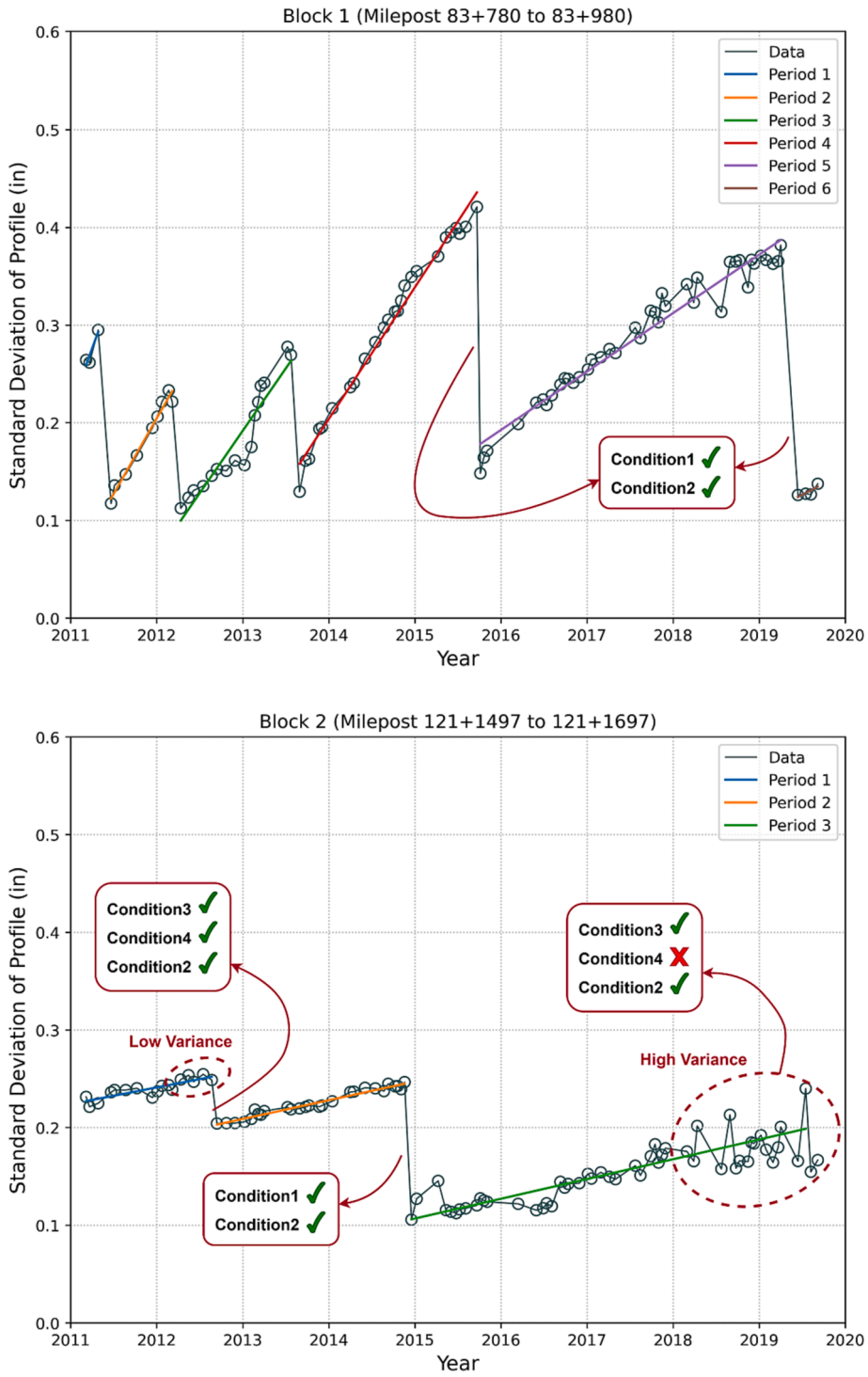

Fig. 7. Performance of automatic maintenance detection on two blocks.

model assumes a constant value of BFI for 3 months before and after dates of collecting BFI and only considers dataset in that period (i.e., March 2011 to Sep 2011 and May 2019 to Nov 2019). Fig. 10 shows the summary of the proposed approach using a flowchart.

4. Results

4.1. Performance metrics

To investigate the performance of the models, different metrics of test datasets in the *k*-fold cross-validation method are calculated. Then, the average value of k iterations is considered as the performance, which is defined by sensitivity, precision, F_1-score, and overall accuracy in classification problems. It should be noted that overall accuracy has valuable information compared to the other metrics when data is unevenly distributed. This is attributed to the fact that classifying all datapoints to the majority class leads to a high overall accuracy while the model performance is poor.

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

```
Start
    for i in total_number_of_blocks:
        for j in total_number_of_inspections_in_a_block:
            if [(SD2/SD1)<0.7] & [(SD1 – SD2)>0.03]:
                maintenance intervention is conducted
            elif [(SD2/SD1)<0.9] & [Variance(four_last_inspections)<4e-6] & [(SD1 – SD2)>0.03]:
                maintenance intervention is conducted
            else:
                maintenance intervention is not conducted
            endif
        endfor
    endfor
Finish
```

Condition 1 Condition 2 Condition 3 Condition 4 Condition 2

Fig. 8. Pseudocode of algorithm developed for detecting maintenance interventions;

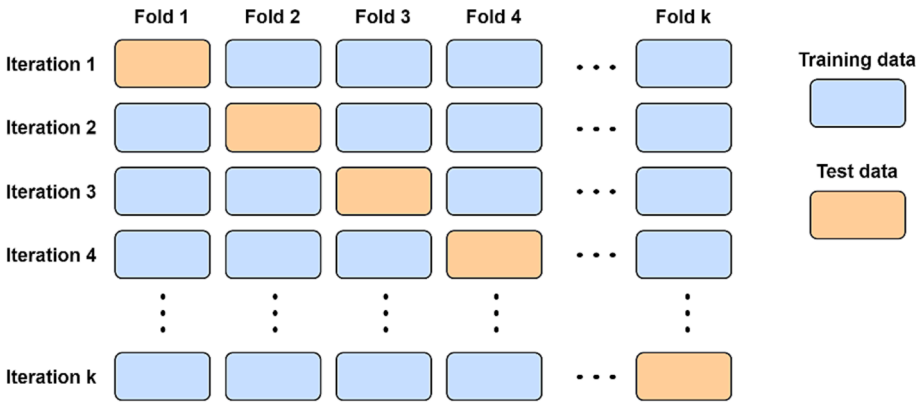

Fig. 9. Train-test split in *k*-fold cross-validation method.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$F_1\text{-score} = 2 \times \frac{Sensitivity \times Precision}{Sensitivity + Precision} \tag{7}$$

$$Overall\,Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{8}$$

where:

True Positive (TP): the block with YDNI that is correctly predicted.
False Positive (FP): the block without YDNI that is wrongly predicted.
True Negative (TN): the block without YDNI that is correctly predicted.
False Negative (FN): the block with YDNI that is wrongly predicted.

4.1.1. Performance of logistic regression

Table 4 presents the confusion matrix of test datasets using logistic regression models with default classification threshold of 50%. The FP and FN that represent the prediction error are relatively small for both models. Notably, the size of model without GPR is roughly 10 times bigger than the one with GPR. It is attributed to the fact that the model without GPR considers 10 years of data rather than 12 months in the model with GPR. Table 5 presents the performance of logistic regression models. The sensitivity of model with GPR is slightly higher than the model without GPR. Indeed, adding BFI to the model provides valuable information that helps in predicting YDNI. If the datasets of two models had the same size, the difference between sensitivities would be more than 0.02. Generally, with an increase in the size of dataset, the robustness of prediction model improves as more types of behaviors are seen in training the model. As stated in the last paragraph, the overall accuracy contains less valuable information compared to the other metrics when the output is imbalance. Overall accuracy is calculated in this study to present a complete list of classification metrics.

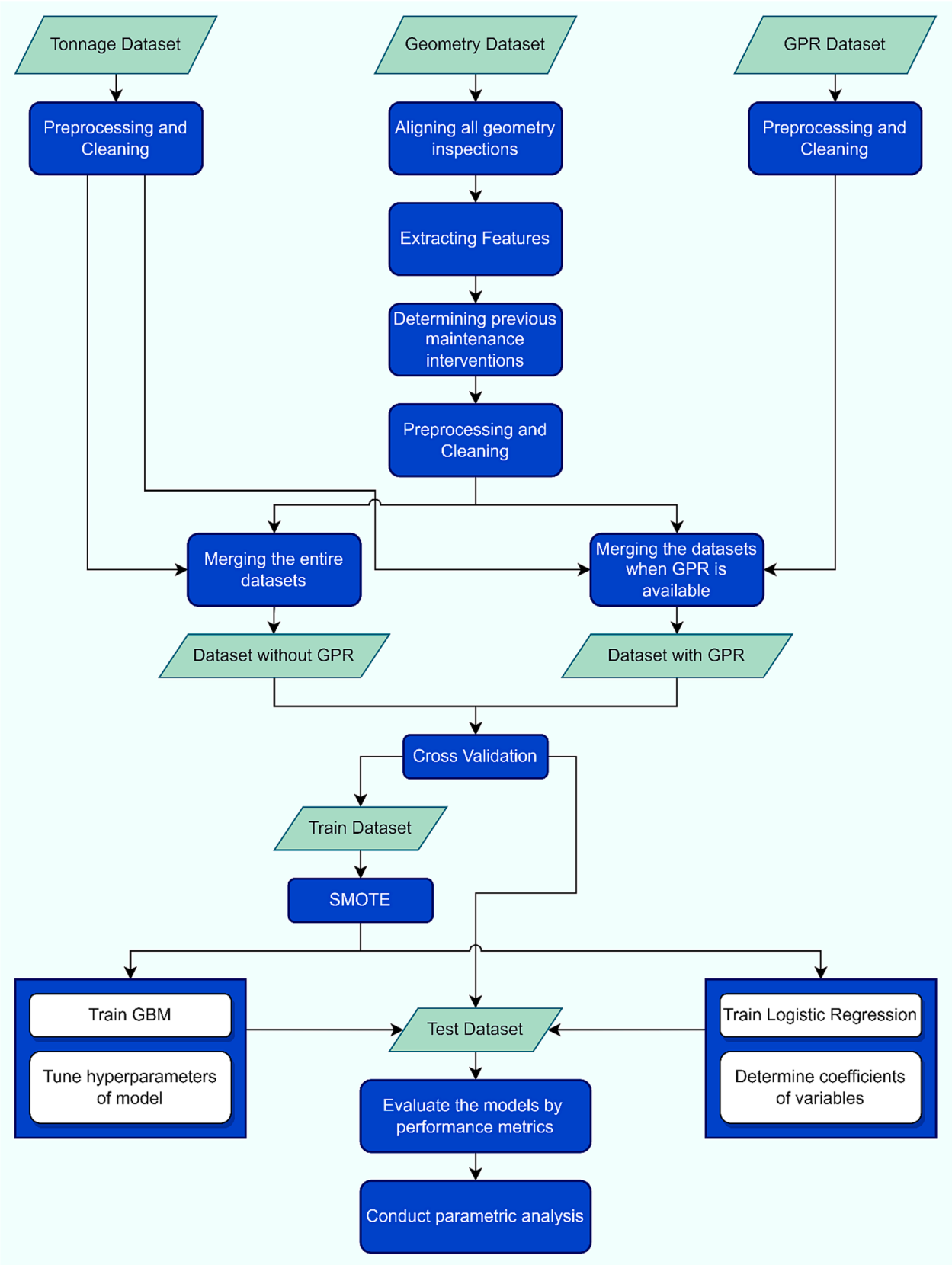

Fig. 10. Flowchart diagram of the proposed modeling approach.

Table 4
Confusion matrix of test datasets using logistic regression models.

Models	Model without GPR data		Model with GPR data	
	Predicted: No	Predicted: Yes	Predicted: No	Predicted: Yes
Actual: No	TN = 50,200	FP = 131	TN = 5,659	FP = 14
Actual: Yes	FN = 279	TP = 1833	FN = 24	TP = 199

Table 5
Performance of logistic regression models.

Models	Sensitivity	Precision	F_1-score	Overall Accuracy
Model without GPR data	0.87	0.93	0.90	0.99
Model with GPR data	0.89	0.93	0.91	0.99

Table 6 presents coefficients of explanatory variables for the estimated logistic regression models. These coefficients can be used in Eq. (9) for predicting the probability of YDNI occurrence. Each class of track is a dummy variable that can take values of 0 or 1 depending on the absence or presence of that class. The coefficient of Class 4 track is negative, which means that the relative probability of YDNI decreases if a block is Class 4. The coefficients of Classes 6, 7, and 8 are similar, which indicates the small difference between the probabilities of YDNI in these classes. The reason for this behavior is shown in Fig. 11. The median of YDNI in Class 4 is much higher than the other classes, which indicates that the SD of this class has to be very high until YDNI occurs. On the other hand, there is a small difference between the median of Classes 6, 7, and 8 showing the behaviors of these classes are similar to each other. The other reason for small differences between coefficients of these classes is presented in Table 2. The FRA safety thresholds for all of these classes are similar and are equal to 1 in.

$$P(YDNI=1)=\frac{e^{\beta_0+\beta_1 SD+\beta_2 DR+\beta_3 SDV+\beta_4 MGT+\beta_5 C4+\beta_6 C5+\beta_7 C6+\beta_8 C7+\beta_9 C8+\beta_{10} BFI}}{1+e^{\beta_0+\beta_1 SD+\beta_2 DR+\beta_3 SDV+\beta_4 MGT+\beta_5 C4+\beta_6 C5+\beta_7 C6+\beta_8 C7+\beta_9 C8+\beta_{10} BFI}} \quad (9)$$

4.1.2. Performance of GBM

The other prediction model that is proposed in this paper is Gradient Boosting Machine (GBM). To increase the performance metrics of GBM, hyperparameters of the model such as depth of tree, number of estimators, and learning rate should be tuned using *k*-fold cross-validation. The complexity of the model increases with increasing the depth of tree. Choosing a high value of tree depth results in over-fitting and low performance of the test dataset. The number of estimators represents the number of sequential trees created in GBM model. Although GBM is robust at a high number of estimators, it can still overfit at a point [48]. After tuning the hyperparameters for each model, the performance of model is investigated. Table 7 presents the confusion matrix of the test dataset using GBM. The number of FN in both models with and without GPR data is lower than logistic regression models, which indicates that GBM can predict YDNI more accurately. Table 8 presents the performance of GBM models. The sensitivity of the model with GPR data is higher than the model without GPR data, which is the similar behavior using the logistic regression model. It is notable that feeding locations that have blocks with high BFI (i.e., higher than 25) to the prediction models does not change their performance. However, the ratio of blocks with defects is higher in blocks with high BFI as shown in Fig. 5.

4.1.3. Comparison between two prediction models

Making a comparison between Table 5 and Table 8 indicates that GBM outperforms logistic regression in terms of sensitivity. The sensitivity of GBM is 5 percentage points greater than logistic regression both for models with and without GPR data. However, GBM and logistic regression perform similarly in terms of precision (0.93). The overall accuracy of models remains 0.99 regardless of employed methods or presence of GPR data. This is attributed to the inability of overall accuracy in measuring the true performance of prediction models when the output is imbalance. The other three metrics are not affected by unevenly distribution of data. Thus, they are better indicators for performance of prediction models in this paper.

There is a trade-off between the complexity and interpretability of prediction models. Logistic regression models are good in terms of interpretability and provide equations that can be utilized for predicting the probability of YDNI occurrence. On the other hand, GBM models that have higher complexity can predict YDNI more accurately than logistic regression models. Thus, each of the proposed models have their own advantages.

4.2. The importance of previous maintenance interventions on performance of predictions

In this section, the impact of correctly detecting previous maintenance interventions on the performance of prediction models will be discussed in detail. Maintenance interventions improve the track condition by significantly decreasing SD and Defect-ratio of track. These interventions should be detected and inspections immediately before them should be filtered to train accurate prediction models. To highlight the importance of correctly detecting interventions, prediction models are trained on four datasets, in which different scenarios are applied for filtering inspections before maintenance interventions. The values of sensitivity for test dataset of these models are presented in Table 9.

The first scenario indicates the results of prediction models while using the current paper's method for detecting maintenance interventions (Section 3.4.1). The data loss in this scenario is 3.43%, which indicates the ratio of filtered inspections immediately before interventions to the total inspections. In the second scenario, maintenance interventions were not detected. Thus, no inspection was filtered from the dataset (i.e., data loss = 0). The sensitivities of trained models on this dataset are much lower. These low values can be explained by Fig. 7. Consider an inspection in this figure that is immediately before intervention; the values of SD and defect-ratio are extremely high and models will predict that the next inspection will have defect. However, YDNI does not occur as the track condition is improved due to intervention. These datapoints are outliers that do not represent the actual behavior of track. Not filtering them from the dataset adversely affects the performance of prediction models. In the third scenario, the method of Cardenas-Gallo et al. [1] is applied. In this method, all inspections that had negative changes in SD were filtered from the dataset regardless of reduction size. The data loss in this method is high (19.67%). The last scenario employs the method that was used in the paper of Soleimanmeiguni et al. [7]. In this method, a constant threshold of 15% reduction in SD is considered for maintenance intervention if the track condition is good. In case of poor track condition, a linear threshold is considered [7]. The comparison between these scenarios indicates that correctly detecting previous maintenance interventions is a key factor in training a prediction model with a good performance.

Table 6
Coefficients of explanatory variables for logistic regression models.

	Intercept (β_0)	SD (β_1)	Defect_ratio (β_2)	SD_variation (β_3)	MGT (β_4)	Class4 (β_5)	Class5 (β_6)	Class6 (β_7)	Class7 (β_8)	Class8 (β_9)	BFI (β_{10})
Model without GPR data	−16.61	14.37	13.11	24.54	0.251	−2.34	0.243	0.661	0.654	0.781	–
Model with GPR data	−13.13	5.03	10.35	2.39	0.344	−0.97	0.088	0.208	0.215	0.324	0.068

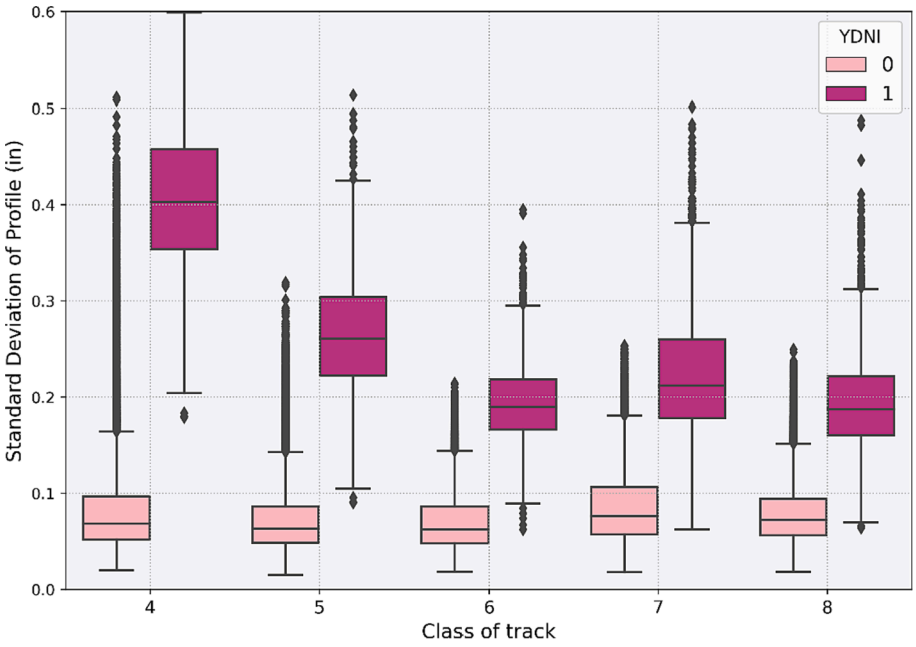

Fig. 11. Boxplot of SD for different classes of track.

Table 7
Confusion matrix of test datasets using GBM models.

Models	Model without GPR data		Model with GPR data	
	Predicted: No	Predicted: Yes	Predicted: No	Predicted: Yes
Actual: No	TN = 50,182	FP = 149	TN = 5,652	FP = 15
Actual: Yes	FN = 165	TP = 1947	FN = 13	TP = 210

Table 8
Performance of GBM models.

Models	Sensitivity	Precision	F_1-score	Overall Accuracy
Model without GPR data	0.92	0.93	0.93	0.99
Model with GPR data	0.94	0.93	0.94	0.99

Table 9
Performance of prediction models that are trained on datasets with different filtering scenarios.

# Scenario	Maintenance detection method	Data loss (%)	Sensitivity	
			Log Regression	GBM
Scenario 1	Current paper	3.43	0.87	0.92
Scenario 2	No detection	0.00	0.73	0.85
Scenario 3	Cardenas-Gallo et al. [1]	19.67	0.74	0.88
Scenario 4	Soleimanmeiguni et al. [7]	5.19	0.78	0.88

4.3. Selecting the optimal classification threshold

The models developed in this paper predict the probability of YDNI occurrence and subsequently classify blocks into two groups (i.e., blocks with and without yellow-tag defects in the next inspection). The default threshold of models for classification is 50%. Thus, the block will be classified as YDNI if the probability of YDNI occurrence is higher than 50%. The classification threshold can be modified to increase the sensitivity of the model. As the sensitivity directly correlates with the track safety, it is rational to increase it even at the cost of decreasing the precision and f1-score. Fig. 12 shows the effect of changing classification threshold on sensitivity, precision, and f1-score of the model. With a decrease in classification threshold, more records are predicted as blocks with YDNI. Thus, more maintenance interventions will be scheduled. Generally, threshold modification can be used for making a balance between maintenance expenses and track safety. Therefore, railway companies can pick a specific threshold according to their maintenance budget and desired level of track safety. As illustrated in Fig. 12, decreasing the classification threshold from 50% to 10% increases the sensitivity from 0.87 to 0.97 at the cost of reducing precision from 0.93 to 0.66. The precision of 0.66 means that 66% of scheduled maintenance are relevant and will be conducted on tracks that truly require maintenance interventions. Another ideal threshold is 41%, in which all three metrics are equal to 0.9.

4.4. Parametric analysis

A parametric analysis is conducted to investigate the effect of each explanatory variable on the probability of YDNI occurrence. In parametric analysis, synthetic datasets should be created in which only one explanatory variable changes while the other variables are kept constant. Then, these synthetic datasets are fed to the prediction models and the probability of defect occurrence is investigated. It should be noted that the amount of changes in explanatory variables must be limited to the range of that variable in the actual dataset. Fig. 13 shows the parametric analysis of the model without GPR data. The probability of occurring YDNI in Class 4 is much lower than in the other classes. It is attributed to the fact that the FRA and yellow-tag defect threshold for this class is much higher than the other classes. Thus, maintenance interventions are usually conducted in this class before the track gets to those thresholds. Among explanatory variables, SD and Defect-ratio have a strong positive correlation with the probability of defect occurrence. Although SD-variation has a direct correlation with the probability of YDNI, its effect is not as significant as the other two variables.

Fig. 14 shows the parametric analysis of the model with GPR data. With an increase in the BFI, the probability of occurring yellow-tag defects increases, which is consistent with the previous studies [8,36,37]. The slopes of the curves increase at BFI higher than 25. A high value of BFI decreases drainage quality of track and water from precipitation remains in the ballast. When dynamic loads of trains are applied to the track, the pore water pressure increases and effective

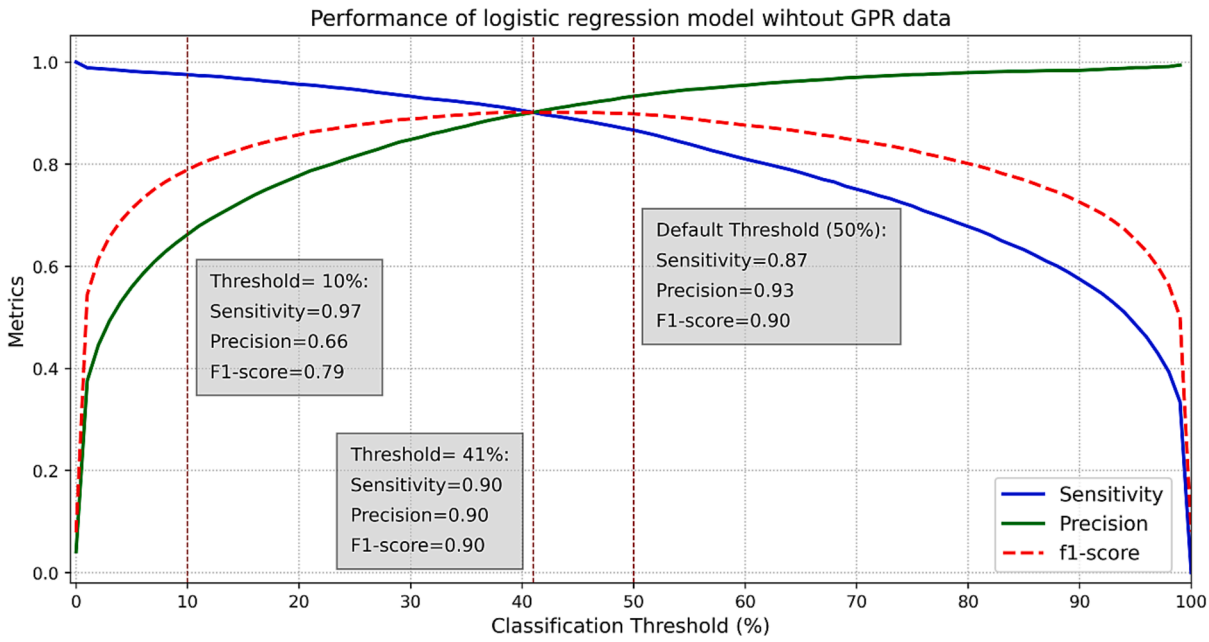

Fig. 12. Effect of changing classification threshold on sensitivity, precision, and F1-score of model.

stress between ballast particles decreases, which results in a high degradation rate [36]. The probability of YDNI occurrence for SD equal to 0.3 in. increases from 0.18 to 0.62 when the BFI increases from 10 to 40.

4.5. Application of model on track

Fig. 15 shows a sample application of the model on milepost 130 to 133. Actual and predicted YDNIs are shown by blue solid circles and red hollow circles, respectively. A combination of solid and hollow circles means that there was an actual YDNI that was correctly predicted. A single solid circle or hollow circle represent FN or FP, respectively. As presented in the confusion matrix, there are 5612 records that were not actual YDNI and were not predicted as defects; these records are not shown in the figure. Out of 85 total YDNIs, the model could correctly predict 79 of them, and could not predict the other 6 defects, which results in the sensitivity of 0.93. As illustrated in milepost 130.7, some YDNIs were detected around 2012. Then, maintenance intervention was conducted in March 2012 that rectified the defects. The same process of detecting defects and maintaining the track was repeated at different years. The ability to predict YDNI allows railway companies to schedule maintenance interventions before defects occur. Notably, the type of maintenance intervention that was conducted in June 2014 is different from the other interventions as it rectified defects for 3 years.

5. Concluding remarks

This research has developed a new data-driven approach for predicting Yellow-tag Defect in the Next Inspection (YDNI), which are indicators of maintenance requirement in the track. Implementing the proposed approach enables the railway companies to efficiently schedule their preventive maintenance interventions. The geometry data that were employed in this paper are from passenger tracks with a length of 155 miles from 2011 to 2020. In addition to geometry data, GPR data were utilized as explanatory variables to enhance the performance of defect prediction and investigate the effect of GPR on defect occurrence. As GPR data were not collected as frequently as geometry data, two prediction models are proposed with respect to the availability of GPR data. The performance metrics of models with GPR data were higher than models without GPR data.

Due to limited number of defects in the dataset, Synthetic Minority Oversampling Technique (SMOTE) is utilized to overcome the issue of imbalance dataset. The employed classification methods in this paper are logistic regression and Gradient Boosting Machine (GBM). Both models can successfully predict the YDNI with high sensitivities of 0.89 and 0.94, respectively for logistic regression and GBM. It was shown that employing the proposed methodology for correctly detecting previous maintenance interventions had a major impact on getting high sensitivities for predictions. The inspections immediately before interventions should be correctly identified and filtered from the dataset as YDNI is 0 (i.e., defects will not occur in the next inspection) in these cases due to maintenance operations. If these inspections were not filtered from the dataset, the prediction model would assume that track condition improves significantly at some random dates.

Enhancing the sensitivity of the model is very important due to the direct correlation between sensitivity and track safety. The sensitivity can be improved by modifying the classification threshold at the cost of reducing the precision and scheduling more maintenance. Railway companies can pick a specific classification threshold for efficiently scheduling their maintenance interventions according to their budget and desired level of safety.

The parametric analysis was conducted to investigate the effect of each explanatory variable on the probability of YDNI occurrence. Results indicate that an increase in Ballast Fouling Index (BFI) increases the probability of YDNI occurrence. The amount of probability increment becomes more significant at BFI values higher than 25, which shows the detrimental effect of high BFI on the degradation of the track.

This paper provides insights into the contributing factors and their impacts on YDNI occurrence. It proposes models that can predict YDNI with very high sensitivities. Railway companies can use these models to predict the probability of YDNI on their track and schedule preventive maintenance. By applying these models, defects will be rectified before they evolve to certain thresholds, which significantly increases track safety.

CRediT authorship contribution statement

Saeed Goodarzi: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization. **Hamed F. Kashani:** Conceptualization, Validation,

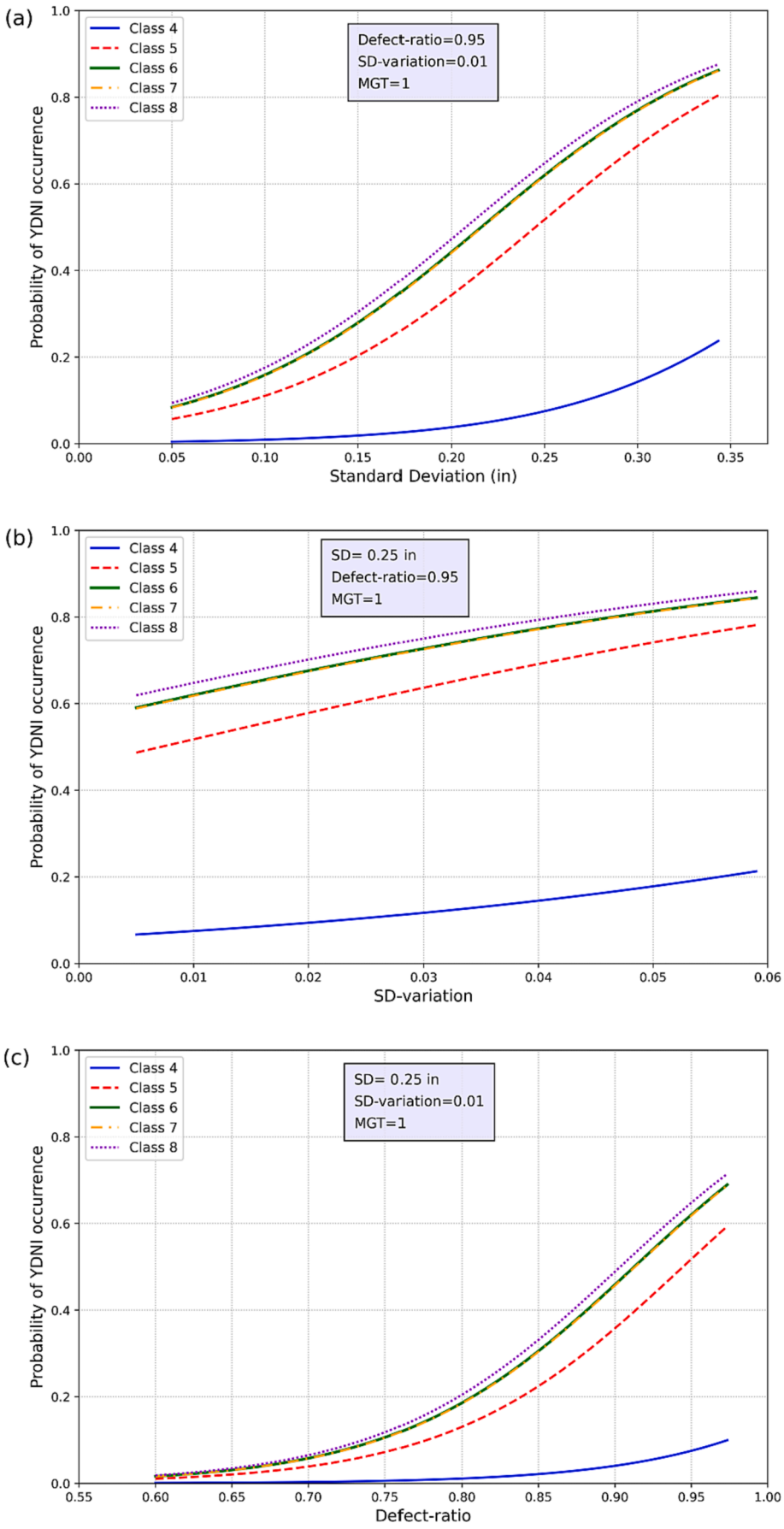

Fig. 13. Parametric analysis of model without GPR data; (a) effect of SD, (b) effect of SD-variation, (c) effect of Defect-ratio.

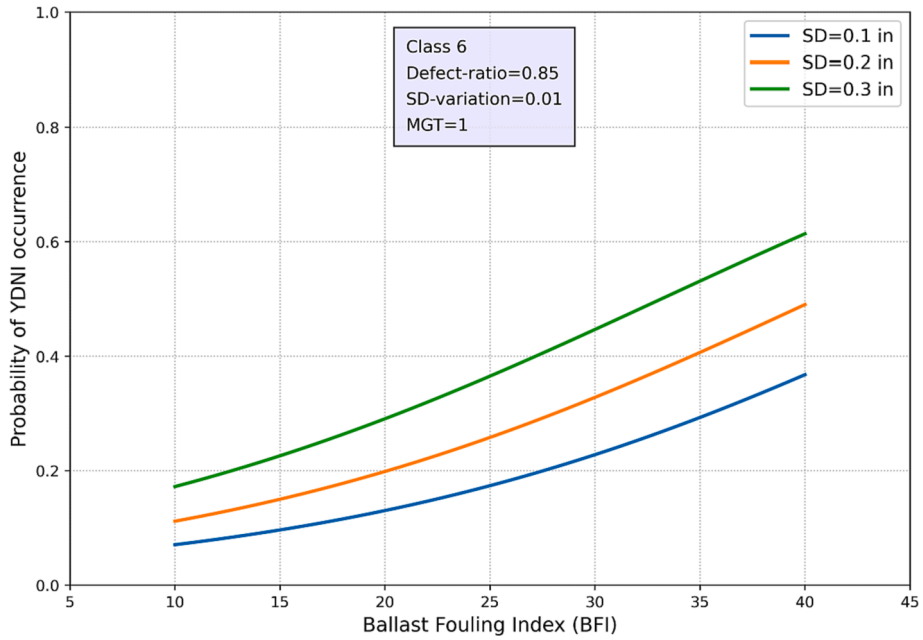

Fig. 14. Parametric analysis of the model with GPR data; effect of BFI on the probability of defect occurrence.

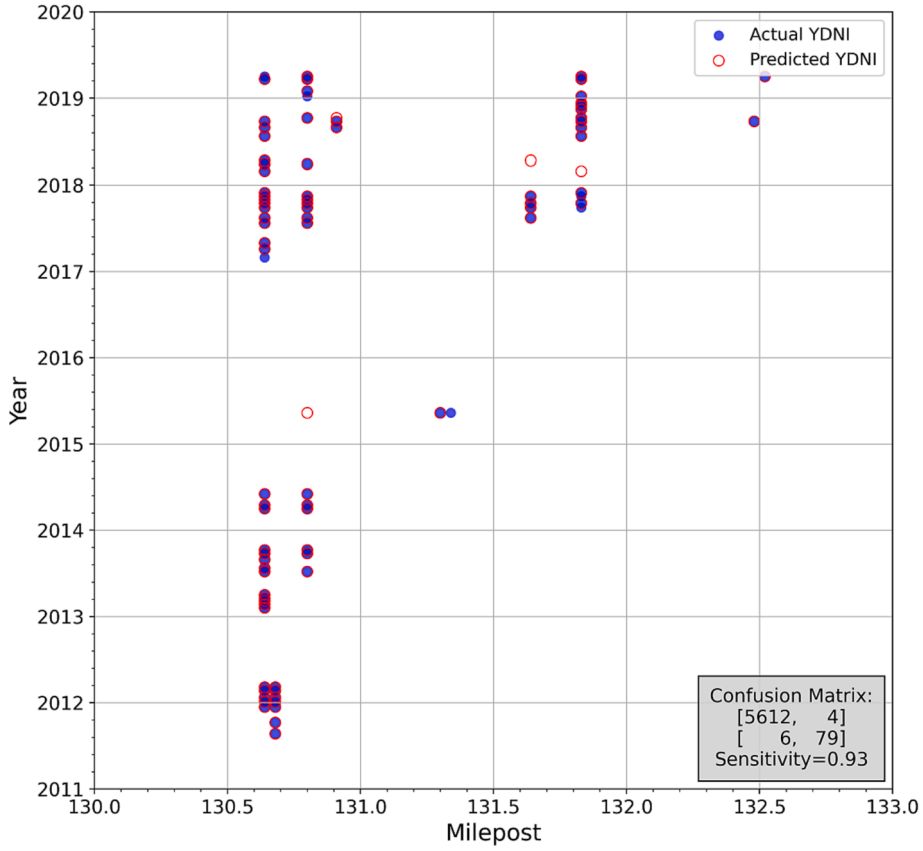

Fig. 15. A sample application of the model on random mileposts of track.

Data curation, Funding acquisition. **Jimi Oke:** Formal analysis, Writing – review & editing. **Carlton L. Ho:** Conceptualization, Formal analysis, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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

The data that has been used is confidential.

Acknowledgment

This work was supported by the Federal Railroad Administration (FRA) under order No. FR19RPD31000000046. The authors would like to thank Jim Hyslip and Aaron Judge of Loram Technologies, Inc for their support and guidance in helping produce this paper.

References

[1] I. Cárdenas-Gallo, C.A. Sarmiento, G.A. Morales, M.A. Bolivar, R. Akhavan-Tabatabaei, An ensemble classifier to predict track geometry degradation, Reliab. Eng. Syst. Saf. 161 (2017) 53–60, https://doi.org/10.1016/j.ress.2016.12.012.
[2] Federal Railroad Administration, Track and Rail and Infrastructure Integrity Compliance Manual: Volume II – Chapter 1 &2, Track Safety Standards, 2018.
[3] F. Peng, Y. Ouyang, K. Somani, Optimal routing and scheduling of periodic inspections in large-scale railroad networks, J. Rail Transp. Plan. Manag. 3 (4) (2013) 163–171.
[4] N.O. Attoh-Okine, Big Data and Differential Privacy: Analysis Strategies for Railway Track Engineering, John Wiley & Sons, 2017.
[5] F. Ghofrani, Q. He, R.M.P. Goverde, X. Liu, Recent applications of big data analytics in railway transportation systems: a survey, Transp. Res. Part C Emerg. Technol. 90 (2018) 226–246.
[6] A. Falamarzi, S. Moridpour, M. Nazem, A review of rail track degradation prediction models, Aust. J. Civ. Eng. 17 (2) (2019) 152–166.
[7] I. Soleimanmeigouni, A. Ahmadi, A. Nissen, X. Xiao, Prediction of railway track geometry defects: a case study, Struct. Infrastruct. Eng. 16 (7) (2020) 987–1001.
[8] D. Yurlov, A.M. Zarembski, N. Attoh-Okine, J.W. Palese, H. Thompson, Probabilistic approach for development of track geometry defects as a function of ground penetrating radar measurements, Transp. Infrastruct. Geotechnol. 6 (1) (2019) 1–20.
[9] J. Neuhold, I. Vidovic, S. Marschnig, Preparing track geometry data for automated maintenance planning, J. Transp. Eng. Part A Syst. 146 (5) (2020) 4020032.
[10] M. Sedghi, O. Kauppila, B. Bergquist, E. Vanhatalo, M. Kulahci, A taxonomy of railway track maintenance planning and scheduling: a review and research trends, Reliab. Eng. Syst. Saf. (2021), 107827.
[11] N. Rhayma, P. Bressolette, P. Breul, M. Fogli, G. Saussine, Reliability analysis of maintenance operations for railway tracks, Reliab. Eng. Syst. Saf. 114 (2013) 12–25.
[12] H. Khajehei, A. Ahmadi, I. Soleimanmeigouni, M. Haddadzade, A. Nissen, M. J. Latifi Jebelli, Prediction of track geometry degradation using artificial neural network: a case study, Int. J. Rail Transp. 10 (1) (2022) 24–43.
[13] A. Lasisi, N. Attoh-Okine, Principal components analysis and track quality index: A machine learning approach, Transp. Res. Part C Emerg. Technol. 91 (2018) 230–248.
[14] C. Vale, S.M. Lurdes, Stochastic model for the geometrical rail track degradation process in the Portuguese railway Northern Line, Reliab. Eng. Syst. Saf. 116 (2013) 91–98.
[15] J.A.P. Braga, A.R. Andrade, Multivariate statistical aggregation and dimensionality reduction techniques to improve monitoring and maintenance in railways: the wheelset component, Reliab. Eng. Syst. Saf. (2021), 107932.
[16] J. Chiachío, M. Chiachío, D. Prescott, J. Andrews, A knowledge-based prognostics framework for railway track geometry degradation, Reliab. Eng. Syst. Saf. 181 (2019) 127–141.
[17] N. Alemazkoor, C.J. Ruppert, H. Meidani, Survival analysis at multiple scales for the modeling of track geometry deterioration, Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. 232 (3) (2018) 842–850, https://doi.org/10.1177/0954409717695650.
[18] S. Sharma, Y. Cui, Q. He, R. Mohammadi, Z. Li, Data-driven optimization of railway maintenance for track geometry, Transp. Res. Part C Emerg. Technol. 90 (2018) 34–58, https://doi.org/10.1016/j.trc.2018.02.019.
[19] Q. He, H. Li, D. Bhattacharjya, D.P. Parikh, A. Hampapur, Track geometry defect rectification based on track deterioration modelling and derailment risk assessment, J. Oper. Res. Soc. 66 (3) (2015) 392–404.
[20] A.R. Andrade, P.F. Teixeira, Statistical modelling of railway track geometry degradation using Hierarchical Bayesian models, Reliab. Eng. Syst. Saf. 142 (2015) 169–183, https://doi.org/10.1016/j.ress.2015.05.009.
[21] S. Bressi, J. Santos, M. Losa, Optimization of maintenance strategies for railway track-bed considering probabilistic degradation models and different reliability levels, Reliab. Eng. Syst. Saf. 207 (2021), 107359.
[22] H. Li, D. Parikh, Q. He, B. Qian, Z. Li, D. Fang, A. Hampapur, Improving rail network velocity: A machine learning approach to predictive maintenance, Transp. Res. Part C Emerg. Technol. 45 (2014) 17–26.
[23] P.C.L. Gerum, A. Altay, M. Baykal-Gürsoy, Data-driven predictive maintenance scheduling policies for railways, Transp. Res. Part C Emerg. Technol. 107 (2019) 137–154.
[24] M. Mehrali, M. Esmaeili, S. Mohammadzadeh, Application of data mining techniques for the investigation of track geometry and stiffness variation, Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit 234 (5) (2020) 439–453.
[25] A. Kasraei, J.A. Zakeri, A. Bakhtiary, Optimal track geometry maintenance limits using machine learning: a case study, Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit 235 (7) (2021) 876–886.
[26] A. Bakhtiary, J.A. Zakeri, S. Mohammadzadeh, An opportunistic preventive maintenance policy for tamping scheduling of railway tracks, Int. J. Rail Transp. 9 (1) (2021) 1–22.
[27] E.T. Selig, J.M. Waters, Track geotechnology and substructure management, Thomas Telford, 1994.
[28] D. Li, J. Hyslip, T. Sussmann, S. Chrismer, Railway geotechnics, CRC Press, 2015.
[29] J. Hyslip, H.F. Kashani, M. Trosino, Ballast State of Good Repair, Am. Railw. Eng. Maint. W. Assoc. AREMA Annu. Conf. (2017).
[30] M. Esmaeili, J.A. Zakeri, S.A. Mosayebi, Effect of sand-fouled ballast on train-induced vibration, Int. J. Pavement Eng. 15 (7) (2014) 635–644.
[31] J. Sadeghi, M.E.M. Najar, J.A. Zakeri, C. Kuttelwascher, Development of railway ballast geometry index using automated measurement system, Measurement 138 (2019) 132–142.
[32] J. Sadeghi, M. Emad Motieyan, J. Ali Zakeri, Development of integrated railway ballast quality index, Int. J. Pavement Eng. 22 (1) (2021) 32–40.
[33] A.R.T. Kian, J.A. Zakeri, J. Sadeghi, Experimental investigation of effects of sand contamination on strain modulus of railway ballast, Geomech. Eng. 14 (6) (2018) 563–570.
[34] F. Khatibi, M. Esmaeili, S. Mohammadzadeh, Numerical investigation into the effect of ballast properties on buckling of continuously welded rail (CWR), Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit 235 (7) (2021) 866–875.
[35] M. Esmaeili, M. Paricheh, M.H. Esfahani, Laboratory investigation on the behavior of ballast stabilized with bitumen-cement mortar, Constr. Build. Mater. 245 (2020), 118389.
[36] H.F. Kashani, J.P. Hyslip, C.L. Ho, Laboratory evaluation of railroad ballast behavior under heavy axle load and high traffic conditions, Transp. Geotech. 11 (2017) 69–81.
[37] H.F. Kashani, C.L. Ho, J.P. Hyslip, Fouling and water content influence on the ballast deformation properties, Constr. Build. Mater. 190 (2018) 881–895.
[38] A.R. TolouKian, J. Sadeghi, J.-A. Zakeri, Large-scale direct shear tests on sand-contaminated ballast, Proc. Inst. Civ. Eng. Eng. 171 (5) (2018) 451–461.
[39] M. Fathali, J. Chalabii, F. Astaraki, M. Esmaeili, A new degradation model for life cycle assessment of railway ballast materials, Constr. Build. Mater. 270 (2021), 121437.
[40] H.F. Kashani, C.L. Ho, C.P. Oden, S.S. Smith, Model track studies by ground penetrating radar (GPR) on ballast with different fouling and geotechnical properties, in: ASME/IEEE Joint Rail Conference, 2015, vol. 56451, p. V001T01A006.
[41] T.R. Sussmann, E.T. Selig, J.P. Hyslip, Railway track condition indicators from ground penetrating radar, NDT e Int. 36 (3) (2003) 157–167.
[42] G.R. Olhoeft, E.T. Selig, Ground-penetrating radar evaluation of railway track substructure conditions, Ninth International Conference on Ground Penetrating Radar 4758 (2002) 48–53.
[43] R. Roberts, I. Al-Audi, E. Tutumluer, J. Boyle, Subsurface Evaluation of Railway Track Using Ground Penetrating Radar, 2008.
[44] H.F. Kashani, C.L. Ho, W.P. Clement, C.P. Oden, Evaluating the correlation between the geotechnical index and the electromagnetic properties of fouled ballasted track by a full-scale laboratory model, Transp. Res. Rec. 2545 (1) (2016) 66–78.
[45] J. Sadeghi, M.E. Motieyan-Najar, J.A. Zakeri, B. Yousefi, M. Mollazadeh, Improvement of railway ballast maintenance approach, incorporating ballast geometry and fouling conditions, J. Appl. Geophys. 151 (2018) 263–273, https://doi.org/10.1016/j.jappgeo.2018.02.020.
[46] L.B. Ciampoli, F. Tosti, M.G. Brancadoro, F. D'Amico, A.M. Alani, A. Benedetto, A spectral analysis of ground-penetrating radar data for the assessment of the railway ballast geometric properties, NDT E Int. 90 (2017) 39–47.
[47] Y. Guo, G. Liu, G. Jing, J. Qu, S. Wang, W. Qiang, Ballast fouling inspection and quantification with ground penetrating radar (GPR), Int. J. Rail Transp. (2022) 1–18.
[48] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, vol. 112, Springer, 2013.
[49] A.R. Andrade, P.F. Teixeira, Unplanned-maintenance needs related to rail track geometry, Proc. Inst. Civil Eng.-Transport 167 (6) (2014) 400–410.
[50] I. Soleimanmeigouni, A. Ahmadi, H. Khajehei, A. Nissen, Investigation of the effect of the inspection intervals on the track geometry condition, Struct. Infrastruct. Eng. 16 (8) (2020) 1138–1146.

[51] H. Khajehei, A. Ahmadi, I. Soleimanmeigouni, A. Nissen, Allocation of effective maintenance limit for railway track geometry, Struct. Infrastruct. Eng. 15 (12) (2019) 1597–1612.
[52] C. Vale, M.L. Simões, Prediction of Railway Track Condition for Preventive Maintenance by Using a Data-Driven Approach, Infrastructures 7 (3) (2022) 34.
[53] I. Balogun, N. Attoh-Okine, Random Forest-Based Covariate Shift in Addressing Nonstationarity of Railway Track Data, ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng. 7 (3) (2021) 4021028.
[54] C. Ding, X.J. Cao, P. Næss, Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo, Transp. Res. Part A Policy Pract. 110 (2018) 107–117.
[55] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[56] S. Touzani, J. Granderson, S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings, Energy Build. 158 (2018) 1533–1543.
[57] J. Zhou, E. Li, S. Yang, M. Wang, X. Shi, S. Yao, H.S. Mitri, Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories, Saf. Sci. 118 (2019) 505–518.
[58] C.P. Oden, C.L. Ho, H.F. Kashani, Man-Portable Real-Time Ballast Inspection Device Using Ground-Penetrating Radar. Railroad Ballast Testing and Properties, ASTM International, 2018.
[59] M. Silvast, A. Nurmikolu, B. Wiljanen, M. Levomaki, An inspection of railway ballast quality using ground penetrating radar in Finland, Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit 224 (5) (2010) 345–351.
[60] J.C.O. Nielsen, E.G. Berggren, A. Hammar, F. Jansson, R. Bolmsvik, Degradation of railway track geometry–Correlation between track stiffness gradient and differential settlement, Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit 234 (1) (2020) 108–119.
[61] S. Goodarzi, H.F. Kashani, S. Chrismer, C.L. Ho, Using large datasets for finding the correlation between the rate of track settlement and changes in geometry indices, Transp. Geotech. 31 (2021), 100665.
[62] S.-A. Mosayebi, M. Esmaeili, J.-A. Zakeri, Dynamic train–track interactions and stress distribution patterns in ballasted track layers, J. Transp. Eng. Part B Pavements 146 (1) (2020) 4019042.
[63] M. Naeimi, J.A. Zakeri, M. Esmaeili, M. Shadfar, Influence of uneven rail irregularities on the dynamic response of the railway track using a three-dimensional model of the vehicle–track system, Veh. Syst. Dyn. 53 (1) (2015) 88–111.
[64] R. Mohammadi, Q. He, F. Ghofrani, A. Pathak, A. Aref, Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects, Transp. Res. Part C Emerg. Technol. 102 (2019) 153–172.
[65] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[66] G. Douzas, F. Bacao, F. Last, Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE, Inf. Sci. (Ny) 465 (2018) 1–20.
[67] J. Lessan, L. Fu, C. Wen, A hybrid Bayesian network model for predicting delays in train operations, Comput. Ind. Eng. 127 (2019) 1214–1222.
[68] Q. Wang, S. Bu, Z. He, Achieving predictive and proactive maintenance for high-speed railway power equipment with LSTM-RNN, IEEE Trans. Ind. Informatics 16 (10) (2020) 6509–6517.
[69] H. Ling, C. Qian, W. Kang, C. Liang, H. Chen, Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine environment, Constr. Build. Mater. 206 (2019) 355–363.
[70] Y. Jung, Multiple predicting K-fold cross-validation for model selection, J. Nonparametr. Stat. 30 (1) (2018) 197–215.