Construction and Building Materials 365 (2023) 130066

Contents lists available at ScienceDirect

Construction and Building Materials

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

Relationship between track geometry defect occurrence and substructure condition: A case study on one passenger railroad in the United States

Xin Wang [a], Xiang Liu [a,*], Todd L. Euston [b]

[a] Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
[b] KCI Technologies, Inc, USA

ARTICLE INFO

Keywords:
Track geometry defect
Passenger railroad
Substructure condition
Data-driven approach
Gradient boosting

ABSTRACT

Analyzing the relationship between track geometry defect occurrence and substructure condition can provide assistance for track inspection and spot maintenance, which contributes to better train operation quality. This paper develops a data-driven approach to estimate the occurrence of track geometry defects on concrete-tie tracks on one passenger railroad in the United States, using substructure data, rail seat abrasion data, infrastructure data, traffic data, track class information, and maintenance data. Feature extraction was implemented to generate input variables for the machine learning models. Recursive feature elimination (RFE) was applied to reduce data dimensionality by recursively considering smaller sets of features. Three data treatment methods, including no resampling, undersampling, and oversampling, were incorporated to address imbalanced data issues. The developed models included logistic regression, artificial neural network, and gradient boosting. The hyperparameters of the proposed models were optimized using Bayesian optimization. The performance of the proposed methods was finally evaluated based on the test dataset generated using random data partitioning. Based on data collected from one passenger railroad, the gradient boosting method with data oversampling shows the highest performance in estimating the occurrence of geometry defects. The F1-score of the model is 0.662, with G-Mean of 0.738. Feature importance identifies that surfacing, traffic, curvature, switch, and rail replacement are the top five factors influencing the predicted probability of track geometry defect occurrence. The proposed model can be used to prioritize maintenance activities on locations prone to track geometry defects and thus further improve infrastructure safety given budgetary constraints.

1. Introduction

1.1. Overview

Track geometry defects are considered to be one of the most important factors affecting the stability and safety of train operations [1,2]. They result in a significant increase in dynamic loads between rail and wheel leading to the reduction of passenger comfort, earlier development of rail fatigue failure, and an associated decrease in rail service life [3,4]. For instance, based on the FRA rail equipment accident database from 2001 to 2010, track geometry is the second largest cause of derailments [5]. While track geometry indicates track deterioration, the substructure information allows observers to identify causal factors of deterioration. Ground penetrating radar (GPR) is a non-destructive method for routine substructure inspections that enables the evaluation of railroad conditions in terms of layers geometry, settlements, fouling level, and drainage assessment [6–8]. The information provided by GPR is crucial for efficient maintenance action and depth of the intervention of track when track geometry defects are caused by substructure conditions.

Due to relatively higher initial costs, concrete-tie tracks are only economical in applications when they have a longer service life and require less maintenance/inspection than wood-tie tracks [9]. To limit the costs of maintenance and duration of traffic interruptions, spot maintenance of track geometry defects on concrete-tie tracks has to be well planned. To enable this, it is necessary to investigate the influence of various factors, such as track substructure, traffic, and infrastructure factors, on the occurrence of track geometry defects on concrete-tie tracks.

Many researchers have previously studied the influence of differing

* Corresponding author at: Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, 500 Bartholomew Road, Room 428D, Piscataway, NJ 08854, USA.
E-mail address: xiang.liu@rutgers.edu (X. Liu).

https://doi.org/10.1016/j.conbuildmat.2022.130066
Received 9 August 2022; Received in revised form 30 November 2022; Accepted 12 December 2022
Available online 21 December 2022

track components on track geometry degradation using physics-based methods or data-driven methods [10–12]. The physics-based models reveal the mechanical behavior as well as the deformation mechanisms of the track geometry. However, it is difficult for mechanical models to simultaneously provide all location-specific risks in one railroad line by considering various influencing factors. On the other hand, data-driven methods allow the consideration of multiple sources of data and predict the location-specific risk along the railroad line with low computational cost [13,15]. In this paper, researchers develop a data-driven framework for estimating the probability of track geometry defects under various factors on concrete-tie track on one passenger railroad in the United States, using substructure data, rail seat abrasion data, infrastructure data, traffic data, track class information, and maintenance data. The proposed framework consists of a pipeline of methodologies including data processing and model development, which has been applied and tested on a passenger railroad.

1.2. Knowledge gaps and intended contributions

Previous studies have shown that track geometry degradation and the occurrence of geometry defects are subject to various factors including multiple substructure conditions [14–17]. Most of the previous data-driven research focuses on analyzing the relationship between track geometry defects and substructure on freight railroads [4,24–26]. But to date, there has been little research on estimating the risk of track geometry defects on passenger railroads using substructure condition and other related factors (e.g., maintenance activities) [23]. Furthermore, the importance level of multiple factors affecting the geometry defects on passenger railroads has not been studied yet. There remains a research gap in analyzing geometry defects on passenger railroads by simultaneously considering multiple substructure conditions measured by GPR, maintenance activities, and other related factors.

Considering the limitations of prior research mentioned above, the main intended contributions of this paper are summarized as follows.

- This paper customizes a framework for estimating the risk of track geometry defects on concrete-tie track on one passenger railroad via a data-driven approach using multi-source data.
- More influencing factors are considered in this paper, including substructure conditions measured by GPR, rail seat abrasion, infrastructure, traffic, track class, and maintenance activities. Some of them have not been considered in the previous studies on geometry defects, such as multiple substructure conditions and maintenance activities.
- To improve model performance, data resampling including undersampling and oversampling are implemented to solve the imbalanced data issue. Multiple meaningful metrics are discussed and reported for each of the prediction methods.
- The importance level of selected factors is identified to provide insights into the significance of factors affecting geometry defects on passenger railroads.

The organization of the article is as follows. Section 2 talks about the previous related work and the state-of-the-art studies. Section 3 provides an overview of the data collected from one passenger railroad. In Section 4, the methodology of this study is presented. Then, the results of different models are compared in Section 5. The conclusions and future research direction are provided in Section 6.

2. Literature review

Previous research identified that the degradation of track geometry is affected by traffic [10–16], curve [10–12], grade [12], maintenance activities [14,16], track class/speed [11,15], and substructure (ballast fouling index) [17]. This section reviews state-of-the-art data-driven models for analyzing factors influencing track geometry defects as well as the relationship between track geometry defects and substructure conditions. Table 1 summarizes studies that are the most related to the context of this paper.

The probability of track geometry defects under different influencing factors has been studied. Sadeghi and Askarinejad [18] applied artificial neural network (ANN) to determine the track geometry defects and track structural conditions including rail, sleeper, ballast, and fastening. Logistic regression was applied to predict the probability that a given track segment will generate a ballast-related track geometry exception as a function of key independent variables such as missing ballast, annual MGT, and curvature [19]. This confirmed the relationship between missing ballast and the development of track geometry defects. Further, a high order polynomial logistic regression model in combination with hierarchical clustering analysis was proposed to determine the probability of a track geometry (profile) defect occurring at segments with substructure conditions [20,21]. The convolutional neural network (CNN) model was developed to estimate the probability of geometry defect as a function of sleeper condition and sleeper positions [22]. Gradient boosting was applied to predict the profile defects using geometry measurement data, traffic density, track class, and ballast fouling index [23].

In terms of the study of the relationship between track geometry defects and substructure conditions, most previous studies focused on freight railroads [4,24–26]. Alsahli et al. [24] found that track segments with geometry defects are associated with poorer tie conditions by comparing statistical distributions of wood tie conditions on track segments with and without geometry defects. Yurlov et al. [20] indicated a statistically significant relationship between track geometry defects and

Table 1
Summary of Reviewed Papers related to Track Geometry Defects.

Objective	Approaches	Factors	Railroad Type	Authors
Relationship with tie condition	Statistical analysis	Wood tie condition	Freight railroad	Alsahli et al. [24]
Relationship with missing ballast		Missing ballast		Zarembski et al. [26]
Relationship with rail defects	MARS	Rail defects		Zarembski et al. [25] Zarembski et al. [4]
Probability of geometry defect	ANN	Rail, sleeper, ballast, and fastening	Not mentioned	Sadeghi and Askarinejad [18]
	Logistic regression	Missing ballast, annual MGT, and curvature	Freight railroad	Zarembski et al. [19]
		Substruction condition (BFI, BLT)		Yurlov et al. [20]; Zarembski et al. [21]
	CNN	Sleeper conditions		Alsahli et al. [22]
	Gradient boosting	BFI, standard deviation of profile, defect ratio, traffic density, and track class	Passenger railroad	Goodarzi et al. [23]
		Multiple substructure conditions (BTI, FDL, BFI, BDM, and TDI), curvature, switch, annual MGT, track class, rail replacement, and surfacing		This paper

key track subsurface conditions. In particular, the ballast fouling index (BFI) and ballast layer thickness (BLT) measured by GPR are well correlated with track geometry defects. Another statistical analysis found that increasing volumes of missing ballast results in increases in geometry defect occurrences [26]. Multivariate regression splines (MARS), a non-parametric function, was applied to account for the quantitative relationship between geometry defects and rail defects [4,25]. The study demonstrated a significant relationship between geometry defects and the occurrence of rail defects. Researchers also found that most crosslevel track defects are usually caused by ballast settlements [50,51].

3. Data overview

The data collected for this study comes from one passenger rail agency in the United States. It has around 220 miles of concrete-tie track containing about 571,000 concrete ties. This research collected track geometry defect data (from 2020 to 2021), substructure conditions (from 2019 to 2020), rail seat abrasion data (from 2020), infrastructure data, traffic, track class, and maintenance activity data (from 2019). Due to data limitations, the dates of collecting geometry data and GPR data are not the same. This research assumes that the substructure conditions do not deteriorate over a short period. Therefore, GPR data is used to analyze its relationship with track geometry defects. The following subsections describe the data in detail.

3.1. Track geometry defect data

Track geometry describes the surface of the track. The amplitudes of the track geometry are inspected by track geometry inspection cars twice a year in the studied railroad network. Track geometry defects occur when the amplitudes of specific measures (e.g., gage) exceed an established threshold in the FRA's Track Safety Standards (49 CFR Part 213) [43]. FRA divides track quality into different classes corresponding to different maximum train speeds [53]. Requirements of geometry prescribed in Track Safety Standards are more stringent for higher classes. For example, the deviation of the mid-offset of alignment on tangent track from a 62-foot line may not be more than 5, 3, 1.75, 1.5, and 0.75 in. for Class 1–5 track, respectively. Gage must be at least 4′8″ for Class 1–5 track but not more than 4′10″ for Class 1 track, 4′9.75″ for Class 2 and 3 track, and 4′9.5″ for Class 4 and 5 track. Relevant features of defects such as location, found date, type, size (length), and other related data are recorded [54]. Three track geometry inspections were collected in spring 2020, fall 2020, and spring 2021. A binary variable was used to indicate if a track section had a geometry defect between 2020 and 2021. Track geometry defects that are presumed to be related to substructure conditions include profile, alignment, gage, crosslevel, excess elevation, twist/warp, and curve speed, which are used for this study.

3.2. Substructure condition data

Substructure condition data in this study refers to ballast conditions.

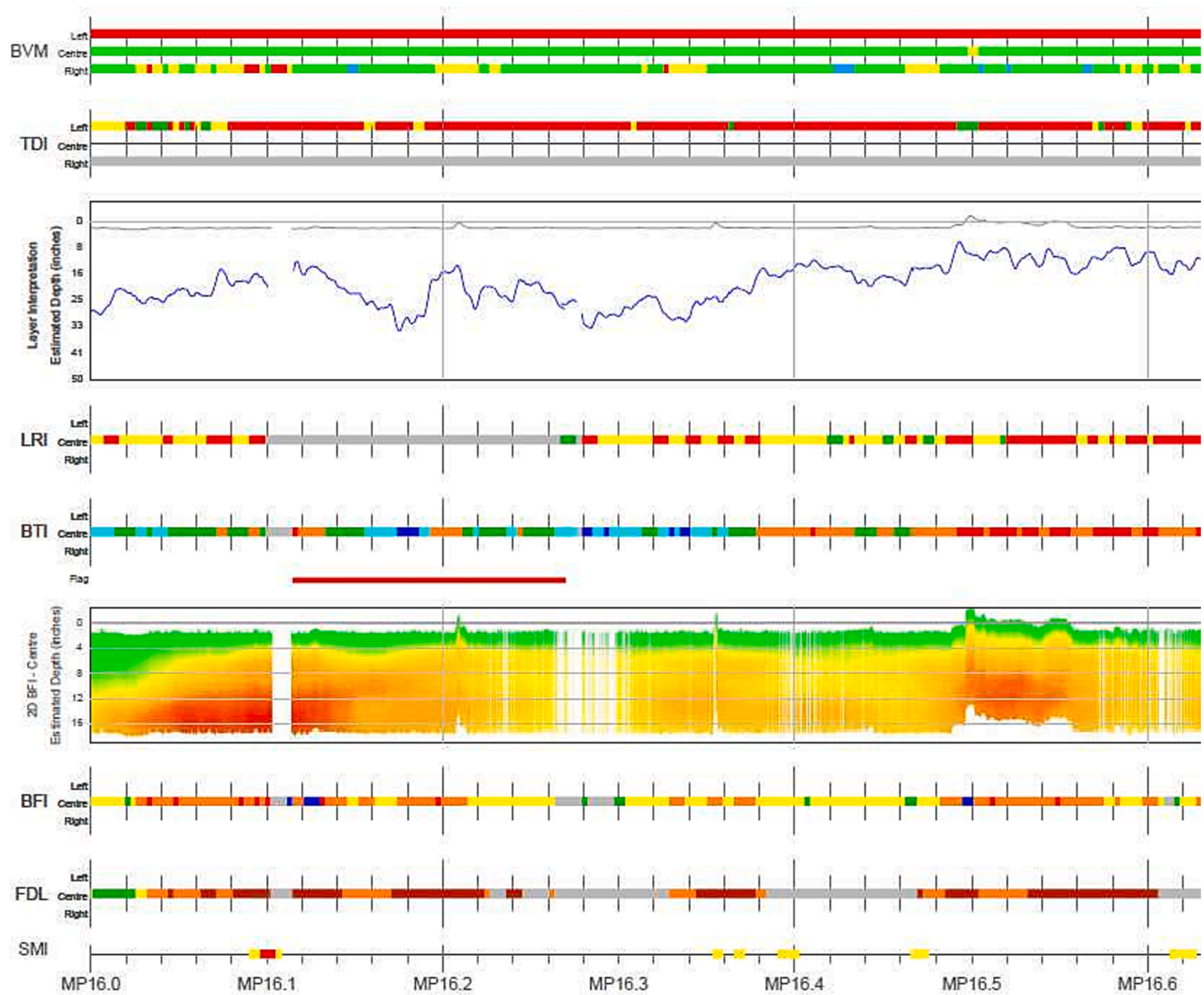

Fig. 1. GPR Data Reports from Milepost 16.0 to Milepost 16.6.

GPR uses a reflection of radar waves in the 300 to 400 MHz range to identify properties in the ballast along the railroad line. Fig. 1 shows the digitization of partial ballast conditions [21,52]. In this research, nine parameters (BFI, FDL, BTI, LRI, BVM, BDM, TDI, SMI, and TCS) of GPR data were collected to indicate the ballast conditions from 2019 to 2020, which are shown in detail below.

Ballast fouling index (BFI) measures the percentage of fine material trapped in the ballast and sub-ballast layer to the modeled average depth. Fouling depth layer (FDL) is a measure of the thickness of clean ballast. Ballast thickness index (BTI) indicates track sections where the thickness of the primary ballast layer falls outside of an optimum range as defined by the standard track bed design thickness. Layer roughness index (LRI) reflects the variation in the depth to the base of the primary track bed layer. Ballast volume metric (BVM) measures the excess or deficit ballast volumes relative to a predefined ballast cross-section. Ballast deficit metric (BDM) shows the deficit in ballast volume relative to a predefined track bed design profile (cross-section). Track drainage index (TDI) reflects the efficiency of the track drainage at a location. Surface mud spot index (SMI) indicates the amount of mud in the ballast or on the sleepers/rails. Track bed condition summary (TCS) provides an indication of the overall quality of track bed by combining various GPR measurements.

Table 2 presents more details of the parameters included in the GPR database.

3.3. Rail seat abrasion data

Rail seat abrasion (RSA) refers to the deterioration of concrete beneath the rail [41,42], as shown in Fig. 2. It is measured using a vision-based inspection system: the Aurora inspection system [27]. This study collected RSA information included in the Aurora dataset which has around 571,000 concrete ties. To be specific, the left-rail-side and right-rail-side RSA values are measured separately and refer to the estimated depth of degradation of rail seat beneath each rail. Greater values indicate worse conditions. Additionally, the RSA database contains concrete tie ID, a binary variable used to indicate whether the tie inspected is a concrete tie.

3.4. Infrastructure data

Infrastructure data in this research includes curvature data and turnout (switch) data. Curvature data shows the curve degree indicating the track's horizontal alignment. The distribution of the curve degree in the studied railroad line is shown in Table 3. It indicates that 52 % of the total railroad network is tangent track.

Turnout data is a binary variable specifying whether a track is within a turnout structure. When the whole network is divided into 71,096 16-feet-long segments (see details in Section 4.2), 2,742 track segments are in turnout area, accounting for 3.9 % of the total network.

3.5. Traffic data

Traffic loading has a significant influence on the deterioration of track geometry [10–13]. Traffic data collected for this research specifies the track segments and corresponding annual gross tonnage. Fig. 3 illustrates the distribution of the annual gross tonnage of the studied network. The portion of the network decreases as the traffic density increases. Around half of the network has traffic density from 15 to 20 MGT while track with traffic density from 35 to 40 MGT only takes up 2.3 % of the whole network.

Table 2
Parameters of Ballast Condition Data.

No.	Parameters	Number of Categories	Inspection Location on Track
1	BFI	5	Left, Center, and Right
2	FDL	4	Left, Center, and Right
3	BTI	5	Left, Center, and Right
4	LRI	3	Left, Center, and Right
5	BVM	5	Left, Center, and Right
6	BDM	3	Left, Center, and Right
7	TDI	3	Left and Right
8	SMI	3	Track
9	TCS	3	Track

3.6. Track class information

The Federal Railroad Administration (FRA) has established a classification system for railroad track quality. The classification of a track indicates specific construction details, including tolerance requirements for the geometrical measurements of the track, which finally determine the speed limits for both freight and passenger trains. FRA track classification information was collected for this research. Table 4 illustrates the distribution of track class in the studied network.

3.7. Maintenance data

Maintenance data in this study reflect maintenance activities from 2019 including tie production, rail replacement, and surfacing. Tie production data records the segments of track where ties were replaced. Rail wear is considered as the main cause of rail replacement [28]. The rail replacement data provides the location and date of replacement activities. Surfacing activities such as tamping and ballast renewal are applied not only to provide a firm foundation for ties but also to bed them so that the track will not be thrown out of line by the lateral thrust of passing trains [40].

In this research, three types of maintenance activity are defined as three binary variables indicating whether a track experienced a certain type of maintenance action in 2019, the year prior to when the geometry data was collected. In this period, 17.0 % of the railroad network experienced maintenance activities. The number of track segments having tie production, rail replacement, and surfacing is 89, 1,943, and 10,024, respectively.

4. Methodology

A data-driven approach is proposed in this paper to estimate the occurrence of track geometry defects in concrete-tie track segments on one passenger railroad. Fig. 4 shows the methodological framework of the proposed research. First, data cleaning was applied to identify concrete-tie track segments, handle missing values, and combine data that have multiple data values (e.g., BDM) at a given longitudinal location into a single value per track location. Then, all datasets were combined into an integrated dataset according to their location information. The GPR data with 16-foot-long segments was used as the base file in the integration process. Feature engineering was implemented to extract features (input variables) from the integrated dataset. Next, the integrated dataset was split into training, validation, and test data by randomly partitioning the integrated data. Feature selection was conducted to reduce the dimensionality of data based on training data. Three data treatment methods were incorporated into data-driven models to address imbalanced data issues. Finally, the model fit on training and validation data was evaluated using test data. The best-fit model was applied to estimate the risk of track geometry defects in the railroad.

4.1. Data cleaning

Data cleaning aims to identify concrete-tie tracks, handle missing values, and combine the information of left side, central, and right side of track. Both concrete-tie tracks and wood-tie tracks are included in the studied railroad network. While the concrete tie ID in the Aurora dataset indicates if the tie inspected is a concrete tie, this could be falsely

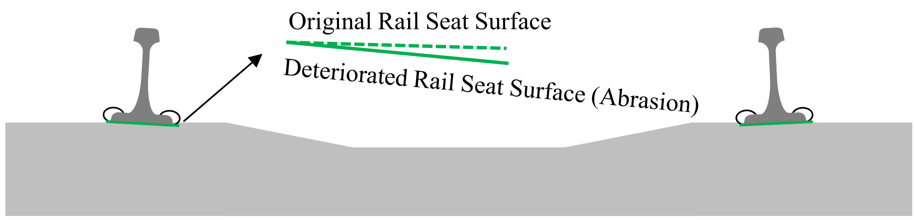

Fig. 2. Concrete Tie Rail Seat Abrasion.

Table 3
Distribution of FRA Track Class in Studied Railroad Network.

Curve Degree	0	0–1	1–2	2–3	3–4	Over 4	Total
Track Miles	114.0	47.9	40.9	10.7	3.6	2.1	219.2
Portion of the Whole Network	52.0 %	21.8 %	18.6 %	4.9 %	1.7 %	1.0 %	100 %

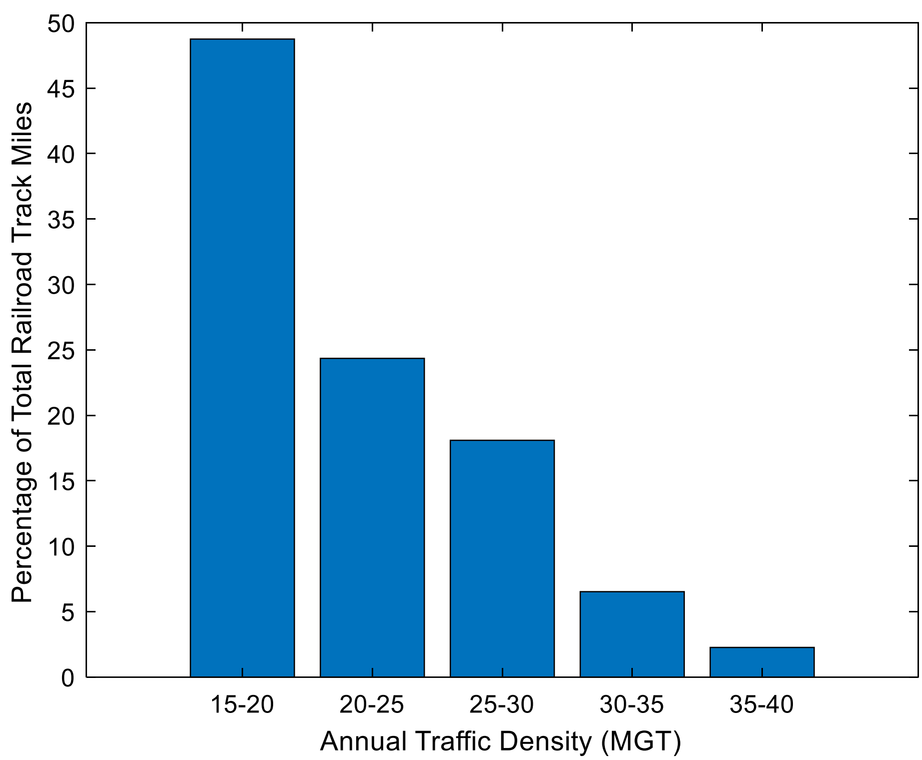

Fig. 3. Distribution of the Annual Traffic Density Over the Studied Railroad Network.

Table 4
Distribution of FRA Track Class in Studied Railroad Network.

FRA Track Class	Class 2	Class 3	Class 4	Total
Track Miles	3.7	57.1	158.4	219.2
Portion of the Whole Network	1.7 %	26.0 %	72.3 %	100 %
Limited Speed for Passenger Railroad (MPH)	30	60	80	–

reported when inspection machines cannot get a clear view of the tie surface due to mud, ballast, leaves, or other obstructions. To minimize the impact of this error, if over 50 % of ties are concrete ties in a segment, this segment is defined as concrete-tie track and is finally included in the studied network. For instance, for a track segment with eight ties, if there are five concrete ties labeled as concrete ties in the Aurora dataset, this segment is defined as a concrete-tie segment.

Missing values were filled with the parameter values of the nearest segments. Sometimes, but not often, errors caused by inspection machines, classification rules, and human factors may cause missing values. BFI, for example, could not be calculated due to high electromagnetic induction (EMI) or the presence of a surface / sub-surface structure, which results in a missing value.

Some parameters differentiate among the left side, central, and right side of the track, which is defined as track-side-level data in this research. For instance, values of BDM of different track sides are separately measured. On the other hand, some datasets are recorded on the track level called track-level data. This paper focuses on track-level analysis because some important factors such as traffic and curvature are track-level data. Therefore, the track-side-level information needs to be transformed into track-level information. The strategy for information combination for different track sides is to retain the records with the worst condition information in the database. When the track-side-level data are inconsistent, they will be integrated into track-level data where the records with the worst condition are finally kept in the

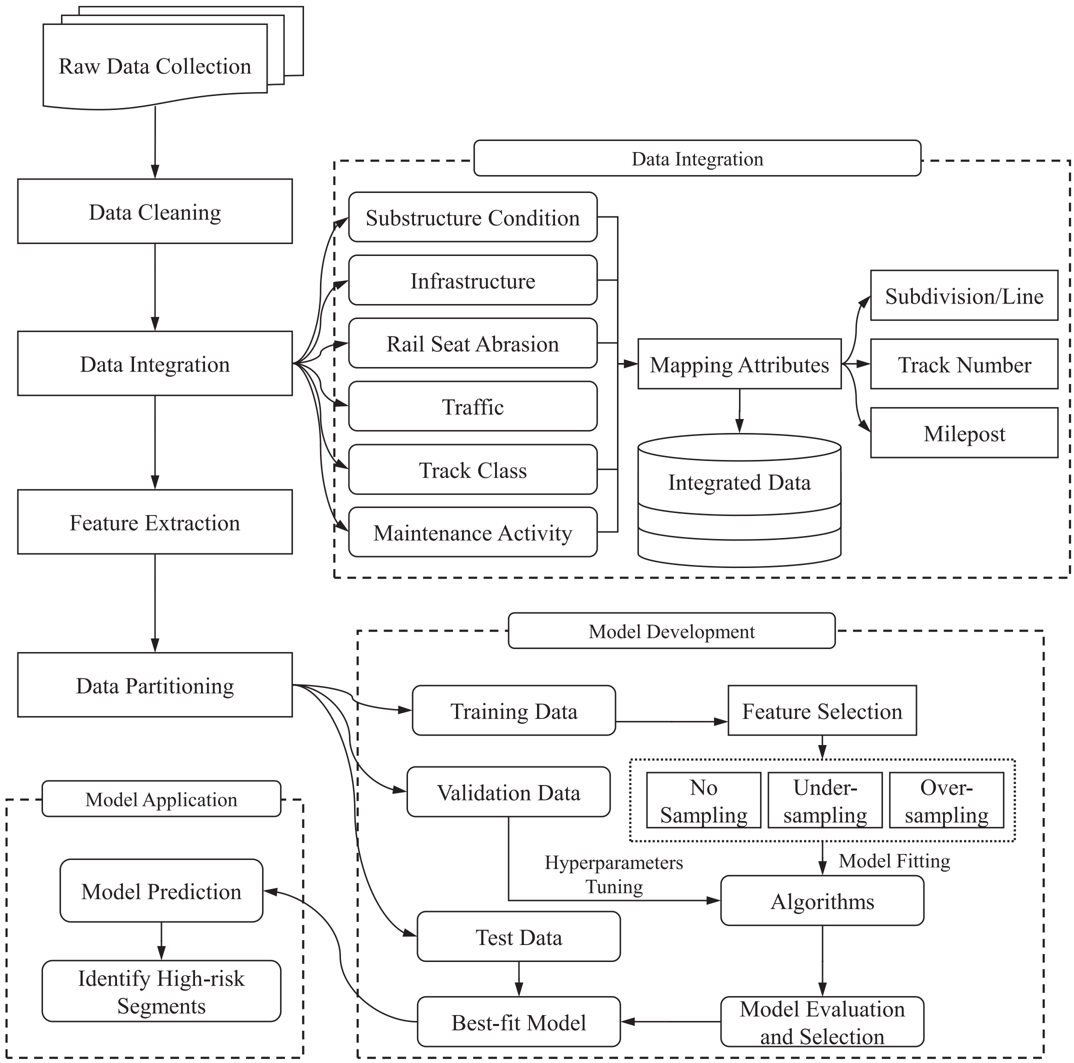

Fig. 4. The Framework for Estimation of Track Geometry Defect Risk.

database, as shown in Fig. 5. For instance, if BDM values measured on the left side, central, and right side of a track at the identical location (same track mile) are different, the minimum category value among different records would be assigned to this segment.

4.2. Data integration and feature extraction

This section introduces the data integration and feature extraction process in the following steps. Spatial information was used to identify segments and map all datasets. Then, feature extraction was applied to transform the integrated data into input variables of data-driven models. Fig. 6 illustrates the data integration process and preliminary input variables generated for the data-driven model.

Step 1: Generate segments for the whole network. The ballast condition database (GPR data) is used as the base file for integration, where each segment is 16 feet long and can be uniquely identified using location information including subdivision/line, track type, and milepost. Railroad staff can easily identify exact track locations that are prone to geometry defects using 16-foot-long segments without excessive inspection efforts.

Step 2: Map RSA for each segment according to location information.

Step 3: Map infrastructure data, traffic density data, FRA track class data, and maintenance data for each concrete-tie segment.

Step 4: Extract features for the data-driven models based on the integrated database.

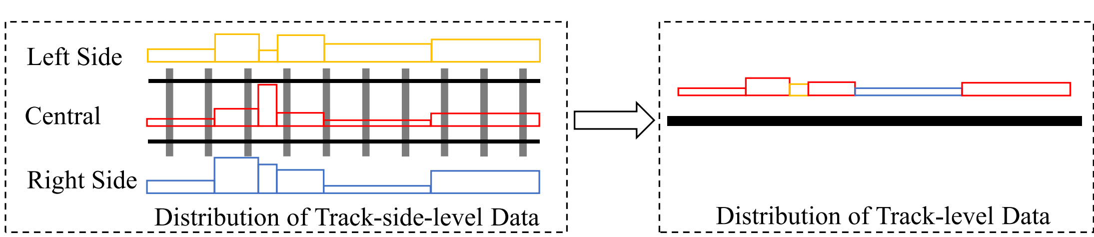

Fig. 5. Diagrammatic Sketch of Information Combination for Rail-level Data.

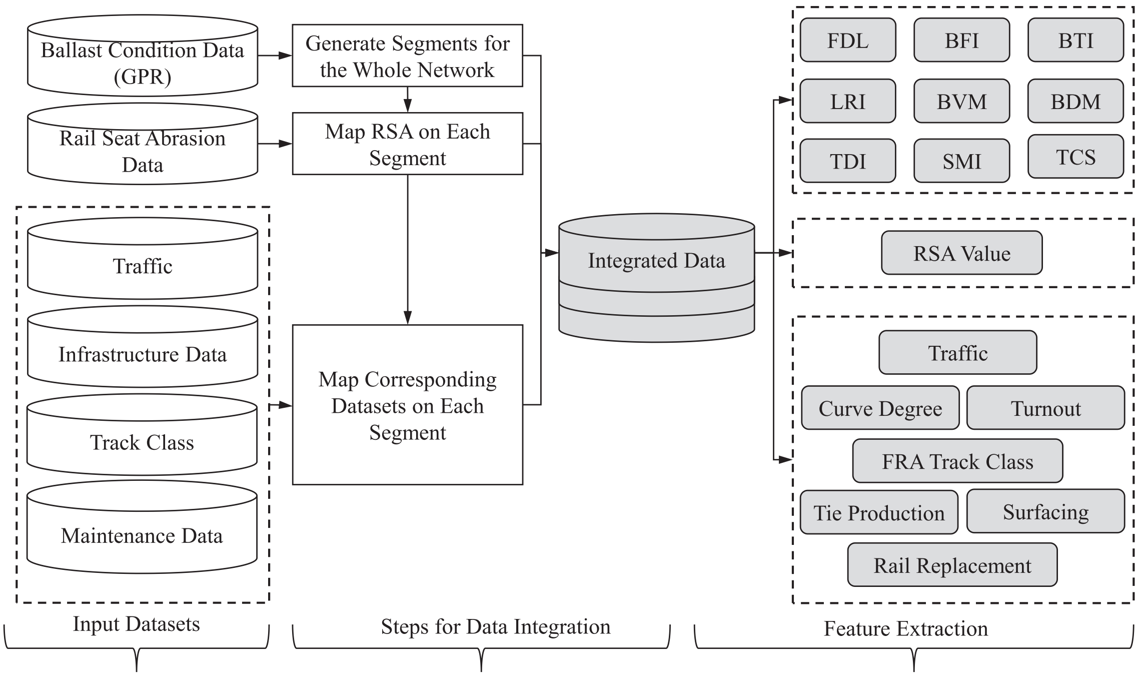

Fig. 6. Steps for Data Integration and Feature Extraction.

4.3. Data partitioning

Random data partitioning was introduced to avoid an over-optimistic estimation of the model by splitting the entire dataset into three parts: training, validation, and test datasets. The training dataset was used to fit the parameters of the data-driven model. The validation dataset was to provide an unbiased evaluation of the trained model during the hyperparameter tuning of the models. The test dataset was used for evaluating the generalization performance of models. This study uses 60 %, 20 %, and 20 % of the entire dataset as training, validation, and test datasets, respectively.

4.4. Feature selection

To lower the computation costs and increase the performance of machine learning models, feature selection is introduced by reducing the number of input variables [44]. The principal component analysis (PCA) [47,48] and regularization techniques (e.g., Ridge or Lasso regularization) [49] were applied in railroad engineering for feature selection.

Recursive feature elimination (RFE) is applied for feature selection in this research because it is easy to configure and effective at selecting the most relevant features contributing to predicted results [45]. It selects features by iteratively considering progressively smaller sets of features. First, the model is fit on initial features. Then, the feature importance is calculated using the trained model. The least important feature is removed from the current feature list because it contributes the least to the predicted results. This procedure is recursively repeated on the pruned set of features until further removal of features cannot improve the models' performance. Eventually, the final input variables for data-driven models are determined.

4.5. Data resampling

Data resampling refers to procedures of undersampling and oversampling, which balances data by reducing the majority class samples or resampling the minority class samples [29–31]. In the studied railroad network, only 2.4 % of total segments have experienced geometry defects. However, misclassifying those segments would incur more costs. Therefore, data resampling was introduced to balance the data. Random undersampling was applied in this research by randomly selecting samples of the majority class without replacement. He et al. [32] developed an adaptive synthetic sampling approach (ADASYN) to focus more on data samples that are hard to learn. It applies a weighted distribution for minority class samples according to their difficulty of learning. More synthetic data is generated for minority class examples that are harder to learn. The synthetic samples are created based on the majority nearest neighbors via the k-NN method. This study adopted ADASYN as the oversampling technique in the machine learning models.

4.6. Data-driven models

For a given dataset $D = \{(x_i, y_i)|i = 1, 2, \cdots, n\}$, x_i is inputs of model having n observations and m features (i.e., $x_i \in \mathbb{R}^m$), and y_i is actual label. In this study, y_i is a binary variable indicating if the observation is associated with a geometry defect. Three data-driven models (i.e., logistic regression, ANN, and gradient boosting) are applied in this research, which is presented in detail as follows.

4.6.1. Logistic regression

Logistic regression fits the logit probability of the response variable as linear function of input variables (Eq.(1)). It not only provides a measure of how appropriate an input variable is, but also indicates the direction of influence of each input variable on the response variable (positive or negative). However, the main limitation of logistic regression is the assumption of linearity between the dependent variable and the input variables. Thus, it cannot be applied to solve the nonlinear problem.

$$P(y|\boldsymbol{x}_i) = \frac{1}{1 + e^{-(\boldsymbol{w}^T\boldsymbol{x}_i)}} \tag{1}$$

where $P(y|\boldsymbol{x}_i)$ is the predicted probability ranging from 0 to 1, $\boldsymbol{w}$ is a parameter vector ($\boldsymbol{w} = [w_0, w_1, \cdots, w_m]$), T is the transpose of a vector. Maximum likelihood estimation (MLE) is applied to estimate the parameters $\boldsymbol{w}$ that maximize the conditional likelihood of $\prod_{i=1}^{n} P(y_i|\boldsymbol{x}_i)$ using Eq. (2).

$$\widehat{w} = \underset{w}{\operatorname{argmax}} - \sum_{i=1}^{n} log\left(1 + e^{-w^T x_i}\right) = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} log\left(1 + e^{-w^T x_i}\right) \tag{2}$$

where $\widehat{w}$ is the estimated parameter using MLE.

4.6.2. Ann

ANN is also known as multi-layer perceptron (MLP) [33]. It consists of multiple layers of neurons where each is fully connected to those in the layers below and above. Different from logistic regression, between the input and output layers of ANN, there can be one or more nonlinear layers. It can be applied to complex nonlinear problems. However, it cannot provide to what extent the response variable is affected by each input variable.

The first layer of ANN is the input layer, and its units take the values of the input variable x_i, as shown in Eq (3). In any feed-forward neural network, any middle layers are called hidden layers because their inputs and outputs are masked by the activation function (Eq. (4)). The last layer is the output layer, and it has a single unit in the case of binary classification, which is calculated by Eq. (5).

$$h_i^{(1)} = \Phi^{(1)}\left(\sum_j w_{ij}^{(1)} x_i + b_i^{(1)}\right) \tag{3}$$

$$h_i^{(k)} = \Phi^{(k)}\left(\sum_j w_{ij}^{(k)} h_j^{k-1} + b_i^{(k)}\right) \tag{4}$$

$$\widehat{y}_i = \Phi^{(K)}\left(\sum_j w_{ij}^{(K)} h_j^{K-1} + b_i^{(K)}\right) \tag{5}$$

where $h_i^{(k)}$ is the units in k layer, $w_{ij}^{(k)}$ is weight parameter for in k layer, $b_i^{(k)}$ is the bias parameter for k layer, $\widehat{y}_i$ is the output of the neural network, $\Phi^{(k)}$ is the activation function for the k layer, k is the number of entire layers, having K in total. The cross-entropy loss is applied as the loss function of ANN model, which measures the distance between the model output $\widehat{y}$ and actual labels y. It is defined in Eq. (6) shown as follows.

$$l(y, \widehat{y}) = -\sum_{j=1}^{n} t_j \log(P_j) \tag{6}$$

where t_j is the actual class, p_j is the probability for the j^{th} class, n is the number of the total classes. For the binary classification problem in this research, n is set to 2. The training process of the ANN is adjusting the connection weights to minimize the loss function l. Using gradient descent, the change in each weight is calculated by Eq. (7–8).

$$\Delta w_{ij}^{(k)} = -\eta \frac{\partial l}{\partial w_{ij}^{(k)}} \tag{7}$$

$$\Delta b_i^{(k)} = -\eta \frac{\partial l}{\partial b_i^{(k)}} \tag{8}$$

where η is the learning rate, $\Delta w_{ij}^{(k)}$ and $\Delta b_i^{(k)}$ are the changes in weight and bias parameters, respectively. This process (Eq. (3)–(8)) is recursively repeated until the loss function is minimized.

In this research, a five-layer MLP is applied, and the number of hidden neurons for each hidden layer is 10, 20, 40,20, and 10, respectively. The rectified linear unit (Relu) function is used as the activation function for the hidden layer. The sigmoid function is applied as the activation function for the output layer.

4.6.3. Gradient boosting

Gradient boosting (GB) creates the ensemble classifier in an iterative fashion. XGBoost is a scalable tree-based ensemble machine learning algorithm that uses a GB framework, as shown in Eq. (9) [34]. It is easy to implement and does not require normalization for input variables because it is a tree-based algorithm.

$$\widehat{y}_i = \sum_{k=1}^{K} \alpha_k f_k(x) \tag{9}$$

where f_k is an independent tree structure, α_k is the weight of the classifier $f_k(x)$, and k is the iteration step having K steps in total. XGBoost is optimized by parallel processing, tree-pruning, and handling regularization to avoid overfitting. The regularized objective function $\mathscr{L}$ is shown in Equation (10), which is to be minimized in the training process.

$$\mathscr{L} = \sum_{i=1}^{n} l(y_i, \widehat{y}_i) + \sum_k \Omega(f_k) \tag{10}$$

$$\Omega(f_k) = \gamma P + \frac{1}{2}\lambda \|\omega\|^2 \tag{11}$$

where ω is the leaf weight in the tree structure model f_k, γ is the penalty value to penalize the complexity of the model, P is the number of terminals or leaves, λ is the scale to perform the regularization process.

GB is trained in an additive manner. Suppose that the prediction of i data sample (x_i) at the t^{th} iteration is denoted by $\widehat{y}_i^{(t)}$. Then, the model traning is to minimize the objective function (Eq.(10)) with f_k added that most improves the model according to Eq.(12).

$$\mathscr{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i)\right) + \sum_{k=1}^{K} \Omega(f_k) \tag{12}$$

4.7. Hyperparameter optimization

The hyperparameter refers to the configuration of the model whose setting cannot be estimated from the training process. For example, XGBoost normally has five hyperparameters need to be tuned including the learning rate, subsample ratio of columns, regularization term on weights, maximum depth of a tree, and the number of gradient boosted trees. They can be optimized by selecting the appropriate set of hyperparameters to avoid overfitting or underfitting. Overfitting occurs when the model learns a function with a very high variance to perfectly model the training data [35]. Underfitting refers to a phenomenon where the model is incapable of capturing the variability of the data [36]. In this paper, Bayesian optimization was applied to find the appropriate set of hyperparameters for each model because it requires fewer trials by reasoning about the best set of hyperparameters based on past trials.

One innovation in Bayesian optimization is using an acquisition function, which the algorithm employs to determine the next point to evaluate [46]. Bayesian optimization works by building a probabilistic model of the objective function that is searched efficiently with an acquisition function. The acquisition function can balance sampling at points that have low modeled objective functions and exploring areas that have not yet been modeled well.

4.8. Model evaluation

4.8.1. Evaluation metric

The main objectives of railroad staff are to correctly predict the occurrence of track geometry defects and identify factors that contribute the most to the occurrence of such defects. In machine learning terminology, binary classification results can be divided into four categories (aka. Confusion matrix): the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). TP and TN are used to measure track segments that are predicted correctly. FP and

FN indicate the instances that are classified incorrectly. Five typical measures can be obtained [37]: 1) true positive rate (recall or sensitivity), where $TPR = \frac{TP}{TP+FN}$, 2) true negative rate (specificity), where $TNR = \frac{TN}{TN+FP}$, 3) false positive rate, where $FPR = \frac{FP}{TN+FP}$, 4) false negative rate, where $FNR = \frac{FN}{TP+FN}$, and 5) positive predictive value (precision), where $PPV = \frac{TP}{TP+FP}$.

Accuracy is the most frequently used metric for estimating the performance of machine learning models in classification problems. It is defined by: $Accuracy = \frac{TP+TN}{TP+FN+TN+FP}$. However, standard evaluation criteria focus on the most frequent cases, which goes against the practical preference in engineering. For example, territory concerned with geometry defects only makes up a small proportion of the whole railroad network. Misclassifying observations of these territories has much more severe consequences than territory observations without a geometry defect.

Several measures were proposed for dealing with the imbalanced data issue [37]. Geometric mean (G-Mean) computes the geometric mean of the accuracy of the two classes. It is calculated by $G-mean = \sqrt{specificity \times Recall}$. Graphical-based metrics were proposed to resolve the imbalanced data issue: the receiver operating characteristics (ROC) curve and area under the ROC curve (AUC). However, AUC may produce misleading estimates because of the existence of variations in the evaluation of AUC [38]. F1-score is the harmonic mean of precision and recall, defined by $F = 2\frac{Precision \bullet recall}{precision+recall}$. It is informative about the effectiveness of a classifier on predicting the samples that matter to the user (minority class samples). Therefore, this research applies F1-score as the evaluation metric of the data-driven models.

4.8.2. Model evaluation

The performance of trained models is evaluated and compared based on the test dataset. The test dataset $U = [X, Y]$ is generated by randomly selecting 20 % of the entire dataset (Section 4.3), where X is the input variables of the test samples, and Y is the corresponding actual labels of the samples. Then, trained models are applied to predict the labels of test dataset given input variables using Eq. (13).

$$\widehat{Y}_{k,\rho} = M_k(X, \rho) \tag{13}$$

where $M_k(k = 1, 2, 3)$ represents three types of trained models (i.e., logistic regression, MLP, and GB), respectively, ρ is the threshold of the probability determining if the data sample is associated with a geometry defect, $\widehat{Y}_{k,\rho}$ is the predicted label of the test dataset using ρ as threshold. Based on the predicted labels $\widehat{Y}_{k,\rho}$, the F1-score can be calculated using actual label and predicted label of the test dataset (Eq. (14)).

$$F_{k,\rho} = F\left(Y, \widehat{Y}_{k,\rho}\right) \tag{14}$$

where $F_{k,\rho}$ is the F1-score of the trained model M_k under probability threshold ρ. After F1-score is calculated, the performance of the three trained models can be compared. The model with the highest F1-score is finally selected as the best-fit model and applied in the proposed research.

5. Case study

In this case study, researchers applied the data analytic methodology to one passenger railroad in the United States. 71,096 16-foot-long track segments (around 220 miles) were generated using the described data integration method. Among these segments, 1,708 segments, accounting for 2.4 % of total segments, had geometry defects from 2020 to 2021.

5.1. Feature selection

17 features (as shown in Section 4.2) were extracted from the integrated dataset and applied for feature selection. Fig. 7 shows the number of features adopted and corresponding values of the F1-score using the RFE algorithm. It indicates that in most cases, when more features are removed from the feature list, F1-score steadily increases because data noise caused by redundant data decreases. However, when the number of features adopted in RFE drops to 11, further removing a feature reduces the F1-score because significant influencing factors are ignored by the algorithm. Particularly, when only one or two features are applied,

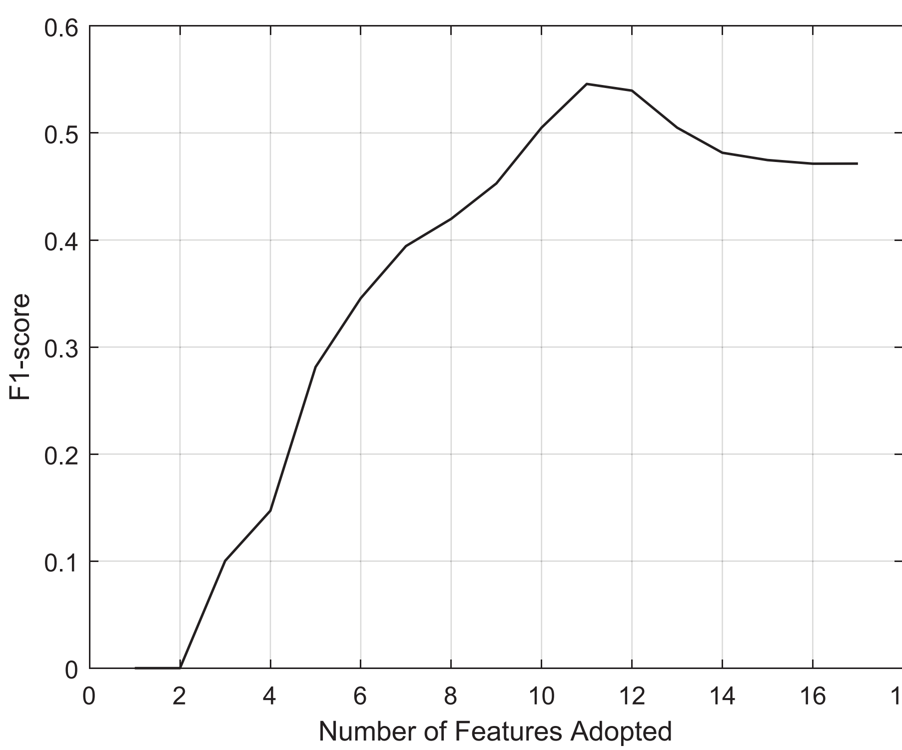

Fig. 7. Number of Features Adopted and F1-score Using RFE Algorithm.

the F1-score drops close to zero. Finally, 11 features selected by RFE include FDL, BFI, BTI, BDM, TDI, curvature, switch, FRA track class, traffic, rail replacement, and surfacing, which are finally applied as the input variables for the machine learning models. In other words, six features (i.e., LRI, BVM, SMI, TCS, RSA, and tie production) that are removed from the feature list contribute the least to the predicted results.

For practical purposes, feature importance has been conducted to identify features that the railroad staff should focus on and analyze more than other features, which is shown in Fig. 8. Particularly, surfacing activity and traffic density have the highest importance rate. Railroads conduct surfacing activity to maintain stable and properly aligned track structure, and to further slow potential track geometry defect growth. The traffic that passes over the rail segments leads to track geometry degradation due to cyclical dynamic loads. Further, rail segments with poor geometry conditions can further increase the dynamic load between rail and wheel when vehicles pass, which accelerates the deterioration of track geometry.

Another practical outcome is that infrastructure-related features, including curvature and the presence of switch, play a crucial role in geometry defect occurrences. Therefore, segments with a certain configuration of infrastructure information may be prone to geometry defect occurrence. An investigation found that isolated track geometry degradations usually occur in spirals at special substructure conditions such as soft subgrade or poor drainage areas [39].

Finally, substructure conditions measured by GPR such as BTI, FDL, and BFI are associated with the occurrence of geometry defects, which is also supported by previous research [15,20,21].

Additionally, the scatter plot matrix is provided to investigate the strength of correlation between 11 input variables, as shown in Fig. 9. It indicates that there is no significant linear correlation between them. Therefore, all these 11 variables are finally used in the data-driven models.

5.2. Model results

Three different data treatment methods, including no sampling, undersampling, and oversampling, were implemented into the training dataset to address the imbalanced classification issue. Three different machine learning methods were separately fit using training and validation datasets with different data treatment methods. The hyperparameters of each model were optimized using Bayesian optimization.

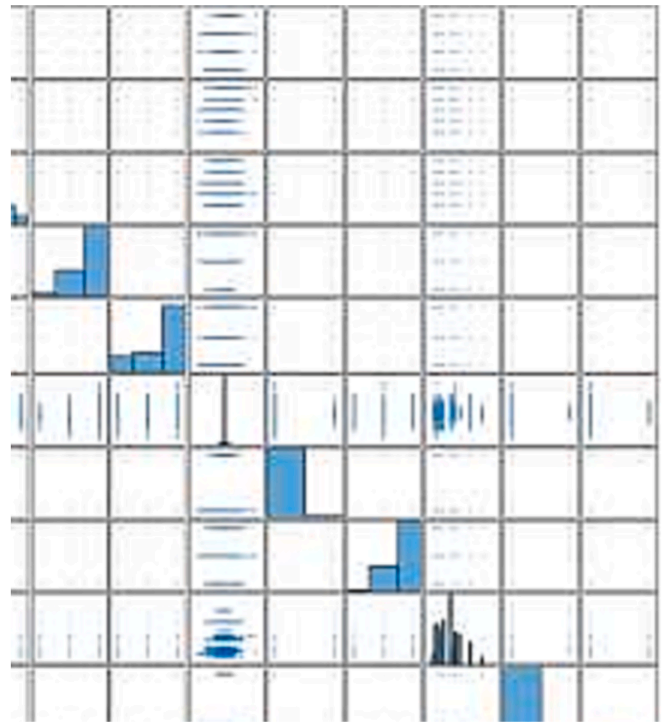

Fig. 9. Scatter Plot Matrix for Input Variables.

To evaluate the performance of the model, various performance criteria were considered. Table 5 includes the precision, recall, G-Mean, and F1-score for all data treatments. It illustrates that classification performance is highest when oversampling is incorporated into the GB algorithm in terms of F1-score. The F1-score of the proposed methods is 0.662, which is better than logistic regression (0.071) and MLP (0.555). With the implementation of oversampling, the F1-score of GB increases from 0.617 to 0.662. It is also observed that oversampling technique can slightly improve the performance of logistic regression and MLP.

Four metrics (i.e., precision, recall, G-Mean, and F1-score) change

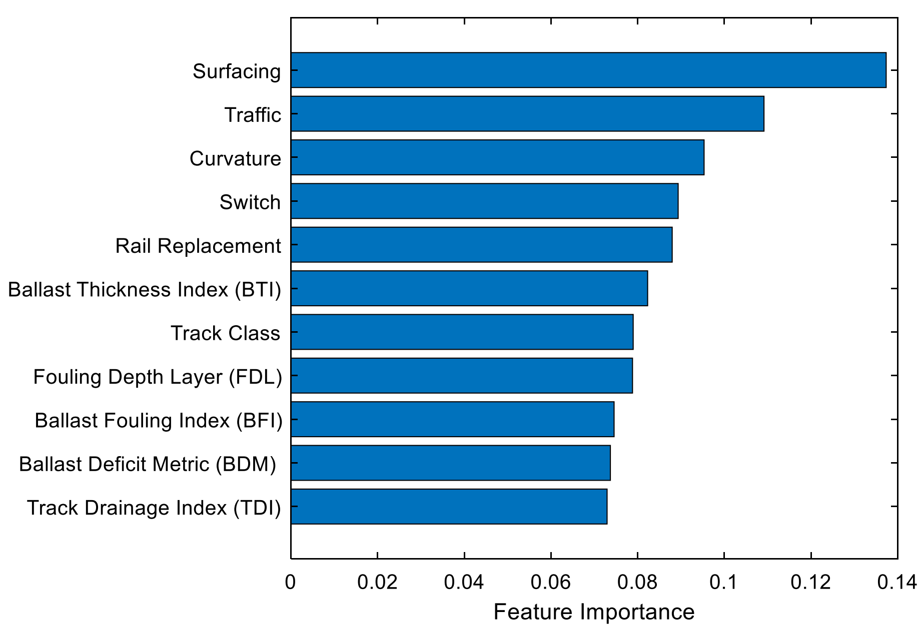

Fig. 8. Importance Rate of Selected Features.

Table 5
Performance of Proposed Methods.

Method	Criteria	No Sampling	Undersampling	Oversampling
Logistic Regression	Precision	0.010	0.038	0.038
	Recall	0.006	0.600	0.533
	G-Mean	0.077	0.617	0.602
	F1-score	0.008	0.071	0.071
MLP	Precision	0.749	0.838	0.790
	Recall	0.432	0.334	0.441
	G-Mean	0.656	0.578	0.663
	F1-score	0.528	0.478	0.555
GB	Precision	0.868	0.828	**0.841**
	Recall	0.479	0.467	**0.545**
	G-Mean	0.691	0.682	**0.738**
	F1-score	0.617	0.597	**0.662**

with different classification threshold [38]. The results in Table 5 are based on the threshold of 0.5. It shows that GB with oversampling achieves relatively good balance between these four metrics. For other methods such as logistic regression with undersampling, even though it has recall rate of 0.600, the precision rate only has 0.038. However, the low recall rate suggests a high number of false negative classifications, which could incur severe consequences. Therefore, the classification threshold can be adjusted to have a higher recall rate to reduce the false negative data samples. Fig. 10 presents the effect of the classification threshold on different metrics of the GB algorithm with oversampling. As the classification threshold decreases, the recall rate gradually increases, indicating that more geometry defects are correctly predicted. When the threshold drops from 0.5 to 0.003, the recall rate increases from 0.55 to 0.88. It is observed that at this point, both the recall and G-mean are equal to 0.88. It also indicates at the intersection point of the recall, precision, and F1-score curves, the classification threshold is 0.21. When the classification threshold decreases from 0.5 to 0.21, three metrics (i.e., precision, recall, and F1-score) are equal to 0.67.

The proposed research can be applied to support the inspection and maintenance prioritization decisions. In the studied network, track geometry is inspected by geometry cars twice a year. This research can be applied to point out the track locations prone to geometry defects, and assist railroad staff in monitoring these high-risk track segments. For example, a portable geometry measurement device can be used to monitor track geometry at a local level. Track segments having higher predicted probability of track geometry defect occurrence should be additionally inspected during the interval of geometry car inspection. Besides, this research can be used for capital planning, such as maintenance prioritization decisions. Track segments prone to geometry defects might be given priority for maintenance, such as tie production and tamping.

To compare the characteristics of the segments with high and low predicted risk, Table 6a and Table 6b show the top 10 segments with the highest and lowest predicted probability of geometry defect occurrence, respectively. It is found that segments associated with higher traffic density and/or sharper curves are prone to track geometry defects, such as #1 and #3 track segments in Table 6a. Track segments with higher traffic volume are prone to vertical geometry defects (e.g., profile) due to larger dynamic loads on tracks. On the other hand, curved tracks are more likely to experience lateral geometry defects (e.g., alignment) because the rail on the curved tracks is consistently subjected to more lateral loads leading to more alignment variations.

However, it is observed that some high-risk segments that are not associated with high traffic density or shape curves usually have poorer substructure conditions measured by GPR. It implies that track geometry defects on those track segments are mainly caused by poor substructure conditions. Both drainage index (TDI) and ballast fouling property (BFI and FDL) reflect the ballast drainage. Track with poor BFI, FDL, and/or TDI indicates that the trackbed drainage is compromised. Mud pumping may occur due to poor drainage, which results in vertical geometry defects including profile, crosslevel, and twist. On the other hand, the ballast thickness and volume are assessed by BTI and BDM, respectively. Track with poor BTI and BDM shows improper thickness and deficit in ballast volume. In this case, there would be a high rate of track geometry degradation in both vertical and lateral directions. For instance, #6 track segment in Table 6a has 6th highest predicted probability of

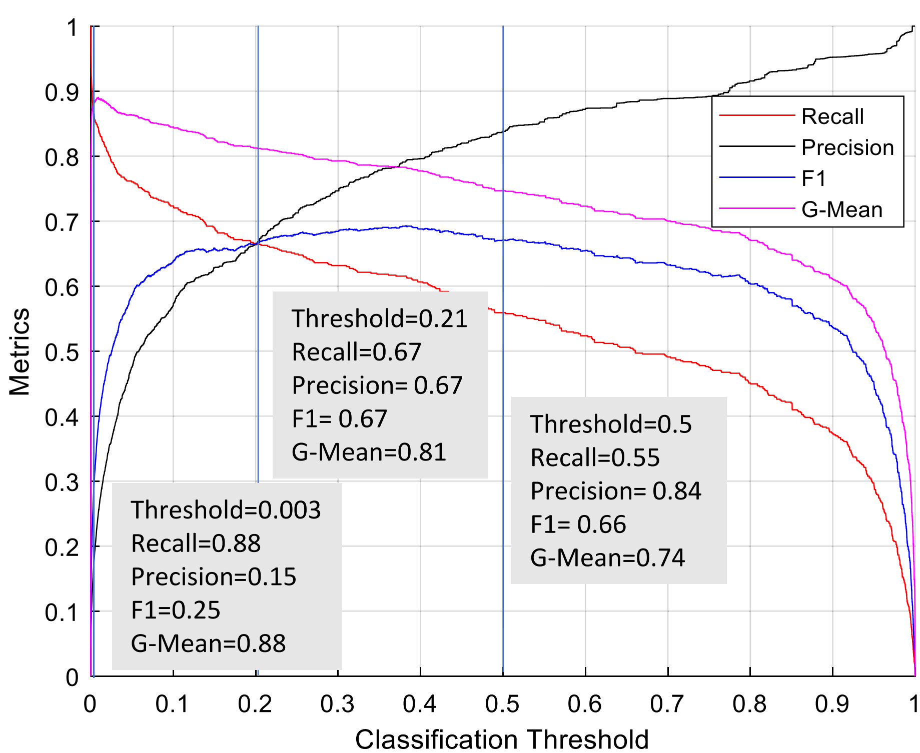

Fig. 10. Influence of Threshold on Different Metrics of Gradient Boosting Model with Oversampling.

Table 6a
Segments with Highest Predicted Probability of Geometry Defect Occurrences.

No.	Surfacing (1 = Yes; 0 = No)	Traffic (MGT)	Curve Degree	Switch (1 = Yes; 0 = No)	Rail Replacement (1 = Yes; 0 = No)	BTI	Track Class	FDL	BFI	BDM	TDI	Actual Geometry Defect Occurrence	Predicted Probability of Geometry Defects
1	0	18.2	2.98	0	0	4	4	4	3	2	1	1	0.99932
2	0	18.2	3.03	0	0	3	4	3	3	3	1	1	0.99669
3	0	18.2	2.99	0	0	3	4	4	4	2	1	1	0.99645
4	1	27.0	2	0	0	3	4	4	5	3	3	1	0.99458
5	1	27.0	1.19	0	0	3	4	4	5	3	1	1	0.98748
6	0	22.7	1.04	0	0	3	4	2	3	2	1	1	0.98310
7	0	22.7	0.71	0	0	4	3	2	4	2	1	1	0.98129
8	0	22.7	0.72	0	0	4	3	2	4	2	1	1	0.98129
9	1	27.0	1.4	0	0	4	4	4	5	3	3	1	0.98053
10	0	30.5	0.01	0	0	2	3	3	5	3	3	1	0.97946

Table 6b
Segments with Lowest Predicted Probability of Geometry Defect Occurrences.

No.	Surfacing (1 = Yes; 0 = No)	Traffic (MGT)	Curve Degree	Switch (1 = Yes; 0 = No)	Rail Replacement (1 = Yes; 0 = No)	BTI	Track Class	FDL	BFI	BDM	TDI	Actual Geometry Defect Occurrence	Predicted Probability of Geometry Defects
1	0	19.2	0.86	0	0	1	4	4	4	3	1	0	2.73E-07
2	1	19.2	0.25	0	0	3	4	4	5	2	3	0	2.43E-07
3	0	19.2	0.86	0	1	4	4	4	5	1	3	0	2.39E-07
4	1	19.2	0	0	0	2	4	3	3	2	3	0	2.38E-07
5	0	19.2	0.88	0	0	1	4	3	3	3	3	0	1.78E-07
6	1	16.9	0.02	0	0	3	3	3	3	2	3	0	1.73E-07
7	0	19.2	0.02	0	0	3	4	4	5	1	3	0	1.67E-07
8	0	18.2	0.04	0	0	1	4	4	5	1	3	0	1.51E-07
9	0	19.2	0.92	0	0	3	4	4	5	3	1	0	1.38E-07
10	0	19.2	0.9	0	0	1	4	3	3	3	3	0	8.12E-08

Note:
BTI: 3 indicates the best condition. Values greater or smaller than 3 represent poorer conditions.
FDL, BFI, and BDM: The greater values indicate better conditions.
TDI: The greater values indicate poorer conditions.

geometry defects. Compared to other track segments with the highest predicted probability, while this track segment is not within a sharp curve and does not experience high traffic density, it has poorer substructure conditions including poor fouling depth (FDL = 2), moderately fouled ballast (BFI = 3), and small deficit in ballast volume (BDM = 2). In other words, the geometry defect on #6 track segment is mainly caused by poor substructure conditions.

Notably, segments that have undergone maintenance activities such as surfacing in the prior year have higher predicted probability in some cases and lower predicted probability in others. Segments associated with larger traffic density and sharper curves tend to have significantly higher predicted probability. In other words, those segments are still prone to geometry defect occurrence even though they have been maintained in the previous year. Meanwhile, segments with smaller traffic density and curve degrees are less likely to experience geometry defects once they have experienced maintenance in the previous year.

6. Conclusions and future work

6.1. Conclusion

Accurate evaluation of geometry defect occurrences under the influence of different factors can assist the maintenance-of-way task and contribute to better operation quality. This research develops a data-driven framework to investigate the impact of various influencing factors on the occurrence of track geometry defects, with a particular application to concrete-tie track on one passenger railroad in the United States. The framework consists of a pipeline of methodologies including data cleaning, data integration, feature extraction, data resampling, feature selection, and hyperparameter optimization. Data cleaning was used to identify concrete-tie track segments and to handle data issues. In data integration, heterogeneous data are combined into one integrated file. Feature extraction was implemented to generate the input variables for the machine learning models. Recursive feature elimination (RFE) was applied to reduce data dimensionality by recursively considering smaller and smaller sets of features. Three data treatment methods, including no resampling, undersampling, and oversampling, were incorporated into the model. Hyperparameters of proposed models were separately optimized using Bayesian optimization to avoid overfitting and underfitting. The performance of the proposed methods was finally evaluated based on a dataset generated by random data partitioning, which can avoid an over-optimistic estimation of the model. Using data collected from one passenger railroad, gradient boosting with oversampling shows the highest performance in predicting the occurrence of geometry defects using data including substructure condition data, rail seat abrasion data, infrastructure data, traffic data, track class information, and maintenance data.

Feature importance identifies that surfacing, traffic, curvature, switch, and rail replacement are the top five factors influencing the predicted probability of track geometry defect occurrence. It is found that track segments with larger traffic density and sharper curves are still prone to geometry defect occurrence even though they have been maintained in the previous year. Meanwhile, segments with smaller traffic density and curve degrees are less likely to experience geometry defects once they have experienced maintenance in the previous year. The results illustrate that the model can be used to prioritize inspection and maintenance activities on locations prone to track geometry defects, thus further improve infrastructure safety given budgetary constraints.

6.2. Future work

This section discusses the limitations and future research directions of this study. First, some critical factors are not included in this data-driven framework due to limited data availability, which results in a not very high recall rate. For example, track segments with tamping interventions are less likely to experience track geometry defects in a short period. In future research, more influencing factors can be collected and incorporated into the model to further improve its performance.

Furthermore, current research does not support time series analysis because the timespan of data is limited. In other words, this study focuses on correlation within relatively short time duration and cannot be used for predicting future track conditions due to data limitations. In the future, as more years of data are collected, future research will focus on forecasting the occurrence of geometry defects in a specific time interval, such as one month or one year in advance, to understand the temporal effects of various factors on track geometry defects. The chronological relationship between the influencing factors and the occurrence of track geometry defects should be identified during data processing. Additionally, time-series data partitioning should be used for splitting the dataset into training, validation, and test datasets, to avoid data leakage during the model training process.

CRediT authorship contribution statement

Xin Wang: Investigation, Methodology, Writing – original draft, Visualization, Data curation. **Xiang Liu:** Conceptualization, Writing – review & editing, Project administration, Supervision. **Todd L. Euston:** Writing – review & editing.

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.

References

[1] C. Higgins, X. Liu, Modeling of track geometry degradation and decisions on safety and maintenance: A literature review and possible future research directions, Proc. Inst. Mech. Eng., Part F 232 (2018) 1385–1397.
[2] Y. Wang, P. Wang, X. Wang, X. Liu, Position synchronization for track geometry inspection data via big-data fusion and incremental learning, Transp. Res. Part C: Emerg. Technol. 93 (2018) 544–565.
[3] 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.
[4] A.M. Zarembski, D. Einbinder, N. Attoh-Okine, Using multiple adaptive regression to address the impact of track geometry on development of rail defects, Constr. Build. Mater. 127 (2016) 546–555.
[5] X. Liu, M.R. Saat, C.P. Barkan, Analysis of causes of major train derailment and their effect on accident rates, Transp. Res. Rec. 2289 (2012) 154–163.
[6] S. Fontul, E. Fortunato, F. De Chiara, R. Burrinha, M. Baldeiras, Railways track characterization using ground penetrating radar, Procedia Eng. 143 (2016) 1193–1200.
[7] M. Solla, V. Pérez-Gracia, S. Fontul, A review of GPR application on transport infrastructures: Troubleshooting and best practices, Remote Sens. (Basel) 13 (2021) 672.
[8] A.M. Zarembski, G.T. Grissom, T.L. Euston, On the use of ballast inspection technology for the management of track substructure, Transp. Infrastruct. Geotechnol. 1 (2014) 83–109.
[9] J.C. Zeman, J.R. Edwards, C.P. Barkan, D.A. Lange, Failure mode and effect analysis of concrete ties in North America, in: Proc of the 9th International Heavy Haul Conference, 2009, pp. 270–278.
[10] A. Hamid, A. Gross, Track-quality indices and track degradation models for maintenance-of-way planning, Transp. Res. Board 802 (1981) 2–8.
[11] J.S. Lee, S.H. Hwang, I.Y. Choi, I.K. Kim, Prediction of track deterioration using maintenance data and machine learning schemes, J. Transp. Eng., Part A: Syst. 144 (2018) 04018045.
[12] Y. Shafahi, P. Masoudi, R. Hakhamaneshi, Track degradation prediction models, using Markov Chain, artificial neural and neuro-fuzzy network, in: 8th World Congress on Railway Research, 2008, pp. 1–9.
[13] 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 (2015) 392–404.
[14] 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 (2022) 24–43.
[15] C. Hu, X. Liu, Modeling track geometry degradation using support vector machine technique. ASME/IEEE Joint Rail Conference, 2016. V001T01A011.
[16] 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.
[17] D. Li, S. Wilk, Recent studies on railway-track substructure at TTCI, Transp. Saf. Environ. 3 (2021) 36–49.
[18] J. Sadeghi, H. Askarinejad, Application of neural networks in evaluation of railway track quality condition, J. Mech. Sci. Technol. 26 (2012) 113–122.
[19] A.M. Zarembski, J.W. Palese, T.L. Euston, Correlating ballast volume deficit with the development of track geometry exceptions utilizing data science algorithm, Transp. Infrastruct. Geotechnol. 4 (2017) 37–51.
[20] 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 (2019) 1–20.
[21] A.M. Zarembski, D. Yurlov, J. Palese, N. Attoh-Okine, Relationship between track geometry defects and measured track subsurface condition, No. DOT/FRA/ORD-20/07, Department of Transportation. Federal Railroad Administration, United States, 2020, p. 147.
[22] A. Alsahli, A.M. Zarembski, N. Attoh-Okine, Predicting track geometry defect probability based on tie condition using pattern recognition technique. ASME International Mechanical Engineering Congress and Exposition, 2020. V014T14A031.
[23] S. Goodarzi, H.F. Kashani, J. Oke, C.L. Ho, Data-driven methods to predict track degradation: A case study, Constr. Build. Mater. 344 (2022), 128166.
[24] A. Alsahli, A.M. Zarembski, J.W. Palese, T.L. Euston, Investigation of the correlation between track geometry defect occurrence and wood tie condition, Transp. Infrastruct. Geotechnol. 6 (2019) 226–244.
[25] A.M. Zarembski, N. Attoh-Okine, D. Einbinder. On the relationship between track geometry defects and development of internal rail defects. 2015; 1:7. Retrieved July 2022, from https://cpb-us-w2.wpmucdn.com/sites.udel.edu/dist/0/12693/files/2022/06/186-WCRR2016-Track-Geometry-and-Rail-Defects.pdf.
[26] A.M. Zarembski, G.T. Grissom, T.L. Euston, J.J. Cronin, Relationship between missing ballast and development of track geometry defects, Transp. Infrastruct. Geotechnol. 2 (2015) 167–176.
[27] J. Belcher, G. Grissom, J. Baciak, Automated crosstie inspection using internal imaging techniques. Proceedings of AREMA Annual Conference and Exposition, Chicago, Illinois, 2014.
[28] D.F. Cannon, K.-O. Edel, S.L. Grassie, K. Sawley, Rail defects: an overview, Fatigue Fract. Eng. Mater. Struct. 26 (2003) 865–886.
[29] N. Japkowicz, The class imbalance problem: Significance and strategies, in: Proceeding of the International Conference on Artificial Intelligence, 2000, pp. 111–117.
[30] M. Kubat, S. Matwin, Addressing the curse of imbalanced training sets: one-sided selection, Int. Conf. Mach. Learning 97 (1997) 179.
[31] V.S. Spelmen, R. Porkodi, A review on handling imbalanced data, in: International Conference on Current Trends towards Converging Technologies (ICCTCT), 2018, pp. 1–11.
[32] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, in: IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 1322–1328.
[33] F. Rosenblatt, Principles of neurodynamics. Perceptrons and the theory of brain mechanisms, Cornell Aeronautical Lab Inc, Buffalo NY, 1961.
[34] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[35] H. Jabbar, R.Z. Khan. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Computer Science, Communication and Instrumentation Devices 2015; 70.
[36] E. Briscoe, J. Feldman, Conceptual complexity and the bias/variance tradeoff, Cognition 118 (2011) 2–16.
[37] P. Branco, L. Torgo, R.P. Ribeiro, A survey of predictive modeling on imbalanced domains, ACM Comput. Surveys (CSUR) 49 (2016) 1–50.
[38] C.E. Metz, Basic principles of ROC analysis, Semin. Nucl. Med. 8 (1978) 283–298.
[39] A. Hamid, K. Rasmussen, M. Baluja, T. Yang. Analytics descriptions of track geometry variations. No.DOT/FRA/ORD-83/03.1. WASHINGTON, DC. 1983.
[40] M. Sol-Sánchez, G. D'Angelo, Review of the design and maintenance technologies used to decelerate the deterioration of ballasted railway tracks, Constr. Build. Mater. 157 (2017) 402–415.
[41] M. Greve, M.S. Dersch, J.R. Edwards, C.P. Barkan, J. Mediavilla, B.M. Wilson, Analysis of the relationship between rail seat load distribution and rail seat

deterioration in concrete crossties. ASME/IEEE Joint Rail Conference. American Society of Mechanical Engineers, 2014.
[42] A.A. Shurpali, J.R. Edwards, R.G. Kernes, D.A. Lange, C.P. Barkan, Improving the abrasion resistance of concrete to mitigate concrete crosstie rail seat deterioration (RSD), Mater. Perform. Charact. 6 (1) (2017) 521–534.
[43] Department of Transportation and the Federal Railroad Administration (FRA). 1998. Track safety standards, Final Rule, 49 CFR Part 213.
[44] J. Cai, J. Luo, S. Wang, S. Yang, Feature selection in machine learning: A new perspective, Neurocomputing 300 (2018) 70–79.
[45] X.W. Chen, J.C. Jeong, December). Enhanced recursive feature elimination, Int. Conf. Mach. Learning Appl. 2007 (2007) 429–435.
[46] V. Nguyen, Bayesian optimization for accelerating hyper-parameter tuning, in: International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 2019, pp. 302–305.
[47] M. Sysyn, U. Gerber, O. Nabochenko, Y. Li, V. Kovalchuk, Indicators for common crossing structural health monitoring with track-side inertial measurements, Acta Polytech. 59 (2) (2019) 170–181.
[48] M. Sysyn, U. Gerber, O. Nabochenko, D. Gruen, F. Kluge, Prediction of rail contact fatigue on crossings using image processing and machine learning methods, Urban Rail Transit. 5 (2) (2019) 123–132.
[49] M. Sysyn, L. Izvolt, O. Nabochenko, V. Kovalchuk, J. Sestakova, A. Pentsak, Multifractal analysis of the common crossing track-side measurements, Civ. Environ. Eng. (2019) 101–114.
[50] Y. Guo, W. Zhai, Long-term prediction of track geometry degradation in high-speed vehicle–ballastless track system due to differential subgrade settlement, Soil Dyn. Earthq. Eng. 113 (2018) 1.
[51] Y. Guo, V. Markine, G. Jing, Review of ballast track tamping: Mechanism, challenges and solutions, Constr. Build. Mater. 20 (300) (2021), 123940.
[52] A. Eriksen, J. Gascoyne, R. Fraser, Ground penetrating radar as part of a holistic strategy for inspecting trackbed, Aust. Geomech. Soc. 46 (3) (2011) 1.
[53] X. Liu, C.P. Barkan, M.R. Saat, Analysis of derailments by accident cause: Evaluating railroad track upgrades to reduce transportation risk, Transp. Res. Rec. 2261 (1) (2011) 178–185.
[54] X. Wang, X. Liu, Z. Bian, A machine learning based methodology for broken rail prediction on freight railroads: A case study in the United States, Constr. Build. Mater. 346 (2022), 128353.