

Data driven machine learning prognostics of buckling failure modes in ballasted railway track

Watcharapong Wongkaew¹ · Wachira Muanyoksakul^{2,3} · Chayut Ngamkhanong^{2,3} · Jessada Sresakoolchai⁴ · Sakdirat Kaewunruen⁵

Received: 20 February 2024 / Accepted: 10 April 2024

Published online: 14 April 2024

© The Author(s) 2024 OPEN

Abstract

This study explores the development and application of a machine learning (ML) approach to predict buckling failure modes in ballasted railway tracks. With the growing demand for safer and more reliable railway systems, the ability to foresee and mitigate track failures is of paramount importance. Our study focuses on harnessing advanced ML algorithms to analyse and interpret complex data sets, aiming to identify potential buckling failures before they occur. The methodology employed involves collecting extensive data from previous advanced numerical studies. Faced with the inadequacy of field data collection on track buckling and the limited availability of data related to track conditions, our study has relied on simulation data for insight and analysis. This data is then processed and analysed using sophisticated ML models, trained to recognise patterns and anomalies indicative of potential buckling failures. A novel aspect of our approach is the integration of environmental factors, acknowledging their significant influence on the likelihood of both snap-through and progressive buckling in railway tracks. We compare the effectiveness of various ML algorithms in accurately predicting these failure modes, evaluating their performance in simulated and real-world scenarios. The findings demonstrate the models' proficiency in identifying early signs of both snap-through and progressive buckling, leading to timely interventions. This capability not only improves railway safety but also aids in efficient maintenance scheduling and asset management. Additionally, a case study in Thailand's railway system demonstrates the model's effectiveness in predicting buckling failures under tropical environmental conditions. This paper contributes a novel perspective to the field of railway infrastructure maintenance. By providing a reliable method for predicting specific buckling failure modes, it paves the way for enhanced operational safety and efficiency in railway networks, particularly in the face of dynamic environmental conditions.

Article Highlights

- A machine learning-based approach has been developed to predict buckling failure modes in ballasted railway tracks, aiming to enhance safety and reliability in railway systems.
- Different machine learning algorithms have been comprehensively evaluated.
- A case study in Thailand's tropical climate has been included to demonstrate real-world applicability.

✉ Chayut Ngamkhanong, chayut.ng@chula.ac.th; Watcharapong Wongkaew, watcharapong.w@chula.ac.th; Wachira Muanyoksakul, 6678002021@student.chula.ac.th; Jessada Sresakoolchai, Jessada.sr@psu.ac.th; Sakdirat Kaewunruen, s.kaewunruen@bham.ac.uk |

¹Chulalongkorn University Transportation Institute, Chulalongkorn University, Bangkok, Thailand. ²Department of Civil Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand. ³Advanced Railway Infrastructure, Innovation and Systems Engineering (ARIISE) Research Unit, Chulalongkorn University, Bangkok, Thailand. ⁴Department of Civil and Environmental Engineering, Faculty of Engineering, Prince of Songkhla University, Songkhla, Thailand. ⁵Department of Civil Engineering, School of Engineering, University of Birmingham, Birmingham B15 2TT, UK.



Keywords Machine learning · Ballasted railway track · Buckling failure · Railway track stability · Railway infrastructure · Early warning detection

1 Introduction

Railway track buckling is a critical safety concern within the railway system [1–3], and with the increasing frequency of heatwaves attributed to climate change, the resilience of railway infrastructure to extreme temperatures has become a pressing issue. Continuous welded rail (CWR), while offering benefits such as smoother rides and lower maintenance compared to traditional jointed rail systems, is particularly susceptible to heat-induced expansion. This can lead to a dangerous build-up of compressive forces, potentially causing the tracks to lose lateral stability and buckle when rail temperatures escalate beyond a certain threshold [4–7]. Instances of track buckling can have severe consequences, including derailments, with the potential for substantial material damages and tragic loss of life [8–10]. It is noteworthy that buckling typically affects conventional ballasted railway tracks, often resulting from inadequate track conditions and misalignments. Extensive research into risks associated with extreme temperatures and track buckling, examining key influencing factors and establishing safe operational temperature ranges, has been undertaken to mitigate these risks [11–15]. It is crucial to emphasise that in hot weather conditions, trains should operate at reduced speeds compared to their usual pace. This adjustment helps mitigate the additional force generated by wheel-rail contact, ensuring safer operation and reducing the risk of track damage or derailment [16–19].

Studies have established that the buckling strength and occurrences in railway tracks are largely influenced by the lateral resistance provided by the ballast and sleepers [20]. To quantify this resistance, the Single Sleeper Push Test (STPT) has proven effective, as it measures the resistance against sleeper movement, a critical indicator of track stability [21–25]. While the resistance provided by the ballast-sleeper interface is crucial, the role of fasteners in offering torsional resistance between the sleepers and rails also factors into the equation, albeit to a lesser extent. The varied and complex influences on rail track performance under extreme temperature conditions cannot be fully captured by traditional beam theory, which is restricted by a limited parameter set for calculating buckling temperatures [26, 27]. Finite Element (FE) modelling has been utilised to include a wider range of parameters and account for more complex structures. However, the intensive computational demand and significant memory usage of FE analysis are notable drawbacks. To enhance the computational efficiency and predictive capabilities derived from FE models, machine learning methods have been proposed as a complementary tool. These methods can efficiently process the intricate array of parameters that impact track stability, providing a more detailed and accurate prediction of buckling phenomena—a task that is often impractical with purely analytical approaches.

Machine learning offers an innovative, data-driven approach for the development of algorithms capable of learning from existing datasets to make predictions about new, unseen data. Unlike traditional statistical models, which often come with stringent assumptions about data distribution, machine learning methods operate without the need for predefined distribution assumptions, offering a more flexible solution for real-world data that may not conform to such expectations. The validity of machine learning outcomes can be rigorously assessed using cross-validation techniques [28].

The utilisation of machine learning has become increasingly prevalent in the civil engineering and railway sectors, particularly for predicting the characteristics of track materials from their constituent ratios, including those of recycled concrete [29] and rubberised concrete [30]. The scope of machine learning extends to diverse railway track applications including machine vision recognition for track inspection and maintenance, defect detection to enhancing safety measures [31–34]. It also includes signal processing tasks such as assessing track dynamics [35], predicting the weight of trains [36], and identifying risks for enhancing railway safety measures [37, 38]. A notable advancement in this field was the pioneering use of artificial neural networks (ANNs) in the prediction of track buckling and safe operating temperatures [39]. ANNs, which are a class of machine learning models, emulate the functioning of biological neural networks and possess the capacity to learn from empirical data, thereby enabling them to solve complex problems with a level of efficiency that mirrors human cognitive processes.

In this paper, we introduce the novel use of machine learning methods to predict buckling failure modes in ballasted railway tracks, employing a suite of classification algorithms. We select a range of machine learning techniques, from basic to advanced, to develop a robust framework capable of accurately predicting buckling failure modes, incorporating a comprehensive array of parameters. This research compares the performance of various machine learning models in their

ability to predict buckling modes, informed by rail temperature and prevailing track conditions. Our datasets are derived from extensive numerical simulations based on Finite Element Analysis [18, 40, 41]. This is because data on track buckling correlated with specific track conditions is scarce, making numerical simulation data a more reliable choice for analysis. The models that exhibit optimal performance demonstrate an advanced capability to navigate the complex relationships between contributing factors and predicted outcomes, which are crucial for the design of track components that can withstand extreme temperatures. These proposed machine learning models represent a significant advancement in the predictive accuracy of buckling phenomena, offering insights that could lead to enhanced maintenance and reinforcement of railway tracks to better withstand the rigours of elevated summer temperatures. An ability to predict the failure modes of buckling ballasted track will enhance the safety of the system, operational reliability and availability, save the maintenance cost through early stage detection, improve the infrastructure longevity, and increase asset management efficiency. The novelties of the proposed approach include the development of advanced asset monitoring techniques, machine learning integration, and the incorporation of climate and weather considerations.

This paper begins with a theoretical exploration of railway track buckling modes, followed by an explanation of the data collection methodology. This includes the creation of a Finite Element (FE) model that integrates new datasets with those from existing literature. The methodology, which includes an oversampling technique and various machine learning approaches, is detailed in Sect. 4. The results section evaluates the performance of the proposed model and introduces a novel diagram illustrating buckling mode transitions. Furthermore, the paper features a case study from Thailand, showcasing the application of the developed machine learning model in a real-world scenario.

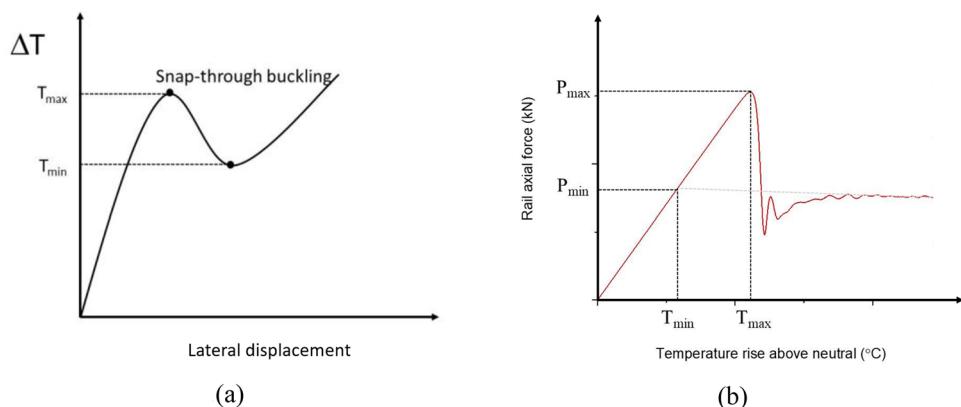
2 Buckling modes

This section describes two buckling modes experienced by rails under temperature variations and axial force increase: snap-through buckling and progressive buckling.

In the pre-buckling phase, the rails are subjected to temperatures above their neutral state, along with a gradual rise in axial force. Snap-through buckling, characterised by an abrupt and explosive track deformation, occurs once the rails hit the maximum critical temperature (T_{\max}) without the need for additional energy (Fig. 1). The track immediately becomes unstable and the axial force sharply decreases. T_{\min} is the temperature threshold below which the track remains stable, essentially acting as a safety margin against buckling. During the post-buckling phase, the track initially experiences large lateral shifts before stabilising and displaying a continuous decrease in axial force. Plotting this diminishing force trend backwards intersects with the pre-buckling force line, indicating the minimal axial force required for buckling. The corresponding temperature at this intersection, above the neutral or safe temperature, is considered the minimum temperature for buckling.

In scenarios where T_{\min} merges with T_{\max} , rendering the peak indiscernible, progressive buckling ensues, as depicted in Fig. 2. This mode is characterised by a steady increase in the track's lateral displacement post-buckling, with the critical temperature designated as T_p . Unlike snap-through buckling, the pinpointing of a maximum temperature at which buckling initiates is not feasible; thus, axial force serves as a reliable proxy for ascertaining the safe temperature threshold. During progressive buckling, the axial force diminishes in a gradual manner rather than abruptly. Notably, the trend line of axial force in the post-buckling phase does not exhibit a definitive intersecting point with the pre-buckling axial

Fig. 1 Snap-through buckling



force. However, it is posited to converge at the point of maximum axial force, leading to the inference that for progressive buckling, T_{\max} is synonymous with T_{\min} . Consequently, the buckling temperature for a progressive buckling failure is determinable at the point of maximum axial force. It is imperative to note that this determination is not directly obtainable from the temperature-lateral displacement curve utilised in the preceding methodology.

3 Data collection

In order to generate datasets, this paper employs Finite Element Analysis (FEA) extensively to generate additional datasets, which are then integrated with existing data from previous studies, primarily consisting of numerical results such as buckling and safe temperature thresholds [15, 18, 39–41]. These previously published datasets have been restructured into new categories of input and output data for further analysis. In this paper, temperature has been introduced as an input variable, while the output data focus on categorising buckling behaviours into three distinct modes: snap-through, progressive buckling, and non-buckling. This section introduces a model of a railway track, along with the input parameters employed for analysing different modes of buckling.

3.1 Railway track model

To investigate the effect of temperature increase on track buckling, an analysis was performed using a 60-m-long model of ballasted railway tracks with a standard gauge, as shown in Fig. 3a. This model comprises 100 spans, each spaced 0.6 m apart. The elements measured for buckling including axial force (at beam 257) and displacement (at node 1360) are positioned at the centre of the model. It is important to note that the chosen length of 60 m is sufficient to accurately represent the phenomena of track buckling and covers the range of buckling lengths typically observed in real-world.

Within this study, the ballasted tracks and their standard gauge components are simulated using the LS-DYNA software. UIC60 rails and sleepers are represented as beam elements to account for shear and bending deformations. The elements are defined using the SECTION_BEAM and MAT_ELASTIC definitions, with the rails further enhanced by the MAT_ADD_THERMAL_EXPANSION property to simulate thermal effects. At the rail seat, a combination of translational springs in three axes and a rotational spring are implemented to mimic the mechanical properties of the rail pad and fastener. It is recommended to use a tensionless support spring for ballast, attaching it to both ends of the sleeper. This approach is preferred over traditional springs as it permits the beam to shift over the support, disregarding tensile support considerations. The spring's stiffness, which determines the sleeper-ballast's lateral resistance, is typically represented by an elastoplastic curve (see Fig. 3b). In this research, the node at the mid-span of the track is crucial for measuring lateral displacement, and the connected beam is analysed for axial force, as shown in Fig. 3a. Furthermore, in nonlinear buckling analysis, introducing initial imperfections or misalignments is essential, since theoretical structures cannot buckle without these imperfections.

For the full railway tracks, an initial misalignment amplitude is introduced at the nodes located at the centre span of both rails as presented in Fig. 3c. It should be noted that, for simplicity, a sine wave is assumed to represent track imperfections, instead of using actual track alignment data. This assumption allows for a more streamlined

Fig. 2 Progressive buckling

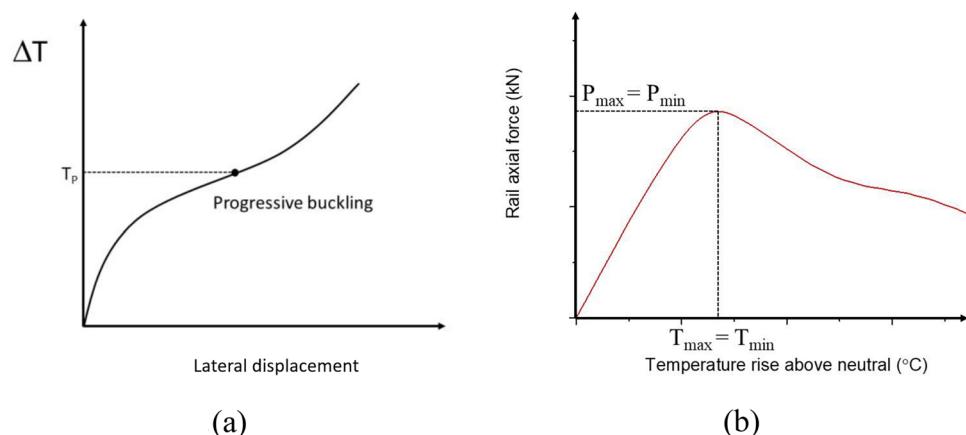
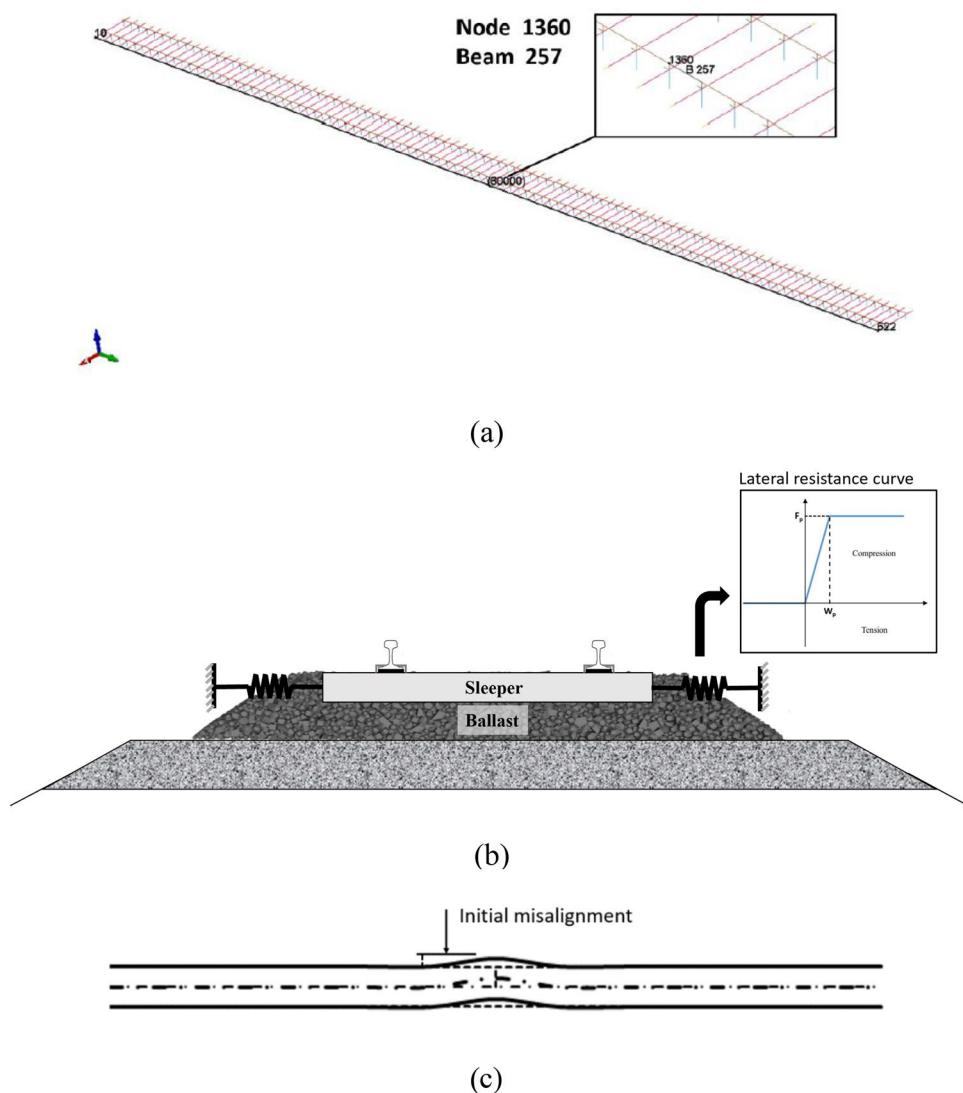


Fig. 3 Railway track model **a**
FE model **b** Idealised lateral
spring model **c** Lateral mis-
alignment



analysis. However, in real-world applications, a preliminary analysis of track alignment is essential to filter out short-wavelength variations. This step is necessary to appropriately fit the sine curve to the actual track data, ensuring that the model remains both simplified and relevant to real track conditions.

The analysis of buckling and safe temperatures is conducted by evaluating the rail's critical axial force through finite element analysis. Snap-through buckling data are derived from the axial force's peak, observed when the maximum (T_{\max}) and minimum (T_{\min}) temperatures coincide. In contrast, progressive buckling is identified in the absence of a distinct peak within the temperature-lateral displacement curve. The resulting datasets are systematically organised in a table featuring three columns: T_{\max} , T_{\min} , and buckling modes, to facilitate clear interpretation and comparison. It is important to note that the models have been previously validated with analytical solutions and finite element (FE) models previously established by other researchers [42, 43]. The validation results can be found in the literature [15, 18, 41].

3.2 Input parameters

In this study, several parameters are meticulously measured and analysed to understand their variations and implications in the system, as presented in Table 1. The initial lateral stiffness, indicative of the resistance force to side-to-side deformation at peak divided by the displacement limit of sleeper (F_p/W_p), exhibited a range from 120 N/mm to 2632 N/mm,

Furthermore, the sleeper's lateral displacement limit (W_p) was scrutinised, revealing a limitation that spanned between 0.5 mm and 2 mm. This is a value where the stiffness becomes zero after this limit. This represents an elastoplastic curve of sleeper-ballast lateral resistance, as shown in Fig. 3b. It is important to note that a lateral resistance force at a displacement of 2 mm is typically used as a benchmark. The range of resistance force used in this study (240 N–5264 N) varies according to track condition [44]. For a freely lying track, the standard values are 800 N for timber sleepers and 1200 N for concrete sleepers. For a loosely ballasted track, especially after cleaning, the resistance force is around 1500 N for timber sleepers and 2500 N for concrete sleepers. Additionally, for tracks that have undergone tamping, the resistance force is around 2850 N for timber sleepers and 5100 N for concrete sleepers. The input data distributions are visualised in Fig. 4.

The torsional resistance, a measure of the fastening system's resistance to twisting, varied from 75 Nm/rad to a maximum of 225 Nm/rad. The study also evaluated the unconstrained length, which manifested values between 6 and 30 m, averaging 24 m with a dispersion measure, or standard deviation, of 8.5 m. Initial misalignment, another critical parameter, showed variations from 8 to 32 mm. Lastly, the investigation shed light on the increased rail temperature, which oscillated between 10 °C and 200 °C. The understanding and quantification of these parameters are quintessential for a holistic comprehension of the buckling behaviour and further research implications.

The dataset encompasses a total of 8000 cases, partitioned into distinct sets for training and testing. The training set is further subdivided into a validation set through the application of the K-Fold Cross Validation technique. A detailed breakdown of these subsets, including the specific allocations for training, validation, and testing, is delineated in Table 2.

4 Methodology

4.1 Oversampling technique

Table 2 presents an imbalanced dataset with a disproportionate number of examples in the snap-through buckling and non-buckling classes. Such an imbalance may adversely affect the predictive performance and capacity of the model. To address this, an oversampling technique is applied to equalise the class frequencies. This method involves generating new samples that are almost identical to the original data, proportionally increasing the representation of the minority class to match that of the majority class, thus balancing the dataset.

The Synthetic Minority Over-Sampling Technique (SMOTE) is employed to rectify class imbalance within the dataset. This approach synthetically augments the minority class by interpolating new samples within the feature space. The process begins by randomly selecting a minority class sample "a," followed by identifying its k nearest neighbours within the same class. A neighbour "b" is then chosen at random, and a synthetic instance is generated by creating a line in the feature space between "a" and "b" and interpolating their features. Through this method, additional synthetic instances of the minority class are crafted, allowing for the generation of the required number of synthetic examples to achieve class balance. The benefits of this technique include improved models' performance and, a reduced risk of overfitting, especially as the number of samples increases. The proportion of data after oversampling is presented in Table 3.

4.2 Machine learning techniques

In this paper, various machine learning methods are applied to assess their effectiveness in predicting the buckling modes of ballasted railway tracks under temperature variations ranging from simple to highly performant. We employ Python 3.11 for constructing models manually and for results visualisation, executed on a system equipped with a

Table 1 Input parameters

Parameters	Range	Mean	Standard Deviation	Unit
Initial lateral stiffness	120–2632	1438.75	714.22	N/mm
Sleeper's lateral displacement limit	0.5–2	1	0.612	mm
Fastening torsional resistance	75–225	150	75	Nm/rad
Unconstrained length	6–30	24	8.5	m
Initial misalignment	8–32	20	8.9	mm
Increased rail temperature	10–200	67.5	34.5	°C

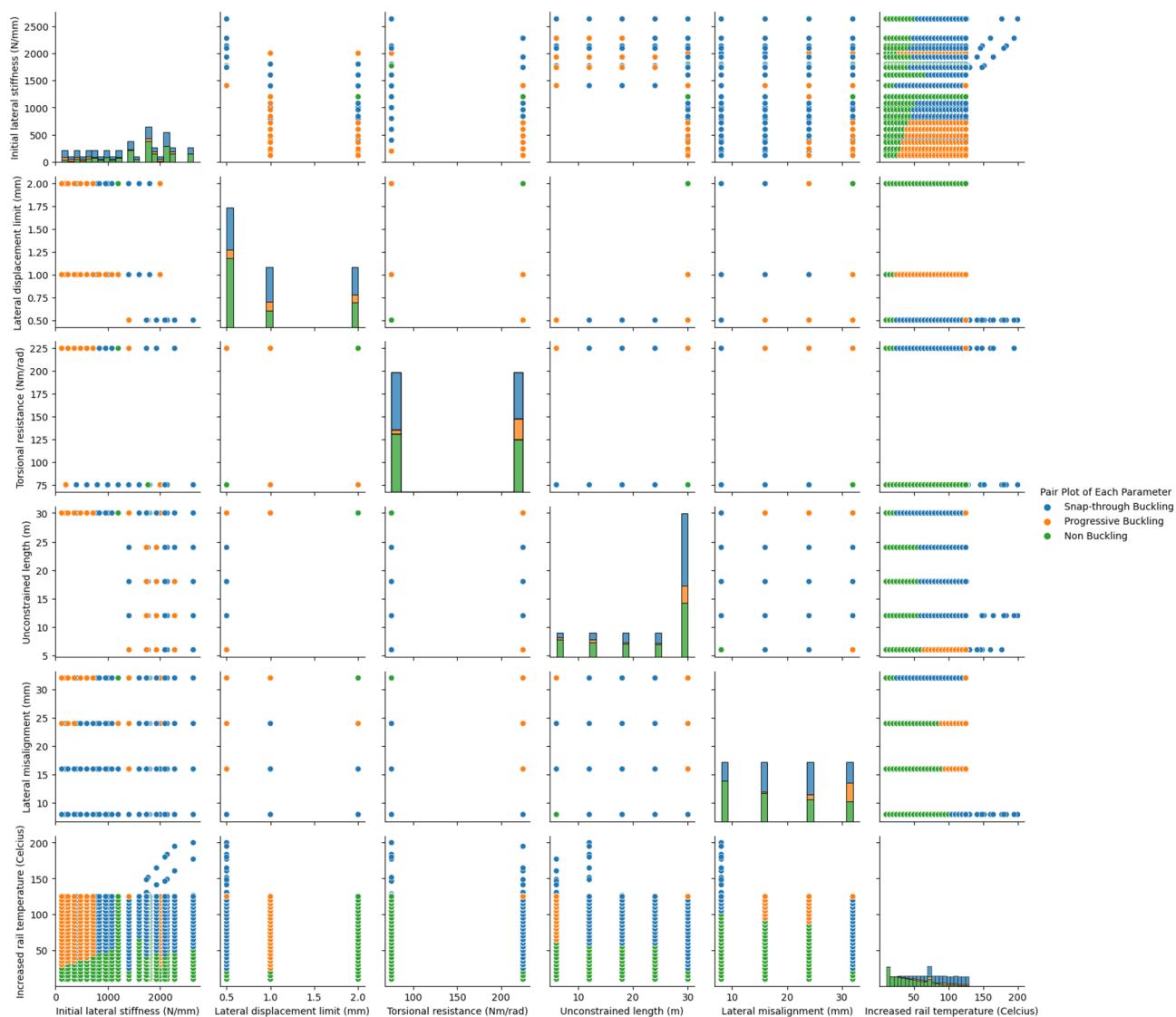


Fig. 4 Input data visualisation

Table 2 Proportion of data splitting before oversampling

Data split	Data instances	Train-test-split ratio	Class split		
			Non-buckling	Snap-through buckling	Progressive buckling
All data	8000	100%	3693	3469	838
Training and validation sets	6400	80%	2954	2775	671
Testing set	1600	20%	739	694	167

12th Gen Intel® Core™ i9-12900K processor at 3.20 GHz, 32.0 GB of RAM, and a 64-bit operating system. Additionally, the K10-Fold Cross Validation technique is used for both training and validation. This method involves partitioning the data into 10 subsets, with each subset being used once as the validation set while the others serve as training data.

Table 3 Proportion of data splitting after oversampling

Data Split	Data Instances	Train-test-split ratio	Class Split		
			Non-buckling	Snap-through buckling	Progressive buckling
All Data	11,079	100%	3693	3693	3693
Training and validation sets	8862	80%	2954	2954	2954
Testing set	2217	20%	739	739	739

4.2.1 Logistic regression

Logistic regression is an algorithm used for binary classification tasks. It models the relationship between a set of input features and a binary target variable by applying the logistic function, which maps a linear combination of features to a probability score between 0 and 1. Through a training process, logistic regression estimates the optimal coefficients to best fit the data, typically using maximum likelihood estimation.

4.2.2 k-Nearest Neighbor (kNN)

The k-Nearest Neighbors (kNN) algorithm, renowned for its adaptability, caters to both classification and regression problems. It operates by considering the k most proximal training exemplars to the query data point within the dataset. For classification, kNN assigns a class label based on the most frequent label among these neighbours. In regression, it predicts a continuous output by averaging the neighbours' values. The default choice for k is 1, although it can be fine-tuned either manually or through automated methods such as leave-one-out cross-validation within predetermined bounds. In our study, we employed the KDTree algorithm from Python 3.11 to implement the kNN model, applying 'uniform' weighting to ensure all neighbours contribute equally to the prediction. The Euclidean distance metric was used to identify the nearest neighbours via a linear search method, with the final output reflecting the mean of the k nearest values.

4.2.3 Decision Tree (DT)

The Decision Tree (DT) algorithm is a prominent machine learning technique used extensively in operations research and decision analysis. This hierarchical method employs a tree-like model to delineate dependencies between dependent and independent variables, making it suitable for both classification and regression tasks. A DT is developed in a top-down approach, initiating with a root node that splits based on attribute values. This process continues through intermediary nodes, each representing attribute tests that lead to further branches. Ultimately, leaf nodes are reached, which provide the predictions. The outcome of a DT is a model where parameters are assigned varying weights indicating their relative importance in predicting the target variable. This methodology is encapsulated in a single decision tree, which serves as a constituent of a Random Forest model that is also analysed in this paper.

In this study, we employed the CART (Classification and Regression Trees) algorithm to develop a classification tree. The construction of the tree was guided by the maximisation of information gain or variance reduction at each node. We further refined the tree through minimal cost-complexity pruning to avoid overfitting. The CART algorithm facilitates the creation of binary trees, selecting the optimal feature and threshold for splitting based on the greatest information gain. It is compatible with continuous numerical variables. Within the Python 3.11 environment, we had the flexibility to configure various hyperparameters such as the minimum number of samples required at a leaf node, the maximum depth of the tree (an adjustment particularly beneficial for boosting algorithms), the criterion for measuring loss, the strategy for splitting nodes, the maximum number of features considered for a split, and the number of folds used in the pruning process.

4.2.4 Random Forest

The Random Forest (RF) algorithm, highlighted in Fig. 5, is a widely-used ensemble machine learning method that aggregates predictions from numerous DTs to produce a singular, consolidated outcome. It constructs an ensemble of DTs, typically 10 or more, where each tree independently predicts an output. In classification tasks, the RF model predicts the class that receives the majority vote from its constituent DTs. Renowned for its versatility and ease of use, RF is a favoured algorithm in the data science community.

This paper discusses the implementation of RF using the `sklearn` library, which involves fitting multiple decision tree classifiers on various sub-samples of the dataset. The DTs are built using the CART algorithm, and refined with minimal cost-complexity pruning to prevent overfitting. The Python 3.11 implementation allows fine-tuning of the RF parameters such as "max_samples" to define sub-sample size, the number of trees in the forest, maximum depth of trees, loss criterion, minimum samples per leaf, and the strategy for splitting nodes and choosing the maximum number of features at each split, similarly to the configuration of individual decision trees.

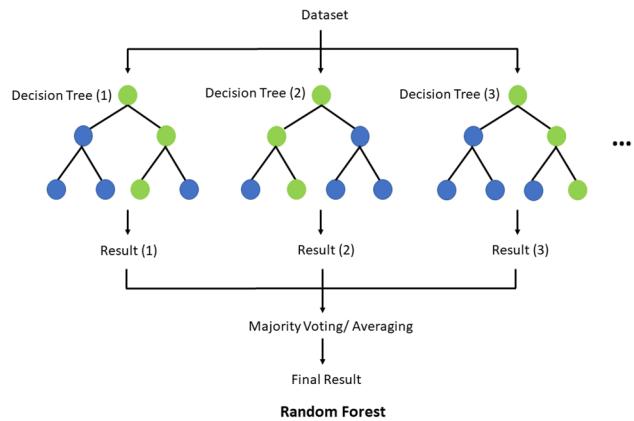
4.2.5 Gradient boosting

Gradient Boosting is a powerful machine learning technique used for both regression and classification tasks. In this work, the authors used this particular technique for classification tasks. It works by combining the predictions of multiple weak learners or trees (typically decision trees) to create a strong predictive model as shown in Fig. 6. The method involves sequential learning and combines all learners to predict final results. Then, to prevent overfitting, regularisation and optimisation methods namely limiting the tree depth, number of weak learners (trees), or other hyperparameters are used. In this study, we use Python 3.11 and Scikit-Learn libraries to implement Gradient Boosting Methods such as XGBoost, and LightGBM.

4.2.5.1 Light Gradient Boosting Machine (LightGBM) Light Gradient Boosting Machine (LightGBM) is a high-performance gradient boosting framework for machine learning. It is specifically designed for handling large datasets and high-dimensional feature spaces efficiently. LightGBM employs a gradient boosting approach to construct an ensemble of decision trees, focusing on the most informative features and using a histogram-based learning technique to speed up the training process. This innovative design leads to faster training times and lower memory usage compared to traditional gradient boosting algorithms.

4.2.5.2 Extreme Gradient Boosting (XGBoost) Extreme Gradient Boosting (XGBoost) is an algorithm operated by creating an ensemble of decision tree models, iteratively refining their predictive capabilities while emphasising the accurate prediction of previously misclassified instances. This is achieved by assigning higher weights to the misclassified samples and sequentially fitting new trees to correct these errors.

Fig. 5 Random Forest



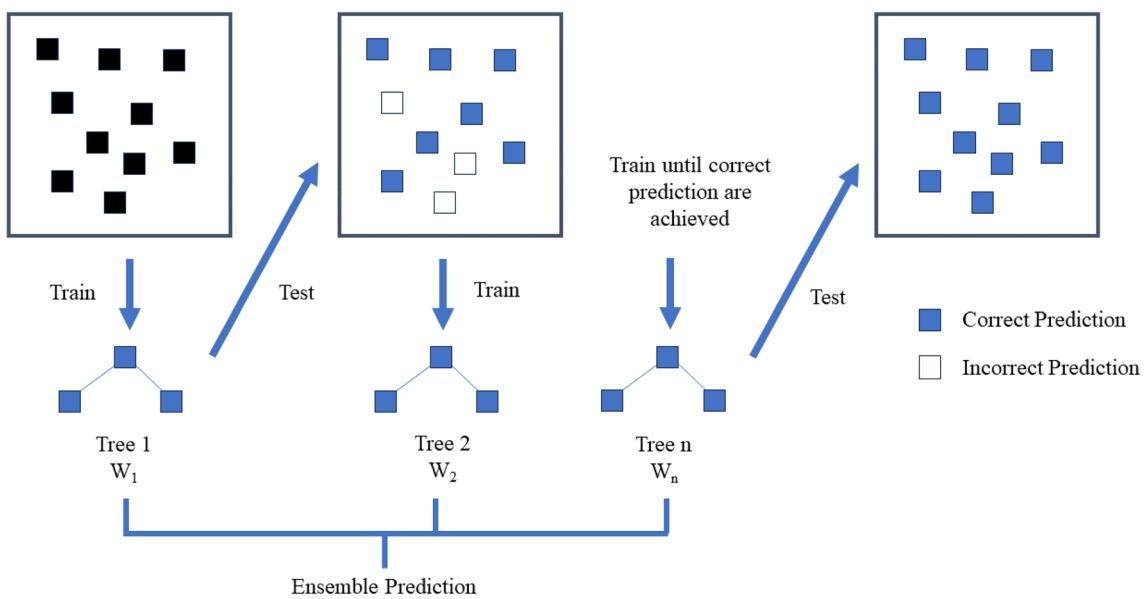


Fig. 6 Gradient boosting

4.3 Performance measures

To evaluate the performance of the models trained in this study, five statistical analyses were employed: Accuracy, F1-Score, Precision, Recall, and Cross Entropy which can be calculated using Eqs. 1–5, respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$F1 - Score = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$H(p, q) = - \sum p(x) \log q(x) \quad (5)$$

where, TP is a test result that correctly indicates the presence of a buckling condition. FP is a test result which wrongly indicates that a certain buckling condition is present. FN is a test result which wrongly indicates that a certain buckling condition is absent. TN is a test result which correctly indicates the absence of a buckling condition. P(x) is the true label probability distribution. Q(x) is the model's predicted probability distribution.

The metrics or performance indicators consist of 5 metrics which are Accuracy, F1-Score, Precision, Recall and Cross Entropy. Accuracy is a metric that captures the overall performance of the model. It calculates from the percentage of all correctly classified observations, ranging from 0, indicating no correctness, to 1 for perfect correctness. The higher accuracy means the overall greater performance of the model. This metric is easy to interpret, but the distribution of data and prediction is unknown through this metric.

Precision is a measure that calculates from correct positive predictions relative to the total positive predictions. It ranges from 0 to 1, the greater value of this metric means there are more relevant results than irrelevant ones. This

metric is used when the consequence of misclassifying an instance as positive is severe which False Positive results have an impact on the metric evaluation, but False Negative results have less impact on the metric evaluation.

Recall is a measure that calculates correct positive predictions relative to the total number of actual positive cases. It ranges from 0 to 1, the greater value of this metric means the greater ability of a model to find all the relevant cases within whole possible positive observations. It is used when the consequence of misclassifying an instance as negative is severe, such as misidentifying a buckling mode as non-buckling. Recall strength lies in that False Negative results have an impact on the metric evaluation, but False Positive has less impact on the metric evaluation.

F1-Score is the metric that combines Precision and Recall by using harmonic mean. This metric score of 1 indicates excellent precision and recall, while a low score indicates poor model performance. It is used when needed a single metric to precisely measure all capabilities of a model. This metric can indicate the distribution of the prediction, but difficult to achieve a high score and interpret.

Cross-entropy often referred to "Log loss" often used as a loss function for training gradient boosting models, including XGBoost and LGBM. It is a loss function that measures the performance of a classification model by quantifying the difference between predicted probabilities and the actual class labels. It is used as a loss function in gradient boosting models and algorithm for classification tasks. Its advantages lie in its calculation as it's designed for classification tasks and works well when needed to estimate class probabilities and regularisation because it discourages models from making overly confident predictions. There was a disadvantage though as values of cross-entropy loss do not have an intuitive interpretation, making it harder to explain the model's performance in simple terms.

5 Results and discussions

5.1 Model performance

Table 4 displays the performance of each machine learning model along with its corresponding rank. Both the Random Forest Classifier and the Decision Tree Classifier show exceptional performance across all metrics within the training dataset. This high level of performance is likely attributed to the discrete nature of the data, which aids in the formation of clear classification rules for the decision algorithm, directly influencing the analysis. However, while these models excel in the training set, their performance diminishes in the testing dataset.

In contrast, models utilising gradient boosting techniques, specifically XGBoost and LGBM, demonstrate superior performance on the testing dataset. Notably, XGBoost and LGBM yield almost identical results across most metrics in the testing dataset, prompting a deeper examination of their true efficacy. To further differentiate their performance, the cross entropy loss metric is employed for comparison, as indicated in the table. This metric reveals that LGBM exhibits a higher cross entropy loss than XGBoost in both datasets, suggesting that XGBoost has a slight edge over LGBM in terms of performance.

The decision tree and random forest models are notably more sensitive among tree-based methods, especially when compared to other ensemble methods. These models categorise cases based on the values of parameters, whereas ensemble methods employ a multitude of learners and values derived from tuning equations to distribute the sensitivity across cases. Given that the results from other metrics are quite similar, a different approach is employed to discern the performance of each model. By using the cross entropy loss metric, a more effective

Table 4 Comparison of metrics between dataset and models

Model	Training set					Testing set					Rank
	Accuracy	Precision	Recall	F1 score	Cross Entropy	Accuracy	Precision	Recall	F1 score	Cross Entropy	
Logistic Regression	0.79/4	0.79/4	0.79/4	0.79/3	–	0.79/5	0.80/5	0.80/5	0.80/5	–	6
k-Nearest Neighbor	0.873/3	0.88/3	0.87/3	0.87/3	–	0.86/4	0.87/4	0.87/4	0.87/4	–	5
Decision Tree	1.00/1	1.00/1	1.00/1	1.00/1	–	0.95/3	0.96/3	0.95/3	0.95/3	–	4
Random Forest	1.00/1	1.00/1	1.00/1	1.00/1	–	0.96/2	0.96/3	0.96/2	0.96/2	–	3
LGBM	0.99/2	0.99/2	0.99/2	0.99/2	0.50/2	0.97/1	0.98/1	0.97/1	0.97/1	0.73/2	2
XGBoost	0.99/2	0.99/2	0.99/2	0.99/2	0.38/1	0.97/1	0.97/2	0.97/1	0.97/1	0.69/1	1

The value expressed as a/b represents the performance value (a) and the rank (b)

comparison of model performance is enabled, allowing us to distinguish between the performances of the various models. In addition, based on Table 4, the matrices' differences between the training and testing sets are not significantly high. Therefore, it can be concluded that there is no overfitting issue in the model development process and the SMOTE technique can be well applied to this study without causing any issue.

The preliminary model selected, as indicated in Table 4, is the XGBoost model. Based on this selection, we present an optimised XGBoost model with tuned hyperparameters using grid search technique. Through meticulous consideration and iterative testing of these parameters, the authors have refined specific hyperparameters, the details of which are presented in Table 5.

Figures 7, 8 illustrate the confusion matrices for the XGBoost model on the training and testing datasets, respectively. In these matrices, the x-axis denotes the true labels, while the y-axis represents the predicted labels. For the training set, from a total of 8862 datasets, the model successfully classifies 8821 datasets, with a mere 41 instances being misclassified. In the case of the testing set, out of 2217 datasets, it accurately classifies 2165 datasets, with only 51 datasets misclassified. Compared to the performances shown in Table 4, it can be seen that the performance of the model is improved after the fine-tuning.

The importance of each parameter is assessed to gain insights into the behaviour of the model and to strategise the maintenance and inspection plan for ballasted tracks. The XGBoost Feature Importance is a technique that assigns a score to every input feature of a given model. These scores indicate the relative 'importance' of each parameter in affecting the model's output. Ranging from 0 to 1, a higher score suggests that a feature significantly influences the model's prediction, while a score of 0 indicates that the feature has no importance or is not used in the model.

Table 5 Hyperparameters of XGBoost model

Hyperparameter	Definition	Default value	Adjusted value
Booster	Method of booster that used	GBTree	GBDart
Maximum tree depth	Maximum depth of each tree	6	5
Learning rate	Step size shrinkage that used to prevent overfitting	0.3	0.1
Gamma	Minimum loss reduction required to make a further partition on a leaf node	0	1
Minimum child weight	Minimum sum of instance hessian weight needed to further partition	1	2.5
Alpha regularisation	L1 regularisation term on weights	0	0.5
Lambda regularisation	L2 regularisation term on weights	1	0.5
Tree Algorithm	Tree construction algorithm used in XGBoost	Auto	Hist (optimised approximate greedy)
Subsample ratio	Subsample ratio of the training instances	1	0.7
Sampling method	Method to use to sample the training instances	Uniform	Gradient Based
Seed (optional)	Random number seed	0	42
Tree grow policy	Method that controls how new nodes are added to the tree	Depthwise	Loss guided

Fig. 7 Confusion matrix of ML models on training datasets (8862 Points)

Legend		True Label (Test Data)		
		Non-Buckling	Snap-Through Buckling	Progressive Buckling
Prediction Label (Model Prediction)	0	2958		
	Non-Buckling	2929	29	2
	Snap-Through Buckling	10	2932	0
	Progressive Buckling	2	0	2958

Fig. 8 Confusion matrix of ML models on testing datasets (2217 Points)

Legend		True Label (Test Data)		
		Non-Buckling	Snap-Through Buckling	Progressive Buckling
Prediction Label (Model Prediction)	0	729		
	Non-Buckling	710	19	3
	Snap-Through Buckling	25	727	0
	Progressive Buckling	3	1	729

Figure 9 illustrates the feature importance of the trained XGBoost model. It reveals that lateral misalignment, torsional resistance, and lateral displacement limit are the most significant factors, in that order. Lateral misalignment plays a crucial role in track buckling, especially in tangent track sections. This is due to the fact that lateral misalignment can introduce an initial curvature to a track section, providing the necessary eccentricity for axial compression forces to induce track buckling. Torsional resistance is another vital factor influencing track buckling behaviour, along with lateral stiffness that composes of lateral displacement limit and lateral resistance.

Furthermore, the XGBoost Model, along with its hyperparameters, can be conveniently saved using the ".save_model" command, which includes "num_iteration" as a hyperparameter. The model can then be reloaded using the ".Booster" command, enabling it to read from .txt or .json files. This process provides the utmost convenience in exporting, training, and utilising the most effective model available. The model can be effectively applied to a set of features corresponding to the physical parameters of rail tracks. By employing a pre-trained XGBoost Model, results can be reliably replicated, achieving high accuracy, F1-score, precision, recall, and cross-entropy scores.

5.2 Buckling mode transition diagram

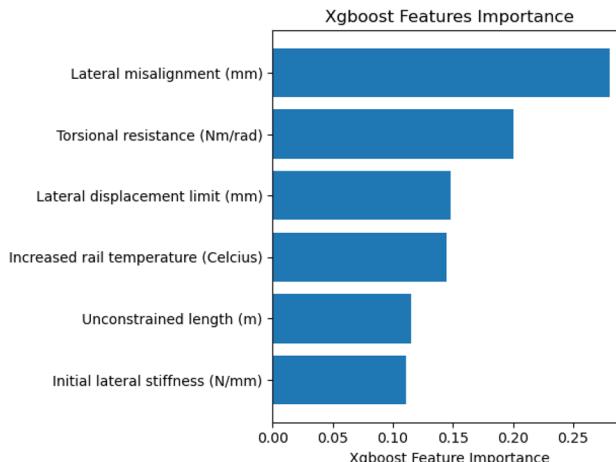
The pre-trained XGBoost model is then used to study the buckling mode-change and their effects on buckling mode. This analysis is visually represented in a diagram that specifically highlights how these buckling modes vary under various conditions and parameters. Figure 10 illustrates the prediction of buckling modes, taking into account lateral resistance and the displacement limit of sleepers while keeping other parameters constant. These plots display a distinct boundary for each type of buckling. The boundary between the buckling and non-buckling areas corresponds to the maximum temperature that can be sustained before buckling occurs. It is evident that progressive buckling modes typically occur when lateral stiffness is low at lower temperatures. As the lateral displacement limit increases, leading to greater lateral resistance force, it becomes apparent that railway tracks are able to withstand higher temperatures, and the occurrence of snap-through buckling becomes more pronounced.

Figure 11 depicts the predicted modes of buckling, taking into account lateral misalignment and initial lateral stiffness, while maintaining a constant lateral displacement limit of 2 mm. The analysis reveals that snap-through buckling typically happens at lower levels of lateral misalignment, albeit at higher temperatures compared to progressive buckling. Progressive buckling, in contrast, tends to occur with the introduction of lateral misalignment. Additionally, it is observed that as lateral resistance increases, snap-through buckling becomes more prominently the sole mode of buckling.

5.2.1 Case study

The case study from Narathiwat in Southern Thailand, illustrated in Fig. 12, presents an instance of symmetrical buckling. In this scenario, we consider a railway system with an approximate unconstrained length of 8 m. Since track alignments were not recorded at the time of the observation (April 20, 2017), the lateral misalignment length was set within the permissible range defined by the model's constraints. Notably, the absence of a ballast shoulder is observed in this scenario, leading to the assumption that we are dealing with a timber sleeper that lacks a ballast shoulder. Therefore, the

Fig. 9 XGBoost Feature Importance



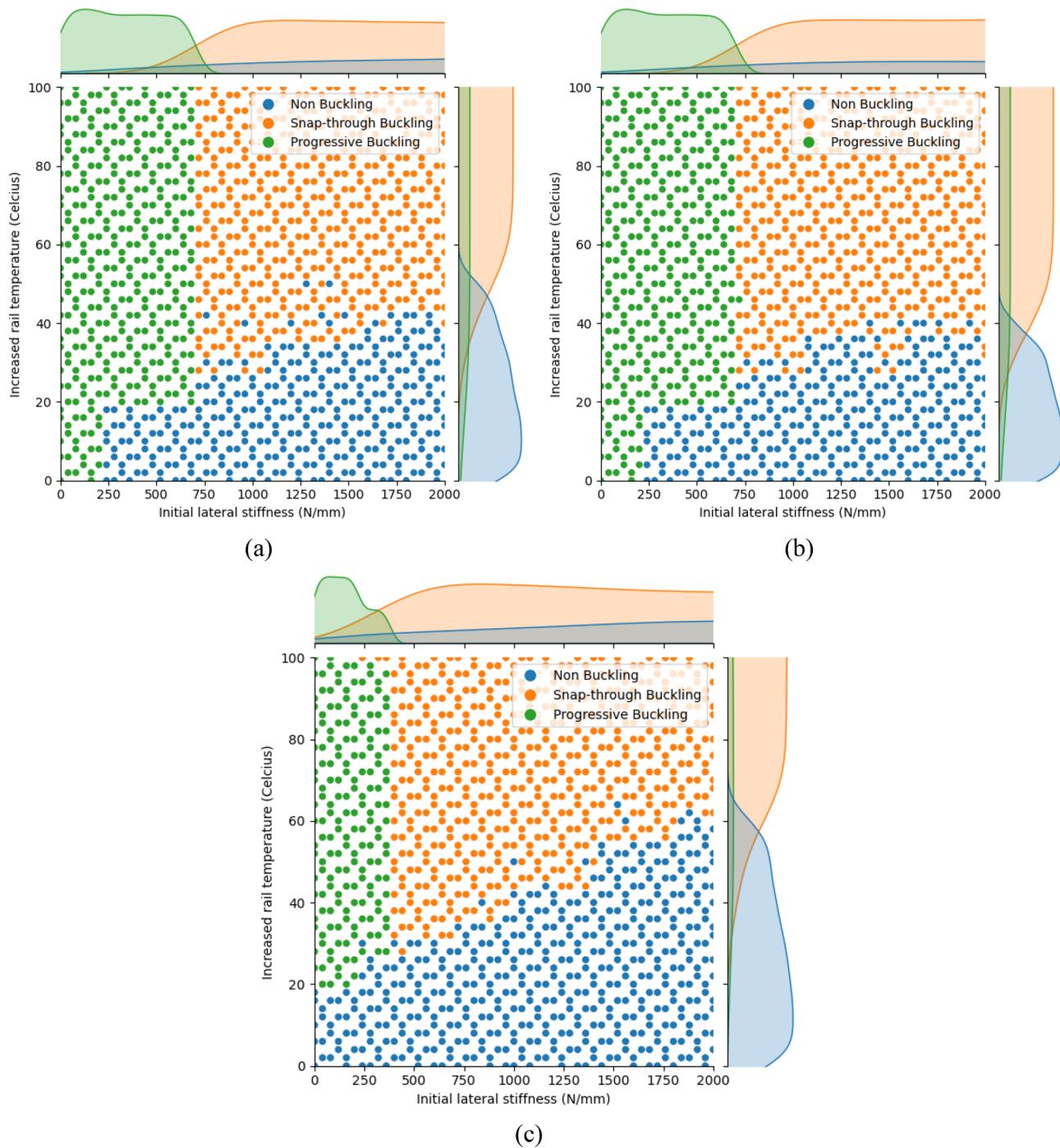


Fig. 10 Buckling mode prediction considering initial lateral stiffness and sleeper's displacement limit **a** 0.5 mm **b** 1 mm **c** 2 mm

model's initial stiffness and the lateral displacement limit are set at 200 N/mm and 1 mm, respectively, based on prior results obtained from STPT test in [24].

On April 20, 2017, historical data indicates that the air temperature was recorded at 36 °C. Based on this, it was assumed that the rail temperature rose to 56 °C. However, it's important to consider that in Southern Thailand, the Stress Free Temperature (SFT) for rails is typically 30 °C. Consequently, the increase in rail temperature was assumed to be from the SFT of 26 °C.

The proposed model is utilised for analysis, setting several parameters based on above assumptions. These include the initial lateral stiffness of the track, the displacement limit for sleepers, and the track's unconstrained length.

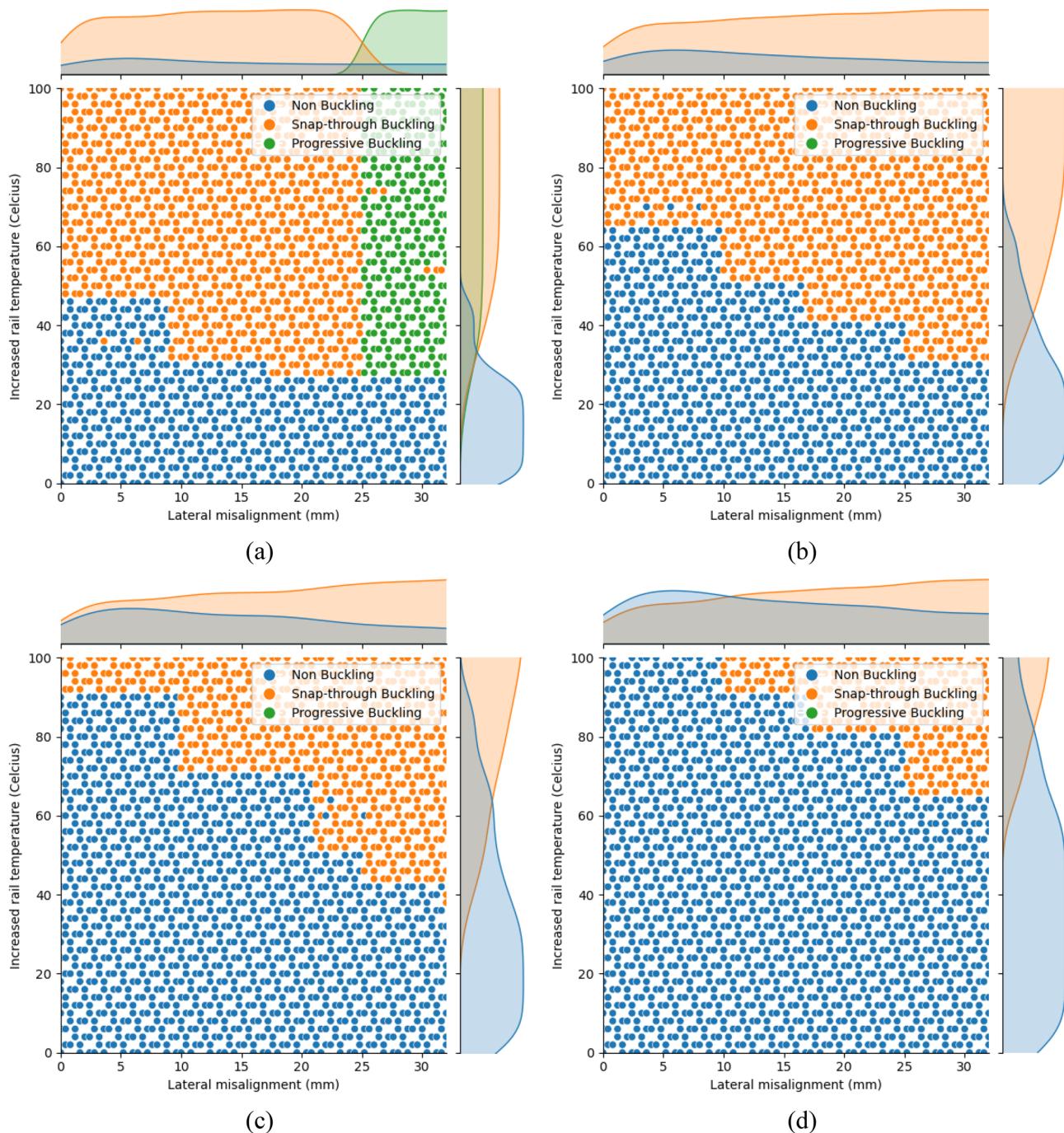


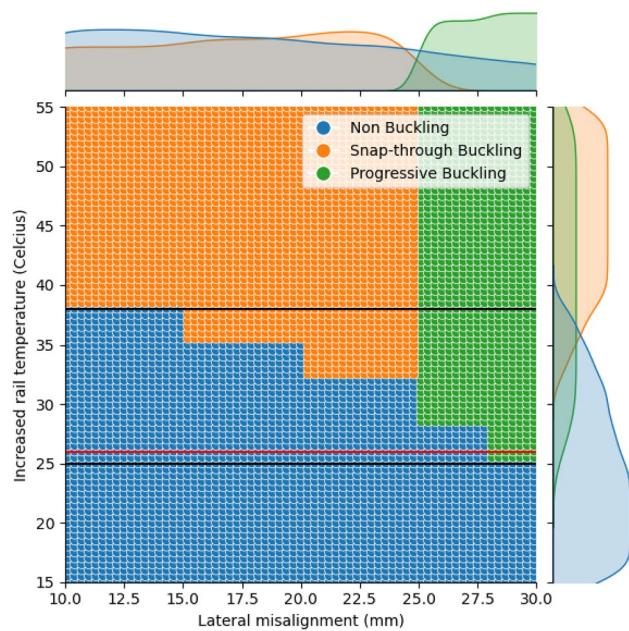
Fig. 11 Buckling mode prediction considering lateral misalignment and lateral resistance **a** 300 N/mm **b** 600 N/mm **c** 1200 N/mm **d** 1800 N/mm

Our focus is on varying the lateral misalignment within a specific range to understand its impact. Additionally, we adjust the rail temperature to examine both the temperature at which buckling occurs and the mode of buckling. Figure 13 illustrates the relationship between the increase in rail temperature and lateral misalignment. We observe that buckling begins at a temperature of 25 °C, exhibiting a progressive buckling mode when the initial lateral misalignment is larger than 28 mm. Notably, the predicted temperature at which buckling occurs aligns closely with temperatures recorded in real-world field conditions with only lower bound difference of 3.9% illustrated in Fig. 13. This also corresponds well with the phenomena of progressive buckling mode, which typically occurs with small

Fig. 12 A case study in Thailand (Photo courtesy of State Railway of Thailand)



Fig. 13 Results of a case study using ML model



lateral displacement after buckling as observed in the field. Therefore, the proposed machine learning model is trustworthy and can be used further for the buckling prediction and early warning system development.

6 Conclusions

This study has demonstrated the substantial potential of machine learning (ML) in enhancing the predictability of buckling failure modes in ballasted railway tracks. Through the deployment of various advanced ML algorithms, including Logistic Regression, Decision Tree, k-Nearest Neighbor, Random Forest and XGBoost, LGBM, we have successfully developed models that offer accurate predictions of track buckling behaviour under a range of conditions. The significant achievement of these models, especially the XGBoost algorithm which recorded an F1 score of 0.97, marks a pivotal advancement in the field of railway engineering. This outcome not only demonstrates the robustness of ML models in complex failure prediction but also opens new avenues for enhancing railway track safety and operational efficiency. We are the first to introduce a detailed analysis of the track buckling phases using ML models, taking into account a variety of different parameters.

However, it is essential to note that the limitations of the current models, particularly the reliance on specific parameters that require prior calculations, such as initial lateral stiffness and sleeper's displacement limit that could be linked

to different physical geometries and material properties of sleeper and ballast. Furthermore, the rail misalignments are approximated using a sine wave, a simplification that contrasts with the more complex realities of actual track alignment. This model focuses on filtering out the short wavelength variations in rail alignment, offering a simplified yet effective representation of the tracks. By adopting this approach, the analysis becomes more practical, despite the acknowledgment that real-world rail deviations may exhibit greater complexity. Thus, prior to incorporating the actual track alignment into this machine learning model, an initial justification is required. This step ensures that the simplified sine wave representation aligns with the intended analysis objectives, balancing simplicity with the need to accurately reflect real-world track conditions. These limitations highlight areas for future research, where efforts could be directed towards refining the models for greater applicability in in-field scenarios. In conclusion, the integration of machine learning into the prediction and management of railway track buckling behaviour represents a significant step forward in railway engineering. The models developed in this study offer a promising direction for future research and practical application, paving the way for safer, more efficient railway systems.

Acknowledgements This project is supported by National Research Council of Thailand (NRCT: N42A650199). This Research is also funded by Thailand Science Research and Innovation Fund Chulalongkorn University. Grants for development of new faculty staff, Ratchadaphiseksomphat Fund, Chulalongkorn University is also gratefully acknowledged. The authors would like to acknowledge the support from Chulalongkorn University for the establishment of Advanced Railway Infrastructure, Innovation and Systems Engineering (ARIISE) Research Unit.

Author contributions W.W., W.M Formal analysis, data curation, visualization. C.N. Conceptualization, data curation, methodology, validation, supervision. J.S. Methodology, validation, supervision. S.K. Validation, supervision. All authors reviewed the manuscript.

Data Availability Data sets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests 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.

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

References

1. Oslakovic IS, et al. Risk assessment of climate change impacts on railway infrastructure. 2013.
2. Ngamkhanong C, Kaewunruen S, Costa BJA. State-of-the-art review of railway track resilience monitoring. *Infrastructures*. 2018;3(1):3. <https://doi.org/10.3390/infrastructures3010003>.
3. Quinn AD, et al. Rail adapt: adapting the railway for the future. A Report for the International Union of Railways (UIC), 2017.
4. Esveld C. Modern railway track, vol. 385. Netherlands: MRT-productions Zaltbommel; 2001.
5. Kish A. On the fundamentals of track lateral resistance. American Railway Engineering and Maintenance of Way Association. 2011.
6. Esveld C. Improved knowledge of CWR track. 1997; p. 8–9.
7. Ahmad SSN, Mandal NK, Chattpadhyay G. A comparative study of track buckling parameters on continuous welded rail. p. 26–28.
8. Ling L, et al. Numerical simulation of dynamical derailment of high-speed train using a 3D train-track model.
9. Ling L, Xiao XB, Jin XS. Development of a simulation model for dynamic derailment analysis of high-speed trains. *Acta Mech Sin*. 2014;30(6):860–75. <https://doi.org/10.1007/s10409-014-0111-0>.
10. Kaewunruen S, Wang Y, Ngamkhanong C. Derailment-resistant performance of modular composite rail track slabs. *Eng Struct*. 2018;160:1–11. <https://doi.org/10.1016/j.engstruct.2018.01.047>.
11. Li Y, et al. Nonlinear responses of longitudinally coupled slab tracks exposed to extreme heat waves. *Eng Struct*. 2023;281: 115789. <https://doi.org/10.1016/j.engstruct.2023.115789>.
12. Cuadrado M, et al. Analysis of buckling in dual-gauge tracks. *Proc Inst Civil Eng Transp*. 2008;161:177–84. <https://doi.org/10.1680/tran.2008.161.4.177>.
13. Villalba I, et al. Methodology for evaluating thermal track buckling in dual gauge tracks with continuous welded rail. *Proc Inst Mech Eng Part F: J Rail Rapid Transit*. 2017;231(3):269–79. <https://doi.org/10.1177/0954409715626>.
14. Yang G, Bradford MA. Thermal-induced buckling and postbuckling analysis of continuous railway tracks. *Int J Solids Struct*. 2016;97:637–49. <https://doi.org/10.1016/j.ijsolstr.2016.04.037>.
15. Ngamkhanong C, Wey CM, Kaewunruen S. Buckling analysis of interspersed railway tracks. *Appl Sci*. 2020;10:3091. <https://doi.org/10.3390/app10093091>.

16. Miri A, et al. Analysis of buckling failure in continuously welded railway tracks. *Eng Fail Anal.* 2021;119: 104989. <https://doi.org/10.1016/j.engfailanal.2020.104989>.
17. Yang G, Bradford MA. On train speed reduction in circumstances of thermally-induced railway track buckling. *Eng Fail Anal.* 2018;92:107–20. <https://doi.org/10.1016/j.engfailanal.2018.02.009>.
18. Machan S, et al. Eigenvalue-based approach for buckling analysis of metre gauge railway tracks incorporating train load effects. *Transp Eng.* 2023;14: 100209. <https://doi.org/10.1016/j.treng.2023.100209>.
19. Pucillo GP. Thermal buckling and post-buckling behaviour of continuous welded rail track. *Veh Syst Dyn.* 2016;54(12):1785–807. <https://doi.org/10.1080/00423114.2016.1237665>.
20. Kish A, Samavedam G. Track buckling prevention: theory, safety concepts, and applications. Cambridge: John A. Volpe National Transportation Systems Center; 2013.
21. Kish A. On the fundamentals of track lateral resistance. In: Annual Conference. 2011: Minneapolis, USA.
22. Jing G, Aela P. Review of the lateral resistance of ballasted tracks. *Proc Inst Mech Eng Part F: J Rail Rapid Transit.* 2020;234(8):807–20. <https://doi.org/10.1177/09544097198663>.
23. Guo Y, et al. Effect of sleeper bottom texture on lateral resistance with discrete element modelling. *Constr Build Mater.* 2020;250: 118770. <https://doi.org/10.1016/j.conbuildmat.2020.118770>.
24. Ngamkhanong C, et al. Evaluation of lateral stability of railway tracks due to ballast degradation. *Constr Build Mater.* 2021;278: 122342. <https://doi.org/10.1016/j.conbuildmat.2021.122342>.
25. Miri A, et al. Mitigation of track buckling in transition zones of steel bridges by geotextile reinforcement of the ballast layer. *Geotext Geomembr.* 2022;50(2):282–92. <https://doi.org/10.1016/j.geotexmem.2021.11.006>.
26. Kerr AD. An improved analysis for thermal track buckling. *Int J Non-Linear Mech.* 1980;15(2):99–114.
27. Kerr AD. Analysis of thermal track buckling in the lateral plane. *Acta Mech.* 1978;30(1–2):17–50. <https://doi.org/10.1007/BF01177436>.
28. Park YS, Lek S. Chapter 7 - artificial neural networks: multilayer perceptron for ecological modeling. In: Jørgensen SE, editor. Developments in environmental modelling. Amsterdam: Elsevier; 2016. p. 123–40. <https://doi.org/10.1016/B978-0-444-63623-2.00007-4>.
29. Mohamad Al iRidho BKA, et al. Recycled aggregates concrete compressive strength prediction using artificial neural networks (ANNs). *Infrastructures.* 2021;6(2):17. <https://doi.org/10.3390/infrastructures6020017>.
30. Huang X, et al. Machine learning aided design and prediction of environmentally friendly rubberised concrete. *Sustainability.* 2021;13(4):1691. <https://doi.org/10.3390/su13041691>.
31. Sresakoolchai J, Kaewunruen S. Detection and severity evaluation of combined rail defects using deep learning. *Vibration.* 2021;4(2):341–56. <https://doi.org/10.3390/vibration4020022>.
32. Nakhaee MC, et al. The recent applications of machine learning in rail track maintenance: a survey. *RSSRail.* 2019. https://doi.org/10.1007/978-3-030-18744-6_6.
33. Li W, et al. Track slab crack detection based on full convolutional neural network. *J Phys: Conf Ser.* 2021. <https://doi.org/10.1088/1742-6596/1848/1/012163>.
34. Khajehei H, et al. Prediction of track geometry degradation using artificial neural network: a case study. *Int J Rail Transp.* 2021. <https://doi.org/10.1080/23248378.2021.1875065>.
35. Do NT, Güll M. Estimations of vertical rail bending moments from numerical track deflection measurements using wavelet analysis and radial basis function neural networks. *J Transp Eng Part A: Syst.* 2021. <https://doi.org/10.1061/JTEPBS.0000489>.
36. Pereira Silva C, Dersch MS, Edwards JR. Quantification of the effect of train type on concrete sleeper ballast pressure using a support condition back-calculator. *Front Built Environ.* 2020;6:214. <https://doi.org/10.3389/fbuil.2020.604180>.
37. Alawad H, Kaewunruen S, An M. A deep learning approach towards railway safety risk assessment. *IEEE Access.* 2020;8:102811–32. <https://doi.org/10.1109/ACCESS.2020.2997946>.
38. Alawad H, Kaewunruen S, An M. Learning from accidents: machine learning for safety at railway stations. *IEEE Access.* 2020;8:633–48. <https://doi.org/10.1109/ACCESS.2019.2962072>.
39. Ngamkhanong C, Kaewunruen S. Prediction of thermal-induced buckling failures of ballasted railway tracks using artificial neural network (ANN). *Int J Struct Stab Dyn.* 2022;22(05):2250049.
40. Ngamkhanong C, Kaewunruen S, Baniotopoulos C. Influences of ballast degradation on railway track buckling. *Eng Fail Anal.* 2021;122: 105252. <https://doi.org/10.1016/j.engfailanal.2021.105252>.
41. Ngamkhanong C, Kaewunruen S, Baniotopoulos C. Nonlinear buckling instabilities of interspersed railway tracks. *Comput Struct.* 2021;249: 106516. <https://doi.org/10.1016/j.compstruc.2021.106516>.
42. Carvalho J, et al. A new methodology for evaluating the safe temperature in continuous welded rail tracks. *Int J Struct Stab Dyn.* 2013;13(2):1350016. <https://doi.org/10.1142/S0219455413500168>.
43. Prud'Homme MA, Janin MG. The stability of tracks laid with long welded rails. *Rail Int.* 1969;46:459–87.
44. Lichtberger B. The lateral resistance of the track. *European Railway Review.* 2007; p. 68–71.

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