Railway defect detection based on track geometry using supervised and unsupervised machine learning

Sresakoolchai, Jessada; Kaewunruen, Sakdirat

DOI:
10.1177/14759217211044492

Railway defect detection based on track geometry using supervised and unsupervised machine learning

Abstract

Track quality affects passenger comfort and safety. To maintain the quality of the track, track geometry and track component defects are inspected routinely. Track geometry is inspected using a track geometry car. Measured values are stored in the machine and processed to evaluate the track quality. However, track component defects require more effort to inspect. Track component defects can be inspected manually which is time- and workload-consuming or using sensors installed at additional cost. This study presents an approach using track geometry obtained by a track geometry car to detect track component defects, namely, rail, switch and crossing, fastener, and rail joint defects. Detection models are developed using several supervised machine learnings. The relationships between track component defects are analyzed to gain insights using unsupervised machine learnings. From the study, the best model for detecting track component defects using track geometry is a deep neural network with an accuracy of 94.31% followed by a convolutional neural network with an accuracy of 93.77%. For the exploration of insights, k-means clustering is used to cluster the track components defects and association rules are used to find the relationships between them. Examples of the insights from applying these two techniques are that switch and crossing defects are usually found where the radius of curvature is less than 2,000 m and the gradient is positive, the most common defects when the radius of curvature higher than 4,000 m are rail defects, or a worn wing rail will be found when the rail section has failed, ties in switches and worn point blades are found with the confidence of 92.17%. The findings of the study can be applied to detect track component defects using track geometry where additional cost is not required and unsupervised machine learning provides the insights that will be beneficial for railway maintenance. The information obtained from machine learning models will be a complementary information to support decision making and improve the maintenance efficiency in the railway industry.

Keywords: Track Geometry, Railway Track Component Defect, Supervised Learning, Unsupervised Learning, Deep Neural Network, Convolutional Neural Network, K-Means Clustering, Association Rules

Highlight

- Foot-by-foot track geometry is used to detect track component defects.
- 4 years of 30 km of track geometry data providing more than 170k samples are used.
- The proposed approach does not require additional costs for developing the detection system.
- The developed detection model has an accuracy of more than 90%.
- Insights of track component defects are explored using unsupervised learning.

1. Introduction

Railway transportation is known to be one of the safest transportation modes. However, if accidents occur, they can result in very serious consequences. Railway defects affect passenger comfort and safety significantly; therefore, maintaining the tracks in good condition is required. The condition of tracks is measured by track geometry and the level of track component defects. For track geometry, a common method of measurement is using a track geometry car (TGC) [1]. The TGC measures deviations from the designed geometrical characteristics [2]. In this study, data obtained from the TGC consists of longitudinal level (surface), alignment, gauge (gage), twist (crosslevel), and superelevation as shown in Figure 1. Track component defects are inspected using manual inspection, laser technology [3], or axle box acceleration [4], etc. It can be seen that component defect inspection is time-consuming or requires the additional installation cost of equipment while track geometry can be measured faster. This study proposes an approach to detect track component defects using the track geometry from the TGC. The process is faster and cost-saving because the TGC is operated along the track with a speed of up to 70 mph and the measurement can be used to evaluate both track geometry and track component defects.

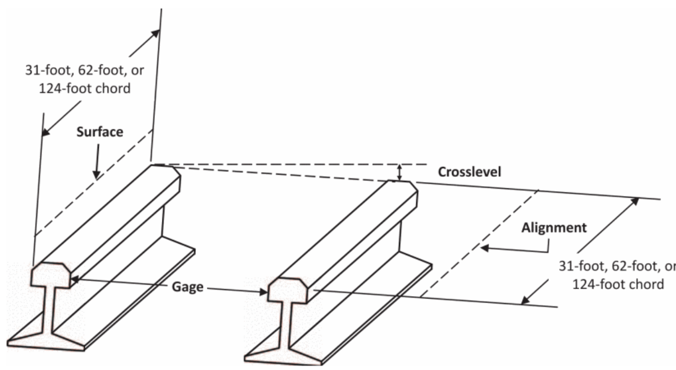

Figure 1 Track geometry [5]

In this study, the TGC measured the track geometry every 1 foot where the total length of the track was 30 km for 4 years, 2016–2019. Therefore, the number of samples was high. To process and develop a detection model, machine learning is appropriate to deal with the large volume of data by which many studies proved that the performance was satisfied. In this study, supervised learning techniques were used to develop models to detect track component defects. There were 70 types of track component defect according to raw data so they were grouped according to the track components consisting of rail, switch and crossing, fastener, and rail joint defects. Some defects that might not have been directly related to the track geometry and samples with missing measurements were removed. Further detail is presented in 3.3.1 and 3.3.2. The supervised learning techniques used in this study were the deep neural network (DNN), the convolution neural network (CNN), multiple regression (MR), the support vector machine (SVM), gradient boosting (GB), a decision tree (DT), and random forest (RF).

Unsupervised learning techniques were used to explore the insights among defects that can improve decision-making. K-means clustering was used to cluster defects into groups using features that were tested to cluster defects. The result from k-means clustering provided insights about which characteristics of the track needed particular attention. Association rules were also used to analyze the relationships between defects. The result from applying association rules makes defects more predictable when particular defects are found in a track section. For example, if two specific defects are detected in a track section, another defect is likely to occur as well.

The results from the study are expected to improve railway maintenance. First, track component defects are accurately detected using the measurement from a TGC which is fast and cost-efficient. Second, the insights from unsupervised learning techniques improve decision-making and defect detection because staff know which sections of the track need special attention. Overall, the developed approach can improve the defect detection system and understanding of track component defects without additional cost. Therefore, any railway operator can apply the results of this study to improve maintenance capabilities.

2. Literature review

Railway track component defects affect passenger comfort and safety significantly. However, accurate inspection requires time and cost to perform. Visual or manual inspection is a traditional method of inspecting component defects. The inspection is then developed to be more automatic and less disruptive to the railway operation. Examples of advanced techniques used to inspect the track defects are laser [6], image processing [7], computer vision [8], and machine learning [9]. However, these techniques need additional installation which creates cost and the installation for a whole route cannot be done easily.

In contrast, track geometry inspection using a TGC is a traditional method widely used by railway operators around the world. Therefore, there are attempts to use the measurement by the TGC to detect track component defects because the TGC is operated regularly to inspect the track geometry. If the measure can be used to detect track components, it will be beneficial to railway operators. Lasisi and Attoh-Okine [10] calculated Track Quality Indexes (TQIs) using selected track geometry parameters. They claimed that this method could be used without losing the relevant information. They applied several machine learning techniques for binary classification (with and without defects). The best model was SVM where accuracy was more than 90%. Martey, Ahmed [2] focused on the substructure of the track. They studied the effect of geocells on track geometry quality using different machine learnings. They also simplified the track geometry using principal component analysis (PCA). In their study, RF performed the best.

Sadeghi and Askarinejad [11] stated that using TGCs to collect data has limitations in identifying track structural defects. Therefore, they aimed to find relationships between track geometry and track structural problems. In the study, they classified the defect severity according to defects' characteristics and components. For example, severities of bent rail was low while rail broken was high. Standard deviation (SD) of track geometry was used to predict track defects. They found that the severities of track defects increased when SD increased. The interesting defects were rail, sleeper, fastening, and ballast. Each component was differently affected by SDs. For example, gauge had the highest effect on fastenings while profile had the highest effect on ballast.

Zarembski, Einbinder [12] studied on relationships between rail and geometry defects and predicted the life time. The technique they used in the study was multivariate regression splines. They found that the relationships between rail and geometry defects had the correlation of 11%.

Soleimanmeigouni, Ahmadi [13] developed a data-driven approach to predict isolated track geometry defects. Data was collected using TGCs. The duration of data collection was four years from 2015-2018 and the distance was 84 km. The study focused on the isolated longitudinal level defects. They found that the developed linear model performed well in detecting defects. They also applied machine learning to detect defects. They developed models using logistic regression when inputs were the standard deviation and kurtosis of longitudinal level.

Mohammadi, He [14] studied the effect of track geometry on the occurrence of rail defects. They also used machine learning techniques consisting of extreme gradient boosting (XGBoost), SVM, RF, and logistic regression for binary classification (with and without defects). They found that XGBoost was the best model for their problem with an accuracy of 80%. Sharma, Cui [15] used the track geometry measurement to predict the geo-defect occurrence probability using RF so their problem was slightly different from other studies because their problem was a regression problem. They claimed that the results of their study could save up to 10% of the maintenance cost.

From the literature review, it can be seen that the use of track geometry to detect rail defects is explored and the results show satisfying performance. However, the use of track geometry to detect track component defects has not been studied. Moreover, the application of unsupervised learning to discover the relationships between defects has not been done. Therefore, there is a gap in this area that should be explored for better railway maintenance.

3. Methodology

Machine learning is the development of computer algorithms to make machines learn from data. Machine learning is mainly categorized into four types, namely, supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. In this study, supervised and unsupervised learning were used. Supervised learning is suitable for pairing features and labeled data. In the study, features are data collected from TGCs and labeled data is track component defects. Supervised learning was used to develop predictive models that the paper aims to predict rail component defects based on data obtained from TGCs. For unsupervised learning, it is used to discover the data insights when data labeling is not needed. In the study, this technique was used because the authors wanted to investigate the relationships between rail defects such as characteristics of track that affect some types of defects and group of defects that tends to occur together. Further detail is presented in the following section.

3.1 Data description

To develop supervised and unsupervised machine learning models, data was collected from a railway operator. The data used in the study consisted of three sets of data, namely, track geometry measurement collected by the TGC, track component defects from site inspection, and the track profile as shown in Figure 2. The length of the studied section is 30 km. Data was collected in2016–2019.

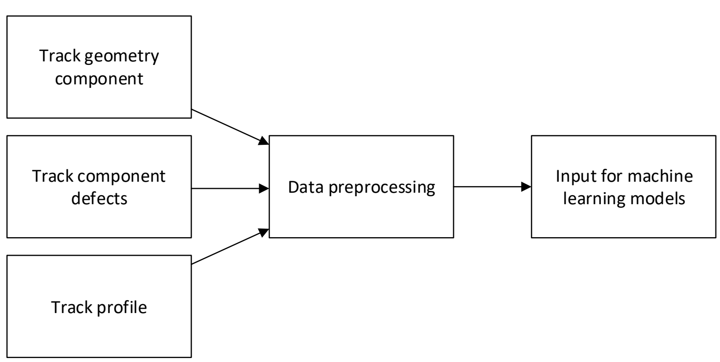

Figure 2 Data components

For track geometry, the measurement is done foot-by-foot so there are more than 95,000 track geometry measurements every year. A track geometry measurement consists of 12 values, namely, superelevation, longitudinal level (10 m chord) of the right rail, longitudinal level (20 m chord) of the right rail, longitudinal level (10 m chord) of the left rail, longitudinal level (20 m chord) of the left rail, alignment (10 m chord) of the right rail, alignment (20 m chord) of the right rail, alignment (10 m chord) of the left rail, alignment (20 m chord) of the left rail, gauge (without load), gauge (with load), and twist (20 m chord).

For track component defects, data was collected indicating the date, location, and types of defect. The locations of the defects were indicated as km-start and km-end of the track so data preprocessing was required with combined track geometry and defects. In the data collection period, more than 1,900 defects were found in the studied section of track section studied and 81 types of defect were indicated. Examples of defects are boltless joints, damaged frogs, broken rails, and broken welds. It can be seen that defects can be grouped as components.

For track profile, the data used in this study indicated the radius of curvature and gradient. They were included because the authors expect that these two variables affect track component defects.

Three sets of data were preprocessed and combined to create the datasets for machine learning models. There were differences in the datasets used in each model. For supervised learning, datasets were the same in every model consisting of two parts, features and labels. For unsupervised learning, datasets were different depending on techniques. Data was preprocessed based on track geometry measurement so a given sample represented a foot of track. Visual Basic for Applications (VBA) was used for data preprocessing. Based on locations and dates of defects, track geometry measurement was extracted. Sections without defects were labeled as defect-free sections. As mentioned, there are 12 values in measurement, and in this study, measurements at the start and the end of a section were extracted from the measurement. Therefore, 24 measurements were extracted and combined with defect data.

Machine learning models for defect detection using track geometry were preliminarily developed to investigate the suitability of features and labels. Different features were initially tested to screen which groups of features provided better performance. It was found that accuracies are improved when 37 features are used. The first 24 features were 12 measurements at the start and the end of each track section, another 12 features were the differences of measurements between the start and the end of each section, and the last feature was a binomial feature indicating if the section is a tangent or curve. Two ends of measurement were used because it was assumed that they could represent conditions of tracks better than only one end of measurement. The preliminary test confirmed this assumption so this set of features was used to develop models. For labels, the authors found that 81 types of defect could not be used as labels directly because the accuracies were poor due to several labels. Therefore, defects were grouped according to components and some defects that were not directly related to track geometry were removed in the data filtering process. Samples with missing values were also removed in this process. From data grouping, there were five classes, namely, switch and crossing defect, fastener defect, railway joint defect, railway defect, and defect-free. In total, there were 172,436 samples in the dataset for developing defect detection models using supervised learning. Training data and testing data were split with a 70/30 proportion. Based on the raw data, the number of samples was more than 400,000 samples. However, it was found that most of samples were defect-free samples. Therefore, the authors randomly removed some samples without defects to keep the samples with and without defects balance. The proportion of samples with and without defects is about 60% and 40% respectively. However, within the defective samples, the number of each type of defect is different because the data is based on the field data. Some types of defect are hardly found compared to others. For example, switch and crossing defects are relatively rare compared to rail defects. The authors want to keep these samples because they represent the field data.

For unsupervised learning, k-means clustering and association rules were used in the study. K-means clustering uses features to cluster data. In this study, the radius of curvature and gradient were used as features to cluster defects. Association rules use only defects to discover the relationships between them. To define sections with defects, a 50 m section was used to determine if defects were included in sections. For example, if an interval of two defects was less than 50 m, these two defects were included in the section and used to determine the relationship. However, if the interval between two defects was more than 50 m, they were considered independent defects which were not related to each other. From this principle, if other defects were within the 50 m distance, they were also included in the section.

3.2 Supervised learning

Supervised learning is an algorithm to map features to labels or input to output. The algorithm is trained using a dataset labeled. Supervised learning is explained in the mathematical equation as (1) [16].

$$using(x_i, y_i)_{i=1}^{l}, find\ f: x \rightarrow y\ = f(x) \tag{1}$$

Where training set x = equation finding the relationship between independent and dependent variables. The algorithm is trained to fit a model to map the input to the output as (2).

$$y^t = g(x^t|\theta) \tag{2}$$

Where $g(\cdot)$ is the model and θ is parameters. During the training, the algorithm minimizes the error using a loss function as (3).

$$\arg\min_{\theta} \sum_t L(r^t, y^t) = \arg\min_{\theta} \sum_t L\big(r^t, g(x^t|\theta)\big) \tag{3}$$

Supervised learning techniques are also categorized as regression if the prediction is continuous and classification if the output is discrete. In this study, supervised learning techniques used were DNN, CNN, MR, SVM, GB, DT, and RF. Features and labels are the same in every technique as mentioned in the previous section.

3.2.1

3.2.8 Grid search

Some parameters are not trained during the training but they are pre-defined. Therefore, to ensure that the models deliver the best performance, hyperparameter tuning is done. A hyperparameter tuning technique used in the study is grid search. Grid search is a tuning technique performed on specific parameters so it can save time and resources in tuning. Hyperparameters which are tuned by grid search of each model are shown in Table 1.

Table 1 Hyperparameter tuning of each model

Model	Hyperparameters	
DNN	Number of hidden layers	Learning rate

	Number of hidden nodes Activation functions Optimizer	Momentum Batch size
CNN	Number of convolutional layers Filter Kernel Number of max pooling layers Pool size	Activation functions Number of hidden layers Number of hidden nodes Optimizer Batch size
MR	Feature selection Min tolerance	Ridge
SVM	Kernel type Kernel cache	Convergence epsilon
GB	Number of trees Maximum depth	Min split improvement
DT	Maximal depth Confidence	Minimal leaf size Minimal size for split
RF	Number of trees Maximal depth	Voting strategy

3.2.9 Performance evaluation

The performance of the prediction models in the study is evaluated using accuracy obtained from the confusion matrix which is popular in classification problems. Values relevant to the confusion matrix are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) where TP means the prediction is positive and it is true, FP means the prediction is positive and it is false, TN means the prediction is negative and it is true, and FN means the prediction is negative and it is false. The accuracy is calculated using (4).

$$Accuracy = \frac{TP + TN}{Total} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

3.3 Unsupervised learning

Unsupervised learning is used when output is not labeled or there are no defined labels. It is used to discover the pattern of data. This technique is widely used to know the insight in data. In this study, k-means clustering and association rules were used.

3.3.1 K-means clustering

K-means clustering was used to group data into a specific number of clusters. The clustering was done using the distance between each data point and centroid of each cluster where data belongs to the cluster with the nearest centroid. The number of clusters was defined according to the suitability of the data. Centroids were then randomly initiated. After that, data was assigned to clusters according to the nearest distance of each centroid using (5) to calculate distance and (6) to determine the cluster which data belonged to where c were centroid and k was the number of clusters. After data was assigned, new centroids were calculated using data belonging to each cluster using (7). Next, all data was assigned to clusters according to new centroids again and these processes were repeated until there were no changes in centroids and members of each cluster [17].

$$dist(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5}$$

$$c_j = \left\{x: \min_k dist^2(x, c_k) = j\right\} \tag{6}$$

$$c_j = \frac{1}{m_j}\sum_{x \in c_j} x \tag{7}$$

In this study, the radius of curvature and gradient were used as features for k-means clustering. These two features were extracted from the track profile data noted in the data description. Defects were the same dataset used in supervised learning techniques.

3.3.2 Association rules

Association rules were used to investigate the relationship between data in the same set. As mentioned in the data description, the study used the 50m range as a section of interest so any defects found in the range of 50 m were classified into the same set. Several parameters were used to explain the relationships between data. In this study, three parameters were used, namely, support, confidence, and lift.

Support indicates how often the relationships are found. It is calculated using (8) where x and y are defects found in a section and N is the number of sections.

$$Support = \frac{frq(x,y)}{N} \tag{8}$$

Confidence indicates how often the rules are true and can be applied to the relationships. It can be calculated using (9).

$$Confidence = \frac{frq(x,y)}{frq(x)} \tag{9}$$

Lift indicates the increase of the probability of finding y when x is found compared to the probability of finding y when the finding of x is not known. If the lift is higher than one, it can be interpreted that defects x and y are dependent and tend to occur together. If the lift is lower than one, it can be interpreted that defects x and y are not often found together or negatively correlated. If the lift is close to one, it can be inferred that defects x and y are independent. In terms of data, features used for association rules are different from supervised learning techniques and k-means clustering because only defects are used. Defect data is processed as transactions or baskets. In the study, a basket could be considered a 50 m-section track containing different defects. The number of defects could be different in each section. Moreover, defects used for association rules were not grouped according to the types of track components but 81 types of defect were used to present the relationships between particular defects instead of grouped defects.

4. Results and discussion

Several sets of data were processed and combined to create the dataset for supervised and unsupervised learning. There were 172,436 samples in the dataset. Seventy percent of data was used as training data and another 30% was used as testing data. Supervised learning was used to develop predictive models for detecting track component defects using track geometry measurement. Thirty-seven features were used to predict labels which consisted of five classes in supervised learning. Unsupervised learning was used to discover the insights in data. Two features, namely, the radius of curvature and gradient, were used for k-means clustering. Baskets of defects were extracted from raw data to explore the relationships of defects using association rules. The results of the study are shown in the following sections.

4.1 Prediction of railway track component defects using track geometry

Different supervised learning techniques are used to develop models. Every model is tuned using the grid search as mentioned in the previous section to ensure that models are structured properly and provide the best performance. From hyperparameter tuning using grid search, the optimal hyperparameters of each model are shown in Table 2.

Table 2 Tuned hyperparameter from grid search

Model	hyperparameter	Tuned value
DNN	Number of hidden layers	2
	Number of hidden nodes	750 (dense1), 1,750 (dense2)
	Activation functions	ReLu
	Optimizer	Adam
	Learning rate	0.001
	Momentum	0.8
	Batch size	128
CNN	Number of convolutional layers	2
	Filter	128 (conv1), 256 (conv2)
	Kernel	4 (conv1), 6 (conv2)
	Number of max pooling layers	2 (after each conv)
	Pool size	1
	Activation functions	ReLu
	Number of hidden layers	2
	Number of hidden nodes	400 (dense1), 200 (dense2)
	Optimizer	Adam
	Batch size	128
MR	Feature selection	M5 prime
	Min tolerance	0.05
	Ridge	1E-8
SVM	Kernel type	Dot
	Kernel cache	200
	Convergence epsilon	0.001
GB	Number of trees	50
	Maximum depth	5
	Min split improvement	1E-5
DT	Maximal depth	10
	Confidence	0.1
	Minimal leaf size	2
	Minimal size for split	4
RF	Number of trees	100
	Maximal depth	10

Model	hyperparameter	Tuned value
	Voting strategy	Confidence vote

The accuracy of each model is shown in Figure 3. From the figure, the accuracy of ANN is highest at 94.3% followed by CNN with an accuracy of 93.8%. However, the accuracy of other models is about 50% or lower while SVM has the worst performance. From the results, it can be inferred that the relationships between features and labels have high non-linear characteristics and are complex so ANN and CNN can perform better than other models. The MR model is clearly a linear model so the performance is not good. Surprisingly, DT performs better than RF and GB although they are all tree-based models and RF which benefits from voting and assembling concepts. For the SVM model, the machine creates a hyperplane to classify samples. However, the number of features used to develop models is 37 so it is complicated to create 37D hyperplane. Although ANN has the highest accuracy, its training time is not significantly higher than other models. The training time of all models is approximately one second/epoch except CNN where training time is 20 seconds/epoch. Therefore, it can be concluded that ANN is the best model for detecting track component defects using track geometry measurement in this study in terms of both performance and resources. For model application, every model can predict without significant difference in terms of time so there is no issue about the model application.

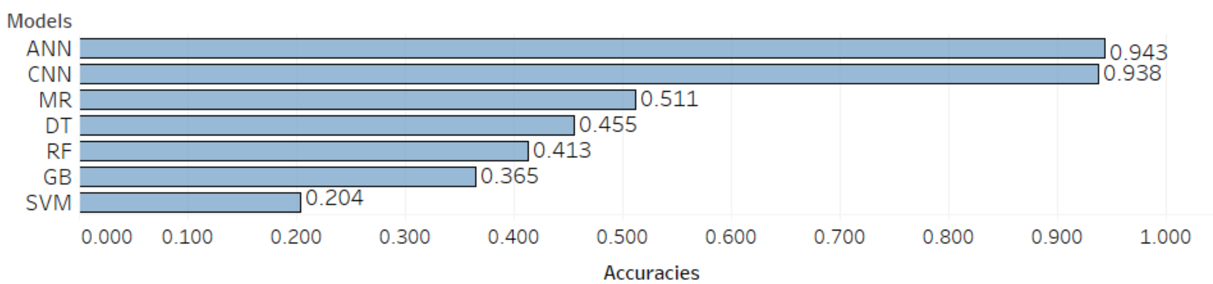

Figure 3 Accuracies of each model

In terms of sensitivity (recall) and specificity, the sensitivity and specificity of the ANN model are shown in the below table. The sensitivity and specificity show that the ANN model performs well in terms of accuracies, sensitivities, and specificities. It shows that the model can detect defects based on the track geometry measurement more than 88% based on the sensitivity. At the same time, the model can avoid incorrect detection more than 97% based on the specificity.

Table 3 Sensitivity and specificity of the ANN model based on classes

Class		Sensitivity	Specificity
Switch and crossing	0	0.88	1.00
Fastener	1	0.89	0.99
Rail joint	2	0.92	0.99
Rail	3	0.96	0.97
None (Defect-free)	4	0.94	0.97

From the results, supervised machine learning can be used to detect component defects using track geometry which saves a lot of time in carrying out the inspection and improves the safety of staff. Moreover, this approach does not require additional cost or installation to detect defects because the

TGC is operated regularly to inspect track geometry. Besides track geometry measurement obtained from the TGC operation, the measurement can also be used to detect defects. The process of data collection is relatively fast because the TGC can operate at speeds of up to 70 mph and does not disturb the normal operation of the track. However, defects predicted in the study are grouped according to track components, namely, rail, switch and crossing, fastener, and rail joint because it is found during the preliminary model development that if the number of classes is too high, the models cannot execute prediction accurately. This issue needs to be improved to give the models more potential to detect defects in detail or to categorize defects better.

4.2 Insight of railway track component defects

K-means clustering was used to group defects. In the study, 4 clusters were used based on the preliminary investigation. Features used for the clustering were the radius of curvature and gradient extracted from track profile data. Labels consisted of four classes, namely, switch and crossing, rail, fastener, and rail joint. From the training, a chart showing the clustering is shown in Figure 4. From the figure, clusters can be categorized using a tree shown in Figure 5 and centroids of each cluster are shown in the figure.

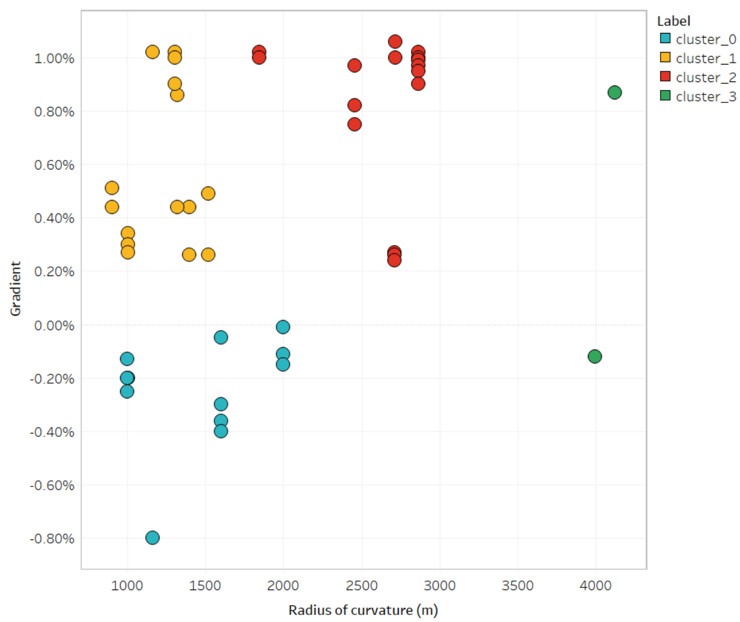

Figure 4 Result from k-means clustering

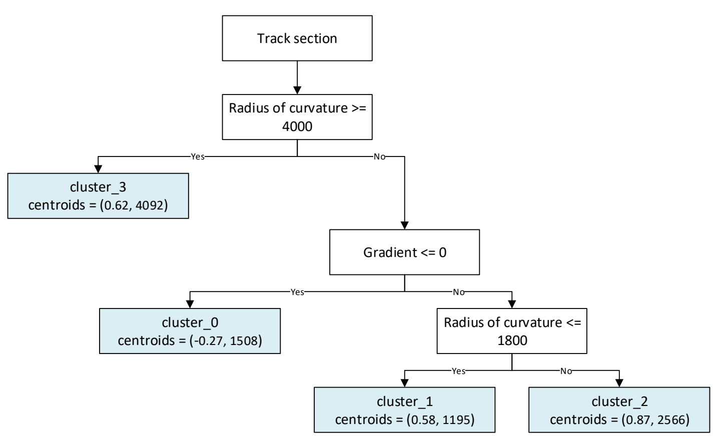

Figure 5 Tree for clustering and centroids

In Figures 4 and 5, tracks with a radius of curvature higher than 4,000 m and tangent tracks are not shown because defects are relatively rare compared to sharper curves. It is noted that the minimum radius of curvature that switch and crossing were installed in the route is 900 m. This might be resulted by the route terrain that tangent tracks were impossible to construct. It can be seen that the smaller radius of curvature results in a higher number of defects. When the radius of curvature is higher than 3,000 m, the number of defects significantly reduces. A positive gradient also results in more defects. From k-means clustering, the radius of curvature and gradient can be used to cluster defects as shown in Figures 4 and 5. To demonstrate the insights from k-means clustering more clearly, the cluster distribution based on defects and clusters is presented in Table 4.

Table 4 Cluster distribution based on component defects and clusters

Based on component defects				
	cluster_0	cluster_1	cluster_2	cluster_3
S&C	12%	73%	8%	8%
Fastener	53%	21%	26%	0%
Rail joint	39%	26%	32%	3%
Rail	38%	38%	17%	7%
Based on clusters				
	S&C	Fastener	Rail joint	Rail
cluster_0	2%	7%	34%	56%
cluster_1	15%	3%	24%	59%
cluster_2	3%	6%	48%	43%
cluster_3	10%	0%	20%	70%

From Figure 4 and Table 4, several insights are discovered. First, when the radius of curvature is between 2,000 and 4,000 m and the gradient is negative, no defect occurs and it can be inferred that tracks with this characteristic are not sensitive to defects. Most of the switch and crossing defects (73%) are in cluster 1 or tracks with the radius of curvature less than 1,800 m and the gradient is positive. It is clear that when the radius of curvature is small, it affects the occurrence of switch and crossing defects significantly. However, the majority of other defects are in cluster 0 where the gradient is negative and it can be inferred that rolling stock might apply brakes and result in track component defects. Cluster 3 has the smallest proportions for all component defects which shows that the radius of curvature affects defect occurrence significantly and there are no fastener defects when the radius of curvature is bigger than 4,000 m. Rail defects are the most common defects because their proportion is the biggest in every cluster except in cluster 2 where rail joint defects have a higher proportion. For cluster 3, rail defects' proportion is up to 70% which and it can be inferred that the bigger radius of curvature reduces the occurrence of every component defect except rail defects. Therefore, the defects which need careful attention when the track is tangent are the rail defects.

From association rules, there are 62 rules where the support is higher than 0.05. To present the insights from association rules, rules with confidence higher than 0.8 are presented in Table 5. In the table, support, confidence, and lift are presented. Support is the proportion of events found in all events. It demonstrates how frequent each event happens in all events. In the table, it can be seen that support might not high because the total number of observations is very high. Therefore, it is better to consider the confidence. Confidence is the proportion of events that left-hand side and right-hand side happen together compared to the number of events that left-hand side events happen. In other words, confidence shows how frequent that right-hand side events happens when left-hand side events happen. Last, lift is the rise of probability to have right-hand side events when left-hand side events are known compared to the case that the rules are not known. In other words, lift can help decision makers to decide better when they apply rules and the higher lift means the better chance to decide correctly. It is noted that every lift is higher than one because it shows the rise of probability in the unit of time.

Table 5 Examples of association rules with the confidence higher than 0.8

Left-hand side	Right-hand side	Support	Confidence	Lift
Failed ties in switch, Surface defect on frog	Rail surface defect	0.056	0.957	1.926
Worn wing rail, Failed ties in switch, Worn point blade	Rail surface defect	0.052	0.953	1.918
Failed ties in switch, Worn point blade	Rail surface defect	0.056	0.948	1.908
Worn wing rail, Worn point blade	Rail surface defect	0.059	0.927	1.867
Rail surface defect, Failed ties in switch, Worn point blade	Worn wing rail	0.052	0.927	6.204
Failed ties in switch, Worn point blade	Worn wing rail	0.054	0.922	6.171
Rail surface defect, Worn point blade	Worn wing rail	0.059	0.898	6.015
Rail surface defect, Worn wing rail, Worn point blade	Failed ties in switch	0.052	0.878	6.896
Failed ties in switch, Worn point blade	Rail surface defect, Worn wing rail	0.052	0.878	7.805
Worn wing rail, Worn point blade	Failed ties in switch	0.054	0.855	6.712
Rail surface defect, Worn point blade	Failed ties in switch	0.056	0.852	6.686
Worn wing rail, Failed ties in switch	Rail surface defect	0.073	0.841	1.694
Worn point blade	Rail surface defect	0.065	0.837	1.684
Surface defect on frog	Rail surface defect	0.074	0.823	1.657
Worn wing rail, Worn point blade	Rail surface defect, Failed ties in switch	0.052	0.815	7.922
Worn point blade	Worn wing rail	0.063	0.810	5.426
Failed ties in switch	Rail surface defect	0.103	0.807	1.625

From Table 5, there are 17 rules with confidence higher than 0.8. The rule with the highest confidence is when a track has failed ties in switch and surface defect on frogs, rail surface defects tends to occur with the confidence of 0.957. These rules can help inspectors to notice defects better based on the defects they found. If defects on the left-hand side occur, inspectors can be aware that defects on the right-hand side might occur depending on the confidence shown in Table 5.

5. Conclusion

This study combined several sets of data, namely, track geometry measurement data, track component defect data, and track profile data to create the dataset for supervised and unsupervised learning. Supervised learning was used to develop models to detect component defects using track geometry. Techniques used in the study are DNN, CNN, MR, SVM, GB, DT, and RF. Each model was tuned using a grid search to make sure that the performance of each model was optimal based on their potential. From the study, DNN had the best performance with an accuracy of 94%. The model can detect defects and categorize them into component levels, namely, rail, rail joint, switch and crossing, and fastener which is more detailed than any other published study. It is noted that there might be an imbalance issue between samples with defects because some defects tend to occur more frequent than other defects. Therefore, more data can improve the completeness of data and confidence in machine learning models.

The study applied unsupervised learning to explore insights of data. K-means clustering and association rules were applied. From k-means clustering, two features, which are the radius of curvature and gradient, were used. The number of clusters was four and each cluster well represented the characteristic of tracks and defects. An example of insight from k-means clustering was the occurrence of defects which was negatively correlated to the radius of curvature. For association rules, several rules were discovered and there were 17 rules wherein the confidence was higher than 0.8. Inspectors can apply these rules to beware of the occurrence of defects that may not detect during the inspection.

The contributions of the study are mentioned as previously. The challenge of future work is how to select features and develop models to better predict track component defects because this study found that the models do not perform well when the number of classes is too high. If models are developed until they can detect particular defects, it will be a significant benefit for the railway industry because it saves a lot of time and cost and also improves staff safety. More data from different years, locations, and routes can be included to add the variability of data to make models perform better. Unsupervised learning can also explore new data and provide new insights from varied data.

6. Acknowledgment

The authors also wish to thank the European Commission for the financial sponsorship of the H2020-RISE Project no.691135 "RISEN: Rail Infrastructure Systems Engineering Network", which enables a global research network that addresses the grand challenge of railway infrastructure resilience and advanced sensing in extreme environments (www.risen2rail.eu).

References

1. Li D, Meddah A, Hass K, et al. Relating track geometry to vehicle performance using neural network approach. *P I MECH ENG F-J RAI* 2006; 220: 273-281.
2. Martey EN, Ahmed L and Attoh-Okine N. Track geometry big data analysis: A machine learning approach. In: *2017 IEEE International Conference on Big Data (Big Data)* 2017, pp.3800-3809. IEEE.
3. Ye J, Stewart E, Zhang D, et al. Method for automatic railway track surface defect classification and evaluation using a laser-based 3D model. *IET Image Process* 2020; 14: 2701-2710.
4. Molodova M, Li Z and Dollevoet R. Axle box acceleration: Measurement and simulation for detection of short track defects. *Wear* 2011; 271: 349-356.
5. Galvan Nunez SA. Hybrid Bayesian-Wiener process in track geometry degradation analysis. 2017.
6. Ye J, Stewart E and Roberts C. Use of a 3D model to improve the performance of laser-based railway track inspection. *P I MECH ENG F-J RAI* 2019; 233: 337-355.
7. Liu S, Wang Q, Luo YJTS, et al. A review of applications of visual inspection technology based on image processing in the railway industry. *Transportation Safety and Environment* 2019; 1: 185-204.
8. Zheng S, Chai X, An X, et al. Railway track gauge inspection method based on computer vision. In: *2012 IEEE International Conference on Mechatronics and Automation* 2012, pp.1292-1296. IEEE.
9. Tsunashima H. Condition monitoring of railway tracks from car-body vibration using a machine learning technique. *Appl Sci* 2019; 9: 2734.

10. Lasisi A and Attoh-Okine N. Principal components analysis and track quality index: A machine learning approach. *Transp Res Part C Emerg Technol* 2018; 91: 230-248. DOI: https://doi.org/10.1016/j.trc.2018.04.001.
11. Sadeghi J and Askarinejad H. An investigation into the effects of track structural conditions on railway track geometry deviations. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 2009; 223: 415-425.
12. Zarembski AM, Einbinder D, Attoh-Okine NJC, et al. Using multiple adaptive regression to address the impact of track geometry on development of rail defects. 2016; 127: 546-555.
13. Soleimanmeigouni I, Ahmadi A, Nissen A, et al. Prediction of railway track geometry defects: a case study. *Structure and Infrastructure Engineering* 2020; 16: 987-1001.
14. Mohammadi R, He Q, Ghofrani F, et al. Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects. *Transp Res Part C Emerg Technol* 2019; 102: 153-172. DOI: https://doi.org/10.1016/j.trc.2019.03.004.
15. Sharma S, Cui Y, He Q, et al. Data-driven optimization of railway maintenance for track geometry. *Transp Res Part C Emerg Technol* 2018; 90: 34-58. DOI: https://doi.org/10.1016/j.trc.2018.02.019.
16. Attoh-Okine NO. *Big data and differential privacy: analysis strategies for railway track engineering*. John Wiley & Sons, 2017.
17. Zaiontz C. K-means Cluster Analysis, https://www.real-statistics.com/multivariate-statistics/cluster-analysis/k-means-cluster-analysis/ (2015).