Report No. RailTEAM UD-12	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Random Forest-Based Covariate Shift in Addressing Non-Stationarity of Railway Track Data		5. Report Date October 6, 2023 6. Performing Organization Code:
7. Author(s) Ibrahim Balogun and Nii O. Attoh-Okine https://orcid.org/0000-0001-5328-5538		8. Performing Organization Report No. UD-12
9. Performing Organization Name and Address Department of Civil & Environmental Engineering University of Delaware 301 DuPont Hall Newark, DE 19716		10. Work Unit No. 11. Contract or Grant No. 69A3551747132
12. Sponsoring Agency Name and Address Office of Research, Development and Technology (RD&T) US Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

For years, track geometry vehicles have been deployed to capture rail defects. However, the limitation associated with the operation is the possibility of non-stationarity of the observed measurements due to external influence. The effect of non-stationarity may lead to the false representation of track conditions and thereby increases the likelihood of false output. For that reason, we considered the possibilities of supervised machine learning techniques for detecting and correcting the track geometry inherent anomalies. The methods include Random Forest (R.F.), Logistic Regression (L.R.), and Support Vector Machine (SVM). To ascertain the discrepancies within the data, we varied the train-test and validation ratio in phases. Conclusively, the developed models' application indicates that the Random Forest is a more practical approach to detecting the non-stationarity of track geometry data. Also, it optimizes the cost of maintenance and supports accurate decision making to improve track safety better.

17. Key Words Non-stationarity of observed measurements of track conditions, supervised machine learning technique	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. http://www.ntis.gov		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 23	22. Price

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

USDOT Tier 1
University Transportation Center
on Improving Rail Transportation
Infrastructure Sustainability and Durability

Final Report UD-12

RANDOM FOREST-BASED COVARIATE SHIFT IN DDRESSING NON-STATIONARITY OF RAILWAY TRACK DATA

By

Ibrahim Balogun
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Delaware
iobalo@udel.edu

and

Nii O. Attoh-Okine, Ph.D., P.E., F.ASCE
Interim Academic Director
UD Cybersecurity Initiative
Professor, Civil and Environmental Engineering
Professor, Electrical and Computer Engineering
University of Delaware
okine@udel.edu

Date: October 6, 2023

Grant Number: 69A3551747132

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

EXECUTIVE SUMMARY

Globally, the railroad industry represents an integral part of any nation through supportive economic recovery and long-term sustainable growth. In the United States, the railway infrastructure accounts for approximately 40% of intercity freight and is worth over $80 billion. Today, many people depend on the railroad for different reasons such as safety, an alternative to traffic congestion and carbon emission. However, substantial efforts need to be expended to position the railway as the lead transportation infrastructure continuously.

On the flip side, no infrastructure ever exits without its associated risks. However, the sustainability of such systems depends on the swift actions of the maintenance agency. A common railroad problem that's has received significant attention amongst rail experts is track geometry defects. The unavailability of requisite techniques to manage the issues often strains on infrastructure safety amidst other pressured concerns. Intuitively, to keep the assets viable, the maintenance officer must capture possible faults that could impact the existing infrastructure's safety, reliability, and operations in real-time.

For years, track geometry vehicles have been deployed to capture rail defects. However, the limitation associated with the operation is the possibility of non-stationarity of the observed measurements due to external influence. The effect of non-stationarity may lead to the false representation of track conditions and thereby increases the likelihood of false output. For that reason, we considered the possibilities of supervised machine learning techniques for detecting and correcting the track geometry inherent anomalies. The methods include Random Forest (R.F.), Logistic Regression (L.R.), and Support Vector Machine (SVM). To ascertain the discrepancies within the data, we varied the train-test and validation ratio in phases.

Conclusively, the developed models' application indicates that the Random Forest is a more practical approach to detecting the non-stationarity of track geometry data. Also, it optimizes the cost of maintenance and supports accurate decision making to improve track safety better.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1.1 Background & Motivation

In recent years, railway infrastructure functioned as a panacea for pressured conventional transportation systems. Railroads have been recognized and promoted as an effective means of mass transit that significantly reduces road traffic, and the resulting emissions (Falamarzi et al. 2019). In general, a US Class I railroad provides a safe, efficient, and cost-effective transportation network that con- veniently pivots the nation's economy (Association of American Railroads 2020). Additionally, climate change, which is considered a global challenge, also contributed to railway transportation rel- evance. Intuitively, its importance could be attributed to reducing fossil fuels that could deplete the sustainable atmospheric organic compound.

Since issues that could affect the rail infrastructure systems must be prevented, track maintenance and accident-related problems have attracted attention amongst the rail experts. Periodic track measurement is required to evaluate track system status -track quality (Pwayblog 2021). Figure 1 shows the track geometry parameters that are measured to ascertain the condition of a rail track. The track lines are described with some deterministic parameters defined in Table 1. Some of the parameters are alignment and gauge in the horizontal and longitudinal planes, respectively, and twist and cross-level in the vertical directions. The maintenance engineer is bound to exercise three fundamental tasks with the railroads, i.e., safety, comfort, and economy. A typical track quality index of a track system is shown in Figure 2. Whilst track geometry recovery actions will improve the track condition; it's almost impossible to rejuvenate the geometry condition to an as-good-as-new state (Sancho et al. 2021). It is essential to adopt a reliable deem algorithm that could dampen the track geometry degradation. According to Andrade and Teixeira (2015), Bayesian statistical models provide a flexible framework to combine prior information from past samples or expert prediction with the new data. Although several factors come into play as the track sections exhibit different types of degradation behaviour over time, the proper understanding of the track geometry parameters needed to avoid possible derailment leads to a careful understanding of track quality indices. Asides from the fact that the parameters must be ascertained before any construction or maintenance of rail track, they are also used as indicators to evaluate the track geometry's performance. Considering the geometry overview, the interoperability of these components is inarguably worth studying.

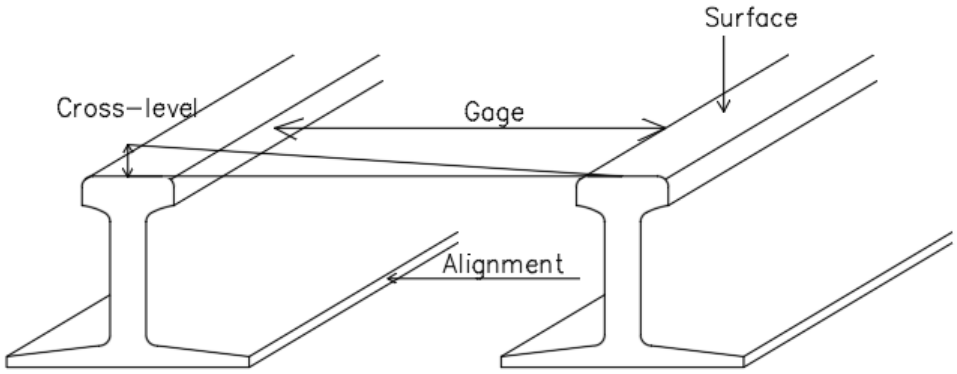

Figure 1. Track Geometry Parameters

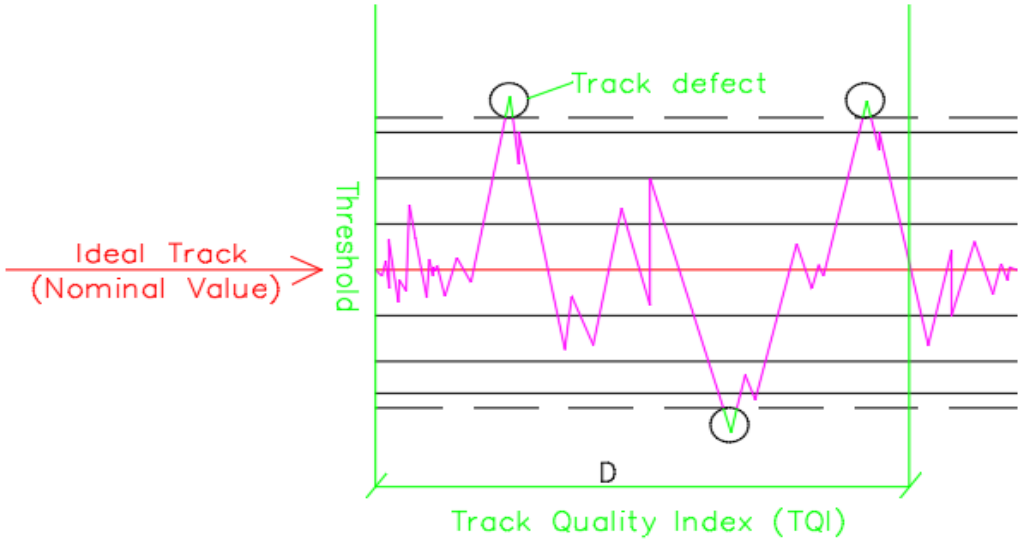

Figure 2. Track Quality Index of a Railway System

1.2 Statement of the Problem

Track geo-defects and track are vital prerequisites for safe railroad operations. Investigations revealed that even when all track geo-defects parameters are observed to conform within the threshold's standards, track accident still occurs. Due to the versatility and complexity of data structure, track managers decide the maintenance scheme that best describes track conditions. Techniques that have existed are the statistical and stochastics method. Grace and Nii-Okine, (2020) considered approximate Bayesian computation techniques to estimate the track quality indices for track degradation. It is interesting to see that ABC, as it is called, estimates the track system's irregularities without computing the likelihood functions, which are computationally expensive (Ashley and Attoh-Okine 2020).

In the past, both statistical and stochastic methods are helpful. However, the recent size of track data has forced analysts to machine learning methods. Lasisi and Okine (2018) apply the Principal Component Analysis (PCA) to determine the track quality index of multi-dimensional geometry data. The study shows the possibility of transforming huge track geometry data into low sample space without compromising the geometry information (Lasisi and Okine 2018).

It is ironic to see that with the level of accuracy of most machine learning techniques, minor shortcomings are still identified. Therefore, no free lunch theorem will always exist. One of the shortcomings of machine learning methods is the low performance with high dimensional data, stream data, and covariate data. The reason is that the techniques are train and test on the same data. However, when they are exposed to new data, their adaptability diminishes. The investigation into the poor performance of the techniques due to the data handling is called data-shift.

Data shift problem has been explained in many fields with real-life problems except railroad. Whenever a shift occurs in data distribution, the accuracy of the solving techniques is usually affected. In railway engineering, considering a track data of 30 years, it will be difficult to isolate the likelihood of train accident for a season since rail track response to seasonal temperature. Thus, track data exploration for possible deterioration should be done with a consideration of external influence.

1.3 Objective of the Study

This research aims to address the non-stationarity of track geometry data using machine learning technique-random forest-based covariate shift. The output of the study is dependent on the following sub-objectives:

- To explore the threshold of all observed measurements using FRA standards (see Table 1).
- To identify non-stationary parameters using the data shift framework (see Figure 3).
- To apply random forest-based covariate shift techniques on the non-stationarity parameters.
- To infuse the re-oriented data in the machine learning algorithm (train, test, and validation) for accurate predictions

Table 1. Track Geometry Parameter Definition

S/N	Parameter	Definition
1	Gauge	The distance between the inner face of two adjacent rails. The standard gauge in North America is 1435mm.
2	Cross-Level	This is the measurement of the difference in elevation between the top surface of the two rails at any point of the track.

3	Alignment	Also known as straightness of the track. It is the projection of the track geometry of each rail or track centerline unto the horizontal plane.

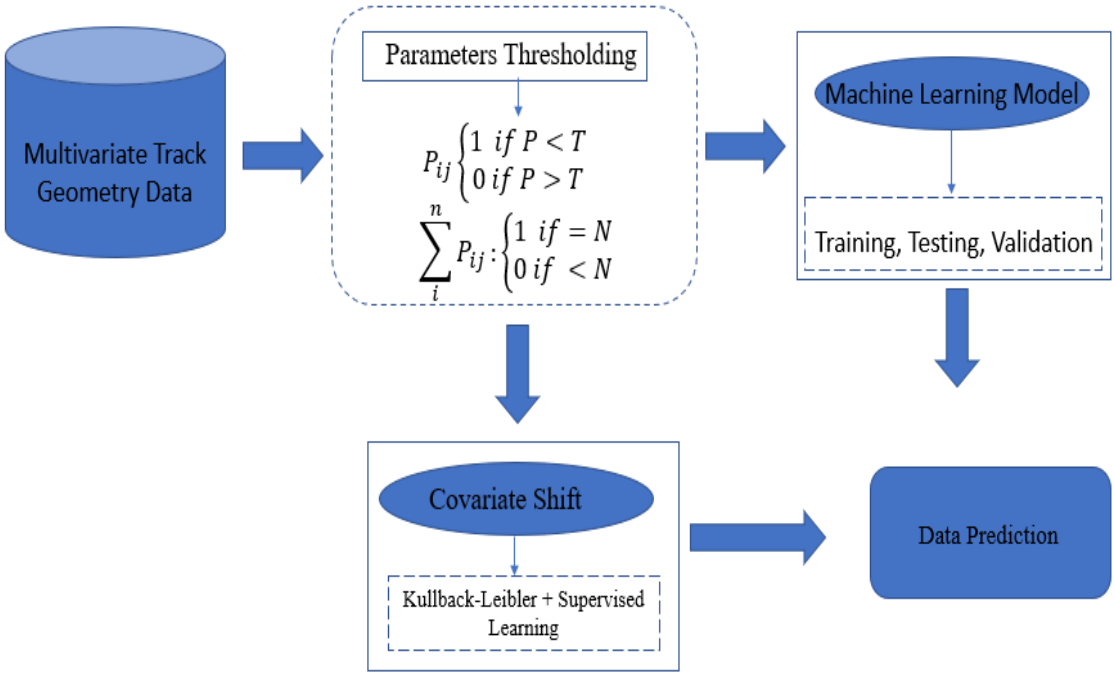

Figure 3. Covariate Shift Framework

NON-STATIONARITY OF TRACK GEOMETRY

There has been a long discussion over the acceptable representation of defects on the rail systems. Although a relationship exists between these defects, research has shown that geometry defects increase dynamic wheel load (fatigue), leading to increased pressure on the rail (Zarembski et al. 2016). These stresses, bending and contact stress, will over time cap to rail defects. Researchers consider studying the non-stationarity of track geometry conditions to better understand the track's dynamic and provide sustainable maintenance planning. The excessive train speed over the initially designed track speed will automatically distort the track geometry, resulting in maintenance related interruptions. Uneasy track geometry can cause the degradation of track components and rolling stock, freight damage, passengers' discomfort, and in extreme cases, derailment. However, a common problem associated with non-stationary is the track data imbalance. For instance, using the traditional method of maintenance, it is possible to mismeasure defects. As such, there would be a possibility of misclassification in the geometry data (Faghih-Roohi et al. 2016). A small patch on the rail may grow into a large squat, thereby increasing the track's amplitude vibration, delivering poor ridership, and ultimately resulting in a derailment in a long time. To better minimize these effects, the track stationarity of track data needs to be

investigated. The question begging for an answer is, “how do we detect the shift in the data, if it ever existed, in a distribution?”

2.1 Covariate Shift

The field of machine learning has been leveraged to account for anomalies in data structures. See Table 2. These anomalies stem from how data are harnessed or collected over a specified period. Many researchers have expressed the concept of data shift in their areas of specialization, using technical terminologies such as “concept shift”, “concept drift”, “changes of classification”, “changing environment”, “contract mining in classification learning”, “fracture points, and fracture between data”. In recent times, McGaughey et al. (2016) have introduced the data shift concept to illustrate the variations of the train test distributions. Since the real problem arises, the covariate shift concept has gained ground as machine learning techniques are forced to process non-independently identically distributed (non-iid) data (Shimodaira 2000).

Table 2. Machine Learning Applications in Railway Track Engineering

Authors	Objective	Machine learning technique.
Jamshidi et al. (2017)	Risk assessment evaluation of rail surface defect using image data.	Deep convolution neural; networks (DCNN)
Heidarysafa et al. (2019)	Rail accident report interpretation for a non-expert reader.	Convolution neural networks (CNN), Recurrent neural networks (RNN), Deep neural networks (DNN), Word to vector (Word2Vec)
Song et al. (2019)	Fasteners detection based on neural networks	Faster R-Convolution neural networks.
Chenariyan Nakhaee et al. (2019)	Application of machine learning in rail track maintenance.	Review of existing applicable machine learning, shortcomings, and solutions.
Qi et al. (2020)	Real-time detection of track fasteners with deep network architecture.	YOLOv3-Tiny

Research has shown how a machine learning-based approach can detect a covariate shift in a feature. In a compact mathematical form, the condition for the covariate shift can be expressed as follows:

$$P_{train}(X) \neq P_{test}(X) \tag{1}$$

$$P_{train}(Y|X) = P_{test}(Y|X) \quad (2)$$

In the above equations, X represents the geometry defects (gauge, alignment, cross-level, surface) for both the training and testing. In a distribution sample, it is expected that the train and test samples should be drawn from the same distribution and independently identically distributed (iid). The distribution plot in Figures 4a & 4b clearly shows the disparity or drift in the gauge dataset even when it is reshuffled. Researchers often make the mistake of predicting unobserved data with a performing algorithm that has been built from a different distribution. This misconception will give a false result and can be catastrophic in some senses, depending on the analysis's application. One way to alleviate the covariate shift problem is by reweighting the log-likelihood terms according to their importance. Another sophisticated approach is by measuring the densities of the training and testing data before the importance is estimated.

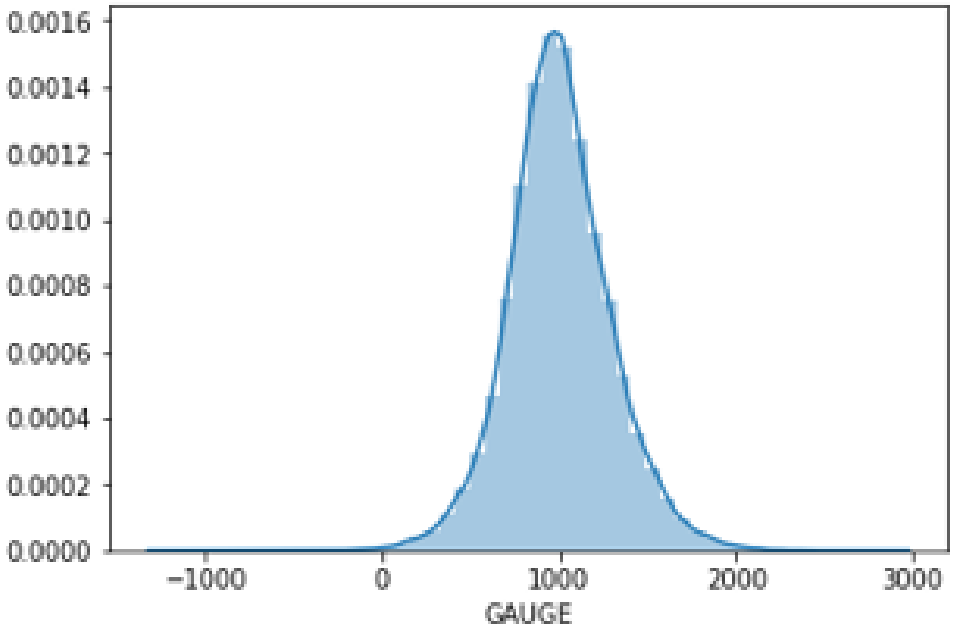

Figure 4a. A gauge plot from training geometry data

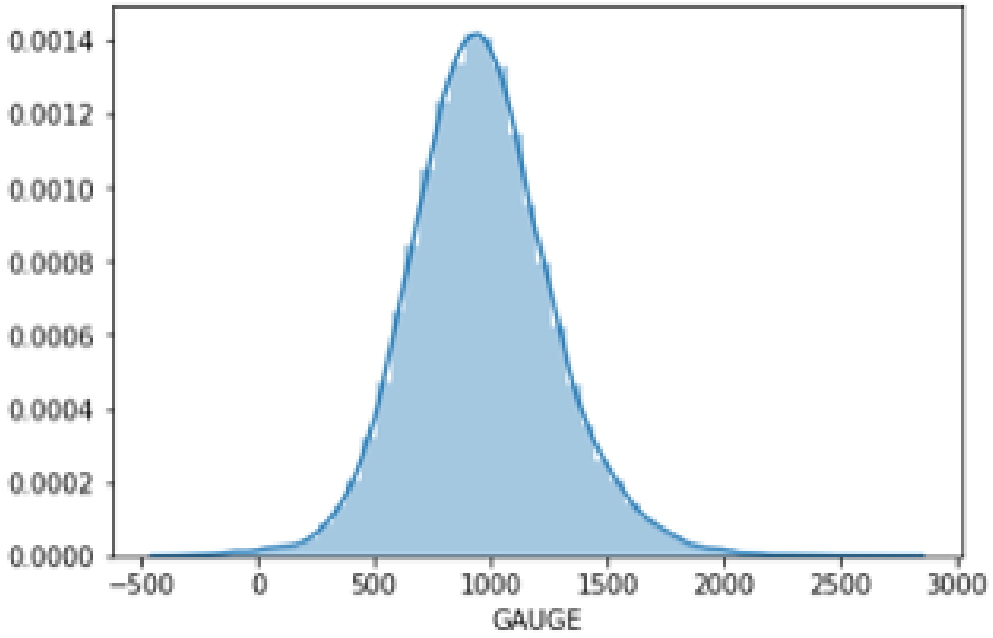

Figure 4b. A gauge plot from testing geometry data

In the space domain D of iid for n features of a real number R, the $P_{tr}(x) > 0$ for all x subset of D. Estimating the densities of the training or testing set, i.e., $P_{tr}(x)$ or $P_{tst}(x)$, should be avoided when estimating the weight importance:

$$w(x) = \frac{P_{tst}(x)}{P_{tr}(x)} \tag{3}$$

where the $P_{tr}(x)$ is the probability of experiencing a feature x in the training set while $P_{tst}(x)$ represent the probability of experiencing a feature x in testing. Intuitively, one will realize that if a sample has a low chance effect in the prediction, it is important to reweight it to affect the prediction sample considerably. From equation (3) above, we can then say that the probability $P_{ts}(x)$, which is analogous to the test input data, can be expressed as follows:

$$P_{ts}(x) = w(x) * P_{tr}(x) \tag{4}$$

The divergence between the test distribution and the weighted predicted test can be expressed by the following:

$$\text{K.L.}\left(P_{tets}(x)\middle|\widehat{P_{tr}}(\text{x})\right) = \int \widehat{P_{tr}}(\text{x}) \log\left(\frac{P_{tets}(x)}{P_{tr}(x)\hat{w}(\text{x})}\right)\text{dx} \tag{5}$$

The importance estimate $w(x)$ is formulated based on the Kullback-Leibler Importance Estimation Procedure (KLIEP). This method was proposed by Sugiyama (2007), and it is formulated from the Kullback-Leibler divergent theorem to minimize the divergence of the training and prediction dataset:

$$\widehat{w}(\mathrm{x}) = \sum_{i=1}^{b} \alpha_i * \varphi_i(x) \quad (6)$$

In the above equation, the α_i represents the weight to be learned, while φ_i represents the basis function. In this process, we realize that it is possible to remove some terms which are independent of $\widehat{w}(\mathrm{x})$. The resulting problem will be a convex optimization problem. We then incorporate this to get the covariate shifted version of the classifier:

$$\left[\sum_{j=1}^{n_p} \log\left(\sum_{i=1}^{b} \alpha_i * \varphi_i(x_j)\right)\right] \quad (7)$$

Constraint

$$\sum_{j=1}^{n_p} \sum_{i=1}^{b} \alpha_i * \varphi_i(x_j) = 1 \quad (8)$$

$$\text{and } \alpha_i, \alpha_i, \ldots\ldots\ldots \alpha_i \geq 0 \quad (9)$$

EXPLORATORY DATA ANALYSIS

3.1 Introduction

This research used the geometry data from a U.S. Class IV railroad's routine maintenance for one year (2018-2019). The Track Geometry Vehicle (TGV), which operates at 180 km/h, focused on individual segments of the track. The track geometry safety thresholds for Track Class IV can be seen in Table 3. The parameters measured cover over 20 features, including gauge, unloaded gauge, alignment, surface, and twist for the examined track segments. Some of the parameters are defined in Table 4.

Table 3. Track Geometry Safety Thresholds for Track Class IV

Track Class IV Thresholds (mm) 62ft Chord				
Superelevation	Cross-level	Alignment	Twist	Profile
85	32	32	44	51

Table 4. Some Selected Parameters from the Railroad Data.

S/N	Parameter	Definition
1	Alignment_31_L	Left 31(ft) wavelength of alignment
2	Alignment_31_R	Right 31(ft) wavelength of alignment
3	Alignment_62_R	Right 62(ft) wavelength of alignment
4	Alignment_62_L	Left 62 (ft) wavelength of alignment
5	Surf_31_L	Left 31(ft) wavelength of the surface
6	Surf_31_R	Right 31(ft) wavelength of the surface
7	Surf_62_L	Left 62 (ft) wavelength of the surface
8	Surf_62_R	Right 62(ft) wavelength of the surface
9	Cant_L	Left Cant
10	Cant_R	Right Cant

3.2 Data Description

The exploratory data analysis of the geometry data exposes the underlying relationships between components. In Figure 5, it is evident that a strong correlation exists between a different wavelength of the same parameter on the same side of the rail (Surf_31_L and Surf_62_L) and (Alignment_31_L and Alignment_62_L). Careful consideration also unearths the hidden relationship between the alignment and the surface for the same wavelength. One may conclude that the correlation results from the longitudinal axis; however, some parameters, such as cant and gauge, also exhibit weak correlations. It is also interesting to see that the superelevation shows no corresponding relationship with any geometry data. In general, the importance of exploratory data analysis has been proven to unravel and further derive meaning from hidden data.

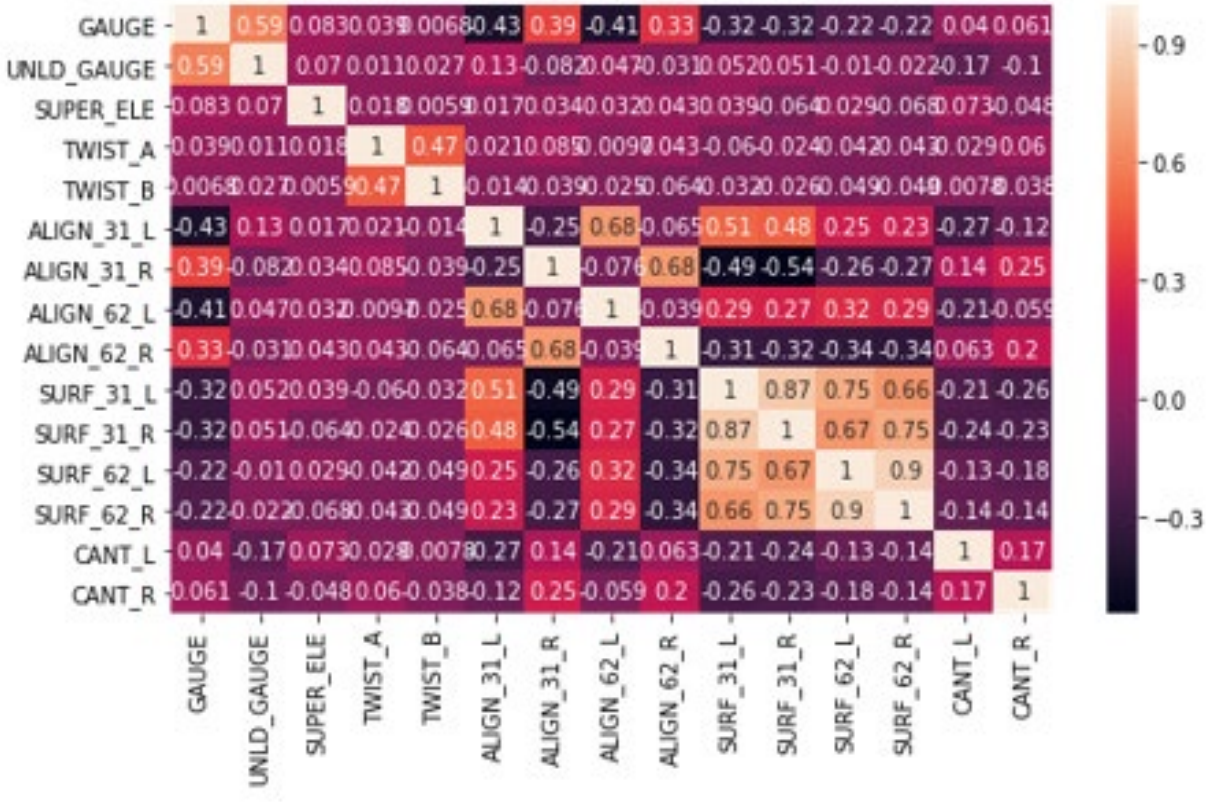

Figure 5. Geometry Track Parameter Correlation Plot.

3.3 Validation Study

In order to ascertain the results of the classifiers, we validated the algorithm with a class I railroad track data. The data consists of one-year track geometry defects measurements for seasonality comparison of track support conditions. Although the data is relatively small, it was relevant for the analysis. We detected a shift (drift) in the track parameters due to the track usage's seasonal variation. Figures 4(a&b) show the bimodal distribution of the gauge parameter. We further established the influence of right and left track parameters on accuracy. We grouped the data into

three sections. In the first section, we considered all the parameters, including right and left components, while in the second section, we considered the only left parameters. In the third section, we considered only the right parameters. The results are shown in Tables (5-8), respectively. The results demonstrate that it is important to consider the position of the parameters when making a maintenance prediction. The accuracy of the classifier confirmed this through the distribution phases.

Table 5. Machine Learning Training Results for Phase Distributions: All components

Predictor	**Accuracy (%)**		
	First Distribution	**Second Distribution**	**Third Distribution**
Logistic Regression	96.76	96.92	97.10
SVM	97.10	96.96	97.00
Random Forest	97.10	97.00	97.11

Table 6. Machine Learning Validation Results for Phase Distributions: All components

Predictor	**Accuracy (%)**		
	First Distribution	**Second Distribution**	**Third Distribution**
Logistic Regression	98.75	96.82	97.81
SVM	98.80	98.74	98.81
Random Forest	98.10	98.92	99.04

Table 7. Machine Learning Validation Results for Phase Distributions: Right component

Predictor	**Accuracy (%)**		
	First Distribution	**Second Distribution**	**Third Distribution**
Logistic Regression	90.36	93.92	95.33
SVM	87.62	89.96	87.40
Random Forest	89.10	80.67	86.22

Table 8. Machine Learning Validation Results for Phase Distributions: Left component

Predictor	**Accuracy (%)**		
	First Distribution	**Second Distribution**	**Third Distribution**
Logistic Regression	91.25	91.33	90.89
SVM	89.61	90.65	91.60
Random Forest	84.91	83.67	88.71

Results and Discussion

Clearly, the advantages of considering a covariate shift in track analysis cannot be downplayed in any engineering application with traces of data irregularities. We realized that various parameters have significantly different behaviours in different track sections, which strongly depend on the track curvature. The discrepancies in the data distribution are detected using the Kullback-Leibler divergent criterion. The training and testing sets are analyzed with supervised machine learning for three training/testing distribution phases. In the first phase, the SVM and the Random Forest performed better than the logistic regression. In the second and the third phase of the distribution, the Random forest performed better than the other learners. It then tells that the Random forest worked best in adjusting the training set to match the testing set and reduce the divergence. Similarly, the ROC graph in Figure 6c shows that the random forest best predicted the defects compared to all the other learners, while in the validation analysis, the Random Forest does not work well with a single component. The result shows that the supervised learning methods gave a better accuracy when all the components were considered compared to a single component.

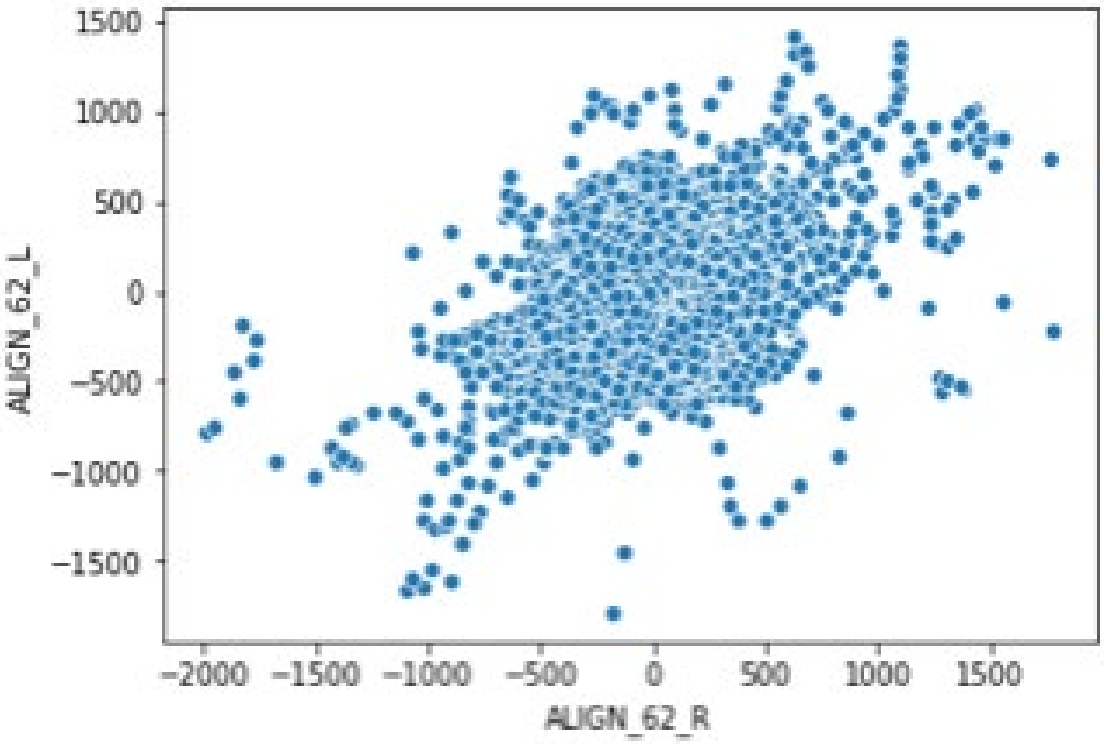

Figure 6a Alignment 62ft left chord scatterplot

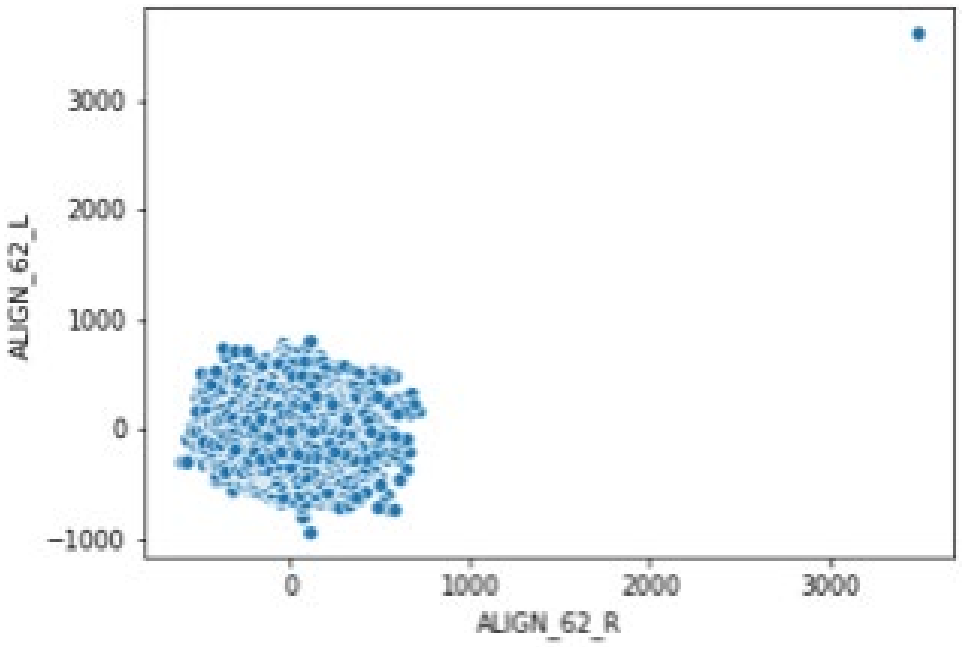

Figure 6b Alignment 62ft right chord scatterplot

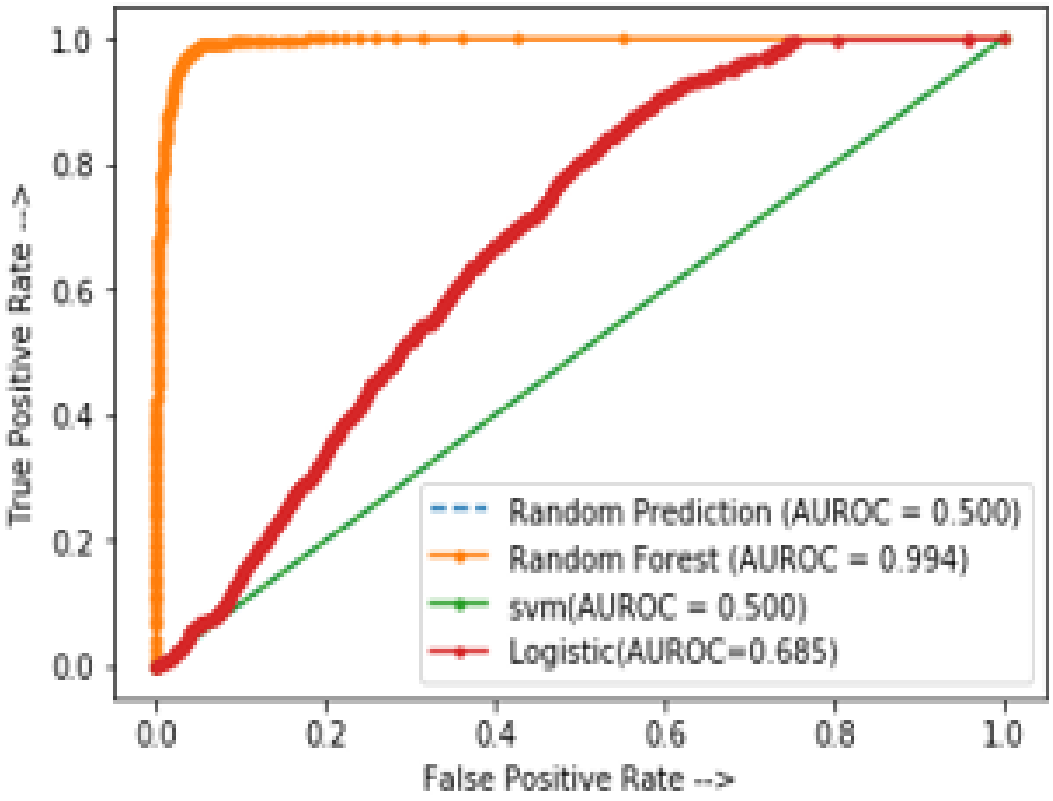

Figure 6c ROC curve of improved predictors

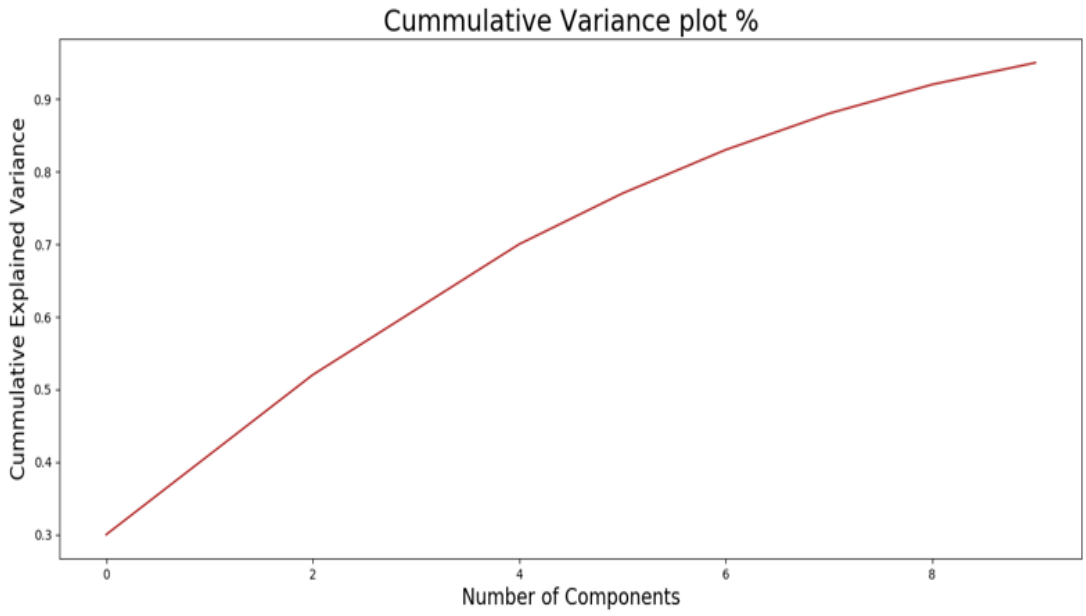

Figure 6d. The graph of percentage variance explained

CONCLUDING REMARKS

Conclusions

Track quality assessment depends on the data obtained from automated inspections. However, a grey area in the chain of the process is data sparsity. Many literary works have outlined the use of machine learning, but they emphasize data orientations. This research attempted to develop a method to establish the data shift in the track geometry data and further process it with supervised machine learning and the Kullback-Leibler divergent approach. The study revealed a divergence, and the less weighted dataset was then reweighted with a minimized error. A data of similar structure was used to validate the model, considering the components holistically.

The findings show that the Random forest-Based Covariate Shift can best handle the imbalance of track geometry data.

Summary/Future Research

In summary, the models formulated in this research can correct possible underlying track geometry discrepancies that could impede TQI predictions. In subsequent research, we hope to explore dimension reduction techniques (TSNE) as a pipeline technique for covariate shift framework. This is because the ML techniques are expected to correct the space collinearity of observed measurements. Additionally, the results will be validated by comparing them with other existing methods.

REFERENCES

1. Falamarzi, A., S. Moridpour, and M. Nazem. A Review of Rail Track Degradation Prediction Models. *Australian Journal of Civil Engineering*, Vol. 17, No. 2, 2019, pp. 152–166. https://doi.org/10.1080/14488353.2019.1667710.
2. Association of American Railroads. Railroad 101. October.
3. Pwayblog. Evaluation of the Track Quality. https://pwayblog.com/2016/09/11/evaluation-of-track-quality/#TQI. Accessed Jan. 11, 2021.
4. Sancho, L. C. B., J. A. P. Braga, and A. R. Andrade. Optimizing Maintenance Decision in Rails: A Markov Decision Process Approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, Vol. 7, No. 1, 2021, p. 04020051. https://doi.org/10.1061/ajrua6.0001101.
5. Andrade, A. R., and P. F. Teixeira. Statistical Modelling of Railway Track Geometry Degradation Using Hierarchical Bayesian Models. *Reliability Engineering and System Safety*, Vol. 142, 2015, pp. 169–183. https://doi.org/10.1016/j.ress.2015.05.009.
6. Wang, B. Z., C. P. L. Barkan, and M. Rapik Saat. Quantitative Analysis of Changes in Freight Train Derailment Causes and Rates. *Journal of Transportation Engineering, Part A: Systems*, Vol. 146, No. 11, 2020, p. 04020127. https://doi.org/10.1061/jtepbs.0000453.
7. Ashley, G., and N. Attoh-Okine. Approximate Bayesian Computation for Railway Track Geometry Parameter Estimation. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2020, p. 095440972097772.

https://doi.org/10.1177/0954409720977726.
8. Lasisi, A., and N. Attoh-Okine. Principal Components Analysis and Track Quality Index: A Machine Learning Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 91, No. January, 2018, pp. 230–248. https://doi.org/10.1016/j.trc.2018.04.001.
9. Hajizadeh, S., A. Núñez, and D. M. J. Tax. Semi-Supervised Rail Defect Detection from Imbalanced Image Data. *IFAC-PapersOnLine*, Vol. 49, No. 3, 2016, pp. 78–83. https://doi.org/10.1016/j.ifacol.2016.07.014.
10. Zarembski, A. M., N. Attoh-Okine, D. Einbinder, H. Thompson, and T. Sussman. How Track Geometry Defects Affect the Development of Rail Defects. 2016.
11. Faghih-Roohi, S., S. Hajizadeh, A. Núñez, R. Babuska, B. De Schutter, S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter. Vancou-Ver, Canada. Vol. 19, 2016, pp. 2584–2589.
12. Shimodaira, H. Improving Predictive Inference under Covariate Shift by Weighting the Log-Likelihood Function. *Journal of Statistical Planning and Inference*, Vol. 90, No. 2, 2000, pp. 227–244. https://doi.org/10.1016/s0378-3758(00)00115-4.

ACKNOWLEDGEMENT

This study was conducted with the support from the USDOT Tier 1 University Transportation Center on Railroad Sustainability and Durability.

ABOUT THE AUTHORS

Nii O. Attoh-Okine, Ph.D., P.E., F.ASCE

Nii O. Attoh-Okine, Professor of Civil and Environmental Engineering, and Electrical and Computer Engineering. He is also the Interim Academic Director of the University of Delaware Cybersecurity Initiative. In the last couple of years, he has authored two books which are defining the direction of research across disciplines: a) Resilience Engineering: Models and Analysis and b) Big Data and Differential Privacy in Railway Track Engineering. He is a founding associate editor for ASCE/ASME Journal of Risk and Uncertainty Analysis. He has served as an Associate Editor on the four ASCE Journals. Attoh-Okine is currently a member of a group of researchers from the United States and Japan working on Smart Cities and various cyber issues related to the Tokyo 2020 Olympic Games.

Ibrahim Balogun

Ibrahim Balogun is a PhD Graduate Research Assistant in the Civil and Environmental Engineering Department at the University of Delaware. He obtained his master's degree in Civil and Environmental Engineering from the University of Lagos. He is a reviewer for Advances in Data Science and Adaptive Analysis and ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering. He is presently a member of Prof. Attoh-Okine's research group.