Proceedings of the 8th European Conference of the Prognostics and Health Management Society 2024 - ISBN – 978-1-936263-40-0

False alarm reduction in railway track quality inspections using machine learning.

Isidro Durazo-Cardenas[1], Bernadin Namoano[2], Andrew Starr[3], Ram Dilip Sala[4], Jichao Lai[5]

[1,2,3,4,5] *Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom*

i.s.durazocardenas@cranfield.ac.uk
bernadin.namoano@cranfield.ac.uk
a.starr@cranfield.ac.uk
salaramdilip@gmail.com
damonlai1996@gmail.com

Abstract

Track quality geometry measurements are crucial for the railways' timely maintenance. Regular measurements prevent train delays, passenger discomfort and incidents. However, current fault diagnosis or parameter deviation relies on simple threshold comparison of multiple laser scanners, linear variable differential transformer (LVDT) and camera measurements. Data threshold exceedances enact maintenance actions automatically. However, issues such as measurement error, and sensor failure can result in false positives. Broad localisation resolution prevents trending/ inferencing by comparison with healthy data baseline at the same position over periodic inspections.

False alarms can result in costly ineffective interventions, are hazardous and impact the network availability.

This paper proposes a novel methodology based on convolutional neural network (CNN) technique for detecting and classifying track geometry fault severity automatically. The proposed methodology comprises an automatic flow of data for quality assessment whereby outliers, missing values and misalignment are detected, restored and where appropriate curated. Improved, "clean" datasets were then analysed using a pretrained CNN model. The method was compared with a suite of machine learning algorithms for diagnosis including k-nearest neighbour, support vector machines (SVM), and random forest (RF).

The analysis results of a real track geometry dataset showed that track quality parameters including twist, cant, gauge, and alignment could be effectively diagnosed with an accuracy rate of 97.80% (CNN model). This result represents a remarkable improvement of 38% in comparison with the traditional threshold-based diagnosis. The benefits of this research are not only associated with maintenance intervention cost savings. It also helps prevent unnecessary train speed restrictions arising from misdiagnosis.

1. Background

Rail transportation's convenience, punctuality and cost-effectiveness have made it the preferred mode for medium distance travelers and freight (Ghofrani et al., 2018; Wang et al., 2018) Train services as well as the total mileage of track is increasing, which poses a considerable challenge for the effective maintenance of railways infrastructure (Durazo-Cardenas et al., 2018). Degraded rail tracks can cause bumps and swaying when trains pass at high speed and can even cause derailments, putting at risk the safety of passengers. In the event of a failure, delays to the network can also cause significant economic losses (Sasidharan et al., 2020).

Wear and degradation are inevitable, and the railways have implemented safe, tolerance limits for track quality parameters (Railtrack PLC, 1998). Today, tracks are regularly inspected and repaired by dedicated infrastructure maintenance teams. This usually implies a combination of sophisticated track quality inspection trains and on-foot crews that validate and repair the defects flagged by the inspection trains. However, this requires experienced technicians working in hazardous environments, while reducing the availability of the network. Clearly, false alarms raised by the inspection trains contribute to further downtime and costs.

1.1. Measurements and data parameters

The New Measurement Train is an automatic inspection train that is currently the primary method of collecting track geometry data on the British Railways (*New Measurement Train (NMT)*, 2024) . It uses multiple laser scanners, linear variable differential transformer (LVDT), gyroscopes and accelerometers and cameras Data from thirteen-time domain track parameters are acquired, plus additional imagery systems, accumulating terabytes of data. The NMT can measure track condition at 125mph and cover up to 115,000 miles in a year.

Cross-level, Gauge and Curvature are the essential features of a track alignment, transverse and vertical deviation. Cross-level is defined as the vertical height difference between the tops of the two tracks, while the distance between the two set of rails is known as the Gauge. Curvature is the radius of the arc of the rail, which describes the degree of curvature of the rail. For straight rails, the desired cross-level value approaches zero. while for curved rails, this closely matches the design value.

The *Twist* parameter combines deviations in vertical, and longitudinal dimensions and is typically measured at 3m intervals (Twist3m). The *Cant* parameter describes the difference between the track cross-level and the design cross-level value on curved track.

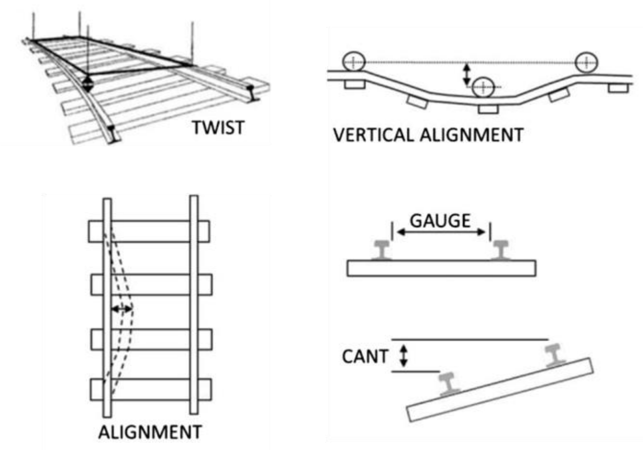


Figure 1 rail track quality parameters. Adapted from D’Angelo et al., (2018).

Top and *Alignment* (AL) account for the deviation between the actual track and the optimal planned path, where top is the vertical distance deviation of the top surface. AL refers to the horizontal distance deviation (Railtrack PLC, 1998). On the other hand, *Dip* is a measure of the depression of the track.

The standard deviation of the measured data parameters is compared to their threshold values for each 1/8 of mile. Exceedances are logged during the train inspection and corrective action notice are issued. Based on the severity of the faults detected and the nominal speed of the line, the health of the track section will be classified as Good, Satisfactory, Poor, Very Poor, and Super Red. Speed restrictions are then issued considering the parameter criticality, with 20 miles per hour being the lowest speed restriction, before track blockage. Network Rail standards(NR/L2/TRK/001/MOD11, 2015 also prescribe the actions to be taken upon threshold exceedances for each parameter, with these ranging from:

1. Block the Line
2. Correct before 36h.
3. Inspect in 72h and correct before 14 to 28 days.
4. Correct before 7 to 14 days.
5. Correct before 14 to 28 days.
6. Add it to the maintenance plan.

1.2. Dataset description

The data used in this this study comprises time series measurement data of the thirteen track parameters described above covering the Southampton-Waterloo line in both directions over a period of one year. This is considered a major line serving many commuter areas including southwestern suburbs of London and the conurbations based around Southampton. Datasets typically comprise CSV acquisition and PDF maintenance team activity logs, and track defect reports. Network Rail reports track quality assessments every 1/8 of mile, with up to 1000 measurements for each parameter acquired. Datasets typically exceed 2 GB.

1.3. Machine learning and related work

Machine learning is often used to analyze large amounts of data and identify connections, offering exceptional potential for anomaly detection analysis (Popov et al., 2022). Recent studies report on machine vision and SVM used to analyze images for track defect detection (Aydin et al., 2021). However, the settings could be significantly costly as it requires high specification tools (cameras, effective fast transmission systems, and efficient storage). Moreover, the large number of images generated bring real time processing challenges hence, affecting on time performance.

Based on time series data, reported accuracy of some traditional machine learning algorithms appears to be relatively low. Considering the disruption, cost and effort involved in railways repairs, higher accuracy is essential. For example, results of SVM algorithm used to detect combined track degradation from car body vibrations reported an accuracy of 80% (Tsunashima, 2019). Lasisi & Attoh-Okine, (2018),used Principal component analysis (PCA) to combine track geometry parameters into a lower-dimensional form and then used SVM, Linear discriminant analysis (LDA), and Random Forest (RF) to detect orbital faults with an accuracy of 92%. However, the true positive rate (precision) is only about 66%, potentially leading to many false alarms.

Several studies have compared the performance of machine learning methods. Sresakoolchai & Kaewunruen, (2019) used a range of supervised and unsupervised machine learning models to analyse track geometry data and sentence faults. The results showed that the non-linear models fitted significantly better, with deep neural network (DNN) having the highest accuracy at 94.3%, followed by convolution neural network (CNN) with 93.8%. The linear models all had accuracy rates below 50%, with SVM being the poorest at 20%. The results showed that the relationship between features and labels is highly non-linear.

In this context, we propose and effective method to automatically detect and classify track defect using data driven approaches. The next section introduces the methodology.

2. METHODOLOGY

The five-step approach employed in this investigation is illustrated in **Error! Reference source not found..** The initial data acquisition step is associated with the acquisition of data. Two different types of data were gathered for the analysis. The first datasets comprise of time series signals representing key track quality parameter measurements such as, the location, time, the CANT, the TOP, the gauge, the AL, the Cross-level and the Curvature. The second dataset represents a set of pdf files containing a threshold-based detection report and human based investigation maintenance logs. These files are useful to annotate the track fault observed in the time series data.

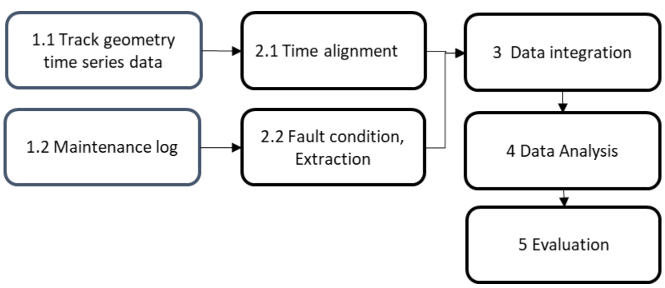


Figure 2 Flow chart of methodology steps.

Step two focuses on the enhancement of individual datasets quality. This was critical step given the issues often encountered with time series data. These datasets typically emanate from a variety of instrumentation sources and are prone to a myriad of consistency issues, including measurement discrepancies caused by instrument malfunctions or user errors. These issues manifest as missing data points, misaligned signals, and a significant presence of noise, each of which can distort the true signal and lead to inaccurate analyses if not properly addressed. To address the absence of data, a localized regression method is implemented. This approach leverages nearby data points to estimate and impute the missing values, assuming that these points observe a similar behavioral pattern. Such assumption is justified given that time series data often exhibits temporal correlation. The efficacy of the imputation process is vital, as it directly affects subsequent analyses. In addition to imputation, the dataset processing phase employs a combination of Dynamic Time Warping (DTW) and Cross Correlation techniques to detect and correct misalignments in the signals. DTW is an algorithm that allows for elastic transformation of time series, enabling the identification of similarities between data sequences that do not align perfectly in time. When used in conjunction with Cross Correlation, it becomes a powerful tool to detect shifts and distortions in the signal, thereby aligning them appropriately for further analysis. Further refining of the dataset includes processing of the maintenance log reports which undergo a procedure to extract and quantify salient features, such as fault type, spatial and temporal coordinates of the occurrences, severity of the detected faults, and the associated maintenance activities required. This information is crucial for understanding the context of the faults and planning preventive measures.

The third stage integrates the outputs of the previous stages, creating a consolidated dataset that is primed for machine learning analytics. This stage is pivotal as it synthesizes the cleaned and aligned time series data with the qualitative information extracted from the maintenance logs, setting the foundation for robust analytical models.

The fourth stage is the heart of the analysis, where two primary categories of machine learning techniques are employed: supervised and unsupervised learning methods. Unsupervised learning techniques, such as the Density-Based Spatial Clustering of Applications with Noise (DBSCAN), are adept at identifying novel fault types that may not have been previously recognized. DBSCAN is privileged for its proficiency in handling noise within the dataset, a common issue in large-scale industrial applications. However, while unsupervised methods excel at detection, they often falter in classification. To counter this, supervised learning methods are applied, harnessing the labelled data produced by unsupervised techniques to train models that can not only detect but also classify fault types. This dual approach ensures that newly occurring faults are not only detected but also categorized correctly. The supervised techniques selected for this stage include robust and widely used algorithms such as Convolutional Neural Networks (CNNs), which are particularly adept at spatial data recognition; k-Nearest Neighbors (kNN), which classifies data based on the proximity to known cases; Random Forests (RF), an ensemble method that improves prediction robustness; and SVMs, which are effective in high-dimensional spaces.

Finally, the last stage focuses on the evaluation of the model's performance, employing three key metrics: precision, recall, and the F1-score. Precision assesses the model's accuracy in predicting fault occurrences, mitigating the risk of false positives. Recall measures the model’s ability to identify all actual fault occurrences, thereby reducing false negatives. The F1-score harmonizes these two metrics, providing a single measure of the model’s accuracy in classification tasks, balancing the trade-off between precision and recall. This comprehensive and iterative process is essential for the identification and classification of faults in complex systems, such as those encountered in Network Rail's infrastructure. By meticulously processing the data and employing a blend of machine learning techniques, the methodology aims to yield a high-performing model capable of detecting and classifying faults, thus enhancing the maintenance and reliability of the rail network.

3. Results and Discussion

3.1. Dataset analysis

The datasets used in this study come from Network Rail and pertains to track geometry data acquired between 2016 and 2017 during inspections run along the Southampton railway line. These datasets consist of 625 pdfs reports, 300 multivariate time series data which includes track geometry measurements, maintenance team activity logs, and track defect reports. The temporal scope extends over 33 days, with an average monthly coverage of three days. The location parameter, although not a track geometry feature, is a crucial piece of information in the dataset because it provides a baseline for comparing data from different measurement time at the same point. The positioning errors in the recording could reach 100 m. The instrument failure can also cause some positions to be recorded as missing points, resulting in various positions for the same serial number in the data sets. Missing data at a position t is imputed by using and interpolation of different data points observed at a location [t-w] and [t+w], where w (set to 5) helps including of neighbouring data points to enhance the imputation accuracy. For the alignment, the data points must first be aligned so that they are of the same length between data and that data points of the same ordinal number are in the same position. To perform the alignment, first DTW method was used to compute pairwise distance between the measurement, and close distance signal were used to compute a reference signal. Hence maximum value between the cross correlation and the reference signal were used to shift the measurement. **Error! Reference source not found.** and **Error! Reference source not found.** show an example before and after the alignment was performed using the Twist3m measurement.

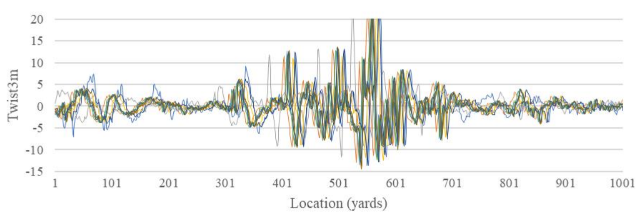


Figure 3 Twist3m measurement before alignment.

After the temporal data alignment, the maintenance log files (shown in **Error! Reference source not found.**) are processed to extract the track identifier (trackid), the mileage, the type of fault (shown in the column "channel"), the peak value observed and the corresponding threshold value. These data are used to locate the fault in the temporal data and hence annotate it. We segmented by file instead of by time series index as the initial experiment on time index split provided imbalanced issues.

The final annotated 300 multivariate time series datasets with about 20000 datapoints each are split into training (60%), validation (20%) and testing (20%).

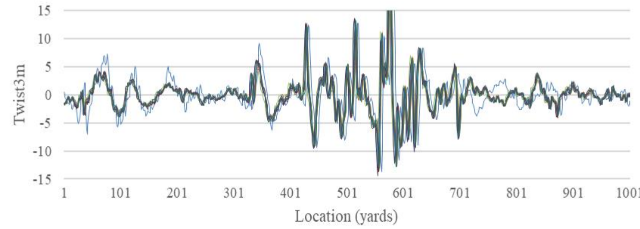
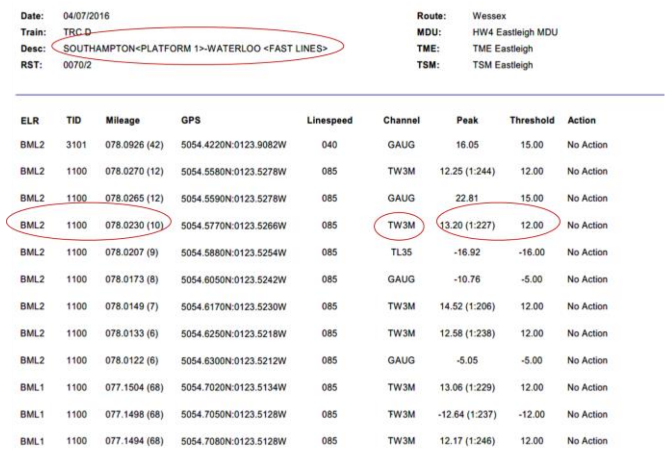


Figure 4 Twist3m after alignment.



Date: 04/07/2016
Train: TRC D
Desc: SOUTHAMPTON<PLATFORM 1>-WATERLOO <FAST LINES>
RST: 0070/2

Route: Wessex
MDU: HW4 Eastleigh MDU
TME: TME Eastleigh
TSM: TSM Eastleigh

ELR	TID	Mileage	GPS	Linespeed	Channel	Peak	Threshold	Action
BML2	3101	078.0926 (42)	5054.4220N:0123.9082W	040	GAUG	16.05	15.00	No Action
BML2	1100	078.0270 (12)	5054.5580N:0123.5278W	085	TW3M	12.25 (1:244)	12.00	No Action
BML2	1100	078.0265 (12)	5054.5590N:0123.5278W	085	GAUG	22.81	15.00	No Action
BML2	1100	078.0230 (10)	5054.5770N:0123.5266W	085	TW3M	13.20 (1:227)	12.00	No Action
BML2	1100	078.0207 (9)	5054.5880N:0123.5254W	085	TL35	-16.92	-16.00	No Action
BML2	1100	078.0173 (8)	5054.6050N:0123.5242W	085	GAUG	-10.76	-5.00	No Action
BML2	1100	078.0149 (7)	5054.6170N:0123.5230W	085	TW3M	14.52 (1:206)	12.00	No Action
BML2	1100	078.0133 (6)	5054.6250N:0123.5218W	085	TW3M	12.58 (1:238)	12.00	No Action
BML2	1100	078.0122 (6)	5054.6300N:0123.5212W	085	GAUG	-5.05	-5.00	No Action
BML1	1100	077.1504 (68)	5054.7020N:0123.5134W	085	TW3M	13.06 (1:229)	12.00	No Action
BML1	1100	077.1498 (68)	5054.7050N:0123.5128W	085	TW3M	-12.64 (1:237)	-12.00	No Action
BML1	1100	077.1494 (68)	5054.7080N:0123.5128W	085	TW3M	12.17 (1:246)	12.00	No Action

Figure 5 Example pdf report.

3.2. Algorithms parameters

The parameters summarised in **Error! Reference source not found.** were empirically tested and configured for this analysis.

For Convolutional Neural Networks (CNN), the setup includes Conv2D, MaxPooling2D, and Dense layers with ReLU and Softmax activations, optimised using adaptive moment estimation (ADAM) with a learning rate of 0.001.

The RF is configured with 100 trees, no maximum depth to allow full growth, a minimum of 2 samples required to split a node and uses the Gini criterion for quality of splits. DBSCAN was set with an epsilon value of 0.5 for maximum neighbourhood distance, and a minimum of 5 samples for core points, while employing Euclidean distance for its metric. DBSCAN being an unsupervised method cannot map automatically with classes, hence we mapped manually the cluster with a script that computes the cluster centre and the known fault centre.

The k-NN algorithm uses 5 neighbours, uniform weights, automatically selects the algorithm for computing, and utilizes the Minkowski metric. Lastly, the Support Vector Machine (SVM) is configured with an RBF kernel, regularization parameter C set to 1.0, 'scale' for gamma, and a polynomial degree of 3, optimizing for non-linear data separation. Each configuration reflects the algorithm's focus, from spatial clustering and decision forests to similarity-based learning and hyperplane optimization, illustrating the adaptability and specificity required for effective machine learning applications.

Table 1 Methods and parameter settings.

Method	Parameter Configuration Example
CNN	Layers: Conv2D, MaxPooling2D, Dense; Activation: ReLU, Softmax; Optimizer: Adam; Learning Rate: 0.001
RF	Number of Trees: 100; Max Depth: None; Min Samples Split: 2; Criterion: Gini
DBSCAN	Epsilon: 0.5; Min Samples: 5; Metric: Euclidean
kNN	Number of Neighbours: 5; Weights: Uniform; Algorithm: Auto; Metric: Minkowski
SVM	Kernel: RBF; C: 1.0; Gamma: Scale; Degree: 3

3.3. Analysis

The evaluation of the employed method on the testing datasets through precision, recall, and f1-score metrics offers a detailed perspective on their performance in predictive modelling tasks. With a multiclass classification problem, average performance results are computed. As highlighted in **Error! Reference source not found.**, CNN showcased a well-rounded performance with a precision of 97.8%, a recall of 97.69%, and an f1-score of 97.73%, indicating a high degree of accuracy and reliability in identifying relevant instances. RF also demonstrated a strong balance between precision (93.6%) and recall (95.21%), culminating in an f1-score of 94.4%, which underscores its effectiveness in handling various data scenarios. DBSCAN, with a precision of 100%, indicates a perfect identification of relevant instances within its clusters, though its lower recall (88.60%) suggests some relevant instances may not be captured within its clusters, reflected in an f1-score of 93.95%. kNN and SVM both achieved high precision rates (97.6% and 100%, respectively) but with slightly lower recall rates (90.37% and 87.95%, respectively), leading to f1-scores of 93.84% and 93.58%, highlighting their precision in classification but at the expense of some sensitivity.

Table 2 Method performance on the datasets.

Method	Precision (%)	Recall (%)	f1-score (%)
Threshold	59.8	53.80	56.64
CNN	97.8	97.69	97.73
RF	93.6	95.21	94.4
DBSCAN	100	88.60	93.95
kNN	97.6	90.37	93.84
SVM	100	87.95	93.58

Figure 6 CNN confusion matrix.

Focusing on CNN, **Error! Reference source not found.**7 provides a confusion matrix table displaying the performance of a classification model on a testing set, summarizing how well the model distinguishes between five fault classes: Gauge Fault, Twist Fault, AL Fault, Cant, and Top. The matrix shows actual class labels on the vertical axis (TARGET) and predicted labels on the horizontal axis (OUTPUT), with each cell containing the count and percentage of instances. Diagonal cells (in green) represent correctly classified instances, while off-diagonal cells (in red) indicate misclassifications. The overall performance is impressive, with the model correctly classifying 97.80% of the instances (2445 out of 2500) and misclassifying only 2.20% (55 out of 2500). Notably, "AL Fault" has the highest accuracy (99.00% correct), while "Top" has the highest misclassification rate (4.00%), indicating a specific area for potential improvement. The table also includes sum totals for

Class Name	Precision	1-Precision	Recall	1-Recall	f1-score
Gauge Fault	0.98	0.02	0.98	0.02	0.98
Twist Fault	0.98	0.02	0.95	0.05	0.97
AL fault	0.99	0.01	0.98	0.02	0.98
Cant	0.98	0.02	0.99	0.01	0.99
Top	0.96	0.04	0.99	0.01	0.97
Accuracy	0.98				
Misclassification Rate	0.02				
Macro-F1	0.98				
Weighted-F1	0.98				

each class, providing a comprehensive overview of the model's classification capabilities. Figure 8 showing detailed performance tabulation was obtained by computing the relevant performance indicators from Figure 7.

Figure 7 CNN detailed performance.

Discussion of these results illustrates the inherent trade-offs between precision and recall metrics across different algorithms. CNN and RF, with their balanced precision and recall, are suited for applications where both false positives and false negatives carry significant consequences, providing a robust option for complex classification tasks. The perfect precision of DBSCAN and SVM suggests their utility in scenarios where the cost of false positives is high, making them ideal for applications requiring high confidence in the prediction of positive instances. However, their lower recall rates indicate a potential shortfall in identifying all actual positive instances, which could limit their application in scenarios where missing any positive instance carries a higher risk. kNN, while slightly less precise than SVM or DBSCAN, offers a good compromise between precision and recall, making it a versatile choice for many practical applications. These results underscore the importance of choosing the right algorithm based on the specific requirements and constraints of the task at hand, considering the balance between identifying relevant instances accurately while minimising false identifications. Although these algorithms have various degree of success, they still overperform traditional thresholds technique which precision is 59.8%.

Comparing the machine learning model performance from the literature with the metrics provided reveals a notable advancement in precision, recall, and F1-score. While the literature highlights variances in SVM accuracy from 20% to 80% across applications, this implementation showcases a substantial leap approaching 100% precision for SVM, underscoring a highly effective application or different context. Similarly, The CNN and kNN models not only surpass some of the literature's DNN and CNN benchmarks with CNN achieving a near parity with the highest reported accuracy of 94.3% (Sresakoolchai & Kaewunruen, 2022) but with superior precision and recall. The inclusion of DBSCAN in this analysis, demonstrating a 100% precision, further highlights the potential of selecting and tuning models to suit specific data characteristics and problem contexts. This synthesis underscores the importance of advanced model fine-tuning, the choice of metrics for performance evaluation, and the adaptability of machine learning algorithms to achieve higher efficacy in complex, non-linear problem spaces, especially in critical applications like fault detection in railway systems. In terms of training time, Figure 6 shows the models average estimated time (in blue) and their standard deviation denoted as Std (orange). Random Forest (RF) and Convolutional Neural Network (CNN) methods demonstrate the shortest training times, approximately 1000 and 1200 seconds respectively, with minimal variability. These observations highlight that while SVM is computationally intensive, RF and CNN are more efficient, making them suitable choices when computational resources or time are limited.

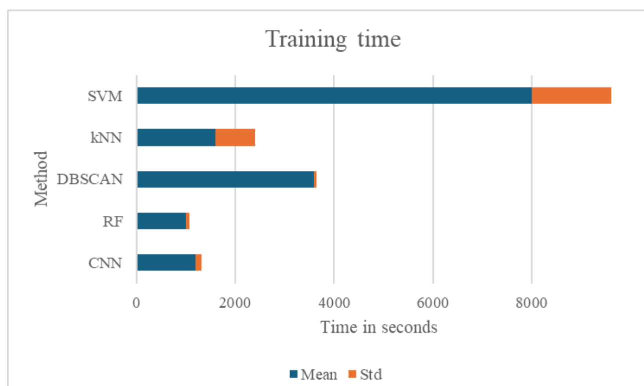


Figure 8 Average Model training time.

4. Conclusion

This paper presents a machine learning methodology that successfully improves false alarm rate of railway track quality inspections by 38%. While the datasets examined in this paper only pertain to one specific route, the methods presented here are applicable to all other railway lines across Britain, since the same NMT inspection vehicle is used.

In the railways, repair interventions are costly, typically requiring a manual confirmation of the fault severity, parts, labour, travel, service disruptions (denial), and penalty fees, as well as being hazardous for the on-foot personnel involved. The ability to correctly diagnose faults also ensures unnecessary speed restrictions are removed, improving journey times and passenger comfort. Evidently, the proposed methodology can potentially have considerable financial impact.

Future work includes an analysis of the potential cost savings achieved using this methodology as well as the integration of context knowledge in the diagnostics.

REFERENCES

Aydin, I., Akin, E., & Karakose, M. (2021). Defect classification based on deep features for railway tracks in sustainable transportation. *Applied Soft Computing*, *111*, 107706. https://doi.org/10.1016/j.asoc.2021.107706

D'Angelo, G., Bressi, S., Giunta, M., Lo Presti, D., & Thom, N. (2018). Novel performance-based technique for predicting maintenance strategy of bitumen stabilised ballast. *Construction and Building Materials*, *161*, 1–8. https://doi.org/10.1016/j.conbuildmat.2017.11.115

Durazo-Cardenas, I., Starr, A., Turner, C. J., Tiwari, A., Kirkwood, L., Bevilacqua, M., Tsourdos, A., Shehab, & Emmanouilidis, C. (2018). An autonomous system for maintenance scheduling data-rich complex infrastructure: Fusing the railways' condition, planning and cost. *Transportation Research Part C: Emerging Technologies*, *89*, 234–253. https://doi.org/10.1016/j.trc.2018.02.010

Ghofrani, F., He, Q., Goverde, R. M. P., & Liu, X. (2018). *Recent applications of big data analytics in railway transportation systems: A survey ☆*. https://doi.org/10.1016/j.trc.2018.03.010

Lasisi, A., & Attoh-Okine, N. (2018). *Principal components analysis and track quality index: A machine learning approach*. https://doi.org/10.1016/j.trc.2018.04.001

New Measurement Train (NMT). (2024). https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/our-fleet-machines-and-vehicles/new-measurement-train-nmt/

NR/L2/TRK/001/MOD11. (2015). *Inspection & Maintenance of Permanent Way: Track Geometry - Inspections and Minimum Actions*. Network Rail.

Popov, K., De Bold, R., Chai, H. K., Forde, M. C., Ho, C. L., Hyslip, J. P., & Hsu, S. (2022). Big-data driven assessment of railway track and maintenance efficiency using Artificial Neural Networks. *Construction and Building Materials*, *349*, 128786. https://doi.org/10.1016/J.CONBUILDMAT.2022.128786

Railtrack PLC. (1998). Section 8: Track Geometry. In *Track Standards Manual*.

Sasidharan, M., Burrow, M. P. N., & Ghataora, G. S. (2020). *A whole life cycle approach under uncertainty for economically justifiable ballasted railway track maintenance*. https://doi.org/10.1016/j.retrec.2020.100815

Sresakoolchai, J., & Kaewunruen, S. (2022). Railway defect detection based on track geometry using supervised and unsupervised machine learning. *Structural Health Monitoring*, *21*(4), 1757–1767. https://doi.org/10.1177/14759217211044492

Tsunashima, H. (2019). Condition monitoring of railway tracks from car-body vibration using a machine learning technique. *Applied Sciences (Switzerland)*, *9*(13), 2734. https://doi.org/10.3390/APP9132734.

Wang, Y., Correia, G. H. de A., van Arem, B., & Timmermans, H. J. P. (Harry). (2018). Understanding travellers' preferences for different types of trip destination based on mobile internet usage data. *Transportation Research Part C: Emerging Technologies*, *90*, 247–259. https://doi.org/10.1016/J.TRC.2018.03.009

BIOGRAPHIES

Isidro Durazo-Cardenas was born in Hermosillo, Mexico. He is a mechanical engineer with an MSc in advanced automation and design and a PhD in precision engineering from Cranfield University. He is an experienced researcher has led several R&D projects in health monitoring and inspection techniques. Recent research has been on railways autonomy of inspection and repair vehicles. He is a Sr lecturer in Life-cycle engineering at Cranfield University.

Bernadin Namoano is a Lecturer with extensive experience software engineering is currently working on developing novel techniques for asset health management using big time series data. He completed his PhD funded by EPSRC and Unipart-Rail/Instrumentel at Cranfield University in the field of condition monitoring applied to railway assets and was awarded the EPSRC Doctoral Fellowship Prize (2021-2023). His current work involves cognitive digital twins and machine techniques development.

Andrew Starr is chair of maintenance systems at Cranfield University. His work in novel sensing, e-maintenance systems, and decision-making strategies has been recognized with grant support such as H2020 Cleansky 2, iGear (PI), iBearing (PI), EPSRC Platform Grant EP/P019358/1, Through-life performance: From science to instrumentation, and EP/J011630 AUTONOM (PI), AMSCI 31587-233189 CHARIOT (PI), and circa £1.5m in projects for Network Rail in H2020 Shift2Rail.

Ram Dilip Sala received his bachelor's degree in civil engineering from Andhra University, India in 2019. He worked for Amazon until 2022 as Tron associate. He obtained an MSc in Management Information Systems from Cranfield University in 2023.

Jichao Lai is a graduate from Cranfield University (2022) and the Nanjing University of Aeronautics and Astronautics. He has completed an Aerospace Manufacturing MSc. His interests include exoskeleton robotics, iindustrial robot technology and machine learning applications.