Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p26424-cd

Predicting Track Geometry Data using Multimodal Methods with Neural Network Models

Nicholas Geere

Resilience Engineering Research Group, University of Nottingham, UK. E-mail: enyng1@nottingham.ac.uk

Dr Rasa Remenyte-Prescott

Resilience Engineering Research Group, University of Nottingham, UK. E-mail: r.remenyte-prescott@nottingham.ac.uk

Dr Joy Egede

School of Computer Science, University of Nottingham, UK. E-mail: joy.egede@nottingham.ac.uk

Robert Dean

London St. Pancras High-Speed, 90 York Way, London, N1 9AG, UK. E-mail: Robert.Dean@stpancras-highspeed.com

Asset management and maintenance of rail track assets often rely on data collected using remote condition monitoring approaches. Through wear due to freight or passenger trains passing over the track, different faults can develop that can lead to serious safety and performance risks. Observing the wear on the track requires a variety of methods to monitor the track adequately, as signs of faults can depend on their type. Some faults may need multiple sources of data from different monitoring approaches to accurately identify the fault and the level of damage. Being able to identify faults in advance is a predictive maintenance goal of railway infrastructure providers, such as Network Rail, due to the potential to avoid paying compensation to train operators from unplanned maintenance affecting train schedules (Xie et al., 2020). Track maintenance engineers work to accomplish this by utilising different modes of data, including digital imagery, track geometry data, and other types of data depending on the scenario (García et al., 2009).

Inspired by recent developments in computer vision, this paper examines the potential viability of incorporating multimodal approaches for prediction of track geometry, using high-definition photographs and historical track geometry readings. The performance of different neural network architectures is compared in the context of track condition monitoring data. This results in predictions of future track geometry data which include historical readings to aid in prediction. Further work would aim to continue developing models to predict dates and classifications of future faults and failures directly.

Keywords: Track geometry prediction, Neural networks, Resilience engineering.

1. Introduction

Railway track geometry is a key factor in determining the safety of rail infrastructure (Mohammadi et al., 2019) and thus needs to be carefully monitored and maintained. Track geometry degradation can exacerbate uneven dynamic loads contributing to or causing derailment (RAI, 2009). Furthermore, failing to maintain track geometry is not only a safety concern; poorly maintained track geometry can lead to expensive repairs and fines for delays affecting railway operator's schedules. Suspected faults with the condition of the track and its geometry are remotely monitored using a variety of methods and monitoring equipment, to detect when degradation must be addressed. However, with the economic demands of being able to run rail transportation safely, being able to predict the areas where geometry will degrade past industry standards can offer significant cost reductions in the maintenance of railway track infrastructure. This is due to the ability of maintenance teams to be able to plan maintenance into operators' schedules in a manner which does not delay any services, when given sufficient time to plan work in advance. To conduct maintenance to prevent further deterioration, reliable predictions of which suspected faults will develop into a severe fault needs to be developed to aid in decision making for determining the order of essential maintenance works. Rail maintenance engineers aim to make these predictions by not only considering historical track geometry data, but data from other sources or modalities, such

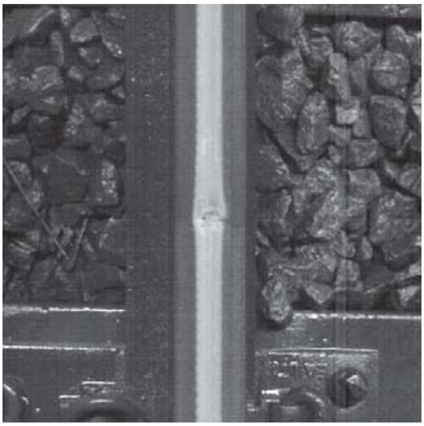


Figure 1. An image of visible degradation known as a squat, courtesy of One Big Circle.

as high-definition bird-eye-view images of the track. This cross-reference of geometry and visual inspection often allows for a clearer understanding of how the track is degrading and helps develop a more informed solution of what repairs are required.

Various types of faults can occur on railways that present differently on certain modalities, which further drives the reliance on multiple sources of information, in order to be able to identify suspected or confirmed faults. One case of a fault, which can develop and be visible through multiple modalities, is squats, which are a type of short wave RCF (rolling contact fatigue) fault (Li, 2009), see Figure 1. These faults are known to be detectable, for example, using ultrasonics (Jamshidi et al., 2017), eddy currents (Song et al. 2011), image processing (Faghih-Roohi et al., 2016), axle box acceleration (ABA) (Molodova et al., 2014), among other methods. However, the ease at which a squat can be identified depends on the stage of degradation of the squat. Very early squats are difficult to detect and require multiple data sources to confirm the existence, whereas much more severe squats can often be detected to a higher degree of accuracy. In the context of being able to predict faults, being able to reliably identify early-stage squats is key to developing a predictive model. Many approaches consider only a single mode in the prediction of degradation, with other non-rail focused multimodal methods focusing on alignment. This allows room for development of multimodal prediction methods for rail degradation prediction, proposed in this paper.

An approach with spatial-temporal neural networks is suggested to link the modalities of track geometry with digital imagery, in order to aid in the prediction of squat development in a multimodal context. In this paper, time series examples of paired track geometry and digital imagery recordings of squats are collected across 34 track locations from the UK's railway infrastructure provider Network Rail (NR) and used to predict the next iteration of track geometry readings.

2. Related Work

Squats were first identified in the UK in the 1960s and 70s on the British Rail (BR) West Coast Mainline (Grassie, 2012). Due to the potential for these faults to lead to rail breaks if unchecked, various methods have been developed to identify and predict squat faults. Squat initiation and progression can be explained and understood by many factors. The geometry condition of the track has shown interesting relationships, for example, squats are associated mostly with gentle curves with a radius between 800m and 1km, less so with severe curves (Zhu et al., 2020). Squats link to many other factors beyond track geometry, relating to the tonnage, speed, grade, the presence of water (or lack of in cases such as tunnels), among others (Li et al., 2008). These factors are often considered as constants in calculations, for simplicity and due to minimal variation in practical applications (Wang et al., 2023).

The prediction of geometry degradation has considered various types of approaches. To aid in the classification of track geometry defects, an ensemble of individual methods to predict degradation including a gamma process, a binary logistic regression model and a support vector machine (Cárdenas-Galloa et al., 2017). Each of the three methods are then combined using aggregation and stacking and then compared to determine which is most appropriate at classifying the severity of track geometry condition. This approach showed that the prediction of track geometry defects (cross level, dip and surface defects) was improved using the ensemble methods of aggregation or stacking over the individual methods. Machine learning approaches have become more prominent in fault identification and prediction. Auto-encoders and KMeans clustering have been used with the goal of localising regions of track containing faults (Popov et al., 2023). This approach used the auto-encoder to determine the regions where change in track geometry has been negligible and where there were improvements from maintenance work or degradation. The KMeans clustering was then applied to classify the areas where there is significant degradation, where there has been maintenance and where there is no significant change in the track geometry. Following this, the paper then linked the percent of certain infrastructure assets (i.e. bridges, etc.) near areas where degradation was identified. Hybrid models have been used for the prediction of foot-by-foot track geometry change over time, which includes a joint Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) model (Wang et al., 2023). This can help to judge the condition of track in future readings, and the patterns which occur within the geometry change at specific locations of track.

For the detection of squats directly using ride data, spiking neural network (SNN) with time-varying weights have been applied to ABA readings to enhance the detection performance of rail squats (Phusakulkajorn et al., 2025). These SNNs offer the ability to handle temporal and spatiotemporal patterns through their unique mechanism of processing data as discrete events or spikes. Squats have also been shown to be detectable through images alone using CNNs, such as the You-Only-Look-Once (YOLO) model series (Hovad et al., 2021). This identified surface defects, as well as good and bad isolation joints with 84%, 74% and 71% recall rate respectively. This was conducted using the YOLOv3 architecture, and there have since been many improvements on this version, but no further work focusing on rail degradation identification using this model. This, along with the larger computational demand of transformer models in multimodal approaches (Girdhar et al., 2023), encourages the use of CNN architectures, explored in this study.

3. Paired Time Series Data

Data is collected from measurement trains on the Network Rail railway network at 34 different locations across the network, and the proposed framework of analysis is based on such data. The data consists of 311 image-geometry pairs in total, of images containing squats and the corresponding track geometry data for the location where the image was taken. At each location, the image-geometry joins all available past pairs from that location to create an irregular time series dataset for each of the locations. Along with the pairs, a time between the previous recorded pair is also included.

Table 1. Definitions of track geometry parameters (Mohammadi et al., 2019).

Parameter	Definition
Gauge	Gauge is the distance between the inner head of the right and left rail.
Horizontal Alignment	Alignment or straightness is the projection of the track geometry of each track centreline onto the horizontal plane.
Twist	Twist is the difference between two cross-level measurements a certain distance apart.
Cant	Cant is the difference in the elevation between the top surfaces of the rails at a single point in a tangent track segment.
Cant Deficiency	The difference between the cant applied to the track and the theoretical equilibrium cant of a train at a certain speed.
Curvature	Curvature, the degree to which a curve deviates from a straight line.
Top	Uniformity of rail’s top surface measured in short distances along the tread of the rails.
Dip	Dip is the largest change in elevation of the centre line of the track within a certain moving window distance. Dip may represent either a depression or a hump in the track and approximates the profile of the centre line of the track.

Time between the recorded pairs is usually 4-6 weeks, with occasional longer gaps of approximately three months due to missing recordings or extremely noisy data, which were excluded from the dataset. The geometry data features, described in Table 1, that were included consisted of twist, horizontal alignment, cross level, gauge, cant, cant deficiency, curvature, top, and dip. The geometry of around 50 yards up and down the track from where the squat was included, based on correlation results found from the literature (Wang et al., 2023). In addition, the speed of the line was also considered.

Data needed to be collected by downloading and inspecting each image-geometry pair manually, due to the common occurrence of missing values and discontinuity within the geometry data in certain areas across the historical readings. Furthermore, images could have been recorded in certain light conditions that heavily interfered with viewing the track condition, which led to discarding the image-geometry pair. As this is a small dataset, the data must be well cleaned to avoid erroneous predictions.

4. Modelling approaches

Forecasting of track geometry using the aligned spatio-temporal data set is compared using two approaches, focusing on an LSTM based method and a deep CNN architecture.

4.1. *Data Alignment*

The geometry data, before being passed to the model, needs to be properly aligned to ensure that accurate predictions can be made, as even small variations can have a large impact on the accuracy. A section of 50 yards of track geometry readings is taken in the area where a squat is found. Then past images are inspected to trace back to when the squat was first imaged and the corresponding geometry readings were collected. The geometry data is already orientated to be in the same direction but requires a shift up or down the track to match recordings from the same place. The alignment is done via determining the phase shift with the maximum Pearson Correlation Coefficient between each pair of historical recordings for each location in the data. This is then visually checked to ensure proper alignment. The structure of the data for each parameter at each location is shown in Figure 2. At each location, the different geometry parameters are then sorted into matrices with rows representing readings taken along the 50-yard section, with columns representing the past historical recording (of varying length).

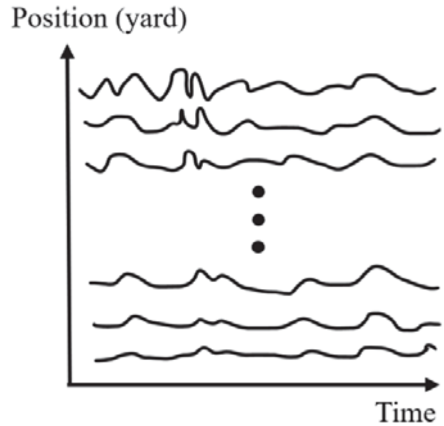
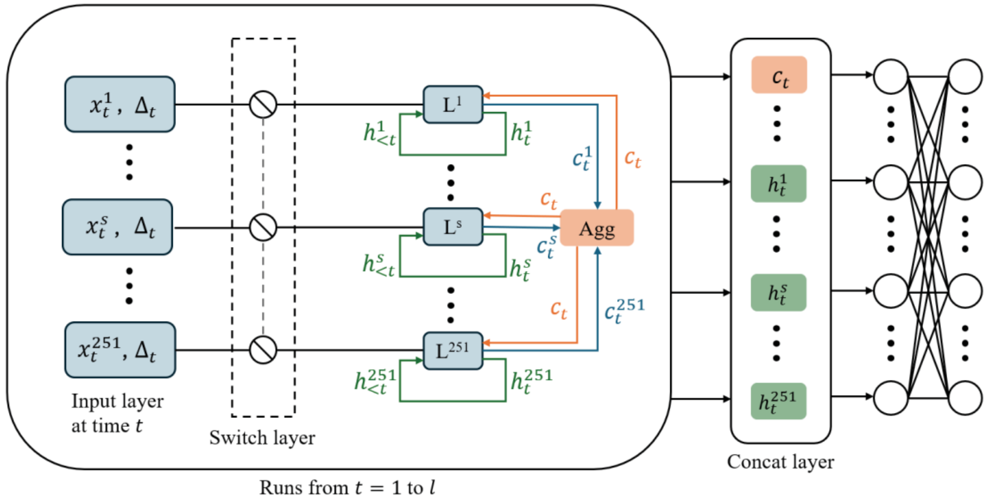


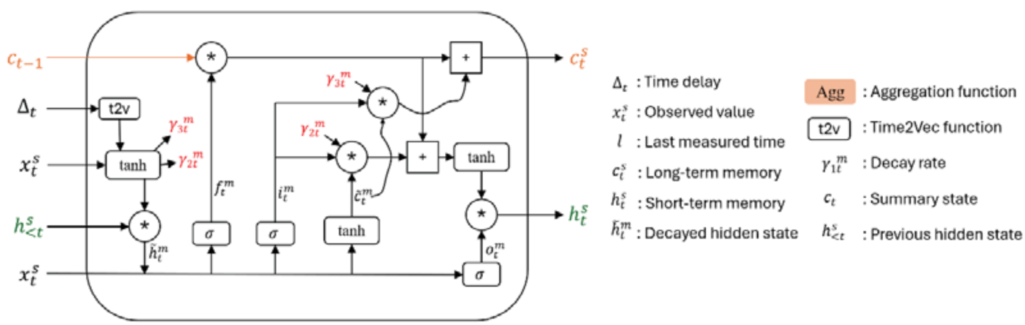
Figure 2. An illustration of the format of track geometry data for each geometry parameter.

4.2. *LSTM Approach*

The LSTM approach incorporates an adjustment on the traditional structure of an LSTM to account for the readings being taken at irregular intervals. The method, denoted as SLAN (Switch LSTM Aggregate Network) from Agarwal et al. (2024), which covers their method in detail. To briefly summarise the important points and adjustments from the paper, this approach incorporates a decay between each time step based on how much time had passed between the recordings. Thus, an extra parameter is passed to the model, Δ, the number of weeks between the geometry measurements, shown in Figure 3. This approach treats the recordings from each yard along the sampled location as inputs for a unique layer of the SLAN network. The switch layer follows the same approach as in the original SLAN architecture, however only continuous 50-yard sections were used in training for this study. In practice, this meant for the aggregate function, an average of the long-term memory values from each LSTM cell is used. The outputs of the short-term memory values and the aggregated value of the



(a) The adapted SLAN architecture.



(b) LSTM block L^s.

Figure 3: (a) The adapted SLAN architecture incorporating a n adjusted switch layer and fully connected final layer (b) the working of an individual LSTM cell using time decay.

long-term memory is then fed into a fully connected layer to predict the values for each parameter along the 50-yard section being predicted. To avoid overfitting, the feed forward network only contains one layer. This approach uses the sigmoid activation function.

4.3. *CNN Approach*

The CNN approach involves three convolutional layers to process the spatio-temporal inputs of each parameter. The details of the architecture are included in Table 2. The model uses three convolutional layers on the data before being passed to an adaptive average pooling layer over the time features to form a single column and a final convolutional layer to predict the next time step. This approach used the ReLu activation function.

4.4. *Results*

Both models were trained for 1200 epochs with a batch size of 2 and a learning rate of 0.001, using the AdamW optimiser in Pytorch. The training dataset included data from 26 of the 34 locations,

Table 2: Structure of CNN based method for a given number of time columns T.

Layer	No. features in	No. features out	Kernel size	Padding
1st Conv2D	1x251xT	32x251xT	(3, 5)	(1, 2)
2nd Conv2D	32x251xT	64x251xT	(3, 5)	(1, 2)
3rd Conv2D	64x251xT	64x251xT	(3, 3)	(1, 1)
Adapt average pool	64x251xT	64x251x1	N/a	N/a
4th Conv2D	64x251x1	1x64x1	(1, 1)	(0, 0)

while the validation and test sets had data from 4 locations in both. As shown in Figure 4, the validation error became greater than the training went beyond the 1200th epoch, so training was halted after this point for both approaches.

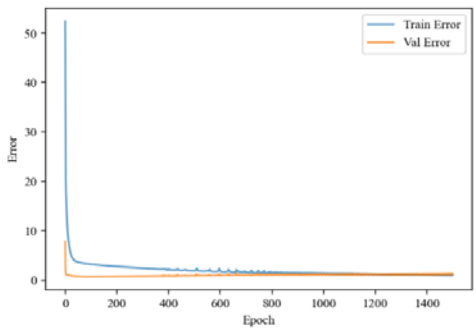


Figure 4: Train set error vs validation set error during training.

For the SLAN architecture, while predictions did pick up large features of the data, such as changes in geometry over several yards, it struggled in predicting finer changes below a yard shown in figure 5. When compared to the CNN based approach, the CNN outperformed the SLAN model. This can be expected as the approach more directly considers the spatial dependence in comparison with the LSTM approach, which only considers a global pooled long term memory parameter between spatial features as shown in figure 6. The results compare the RMSE and the average PCC (Pearson Correlation Coefficient) on the average

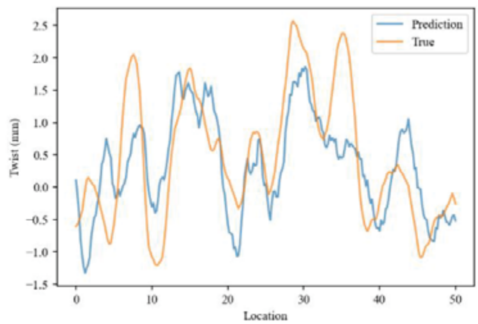


Figure 5: The prediction of future twist values along 50 yards at a particular section of track containing degradation using the LSTM based method.

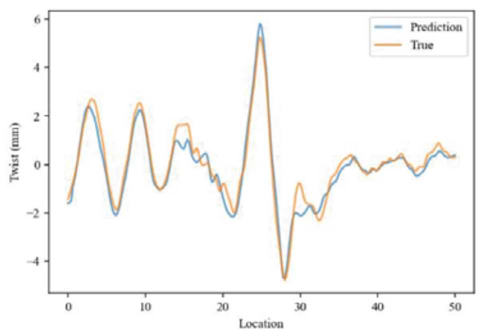


Figure 6: The prediction of future twist values along 50 yards at a particular section of track containing degradation using the CNN based method.

performance of the model predictions across the parameters of twist, gauge, alignment, cant and top. The CNN based approach achieved an RMSE (readings measured in mm) of 0.883 and an average PCC score on the test set of 95.9% based on a 1-month prediction period, whereas the SLAN approach achieved an RMSE (readings also measured in mm) of 2.04 and an average PCC of 57.8%.

The predictions for each geometry parameter incorporate only curvature in the contextual data. Uncertainty in the progression of track geometry degradation will arise due to lack of additional conditional data such as line speed and tonnage. This data was unavailable to include within this paper but will look to be included in further works. To ensure predictions of geometry degradation remain accurate further into the future, predictions can be validated with data collected from future inspections. Predictions can then be compared to the actual data and used to calculate what adjustments need to be made to existing model parameter values to improve the predictions. It is important to consider that any maintenance work conducted will significantly change the profile for a section of track. Estimating the expected impact of maintenance work is outside the scope of this paper.

5. Conclusions

In order to determine what is the most effective way to conduct maintenance work, effective planning needs to be carried out. Being able to accurately predict at what point a level of degradation will occur is largely beneficial to be able to plan maintenance schedules. Our study demonstrates that track geometry predictions are more accurate when incorporating spatial and temporal features into the analysis. This is highlighted with the approach of using a convolutional kernel over a spatio-temporal input, compared to an LSTM based approach. The LSTM method places less focus on the spatial features of the input data, which can help to explain the difficulty in determining finer geometry patterns in small adjacent regions by this approach The CNN based approach was able to predict with greater accuracy but showed errors when describing the height of peaks in geometry data. The method could be enhanced in further models and work by several approaches: more methods can be implemented around deeper CNNs, using architectures more focused on predicting spikes in geometry data, and hybrid methods which incorporate layers for temporal features and spatial features individually. Furthermore, the data could be utilised in a more multimodal approach to use input images along with the geometry data to aid in predicting the rate of degradation.

In future work, more options will be explored with greater amounts of training data. For a wider application to influence decision making in railway maintenance schedules, predictive methods need to incorporate all available data to be able to provide the most reliable track geometry predictions possible.

Acknowledgements

Data is provided through Network Rail and One Big Circle's AIVR dashboard. Funding for research is supported by London St Pancras High Speed and the University of Nottingham.

Data Availability

Data will not be available due to a non-disclosure agreement with the data providers. Real data used has kept exact location details of railway track readings anonymous.

References

Xie, J., J. Huang, C. Zen, S-H. Jiang, and N. Podlich (2020). Systematic Literature Review on Data-Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering. Geosciences, 10(11):425.

García Márquez, FP., C. Roberts, and AM. Tobias (2009). Railway point mechanisms: Condition monitoring and fault detection. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 224(1):35-44.

Mohammadi, R., Q. He, F. Ghofrani, A. Pathak, and A. Aref (2019). Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects, Transport. Res Part c: Emerg. Technol. 102 153–172.

RAI (2009), Derailment at Ely Dock junction 22nd June 2007. Department for Transport.

Li, Z. (2009). Chapter 13 - squats on railway rails. in: R. Lewis, U. Olofsson (Eds.), Wheel–Rail Interface Handbook, pp. 409–436. Woodhead Publishing.

Jamshidi, A., S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, R. Dollevoet, Z. Li, B. De Schutter (2017). A big data analysis approach for rail failure risk assessment. Risk Anal. 37. 1495–1507.

Song, Z., T. Yamada, H. Shitara, Y. Takemura (2011). Detection of damage and crack in railhead by using eddy current testing. Electromagn. Anal. Appl. 3. 546–550.

Faghih-Roohi, S., S. Hajizadeh, A. Núñez, R. Babuska, B. De Schutter (2016). Deep convolutional neural networks for detection of rail surface defects. in: Proceedings of the International Joint Conference on Neural Networks 2016. pp. 2584–2589.

Molodova, M., Z. Li, A. Núñez, and R. Dollevoet (2014). Automatic detection of squats in railway infrastructure. IEEE Trans. Intell. Transp. Syst. 15. 1980–1990.

Grassie, S L. (2012). Squats and squat-type defects in rails: the understanding to date. Proceedings of the Institution of Mechanical Engineers. Part F: Journal of Rail and Rapid Transit, 226, 235-242.

Zhu, H., H. Li, A. Al-Juboori, D. Wexler, C. Lu, A. McCusker, J. McLeod, S. Pannila, and J. Barnes (2020). Understanding and treatment of squat defects in a railway network. Wear. 442-443. 203139.

Li, Z., X. Zhao, R. Dollevoet, and M. Molodova, (2008). Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects. International Journal of Vehicle Mechanics and Mobility 46. 237-246

Hovad, E., T. Wix, M. Khomiakov, G. Vassos, A. Filipe da Silva Rodrigues, A. de Miguel Tejada and L H. Clemmensen (2021). Deep Learning for Automatic Railway Maintenance. In: Galeazzi, R., Kjartansson Danielsen, H., Kjær Ersbøll, B., Juul Jensen, D., Santos, I. (eds) Intelligent Quality Assessment of Railway Switches and Crossings. Springer Series in Reliability Engineering.

Wang, X., Y Bai, X Liu (2023). Prediction of railroad track geometry change using a hybrid CNN-LSTM spatial-temporal model, Advanced Engineering Informatics 58. 102235.

Cárdenas-Galloa, I., C. Sarmientoa, G. Moralesa, M. Bolivara, R. Akhavan-Tabatabaeia (2017). An ensemble classifier to predict track geometry degradation. Reliability Engineering and System Safety 161. pp. 53-60.

Popov, k., R. De Bold, H.-K. Chai, M.C. Forde, C.L. Ho, J.P. Hyslip, H.F. Kashani, R. Kelly, S.S. Hsu, M. Rippin (2023). Data-driven track geometry fault localisation using unsupervised machine learning. Construction and Building Materials 377. 131141.

Phusakulkajorn, W., J. Hendriks, Z Li, A. Núñez (2025). Spiking neural network with time-varying weights for rail squat detection. Applied Soft Computing 184. 113689.

Li, Y., Y Du, Y Liu, F Feng, M X Feng,Y Wu (2025). On Support Samples of Next Word Prediction. arXiv. 2506.04047.

Agarwal, R., A. Sinha, A. Vishwakarma, X. Coubez, M. Clausel, M. Constant, A. Horsch, D. K. Prasad (2024), No Imputation Needed: A Switch Approach to Irregularly Sampled Time Series. arXiv. 2309.08698.