iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49069

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Wheel Defect Detection with Advanced Machine Learning

Sumit Kumar Das
KIIT University, Bhubaneswar

Abstract: ***Wheel defects on railway wagons have been identified as an important source of damage to the railway infrastructure and rolling stock. They also cause noise and vibration emissions that are costly to mitigate. We propose two machine learning methods to automatically detect these wheel defects, based on the wheel vertical force measured by a permanently installed sensor system on the railway network. Our methods automatically learn different types of wheel defects and predict during normal operation if a wheel has a defect or not. The first method is based on novel features for classifying time series data and it is used for classification with a support vector machine. To evaluate the performance of our method we construct multiple data sets for the following defect types: flat spot, shelling, and non-roundness. We outperform classical defect detection methods for flat spots and demonstrate prediction for the other two defect types for the first time. Motivated by the recent success of artificial neural networks for image classification, we train custom artificial neural networks with convolutional layers on 2-D representations of the measurement time series. The neural network approach improves the performance on wheels with flat spots and non-roundness by explicitly modelling the multi sensor structure of the measurement system through multiple instances learning and shift invariant networks.***

I. INTRODUCTION

Measurement System and Defect Types

A. Wheel Load Checkpoint

As part of this system, the wheel load checkpoints (WLC) measure vertical force through strain gauges installed on the rails. These devices are used for observing maximal axle load, maximal train load, load displacement and grave wheel defects. Our study investigates the use of machine learning methods to defect and classify wheel defects based on the data obtained through these wheels load checkpoints.

Each WLC consists of four 1m long measurement bars with four strain gauges (referred to as sensors in the following) per measurement bar. Since on each side two measurement bars with 4 sensors are installed, each wheel that runs over the WLC is measured eight times at different parts of the wheel.

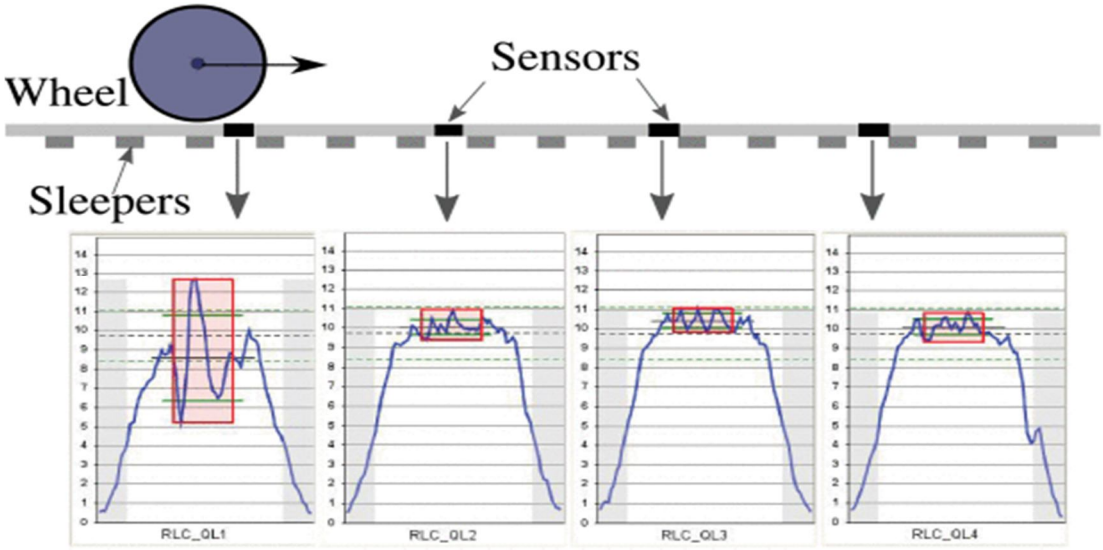


Fig. 1.

Multiple vertical wheel force measurements of a train wheel by the four sensors of one measurement bar. The wheel is affected by a discrete defect that manifests itself in the measurement of the first sensor. The remaining sensors do not directly observe the defect.

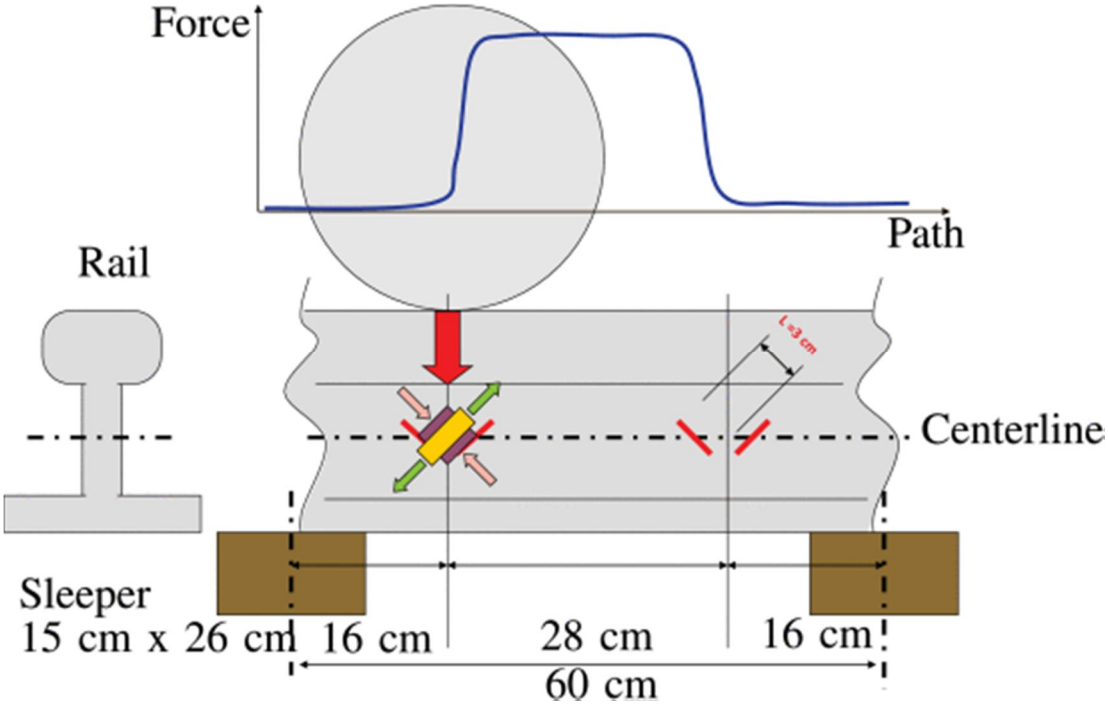


Fig. 2.

Diagram of one sensor on a measurement bar of the WLC. The strain gauges are attached to the side of the wheel between two sleepers and cover 28cm of vertical wheel force of the wheel rolling on the track.

B. Railway Wheel Defects

A relatively well understood wheel defect type is the flat spot or wheel flat. This defect occurs when the wheel stops rotating (for instance during an emergency brake) and is dragged along the track. Apart from flat spot, other common wheel defects on railway vehicles are *non-roundness* and *shelling*. Wheels with non-roundness have a high influence on the vibration and noise emitted by a passing train and, therefore, they are an important type of defect to detect. Non-roundness, in contrast to shelling and flat spot, is a non-discrete type of defect. This characterization means that the defect affects a large part of the wheel and changes its shape in a non-local way.

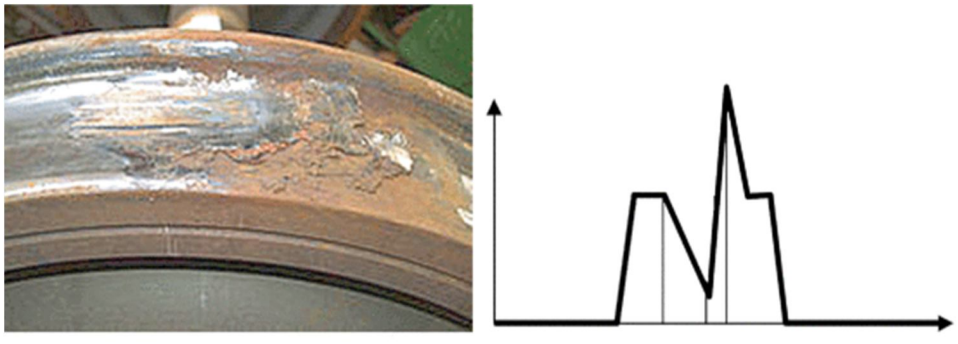


Fig. 3.

Picture of a serious flat spot on a train wheel of SBB (a) and the resulting idealized wheel load measurement (b). (Picture taken from Wikipedia/Bobo11)

II. BACKGROUND

A. Data Sets and Models

Two data sets from different sources are assembled to evaluate the performance of different methods for wheel defect detection and classification and to train various classifiers. For both data sets the signals that we use to predict a wheel defect are measured by the wheel load checkpoint.

1) Models and Features

On the first data set we compare the Wavelet-SVM with benchmark flat spot prediction methods. We show that it greatly outperforms prior art based on thresholding the dynamical coefficient (Eq. 10 below) and also on multiple instance learning with dynamic time warping.

The second data set serves to demonstrate that the Wavelet-SVM can accurately classify all three defect types. We also compare the performance of the deep learning models on different time series representations by showing that the cyclic permutation network outperforms the simpler neural networks and also the Wavelet-SVM for non-roundness. For flat spots, the neural network with features learned on the 2D time series representation also outperforms the Wavelet-SVM.

2) Data Set 1: Calibration Run

To acquire a first training data set for flat spots, two wheels on different wagons were artificially damaged. The wagons were then added to a calibration train that was run over different measurement sites with different velocities and from both directions to calibrate the wheel load check points. This resulted in 1600 measurements, 50% of which are from a wheel with a flat spot.

We also consider another method to detect flat spots in this data set, that is not based on machine learning. It is a conservative threshold on the dynamic coefficient: a general measure of spread within one time series. For each sensor this coefficient is given by

$$d_{BW}(x) = \frac{\max(x)}{x`} \quad \text{----------(10)}$$

where max and x` refer to the maximum and average value of a sequence of measurements x, respectively.

3) Data Set 2: Reprofile Events

To generate data for training and testing a classifier that can predict additional types of wheel defects, we aggregated the time and date of reprofile events and linked them to railway wagons. We used two sources for these events: the protocols of repair workshops of freight trains and the regular maintenance measurements of passenger trains. These were annotated with a defect class by an expert before re-profiling the defective wheels. Using this procedure, we were able to obtain a large data set of annotated measurements from wheels of different defect classes over the span of multiple years. 1836 measurements are evaluated for flat spot detection, where 588 cases are classified as defective. For shelling, we received 6070 measurements, with 2678 being defective. For the non-roundness defect class, 688 cases out of 920 measurements are defective.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

For performance evaluation of the methods, we compute three metrics: accuracy, precision and recall. Whereas accuracy gives the total fraction of correctly classified wheels, precision measures the fraction of correctly predicted defects out of all predicted defects and recall the fraction of correctly predicted defects out of all defects.

A. Model Selection and Evaluation

To make the evaluation robust against chance we repeat each experiment multiple times on new random train/test splits and report average and standard deviation over these repetitions. For data set 1 we only report the average as the standard deviation was not reported for the benchmark method. For data set 1 50% of the data is hold out for testing, for data set 2 20%. For the Wavelet-SVM the average performance is computed over 10 repetitions, for the DNNs over three repetitions.

B. Data Set 1

In a study prior to this publication [27], this data set was used to empirically demonstrate the effectiveness of a new algorithm for MIL [26].

Using the features described in Section III with a SVM (Section V) we were able to improve accuracy to 92% (Table I).

Table I Test Set Performance on Data Set 1

Method	Accuracy (%)	Precision (%)	Recall (%)
Wavelet-SVM (ours)	92	94	93
eMIL	70	-	-
Dynamic coeff.	60	100	22

With the current operational threshold of θ=3 on the maximal dynamic coefficient (Eq. 10) an accuracy of 60% is achieved. This is relatively low, as with random guessing already 50% accuracy could be achieved. It is thus important to note that the precision of this method is perfect with 100% of reported wheels being defective. So even though the method misses defective wheels it never raises a false alarm.

C. *Data Set 2 – SVM*

The SVM classifier (Section V) are trained on the labels obtained by this method for the defect types flat spot, shelling and non-roundness.

In Table II the performance on the reserved test set is reported for each defect type including standard deviation over the permutations.

Table II Test Set Performance of the Wavelet-SVM on Data Set 2

Defect	Accuracy (%)	Precision (%)	Recall (%)
Flat spot	87 ± 3	89 ± 4	86 ± 6
Shelling	92 ± 2	92 ± 3	93 ± 3
Non-roundness	87 ± 6	87 ± 10	89 ± 4

1) This observation can be explained by the fact that the training set for this defect type was by far the largest, so we were able to train a classifier with higher accuracy.
2) This defect type also affects the wheel globally, so it is harder to miss for the sensors than a flat spot.
3) To improve the performance on flat spot and non-roundness we trained custom deep neural networks and give the results in the next section.

D. *Data Set 2 - Deep Learning*

Using the same data set as in the previous section we evaluate the deep learning method (Section IV) on the two defect types flat spot and non-roundness. To simplify the experiments, we do not include additional features like speed, measurement site or template fit, but only consider the wheel vertical force measurements from the WLC sensors. Therefore, the performance of the SVM is slightly worse compared to the previous section.

To compute the 2D image of the time series we proceeded as following:

1) The recording from each of the 8 channels have been pre-processed via PAA [24], with bin number N=156.
2) The GAF encoding as well as the 2D graph were computed for each channels (we took the following parameters for the 2D graph: Vmin=−4,Vmax=6 as the window captures more than 99.9% of all the values, and r=Vmax−VminN=10N to generate square pictures of size N×N).
3) Finally, the picture size was further reduced by averaging every 2×2 non-overlapping pixels for computational reasons, resulting in 8 channels of size 78×78 for both GAF and 2D graph encoding.

The learning rate was set to decay inversely proportional to the number of epochs.

a) Flat Spots

In Table III we compare the performance of the different DNN models and the Wavelet-SVM. The only deep model that is able to out-perform the accuracy of the Wavelet-SVM is based on the 2D image of the time series. All of the deep models have smaller standard deviation and higher precision.

Table III Test Set Performance of the Deep Models on Flat Spots in Data Set 2

Model	Accuracy (%)	Precision (%)	Recall (%)
Deep 1D	88 ± 1	96 ± 2	79 ± 3
Deep 2D	89 ± 1	93 ± 2	85 ± 2
Deep GAF	87 ± 2	91 ± 1	81 ± 5
Wavelet-SVM	87 ± 3	87 ± 2	86 ± 5

b) Non-Roundness

In Table IV we compare the performance of the cyclic DNN with the DNN used for flat spot prediction (Deep MIL), a DNN that is trained on the concatenation of all the sensors (Deep Concat) and the Wavelet-SVM. Remember that the MIL-DNN used for flat spot prediction is trained by looking at the time series of each sensor individually and computing the loss on sensor with highest probability of observing the defect.

Table IV Test Set Performance of the Deep Models on Non-Roundness in Data Set 2

Model	Accuracy (%)	Precision (%)	Recall (%)
Deep MIL	81 ± 1	89 ± 3	71 ± 3
Deep Concat	81 ± 2	82 ± 2	78 ± 3
Deep Cyclic	88 ± 1	93 ± 1	82 ± 1
Wavelet-SVM	84 ± 9	80 ± 13	95 ± 3

In comparison with the Wavelet-SVM the cyclic DNN shows higher accuracy and precision and reduced variance.

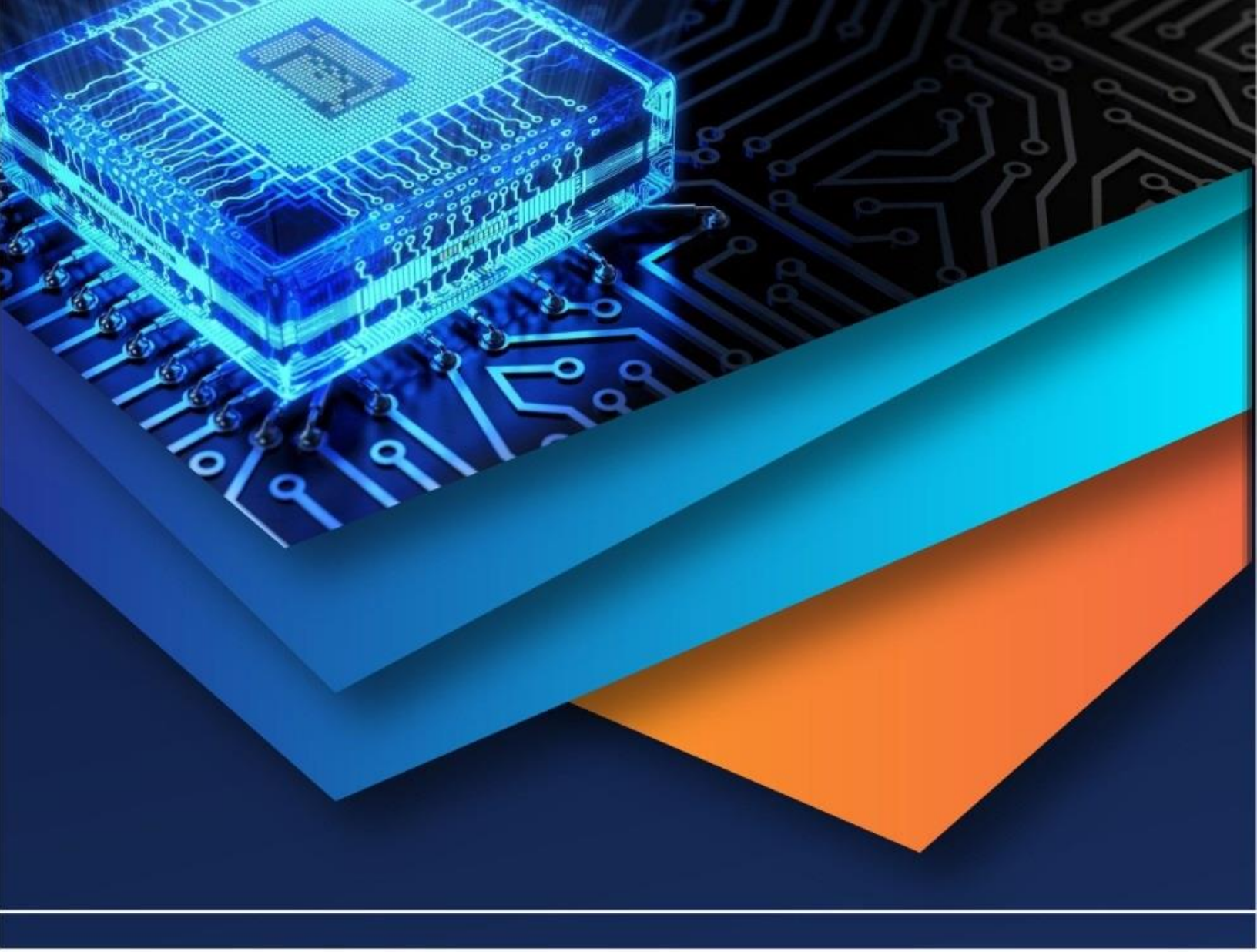
IV. CONCLUSION

We have presented two machine learning methods for defect detection on railway train wheels. The methods analyse multiple time series of the vertical force of a wheel under operational speed and output if a wheel has a defect or not. Both methods are trained automatically on measurements gathered from defective and non-defective wheels. The first method is based on novel general wavelet features for time series. The second method employs deep convolutional neural networks to automatically learn features from the time series directly or from a 2-dimensional representation.

The methods that were developed for this work are currently being implemented as part of the SBB wayside train monitoring system. To improve the quality of the training and test data RFID tags will be deployed to enable perfect association between defect labels and measurements. Further future work consists of integrating external features into the deep learning models, optimizing for precision and predicting severity scores for the defects.

REFERENCES

[1] A. Bracciali, "Railway Wheelsets: History, Research and Developments", International Journal of Railway Technology, 5(1), 23-52, 2016. http://dx.doi.org/10.4203/ijrt.5.1.2
[2] https://www.sciencedirect.com/topics/engineering/railway-wheel
[3] https://www.ijser.org/researchpaper/analysis-of-railway-wheel-to-study-thermal-and-structural-behaviour.pdf
[4] Tao, G., Wen, Z., Jin, X. et al. Polygonisation of railway wheels: a critical review. Rail. Eng. Science 28, 317–345 (2020). https://doi.org/10.1007/s40534-020-00222-x



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)