A TECHNICAL SEMINAR REPORT ON

Wheel Defect Detection With Machine Learning

Submitted in partial fulfilment of the
Requirements for the award of the Degree of

BACHELOR OF ENGINEERING

IN

INFORMATION TECHNOLOGY

Submitted by

T. HARSHA SRI(17B81A1219)

Under the esteemed guidance of

Dr. Bipin Bihari Jayasingh

Professor , IT Department

Department of Information Technology

CVR College of Engineering

ACCREDITED BY NBA, AICTE & Affiliated to JNTU-H

Vastunagar, Mangalpally (V), Ibrahimpatnam (M), R.R. District, PIN-501 510

2020-21

Cherabuddi Education Society's
CVR COLLEGE OF ENGINEERING
(An Autonomous Institution)
ACCREDITED BY NATIONAL BOARD OF ACCREDITATION, AICTE
(Approved by AICTE & Govt. of Telangana and Affiliated to JNT University)
Vastunagar, Mangalpalli (V), Ibrahimpatan (M), R.R. District, PIN - 501 510
Web : http://cvr.ac.in, email : info@cvr.ac.in
Ph : 08414 - 252222, 252369, Office Telefax : 252396, Principal : 252396 (O)

DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the Technical Seminar Report entitled **"Wheel Defect Detection Using Machine Learning"** is a bonafide work done and submitted by **T . Harsha Sri(17B81A1219)** during the academic year 2020-2021, in partial fulfillment of the requirement for the award of Bachelor of Technology degree in Information Technology from Jawaharlal Nehru Technological University Hyderabad, is a bonafide record of work carried out by them under my guidance and supervision.

Certified further that to my best of the knowledge, the work in this dissertation has not been submitted to any other institution for the award of any degree or diploma.

INTERNAL GUIDE	HEAD OF DEPARTMENT
Dr.Bipin Bihari JayaSingh	Dr.Bipin Bihari JayaSingh
Professor,IT Department	Professor,IT Department

TECHNICAL SEMINAR COORDINATOR

Mr. Hussian Ahmed Choudhury

Assistant Professor, IT Department

ACKNOWLEDGMENT

The satisfaction of completing this Technical Seminar would be incomplete without mentioning our gratitude towards all the people who have supported me. Constant guidance and encouragement have been instrumental in the completion of this seminar.

First and Foremost, I thank the Chairman, Principal, Vice Principal for availing infrastructural facilities to complete this.

I offer my sincere gratitude to our internal guide**, Dr. Bipin Bihari Jayasingh**, Professor, IT Department, CVR College of Engineering for his immense support, timely co-operation and valuable advice throughout the seminar.

I would like to thank Professor In-Charge of seminars, **Prof.U.V. RamanaSarma,** Information Technology for his valuable suggestions in implementing the seminar..

I would like to thank the Head of Department, Professor **Dr. Bipin Bihari Jayasingh**, for his meticulous care and cooperation throughout this work.

I am thankful to **Mr. Hussian Ahmed Choudhury, Seminar Coordinator, Assistant Professor, IT Department,** CVR College of Engineering for her supportive guidelines, and for having provided the necessary help for carrying forward this without any obstacles and hindrances.

I also thank the **Project Review Committee Members** for their valuable suggestions.

DECLARATION

I hereby declare that the Technical Seminar report entitled **"Wheel Defect Detection using Machine Learning"** is an original work done and submitted to IT Department, CVR College of Engineering, affiliated to Jawaharlal Nehru Technological University Hyderabad, Hyderabad in partial fulfillment of the requirement for the award of Bachelor of Technology in **Information Technology** and it is a record of Technical Seminar work carried out by us under the guidance of Professor **Dr. Bipin Bihari Jayasingh**, Professor,IT Department of Information Technology.

I further declare that the work reported in this Technical Seminar has not been submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other Institute or University.

T.Harsha Sri(17B81A1219)

TABLE OF CONTENTS

1.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. Author 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. This 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 these method author construct multiple data sets for the following defect types: flat spot, shelling, and non-roundness. Author 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, author 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 modeling the multi sensor structure of the measurement system through multiple instance learning and shift invariant networks.

2. INTRODUCTION

Early detection of serious wheel defects on freight trains are an essential part in preventing damage to the railway infrastructure and in providing the train operators with timely information on necessary repairs, that can prevent further deterioration of the wheels. Wheel defects of railway vehicles directly cause an increase in attrition of and damage to the railway infrastructure, e.g., the track systems or the civil engineering works, thereby adding additional costs to maintenance and repair and leading to a reduced lifetime and availability of rolling stock.

The life span of the railway infrastructure is significantly shortened by the negative effects of wheel defects. The life span of railway bridges for instance is calculated with an assumed maximal dynamical load of 21 tons. Due to wheel defects the actually occurring dynamical load can be up to 50 tons, or 270% higher than the theoretically assumed maximum, thus shortening the life span. Wheel defects also accelerate crack-growth on the rail tracks and lead to premature failure of the rail system. Another important effect caused by wheel defects are ground vibration and noise emissions. In the European Union (EU) Project "Railway Induced Vibration Abatement Solutions" (RIVAS)1 27 partners from nine countries investigated the source and mitigation measures for noise and vibration emissions. They found that reducing wheel defects by wheel maintenance significantly reduces vibration and noise emissions directly [1]. Therefore, it is recommended to use timely and targeted maintenance of train wheels as an economic means to reduce emissions [2]. This measure is all the more important as the density and usage of modern railway networks is steadily increasing and failures quickly disrupt operation of the whole network or parts of it. Since 2008, all states in the EU are advised to employ noise emission ceilings. Switzerland started a noise abatement program based on emission ceilings that requires the infrastructure manager to curb emissions above the ceiling. This abatement programme leads to total costs of 1.5 billion CHF [3].

In this paper author propose a method of detecting defective wheels. This classification method promises to increase the reliability of the railway infrastructure, to reduce the cost of freight train operation and to save additional investments on noise protection measures. To reach this goal without the costly construction of further measurement sites or newly built sensors, author propose the use of statistical methods that allow us to automatically inspect the existing data and extract the information about defective wheels that is already present.

Our proposed methods do neither require a model of the measurement system, nor of train dynamics or wheel defects. The methods enable us to predict defects on wheels where there is no prior understanding of how these defects manifest themselves in the measurements. The methods detect and classify different types of defects based on measurements during normal operation where the trains pass the measurement sites in full operational speed. The features that author have developed for the use in supervised learning are general and can in principle be used for any time series data and are not restricted to specific defect types. In a second step author automatically learn features directly from the raw measurement signal.

3. LITERATURE SURVEY

While there has been research on machine learning methods for railway track inspection [4]–[6] or condition based maintenance [7], to our knowledge machine learning methods for railway wheel defect detection have not been developed so far. There has been some research on sensor systems for wheel defect detection on freight trains. Nenov et al. [8] analyses the signal from acceleration sensors and demonstrate that they can visually see a difference between the measurements of wheels with flat spots and good wheels but they do not propose a method for detection. Another related work [9] advocates the use of Fibre Bragg Gating sensors for defect detection of rails to monitor track conditions. The authors investigate the wavelet decomposition of pressure signals but they do not propose a method or threshold for automatic defect detection. Jianhai et al. [10] use continuous wavelet analysis of acceleration sensor data to visually inspect the measurements and conclude that there is a difference in the coefficients for wheel with flat spots and defect-free wheels.

Different kinds of track scales are in use in the field. They can in principle be used to detect flat spots. But to our knowledge they do not use machine learning to train a defect classifier. A general advantage of our proposed system is that the measurement system is relatively inexpensive, but they can show that it can still be used to detect wheel defects, thanks to our proposed machine learning methods.

4. PROBLEM STATEMENT

Wheel defects on railroad carts have been distinguished as a significant wellspring of harm to the railroad framework and moving stock. They likewise because clamor and vibration outflows that are exorbitant to relieve. This paper propose two machines learning to automatically identify these wheel defects, given the wheel vertical power estimated by a forever-introduced sensor framework on the railroad organization. These techniques automatically learn various sorts of wheel defects and foresee during ordinary activity if a wheel has a deformity or not.

5. METHODOLOGY

5.1 Introduction

The main contribution is two methods for automatic railway wheel defect detection and classification through vertical force measurements of trains running in full operational speed. For the first method authors design novel wavelet features for time series data from multiple sensors and learn a classifier using a support vector machine. For the second method Author design and train convolutional neural networks for different wheel defect types by deep learning. Author evaluate our novel and other classical methods for wheel defect detection on two labeled data sets with different types of wheel defects, that are constructed from calibration runs and from maintenance reports.

5.2 Measurement System and Defect types

In this section, we will describe first the measurement system to detect the defect in railway wheels and next we specify different types of defects used for classification

5.2.1 Wheel Load Checkpoint

As part of the SBB infrastructure for automatically monitoring trains, 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 wheel 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 shows schematically the measurement of one wheel by one measurement bar. In this example a defect is directly observed by the measurement of the first sensor.

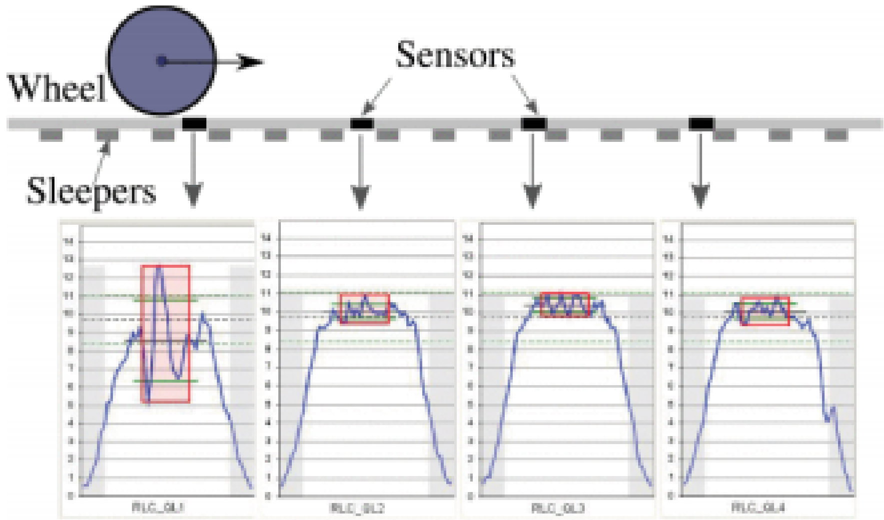


Fig. 1.

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.

See Fig. 2 for a diagram of one sensor. The strain gauges are installed perpendicular on the centerline of the railroad track and they are combined into one vertical wheel force measurement. One sensor covers approximately 30cm of the wheel circumference. The wheel load checkpoints are installed on multiple strategic sites on the railway network.

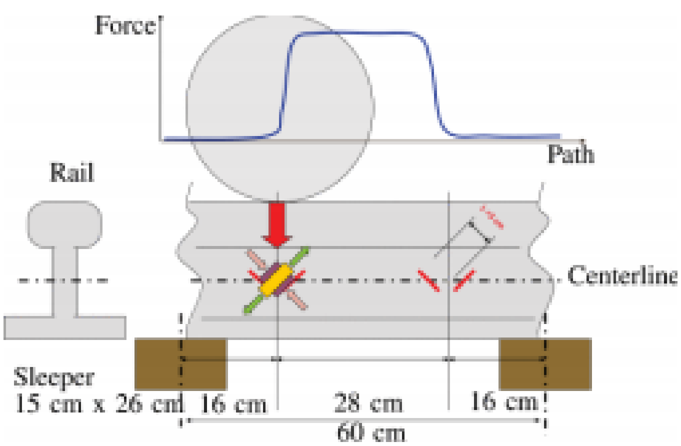


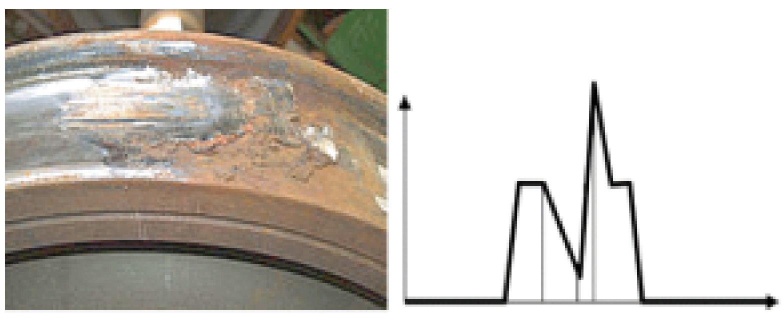
Fig. 2.

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.

5.2.2 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. Fig. 3 shows an image of a flat spot on a railway wheel of SBB and the corresponding idealized measurement obtained by the WLC if the flat spot directly hits a sensor of the measurement system. Grave wheel flats can be detected by looking at simple statistics of the measurement if the defect hits the sensor perfectly. To be able to detect flat spots that are less grave or that do not hit a sensor directly, more advanced machine learning methods are required.

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. Author create an additional data set that contains the defect types flat spot, non-roundness and shelling and then, compare the performance of our two machine learning methods in predicting these three defect types



(a) (b)

Fig. 3.

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

5.3 Time Series Representation for Defect Detection

An important step in any machine learning method is finding a representation of the original measurements that supports discrimination between different classes. In this paper Author suggest to decompose the signal by a multiscale wavelet analysis in order to extract indicative frequency features for time series data.

5.3.1 Wavelet Transform

The symmetrical wavelets given by definition at various scales settle the first sign at various goals. The DWT would thus be able to be utilized to develop a multiresolution signal guess. An equal method of figuring the DWT is bypassing the first sign through a progression of suitable high-pass and low-pass channels and sub-inspecting tasks, where at each level, the yield of the high-pass channel is put away as the detail coefficients for that level and the yield of the low-pass channel is disintegrated further at the following level until level T =log(n) is reached, where n is the length of the first sign[9]. If the high-pass and low-pass channels in this channel bank are gotten from the kid wavelets in Equation 1, the detail coefficients (C1. . . CT) compare precisely to the wavelet coefficients. The wavelet change has been broadly utilized in fields going from biomedical sign preparation, geosciences to picture pressure. Since weight estimation signals and the deformity impacts on the sign are both restricted as expected and recurrence, the wavelet change unequivocally encodes this neighborhood annoyance and, this way, has an advantage over the Fourier change in our application.

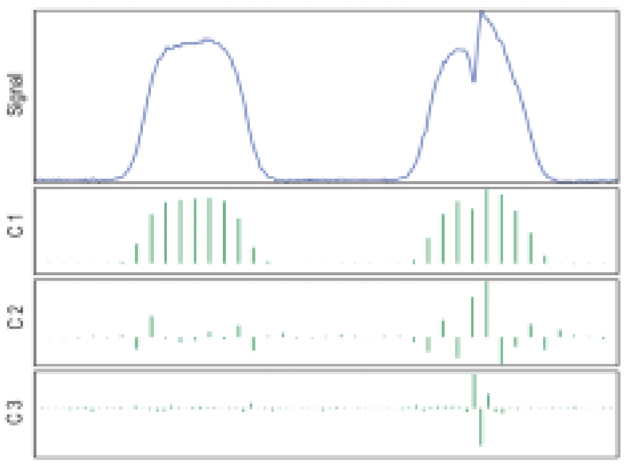


Fig. 4.
Signals and wavelet coefficients at different levels (C1 to C3) of a defective (right) and non-defective (left) wheel. The power in the high frequency coefficients C2-C3 reveal the defect.

5.3.2 Measurement Site

Each wheel load checkpoint exhibits different physical characteristics due to small differences in the ground below the tracks and the curvature of the track before the checkpoint. These characteristics change the wheel load measurement. Small unevenness in the tracks also manifest themselves as noise or small bumps in the signal. Therefore, we add the site of the wheel load checkpoint as additional feature to enable different predictions based on the origin of the measurement site. Author encode this information as a unary code or a one-hot vector, where every dimension represents a site and is 1 only for measurements from that site. When in the future a new measurement site would be built on the railway network, training data for the new site would need to be collected.

5.3.3 Load

A train with different load, but the same waggons results in different wheel measurements for the same defect types, since the weight of the train plays a significant role how the defect exerts its pressure on the sensors. Another important reason to add information about the load to the feature set arises from the following observation: certain defect classes like nonroundness mostly change the average of a sensor reading, but only marginally affect higher order information. An oval wheel for instance will result in higher load measured by some of the sensors and lower load by others, but will not be detected as a defect wheel by individual load normalized measurements. The mean load of all the sensors, standard deviation over the mean load per sensor and the mean load for each sensor are added to the feature set.

5.4 Classification of Wheel Defects

Recognition and classification of wheel defects add up to gather from a vertical power estimation x of a wheel if a wheel is faulty or not. Numerically, a capacity f(.)either encode the double data that a deformity is available or missing or its imperfection class when we can separate the deformity classification. To accomplish this objective, Author use sets of estimations of wheels to prepare choice capacities for certain deformity types and non-deficient wheels. At that point, we utilize this preparation set of estimations and marks (the sort of imperfection) to automatically discover a capacity that is relied upon to anticipate the defects of wheels not seen during preparation precisely.

5.4.1 Support Vector Machine

One of the most mainstream models to discover such a capacity is Support Vector Machines (SVM). An SVM finds a straight capacity defined by the vector w that maximally isolates the two classes during preparation. It accomplishes this partition by boosting the edge between the two classes' purposes in highlight space, or comparably by limiting the regularized exact risk where we limit the experimental risk over the boundaries (w, b) that encode the hyperplane isolating the two classes. $y_i \in (-1,+1)$ is the name (class enrollment) of the i th model in the preparation set, x_i signifies the element vector of the i th estimations, and $\max(0, 1 - z)$ is the pivot misfortune. Estimations of another wheel x would now be able to be arranged with the accompanying choice principle:

$$y:=\mathrm{sgn}(w^{\top}x+b). \quad (1)$$

This decision rule (1) communicates its reliance simply by a scalar item between loads w and the component vector x. Along these lines, it can demonstrate non-straight choice capacities by supplanting the scalar item with a kernel. A helpful decision is a Gaussian outspread premise kernel function of the $k(x_i,x_j)=\exp(-\gamma\|x_i-x_j\|^2)$ Feature vectors x_i,x_j . We would now be able to communicate the minimization issue above in the double and utilize the kernel trick to learn boundaries αi and get the new classification rule.

$$y=sgn(\sum i=1n\alpha iyik(xi,x)+b) \quad (2)$$

To decide the ideal boundaries for regularization λ also, scale γ, we amplify exactness on cross-validation folds.

5.4.2 Classification with DNN

With the logistic loss function $\log(1+\exp(-yiw^{\top}x))$ we get regularized logistic relapse. This enhancement issue has the bit of leeway that advancement calculations gauge probabilities of the class probabilities in expansion to the twofold marks. Utilizing the softmax function rather than the logistic loss, this advantage can be summed up to a self-assertive number of classes. We will utilize these likelihood gauges through a Soft Max-layer in our DNN to join the yield of multiple classifiers for various estimations of a similar wheel. For a given info and C classes, its log-likelihood for having a place with class i equals to

$$p(v|i)=\log(\exp(vi)\sum Cj=1\exp(vj)) \quad (3)$$

Where(vi)1≤i≤C is the top-layer features of the network. The delicate max function above is not just utilized for DNNs however, likewise in numerous multiclass classification methods. For logistic regression or in dynamical framework assessment with the various model adaptive assessment.

5.5 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. These data sets contain information about different types of defects as described in the following. Author also describe what models and features they will use for the respective data sets in this section.

5.5.1 Models and Features

On the first data set Author compare the Wavelet-SVM with benchmark flat spot prediction methods. They show that it greatly outperforms prior art based on thresholding the dynamical coefficient and also on multiple instance learning with dynamic time warping. The second data set serves to demonstrate that the WaveletSVM can accurately classify all three defect types. Author also compare the performance of the deep learning models on different time series representations

Author use different models and features for different defect classes, as this allows us to model network structure and feature construction adaptively to the effects the defects have on the measurements. As there are no known methods to predict nonroundness or shelling Authors compare to baseline methods on a data set with flat spots (data set 1). To evaluate our Wavelet-SVM on non-roundness and shelling as well Authors use data set 2 to estimate classification performance on all three defect classes. Authors have proposed two different DNNs for defect detection the cyclic permutation network (cyclic DNN) and the MIL-DNN. Authors use the cyclic DNN to predict non-roundness as this is a non-discrete defect type with large-scale effects. We take the maximum probability of defectiveness over multiple inputs. The MIL-DNN is used to predict flat spot on data set 2 as the multiple instance learning setting lends itself nicely to this defect type.

5.5.2 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.

Authors 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

$$dBW(x)=max(x)/\bar{x}$$

where max and $\bar{x}$ refer to the maximum and average value of a sequence of measurements x , respectively

5.5.3 Data Set-2 : Reprofile Events

To generate data for training and testing a classifier that can predict additional types of wheel defects, Authors aggregated the time and date of reprofile events and linked them to railway wagons. Authors 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. Authors then categorized measurements of the wheel load checkpoints of the same wagons around the date of re-profiling. Measurements up to a week before re-profiling were considered defective (according to the class label given by the expert), while measurements up to a week after re-profiling were considered defect free. Using this procedure Authors 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, Authors received 6070 measurements, with 2678 being defective. For the non-roundness defect class, 688 cases out of 920 measurements are defective.

6. RESULTS AND ANALYSIS

For performance evaluation of the methods Authors 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.

6.1 Model Selection and Evaluation

For all the experiments in this section the performance shown is computed on a test set that was not used for training or model/parameter selection. To make the evaluation robust against chance Authors 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. Using less experiments for the DNNs is due to computational reasons and justified by the low standard deviation over repetitions in all experiments <=2%. For the Wavelet-SVM three-fold cross-validation is performed on the training set to find the optimal hyper-parameters of the SVM and the Gaussian rbf kernel with grid-search on an exponentially spaced grid. For the DNN 10% of the training set were set aside as a validation set to benchmark performance online and decide on when to stop training.

As the class proportions for data set 2 are not balanced training and evaluating the classifiers directly on this data would lead to bias and higher classification probability for the over-represented class. It would also make judging accuracy and comparing the methods and data sets hard, as the baseline for random chance would not be 50%. Therefore as a first step in all experiments we re-balance the class proportions of the data sets by randomly over-sampling the smaller class through sampling with replacement. While balanced data sets are useful for comparing methods and data sets, in a real-world setting the true proportion of the classes is important and mistakes for different types of error might have different cost. Therefore Authors recommend to give class probability estimates for each class when implementing such a system and then adapting a threshold for raising an alarm iteratively based on the test performance of the system.

6.2 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]. Krummenacher et al. [27] beat stateof-the art MIL algorithms on this data set and get a classification accuracy of 70% with ellipsoidal multiple instance learning (eMIL). In this study features based on the Global Alignment (GA) kernel for time-series [32], [33] were used. Using the features described with a SVM Authors were able to improve accuracy to 92% (Table I).

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

TABLE I Test Set Performance on Data Set 1

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.

6.3 Data Set-2 SVM

Equipped with our general method of constructing features from multiple wheel vertical force measurements and learning a classifier from them we are now ready to predict other types of wheel defects as well. Authors also evaluate the DNN based method in this section.The SVM classifier 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. The performance on shelling is the best out of the three defect types. This observation can be explained by the fact that the training set for this defect type was by far the largest, so Authors were able to train a classifier with higher accuracy. This defect type also affects the wheel globally, so it is harder to miss for the sensors than a flat spot. To improve the performance on flat spot and non-roundness Authors trained custom deep neural networks and give the results in the next section.

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

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

For the defect type non-roundness, the load normalized features based on the load observed by individual sensors substantially contributed to an increase in accuracy. This effect can be explained by the observation that wheel non-roundness errors do not cause a large variation on the within measurement time series since they are a non-discrete type of wheel defects. They do introduce variations between the different measurements per wheel on the other hand and so features based on averages per measurement sequence are important. We will improve the classification performance for flat spot and non-roundness in the next section by using a custom deep neural network (DNN) that is cyclic-shift invariant for classification of these defect types.

6.4 Data Set-2 Deep Learning

Using the same data set as in the previous section we evaluate the deep learning method on the two defect types flat spot and non-roundness. To simplify the experiments Authors 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.

6.4.1 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.

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

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

6.4.2 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 the sensor with highest probability of observing the defect. The performance of the different methods on the test set shows that MIL is an inadequate model for this type of defect since a wheel with a non-roundness defect can not be reliably identified on the basis of only one sensor measurement. This non-local behavior is in contrast to the challenge of predicting flat spot. Concatenating the sensors as is and not looking at the possible cyclic permutations resulted in training set accuracy similar to the cyclic shift network, but performance on the test set is significantly worse (Table IV). Intuitively ignoring the permutations leads to over-fitting as the measurements in the test set might be shifted arbitrarily.

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

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

In comparison with the Wavelet-SVM the cyclic DNN shows higher accuracy and precision and reduced variance. Unlike the DNN for flat spot Authors only trained the cyclic DNN for non-roundness directly on the 1D time series, as the increase in parameters due to the concatenation of measurements of the sensors prohibited efficient training of the model on the 2D representation.

7. CONCLUSION

Authors 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. Authors design cyclic shift invariant artificial neural networks for the detection of wheel flats and non-round wheels that model the relationship between the measurements inherent to these defects. To evaluate our methods Authors collected two data sets from different sources and demonstrate improved performance for predicting flat spot, shelling and non-roundness.

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. For the prediction of severity scores Authors obtained promising preliminary results on regressing the flat spot length using support vector regression [36] and the wavelet features.

8. REFERENCES

1. R. Müller, D. Leibundgut, B. Stallaert and L. Pesqueux, "Validation of wheel maintenance measures on the rolling stock for reduced excitation of ground vibration", 2013.

2. P. Huber, B. Nélain and R. Müller, "RIVAS—Mitigation measures on vehicles (WP5); experimental analysis of SBB ground vibration measurements and vehicle data" in Noise and Vibration Mitigation for Rail Transportation Systems, Berlin, Germany:Springer, pp. 531-538, 2015.

3. E. Verheijen and F. Elbers, "Future european noise emission ceilings: Threat or solution? A review based on Swiss and dutch ceilings" in Noise and Vibration Mitigation for Rail Transportation Systems, Berlin, Germany:Springer, pp. 71-78, 2015.

4. Y. Li and S. Pankanti, "Anomalous tie plate detection for railroad inspection", *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, pp. 3017-3020, 2012.

5. Y. Li, H. Trinh, N. Haas, C. Otto and S. Pankanti, "Rail component detection optimization and assessment for automatic rail track inspection", *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 760-770, Apr. 2014.

6.X. Gibert, V. M. Patel and R. Chellappa, Deep multitask learning for railway track inspection, 2015, [online] Available: https://arxiv.org/abs/1509.05267.

7.T. Idé, "Formalizing expert knowledge through machine learning" in Global Perspectives on Service Science: Japan, New York, NY, USA:Springer, pp. 157-175, 2016.

8.N. Nenov, E. Dimitrov, V. Vasilev and P. Piskulev, "Sensor system of detecting defects in wheels of railway vehicles running at operational speed", *Proc. 34th Int. Spring Seminar on Electron. Technol. (ISSE)*, pp. 577-582, 2011.

9.T. K. Ho et al., "Signature analysis on wheel-rail interaction for rail defect detection", *Proc. 4th IET Int. Conf. Railway Condition Monitor.*, pp. 1-6, 2008.

10.Y. Jianhai, Q. Zhengding and C. Boshi, "Application of wavelet transform to defect detection of wheelflats of railway wheels", *Proc. 6th Int. Conf. Signal Process.*, vol. 1, pp. 29-32, 2002.

11.W. Badran and U. Nietlispach, "Wayside train monitoring systems: Networking for greater safety", *Eur. Railway Rev.*, vol. 17, no. 4, pp. 14-21, 2011.

12.J. C. Nielsen and A. Johansson, "Out-of-round railway wheels-a literature survey", *Proc. Inst. Mech. Eng. F J. Rail Rapid Transit*, vol. 214, no. 2, pp. 79-91, 2000.

13.J. Nielsen, "Out-of-round railway wheels" in Wheel–Rail Interface Handbook, Sawston,

U.K.:Woodhead Publishing, pp. 245-279, 2009.
14.S. Mallat, A wavelet Tour of Signal Processing, San Diego, CA, USA:Academic, 1999.

15.M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications", *Proc. IEEE*, vol. 84, no. 4, pp. 626-638, Apr. 1996.

16.P. Kumar and E. Foufoula-Georgiou, "Wavelet analysis for geophysical applications", *Rev. Geophys.*, vol. 35, no. 4, pp. 385-412, 1997.

17.A. Skodras, C. Christopoulos and T. Ebrahimi, "The JPEG 2000 still image compression standard", *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36-58, Sep. 2001.

18.I. Daubechies et al., Ten Lectures Wavelets, Philadelphia, PA, USA:SIAM, vol. 61, 1992.

19.C. Szegedy et al., "Going deeper with convolutions", *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1-9, 2015.

20.L. Deng and D. Yu, "Deep learning: Methods and applications", *Found. Trends Signal Process.*, vol. 7, no. 3, pp. 197-387, Jun. 2014.

21.A.-R. Mohamed, G. E. Dahl and G. Hinton, "Acoustic modeling using deep belief networks", *IEEE Trans. Audio Speech Language Process.*, vol. 20, no. 1, pp. 14-22, Jan. 2012.

22.D. Silver et al., "Mastering the game of go with deep neural networks and tree search", *Nature*, vol. 529, no. 7587, pp. 484-489, 2016.

23.A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", *Proc. Adv. Neural Inf. Process. Syst.*, pp. 1097-1105, 2012.

24.Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks", *Proc. 29th AAAI Conf. Artif. Intell. Workshops*, 2015.

25.K. He, X. Zhang, S. Ren and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imageNet classification", *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1026-1034, Dec. 2015.

26.T. G. Dietterich, R. H. Lathrop and T. Lozano-Pèrez, "Solving the multiple instance problem with axis-parallel rectangles", *Artif. Intell.*, vol. 89, no. 1, pp. 31-71, 1997.

27.G. Krummenacher, C. S. Ong and J. Buhmann, "Ellipsoidal multiple instance learning", *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, pp. 73-81, 2013.

28.C. Cortes and V. Vapnik, "Support-vector networks", *Mach. Learn.*, vol. 20, no. 3, pp.

273-297, 1995.
29.M. Imani and U. Braga-Neto, "Optimal gene regulatory network inference using the Boolean Kalman filter and multiple model adaptive estimation", *Proc. 49th Asilomar Conf. Signals Syst. Comput.*, pp. 423-427, 2015.

30.P. S. Maybeck and P. D. Hanlon, "Performance enhancement of a multiple model adaptive estimator", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 4, pp. 1240-1254, Oct. 1995.

31.C. J. V. Rijsbergen, Information Retrieval, Newton, MA, USA:Butterworth-Heinemann, 1979.

32.M. Cuturi, "Fast global alignment kernels", *Proc. ICML*, pp. 929-936, 2011.

33.M. Cuturi, J.-P. Vert, O. Birkenes and T. Matsui, "A kernel for time series based on global alignments", *Proc. ICASSP*, vol. 2, pp. II-413-II-416, 2007.

34.N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929-1958, 2014.

35.I. Sutskever, J. Martens, G. Dahl and G. Hinton, "On the importance of initialization and momentum in deep learning", *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, pp. 1139-1147, 2013.

36.A. J. Smola and B. Schölkopf, "A tutorial on support vector regression", *Statist. Comput.*, vol. 14, no. 3, pp. 199-222, 2004.