The Recent Applications of Machine Learning in Rail Track Maintenance: A Survey

Muhammad Chenariyan Nakhaee[1(✉)], Djoerd Hiemstra[1], Mariëlle Stoelinga[2,3], and Martijn van Noort[4]

[1] Data Science, University of Twente, Enschede, The Netherlands
{m.cnakhaee,d.hiemstra}@utwente.nl
[2] Formal Methods and Tools, University of Twente, Enschede, The Netherlands
m.i.a.stoelinga@utwente.nl
[3] Software Science, Radboud University Nijmegen, Nijmegen, The Netherlands
[4] ProRail, Utrecht, The Netherlands
martijn.vannoort@prorail.nl

Abstract. Railway systems play a vital role in the world's economy and movement of goods and people. Rail tracks are one of the most critical components needed for the uninterrupted operation of railway systems. However, environmental conditions or mechanical forces can accelerate the degradation process of rail tracks. Any fault in rail tracks can incur enormous costs or even results in disastrous incidents such as train derailment. Over the past few years, the research community has adopted the use of machine learning (ML) algorithms for diagnosis and prognosis of rail defects in order to help the railway industry to carry out timely responses to failures. In this paper, we review the existing literature on the state-of-the-art machine learning-based approaches used in different rail track maintenance tasks. As one of our main contributions, we also provide a taxonomy to classify the existing literature based on types of methods and types of data. Moreover, we present the shortcomings of current techniques and discuss what research community and rail industry can do to address these issues. Finally, we conclude with a list of recommended directions for future research in the field.

Keywords: Rail track · Machine learning · Maintenance · Deep learning

1 Introduction

Railway systems are one of the most important means of transportation and play a crucial role in the world's economy [1]. Compared to other means, railways provide a more comfortable experience. Besides, they are more affordable, which make them one the most popular way of commuting. Railway tracks are one of the most important components of railway systems. However, the continuous impact of repetitive passing of trains, high railroad network velocity, axle loads and environmental conditions cause rail deterioration. The presence of even a small flaw in rail tracks might introduce more severe defects and broken rails which can lead to huge maintenance costs and reduce the reliability and availability of the system [2]. But more importantly, broken rail track

S. Collart-Dutilleul et al. (Eds.): RSSRail 2019, LNCS 11495, pp. 91–105, 2019.
https://doi.org/10.1007/978-3-030-18744-6_6

can lead to train derailments which subsequently endanger the safety of the passengers and train crews [3]. For example, over the past decade, around one-third of all railroad accidents in the US have been caused by track related defects [4]. Thus, to avoid risks and system disruptions, rail tracks need to be monitored and maintained regularly [5, 6]. However, railway track maintenance is one of the most expensive maintenance activities in railway engineering. For instance, the estimates reveal that approximately each year half of the maintenance budget in the Netherlands is spent only on railway track maintenance activities [7]. Therefore, to reduce the costs and risk associated with rail track failures and to improve the safety and maintenance operations novel techniques and approaches should be developed and be adopted.

Nowadays due to the rapid technological advances and the extensive deployment of low-cost connected devices and sensors, the industrial Internet of Things (IoT) plays an increasing role in the effective implementation of maintenance strategies across a wide range of industries [8]. The railway industry has also embraced the integration of connected devices, sensors and big data technologies to improve their daily maintenance operations [9]. Over the past two decades, machine learning (ML) has revolutionized a wide range of fields such as computer vision, natural language processing, and speech recognition. With the explosion in the amount of data collected by advanced monitoring devices such as wireless sensor networks or high resolution video cameras which are being widely used to inspect critical railway infrastructure, machine learning is also gaining in popularity to improve the operations and reliability of railway systems, and to minimize the daily maintenance costs and risks [10].

To address this demand from the rail industry, a great deal of research has been done over the past few years and various machine learning models have been employed for condition monitoring of rail tracks. Although the application of machine learning for maintenance has been reviewed in other domains such as machine health monitoring [8] and wind turbines [11], to the best of our knowledge no other paper has surveyed the existing literature on the application of machine learning in the rail track maintenance. The aim of this paper is to provide a thorough literature review on current machine learning techniques used for the condition monitoring of rail tracks while also discussing drawbacks of these methods along with what researchers and industry can do to improve the performance and trustworthiness of existing approaches.

This paper is organized as follows: In Sect. 2, the paper introduces different paradigms of machine learning. Section 3 discuss what kinds of flaws can be observed in rail tracks and which types of tools are utilized to inspect rail defects. Section 4, explores the existing machine learning algorithms used in the context of rail track maintenance. In Sect. 5, we describe the shortfalls of current techniques and present a set of new research directions. Finally, in Sect. 6 we present our conclusion.

2 A Brief Introduction to Machine Learning

An ML algorithm usually defined as an algorithm that can learn the underlying patterns from data without being explicitly programmed by human experts. Supervise learning algorithms are a subset of ML models that can learn to predict a target variable from a set of predictive variables also called as features or attributes. On the other hand, unsupervised learning techniques try to infer the inherent structure or represent the input data into a more compressed and interpretable way without being provided with labeled datasets. For instance, principal components analysis (PCA) which is one of the most widely-used unsupervised techniques, takes a dataset stored as a set of potentially correlated variables and compress the dataset by generating a set of new variables that have no linear correlation. Machine learning techniques can also be divided into shallow algorithms and deep algorithms. The main distinction between shallow and deep learning algorithms is in their level of representation. Shallow learning-based techniques use hand-crafted features, manual feature extraction/selection techniques and algorithms such as Support Vector Machines (SVM) [12], Decision Trees [13] and Random Forests [14] for learning the mapping between predictive variables and the target [8]. Moreover, this set of algorithms often use structured datasets such as tables as an input. For example, a decision tree algorithm incrementally learns a set of decision- rules represented as decision nodes and leaf nodes from a dataset that has multiple rows and columns. At each decision node, the decision tree algorithm splits the observations into smaller subsets based on a feature in the dataset that gives higher homogeneity among observations in each subset. A random forest algorithm is an ensemble of multiple decision trees. In each iteration of a random forest algorithm, a decision tree model is trained on a subset of features and a subset of data samples. Then, the algorithm aggregates the outputs of individual trees to make a prediction. Random forests can be an extremely powerful machine learning technique since they add an extra randomness element to a simple decision tree and they combine the predictions of multiple decision trees.

However, deep learning algorithms rarely require hand-engineered features and they can learn the representation directly from the data (e.g. raw images). For this reason, deep learning is sometimes referred to as “representation learning” [15]. This property partially eliminates the need for feature engineering, which gives deep learning algorithms an edge over shallow learning algorithms. Over the past couple of years, the research community has also taken advantage of deep learning for rail defect inspection and monitoring. Even some researchers believe deep learning may become a potential element in the ultimate fully automated rail inspection systems [6].

Convolutional neural networks (CNN) are a special case of deep artificial neural networks (ANN) which have been especially used for computer vision tasks. In CNN models, the fully connected layers in normal neural networks are replaced by convolutional layers. The main difference between the fully-connected and convolutional layer is that in a convolutional layer each neuron is not connected to all neurons in the previous and next layers and the weights are shared between groups of layers [16]. It has been shown that this difference give CNNs a unique property. The early layers of CNNs store low-level feature like edges and curves, while the last layers of a CNN contain the information about the more complex features such as eyes [17]. This is considered to be an interesting characteristic of CNNs as it gives the CNN the ability to use the knowledge (weights) learnt from solving a problem to solve a new problem, also widely known as transfer learning. For example, the weights of a CNN trained on a very large dataset such as ImageNet database [18] can be used to train a new CNN network for detecting tumors in medical applications [19]. CNNs have been successfully applied to various computer vision problems and even beat both humans and the existing algorithms in tasks such as image classification and object detection [20, 21]. In the following section we can also see a surge in the number of publications that trained CNNs to recognize faults in rail tracks.

Besides CNNs there are other classes of deep learning algorithms which have been widely used in the literature to predict time series data [22]. For example, long short-term memory (LSTM) networks, a variant of recurrent neural networks (RNNs) [23], can learn the long-term temporal dependencies by utilizing special mechanisms called memory cells [24]. Lately LSTM networks have shown promising results in predicting the remaining useful life of industrial equipment using IoT data [25].

3 Rail Track Data

The rail inspection data can differ based on different rail defects and measurement methods. In addition, rail data can be stored as structured, semi-structured and unstructured formats. These differences determine which kind of processing techniques and algorithms are more suited for a certain problem. For instance, rail track data such as records of previous maintenance activities collected by human operators can be stored as a structured table and later used by shallow learning algorithm such as random forests. On the other hand, deep learning algorithms are the natural choice for dealing with unstructured data like images. Therefore, in this section we draw a distinction between different defects and data sources which later will be used in our proposed taxonomy.

3.1 Type of Rail Track Faults

Rail defects can develop and grow in different parts of a railway track and therefore they have been categorized in different ways by the researchers. However, in general, rail track defects can be divided into structural defects and track geometry irregularities [1]. Track geometry defects such as rail misalignments are characterized by undesirable deviation of rail geometric parameters from their designed value. Structural defects describe the structural degradations of rail track components such as rail, ballast and fasteners [26]. However, It should be noted that not only track geometry irregularities are responsible for train accidents and directly impact the safety of the rail network but they can also lead to the birth of structural defects [4, 27]. More information on different geometry defects can be found in [28]. Readers can also refer to [29] to find a more complete overview of different structural rail track defects.

3.2 Rail Inspection Methods and Tools

Numerous non-destructive methods and tools are utilized in the rail industry to inspect the condition of rail tracks and data collection. These techniques include manual inspection, ultrasonic devices, high resolution video cameras, 3D-laser cameras, eddy current inspection, magnetic flux leakage etc. A more comprehensive description and comparison of rail inspection tools and methods can be found in [10, 30]. While each method can be used to detect failures in different parts of the rail track and collect specific information about the condition of the rail, not all of them have been used in machine learning literature. However, in recent years, visual inspection systems and particularly video cameras have become one of the most important and effective inspection tools for automatic and flexible rail track monitoring [2]. Video cameras mounted on specialized trains can capture high-resolution images of rail tracks from different angles. In that case, a large number of images are collected which later can be used to train machine learning algorithms to detect anomalies in the rail track. However, large scale deployment of video cameras can present some technical challenges as they require a key infrastructure for efficient storage and processing of streaming data. For instance, each year video cameras collect roughly 10 terabytes of image data in the Dutch railway system [31]. Moreover the existence of some residuals such as oil and dust which might be present in the collected images can have a negative impact on the performance of machine learning algorithms [32].

4 Machine Learning for Track Defect Detection

In this section, we summarize different machine learning techniques adopted by researchers to help the rail industry overcome its maintenance challenges. The current literature has been divided into two major classes of techniques based on the taxonomy we have presented throughout the paper (Table 1). The first group represents the experiments that were carried out with shallow learning algorithms and the second group specifically includes deep learning-based approaches. Further, Table 1 offers more information on other parts of our taxonomy.

4.1 Shallow Learning-Based Algorithms for Rail Track Maintenance

Before 2012 and when deep learning made its first breakthrough in the field of computer vision by AlexNet [33], researchers mainly used complex features extracted manually from images and then trained a shallow learning algorithm such as SVM for image classification and object detection [15]. Likewise, in classical defect detection literature and before the emergence of deep learning techniques, various feature extraction and transformation techniques such as histogram of oriented gradients (HoG) have been applied to image datasets [34]. For instance, Xia et al. [35] extracted Haar-like features to detect broken fasteners in the railway network by an AdaBoost algorithm. To reduce the dimensionality of the input data, Santur et al. [36] first applied various feature extraction techniques such as PCA, kernel principal component analysis (KPCA), singular value decomposition (SVD) and histogram match (HM) techniques to a dataset which comprised a number of non-defective image and an artificially generated image dataset of non-defective images. Next, they trained a random forest algorithm on a set of extracted features. They concluded that features created by PCA provided the most accurate result. Gao et al. [37] merged three different data sources in what they described as 'combined systems method' which comprised of ultrasonic, eddy current and surface imaging video measurements. Then they fed the features extracted from applying a clustering algorithm on their database to an SVM algorithm for detecting squats. Sadeghi et al. [38] employed four neural network models with each with one hidden layer to predict the defect density of rail tracks which was defined as the fraction of a rail segment that is defective. To train the neural network models they combined various attributes such as track quality index of gauges collected through manual inspection.

In the case of rail geometry defects, using a subset of RAS Problem Solving Competition 2015 dataset, Hu et al. [39] attempted to use an SVM algorithm to forecast when a less severe track defect will develop into a more severe type of defect. Famurewa et al. [40] presented a systematic data methodology for rail condition monitoring which consists of descriptive, diagnostic, predictive, and prescriptive steps. As a part of their descriptive and diagnostic strategy, the authors aimed their attention to detect anomalous patterns in sharp curves using PCA algorithm and data acquired by manual inspection from the Swedish railway network. Jiang et al. [41] proposed a hybrid approach to recognize rolling contact fatigue from data obtained in laser ultrasonic experiments. In their proposed approach, the measurement signals were decomposed into a new set of features using a wavelet packet transform (WPT). Next, to reduce the dimensionality of data and to remove the effect of correlated features. Similarly, to better understand and visualize high dimensional track geometry data into a more compressed representation, Lasisi et al. [4] applied PCA, a well-known dimensionality reduction algorithm, to a dataset of 31 features collected from a section of US Class I railway network. KPCA, a nonlinear variant of PCA technique, was applied to new features. Finally, the output of the KPCA algorithm was used as an input to an SVM model to detect four kinds of surface defects.

Lee et al. [42] made use of artificial neural networks (ANNs) and SVM algorithms to predict track quality index (TQI) based on simulation data generated from various important track parameters such as type of curvatures. They concluded that while the ANN algorithm slightly performs better that the SVM algorithm, the difference between these two algorithms is mostly insignificant. Furthermore, they stated that at least two years of data is required for more stable predictions.

In some real-world cases, the railway defect dataset might consist of only positive (defective) and unlabeled observations which essentially means that the conventional classification metrics cannot be computed accurately. Motivated by this problem, Hajizadeh et al. [43] introduced a new metric called Positive and Unlabeled Learning Performance (PULP) to assess the performance of classifiers on datasets with only defective observations. They tested their proposed metric on a rail vibration datasets using two SVM models and stated that a model with a better PULP performance can detect more failures compared to a model with inferior PUPL performance. In another similar work, Hajizadeh et al. [44] proposed a semi-supervised technique which added the unlabeled observations to the training dataset to improve the balance between the two classes of squat defects and non-defects.

4.2 Deep Learning-Based Algorithms for Rail Track Maintenance

One of the earliest attempts to employ deep learning techniques for rail defect detection was carried out by Soukup et al. [45]. They designed a CNN network with two layers to distinguish defective and non-defective cases using photometric stereo images. Since they had a relatively small dataset, and the methodology appeared to be vulnerable to over-fitting, sparse autoencoders and data augmentation were also used in their experiment to tackle this issue. After the successful implementation of CNNs for rail defect detection, other researchers gradually started to apply CNNs to other image databases. In [46], Gibert et al. applied a CNN network with 4 convolutional layers to a set of manually annotated images collected on US Northeast Corridor and classified rail track materials. Then they used the trained parameters of the CNN model for defect detection and semantic segmentation of railroad ties. As an extension of their previous research and based on their proposed approach in [47] which used an SVM to classify fastener defects, Gibert et al. [6] designed and trained a custom CNN architecture with five convolutional layers on the same dataset to categorize the condition of rail fasteners as missing, broken or good. To make their machine learning model more robust against unusual situations, they also used data augmentation and used re-sampling to add more hard-to-classify images to their training dataset.

To provide a tool for automatic defects detection in rail surface, Faghih-Roohi et al. [34] trained 3 different-sized CNN architectures on a manually labeled image dataset collected from approximately 700 km of rail tracks in the Netherlands. Based on the results of their experiment, they concluded that the deepest architecture outperforms the other two models on the multi-class classification of squat defects. The designed architecture for the medium-sized CNN network proposed in this paper is shown in Fig. 1. Jamshidi et al. [31] also classified squat defects with different levels of severity using a simple CNN architecture and a real-world image dataset. They also assessed the visual growth of a defect and its severity using an image database. However, the interesting contribution of their work is that not only they used image data for squat defect classification but they also analyzed crack growth using data collected from ultrasonic measurements and then combined it with image analysis results to provide a failure risk model.

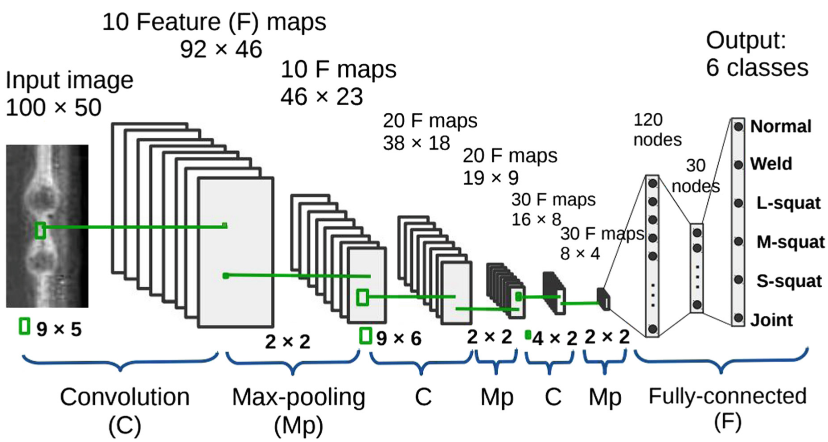

Fig. 1. Illustration of the medium-sized CNN network proposed in [34]

Table 1. An overview of current ML publications for rail track maintenance

Year	Authors	Defect type	ML class	ML algorithm	Data Source
2010	Xia et al. [35]	Structural	Shallow learning	Ada-Boost	Video cameras
2012	Sadeghi et al. [38]	Structural	Deep learning	ANN	Manual inspection
2014	Soukup et al. [45]	Structural	Deep learning	CNN/ Autoencoders	Photometric sensors
2014	Hajizadeh et al. [43]	Structural	Shallow learning	SVM	Video cameras
2015	Gibert et al. [6]	Structural	Shallow learning	SVM	Video cameras
2015	Gibert et al. [46]	Structural	Deep learning	CNN	Video cameras
2016	Hajizadeh et al. [44]	Structural	Shallow learning	SVM	Video cameras
2016	Hu et al. [39]	Geometry	Shallow learning	SVM	Manual inspection
2016	Faghih-Roohi et al. [34]	Structural	Deep learning	CNN	Video cameras
2017	Santur et al. [36]	Structural	Shallow learning	PCA/KPCA/ SVD/HM/ Random forest	Video cameras
2017	Gibert et al. [6]	Structural	Deep learning	CNN	Video cameras
2017	Santur et al. [48]	Structural	Deep learning	CNN	3D-laser cameras
2017	Famurewa et al. [40]	Geometry	Shallow learning	PCA	Manual inspection
2017	Jamshidi et al. [31]	Structural	Shallow learning	CNN	Ultrasonic/ Video cameras
2018	Gao et al. [37]	Structural	Shallow learning	SVM	Ultrasonic/ Eddy current/ Video cameras
2018	Lee et al. [42]	Geometry	Shallow learning	ANN	Simulation
2018	Santur et al. [49]	Structural	Deep learning	CNN	Video cameras
2018	Rauschmayr et al. [50]	Structural	Deep learning	Faster R-CNN/ GAN	Video cameras
2018	Wang et al. [51]	Structural	Deep learning	Pre-trained CNN	Video cameras
2018	Lasisi et al. [4]	Geometry	Shallow learning	PCA	Manual inspection
2018	Jamshidi et al. [27]	Structural	Deep learning	CNN	Video cameras
2018	Ritika et al. [52]	Geometry	Deep learning	Pre-trained CNN	Video cameras
2019	Jiang et al. [41]	Structural	Deep learning	KPCA/ SVM	Laser ultrasonic

In one of their other works, Santur et al. [48] proposed 3D laser cameras as a viable solution for fast and accurate rail inspection. To test their approach they described training a CNN model on data collected through 3D laser cameras to classify rail tracks as either “faulty” or “healthy”. However, the specification of the CNN architecture (e.g. the number of convolutional layers) was not mentioned in their research. However, in their next experiment, Santur et al. [49] used normal video cameras and proposed a three-stage pipeline with a blur elimination step and trained a three-layers CNN model.

As a part of a more comprehensive big data-oriented methodology, Jamshidi et al. [27] in their recent analysis, trained a CNN network on both Axle Box Acceleration (ABA) inspection data and a manually labeled image dataset collected from a specific section of the Dutch rail network. In the other major contribution of this paper, the output of deep learning model, designed to classify the state of rail tracks as a normal, light squat defect and sever squat defect, was later used along with input from analysis of degradation factors and domain experts to define an optimal maintenance strategy.

Lately, the research community has also adopted more advanced deep learning techniques in railway engineering. For instance, to reduce the maintenance expense and enhance the safety of Swiss Federal Railways (SBB) system, Rauschmayr et al. [50] employed several state-of-the-art deep learning algorithms to detect defect and to locate the defective parts on the tracks. First of all, by using a pre-trained faster R-CNN model, they segmented track surfaces and clamps to identify anomalies. Then they made use of Generative Adversarial Networks (GAN) to cluster normal and anomalous observations. In this case, if an observation does not belong to certain clusters, more likely it will be a defect. Further, they discussed the feasibility of this approach as an alternative to replace the manual labeling. Wang et al. [51] also performed an experiment with two well-known deep learning architectures and transfer learning, known as AlexNet and ResNet, to recognize fasteners defects using a hand-annotated image dataset acquired from two separate lines of rails in the US. They concluded that the pre-trained ResNet not only achieved more accurate and reliable results, but it could generalize well on classification of different track lines. To detect geometry defects, Ritika et al. [52] applied several data augmentation techniques to generate artificial images with sun kinks defects. Then, they used a pre-trained Inception V3 CNN architecture to identify sun kinks in rail tracks.

5 Discussion

As one can observe in Table 1, deep learning algorithms have been the most extensively used technique for the detection of structural defects. That has happened thanks to the large-scale usage of video cameras by the industry which subsequently provides the research community with a vast amount of data to experiment with more advanced methods. The table further demonstrates the extensive applications of shallow learning techniques for geometry irregularities. Yet the current state of the literature on the applications of machine learning in rail track maintenance suffers from a few shortcomings. To accelerate the machine learning research progress and machine learning adoption in the railway systems, it is the responsibility of both the research community and the industry to focus on what they can do to address for these shortcomings:

- **Small number of defective observations:** One major property of rail defects datasets is the highly skewed distribution of defective and non-defective classes. In general, a substantial majority of observations belong to the non-defective components while only a slim portion of observations are in fact defective (often less than 1 percent). This can negatively affect the performance of machine learning models as they often favor the majority class [53]. In machine learning literature, various techniques have been proposed to deal with imbalanced datasets. For instance, under-sampling and over-sampling are the two most common approaches used to mitigate the effect of the imbalanced number of classes on training machine learning algorithms [54]. However, in rail maintenance literature only a few number of attempts have been made to address the class imbalance problem or to study the effectiveness of current techniques on rail data. The only known research concerning this issue are carried out by Hajizadeh et al. [44]. Thus, the research community needs to focus more on developing or applying new techniques for overcoming the problem of imbalanced observations in rail defect datasets.

- **Availability of labeled datasets:** The performance of machine learning models heavily depends on the availability and quality of a sufficiently large and labeled dataset. However, while due to the huge amount of measurements most of the time the size of a dataset is not a problem, the presence of enough labeled samples can pose a more serious challenge. Especially this issue becomes more visible in image datasets since manual labeling of the rail track images is a labor-intensive and expensive process, and requires a high level of expertise and domain knowledge. As a result, often the existing datasets cannot satisfy the amount of data needed for machine learning systems. Although several research papers have been published and a few tools have been developed to partially automate the dataset labeling problem, these issues have been overlooked by researchers in the rail domain and in the intelligent maintenance community. So far, only Rauschmayr et al. [50] and Hajizadeh et al. [44] have tried to develop techniques to automatically label rail images.
- **Lack of a public benchmark dataset:** There are several well-known public datasets that have been widely used and studied as a benchmark for comparing different techniques and approaches in other maintenance domains [55]. However, only a few small datasets are available for rail track defects and often the datasets used by researches are proprietary and not sharable. This issue makes training, evaluating and comparing the results of machine learning algorithms more challenging. Thus, as long as there is no public dataset available, not all machine learning researchers outside the rail industry can contribute to the research progress in this domain which subsequently could slow down the progress and stifle the innovation in the domain. Therefore, it is necessary that the rail industry grants the academia access to the rail track data.
- **Explainability of machine learning models:** As mentioned at the beginning of this section, a significant number of papers published in rail maintenance domain exploited CNN models and recommended the use of CNNs for automatic defect detection in real-world scenarios. However, CNNs are considered to be black-box models and are not inherently interpretable. In other words, the machine learning researcher is not able to explain how a CNN model came up with its predictions or prove its trustworthiness to the end user [56]. So far the question of how we can trust ML models has not been addressed by the research community. Therefore, developing accurate black-box machine learning algorithms should not be the only goal but actually how these algorithms classify defects needs to be taken into consideration.

- **Combining domain knowledge with machine learning models:** How defects evolve, which factors contribute to the degradation of rail track components and domain expert knowledge can significantly influence the effective scheduling of rail maintenance operations [27]. For instance, rail track areas with a high concentration of light squats can be fixed by a grinding process. However, if these light squats develop into more severe defects, not only a replacement is needed to fix rail track faults, but the risk of more serious damages also increases [57]. Fault tree analysis (FTA) is a powerful model-based method for risk assessment of complex systems. Fault trees have been used by a vast array of industries, to model how malfunctions in system components lead to the failure of the system [58]. ML techniques can be used together with fault trees to better learn how a system fails [59].

6 Conclusion

This paper has reviewed major machine learning techniques for fault detection. First of all, we have found that especially in the past few years, deep learning algorithms have become the prevailing tool for identifying structural rail defects. Similarly, the results of our survey show that video cameras are the most popular data source for machine learning applications.

However, the current research publications are exposed to a number of shortcomings that we have highlighted throughout our paper. Data quality issues such as highly imbalanced datasets, limitation of manual labeling process and the absence of a comprehensive public database for training and evaluating different approaches is slowing down the progress on the side of research community. The issues related to explaining how an algorithm identifies defects which is absolutely necessary to earn the trust of the industry and incorporating the domain knowledge in ML approaches hinder the progress on the deployment side of ML research. To overcome these shortcomings several research directions and suggestions have been proposed. We believe that the research community needs to focus more on issues including data quality, explainability and trustworthiness of machine learning algorithms and combining the expert knowledge with their machine learning models while the industry should provide the academia the access rail track datasets to facilitate the progress of ML research and to encourage more researchers to contribute and improve the existing methods.

Acknowledgment. This research is supported by ProRail and the Netherlands Organization for Scientific Research (NWO) under the Sequoia project.

References

1. Sharma, S., Cui, Y., He, Q., Mohammadi, R., Li, Z.: Data-driven optimization of railway maintenance for track geometry. Trans. Res. Part C: Emerg. Technol. **90**, 34–58 (2018)
2. Zhuang, L., Wang, L., Zhang, Z., Tsui, K.L.: Automated vision inspection of rail surface cracks: a double-layer data-driven framework. Transp. Res. Part C Emerg. Technol. **92**, 258–277 (2018). https://doi.org/10.1016/j.trc.2018.05.007

3. Liu, X., Saat, M., Barkan, C.: Analysis of causes of major train derailment and their effect on accident rates. Transp. Res. Rec. J. Transp. Res. Board. **2289**, 154–163 (2012). https://doi.org/10.3141/2289-20
4. Lasisi, A., Attoh-Okine, N.: Principal components analysis and track quality index: a machine learning approach. Transp. Res. Part C Emerg. Technol. **91**, 230–248 (2018). https://doi.org/10.1016/j.trc.2018.04.001
5. Durazo-Cardenas, I., et al.: An autonomous system for maintenance scheduling data-rich complex infrastructure: fusing the railways' condition, planning and cost. Transp. Res. Part C Emerg. Technol. **89**, 234–253 (2018). https://doi.org/10.1016/j.trc.2018.02.010
6. Gibert, X., Patel, V.M., Chellappa, R.: Deep multitask learning for railway track inspection. IEEE Trans. Intell. Transp. Syst. **18**, 153–164 (2017). https://doi.org/10.1109/TITS.2016.2568758
7. Zoeteman, A., Dollevoet, R., Li, Z.: Dutch research results on wheel/rail interface management: 2001–2013 and beyond. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. **228**, 642–651 (2014). https://doi.org/10.1177/0954409714524379
8. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X.: Deep learning and its applications to machine health monitoring: a survey, **14**, 1–14 (2016). https://arxiv.org/abs/1612.07640
9. Thaduri, A., Galar, D., Kumar, U.: Railway assets: a potential domain for big data analytics. Proc. Comput. Sci. **53**, 457–467 (2015). https://doi.org/10.1016/j.procs.2015.07.323
10. Li, Q., Zhong, Z., Liang, Z., Liang, Y.: Rail inspection meets big data: methods and trends (2015)
11. Stetco, A., et al.: Machine learning methods for wind turbine condition monitoring: a review. Renew. Energy. **133**, 620–635 (2018). https://doi.org/10.1016/j.renene.2018.10.047
12. Widodo, A., Yang, B.S.: Support vector machine in machine condition monitoring and fault diagnosis. Mech. Syst. Signal Process. **21**, 2560–2574 (2007). https://doi.org/10.1016/j.ymssp.2006.12.007
13. Sun, W., Chen, J., Li, J.: Decision tree and PCA-based fault diagnosis of rotating machinery. Mech. Syst. Signal Process. **21**, 1300–1317 (2007). https://doi.org/10.1016/j.ymssp.2006.06.010
14. Cerrada, M., Zurita, G., Cabrera, D., Sánchez, R.V., Artés, M., Li, C.: Fault diagnosis in spur gears based on genetic algorithm and random forest. Mech. Syst. Signal Process. **70–71**, 87–103 (2016). https://doi.org/10.1016/j.ymssp.2015.08.030
15. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015). https://doi.org/10.1038/nature14539
16. LeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**, 541–551 (1989)
17. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
18. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**, 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
19. Kim, S., Kim, W., Noh, Y.K., Park, F.C.: Transfer learning for automated optical inspection. In: International Joint Conference on Neural Networks (IJCNN), May 2017, pp. 2517–2524 (2017). https://doi.org/10.1109/ijcnn.2017.7966162
20. Karpathy, A.: What I learned from competing against a convnet on imagenet (2014). http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet
21. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks, pp. 1–14 (2017). https://doi.org/10.1109/CVPR.2018.00745

22. Lipton, Z.C., Kale, D.C., Elkan, C., Wetzel, R.: Learning to Diagnose with LSTM Recurrent Neural Networks. 1–18 (2015). https://arxiv.org/abs/1511.03677
23. Lipton, Z.C., Berkowitz, J., Elkan, C.: A critical review of recurrent neural networks for sequence learning, pp. 1–38 (2015). https://arxiv.org/abs/1506.00019
24. Hochreiter, S., Urgen Schmidhuber, J.: Ltsm. Neural Comput. **9**, 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
25. Zhang, W., et al.: LSTM-based analysis of industrial iot equipment. IEEE Access. **6**, 23551–23560 (2018). https://doi.org/10.1109/ACCESS.2018.2825538
26. Ghofrani, F., He, Q., Goverde, R.M.P., Liu, X.: Recent applications of big data analytics in railway transportation systems: a survey. Transp. Res. Part C Emerg. Technol. **90**, 226–246 (2018). https://doi.org/10.1016/j.trc.2018.03.010
27. Jamshidi, A., et al.: A decision support approach for condition-based maintenance of rails based on big data analysis ☆. Transp. Res. Part C **95**, 185–206 (2018). https://doi.org/10.1016/j.trc.2018.07.007
28. Soleimanmeigouni, I., Ahmadi, A., Kumar, U.: Track geometry degradation and maintenance modelling: a review. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit. **232**, 73–102 (2018)
29. Alahakoon, S., Sun, Y.Q., Spiryagin, M., Cole, C.: Rail flaw detection technologies for safer, reliable transportation: a review. J. Dyn. Syst. Meas. Control **140**, 020801 (2017). https://doi.org/10.1115/1.4037295
30. Papaelias, M.P., Roberts, C., Davis, C.L.: A review on non-destructive evaluation of rails: state-of-the-art and future development. Proc. Inst. Mech. Eng. **222**, 367–384 (2008). https://doi.org/10.1243/09544097JRRT209
31. Jamshidi, A., et al.: A big data analysis approach for rail failure risk assessment. Risk Anal. **37**, 1495–1507 (2017). https://doi.org/10.1111/risa.12836
32. Santur, Y., Karaköse, M., Akın, E.: Condition monitoring approach using 3D-modelling of railway tracks with laser cameras (2017)
33. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105. Curran Associates, Inc. (2012)
34. Faghih-Roohi, S., Hajizadeh, S., Nunez, A., Babuska, R., De Schutter, B.: Deep convolutional neural networks for detection of rail surface defects. In: Proceedings International Joint Conference on Neural Networks (IJCNN), October 2016, pp. 2584–2589 (2016). https://doi.org/10.1109/ijcnn.2016.7727522
35. Xia, Y., Xie, F., Jiang, Z.: Broken railway fastener detection based on adaboost algorithm. In: Proceedings - 2010 International Conference Optoelectronics and Image Processing, ICOIP 2010, vol. 1, pp. 313–316 (2010). https://doi.org/10.1109/icoip.2010.303
36. Santur, Y., Karakose, M., Akin, E.: Random forest based diagnosis approach for rail fault inspection in railways. In: 2016 National Conference on Electrical, Electronics and Biomedical Engineering (ELECO) (2017)
37. Gao, S., Szugs, T., Inspection, E., Ahlbrink, R.: Use of combined railway inspection data sources for characterization of rolling contact fatigue (2018)
38. Sadeghi, J., Askarinejad, H.: Application of neural networks in evaluation of railway track quality condition. J. Mech. Sci. Technol. **26**, 113–122 (2012). https://doi.org/10.1007/s12206-011-1016-5
39. Hu, C., Liu, X.: Modeling track geometry degradation using support vector machine technique (2016)

40. Famurewa, S.M., Zhang, L., Asplund, M.: Maintenance analytics for railway infrastructure decision support. J. Qual. Maint. Eng. **23**, 310–325 (2017). https://doi.org/10.1108/JQME-11-2016-0059
41. Jiang, Y., Wang, H., Tian, G., Yi, Q., Zhao, J., Zhen, K.: Fast classification for rail defect depths using a hybrid intelligent method. Optik (Stuttg). **180**, 455–468 (2019). https://doi.org/10.1016/j.ijleo.2018.11.053
42. Lee, J.S., Hwang, S.H., Choi, I.Y., Kim, I.K.: Prediction of track deterioration using maintenance data and machine learning schemes. J. Transp. Eng. Part A Syst. **144**, 4018045 (2018). https://doi.org/10.1061/JTEPBS.0000173
43. Hajizadeh, S., Li, Z., Dollevoet, R.P.B.J., Tax, D.M.J.: Evaluating classification performance with only positive and unlabeled samples. In: Fränti, P., Brown, G., Loog, M., Escolano, F., Pelillo, M. (eds.) S+SSPR 2014. LNCS, vol. 8621, pp. 233–242. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44415-3_24
44. Hajizadeh, S., Núñez, A., Tax, D.M.J.: Semi-supervised rail defect detection from imbalanced image data. IFAC-PapersOnLine. **49**, 78–83 (2016). https://doi.org/10.1016/j.ifacol.2016.07.014
45. Soukup, D., Huber-Mörk, R.: Convolutional neural networks for steel surface defect detection from photometric stereo images. In: Bebis, G., et al. (eds.) ISVC 2014. LNCS, vol. 8887, pp. 668–677. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-14249-4_64
46. Giben, X., Patel, V.M., Chellappa, R.: Material classification and semantic segmentation of railway track images with deep convolutional neural networks. In: Proceedings International Conference Image Processing ICIP, December 2015, pp. 621–625 (2015). https://doi.org/10.1109/icip.2015.7350873
47. Gibert, X., Patel, V.M., Chellappa, R.: Robust fastener detection for autonomous visual railway track inspection. In: 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 694–701 (2015)
48. Santur, Y., Karaköse, M., Akin, E.: A new rail inspection method based on deep learning using laser cameras (2017)
49. Santur, Y., Karakose, M., Akin, E.: An adaptive fault diagnosis approach using pipeline implementation for railway inspection. Turk. J. Electr. Eng. Comput. Sci. **26**, 987–998 (2018). https://doi.org/10.3906/elk-1704-214
50. Rauschmayr, N., Hoechemer, M., Zurkirchen, M., Kenzelmann, S., Gilles, M.: Deep Learning Of Railway Track Faults using GPUs Swiss Federal Railways (SBB) Swiss Center for Electronics and Microtechnology (CSEM) (2018)
51. Dai, P., Du, X., Wang, S., Gu, Z., Ma, Y.: Rail fastener automatic recognition method in complex background. In: Tenth International Conference Digital Image Processing (ICDIP) 2018, vol. 314, p. 1080625 (2018). https://doi.org/10.1117/12.2503323
52. Ritika, S., Rao, D.: Data augmentation of railway images for track inspection (2018)
53. García, V., Sánchez, J.S., Mollineda, R.A.: On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowl.-Based Syst. **25**, 13–21 (2012). https://doi.org/10.1016/j.knosys.2011.06.013
54. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**, 1263–1284 (2009). https://doi.org/10.1109/TKDE.2008.239
55. Nectoux, P., et al.: PRONOSTIA : an experimental platform for bearings accelerated degradation tests. In: IEEE International Conference on Prognostics and Health Management, pp. 1–8 (2012)
56. Ribeiro, M.T., Guestrin, C.: Why should I trust you ? Explaining the predictions of any classifier (2016)

57. Jamshidi, A., Nunez, A., Li, Z., Dollevoet, R.: Maintenance decision indicators for treating squats in railway infrastructures. In: Transportation Research Board 94th Annual Meeting (2015)
58. Ruijters, E., Stoelinga, M.: Fault tree analysis: a survey of the state-of-the-art in modeling, analysis and tools. Comput. Sci. Rev. **15–16**, 29–62 (2015). https://doi.org/10.1016/j.cosrev.2015.03.001
59. Nauta, M., Bucur, D., Stoelinga, M.: LIFT: learning fault trees from observational data. In: McIver, Annabelle, Horvath, Andras (eds.) QEST 2018. LNCS, vol. 11024, pp. 306–322. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99154-2_19