

# Predictive Maintenance for Railway Domain: A Systematic Literature Review

—MARIO BINDER

Institute of Industrial Management,  
FH Joanneum University of Applied Sciences,  
8605 Kapfenberg, Austria

—VITALIY MEZHUYEV 

Institute of Industrial Management,  
FH Joanneum University of Applied Sciences,  
8605 Kapfenberg, Austria

Senior Member, IEEE

—MARTIN TSCHANDL 

Institute of Industrial Management,  
FH Joanneum University of Applied Sciences,  
8605 Kapfenberg, Austria

(Corresponding author: Vitaliy Mezhuyev.)

IEEE DOI 10.1109/EMR.2023.3262282

**Abstract**—Railways are considered to be an environmentally friendly and efficient means of transport for people and goods with increasing importance in the transport policies of many countries. However, the infrastructure and the substantial demand for maintenance create additional costs for railway operators. To overcome outdated maintenance modes, implementation of new solutions, optimization of maintenance activities, and resource utilization are required. Through a systematic literature review, this article evaluates new approaches toward implementing predictive maintenance in the railway domain. A comprehensive search, including the IEEE Xplore, Science Direct, and ACM Digital Library, has been conducted, focusing on papers related to predictive maintenance and railway systems, published in peer-reviewed journals since 2016. The selected papers were analyzed and grouped to allocate the research purposes as well as the considered assets, components, predicted defects, and maintenance conditions. Furthermore, the utilized predictive maintenance algorithms and their limitations are structured and evaluated. Analysis shows that a great variety of algorithms were used for either defect detection or the prediction of conditions of 20 different components, which are critical for the safety and availability of railway operations. The study shows that the proposed approaches were successfully tested and yielded great potential for predictive maintenance solutions. Researchers state to enhance proposed solutions within their future work, increasing accuracy and performance and widening the area of application in the railway domain.

**Key words:** Predictive maintenance, railway, systematic literature review (SLR)

## I. INTRODUCTION

THE goals of a company depend on the influence of its shareholders and the bargaining power of its stakeholders, including employees, customers, business partners, and many more. But the long-term goal of any business operating in the markets is profit maximization. This point of view is strongly criticized by proponents of the stakeholder approach, often resulting in adaptations to include social and ecological goals to avoid a potential conflict [1]. While maintaining social stability and ecological balance may be increasingly important for businesses, the main goal is to maximize the quality

and quantity of products and services while minimizing costs and optimizing profitability [1]. It is crucial to ensure a high degree of availability with minimum downtime and efficiency in compliance with the required quality standards when optimizing a production facility. This implies the need for effective and efficient maintenance strategies to achieve this goal without wasting precious time and resources [2].

Maintenance and its associated costs have a huge impact on the total production costs, varying from 15% to 60% depending on the industrial sector. A closer look at the effectiveness of maintenance

indicates that about one-third of the money spent is wasted, impacting the cost efficiency and profitability of companies. A direct impact on cost efficiency is the loss in performance through maintenance activities, hindering production and thereby decreasing the efficiency of the whole plant [3].

Historically, maintenance was viewed as a necessity to keep systems running and therefore not much effort was put into considering maintenance strategies, as it is the nature of running machines to be repaired and replaced from time to time. Throughout the years though, that view has changed drastically. Where in the past, systems were run until maintenance became so critical that further operations were no longer possible, today’s emphasis is on maintaining availability, minimizing costs, and optimizing maintenance resources [4].

Due to the high potential for cost savings and improving reliability, strategies for maintenance, replacing parts, and inspection have rapidly developed in the last decades [5]. The latest leap forward is the implementation of Industry 4.0 tools and methodologies into maintenance strategies, such as advanced analytics, Big Data, and the Internet of Things (IoT) [6].

Maintenance, as defined by DIN 13306 2019 standard, includes all technical administrative and management actions during the life cycle of an object with the goal to keep it in a state, where its required functionalities are fulfilled [45]. Maintenance approaches can be categorized into two groups: those which do not utilize sensing and computing technology are grouped into reactive maintenance, proactive maintenance, and preventive maintenance, and the approaches with sensing and computing technology are divided into condition-based maintenance, predictive

maintenance, and prescriptive maintenance [46].

The term knowledge-based maintenance describes a system-oriented, comprehensive approach that focuses on the identification of critical elements and analyses possible measures and their potential effects on results. Maintenance strategies, following a knowledge-based approach, do not only focus on keeping a needed functionality of an available object but include the perspectives of maintenance management, plant condition, and economic consequences. Knowledge-based maintenance can be divided into four maturity levels (see Fig. 1): descriptive maintenance, focusing on what happened; diagnostic maintenance, analyzing the cause of failure; predictive maintenance, determination of what faults will happen and when; and prescriptive maintenance, which guides how maintenance should be carried out [47].

An industry that requires substantial and increasing amounts of maintenance is the railway domain. Higher speeds for passenger transport and greater tonnages of freight require more frequent maintenance, to keep systems reliable and avoid breakdowns. Currently, maintenance in the railway industry often lacks optimization

measures, because the only factors considered are tonnage, time, and predetermined standards, which were historically developed within the railway company [7].

*Related work.* There are several systematic literature reviews (SLR) devoted to the applications of data-driven approaches, artificial intelligence (AI), and specifically machine learning (ML) for the railway domain. There are reviews of public datasets for railway applications [48], Industry 4.0 technologies, applied to the rail transportation industry [49], resilience in transportation systems [50], the effectiveness of safety management systems in transport [51], adoption of ML for failure prediction in industrial maintenance [52] and more general, SLR for AI applications in railway systems [53] and SLR on AI models and methods in automotive manufacturing [54]. Two specific surveys on data-driven predictive maintenance for the railway industry [55] and data-driven models for predictive maintenance of railway tracks [56] do not consider specific railway assets, components, predicted defects, and maintenance conditions, which are critical for the safety and availability of railway operations. The algorithms used in the proposed solutions and their limitations were also outside the scope of the previous research.

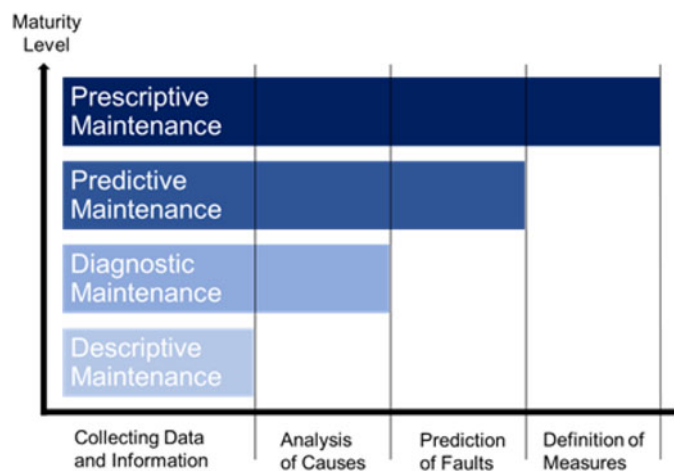


Fig. 1. Levels of knowledge-based maintenance.

This article analyses different approaches to predictive maintenance for the railway domain through an SLR by answering the following research questions.

- RQ1. *What is the research purpose, the study was conducted for?*
- RQ2. *What components (or assets) and which failure modes (or defects) are considered?*
- RQ3. *Which data sources and what algorithms are used for the proposed solutions and what results were achieved?*
- RQ4. *What are the limitations of the proposed solutions and what are potential fields of interest for future research?*

The rest of this article is structured as follows. Section II describes the methodology of SLR including the definition of criteria to evaluate the relevancy of the papers, the scientific databases used, and the general conditions concerning the selection of papers. Each of the following chapters is dedicated to answering the corresponding research questions. Finally, Section VII concludes the article based on the findings and gives further directions for research.

## II. METHODOLOGY OF SYSTEMATIC LITERATURE REVIEW

A literature review is a process of evaluating, condensing, and critically reviewing a specified selection of documents in a certain field [8]. O'Leary [9] stated that the purpose of a literature review can be composed of three main points: to inform the reader about new findings, developments, and conclusions within the field; to establish research credibility through critical evaluation of the review literature; and to

highlight potential gaps or evaluate used methodologies.

An SLR combines a systematic process of comprehensive search with the strength of a critical review. SLR has become increasingly important with the enlarged number of publications available on the Internet [8], [10].

The comprehensive approach of the SLR methodology ensures that evidence is gathered through the review of many different sources, to confirm that a controlled picture of the research topic is provided. But without a clearly defined and strictly conducted process of exclusion and inclusion of literature, the result may still be subjective and may support a predefined line of argument [11].

### A. Process of Systematic Literature Review

An SLR follows a defined model with different steps that provide structure to the process. This article follows the approaches described in [8], [10], and [12] by condensing them into one process, covering the main steps of all referenced guidelines.

1) *Scope of Research:* The first step is the identification of the area of research through scoping activities. It includes databases, scanned through a keyword search for available sources of potential interest for the study. The reason for this is to identify knowledge gaps and evaluate the availability of relevant literature. In this phase, the research questions are defined and the search protocol with a compiled list of keywords is prepared. Inclusion and exclusion criteria are defined to narrow down the number of relevant papers. In addition, a data sheet for the extraction and collection is designed.

2) *Comprehensive Search:* Based on the previously defined search criteria and keywords, a comprehensive search in scientific

databases is conducted and the result is documented. At this point, the first refinements to the initial keywords and search criteria can be made based on the results of the first searches.

### 3) Literature Selection:

All papers found and documented in the comprehensive search are evaluated based on predefined criteria to exclude or include them in the further process. This ensures that only literature relevant to the research is included in the detailed part of the analysis and that sources of bad quality are excluded. Depending on the number of sources, a refinement of the inclusion and exclusion criteria can be done to widen or narrow the point of view.

### 4) Data Collection:

After all improper sources are eliminated, consistent data are extracted from the selected ones. The data are organized in tables, structured in line with the formulated research questions.

### 5) Synthesis:

The extracted data are condensed into a single paper connecting all sources to find an answer to research questions.

### 6) Conclusion:

Based on the findings described in the synthesis, recommendations for further research are made and conclusions are drawn.

## B. Scope of Research

### 1) Definition of Search Criteria:

The specified keywords in Table 1 are used in a search string defined as "A AND B" to scan the selected databases for relevant papers. The terms were purposely chosen to be

A	B
Predictive maintenance	Railway
Machine learning	Rail

very specific to reduce the number of papers that can be considered relevant upfront.

Predictive maintenance is often described in a relation to ML; therefore, this term is included as a keyword to not miss potentially relevant literature sources. The term rail was used as a synonym in addition to the term railway.

2) *Definition of Inclusion and Exclusion Criteria:*

To narrow down the results, the inclusion and exclusion criteria displayed in Table 2 are applied.

**C. Comprehensive Search** A comprehensive search was conducted by scanning digital libraries represented in Table 3. The number of papers found in each library with the corresponding search string is listed.

The advanced features of the search engines were used to select papers, published from 2016 onward.

**D. Literature Selection** The identified sources were cleared from duplicates and screened based on the title and abstract. Afterward, the remaining sources were analyzed for their eligibility for the research questions. Two reviewers independently screened each record.

The results of the whole process, beginning with the comprehensive search and ending with the final papers that are included in this research, are summarized in a PRISMA diagram [13] (see Fig. 2).

**E. Data Collection** Despite the initial number of 1264 papers recorded by the comprehensive search, only 24 of them focus on predictive maintenance or closely related topics within the railway domain. The papers were assessed by two reviewers independently as being relevant to this research and

the key information was collected within a table. Besides general information, based on the research questions records were collected (see Table 4).

The following chapters answer the specific research questions. Section III determines the research purpose. Section IV describes the components and their failure modes. Section V elaborates on the data sources and utilized algorithms. Section VI analyzes limitations and future work.

All collected records were compatible with the selected measures of analysis. In several cases, the limitation or the future work section was absent, which does not limit the generality of the analysis.

**F. Bias Assessment and Result Synthesis** Two reviewers worked independently to assess the risk of bias in the included studies. To decide if a study was eligible for synthesis, each study was compared based on the metrics represented in Table 4. The method of quantitative data synthesis was used. Tabulation and visual display of the results were specified in the protocol alongside the synthesis. In addition to the diagrams, the tag cloud representation was used.

**III. RESEARCH PURPOSE**

The reviewed papers state different reasons why improvements and new approaches are necessary for the railway sector. While the papers provide a variety of different aspects, the motivation-related statements can be clustered into three main groups (see Table 5), relating to economic (acknowledged in 20 reviewed papers), safety (14 papers), and technical (11 papers) concerns.

**A. Economic Concerns** The major concern is related to economic losses and potential cost reduction through the implementation of new approaches. Economic concerns mainly focus on the reduction of costs by decreasing resource consumption in maintenance activities, increasing efficiency, or decreasing demand. This can be achieved through an extended life cycle of parts which are subject to regular wear and tear or simply through lowering costs related to accidents, sudden breakdowns, and interruptions of operation.

1) *Traffic Interruption and Availability:*

One of the main issues with a huge economic impact is interruptions of operation, caused by accidents,

**Table 2. Inclusion and Exclusion Criteria.**

Phase	Inclusion/Exclusion criteria
1	Date of the publication since 2016
2	Were published in a peer-reviewed journal
3	Not a duplicate
4	Published in the English language
5	Full text is accessible

**Table 3. Digital Libraries and Initial Results of Comprehensive Search.**

Library	Search String	Num. of Papers
IEEE Xplore <a href="https://ieeexplore.ieee.org/">https://ieeexplore.ieee.org/</a>	((“predictive maintenance” OR “machine learning”) AND (“railway” OR “rail”))	443
Science Direct <a href="https://www.sciencedirect.com/">https://www.sciencedirect.com/</a>	Title, abstract, keywords: (“predictive maintenance” OR “machine learning”) AND (“railway” OR “rail”)	84
ACM Digital Library <a href="https://dl.acm.org/">https://dl.acm.org/</a>	[[All: “predictive maintenance”] OR [All: “machine learning”]] AND [[All: “railway”] OR [All: “rail”]]	737

planned or unplanned maintenance activities. Railway operators need to avoid these interruptions, as they reduce the availability of the railway network, causing delays and leaving customers unsatisfied. Furthermore, sudden interruptions are costly, as urgent repairs and maintenance often come at a higher cost compared to scheduled activities. Every so often the sites need to be revisited for extensive repair work, as the first solution is often only provisional to allow an operation to continue. Hence a reliable and properly maintained railway network is key to customer satisfaction and avoiding economic loss through traffic disturbances [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24].

### 2) Increasing Efficiency:

Increasing efficiency while decreasing the waste of resources is another proposed solution. Authors criticize that commonly applied time-based maintenance strategies are inefficient, as they consume a lot of resources and do not analyze information to determine

an actual demand. The usage of modern information technologies would allow for more accurate resource allocation to avoid the issues related to the modes of corrective and preventive maintenance, conducting maintenance activities either too late or probably too early [15], [20], [25], [26], [27].

Inspections are one of the main topics criticized in the current strategies, because they, in most cases, rely on human labor to perform the tasks. These inspections need to be done by trained experts and the tiring nature of such tasks makes them prone to mistakes [19], [23], [26], [28], [29].

One of the reviewed papers highlights the potential issues of proposed solutions, which require additional equipment to be installed, i.e., efforts, generating additional costs [27].

### 3) Affecting Other Components:

Researchers propose solutions for detecting defects on wheels and current collector strips as well as

ballast stabilization and prediction of rail pad conditions, as damage to these components significantly affects the condition of other critical parts. For example, defective wheels cause increased vibrations, which leads to more wear on rails and components of the axle bogie. Worn current collector strips may cause damage to the catenary system and ballast degradation; elastic rail pads strongly influence track stiffness, a key parameter to determine a railroad's overall condition. In conclusion, properly maintaining these components indirectly conserves the condition of others, therefore reducing overall demand and costs [30], [31], [32], [33].

**B. Safety Concerns** Concerns about safety and related topics are the second most important reason to propose new solutions. The safety concerns mainly focus on accidents, their causes, and how to prevent them through defect detection and prediction. The role of humans, as a

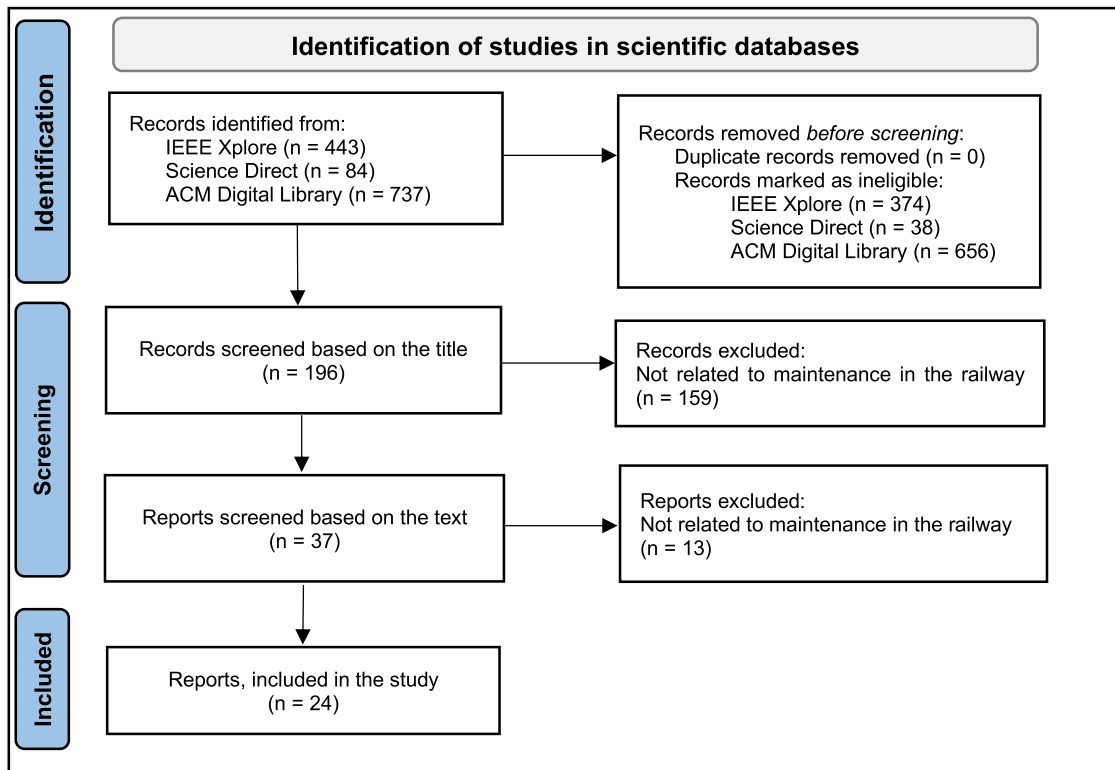


Fig. 2. PRISMA diagram.

potential source of mistakes and causes of accidents, is a widely discussed matter of concern.

1) *Avoiding Accidents:*

Severe defects in railway components and assets threaten operational safety, thus it is important to propose new approaches for more precise detection and prediction in the railway. Defects of infrastructure components are the predominant cause of accidents, second only to human error.

Derailments are a common threat that can occur under various circumstances, e.g., induced by poorly maintained track geometry, because of layer settlement or degradation of the ballast, or broken rails. Rail breaks are frequent, as they develop from cracks and other weak points within the rails, such as rail welds, which can be found within an average distance of every 25–100 m in continuously welded tracks [14], [16], [17], [21], [22], [25], [26], [29], [32], [34], [35].

2) *Human Mistakes:*

Visual inspections, done by personnel, are by nature subjective and therefore prone to mistakes. Undetected defects may grow and develop into more serious ones, posing a potential threat to the safety of operation, which is not the only concern, as personnel

patrolling the tracks is exposed to danger from passing trains, the high voltage within the catenary system, and threats from various other potential sources [19], [28].

**C. Technical Concerns** These concerns are mainly related to an increase in vibrations and noise emissions, the implementation of technological advances, and the discovery of new failure modes.

1) *Increased Emissions and Affecting Comfort:*

Malfunctions of railway tracks, as well as damage to the wheels, increase vibration, which is not only a concern related to increased wear of components but also to noise emissions, which can be a disturbance, especially in urban areas, for people living close by. In addition, higher vibration and noise levels reduce the ride comfort of passengers, decreasing customer satisfaction [29], [30], [32], [33], [38].

2) *Technical Advances:*

New developments and greater accessibility of technologies as well as successful demonstrations of new techniques allow the enhancement of existing systems and grant the possibility to develop new

approaches. Improvements in sensor techniques in combination with increasing demand for automation encourage the development of faster and more precise defect detection and prediction models. In addition, approaches focusing on defect detection based on vibration signals have been successfully deployed in other industries. Furthermore, existing approaches demand additional equipment to be installed, not only increasing the cost but also the barrier to wider adoption [16], [24], [27], [34], [37], [38].

3) *New Failure Modes:*

A reviewed paper mentioned an increased complexity in newly adopted AC/DC combined power systems, with potentially new and unexplored failure modes as a reason for their research motivation. These new systems are still facing unresearched issues due to higher complexity, leading to increased numbers of potential points of failure, resulting in higher failure rates and even yet unknown failure modes [17].

**IV. ASSETS, COMPONENTS, DEFECTS, AND PREDICTED CONDITIONS**

The papers were analyzed regarding their assets and components addressed, as well as the failure modes and defects. The papers were classified based on their approach to predictive maintenance in three different groups, which are described in Table 6.

The classification can be used to analyze the focus of studies in the field of predictive maintenance for the railway, to get a better understanding of the different proposed approaches.

**A. Considered Components and Assets**

The key components of railway infrastructure are considered by the reviewed papers, focusing on the elements that experience a high amount of stress and wear, thus

**Table 4. Structure of Collected Information.**

Category	Data
General information	Title Author(s) Year of Publication Source Pages
Research information	Related to RQ1: <ul style="list-style-type: none"> <li>● Motivation for research</li> </ul> Related to RQ2: <ul style="list-style-type: none"> <li>● Concerned assets and components</li> <li>● Considered detects, faults, and predicted conditions</li> </ul> Related to RQ3: Data Sources <ul style="list-style-type: none"> <li>● Used Algorithms</li> <li>● Results</li> </ul> Related to RQ4: <ul style="list-style-type: none"> <li>● Limitation</li> <li>● Future Work</li> </ul>

**Table 5. Classification of Research Motivation.**

Classifier	Description
Economic Concerns	The paper mentions economic or cost-related issues, such as inefficiencies or interruption of operation.
Safety Concerns	The paper mentions safety concerns or potential accidents which are reduced or avoided with the proposed approach.
Technical Concerns	The paper mentions issues related to technical improvements or side effects of defects.

**Table 6. Classification of Approaches.**

Classifier	Description
Detection of Defects	The paper focuses on inspection or monitoring techniques for certain components.
Prediction of Defects	The paper focuses on predicting future failures of certain components or assets, based on models and/or existing data.
Maintenance Solution	The paper proposes or develops a unique approach for the implementation of predictive maintenance.

demanding particularly high amounts of maintenance activities.<sup>1</sup>

The rails themselves are the components of railway infrastructure addressed the most within the review studies. They were classified as a safety-critical component and the major cause for derailments in case of failure, being second to humane misbehavior, as well as exerting a high influence on the maintenance costs of railway operators [20], [21], [29], [35].

The train bogies, including different subsystems and components, were the second most addressed component group, being considered within four different papers, as they are safety-critical, significantly affecting reliability and directly influencing the vehicle's motion, as they connect the vehicle body to the rails [16], [18], [34], [38].

The third most considered components are the rail joint and the track geometry. The reason that rail joints were addressed is that they are the weakest parts of the tracks, and in case of a break, they can lead to derailments. In

<sup>1</sup>Maintenance activities describe all actions, necessary to maintain an item, including inspection, condition monitoring, compliance test, overhaul, fault diagnosis, fault localization, restoration, repair, task preparation and scheduling [45].

addition, fully welded tracks consist of 25–100 m long rails, which leads to high numbers of rail joints and makes them a common point of failure [14], [25]. The track geometry was considered, as the geometry defects cause a significant number of derailments, hence significantly contributing to railway accidents and costs associated with properly maintaining it, to avoid those accidents [20], [24].

Other components which were addressed within the reviewed papers include railway plugs, the catenary system, power equipment, power systems of trains, trackside equipment, current collector strips, ballast, railway tunnels, switch machines, rail pads, and railway switches. To offer a better overview of the considered components and assets, they are visually summarized in Figs. 3 and 4.

## B. Defects and Predicted Conditions

### 1) Detection of Defects:

The core of defect detection is fault diagnosis, which describes the actions taken, to recognize, localize, and identify the fault. This can either be achieved through inspections, including examination for conformity through observing, measuring, or testing an item on its relevant

characteristics or through condition monitoring, describing the activities or tasks to measure the characteristics and parameters of the physical state of an item [45].

Most of the reviewed papers, 14 to be exact, focus on the detection of defects in certain components or assets of the railway infrastructure. Those papers can further be split into two groups, based on the approach they had chosen, either to detect a general kind of defect within a group of components or to detect specific defects and faults on certain components.

a) *General defects* The components, component groups, or assemblies are monitored for general defects, faults, or failures, not detailing the different failure modes of each component, but rather of the whole assembly group. These approaches are either based on vibration emissions, changes in current or voltage, or general changes in the appearance of an object.

Emission-based approaches use vibrations as their source of information and detect defects based on the analysis of these signals in certain scenarios. Any kind of defect, that distort the soundness of moving components, usually increases the vibration output or modifies its frequency [16], [30], [34], [38].

Approaches that are based on voltage or current analysis, determine defects and faults by observing those signals and noticing either certain unusual patterns, such as the grounding of the signal, or slow changes in the baseline of the signal, such as permanently increased current consumption [17], [22].

Visual-based approaches try to mimic the human behavior of inspections. They detect faults and defects through a comparison of visual signals, such as pictures or laser scans. The comparison can either be between the inspected

object and examples representing defects and damaged items or between pictures of the same component, that were taken at different points in time, determining defects through the change of the visual appearance [14], [19], [26], [28].

*b) Specific defects* Compared to the general approaches, the specific ones focus on certain defects with the goal to determine the conditions of defects and faults of individual components. The dominant approaches for the detection of individual defects are based on visual inspection systems (VIS), with one reviewed paper perusing a vibration-emission-centered idea.

The most common approaches for identifying individual defects are VIS,

where camera footage is analyzed for a visually detectable defect. Such systems not only grant the opportunity to find and identify faults but also assess the urgency of required maintenance activities, based on type, size, and number, for precise scheduling. For example, surface discrete defects on rails can be detected through the analysis of images captured by cameras mounted underneath a train, which allows the classification of conditions for defined track sectors [29], [31], [37].

A different approach was proposed by Carboni and Crivelli [18], perusing the detection of fatigue cracks within railway axels based on the sound signature that these cracks emit during their development. This allows the diagnosis of cracks not visible on the surface, at a very early stage.

2) *Prediction of Defects:*

Predictive maintenance is condition-based maintenance that follows a forecast based on known conditions and the analysis of key parameters indicating the degradation of an item. Derived from this definition, the key to predictive maintenance is understanding the significant parameters affecting the conditions of a component or asset to predict future conditions at defined parameters [45].

Based on such parameters prediction of defects is considered in eight reviewed papers. The proposed approaches use various sources as their input, including information on maintenance and inspection logs, operation data as well as general data like track parameters. Furthermore, some papers used models, based on the real-life system, to generate additional data for training

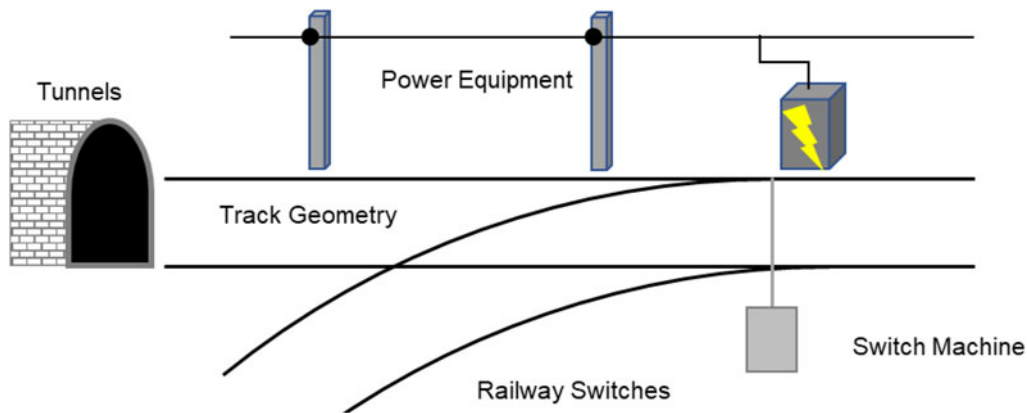


Fig. 3. Components/assets—part 1.

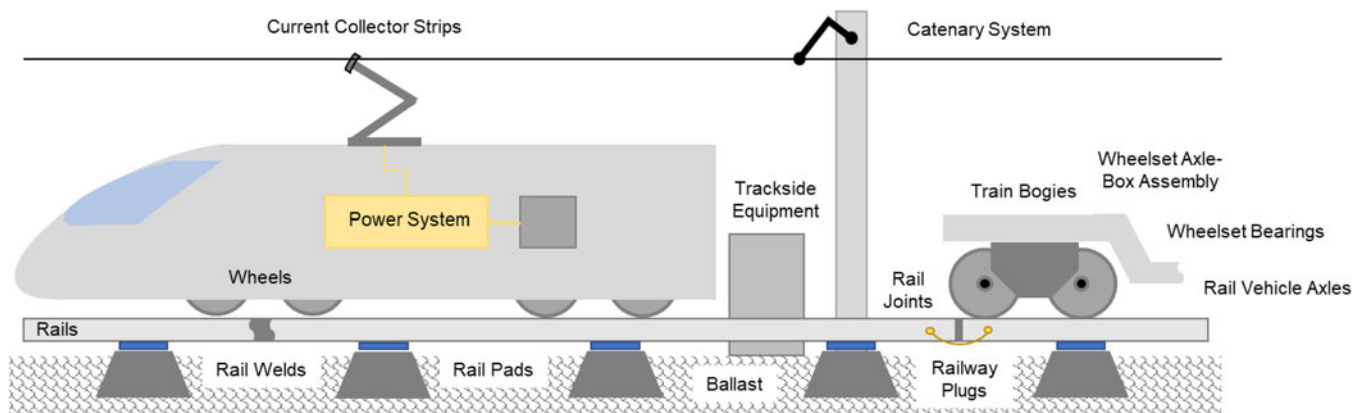


Fig. 4. Components/assets—part 2.



and testing, as quality data was insufficiently available [20], [21], [24], [25], [27], [33], [35].

In comparison to defect detection, either focusing on the detection of general or certain defects, the papers perusing the predictive approach only focus on general defects, mainly to avoid unnecessary complexity of the approach. The reviewed papers focused on the prediction of rail weld defects in continuously welded tracks, failures of power equipment, condition of track geometry, defects in rails, and failures of switch machines [20], [21], [24], [25], [27], [33], [35].

While papers focus on the prediction of defects or faults within certain components, one paper predicted the mechanical properties of rail pads based on their operation condition, including temperature, axle loads, frequency, and toe load. Hence, the proposed approach does not allow for direct defect and fault identification but enables the possibility to predict one of the fundamental parameters for track maintenance, the track stiffness [33].

In addition to the prediction of defects, three of the review papers proposed implementing the classification of the predicted faults, to make the results easier to interpret for maintenance personnel. Two papers proposed a track quality index (TQI) that narrows the results down to a single dimension, which provides information on the conditions within a certain sector. One paper followed the approach of classifying the predicted defects based on their severity for the railway operation into two groups: red defects, demanding immediate maintenance activities, and yellow defects or faults,

which allow operations to continue, but demand maintenance activities later [20], [24], [27].

### 3) *Maintenance Solution:*

Two papers proposed unique approaches for the integration of predictive maintenance solutions. One of them proposed a procedure to extend the time between maintenance activities, hence reducing the maintenance demand. This approach follows the idea that the preemptive slowdown of degradation mechanisms reduces costs upfront. Ballast settlement and particle degradation are the two main mechanisms, induced by the cyclical loads of the traversing trains, that reduce the functionality of the ballast, directly impacting safety and ride comfort [32].

Another paper focused on the implementation of the group of approaches to predictive maintenance, instead of focusing on solutions for individual problems. It explores utilizable ways to implement predictive maintenance solutions and combine them into a whole system, solving multiple maintenance-related digitalization issues [23].

## V. DATA SOURCES AND ALGORITHMS

A variety of different algorithms were used within the proposed solutions; several papers tested multiple algorithms to determine the one with the highest accuracy and best overall performance. Furthermore, different kinds of data sources are used to feed the systems, depending on the persuaded method as well as on the

availability of existing data that can be utilized.

These approaches are categorized into three main groups, either using sensor data of an additionally installed device, using data, gathered within existing systems, or utilizing a physical model to study the system (see Table 7).

While different algorithms were used for either detection or prediction, researchers mainly present the results of the best-performing ones. The algorithms that are the core of the proposed solutions are extracted and described in close relation to the achieved results.

**A. Data Sources** Quantitative analysis of the papers regarding the source of data shows that some papers focused on only one source of data, for example, sensor data or existing data, whereas others utilized multiple sources. Thirteen considered papers solely used sensor data, six focused on existing data, three utilized existing data in combination with physical models, and two combined sensor data with existing sources for the solution.

Fig. 5 illustrates the different data sources used by the reviewed papers. The difference between sensor data and existing data is that sensor data are collected solely for the proposed approach while existing data were already collected before for other reasons. Physical models are used to emulate the concerned failure modes and to study defect development, which is utilized within prediction models or for the creation of additional training and test datasets.

The distribution of the reviewed sources is displayed in Fig. 6. Note that many papers utilized multiple data sources, hence the sum of papers using each source is greater than the total number of reviewed studies.

**Table 7. Classification of Used Data.**

Classifier	Description
Sensor Data	The source of data is sensors, commonly providing real-time signals.
Existing Data	The proposed solution utilizes available data from existing sources and systems, which is already collected for other purposes.
Physical Model	A physical model emulating failure modes or defect development is used to create data samples.

1) *Sensor Data:*

Out of the reviewed papers, 14 propose a detection-based approach focused on the utilization of sensors to detect anomalies or patterns that indicate defective behavior of components. The most common are visual data as they are the core of visual-based inspection systems, proposed by eight papers. Others utilize electric signals, such as current or voltage, appearing forces, vibration, and acoustic emission.

a) *Visual data* The reason for the common usage of visual data is that it is the base of VIS, which is extremely

robust against the influence of environmental factors in comparison to other detection and testing methods. Furthermore, VIS are widely used as they allow efficient, precise, and objective defect detection within a large database of acquired images [29], [37].

VIS consists of two main parts, the image acquisition system (IAS), the hardware to capture the image or video footage, and the image processing subsystem including the algorithms to detect defects. IAS generates the data for further analysis [29].

Because pictures captured by digital cameras are the simplest form of

visual data, this approach was used in most of the reviewed papers. Not all papers considered their systems to be implemented for permanent usage and hence do not need a complete IAS. In this case, test and training data were often gathered manually with commonly used digital cameras and a light source to increase visibility and contrast. Several solutions incorporated the installation of digital cameras underneath a train to capture continuous series of pictures at fixed distances [28], [29], [31], [37].

In cases, where the alignment of pictures is important, for example, to compare the same object within different periods, GPS data were utilized to align the images or to find the objects of interest within a continuous collection of images. In these cases, GPS is mounted onto the train, next to the cameras, and every time a picture is taken, GPS data were stored alongside it [26].

One paper proposed a thermography-based method and analyzed thermal images captured by cameras mounted on the train. This allows the interpretation of connections between components and the allocation of small gaps, as those become greatly visible in a thermographic spectrum [14].

Another proposed approach uses technology based on a LiDAR mobile laser scanner, which allows for 3-D reconstruction of the scanned environment and makes it possible to scan large infrastructure objects in short periods. This special process requires the vehicle to follow a predetermined path, alongside which the laser scanner gathered data points with 3-D coordinates and represents the scanned objects within a data cloud [19].

b) *Acoustic and vibration emissions* Three out of the fourteen papers proposed approaches utilize

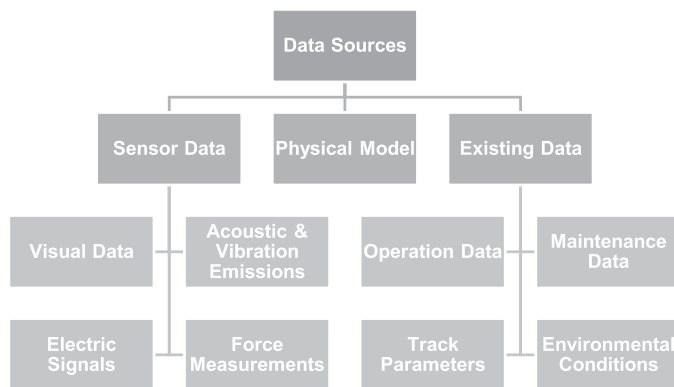


Fig. 5. Classification of data sources.

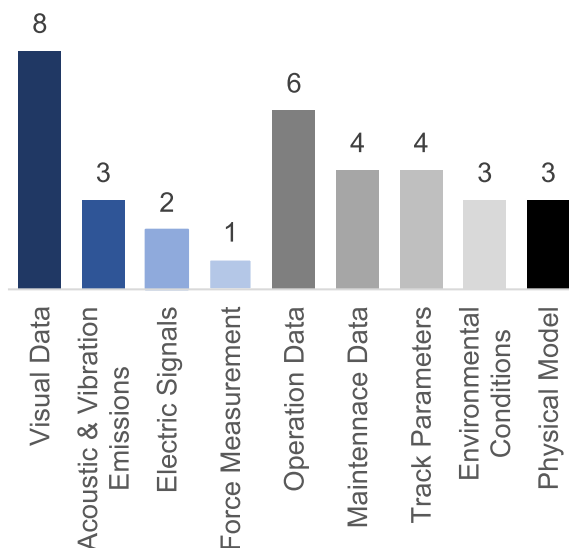


Fig. 6. Distribution of data sources.

sensor data based on measurable emissions, including acoustic and vibration emissions. Within the reviewed papers, these approaches were used for moving objects that are exposed to accelerations, such as axels, bogies, and wheelset bearings.

Vibration sensors mounted near the object of interest, collect vibration data which are analyzed to detect patterns created by defective components. The output of such a sensor may be 1-D or multiple-dimensional. In comparison, a similar approach utilizes multiple 3-D vibration sensors, being able to identify vibration in vertical, lateral, and longitudinal directions. In addition, thermal sensors are mounted on heat-sensitive components, to monitor the temperature development during operation, with a special focus on increases in vibration-induced temperature [16], [34].

Based on acoustic emissions, the defect detection approach has an advantage in that defects can already be detected while they develop. This is based on observations of developing damage, e.g., fracturing or plasticization, which releases energy in the form of bursts of ultrasonic elastic waves, within a certain bandwidth. Therefore, piezoelectric sensors are mounted onto the surface of the monitored object to detect the acoustic emissions of these burst events [18].

*c) Electric signals* Electrical components of the power equipment can conveniently be monitored by measuring existing electric signals, including voltage or current. Two papers focusing on electrical components, proposed such approaches since such can easily be achieved through noninvasive sensors. Time series data of current or voltage signals are collected and analyzed for abnormalities or unique patterns that indicate defects [17], [22].

*d) Force measurements* Kruppenacher et al. [30] proposed a solution, which focuses on the utilization of checkpoints, consisting of four measuring bars with 1 m in between them on both rails. Each of those bars observes traversing vehicles and measures maximum vehicle load, maximum axle load, and load distribution on different contact points of the wheels. The analysis of the time series data collected allows for identifying defective wheels, as common wheel defects like flat spots disturb the homogeneous signal, created by nondefective wheels.

### 2) Existing Data:

Many railway operators gather huge amounts of data with the goal to predict the behavior of their systems. The collected data often yield huge potential for additional applications. Hence, several reviewed papers proposed to utilize existing data, including maintenance logs, defect statistics, operation data, periodically measured track geometry parameters, quality indexes, or data describing the structure of rail segments.

*a) Operation data* Operation data are commonly collected by railway companies through internal processes. The gathered data include the number of wagons, axels, and wheels with their approximate load conditions, including axel load and resulting toe loads on the rails, as well as the frequency of vehicles traversing each track segment and the accumulated traffic load, known as million gross tonnage (MGT) [15], [20], [21], [25], [32].

*b) Maintenance data* Information about maintenance activities is usually collected within maintenance reports or logs that describe the time, position, and tasks carried out by the maintenance crew. Reviewed papers [20], [25], [27], [32] use such data, including information about

performed grinding activities or tamping and renewal of ballast. Furthermore, these reports give information about changed components as well as the type and severity of defects and allow an analysis of the frequency and demand of maintenance activities.

*c) Track parameters* Studies [20], [24], [25], [35] proposed the usage of track parameters, which can be divided into two types. Static parameters offer general information about the layout and structure of the tracks, for example, the number of curves and slopes, the length of segments or sectors, and the number of welding joints. These parameters do not change during operation, compared to others that need to be checked periodically, including gauge, cant, and superelevation, just to name a few. Because many parameters are determining the condition, they are commonly combined into a 1-D TQI, making it easier to be interpreted by maintenance personnel.

*d) Environmental conditions* Studies [14], [25], [33] considered environmental conditions, including temperature as well as daylight as important factors for their approaches. Temperature significantly changes the mechanical properties of components, affecting the stiffness of rail pads and tensions within rails and rail joints, hence being an important parameter when considering these components. Additionally, temperature and daylight information must be considered when sensors are sensitive to changes in those parameters.

### 3) Physical Models:

Several authors reported a lack of data or bad quality of data and chose to overcome those issues by creating physical simulation models. Another approach allows to study of failure modes and increases the precision of

the prediction models through higher amounts of training data, generated by sample generators based on a physical model.

Two papers [21], [32] chose to develop models of the physical process causing the concerned failure modes. A finite element model was used to simulate the development of rail breaks, starting from little cracks within the railheads, under different load cycles. The results of the physical model are used to determine the behavior of crack development under certain operating conditions. Other researchers utilized existing knowledge about the degradation of ballast and incorporated it into their model, creating a function to calculate the conditions based on operation data.

The research in [45] developed a sample generator based on a physical model, based on the probability and stochastic processes to create additional data samples for training and testing their prediction model. Expected damage occurrence, degradation threshold, and shock arrival rates were parameters integrated into the generator, just to name a few.

**B. Algorithms** A variety of different algorithms and computational models were used for processing data to predict or determine conditions or defects.

#### 1) Convolutional Neural Network (CNN):

CNNs are artificial neural networks (NNs) with a reduced number of parameters, allowing for solving complex tasks. CNNs consist of multiple layers, including convolutional layers, nonlinearity layers, pooling layers, and fully connected layers, with the convolutional and fully connected layers being the only ones with parameters. CNNs are especially potent in dealing with large image data [39].

CNNs were utilized for defect detection of railway plugs, wheels, current collector strips of pantographs, and train bogies [16], [28], [30], [31]. For defect detection on railway plugs, the model achieved good results on test data, however, the actual inspection for defects needs to be done manually. The CNN was used to analyze the vertical wheel force to detect nonround wheels with flat spots and the model developed to detect wear on current collector strips both managed to identify defective parts with an accuracy of about 70% in test scenarios. Furthermore, promising results were achieved by the CNN-based systems used for fault diagnosis of train bogies, managing to correctly classify defects, and reaching an accuracy of up to 100% on test datasets.

#### 2) Multitask 1-D CNN (MT-1DCNN):

MT-1DCNNs are CNNs that utilize parallel processing of auxiliary tasks, and share learned features through a so-called trunk network. The features learned by the trunk network are the input of the individual branches, hence every branch utilized the features learned by the trunk and fully understands the characteristics of the data, improving the precision of the whole network [34].

An MT-1DCNN was utilized for the fault diagnosis of wheelset bearings, based on operation data. An experimental comparison of five peer networks proved the new model advantageous in terms of precision and recall [34].

#### 3) U-Net:

A U-Net is a specialized CNN, which is optimized for image segmentation with fewer training images and higher precision. It utilizes data augmentation techniques, such as elastic deformations, to use the available samples more efficiently [40].

A system based on a U-Net was developed that manages to detect

surface discrete rail defects with an accuracy of up to 99.76%, which is an improvement, over comparable existing models, of up to 6.74% [29].

4) *Deep Neural Network (DNN)*: DNNs are NNs with multiple hidden layers between the input layer and output layer, which allows modeling complex relationships.

A DNN was utilized for fault diagnosis on wheelset-axle box assemblies, by analyzing image data created by the SMTF algorithms, which transform the vibration signal into a visual form, significantly reducing complexity and computational burdens [38].

#### 5) Recurrent Neural Network (RNN):

RNNs are special types of NNs, which utilize their lagged output as an additional input, to recursively calibrate the prediction accuracy. Therefore, RNNs produce different results every time they are run. RNNs are specially designed to deal with time-series data, as they can memorize dependencies between adjacent time steps [20].

An RNN was utilized for the prediction of red and yellow classified rail defects based on maintenance and operation data and was able to achieve an accuracy of 82% on a test dataset. In addition, the results are used for maintenance activity scheduling [20].

#### 6) Long Short-Term Recurrent Neural Networks (LSTM-RNN):

RNNs face the problem of gradient exploding or vanishing, as they have limited capabilities in long-term dependence memory. This problem is solved by LSTM-RNNs, which are a special type of RNNs, that incorporate such long-term dependencies, through the tradeoff of having a more complicated structure [15].

An LSTM-RNN was used to create a model to predict future maintenance

timing, based on historical maintenance sample data [15].

7) *Gated Recurrent Unit (GRU)*: GRU networks are a specific variant of the long short-term memory (LSTM) algorithm, that reduces the short-term memory of long sequences in RNN. This solves the issue of conventional RNNs of vanishing or exploding gradients through sudden, substantial changes, within the input data, by learning through long-term historical information of the time-series data [17].

A GRU-based system for fault detection and isolation for AC/DC power systems was developed, tested, and compared to other NNs under real-time conditions. The GRU-based system outperformed the others in terms of hardware utilization, runtime, and accuracy [17].

8) *Self-Organizing Map (SOM)*: A SOM is a special type of NN that groups similar inputs into classes. It was successfully used for unsupervised classification of acoustic emission events, separating background noise with actual defect development-related noise emissions, for the detection of damage within railway axels [18].

9) *Extreme Learning Machine (ELM)*: An ELM is a single-hidden-layer feedforward NN algorithm that is effective and easy to use. In comparison to traditional NNs, ELMs generally provide faster results and have increased learning speeds and more generalization [25].

An ELM was utilized to predict rail weld defects with a 100% recall rate, based on a test dataset, allowing for decreased workloads, related to inspections, while maintaining the safety requirements [25].

10) *Complex Fuzzy System*: A complex fuzzy system is based on fuzzy logic and represents a mathematical model that mimics the human decision-making process. A complex fuzzy system consists of four main parts, a blur zone, where input variables are converted into fuzzy values, the knowledge base, where the decision rules are located and the data are stored, the subtraction section, where the results are created by applying the rules to the input data and last, the defuzzification section, which converts fuzzy values back into numerical ones [14].

A system based on complex fuzzy logic was developed, to determine the conditions of the catenary systems as well as the rails, through thermal image analysis [14].

11) *Random Forest (RF)*: RF is a regression model that combines independent regressors, the so-called decision trees. These decision trees are created by random selection of a subset of features from the set including all of them and by applying an aggregation technique. The prediction is done by averaging the outputs of all decision trees [20].

The RF algorithm was utilized for a model to predict red and yellow classified rail defects based on operation and maintenance information. Furthermore, these predictions were used to schedule maintenance activities [20].

12) *Approximate Bayesian Computation (ABC)*: ABC models stem from Bayesian statistics in which probability expresses a degree of belief within an event. These models are widely used for making assumptions and approximations as well as parameter estimation and model selection. ABCs are very well suited for complex models, as it bypasses the evaluation of the likelihood function, and its

solving might be computationally very costly [42].

A model to predict rail break arrival rate, based on ABC was developed, tested, and compared to an existing negative binomial-based system, with the ABC-based one having a 25% lower mean absolute error rate, on a test dataset [21].

13) *K-Means Clustering*: The k-means clustering algorithm is used to identify and group points of datasets into clusters, which are regions with higher local point density in comparison to others. The algorithm aims to identify a center for each respected group [22], [43].

K-means clustering was used for grouping and pattern identification, which is visualized for specialists who analyze the displayed results to determine defective railroad switches [22].

14) *Recursive Feature Elimination (RFE)*: RFE algorithms are used to reduce the dimensions of datasets based on the contribution of each feature toward the accuracy of the results. It eliminates features that do not significantly impact the outcome, such as highly correlated features, and eliminates them, reducing the complexity.

RFE was utilized to identify MGT as the significant parameter for rail defects, with the developed prediction model being able to predict 83% of rail defects based on track geometry and tonnage data [35].

15) *Support Vector Machine (SVM)*: SVMs are the algorithms for clustering and regression analysis. SVMs try to create a so-called hyperplane that divides data points into two groups, with as much distance between them as possible. Enough space between data sets ensures that data that deviate from

the training data are correctly classified [24].

An SVM-based model proved itself as an efficient tool for predicting geometry defects and classifying track sections, which allows the planning of maintenance activities, such as tamping, stone blowing, and gage correction activities [24].

16) *Gradient Boosting (GB)*: Gradient tree boost and stochastic GB are two algorithms that create regression models by sequentially fitting a parameterized function, to so-called pseudo-residuals at each iteration. The parameterized function is the base learner, which is fitted to the gradient derived from the minimizing of the loss function. This process respects the values of the model at each training data point assessed at the current step [41].

GB was utilized for the development of a prediction model that forecasts the dynamic stiffness of base plates, based on environmental and operation conditions, with an error value of less than 5%. GB tree was also used to predict the maintenance demand of railway switches, which was achieved with an accuracy of 86%, outperforming a similar RF-based model [27], [33].

17) *Speeded-Up Robust Features (SURF)*: The SURF algorithm is a feature detector and descriptor, used for object recognition, image matching, and other machine vision tasks. On a fundamental level, it works by matching pictures, obtained by the detector, to image features, described and contained within the descriptor [26].

The SURF algorithm was used for pixel-level image matching to detect irregularities through the comparison of images of the same object within different periods. The system was

tested with good results, however with a high rate of false positives [26].

18) *Random Sample Consensus (RANSAC)*: The RANSAC algorithm utilizes an iterative procedure to cluster data points, to estimate a mathematical model that describes them. RANSAC is often used for the evaluation of sensor data that has many outliers, as its results are not affected by them [19].

The RANSAC algorithm is used to detect outliers and inliers within point clouds, gathered with LiDAR technology, to determine if a point is related to representing a certain object or not. Furthermore, RANSAC is utilized for the alignment of rail images to position them correctly before comparison [19], [26].

19) *Background Oriented Defect Inspector (BODI)*: A BODI generates a constantly updating background model by randomly selecting sets of pixels along a longitudinal line, for example, along rails. Therefore, changes within the background are taken into consideration, increasing the prediction accuracy [37].

A BODI-based system was successfully used for the inspection and detection of rail surface discrete defects, through the analysis of images. Experimental results show promising results, with the BODI system detecting 100% of the defects with a low false-positive rate [37].

## VI. LIMITATIONS AND FUTURE WORK PRESENTED IN THE REVIEWED STUDIES

Limitations and areas of future work are important parts of a research paper, as they allow insights into encountered obstacles as well as the continuation of work. The papers were scanned regarding statements

indicating limitations and future work, yet not all papers mentioned such, as researchers did not encounter limitations or do not plan to continue research as consider their work as completed.

**A. Limitations** Eight of the reviewed papers discuss limitations that occurred during the development or testing processes and thus restricted their proposed solution. Several general reasons can be allocated as the reasons for limitations, with four papers mentioning physics-related difficulties, three papers mentioning problems related to data, and one mentioning missing standardization.

Four papers mentioned limits of the capability of their model, depending on the component size or other physical boundaries. Vibration signals for example are not only affected by a single damaged component but rather by the condition of the whole assembly and all connecting components. In particular, the vibration emissions measured on the axel, or the bogie are heavily affected by the conditions of the rails, thus increased vibration, which may be traced back to wear on components in the axel box assembly may simply be induced by poorly maintained tracks [17], [28], [34].

Another physics-related issue, that specifically affects VIS, is the variety in size of different components, as the images gathered must be detailed enough to display the items of interest in necessary quality, but also provide enough angle to capture the whole component. While this may not be an issue for approaches focusing on single components, like rails or pantographs, it is a challenge for the detection of defects on track-side equipment [26].

Data-related limitations concern either the size of the data itself, as ML algorithms need a significant size of data to be trained and tested on; or the quality of the available data, as it is often incomplete or in the case of data generated via a simulation, only partly representative of the real-work scenario. Furthermore, this made it necessary to process the data, removing missing or not valid values to avoid falsification of the ML algorithm, or to create additional data samples for testing and validation of the developed model [29], [15], [31].

Researchers [24] stated that missing standardization of parameters significantly limits the usability of their developed solution, as track parameters, especially TQIs are determined differently by railway companies.

**B. Future Work** Sixteen of the reviewed papers mentioned fields of interest and goals that are planned to be reached within future research. The papers are clustered into three groups, based on their statements, focusing on either the improvement of the proposed solutions, the development of a new one or adding additional features based on the gathered experience, and investigation of other algorithms and data sources, not used in their current paper (see Table 8).

#### 1) *Enhancement:*

The continuation of the research by enhancing the proposed solution is what eight papers stated as their future work. Their goal is to optimize and improve the proposed models through additional data, to increase

accuracy, and consistency and enable their utilization for a wider area of applications.

Five papers stated to focus on the improvement of prediction or detection accuracy as well as recall rate and general optimization of their model. This will be achieved by using extended datasets for additional training, through the utilization of additional input dimensions or the integration of ML algorithms [16], [26], [29], [30], [31].

Two papers stated to widen the application area of their system to allow for deployment in a greater variety of settings. The diagnosis of axle box assemblies under varying velocities will be developed, and the prediction of ballast settlement for tracks with heterogeneous traffic will be complemented [32], [38].

One paper mentioned considering additional input sources to strengthen the reliability of the proposed system. In detail, the usage of conventional images in addition to thermal ones is planned for the detection of surface discrete defects of rails [14].

#### 2) *Further Development:*

Development of additional features or new models, that complement the functionality of the proposed approaches, based on the experience gathered, is what seven papers considered for their future work.

Five papers stated to improve their proposed solutions by adding additional features. These features include the implementation of diagnosis and prognostic tools to

allow better evaluation of a component's condition, the implementation of regression models to predict the ideal timing to conduct maintenance activities, the development of a threshold for principal component indices, utilized in track quality indices, the implementation of severity scores for the determination of sizes of detected flat spot on wheels as well as the additional prediction of the damping capacity of rail pads [18], [24], [27], [30], [33].

Two papers mentioned improvements in their proposed system through the implementation of additional models that enhance their capabilities. To increase the precision of a VIS, researchers developed a model that considers the relationship between neighboring pixels to complement their proposed BODI system. Another paper states to develop a model for the detection and classification of individual components within point clouds gathered through LiDAR technology [19], [37].

#### 3) *Investigation of Different Approaches:*

Analysis of algorithms and data sources that may yield better results and performance are what three papers stated to be their future work. One study stated to investigate the potential of other signals, such as ultrasound, for the detection of rail defects [37].

Two papers stated to investigate other network architectures and algorithms to potentially find a better suiting alternative or to combine it with the developed solution to improve its performance. The algorithms DBSCAN, mean shift, affinity propagation, and Gaussian mixture will be analyzed [22], [34].

Fig. 7 summarizes the limitations and future work presented in the reviewed studies.

**Table 8. Classification of Future Work.**

Classifier	Description
Enhancement	The paper states that the enhancement or improvement of the proposed approach is planned.
Further Development	The paper states that further development or adding new features, complementing the proposed solution, is planned.
Investigation	The paper states that the investigation and analysis of other algorithms and data sources are planned.

## VII. DISCUSSION

A variety of 20 different components, assets, and assembly groups were considered by the studies, with the rails and joints being the most addressed component. They were considered as the safety-critical components, as rail breaks pose a substantial safety threat, being a major cause for derailments, second to human error. The second most considered components were the parts connecting the vehicle body with the tracks, including train bogies, wheels, axel-box assemblies, axels, and wheelset bearings, as they are directly affecting the vehicle’s motion, hence being critical for safety, availability, and ride comfort.

Two main approaches for the integration of predictive maintenance solutions were observed. One is to automate inspection and defect detection, hence replacing the need for workers to manually carry out the periodic and time-consuming acts of inspecting components to evaluate their functionality. Many solutions were proposed utilizing a variety of different sensors, including visual cameras, vibrations sensors, sensors measuring electric signals, such as current or voltage, and force sensors.

The most proposed solutions for defect detection are VIS, as they are recognized as robust and versatile, which allows them to be utilized for a wide range of applications. However, the drawback is that they are only able to detect surface discrete defects, of components that are clearly visible.

The second main approach focuses on the prediction of defects and conditions, to enable the possibility to plan future demands of maintenance activities. Therefore, defect and component failures are predicted by utilizing existing data such as maintenance logs, including information about defects, and performed tasks, operation data, such as speed and MGTs, and environmental data and track geometry parameters.

Besides the different goals of these two approaches, they utilized significantly distinct sources of data. While the defect detection approaches solely use sensor data, therefore demanding additional sensors to be installed and maintained, creating additional costs, and requiring the implementation of new systems to process the gathered data, the prediction-centered

approaches focused on existing data. This has the advantage that data, which has already been gathered for different purposes, is used. Hence, there is no need for new systems to gather, manage, and store data, only requiring an interface to the existing system. In conclusion, this demands less effort and reduces the threshold for adoption within the industry.

But there is one downside related to the nature of prediction systems. While they yield great potential as tools for additional information, incorporated in the long-term planning and scheduling of maintenance activities, they cannot replace inspections. Consequently, prediction and detection systems should be used together, complementing each other’s functionality.

Additionally, physical models were utilized to emulate the defect development processes, to support understanding of defect development, including them in their prediction system, and to generate additional data samples to overcome one of the main limitations, namely the poor availability and quality of data. Even though existing data are available, they might be of poor quality, incomplete, and containing not valid values, which need to be removed, as they distort the used algorithms leading to wrong results.

Various algorithms were utilized within the proposed systems for defect detection or prediction, as, depending on the data source and goal of the approach, different kinds are more suitable than others and promise better results. These are classified into three groups, with NNs being the most utilized ones, followed by classification algorithms and feature selection algorithms.

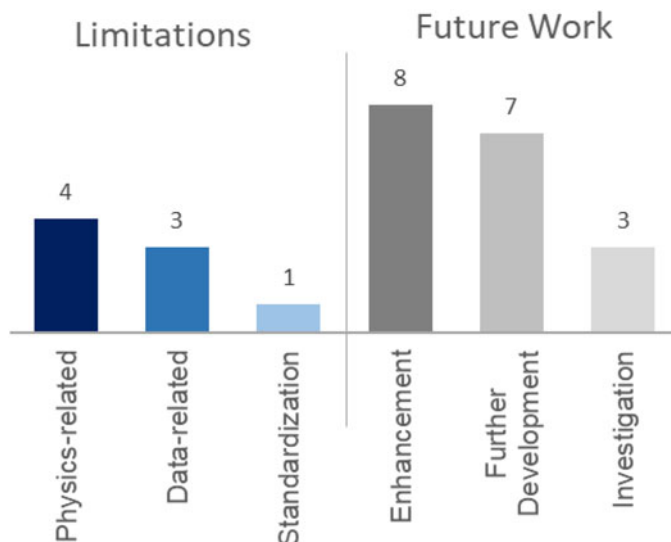


Fig. 7. Limitations and future work presented in the reviewed studies.

Fig. 8 demonstrates the tag cloud of the used algorithms.



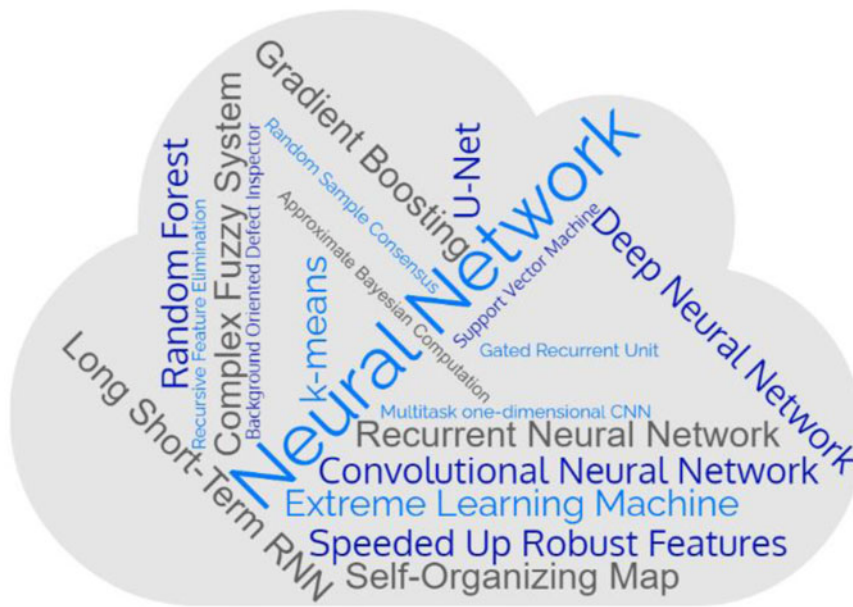


Fig. 8. Tag cloud of the used algorithms.

In addition to the data-related limitations, physics-related issues need to be overcome. The generalization of VIS, to function for a variety of components at the same time, widens its area of application but creates the demand for the system to consider components with great differences in size which weakens the system's precision.

Many of the developed systems yield great potential and have shown their capabilities in tests or have already been put into service. Therefore, many papers stated to enhance them by improving the accuracy and reliability, as well as widening their area of application and developing additional features that grant added functionality.

## VIII. CONCLUSION

There are several theoretical contributions and practical implications of the SLR. In addition to

the analysis of the research purpose and concerns of the papers, the railway assets, their components, defects, failure modes, and predicted conditions were considered. Analyzed data sources and algorithms have been mapped into existing solutions, focusing on their limitations, which will help practitioners with specific implementations.

The implementation and utilization of new predictive maintenance solutions show great potential for saving costs through optimizing resource consumption, reducing the labor intensiveness of maintenance activities, and decreasing the demand for labor-intensive tasks such as inspections and patrolling. Furthermore, a well-maintained railway network offers improved availability, decreases costs related to service interruptions, and simultaneously improves customer satisfaction. Improving the safety of

the railway network through early detection of defects reduces the risk of component-related accidents, as well as the risk of mistakes related to the subjectivity of human nature, for example leaving defects undetected, during long and tiring inspections or patrolling tasks. In addition, there is greater workplace safety, as there is less need to be on-site. Hence there is a lower chance of potential accidents occurring during maintenance activities.

Future work considers the further analysis of the proposed solutions to assert the feasibility of practical implementation including the evaluation of requirements that need to be met. The transferability of predictive maintenance approaches from production and operations management to the railway industry will be considered. Furthermore, a comparison of the required efforts and the potential benefits of a performance assessment may be conducted.

## DECLARATIONS

### Competing Interests

We acknowledge that authors have no competing interests.

### Authors' Contributions

We acknowledge that the authors have contributed significantly and are in agreement with the content of the manuscript.

## AVAILABILITY OF DATA AND OTHER MATERIALS

Data collection forms and data extracted from included studies and used for analysis are available upon request to the authors.

## REFERENCES

- [1] Wöhe Günter, Döring Ulrich, and Brösel Gerrit, *Einführung in die Allgemeine Betriebswirtschaftslehre*, 26th Edition, München: Springer-Verlag, 2016.
- [2] Motaghare Omkar, Pillai Anju S., and Ramachandran K. I., "Predictive maintenance architecture," in *Proceeding of the IEEE International Conference on Computational Intelligence and Computing Research*, 2018, pp. 1–4.
- [3] Mobley Keith R., *An Introduction to Predictive Maintenance*, 2nd Edition, Amsterdam, The Netherlands: Elsevier, 2002.
- [4] Wireman Terry, *Maintenance Work Management Processes: Maintenance Strategy Series*, 3rd Edition, Fort Myers, FL, USA: Reliability Web, 2008.
- [5] Wang Honzhou and Pham Hoang, *Reliability and Optimal Maintenance*, London, U.K.: Springer-Verlag, 2006.
- [6] Mckinsey & Company, "Ask an expert: What Industry 4.0 can do for maintenance," 2017. Online. [Available]: <https://www.mckinsey.com/business-functions/operations/our-insights/ask-an-expert-what-industry-40-can-do-for-maintenance> Accessed: Feb. 9, 2021.
- [7] Sasiidharan M., Burrow M. P. N., and Ghataora G. S., "A whole life cycle approach under uncertainty for economically justifiable ballasted railway track maintenance," *Research in Transportation Economics*, vol. 80, 2020, pp. 805–824.
- [8] Ridley Diana, *The Literature Review: A Step-by-Step Guide for Students*, 2nd Edition, London, U.K.: Sage, 2012.
- [9] O'leary Zina, *The Essential Guide to Doing Your Research Project*, 3rd Edition, London, U.K.: Sage, 2017.
- [10] Jesson Jill K., Matheson Lydia, and Lacey Fiona M., *Doing Your Literature Review: Traditional and Systematic Techniques*, London, U.K.: Sage, 2011.
- [11] Grant Maria J., Booth Andrew, "A typology of reviews: An analysis of 14 review types and associated methodologies," *Health Information and Libraries Journal*, vol. 26, 2009, pp. 91–108.
- [12] Littell Julia H., Corcoran Jacqueline, and Pillai Vijayan, *Systematic Reviews and Meta-Analysis*, New York, NY, USA: Oxford University Press, 2008.
- [13] Page Matthew J. et al., "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," *BMJ Journals*, vol. 372, 2021, Art. no. 160.
- [14] Karakose Mehmet and Yaman Orhan, "Complex fuzzy system based predictive maintenance approach in railways," *IEEE Transactions on industrial informatics*, vol. 16, no. 9, Sep. 2020, pp. 6023–6032.
- [15] Wang Qi, Bu Siqi, and He Zhengyou, "Achieving predictive and proactive maintenance for high-speed railway power equipment with LSTM-RNN," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, Oct. 2020, pp. 6509–6517.
- [16] Kou Linlin, Qin Yong, Zhao Xuejun, and Chen Xin'an, "A multi-dimension end-to-end CNN model for rotating devices fault diagnosis on high-speed train bogie," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, Mar. 2020, pp. 2513–2524.
- [17] Liu Qin, Liang Tian, and Dinavahi Venkata, "Real-time hierarchical neural network based fault detection and isolation for high-speed railway system under hybrid AC/DC grid," *IEEE Transactions on Power Delivery*, vol. 35, no. 6, Dec. 2020, pp. 2853–2864.
- [18] Carboni Michele and Crivelli Davide, "An acoustic emission-based structural health monitoring approach to damage development in solid railway axles," *International Journal of Fatigue*, vol. 139, 2020, pp. 1–11.
- [19] Sánchez-Rodríguez A., Riveiro B., Soilán M., and González-Desantos L. M., "Automated detection and decomposition of railway tunnels from mobile laser scanning datasets," *Automation in Construction*, vol. 96, 2018, pp. 171–179.

- [20] Gerum Pedro C. L. and Altay Ayca, Baykal-Gürsoy Melike, "Data-driven predictive maintenance scheduling policies for railways," *Transportation Research Part C*, vol. 107, 2019, pp. 137–154.
- [21] Ghofrani Faeze, Yousefianmoghadam Seyedsina, He Qing, and Stavridis Andreas, "Rail breaks arrival rate prediction: A physics-informed data-driven analysis for railway tracks," *Measurement*, vol. 172, 2021, Art. no. 108858.
- [22] Soares Nielson, De Aguiar Eduardo P., Souza Amanda S., and Goliatt Leonardo, "Unsupervised machine learning techniques to prevent faults in railroad switch machines," *International Journal of Critical Infrastructure Protection*, vol. 33, 2021, Art. no. 100423.
- [23] Gbadamosi Abdul-Quayyum et al., "IoT for predictive assets monitoring and maintenance: An implementation strategy for the UK rail industry," *Automation in Construction*, vol. 122, 2021, Art. no. 103486.
- [24] Lasisi Ahmed and Attoh-Okine Nii, "Principal components analysis and track quality index: A machine learning approach," *Transportation Research C*, vol. 91, 2018, pp. 230–248.
- [25] Yao Nan, Jia Yuejun, and Tao Kai, "Rail weld defect prediction and related condition-based maintenance," *IEEE Access*, vol. 8, 2020, pp. 103746–103758.
- [26] Gao Limin, Jiu Yunyang, Wie Xiang, Wang Zhongchuan, Xing Weiwei, "Anomaly detection of trackside equipment based on GPS and image matching," *IEEE Access*, vol. 8, 2020, pp. 17346–17355.
- [27] Bukhsh Zaharah A., Saeed Aaqib, Stipanovic Irina, and Doree Andre G., "Predictive maintenance using tree-based classification techniques: A case of railway switches," *Transportation Research Part C*, vol. 101, 2019, pp. 35–54.
- [28] Du Xinyu, Cheng Yu, and Zichen Gu, "Change detection: The framework of visual inspection system for railway plug defects," *IEEE Access*, vol. 8, 2020, pp. 152161–152172.
- [29] Lu Jun et al., "SCueU-Net: Efficient damage detection method for railway rail," *IEEE Access*, vol. 8, 2020, pp. 125109–125120.
- [30] Krummenacher Gabriel, Ong Cheng Soon, Koller Stefan, Kobayashi Seijin, and Buhmann Joachim M., "Wheel defect detection with machine learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, Apr. 2018, pp. 1176–1187.
- [31] Karaduman Gulsah and Akin Erhan, "A deep learning method for detecting of wear on the current collector strips' surfaces of the pantograph in railways," *IEEE Access*, vol. 8, 2020, pp. 183799–183812.
- [32] D'angelo Giacomo, Bressi Sara, Giunta Marinella, Presti Davide Lo, and Thom Nick, "Novel performance-based technique for predicting maintenance strategy of bitumen stabilized ballast," *Construction and Building Materials*, vol. 161, 2018, pp. 1–8.
- [33] Ferreno Diego et al., "Prediction of mechanical properties of rail pads under in-service conditions through machine learning algorithms," *Advances in Engineering Software*, vol. 151, 2021, pp. 1–11.
- [34] Liu Zhiliang, Wang Huan, Liu Junjie, Qin Yong, and Peng Dandan, "Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021, Art no. 3501711.
- [35] Mohammadi Reza, He Qing, Ghofrani Faeze, Pathak Abhishek, and Aref Amjad, "Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects," *Transportation Research Part C*, vol. 102, 2019, pp. 153–172.

- [36] Macedo Rita, Benmansour Rachid, Artiba Abdelhakim, Mladenovic Nenad, and Urosevic Dragan, "Scheduling preventive railway maintenance activities with resource constraints," *Electronic Notes in Discrete Mathematics*, vol. 58, 2017, pp. 215–222.
- [37] Gan Jinrui, Wang Jianzhu, Yu Haomin, Li Qingyong, and Shi Zhiping, "Online rail surface inspection utilizing spatial consistency and continuity," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 7, Jul. 2020, pp. 2741–2751.
- [38] Bai Yongliang, Yang Jianwei, Wang Jinhai, Zhao Yue, and Li Qiang, "Image representation of vibration signals and its application in intelligent compound fault diagnosis in railway vehicle wheelset-axlebox assemblies," *Mechanical Systems and Signal Processing*, vol. 152, 2021, pp. 107421–107434.
- [39] Albawi Saad and Mohammed Tareq Abed, "Understanding of a convolutional neural network," in *Proceeding of the International Conference on Engineering and Technology*, 2017, pp. 1–6.
- [40] Ronneberger Olaf, Fischer Philipp, and Brox Thomas, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceeding of the Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [41] Friedman Jerome H., "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, 2002, pp. 367–378.
- [42] Sunnaker Mikael, Busetto Aberto G., Numminen Elina, Corrande Jukka, Foll Matthieu, and Dessimoz Christophe, "Approximate bayesian computation," *PLOS Computational Biology*, vol. 9, 2013, pp. 1–10.
- [43] Likas Aristidis, Vlassis Nikos, and Verbeek Jakob J., "The global k-means clustering algorithm," *Pattern Recognition*, vol. 36, 2003, pp. 451–461.
- [44] Prisma, "Prisma: Prisma flow diagram," PRISMAStatement; FlowDiagram, 2020. Accessed: May 10, 2021. <http://www.prisma-statement.org>
- [45] Önorm EN 13306, 2018. Accessed: May 10, 2021. [https://shop.austrian-standards.at/action/de/public/details/623612/OENORM\\_EN\\_13306\\_2018\\_01\\_15](https://shop.austrian-standards.at/action/de/public/details/623612/OENORM_EN_13306_2018_01_15)
- [46] Schenkelberg Kai, Seidenberg Ulrich, and Ansari Fazel, "Supervised machine learning for knowledge-based analysis of maintenance impact on profitability," *IFAC PapersOnLine*, vol. 53, no. 2, 2020, pp. 10651–10657.
- [47] Pawellek Günther, *Integrierte Instandhaltung und Ersatzteillogistik*, 2nd Edition, Berlin, Germany: Springer-Verlag, 2016.
- [48] Pappaterra M. J., Flammini F., Vittorini V., and Bešinović N., "A systematic review of artificial intelligence public datasets for railway applications," *Infrastructures*, vol. 6, no. 10, 2021, Art. no. 136.
- [49] Laiton-Bonadiez C., Branch-Bedoya J. W., Zapata-Cortes J., Paipa-Sanabria E., and Arango-Serna M., "Industry 4.0 technologies applied to the rail transportation industry: A systematic review," *Sensors*, vol. 22, no. 7, 2022, Art. no. 2491.
- [50] Wan C., Yang Z., Zhang D., Yan X., and Fan S., "Resilience in transportation systems: A systematic review and future directions," *Transport Reviews*, vol. 38, no. 4, 2018, pp. 479–498.
- [51] Thomas M. J., *A Systematic Review of the Effectiveness of Safety Management Systems*, Canberra, ACT, Australia: Australian Transport Safety Bureau, 2012.
- [52] Leukel J., González J., and Riekert M., "Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review," *Journal of Manufacturing Systems*, vol. 61, 2021, pp. 87–96.
- [53] Tang R. et al., "A literature review of artificial Intelligence applications in railway systems," *Transportation Research Part C: Emerging Technologies*, vol. 140, 2022, Art. no. 103679.

- [54] Mueller C. and Mezhuyev V., "AI models and methods in automotive manufacturing: A systematic literature review," in Al-Emran M., Shaalan K., Eds., *Recent Innovations in Artificial Intelligence and Smart Applications. Studies in Computational Intelligence*, vol 1061, Berlin, Germany: Springer-Verlag, 2022, doi: [10.1007/978-3-031-14748-7\\_1](https://doi.org/10.1007/978-3-031-14748-7_1).
- [55] Davari N., Veloso B., Costa Gustavo de Assis, Pereira P. M., Ribeiro R. P., and Gama J., "A survey on data-driven predictive maintenance for the railway industry," *Sensors*, vol. 21, no. 17, 2021, Art. no. 5739, doi: [10.3390/s21175739](https://doi.org/10.3390/s21175739).
- [56] Xie J., Huang J., Zeng C., Jiang S.-H., and Podlich N., "Systematic literature review on data-driven models for predictive maintenance of railway track: Implications in geotechnical engineering," *Geosciences*, vol. 10, no. 11, 2020, Art. no. 425, doi: [10.3390/geosciences10110425](https://doi.org/10.3390/geosciences10110425).

**Mario Binder** received the B.Sc. degree in engineering in 2021 from FH Joanneum University of Applied Sciences, Kapfenberg, Austria, where he is currently working toward the master's degree in international industrial management. Since 2017, he has been in a professional position in the railway domain in engineering, design, and project management at Voestalpine Railway Systems, Leoben, Austria.

**Vitaliy Mezhuyev** (Senior Member, IEEE) received the Ph.D. degree in educational technology from Kyiv National Pedagogical University, Kyiv, Ukraine, in 2002, and the Sc.D. degree in information technology from Odesa National Technical University, Odesa, Ukraine, in 2012. From 2004 to 2014, he was a Professor of Informatics with Berdyansk State Pedagogical University, Berdyansk, Ukraine. From 2014 to 2019, he was a Professor of Informatics with the University of Malaysia Pahang. From 2019 with the Institute of Industrial Management at FH Joanneum, Kapfenberg, Austria. Visiting professor at six international universities. He participated in multiple international scientific and industrial projects, devoted to the design, development, and formal verification of computer systems. He authored more than 140 scientific papers in peer-review journals, including highly reputed venues such as *Complexity*, *Computers & Education*, *Cybernetics and Systems*, *Education and Information Technologies*, *IEEE Access*, *Information Management*, and *Technology in Society* among many others.

**Martin Tschandl** received the Ph.D. degree in social and economic sciences from Karl-Franzens-Universität Graz, Austria, in 1993. He completed his Research Semester with New York University, NY, USA. From 1991 to 1998, he was with industrial and service companies where he was responsible for finance and IT. Then he became a Professor of business administration and controlling. Since 2001, he has been the Head of the Industrial Engineering Institute "Industrial Management," FH JOANNEUM - University of Applied Sciences, Austria. He was a Visiting Professor with four international universities. He is the Head of the Austria II working group of the International Controller Association (ICV), Member of the International Group of Controlling (IGC), and Member of the board of the Logistics Network Association (VNL) and the Austrian Association of Industrial Engineers. He is involved in more than 70 research and industry projects on these topics. He has authored or coauthored more than 140 scientific publications, and edited the book series *Industrial Management*. His research interests include planning and control, strategic corporate development, management of logistics, and digital finance (forecasting, RPA).