Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

Adrian Matic

Machine Learning in Predictive Maintenance of Railway Infrastructures: Implementations and Challenges

Master's thesis in Global Manufacturing Management
Supervisor: Fabio Sgarbossa
June 2022

NTNU
Norwegian University of
Science and Technology

Adrian Matic

Machine Learning in Predictive Maintenance of Railway Infrastructures: Implementations and Challenges

Master's thesis in Global Manufacturing Management
Supervisor: Fabio Sgarbossa
June 2022

Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU
Norwegian University of
Science and Technology

Preface

This thesis is written as the final assignment of the Master of Science program in Global Manufacturing Management at the Norwegian University of Science and Technology (NTNU) in Trondheim. It is submitted in the course TPK4930 – Production Management, Master's Thesis, at the Department of Mechanical and Industrial Engineering.

I am thankful to my supervisor at NTNU, Fabio Sgarbossa, for guiding me during the process of this master's thesis. His feedback and guidance have been of great support for completing the assignment. I sincerely appreciate his ability to answer questions and provide support at any time. In addition, he connected me with one of the interview participants, which I am grateful for.

Furthermore, I want to direct my gratitude to my former colleague, Michelle Karlsson, for arranging another interview participant. Additionally, I am thankful for her thoughts on the problem formulation at the beginning of the project.

I also want to sincerely thank the interview participants, Veronica Brizzi from MIPU and Kristine Tveit from Bane NOR, for participating in the interviews. Their expertise gave further insights to the topic and was of great importance for the project.

Trondheim, June 2022

Adrian Matic

Summary

The recent technological advancements of Industry 4.0 technologies have shifted towards machine learning (ML) applications in predictive maintenance (PdM) and have been implemented in many industries. However, research on ML implementations to PdM for railway infrastructure assets is limited.

This master thesis investigates the topic by these research questions; (1) Which ML algorithms are applied to PdM of railway infrastructure assets? (2) Which railway infrastructure assets are subjected to ML algorithms in PdM? (3) What inspection methods and tools are utilized for data acquisition for ML in PdM of railway infrastructure assets? (4) What are the main challenges of ML in PdM of railway infrastructure assets?

To answer these research questions the following research methods are conducted:

1. Literature review (LR) to collect necessary theoretical information on railway infrastructure maintenance, ML methods and algorithms, and challenges related to ML in PdM.
2. A systematic literature review (SLR) for identifying and extracting materials and data from all relevant papers on the topic.
3. Semi-structured interviews (SSI) with experts of railway infrastructure maintenance for further information on the progress of ML implementations to PdM.

The main purpose was to uncover the progress towards ML in PdM of railway infrastructure assets. The papers identified in the SLR were screened against established eligibility criteria. This led to 20 relevant papers of the subject that were further analyzed. The analyses proved that the research on ML applications in PdM of railway infrastructure assets is relatively new and trending.

The results found that 17 different algorithms are applied to railway infrastructure PdM purposes. The algorithms consists of supervised, unsupervised, ensemble, and deep learning (DL) methods. Railway tracks and switches are the most dominating assets of the SLR. Inspection methods of railway infrastructure assets are mainly inspection vehicles and sensors installed in the infrastructure. The most utilized datatypes are geometry, asset properties, and historical data. ML algorithms are trained with data gathered from these methods, in combination with a variety of other data. The main challenges are associated with foundations, data collection, data quality, and data knowledge. Since ML implementations to PdM of railway infrastructure assets is a new research topic, several suggestions for further research are proposed.

Sammendrag

De nylige teknologiske fremskrittene av Industry 4.0 teknologier har skapt et skifte mot applikasjoner innen maskinlæring (ML) for prediktivt vedlikehold (PdM) og har blitt tatt i bruk av mange bransjer. Derimot finnes det lite forskning på implementasjon av ML til PdM for komponenter i jernbaneinfrastrukturen.

Denne masteroppgaven undersøker dette emnet med disse forskningsspørsmålene; (1) Hvilke algoritmer innen ML brukes i PdM av komponenter i jernbaneinfrastrukturen? (2) Hvilke komponenter i jernbaneinfrastrukturen har blitt underlagt av ML algoritmer i PdM? (3) Hvilke inspeksjonsmetoder og verktøy brukes for datainnsamling for ML i PdM av komponenter i jernbaneinfrastrukturen? (4) Hva er hovedutfordringene til ML i PdM av komponenter i jernbaneinfrastrukturen?

For å svare på forskningsspørsmålene er følgende forskningsmetoder utført:

1. Gjennomgang av relevant litteratur (LR) for å samle nødvendig teoretisk informasjon knyttet til vedlikehold av jernbaneinfrastrukturen, metoder og algoritmer innen ML, og utfordringer knyttet til ML i PdM.
2. En systematisk litteraturgjennomgang (SLR) for å identifisere og trekke ut materialer og data fra alle relevante artikler i temaet.
3. Semistrukturerte intervjuer (SSI) med eksperter i vedlikehold av jernbaneinfrastruktur for ytterligere informasjon om fremdriften av ML implementeringer i PdM.

Hovedformålet var å avdekke fremgangen av ML i PdM av komponenter i jernbaneinfrastrukturen. Artiklene som ble identifisert gjennom SLR ble screenet mot etablerte kvalifikasjonskriterier. Dette førte til 20 relevante artikler som ble videre analysert. Ytterligere analyser viste at forskningen innen ML i PdM av komponenter i jernbaneinfrastrukturen er relativt ny og voksende.

Resultatene viser at 17 forskjellige ML algoritmer brukes til PdM av jernbaneinfrastrukturen. Algoritmene består av «supervised», «unsupervised», «ensemble»- og dyp lærings (DL) -metoder. Jernbanespor og sporveksler er de mest dominerende komponentene. Inspeksjonsmetoder for komponentene er hovedsakelig inspeksjonskjøretøy og sensorer installert i infrastrukturen. De mest

brukte datatypene er komponentegenskaper, geometriske og historiske data. ML algoritmer trenes med data samlet fra disse metodene, i kombinasjon med en rekke andre data. Hovedutfordringene er knyttet til fundamentale grunnlag, datainnsamling, datakvalitet, og datakunnskap. Siden implementering av ML for PdM av komponenter i jernbaneinfrastrukturen er et nytt forskningstema, er flere forslag til videre forskning foreslått.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

ANN	Artificial Neural Network
CART	Classification and Regression Trees
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
GB	Gradient Boosting
GBT	Gradient Boosting Tree
GRU	Gated Recurrent Unit
kNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LR	Literature Review
LnR	Linear Regression
LgR	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MR	Multiple Regression
PCA	Principal Component Analysis
PCR	Principal Component Regression
PdM	Predictive Maintenance
PL	Probabilistic Learning
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF	Random Forest
RNN	Recurrent Neural Network
RUL	Remaining Useful Life
SLR	Systematic Literature Review
SSI	Semi-Structured Interview
SVM	Support Vector Machine
SVR	Support Vector Regression

1 Introduction

This study is motivated by the project thesis written by the author in the fall of 2021. The project thesis studied the methods for implementing predictive maintenance (PdM) in railway infrastructure maintenance. Another objective was to identify the challenges associated with the maintenance, and methods for collecting and measuring data. The author found that PdM can beneficially impact the infrastructure performance, maintenance efficiency, and mitigate the potential consequences of the challenges. The project thesis established future research to be conducted. Amongst the suggestions was research on how Industry 4.0 technologies can further improve the performance and efficiency of PdM of railway infrastructures (Matic, 2021). Therefore, the author chose to pursue the artificial intelligence technology named machine learning (ML) and uncover the progress of ML in PdM specifically in the railway industry. The following chapters defines the problem definition, research scope, thesis structure, and previous research.

1.1 Problem Definition

Large amounts of data are collected in the railway sector by the current inspection methods and tools (Kalathas and Papoutsidakis, 2021). This data must be reviewed, analyzed, and deployed to support decision-making procedures of maintenance. PdM is crucial for utilizing the data to predict and avoid faults in the railway infrastructure. Faults lead to replacement of assets and additional costs. Furthermore, it is critical to conserve resources, maintain customer service and passenger safety. The railway infrastructure consists of numerous assets, e.g., tracks, switches, embankment, catenary, and signaling systems. These require long-term and sustainable maintenance strategies to avoid and mitigate faults. Due to several differences in properties of assets, they need different maintenance plans and actions, including inspection methods for data collection. PdM can increase maintenance efficiency and support the existing challenges (Matic, 2021).

The ongoing industrial revolution called Industry 4.0 is linked to integrating physical and digital systems. This integration leads to larger amounts of data. Moreover, the emerging technologies from Industry 4.0 integrate a better interaction between people and machines, allowing a quicker and more centered information-sharing platform. Alpaydin (2016) defines ML as “programming computers to optimize a performance criterion using example data or past experience”. As a result, the ML approach has become a vital tool for developing proper PdM strategies (Carvalho et al., 2019). With support from ML, PdM for infrastructure asset management can more accurately determine if an asset will fail within a specified time frame, known as failure prediction. Moreover, the technology can estimate the remaining useful life (RUL) of an asset (Bukhsh and Stipanovic, 2020, Matic, 2021). However, according to a survey by Haarman et al. (2017), only 11% of the participating companies have applied ML in PdM. This indicates that the implementation has much potential, but also implies that there are some inherent challanges

With this background, the master thesis will research the PdM strategies with ML approaches in railway maintenance. With respect to the railway sector, the author seeks to establish the most suitable ML algorithms for PdM. In addition, researching which data collection methods (including data types and asset dimensions) are utilized, what infrastructure assets are most exposed to PdM with ML, and identifying the current inherent challenges. To achieve this purpose, the author will gather information on railway infrastructure PdM and ML methods through a LR, followed by research attempting to uncover all papers related to ML implementations in PdM of railway infrastructure assets. This is accomplished with SLR. Parallelly, to gain more knowledge and insight into the problem, interviews with experts in ML for railway PdM has been conducted. At the end, the author will present the results and analysis, followed by a discussion. Lastly, a conclusion of the whole report is conducted.

1.2 Research Scope

The primary objective is to identify ML algorithms applied to PdM for railway infrastructure assets. Regarding their application, railway infrastructure assets, inspection methods, and tools need to be addressed. Moreover, this research seeks to classify the data acquisition sources, description of datasets, and asset dimensions. In addition, to identifying ML algorithms, another objective is to determine which infrastructure assets are exposed to ML applications. Previous research implies that ML may benefit PdM, but companies still struggle with the implementation, as stated in the problem definition. Therefore, the last objective is identifying the current challenges related to ML in PdM. Centered around these objectives, the following research questions (RQs) have been formulated:

RQ1: Which ML algorithms are applied to PdM of railway infrastructure assets?

RQ2: Which railway infrastructure assets are subjected to ML algorithms in PdM?

RQ3: What inspection methods and tools are utilized for data acquisition for ML in PdM of railway infrastructure assets?

RQ4: What are the main challenges of ML in PdM of railway infrastructure assets?

1.2.1 Limitations

The scope of the project is limited to ML in PdM of the railway industry. More precisely, ML applications to other assets (e.g., trains and inspection vehicles) than the railway infrastructure will not be researched. In terms of PdM, this thesis strictly focuses on methods that predict failure, not methods for detecting anomalies, which is sometimes related to PdM. Regarding the ML algorithms, the author will present their general functions, but not study them deeply in terms of programming implementation. In addition, the author will not describe the identified papers from the SLR in every detail, but provide an overview of inspection methods and tools, main findings, and expressed challenges.

1.3 Thesis Structure

Chapter 1 – Introduction	This chapter introduces the project by expressing the motivation and background, justifying the research topic. It presents the problem definition, research scope, thesis structure, and previous research.
Chapter 2 – Methodology	Presents the design of the conducted research. It consists of three parts, a LR, SLR, and a SSI, that are used to achieve the purpose of this thesis.
Chapter 3 – Railway infrastructure maintenance	Describes the relevant information of railway infrastructure maintenance found from the LR. The chapter investigates several sub-topics in the area.
Chapter 4 – Machine Learning	Provides the theory explored in the LR. Establishes definitions of ML topics, and describes them from a general perspective, and presents the challenges related to ML in PdM.
Chapter 5 – Results	Presents the results from the SLR and the SSIs.
Chapter 6 – Discussion	This chapter discusses the results from chapter 5 in correlation to the research questions, previous work, and weaknesses.
Chapter 7 – Conclusion	Concludes the thesis by presenting the key results and discussion of the problem definition. In addition, suggestions for further research areas are provided.

1.4 Previous Research

Previous research in the literature of ML encompasses a wide range of models and methods for each learning type. Regarding PdM applications, the most prevalent ML types that may be utilized for failure prediction and RUL, are supervised and unsupervised learning (Arena et al., 2022). From a comparative study conducted by Ouadah et al. (2022), Random Forest (RF), Decision Tree (DT), and k-Nearest Neighbors (kNN) were selected as the most applied algorithms for PdM purposes based on a set of criteria. This study found that RF and DT achieved relatively similar accuracy and that kNN is a more robust classification algorithm for large datasets, but RF performs better for small datasets.

Several research papers on PdM applications with ML approaches have been published. To the knowledge of the author, other SLRs are lacking in the railway infrastructure assets field. Nakhaee et al. (2019) focused on ML algorithms applied to purely railway tracks and no other assets. In contrast to failure prediction, the study was fixated on fault detection. Similarly, Xie et al. (2020) performed an SLR focused on railway tracks, but the research was centered on PdM. They found that unsupervised learning, DL, and ensemble methods for PdM are growing, but anomaly detection models will not fade away in the near time. Davari et al. (2021) conducted an SLR for ML-PdM applications in the railway industry. However, the research only obtained five papers for railway infrastructure assets, and the rest were outside of railway infrastructure assets, e.g., train components, train speed, and other industries.

These observations gave the author further motivation and will to conduct the research.

2 Methodology

This chapter presents the research design of the project. It consists of three parts: LR, SLR, and SSI. The LR is conducted to establish a solid theoretical background on the topic. The SLR is the primary method and is, along with the LR and the interviews, carried out to collect data and answer the RQs to support the findings of the project.

2.1 Literature Review

The LR was conducted to gain theoretical knowledge of primarily railway infrastructure maintenance, ML methods and algorithms, and challenges related to ML in PdM. A LR can be either qualitative or quantitative (Purssell and McCrae, 2020). A qualitative LR of books and peer-reviewed papers were selected. The LR was done primarily through NTNUs bibliographic database, Oria, but the supervisor provided a few sources, and some sources from the project thesis were used.

The search was organized with two primary search words to obtain the relevant literature (see Table 1). The primary and secondary search words were combined differently to obtain the needed literature. Primary search words were utilized alone and combined with one or more secondary search words. The theory gathered from the LR is presented in chapters 3 and 4.

Primary search words	Secondary search words	
Railway infrastructure maintenance	Assets	Methods
	Predictive maintenance	Tools
	Inspection	Process
	Characteristics	Planning
Machine learning	Methods	Type
	Algorithm	Task
	Predictive maintenance	Challenges
	Definition	Issues
	Process	

Table 1: Search words used in the literature review

2.2 Systematic Literature Review

Siddaway et al. (2019) describe a SLR as a review of a clearly formulated question(s) that employs systematic and explicit approaches to locate, select, and critically evaluate relevant research. Furthermore, to gather and analyze data from the studies. A SLR consists of a systematic search for all relevant works. The located work should address one or more of the research questions, and the criteria and takeaways of the search should be presented and summarized systematically.

Prior to conducting the SLR, it is essential to consider if an identical SLR has been published previously (Purssell and McCrae, 2020). From what was found in 1.4, this was not the case, and the author proceeded with the SLR.

Based on methodological guidelines from Siddaway et al. (2019) and Purssell and McCrae (2020), Figure 1 presents the following developed SLR framework. This framework represents step-by-step the practice of conducting this SLR. The research scope is presented in chapter 1.2, and the results, discussion, and conclusion are elaborated in chapters 5, 6, and 7, respectively.

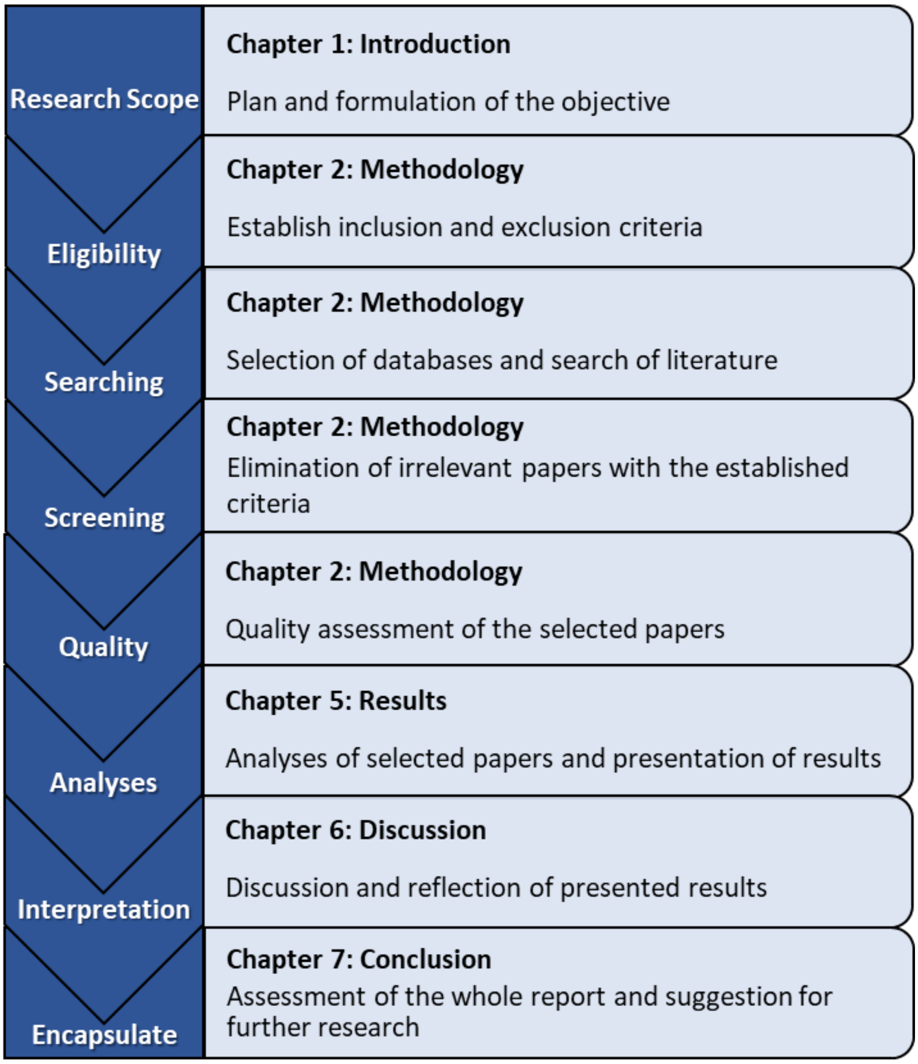

Figure 1: SLR Framework

2.2.1 Eligibility Criteria

Before the SLR can begin, inclusion and exclusion criteria must be established. As expressed by Siddaway et al. (2019), they should be formulated purely based on the RQs and should be reviewed consistently throughout the search process. The RQs presented in chapter 1.2 provide inclusion and exclusion criteria, shown in Table 2. In addition, the selected criteria are justified and built on theoretical grounds.

Inclusion Criteria	Exclusion Criteria	Justification
Papers related to PdM, ML, and railway infrastructure assets	Papers not related to PdM, ML, and railway infrastructure assets	The selected paper must be related to using ML-PdM for railway infrastructure assets. This is demanded to answer the stated RQs.
Presentation of model/method/technique, including testing and results	Papers that only present an idea/hypothesis without any experimentation	The criteria are set to ensure that the proposed methods have been experimentally tested, and that the results are expressed. The presented method is not helpful to this research without it being tested.
Peer-reviewed papers	Grey literature (unpublished work and not peer-reviewed)	Peer-reviewed papers have been published in scientific journals and have gone through a critical filter. This will improve the quality assessment and reduce biased papers (Purssell and McCrae, 2020).
Papers published after 2011	Papers published before 2011	The author chose to set a time restriction, due to the introduction of digital technologies resulting in advancements in PdM and ML, was in 2011 (Meindl et al., 2021)
Primary Sources	Secondary Sources (Interprets and comments primary sources)	This SLR targets papers that involve original research and new findings and is therefore restricted to primary sources.
English papers	Not English Papers	The search was limited to the English language. While including other languages in the search can potentially improve the SLR, it is also challenging and time-consuming to obtain and translate. Therefore, restricting to English papers is commonly accepted (Purssell and McCrae, 2020).

Table 2: Inclusion and Exclusion Criteria

Note that papers that did not involve the implementation of ML for railway infrastructure assets were excluded. One reason was to pursue the project report written by the author in the fall of 2021 (Matic, 2021), which focused on PdM implementation in railway infrastructure maintenance. Another primary reason was that the research would be too comprehensive if the field of interest were expanded. Thus, papers associated with ML implementations in PdM, e.g., inspection vehicles or components of a train, were not included in this research (See chapter 3 for a description of the railway infrastructure). For instance, Ribeiro et al. (2016) implemented a SVM algorithm to predict failures on train doors, and Kalathas and Papoutsidakis (2021) utilized a software built on DTs (amongst others) to predict failures on braking systems of trains. These papers are not related to the infrastructure of railways and were thereby excluded from the SLR.

2.2.2 Searching

When conducting a SLR, at least two different literature databases should be searched (Siddaway et al., 2019). For this method, the selected databases were Oria (NTNUs bibliographic database) and Scopus, which are relevant for the topic area. Searching in Google Scholar was also considered, but since Oria resulted in a comprehensive list of matches, a search in Google Scholar was deemed unnecessary. As priorly mentioned, peer-reviewed papers give better insurance of the quality of the paper. Both selected databases have a filter for showing only peer-reviewed papers. The search process was conducted using a systematic literature search. The search terms, databases, and procedures are organized and preplanned in a systematic literature search, and the author must evaluate the generated results throughout the process.

The initial search was conducted from this formulated search string: (“railway infrastructure” AND “predictive maintenance” AND “machine learning”). This search string resulted in a total of 60 matches on Oria and Scopus (18.03.2022). After investigating a large portion of the search results, the author realized that the search string needed to be expanded and decided to update the search string due to the listed reasons:

- “Railway infrastructure” was too specific and did not necessarily include articles that used different terminology
- A synonym for “predictive maintenance” was discovered that could increase the matches of relevant articles
- Not all papers referred to “machine learning”, but used “deep learning” instead. DL is presented in chapter 4.3.

Due to the listed reasons, the search string was updated, and other search operators were applied. The search term “railway infrastructure” was changed to “rail*”. The truncation symbol “*” is used to look for all terms that begin with a specific letter combination (Purssell and McCrae, 2020). Therefore, the search term “rail*” would not only include both “rail” and “railway”, but also produce results that could include “railway (or rail) infrastructure”. The boolean search operator “OR”, was added to cover more similar search terms. The updated search string was: (“rail*” AND (“predictive maintenance” OR “failure prediction”) AND (“machine learning” OR “deep learning”)). This search resulted in 244 matches in Oria and Scopus (20.03.2022).

2.2.3 Screening

The screening process was conducted to eliminate all papers that are either irrelevant or unnecessary for the research. This procedure consisted of three steps that were performed with the established eligibility criteria taken into consideration. In the first step, a good amount of the obtained papers were clearly irrelevant and eliminated by simply reading the titles. The abstracts of the papers were viewed in the next step to decide if the papers were relevant. This removed a significant number of papers. In the last step, the author was required to read various parts of the remaining papers, to be assured whether they should be included or excluded (Purssell and McCrae, 2020). This selection process is presented in Figure 2, which illustrates the elimination of papers in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart based on the eligibility criteria. Purssell and McCrae (2020) describe the flowchart as an important tool to illustrate the screening process. It is a top-down flowchart where all papers from both databases of the original search result are placed at the top. The level below illustrates the papers remaining after removing duplicates, executed in EndNote. The next level down presents the number of papers remaining (and removed) after being screened (by title and abstract) with eligibility criteria. Then the remaining papers were assessed and eliminated reviewing the full-text. The final level shows the papers included in the SLR.

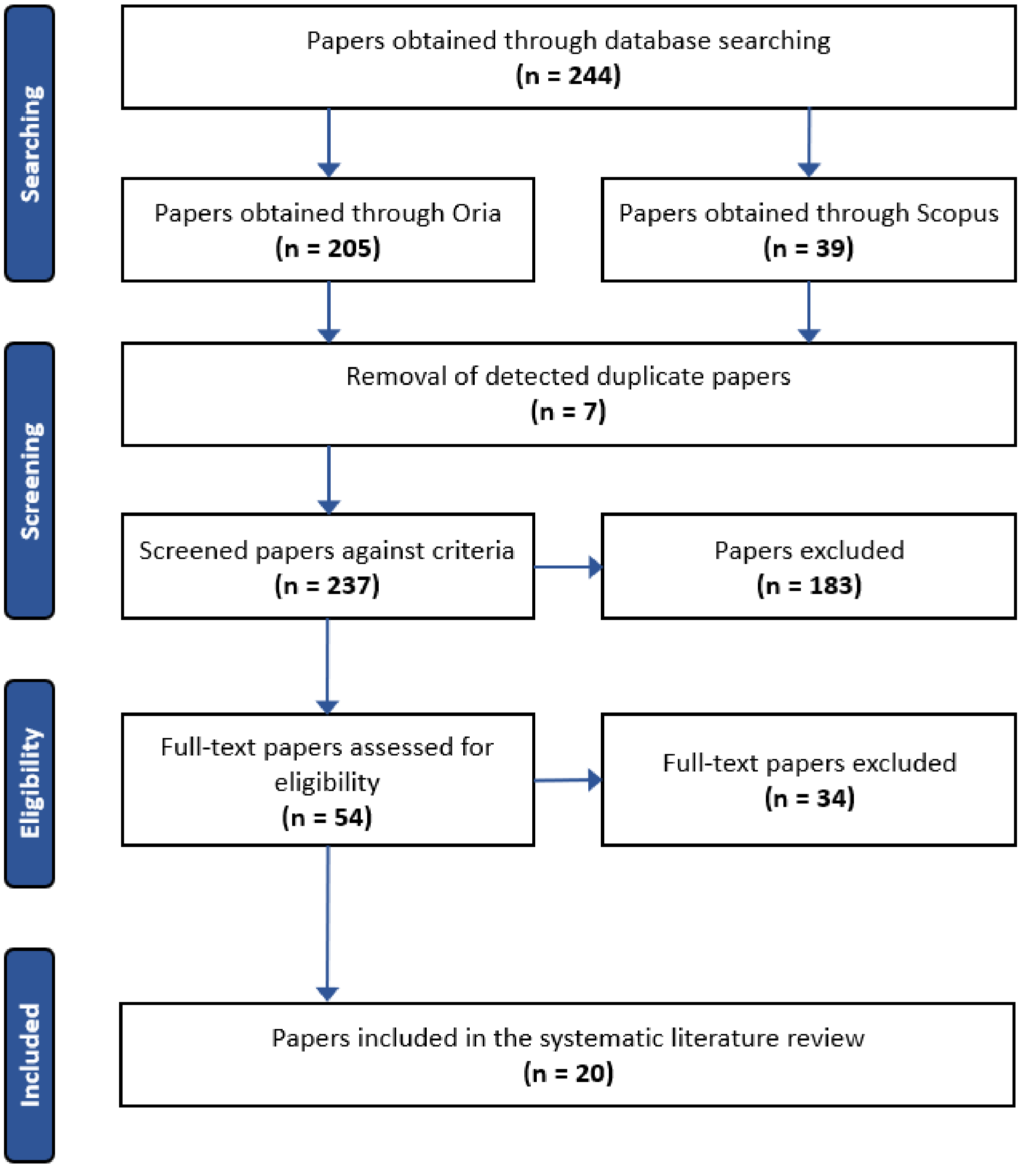

Figure 2: Selection process of papers - a PRISMA flowchart

The included papers were exported and listed in a spreadsheet. They were distributed by several categories of relevant information, e.g., publication date, author(s), inspection methods and data, main findings, and challenges. The full spreadsheet can be found in the appendix (9.3). The chosen categories were selected and updated to support the aim of this research. This spreadsheet was necessary for the author to get a structured overview of the included papers, facilitating the later analyses.

A citation analysis of the selected papers was also performed, measuring the journal impact factor. This analysis measures the total number of citations for each paper. The citations were obtained from Google Scholar, including citations by peer-reviewed articles and academic theses (Purssell and McCrae, 2020).

2.2.4 Quality Assessment

Inclusion and exclusion criteria are determined to ensure that all relevant papers have been identified (Siddaway et al., 2019). Regardless, the papers will have a different quality, and the possibility of some relevant papers being unidentified in the SLR is probable. Although the utilized search string is comprehensive and captures many papers, perhaps more relevant papers could have been identified with other search terms, in other databases, or by using more languages. This could have potentially changed the results and outcome of this SLR.

Purssell and McCrae (2020) describe critical appraisal as "the balanced assessment of a piece of research, looking for its strengths and weaknesses and then coming to a balanced judgment about its trustworthiness and its suitability for use in a particular context". The papers that had potential of being included were examined through a quality assessment form (see Appendix 9.1). This form focused on the quality of the methods, meaning an investigation of the research design of each paper and how well it was conducted (Purssell and McCrae, 2020). It comprises questions about the purpose and scope of the publication, research design, data collection, and outcomes, to mention a few.

Most of the obtained papers in the SLR appeared strong by being innovative and presenting strong results. Therefore, limitations, challenges, and future work stated in each paper were thoroughly examined and listed in the spreadsheet, leading to an overview of the weaknesses of the papers. This created a balance between the strengths and weaknesses and established a weight from stronger to weaker papers. Additionally, as mentioned previously, this SLR was restricted to peer-reviewed papers, which should have ensured reliable and valid sources.

Risk of Bias

The mindset of this SLR is to be objective towards all conducted research and gather all relevant papers on the topic of interest. However, it is inevitable to prevent the research from being completely unbiased. Purssell and McCrae (2020) define bias as "a systematic error, or deviation from the truth". Regarding bias in this SLR, the researched topic does not provide a vast amount of papers. Therefore, it was inconvenient for the author to select papers that favored the outcome of this research. The criteria in 2.2.1 created a solid theoretical foundation for which papers to include and prevented a biased selection. Since the research was restricted to English might have created a bias. The fact that the SLR was conducted by the author alone, is another potential risk of bias and weakness. Ideally, it should be reviewed by more than one, leading to discussions of which papers to include and a cooperative evaluation of the risk of bias. The author may be biased by his thoughts on what methods and analyses should be utilized to produce the correct results and to reach the final goal of the SLR.

The risk of bias within each included paper was also considered. This is not easily detected. When it is detected, the extent of the bias remains uncertain. However, the quality assessment form was applied to mitigate the risk of bias, e.g., bias in the selection of results, missing outcome measurements, and diversion from the initial scope of the research (Purssell and McCrae, 2020).

2.3 Semi-Structured Interview

During the project, two interviews were conducted. The first interview was carried out on 14.03.2022 and the second on 20.05.2022. Both were executed online through Microsoft Teams. The author desired to keep an open but structured dialogue with the participants, and therefore chose to arrange SSIs. A SSI is a qualitative approach for data collection where the questions are preplanned but open-ended, usually with follow-up questions, with one respondent at a time (Adams, 2015). In this type of interview, the researcher has more control over the interview topic than in an unstructured interview, but unlike structured interviews or surveys that incorporate closed questions, there is no set range of answers to each question (Given, 2008).

SSIs are acknowledged as a conversation that might wander around the agenda topics and possibly uncover unexpected issues, but still be structured enough to not derail from the topic of interest (Adams, 2015). This can be seen as one of the major strengths of SSIs. On the other hand, SSIs also bear weaknesses. They require that the interviewer is knowledgeable to make the process effective and to produce reliable results. In contrast to the benefits of follow-up questions, they also reduce the reliability of the findings since each interview might vary from one to another (Adams, 2015).

To mitigate the degree of the abovementioned weaknesses, an interview guide should be constructed. Adams (2015) suggests the following recommendations for structuring an interview guide:

1. Set aside enough time to formulate the questions
2. Do not have too many questions on the agenda, but focus on the critical ones
3. Closed questions can be ideal entrances to follow-up questions
4. The order of the questions can change during the interview. If this happens, proceed with the topic of interest, and return to the question that was skipped later

These recommendations were used when formulating the interview guide, which can be found in the appendix (9.2). The guide presents the main structure of the interview, but the author did not hesitate to ask relevant and unforeseen follow-up questions. Due to unforeseen follow-up questions in the first interview, a few questions were added in the second interview.

2.4 Evaluation of Selected Methods

The author spent the first weeks of the project to gain knowledge on the topic, mainly ML types, tasks, and algorithms. These areas were studied through the LR. This was beneficial for both the author and the report in its entirety. However, the time spent on the LR led to less time for conducting and analyzing the SLR. In addition, perhaps more interviews could have been performed, given this time. The following section reviews the reliability and validity of the selected methods.

2.4.1 Reliability and Validity

Purssell and McCrae (2020) define reliability as “the degree of which results obtained by a particular measurement can be replicated”. In terms of the SLR, the conducted research has an excellent level of reliability. The process of identifying relevant literature has been described thoroughly, by a detailed explanation of the eligibility criteria, search and screening process, and quality assessment. Purssell and McCrae (2020) define validity as “the degree to which evidence and theory support the adequacy and appropriateness of the interpretations and actions that come from the results”. The validity of each paper was studied through the quality assessment form, mitigating the risk of bias. In addition, the search string was restricted to peer-reviewed papers and built on the problem definition of this report, ensuring a collection of valid and reliable material to pursue the purpose of the report. The findings are valid and reliable, but one can discuss the potential negative influence of the author conducting the SLR alone. As mentioned above, a SLR conducted alone can impact the selection of papers, methods for analysis, and results. The author has examined the analysis and findings several times, attempting to secure their reliability and validity. Despite this, it can be argued that the level of reliability and validity would be improved if more individuals had been involved to eliminate possible mistakes.

Regarding the SSIs, reliability is mentioned as an issue. Reproducing the interview is difficult due to the nature of an SSI being somewhat loose in structure. The answers from the participants may gradually change over time, since the questions consider technological advancement. However, the SSI structure was based on an interview guide that slightly increased the reliability. Another issue that can be related to the SLR, is that the interview was conducted by the author alone. The outcome of the interview can be affected by personal bias. To mitigate this, the author sent a copy of the final contribution of the interview to the participants for confirmation, also increasing the validity of the SSI.

The LR is reliable since the searches have been presented in detail. To ensure validity, the identified sources were constantly evaluated by, e.g., comparing one source to another on the same topic. Many sources were investigated that backed the relevant theory needed. Peer-reviewed sources were used since they have high validity and reliability. The sources provided by the supervisor are assumed to be quality assessed, and the author has previously verified the sources from the project thesis.

3 Railway Infrastructure Maintenance

The railway conditions impact railway infrastructure maintenance, necessitating expert knowledge of each railway asset, its relationship to other assets, and their degradation patterns. As a result, scheduling maintenance for the asset groups is demanding, leading to challenging maintenance operations. For planning and scheduling maintenance and operations, geographic and geological characteristics, terrain, and climate conditions, among other factors, are critical. During maintenance planning, other aspects such as availability, pricing, resources, downtime, and more are actively considered. All these considerations are crucial for maintenance operations to avoid infrastructure disruptions, maximize capacity utilization, and minimize costs. The goal of railway maintenance is to accomplish the best approach for ensuring and optimizing efficiency, availability, and safety (Espling and Kumar, 2007, Patra, 2009, Connor, 2019, Lamberts, 2009, Matic, 2021). This chapter briefly presents the railway infrastructure, its belonging assets, and PdM in railway infrastructures.

3.1 Railway Infrastructure Assets

An asset is defined differently depending on the railway infrastructure organization, and it might include both physical and non-physical assets. The following definition for an asset is provided by ISO 55000 (UIC, 2016):

"An item, thing or entity that has potential or actual value to an organization."

The railway infrastructure is comprehensive, and the assets that make up the infrastructure can be separated into various groups. One group is superstructure, and it consists of, e.g., switches, sleepers, and rails. The next group is substructure, including assets such as track ground foundation, embankment, drainage elements, and fences. Another group is signaling systems for railway traffic control. Other groups are electrical assets such as catenary masts, cables, heating and illumination elements, and telecom systems, including systems for communication, detectors, and more (BaneNOR, 2020, Matic, 2021). Some of the railway infrastructure assets are depicted in Figure 3.

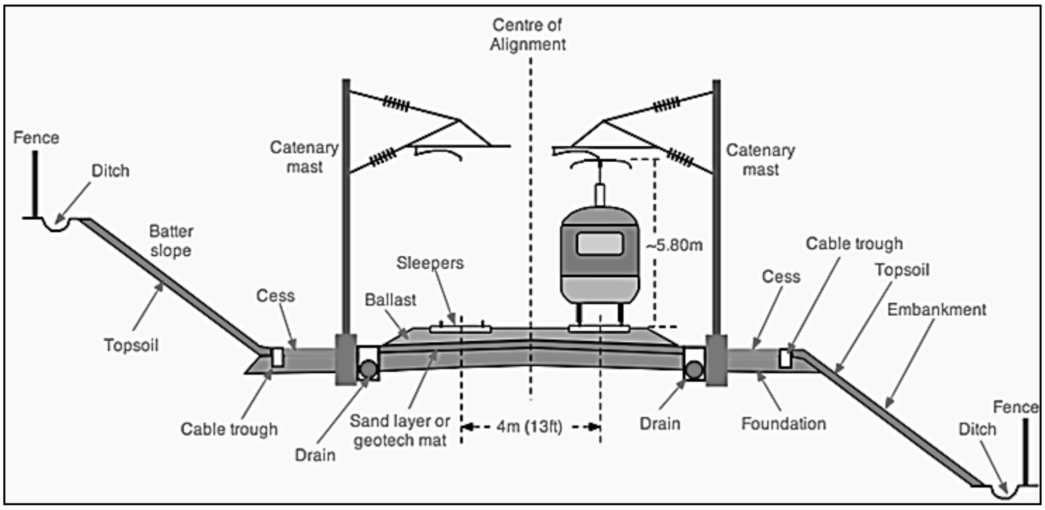

Figure 3: Railway infrastructure assets (adapted from Connor, 2019)

Railway Infrastructure assets are increasing in complexity with regards to their quality, reliability, efficiency, and availability. This is the result of the continuous focus on globalization and technological advancement (Stenström, 2014). Therefore, the maintenance industry must adapt to emerging developments and offer new inventive solutions. Due to the same reasons, customer service is becoming increasingly important. Customer service is critical, which creates additional strain on railway maintenance activities in terms of planning, scheduling, and duration. To meet these asset requirements, a proper maintenance strategy is necessary (Matic, 2021, Pintelon and Parodi-Herz, 2008). According to Xie et al. (2020), railway tracks and switches are the most critical assets of the infrastructure. These assets are exposed to intense traffic levels, high axle loads, and different climate conditions, indicating that a minor fault may lead to infrastructure disruption. Disruptions cause stoppage of traffic, wasting time and resources, and can lead to potentially high costs. Railway track and switches are described below, each consisting of several complex sub-assets.

3.1.1 Railway Track

The main purposes of a railway track are to guide the vehicle and carry and distribute the weight safely. Figure 4 illustrates the assets of a railway track, which are rails, sleepers, fasteners, ballast, and joints. Rails are the longitudinal steel components that evenly and continuously guide the train wheels. Joints attach two rail segments, and sleepers are spans that extend over the track and connect the two rails. The main functions are to hold the fasteners to sustain the correct track gauge, carry the rail load, and distribute it across the underlying ballast at an adequate pressure level. Rail fasteners keep the rails on the sleepers, maintain track gauges, and withstand vertical, lateral, longitudinal, and overturning rail motions. Sleepers are laid on top of crushed stone called ballast (Tzanakakis, 2013).

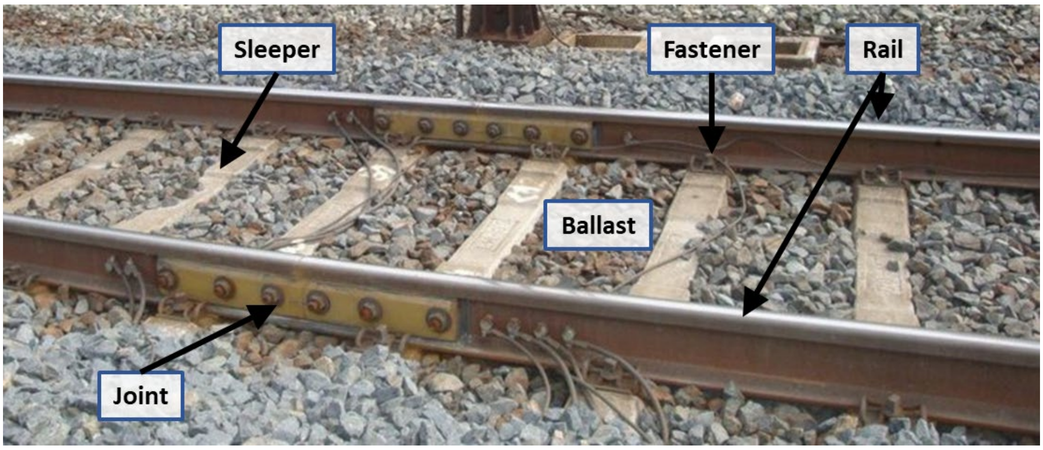

Figure 4: Railway track assets (adapted from Németh and Fischer, 2019)

Railway track defects can be categorized into structural defects and track geometry defects. Structural degradations of railway track assets, e.g., rails, sleepers, ballasts, and joints, refers to structural defects. For instance, joint and ballast deterioration, corrugation and wear, and loose or missing fasteners (Tzanakakis, 2013). Railway track geometry defects are associated with anomalies from desired values of track geometric parameters. For instance, anomalies in alignments, longitudinal level, twist, and gauge (Nakhaee et al., 2019).

3.1.2 Railway Switches

Railway switches (also known as turnouts, and switches and crossings) come in many types, but they all provide the same main function, which is to redirect the vehicle onto a different track (BaneNOR, 2020). Switches consist of several components (see Figure 5), each requiring complicated engineering (Dindar and Kaewunruen, 2017).

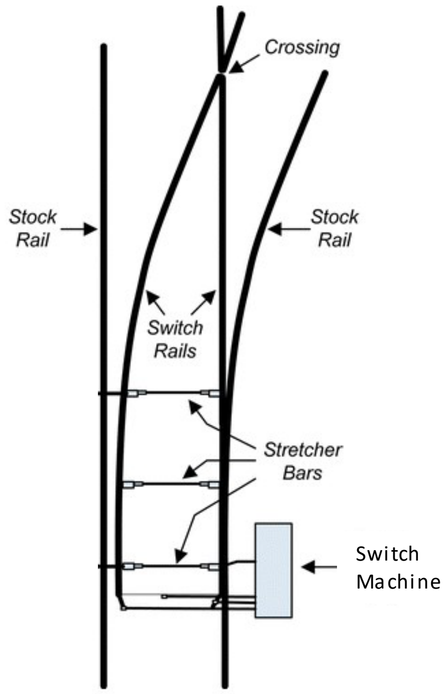

Figure 5. Common layout of railway switches (Rama and Andrews, 2013)

Rama and Andrews (2013) and Dindar and Kaewunruen (2017) provide the following description of the main components of switches:

- A switch machine (also known as a point machine) is an electric, hydraulic, or pneumatic mechanism to perform the motion of switch rails to the redirected track
- Stretcher bars are steel bars to maintain the rails in the correct position for passing railway vehicles
- Switch rails that can be moved to direct the vehicle to the desired track
- Stock rails have the same function as standard rails, keeping the rails at a correct distance
- Crossing refers to the point where two rails cross paths
- Switch heater (or point heater) maintains secure operability during extreme weather conditions. These weather conditions can be detected by utilizing sensors.

Common failures to railway switches are faults related to the switch, stretcher bars, rails, geometry, ballast, and signal systems. Faults caused by the switch can be related to the switch being damaged, e.g., broken, worn, bent, or disconnected. Other faults can be linked to the stretcher bars not working sufficiently, and similar to geometry, ballast, and rails as with standard railway tracks. Signal system failures are also common, caused by the failure of signaling and sharing crucial information about the status of the railway switch (Dindar and Kaewunruen, 2017).

3.2 Predictive Maintenance in Railways

The EN 13306:2017 defines PdM as (BSI, 2017):

> "Condition-based maintenance carried out following a forecast derived from repeated analysis or known characteristics and evaluation of the significant parameters of the degradation of the item."

PdM has grown parallel to maintenance becoming increasingly digital, transforming from purely corrective and preventive maintenance (Matic, 2021). One big contributor to this change is cheaper and more accessible condition monitoring technology (Stenström, 2014). Condition monitoring equipment that monitors, analyzes, and evaluates railway assets, results in more frequent PdM intervals. A monitoring system identifies possible deterioration by determining an anomaly to the desired value (Tzanakakis, 2013, Matic, 2021). Compared to many other industries, PdM techniques in the railway sector are relatively new (Davari et al., 2021).

The purpose of PdM is the failure prediction of an asset is to perform preventative activities before asset failure. Failure prediction investigates the chance of an asset fault occurring within a given time. This process leads to an extended lifetime for the asset and eliminates unexpected failures that may cause a breakdown of the entire system, reducing both cost and downtime. Another purpose is to estimate RUL of one asset or a set of assets. RUL uses data to identify when an asset is likely to fail (Davari et al., 2021, Tzanakakis, 2013, Matic, 2021, Bukhsh and Stipanovic, 2020).

In contrast to these advantages, PdM also have challenges related to data collection and management, and implementation. Data is collected in huge amounts by different inspections and sensors, and is difficult to efficiently manage and process by using digital tools. Implementation of PdM brings more new technologies into the spotlight, and due to insufficient knowledge of prediction models, the output can be difficult to comprehend (Bukhsh and Stipanovic, 2020).

Predictive railway infrastructure maintenance must be supported by proper tools for monitoring and analyzing the condition (Matic, 2021).

3.2.1 Inspection Methods and Tools

Railway infrastructure inspection consists of several methods and tools for collecting relevant data. Certain inspections are carried out more frequently than others. The methods can be separated into manual inspection, vehicle inspection, and inspection by fixed sensors (Xie et al., 2020). Manual inspection is performed by operators visually inspecting the infrastructure to possibly detect faults or anomalies, such as missing or defect catenaries, fasteners, or rails. This inspection type is not related to failure prediction, but correlates to corrective maintenance. In vehicle inspection, different vehicles installed with measurement devices are exploited. These devices can be, for instance, cameras or sensors. Examples of sensors are thermals, accelerometers, and lasers. Since manual walking inspections are conducted along the track, they are somewhat dangerous and inefficient. Therefore, railway agencies have widely adopted measurements with cameras to detect surface defects or missing components. However, inspections with human eyes or with cameras are not able to detect faults within the assets. For this issue, ultrasonic testing and different sensors are applied. Sensors installed along the track use the inherent technologies to capture and measure relevant data such as vibration, data, temperature, sound, and geometry. This data is used for failure prediction and RUL in PdM (Xie et al., 2020, Jing et al., 2022).

4 Machine Learning

ML has been given numerous formal definitions. Alpaydin (2016) defined ML as "programming computers to optimize a performance criterion using example data or past experience". Burkov (2019) defined ML in his book as "a subfield of computer science that enables computer programs to perform prediction, diagnosis, planning, and recognition of behavior patterns by learning from historical data, i.e., without prior knowledge". The various definitions all share the idea of teaching computers to execute tasks besides traditional calculations by learning from their surroundings through repeated instances.

4.1 Machine Learning Types and Tasks

A ML algorithm is a data processing technique that utilizes input data to accomplish a task without being explicitly designed to produce a particular output. These algorithms are programmed with the ability to automatically adjust or adapt their structure through repetition to progressively improve at accomplishing the desired task. This process is known as training, and it involves samples of input data along with targeted outputs. Afterwards, when given new and previously unknown data, the algorithm optimizes itself to produce wanted results. Over time, the algorithm proceeds to improve and learn from its mistakes (El Naqa and Murphy, 2015).

ML algorithms have a variety of methods to respond and adapt to training. For example, the selection and weight of the input data, iterative optimization to alter the variable numerical parameters of the algorithm, and/or structuring pathways for the best results. As Figure 6 shows, ML can be classified as supervised, unsupervised, semi-supervised, or reinforcement learning (Kang et al., 2020, El Naqa and Murphy, 2015).

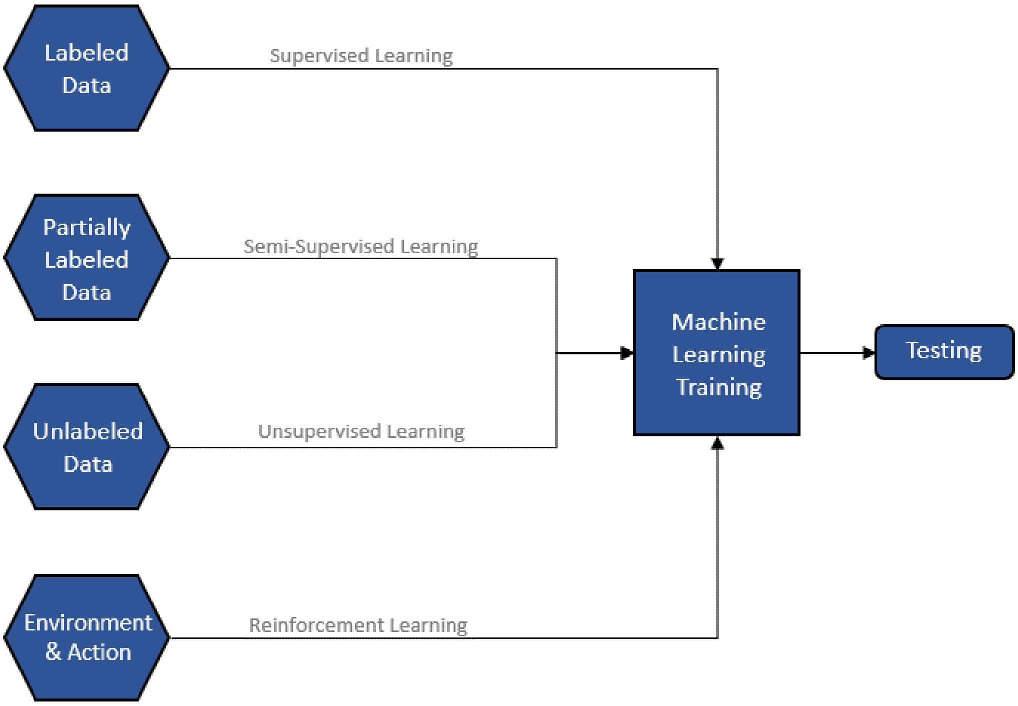

Figure 6. Machine learning types

- <u>**Supervised Learning**</u>

  Based on the assumption that all previous training examples are labeled, the computer program develops a function mapped by input(s) and output(s) from a collection of labeled training data (e.g., classification and regression). Human interaction is crucial in supervised learning. People choose (based on assumptions) the features, algorithms, and control parameters, in addition to labeling the output for the training set. This learning type is frequently mentioned and utilized in sectors where people have specific expertise for a model (El Naqa and Murphy, 2015, Kang et al., 2020, Jo, 2021).

- <u>**Unsupervised Learning**</u>

  In unsupervised learning, all training examples are assumed to be unlabeled, and it is only supplied with input data. This learning type is commonly used to define the similarity metric between unlabeled data. While supervised learning provides an output value, unsupervised learning defines a pattern of input variables and often displays several clusters based on the input data. Therefore, Data Clustering is a common task where unsupervised learning algorithms are applied (Jo, 2021, Kang et al., 2020, El Naqa and Murphy, 2015).

- **<u>Semi-supervised Learning</u>**

  Text/image extraction systems are examples of semi-supervised learning types. It is a mixture of supervised and unsupervised learning. Here, a data segment is partially labeled, and the labeled segment is used to deduce the unlabeled segment. The goal is to utilize unlabeled examples, which are cheaper to obtain, in addition to the labeled examples for training the learning algorithms. By doing this, the model can be trained to improve accuracy compared to supervised learning, which operates with very limited labeled data (El Naqa and Murphy, 2015, Kang et al., 2020, Jo, 2021).

- **<u>Reinforcement Learning</u>**

  This type is defined as the interaction between the agents and the environment. The external environment provides the input, and the output is formed as an action. Positive actions are rewarded, and negative actions are penalized by the environment. To maximize the rewards and minimize the penalties, the parameters are updated whenever possible (Jo, 2021, Kang et al., 2020).

ML algorithms are applied to perform different tasks, depending on what is desired. The tasks can be divided into four common tasks, as illustrated in Figure 7.

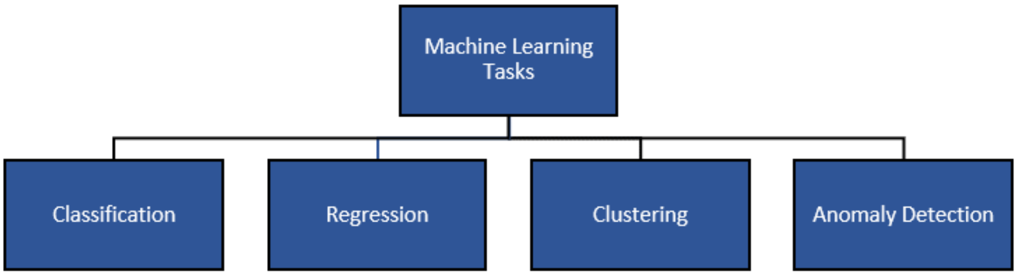

Figure 7: Machine learning tasks

- **<u>Classification</u>**

  The action of assigning one or more of the predetermined classes to each item is known as classification (Jo, 2021). In classification, input features are mapped to one of the discrete output variables. The output variable represents the underlying problem (Kang et al., 2020). The classification types are depicted in Figure 8. They can be binary or multiple, where binary classification is the least complicated. In binary classification, the output variable consists of two classifiers, positive and negative (or one and zero). In multiple classification, more predefined classes can be utilized (Kang et al., 2020, Jo, 2021).

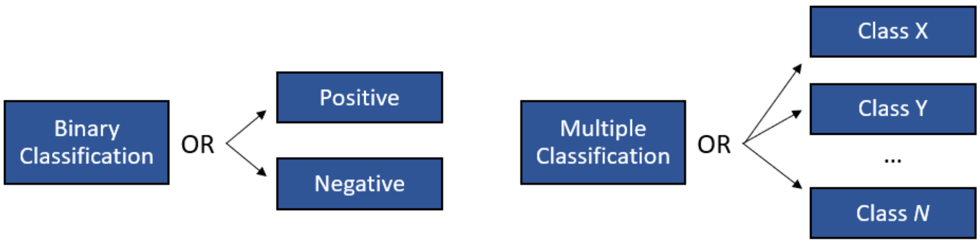

Figure 8: Classification types

- **Regression**

  While classification predicts discrete outputs, the goal of regression is to predict continuous outputs through mapped input features (Bi et al., 2019). Regression can be defined as the action of estimating an output value consisting of several factors. The estimated output can be an integer or a floating-point number (Kang et al., 2020). There are two forms of regression, univariate and multivariate (Figure 9). The univariate regression estimates only one output value, and the multivariate regression estimates more than one output value. As stated previously, regression and classification are tasks applied in supervised ML (Jo, 2021).

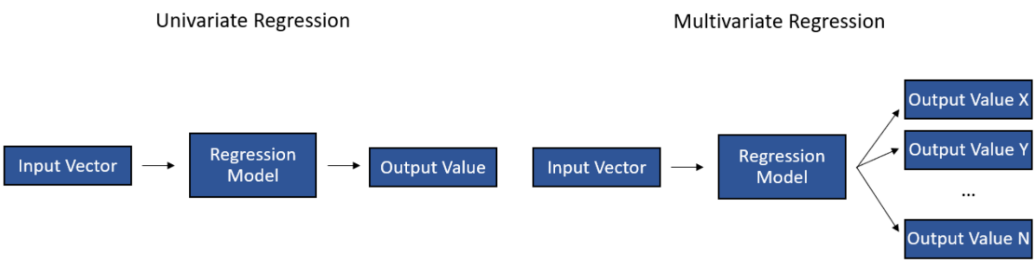

Figure 9: Regression types

- **Clustering**

  Clustering is the action of segmenting data into groups, where each group contains data based on similar data characteristics (Kang et al., 2020). The data clustering is performed by creating clusters based on data structure similarities. The clusters or classes developed are then labeled. When trained, the algorithm adds new unseen data to respective clusters (Alzubi et al., 2018). By automatically collecting labeled training examples through clustering, it can be integrated with classification to classify datasets (Jo, 2021). Contrary to classification and regression which are utilized in supervised learning, clustering is a common unsupervised learning implementation (Bi et al., 2019). A general view of a clustering model is illustrated in Figure 10.

Figure 10: General view of a clustering model

- **<u>Anomaly detection</u>**

  Anomaly detection groups the data similarly to clustering. It is the process of analyzing an established pattern and detecting anomalies or changes to this pattern. These outliers are identified in the dataset through specific algorithms. This detection task is commonly used in unsupervised learning (Kang et al., 2020, Alzubi et al., 2018). Figure 11 presents the general view of an anomaly detection model.

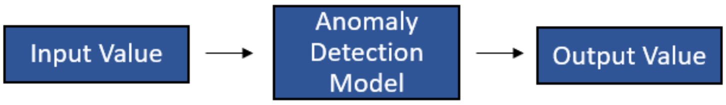

Figure 11: General view of an anomaly detection model

4.2 Machine Learning Algorithms

As stated by Arena et al. (2022) in chapter 1.4, supervised and unsupervised learning techniques are the most common for PdM applications. Xie et al. (2020) argued that unsupervised learning, DL, and ensemble methods are growing. Therefore, these methods and algorithms are presented in the next subchapters. DL algorithms are presented in chapter 4.3.

4.2.1 Artificial Neural Network

An artificial neural network (ANN) is a computational model inspired by biological neurons and their way of processing input data through several layers of interconnected neurons or nodes, to the final output (Baloglu et al., 2021, Çinar et al., 2020). This structure is the foundation for ANNs analysis of intricate interactions between a number of measurable variables to predict an output (Bi et al., 2019). ANNs can simultaneously execute any amount of classification and/or regression tasks, but one task is common for each network (Ouadah et al., 2022). The layers consist of an input layer, one hidden layer, and an output layer, as shown in Figure 12. Depending on the learning type, the algorithm can be classified as a supervised neural network or an unsupervised neural network (Alzubi et al., 2018). ANNs learn the basic principles from a series of provided symbolic situations, rather than following the set of laws established by human experts. Due to its capability to learn from examples, ANN models are broadly utilized in many industries. The algorithm is advantageous for systems with large quantity of complex and ambiguous information (Çinar et al., 2020).

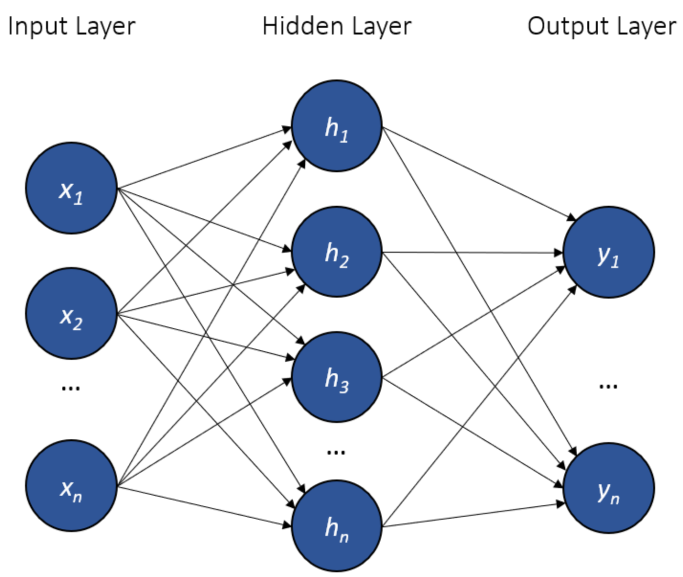

Figure 12: Diagram of Artificial Neural Network

4.2.2 Decision Trees

A decision tree (DT) groups the domain into several linear areas and predicts outputs using a rule-based method (Suthaharan, 2016). The grouping is based on attributes and their respective values, and the tree-network is built up of nodes and branches (Ouadah et al., 2022). The nodes signify attributes or categories, and extending branches from a node corresponds to an attribute-related value or a value interval (Jo, 2021). As shown in Figure 13, the tree structure consists of a root node, sub-trees, decision nodes, and final nodes. Classification trees and regression trees (CART) are two types of ML decision trees. Regression trees predict continuous outputs, while classification trees predict discrete outputs (Bi et al., 2019).

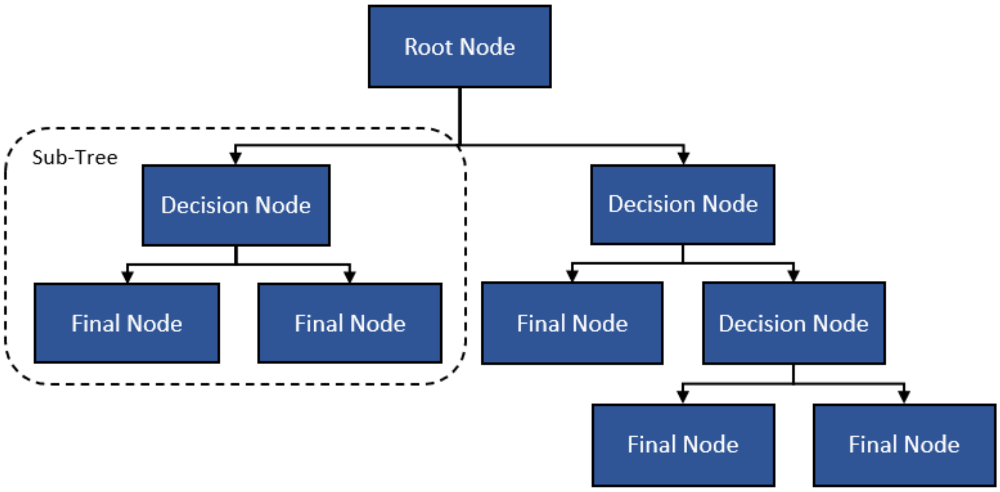

Figure 13: Structure of decision trees

4.2.3 Random Forest

Random Forest (RF) is an ensemble method that collects decision trees using a random data set with replacements (Ouadah et al., 2022). The trees can be classification or regression trees, and therefore RF can be utilized for classification tasks and regression tasks. In contrast to DT, which provides one trained decision tree, the RF provides several trained decision trees (Suthaharan, 2016). The final decision trees and output category are determined by the combined output of all prior decision trees in the RF (Alzubi et al., 2018). The RF process is illustrated in Figure 14. The training dataset is divided into subsets, and each subset is used to build the decision tree. The trees then classify the data and the final output is determined by voting or averaging the decision tree outputs (Jo, 2021).

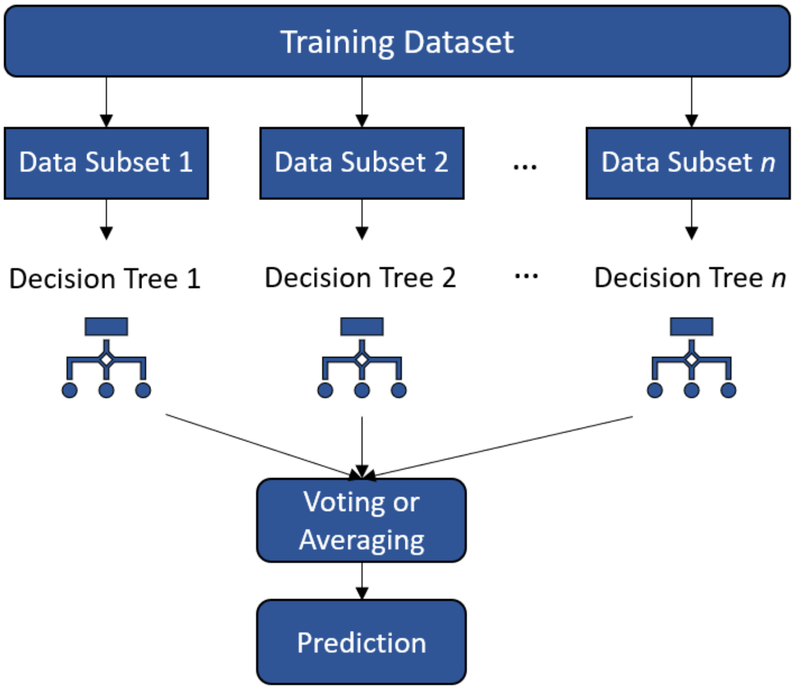

Figure 14: The process of random forests

4.2.4 Gradient Boosting

Boosting is an ensemble method that improves the robustness of a single estimate by combining the predictions of numerous base estimators, creating a strong predictor (Ouadah et al., 2022). Gradient boosting (GB) utilizes gradient descent to improve the performance of classifiers. Gradient boosting tree (GBT) decision tree based on GB (Bi et al., 2019).

4.2.5 Support Vector Machines

Support Vector Machines (SVM) is a supervised learning algorithm that can be used for both classification and regression tasks. SVMs exercise around the theory of margin calculation and construct an optimal hyperplane that divides the dataset into two groups. The data is divided by determining their value of a number of features in a dimensional space, as illustrated in Figure 15 (Alzubi et al., 2018). Although, several data observations regularly need to be transformed before the hyperplane can separate them (Bi et al., 2019). It has been proven that maximizing the margin and therefore generating a big distance between the hyperplane and the data points on either side, reduces the predicted generalization error (Ouadah et al., 2022). SVMs provide high accuracy when solving big data problems, but they are computationally costly and time-consuming as it is mathematically complex

(Suthaharan, 2016). Support vector regression (SVR) is an expansion of SVM to predict continuous outputs (Bi et al., 2019).

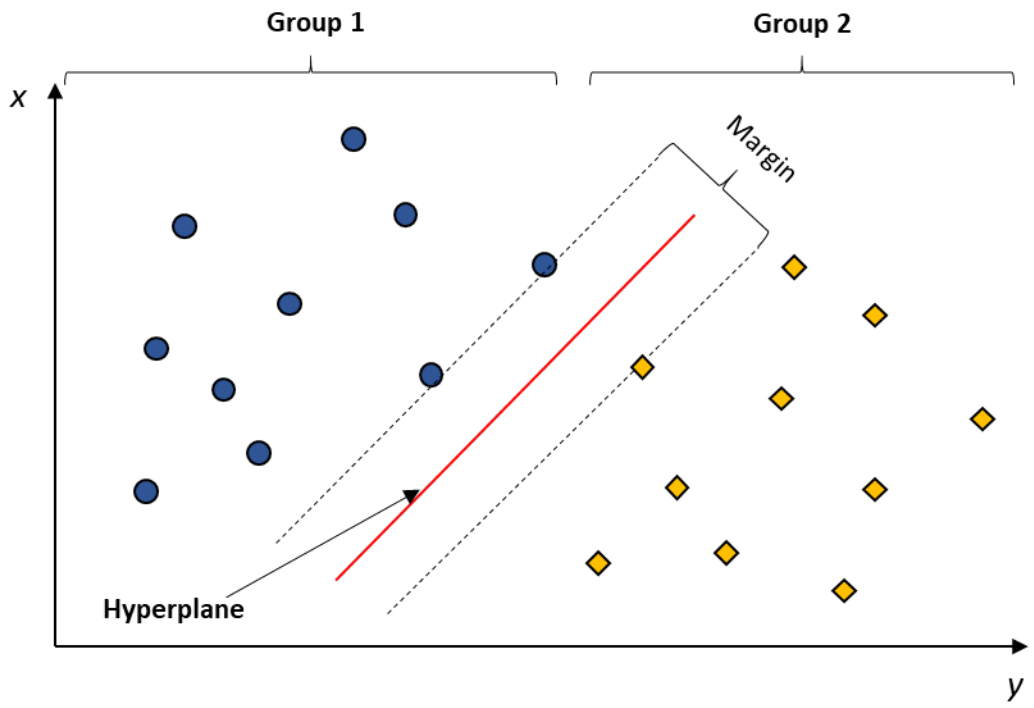

Figure 15: Functional view of support vector machine algorithm

4.2.6 K-Nearest Neighbors

K-Nearest Neighbors (kNN) is a supervised ML algorithm that can be used for both regression and classification tasks, but is commonly more used to solve classification problems (Ouadah et al., 2022). It is one of the simplest algorithms to implement and interpret. The objective is to store the training dataset and subsequently predict the label of all new instances based on its nearest neighbor label in the training dataset (Shai and Shai, 2014). The given input training datasets consist of k-values that are nearest to the new variable used in the featured set. The output depends on whether kNN is used as a classification or regression algorithm (Alsharif et al., 2020). Figure 16 illustrates the functionality of kNN.

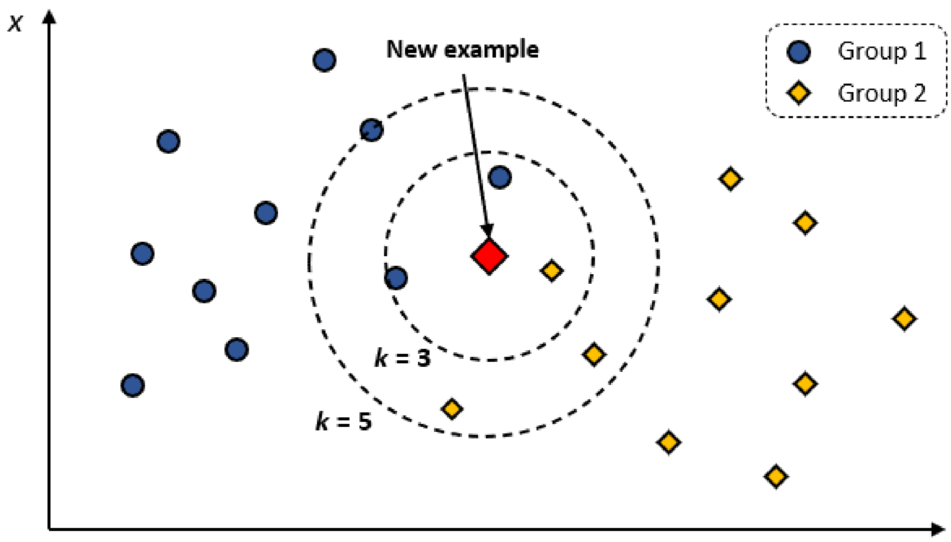

Figure 16: Functional view of the k-nearest neighbor algorithm

4.2.7 K-Means Clustering

K-Means Clustering is an unsupervised ML clustering algorithm utilized for analyzing datasets and grouping them into clusters. In this algorithm, "k" indicates the number of clusters that should be determined prior to the algorithm (Alzubi et al., 2018). Figure 17 gives a simple representation of how an unlabeled dataset is grouped into labeled clusters. The grouping of x observations is randomly selected as the initial mean vectors. Each observation is sorted into the cluster with the most equivalent mean vector. The center of the cluster is established by the mean of the observations in the cluster, respectively. The mean vectors and sorting of observations of the cluster are updated and iterated until the mean vectors converge (Jo, 2021).

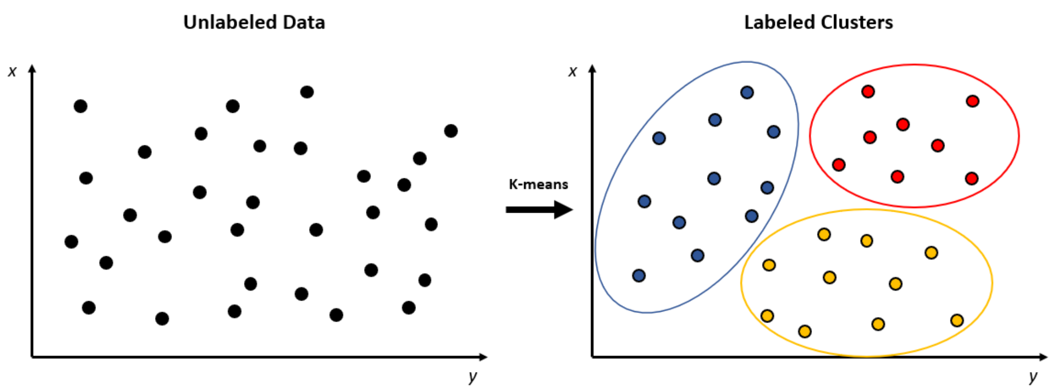

Figure 17: Functional view of k-means clustering algorithm

4.2.8 Probabilistic Learning (Naïve Bayes)

Probabilistic learning (PL) is a type of supervised ML that utilizes the Bayes rule (see equation 1) when calculating the probability of a group given an observation (Jo, 2021). Naïve Bayes is a classification algorithm that assumes independence between predictive variables (Bi et al., 2019). It is derived from a set of probabilistic classifiers, tolerates data with high dimensionality, and requires fewer datasets for training (Alsharif et al., 2020). The Bayes rule calculates the conditional probability, P(A|B), which calculates the possibility of event A, given that event B occurs (Alzubi et al., 2018).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

Afterwards, it selects the optimal group with the highest probability (Bi et al., 2019). The probability of an event occurring generates trees that are called Bayesian networks. The structure of these networks consists of an original node and several sub-nodes, with the same assumption of independence (Ouadah et al., 2022).

4.2.9 Linear Regression and Logistic Regression

As previously mentioned, regression is used to predict a continuous value, and two common algorithms are linear regression (LnR) and logistic regression (LgR). Given a set of independent variables, linear regression estimates continuous output, while logistic regression provides discreet output (Çinar et al., 2020). Another difference is that linear regression is applied to regression tasks, and logistic regression is applied to classification tasks (Ouadah et al., 2022). Another type is called multiple regression (MR), which compromises more input variables than LnR to train the model.

4.2.10 Dimensionality Reduction Algorithms

Modern ML algorithms are inefficient and rigid when dealing with enormous amounts of data. Data with high dimensionality has been shown to be a burden in data processing. The sparsity of the data is another problem, and finding an optimum for such data is costly and time-consuming. Dimensionality reduction algorithms decrease the computing cost by lowering the number of dimensions of the data. This is accomplished by decreasing excessive and unrelated data and thereby increasing accuracy (Alzubi et al., 2018). Some examples of dimensionality reduction algorithms are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Principal Component Regression (PCR).

4.3 Deep Learning

DL is a part of ML and utilizes ANNs on a "deeper" level. Instead of solving nonlinear problems with a single nonlinear model, it applies multiple steps of linear models. This means that it can have multiple hidden layers from the input data to the output value, rather than having one (such as ANN) hidden layer (Jo, 2021, Xiao and Sun, 2021c). The reason for the growth of deep neural networks is due to innovative technologies and advancements. For example, sensors and IoT creates large amounts of data making it more demanding for traditional ML algorithms. Therefore, DL approaches will be one of the future PdM solutions (Davari et al., 2021). Hence, the following subsections present some DL algorithms.

4.3.1 Deep Neural Network

A Deep Neural Network (DNN) is an ANN with more than two hidden layers between the input and output layer (see Figure 18). The network calculates the probability of each output value through the flow that moves from one layer to the latter (Davari et al., 2021). Several DNN types are developed depending on the different forms of input data. For instance, a Convolutional Neural Network (CNN) is applied to modeling grid data (for example, images or time series), and Recurrent Neural Networks (RNN) are convenient for modeling sequential such as text sequences (Xiao and Sun, 2021b).

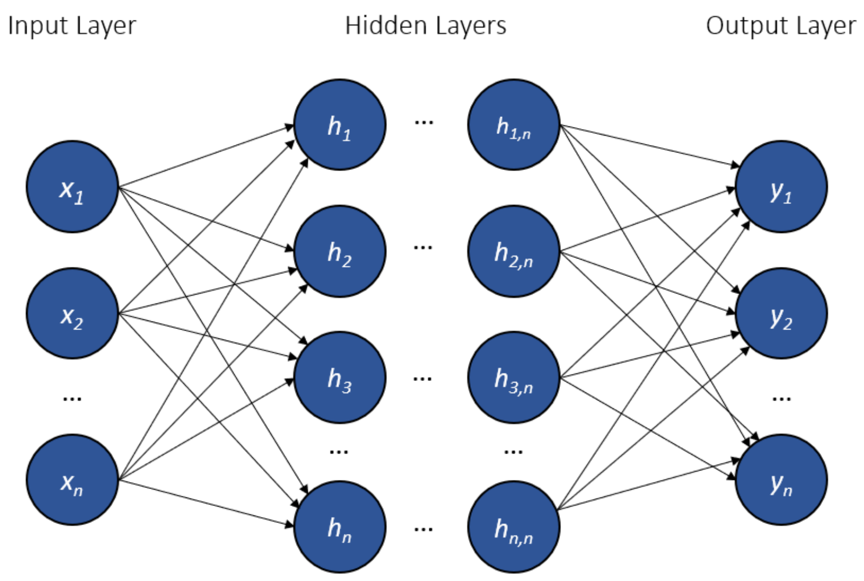

Figure 18: Diagram of deep neural network

4.3.2 Convolutional Neural Network

CNN is a type of DNN that processes large-scale images by learning its features to perform recognition and classification. As shown in Figure 19, it consists of layers of convolution, pooling layers, followed by fully connected layers (Xiao and Sun, 2021a). The convolution layer collects features from the input data and creates feature maps, while the pooling layer reduces the input samples and the dimension. (Rengasamy et al., 2020). This step of CNNs is often repeated multiple times before employing the fully connected layers to combine all features, generating more abstract features for classification tasks (Xiao and Sun, 2021a). CNNs are applied to PdM practices such as RUL and fault diagnosis (Davari et al., 2021).

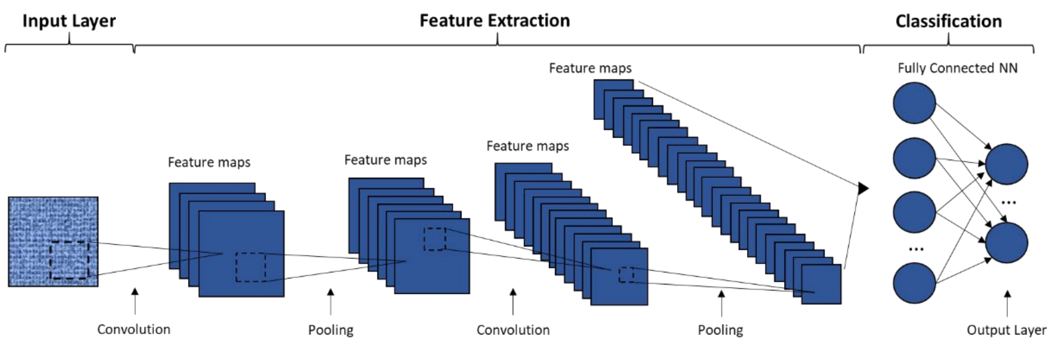

Figure 19: Functional view of a convolutional neural network

4.3.3 Recurrent Neural Network

Recurrent Neural Networks (RNN) are supervised types of deep neural networks for classification of sequential data or time-series data such as text records. RNN models carry input data as a series of vectors, and each vector is mapped to the related hidden layer. The hidden layer produces the output layer (Xiao and Sun, 2021d). In RNN, internal feedback loops are possible in contrast to CNN, which uses pooling layers (Davari et al., 2021). These internal loops cause recursive behaviors in the networks, introducing delayed initiation dependencies throughout the network's processing elements (Marhon et al., 2013). Long short-term memory (LSTM) and gated recurrent unit (GRU) are two types of RNN that have grown in popularity, due to their ability to solve technical challenges associated with collecting long-term dependencies and managing sensor data to perform predictions (for interested readers, see: Davari et al. (2021), Xiao and Sun (2021d), Rengasamy et al. (2020)).

4.4 Towards ML in PdM and Inherent Challenges

The recent advancements in ML have been supported by novel algorithms and theory, as well as the continuous growth of more available and less expensive digital technologies (Çinar et al., 2020, Kumar and Galar, 2020). ML is a vital factor for analyses of massive datasets, due to its ability to generate predictions and uncover hidden information. ML is utilized in PdM for further analysis automation of railway inspection and condition monitoring data. Additionally, to reduce the bias of manual condition evaluations (Kumar and Galar, 2020). ML presents many changes to the original approach of companies and creates a significantly greater dynamic environment, which introduces new challenges (Dalzochio et al., 2020).

A major challenge with ML in PdM is to develop the foundations to implement ML models, due to the absence of a general model that applies to various industries (Dalzochio et al., 2020). Firstly, ML algorithms need input data to work, and input data requires suitable inspection methods and tools for collection (Çinar et al., 2020, Kumar and Galar, 2020). Industries use different monitoring methods and tools, resulting in a varying selection of ML algorithms. Numerous ML algorithms exist, each of which has its advantages and disadvantages

(Alsharif et al., 2020). Therefore, choosing the most suitable ML algorithm for the PdM application can be a significant challenge, as wrong selection leads to a loss in time and costs (Dalzochio et al., 2020, Çinar et al., 2020).

Another challenge comes from the data collection. Data collected through inspection procedures are difficult to acquire, organize and process using digital technologies. A significant amount of the data can consist of errors, e.g., missing labels or values, and inaccuracies. The parameters must be determined appropriately, and the data requires thorough processing and analysis. The sensors and other devices acquire massive volumes of data from real-time and continuous monitoring of railway assets. Furthermore, collected data accumulates rapidly over time and spreads across several systems. Another issue is related to the high levels of scalability and network bandwidth that are required for real-time monitoring and decision-making (Matic, 2021, Bukhsh and Stipanovic, 2020, Tretten et al., 2021, Dalzochio et al., 2020).

Data quality is also an issue that consists of several problems. One of the issues is the consequence of several differences in datasets which leads to imbalanced data. Therefore, the process of preprocessing the data is critical (Dalzochio et al., 2020, Kaur et al., 2019). Another issue is acquiring data that indicates the probability of normal state behavior to failure. This data type is critical for model training since it can be necessary to use information related to failure (Dalzochio et al., 2020). Dealing with data of a wide variety of data types and formats is an issue, which is called data heterogeneity. This can have a negative impact on the model training. Both the small and large degrees of heterogeneity can affect the predictability of the algorithms (Dalzochio et al., 2020).

A lot of time and resources are invested in establishing ML solutions. Therefore, understanding the data, which requires expert knowledge, is crucial, and it is essential to practice on various datasets to improve the ML knowledge. As a result, each application requires different nuance, data preprocessing, and modeling strategy. Additionally, the identification of required data to collect is a challenge. The company needs to have clear business goals, planning methods, and data evidence that provides value (Çinar et al., 2020, Dalzochio et al., 2020).

5 Results

The following subchapters present the results from the SLR and the SSIs.

5.1 Results from SLR

A total of 244 papers were identified through Oria (205 papers) and Scopus (39 papers). Seven articles were removed for being duplicates, and 183 papers were eliminated by viewing the titles and abstracts. The remaining 54 papers were assessed for eligibility by reviewing the full text. This led to the exclusion of 34 papers, and the author was left with 20 papers obtained through the SLR. This selection process is presented as a PRISMA flow chart in Figure 2. Afterwards, the 20 remaining papers were gathered in a spreadsheet for further analysis (2.2.3)

The analyses conducted in the spreadsheet assembled an overview of the papers in a table (see 9.3), with relating individual information distributed across columns. The table columns consists of references, PdM aim (RUL or failure prediction), ML algoritm(s), asset(s), acquisition source(s), data type(s), approach and main findings, bibliographic database, and citations.

Aim	ML Algorithm(s)	Asset(s)	Data Acquisition Source(s)	Reference
Failure Prediction	RF, RNN, k-means	Railway Track	Rail defects database and inspections database	Lopes Gerum et al. (2019)
	DT, RF, GBT	Switches	ERP[1] maintenance request process	Allah Bukhsh et al. (2019)
	RNN	Gas-insulated Switchgear [2]	Data from power equipment	Wang et al. (2020)
	SVM, RF, LDA, PCA	Railway Track	Manual inspection and inspection vehicles	Lasisi and Attoh-Okine (2018).
	AM[3], CNN, GRU	Railway track	Inspection trains	Hao et al. (2022)
	CNN	Rails and joints	Line scan cameras on recording car	Hovad et al. (2021)
	PCA, k-means	Switches	Repair records and track recording car	Vassos et al. (2021)
	RF	Railway track	Public data	Sharma et al. (2018)
	DNN, CNN, MR, SVM, GB, DT, k-means	rail, rail joint, switches, and fastener	Track geometry car	Sresakoolchai and Kaewunruen (2022)
	DNN, DT, RF	Rail	Track geometry car	Mercy and Srinivasa Rao (2018)
	PCA, k-means, Bayesian method	Switches	Switch machine sensors	Soares et al. (2021)
	ANN, SVM, naïve bayes, DT	Switches	Switch machine sensors	Arslan and Tiryaki (2020)
RUL	PCA, LgR	Railway Track	Historical track condition data and inspection vehicle	Vale and Simões (2022)
	ANN	Railway track	Vehicle and various devices	Guler (2014)
	kNN	Ballast	Multivariate sensors from Instrumented revenue vehicle	Tan et al. (2017)
	LgR, SVM	Railway track	Track geometry vehicles	Cárdenas-Gallo et al. (2017)
	Bayesian Method	Track slab[4]	Fiber bragg grating sensor	Wang and Ni (2019)
	RNN (LSTM), CNN (ResNet[5])	Switches	Sensors (accelerometer) on track	Najeh et al. (2021)
	ANN, LnR	Railway track	Database and system	Khajehei et al. (2022)
	ANN, SVR	Railway Track	ERP system	Lee et al. (2018)

Table 3: ML applications for PdM in the Railway Infrastructure

[1] Enterprise Resource Planning
[2] Electrically powered asset
[3] Attention Mechanism
[4] Track embankment of concrete or asphalt instead of ballast
[5] Residual Neural Network which is a type of CNN

Additionally, challenges experienced of ML in PdM from the papers were noted and are discussed in 6.4. Table 3 presents a compromised version of the entire table. Twelve selected papers aimed at failure prediction, while eight aimed to predict RUL. The paragraphs below briefly summarize each paper, sorted by the respective assets.

Railway tracks

Lopes Gerum et al. (2019) presented a novel framework for optimal scheduling and rail and geometry defects prediction. They used k-means for feature selection and identifying the optimal number of clusters. For the failure prediction, they applied RNN and RF based on discrete rail value data and data from inspections. In addition, they created a framework for integrating prediction with inspection and maintenance activities. The results showed that the framework effectively predicted defects and developed long-term maintenance scheduling strategies with real-time track conditions.

Lasisi and Attoh-Okine (2018) demonstrated how combined track geometry parameters with track quality indices simplified the track properties without losing variability in the data. SVM proved to be the most effective algorithm to predict track defects. In addition to SVM, they experimented with RF, LDA, and PCA.

A model for predicting track irregularities with track geometry data, vehicle body acceleration data, and vehicle speeds was presented in Hao et al. (2022). The model was built of AM, CNN, and GRU to learn sequential features and shape features, to focus on the most crucial features. This model performed better than models without AM and models purely built on GRU.

Hovad et al. (2021) created a computer vision system to predict and detect defects on rails and joints using CNN image classification. The recall rate for defects was 84%, rarely missing surface defects. The results showed great versatility of the model, which can potentially reduce the need for visual inspections.

Sharma et al. (2018) developed a failure prediction policy for track geometry defects with track geometry data. The prediction tool applied RF to predict the spot geometric defect occurrence probability. The policy produced approximately 10% savings in total maintenance costs for every 1 mile of track.

By applying DNN, RF, and RT, Mercy and Srinivasa Rao (2018) presented a method for predicting rail surface defects using rail geometry data. DT had the highest accuracy for predicting twists and unevenness of right and left rails. RF and DT had equal accuracy when predicting alignment left. RF has the best prediction accuracy for alignment right and gauge. DNN underperformed compared to DT and RF. The presented method did not add any additional development costs.

Vale and Simões (2022) estimated the RUL of track based on geometry data, using PCA and LgR. They utilized predictors of longitudinal level and alignment of the left and right rail, the time interval between inspection actions, and the sequential number of track segments. Their easily implemented model resulted in a 91.1% success rate.

Guler (2014) modeled the track deterioration with ANN by data related to the track structure, layout, geometry, traffic characteristics, environmental factors, maintenance, and renewal. They argue that ANN is one of the best ways to predict track geometry degradation linked with considerable inherent complexity, and their model is designed without expert knowledge.

Cárdenas-Gallo et al. (2017) created a method for developing an ensemble classifier using geometry data to predict the evolution of track geometry defects. The ensemble classifier consisted of LgR and SVM algorithms and outperformed the individual algorithms in every case. Their approach can be implemented to other tracks and other assets.

Khajehei et al. (2022) presented a model for predicting track degradation rate by collecting track geometry data, asset information, and maintenance history. They applied LnR and ANN, which resulted in adequate performance. Historical maintenance data was the most critical contributor to the prediction, and the number of tamping activities highly influences the degradation rate due to its destructiveness.

Lee et al. (2018) applied SVR and ANN to predict track deterioration with data from track properties, tamping ratios, measurement data, and track quality indices. They performed two case studies with different input parameters. ANN performed marginally better than SVR due to the retraining process. The results stated that at least two years of maintenance data are needed to develop a stable prediction. Furthermore, they can be used to support the maintenance decision-making process.

Sresakoolchai and Kaewunruen (2022) created a model for predicting several asset defects and tested DNN, CNN, MR, SVM, GB, and DT. The assets were rails, joints, switches, and fasteners. K-means investigated the insights of track defects. A track geometry car collected geometry data, asset defect data, and track profile data. DNN performed with the best accuracy of 94.3%, followed by CNN with 93.8%. Other algorithms had approximately a prediction accuracy of 50% or lower. The algorithm detected defects and categorized them into rails, joints, switches, and fasteners.

Railway Switches

Allah Bukhsh et al. (2019) utilized DT, RF, and GBT for failure prediction of railway switches with historical data from visual inspections, maintenance records, and condition data. GBT was the best performer for predicting maintenance, while RF was the most accurate algorithm for the prediction of activity type and trigger status. They state that the approach is applicable to or discrete types of infrastructure assets.

Vassos et al. (2021) split data into two clusters (from historical repair data and switch geometry data) to predict the maintenance need of switches, supported by k-means and PCA. However, no clear boundary existed between the clusters, and it was difficult to assign a specific cluster label successfully.

Soares et al. (2021) created a PdM model to predict failures of switch machines. They used k-means, PCA, and a gaussian method on data extracted from switch machine sensors. The results of the approach were impressive, and its success depended on the attention to necessary procedures conducted of the ones classified as faults.

Arslan and Tiryaki (2020) predicted switch failures by utilizing SVM and ANN. They studied data consisting of switch movement, movement time, position, and status. The algorithms outperformed both DT and Naïve Bayes. However, ANN was more appropriate in the implementation face. If inputs were to change, the presented method would correctly determine the outputs and predict failures.

Najeh et al. (2021) developed an RNN and CNN framework to design an effective prediction system for switch deterioration using vibration data. Their system had an acceptable accuracy of wear estimation in the middle section of switches.

Other railway assets

Wang and Ni (2019) utilized real-time monitoring data to develop an online system based on a Bayesian algorithm to predict track slab deformation. The experimental results show improvements in terms of regression compared to the current algorithm. The method improves the prediction accuracy of all sensors for track slab monitoring.

Tan et al. (2017) presented a kNN method for predicting ballast tamping effectiveness, achieving 68% accuracy. Sensors implemented on the vehicle collected data to predict effectiveness 12 weeks prior to tamping.

Maintenance prediction of gas-insulated switchgear (electrical asset) was attempted by Wang et al. (2020). They developed a maintenance predictor powered by LSTM-RNN on historical data. The results were positive and proved the possibility of predicting future maintenance activities.

5.1.1 Publication Year and Citation Analysis

Figure 20 illustrates the papers sorted by publication year, with a descriptive trend line. Papers relating to the topic have been published from 2014 until this year, with a gap between 2014 and 2017. 2014 had the lowest number of published papers, and 2018, 2021, and 2022 had the highest amount, with one and four publications, respectively. There were four publications in 2022, before the 20th of March when the search was conducted (2.2.2). The average number of publications per year from 2014 to 2017 is 0.75, and 3.4 publications per year from 2018 to 2022.

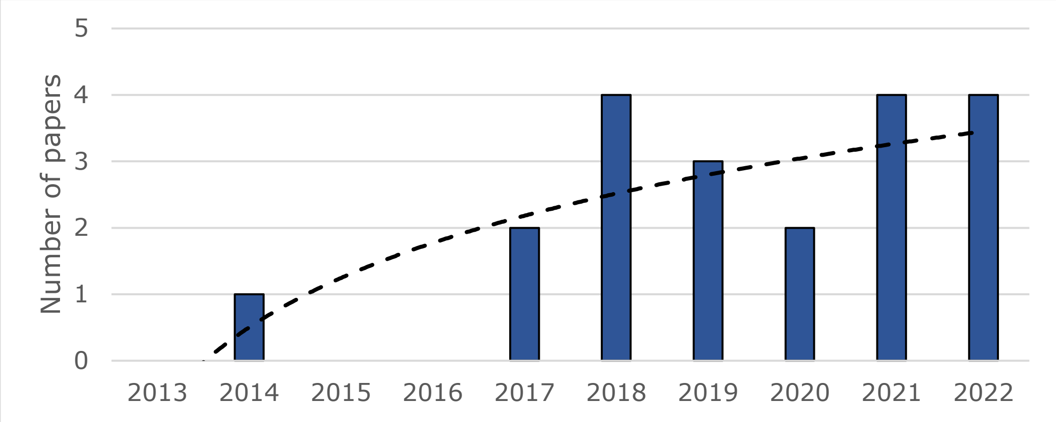

Figure 20: Distribution of publication years of the selected papers

Table 4 shows the top 10 most cited papers, and the average number of citations of the 20 papers is 24.15. Guler (2014) and Sharma et al. (2018) have the highest number of citations, with respectively 89 and 80 citations. Guler (2014) studied the RUL of railway tracks using a predictive model based on ANN, while Sharma et al. (2018) developed a failure prediction policy based on RF to predict track geometric defects.

Reference	Title	Citations
Guler (2014)	Prediction of railway track geometry deterioration using artificial neural networks: a case study for Turkish state railways	89
Sharma et al. (2018)	Data-driven optimization of railway maintenance for track geometry	80
Lasisi and Attoh-Okine (2018)	Principal components analysis and track quality index: A machine learning approach	73
Allah Bukhsh et al. (2019)	Predictive Maintenance using tree-based Classification techniques: A Case of Railway Switches	58
Lopes Gerum et al. (2019)	Data-driven Predictive Maintenance Scheduling Policies for Railways	45
Cárdenas-Gallo et al. (2017)	An ensemble classifier to predict track geometry degradation	43
Wang et al. (2020)	Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment With LSTM-RNN	29
Lee et al. (2018)	Prediction of Track Deterioration Using Maintenance Data and Machine Learning Schemes	19
Wang and Ni (2019)	Measurement and Forecasting of High-Speed Rail Track Slab Deformation under Uncertain SHM Data Using Variational Heteroscedastic Gaussian Process	17
Arslan and Tiryaki (2020)	Prediction of railway switch point failures by artificial intelligence methods	7

Table 4: Citation analysis

5.1.2 Distribution of ML Algorithms and Assets

The distribution of papers per ML algorithm applied to PdM of railway infrastructure assets is illustrated in Figure 21. RF is the most utilized ML algorithm, with five papers, while the second placeholders consist of SVM, CNN, PCA, ANN, DT, k-means, and RNN. PL, which includes Naïve Bayes and Gaussian methods, are ranked third. DNN, GB, and LgR were applied in two papers each, and LDA, kNN, MR, LnR, and SVR are ranked last with one paper each.

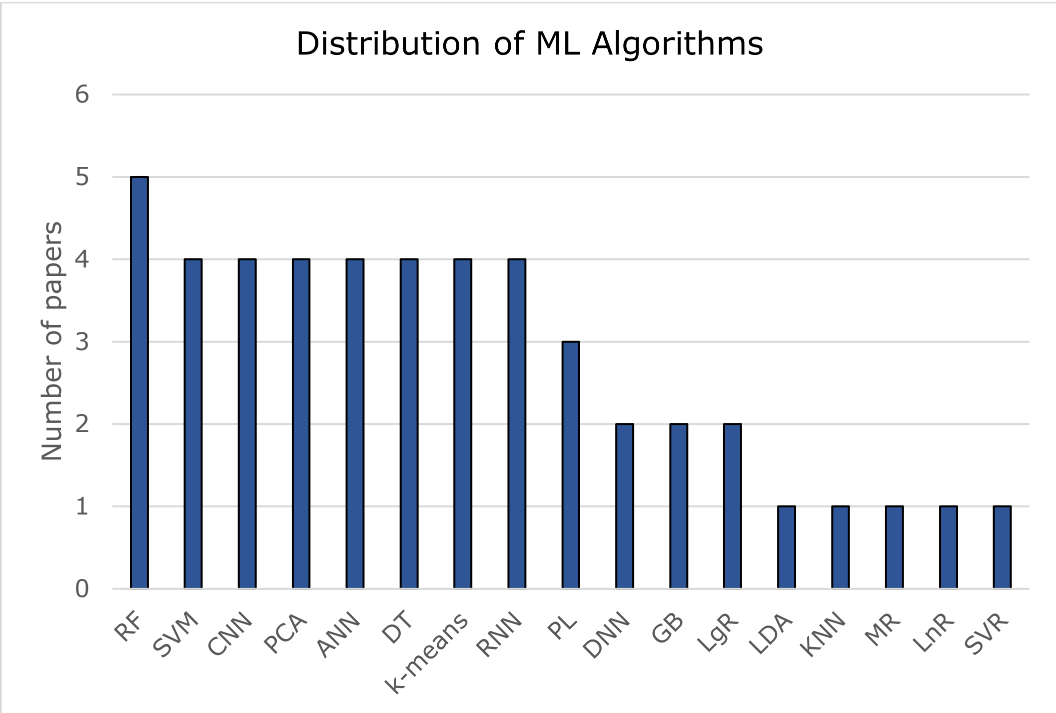

Figure 21: Number of papers per ML algorithm

More than half of the papers have applied ML algorithms to railway tracks, as illustrated in Figure 22. This group of assets predicts faults to rails, sleepers, joints, fasteners, and track geometry. Slightly more than a quarter of the studies implemented ML algorithms to switches. The "Others" group consists of algorithms applied to the track embankment (ballast or slab) and electrical equipment, resulting in a minority of research.

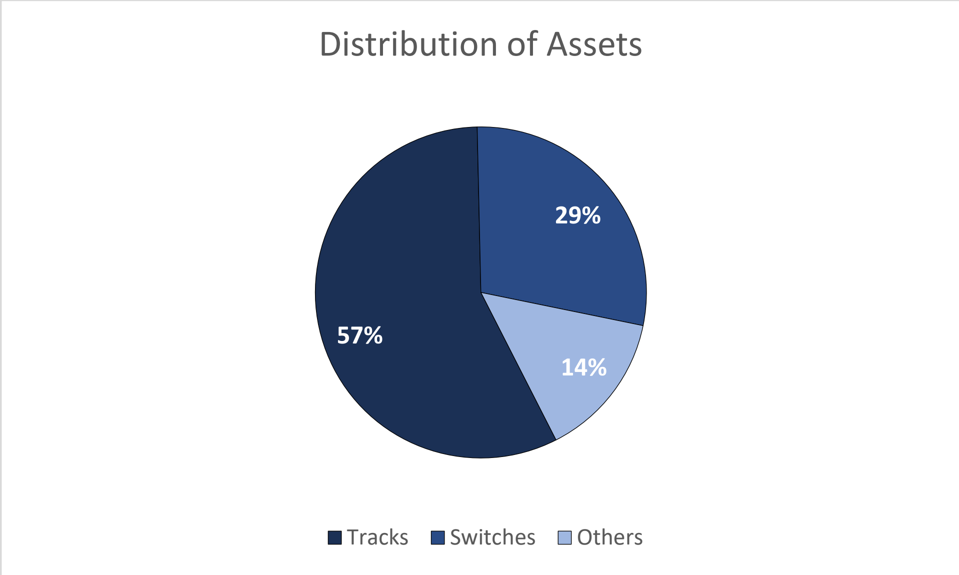

Figure 22: Distribution of assets subjected in the papers

Figure 23 is an extension of Figure 21, showing the allocation of assets to the ML algorithms. There is a slight difference between the figures. The reason is that one paper applied ML to both tracks and switches. Many algorithms were applied to both tracks and switches, except PL, LgR, LnR, LDA, SVR, and kNN. PL was used for switches and other assets. LgR, LnR, LDA, and SVR were purely related to tracks. On the contrary, kNN was associated with a single asset, which was ballast in this case. RNN is the only algorithm applied for every asset group out of all the algorithms.

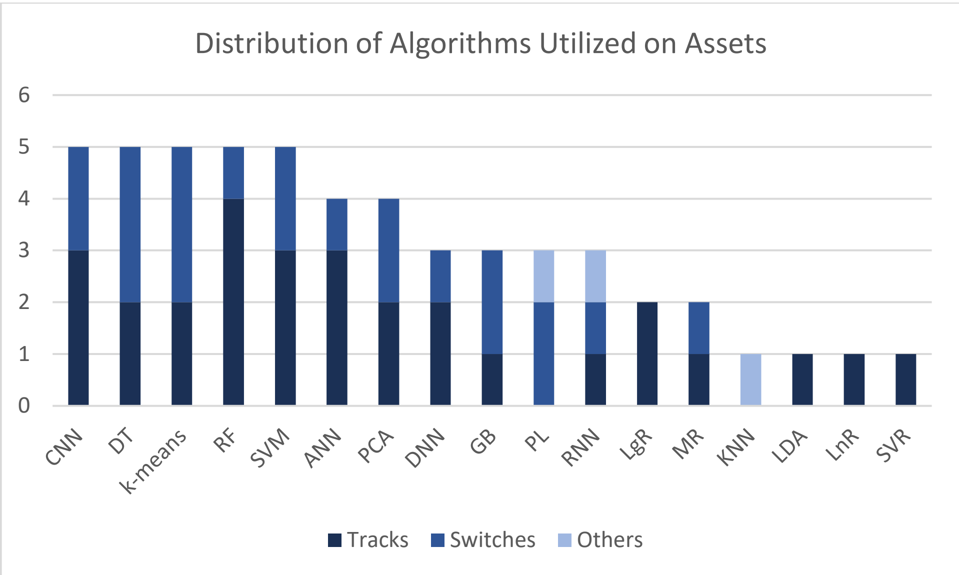

Figure 23: Distribution of ML Algorithms on Assets

5.1.3 Data Sources, Data types, and Asset Dimensions

The data acquisition sources for each paper are presented in Table 3. The sources were from vehicle or sensor inspections, historical data, or a particular maintenance system or database. Certain papers acquired data from vehicle inspections, and other research acquired data from sensors on the track. No studies used purely manual inspection methods. However, Lasisi and Attoh-Okine (2018) gathered data from inspection vehicles together with manual inspections.

Table 5 presents the types of data used in datasets for ML algorithms, with the referring number of papers. The numbers of papers are higher due to some papers applying multiple data types in their dataset. In addition, Wang et al. (2020) and Wang and Ni (2019) used artificial data and were excluded from this analysis. Geometry data, asset properties, and historical data were utilized in datasets by, in the subsequent order, 67%, 39%, and 33% of the papers. Geometric values are presented in chapter 3.1.1. Historical data consists of data related to maintenance, inspections, defects, and condition states. Fewer articles considered the remaining data types. Hovad et al. (2021) was the only paper that applied images to their algorithm. Furthermore, Guler (2014) was the single work that considered traffic characteristics and environmental factors.

Data types	Number of papers
Geometry data	12
Asset properties	7
Historical data	6
Acceleration/Vibration data	2
Tamping data	2
Switch machine data	2
Images	1
Traffic characteristics	1
Environmental factors	1

Table 5: Distribution of papers of data types used in datasets applied to ML algorithms

The utilized dimensions related to ML in PdM of switches consisted of several parameters from many switches. Arslan and Tiryaki (2020), Soares et al. (2021), Najeh et al. (2021) collected data from six switches. Allah Bukhsh et al. (2019) collected data from 802 switches, while Vassos et al. (2021) gathered data from 166 switches. Sresakoolchai and Kaewunruen (2022) acquired both track and switch defect data from a 30 km long track without specifying the number of switches.

In dimensions for ML in PdM of railway tracks, the length of the measured tracks varied, as Table 6 depicts. Hao et al. (2022), Khajehei et al. (2022), and Guler (2014) applied tracks more than 150 km in their cases. On the contrary, Hovad et al. (2021) and Lasisi and Attoh-Okine (2018) measured significantly shorter tracks. Cárdenas-Gallo et al. (2017) studied four track segments for predicting RUL of tracks, but did not state a total track length. Similarly, Lee et al. (2018) and Mercy and Srinivasa Rao (2018) did not state the total length of the measured track.

Reference	Track length (km)
Hao et al. (2022)	300
Khajehei et al. (2022)	216
Guler (2014)	180
Sharma et al. (2018)	80
Vale and Simões (2022)	51
Sresakoolchai and Kaewunruen (2022)	30
Lopes Gerum et al. (2019)	17
Hovad et al. (2021)	5,4
Lasisi and Attoh-Okine (2018)	1,6

Table 6: Track lengths used in datasets by each study for PdM of railway tracks

Wang et al. (2020) was the only paper related to this topic that applied ML in PdM of electrical equipment. Their case study used deterioration and failure data to build a sample generator, resulting in a dataset consisting of 10^5 samples with 20 observations in each sample.

Wang and Ni (2019) and Tan et al. (2017) presented models for predicting RUL embankments (slab and ballast tracks). Wang and Ni (2019) collected data using five concrete blocks to represent slab tracks and field test data. Tan et al. (2017) performed their data collection with a sequence of 50 m blocks of track.

5.2 Results from Interviews

This chapter presents the results from the two conducted interviews.

5.2.1 Interview 1 – Veronica Brizzi

The first participant was Veronica Brizzi, who is currently the Team leader and Data Scientist in the research and development team of MIPU – Predictive Hub. Veronica has a long industry experience and specializes in ML, energy, and PdM. MIPU is an Italian company developing ML and PdM solutions to improve industrial operations. One of their solutions to ML in PdM, is a CNN image classification model. The model analyzes images acquired by inspection vehicles on the track and categorizes them into 21 assets with an accuracy of nearly 90%. Additionally, with unsupervised learning and classification models based on current and voltage data, they can accurately predict an electromechanical failure of switches 2-3 months in advance.

Veronica mentioned several challenges related to ML-based PdM of railway infrastructure assets, including data collection and quality. She said that it is essential to understand the data and define the needs with respect to the technology. Furthermore, she said that the data collection captures an extreme amount of data, commonly unbalanced data. Thus, involving people with data knowledge skills and technical infrastructure competence is crucial. She further explained that assets themselves are not very digitalized and therefore, can lack support to ML. In addition, she mentioned the existence of technology issues from edge computing to the cloud regarding the amount of data across many systems. She further mentioned that certain asset shapes are more difficult to classify than others in terms of image classification. When selecting algorithms, they sometimes train three potential algorithms before deciding. If the results are not sufficient, they do more research.

5.2.2 Interview 2 – Kristine Tveit

The next participant was Kristine Tveit from Bane NOR. Bane NOR is responsible for the railway infrastructure in Norway. She is the Smart Maintenance Team manager and has a long experience with railway infrastructure maintenance. The Smart Maintenance team aims to predict faults before they occur in the infrastructure to avoid train disruptions.

When asked about the current situation of ML in PdM in Bane NOR, she responded that, as of now, the maintenance is not very predictive, but more detection based. They use different monitoring equipment to receive alarms of incipient faults and anomalies of various assets, aiming to avoid breakdowns. However, she mentioned they will now begin an ML task to predict switch faults. This will be inserted in a criticality score, simplifying the prioritization of switch maintenance, and received alarms. Furthermore, she expressed that they have attempted implementing ML to track circuits. In their cases, there were issues regarding the sensors picking up too much noise, and instead of detecting faults, the algorithms detected track circuits working correctly.

Regarding challenges related to ML, she emphasized that data quality is a concern. It is sometimes difficult to acquire documentation on what the actual errors are. With railway switches, the data must be compared to another identical switch machine, hence fewer observations. Additionally, there are issues related to the training of the ML algorithms due to a lack of data. She expressed that, apart from the many advantages of ML, some assets are very troublesome to predict with the current sensors, for instance, cable breakages. They have attempted prediction with other sensors gathering vibration data, but without success in this case.

6 Discussion

This chapter discusses the findings from the results chapter. The first chapter discusses the growth of ML in PdM. The second chapter evaluates how methods for inspection and data should be selected, followed by a chapter on selecting ML algorithms. The next chapter discusses the challenges related to ML in PdM, and the last chapter elaborates on the weaknesses of this study.

6.1 Growing Technology

Even though research of ML in PdM is growing, this SLR resulted in only 20 papers exclusively connected to the railway sector. This proves that the research on this topic is somewhat limited and suggests that challenges are delaying the development. The distribution of publication year for each paper from the SLR validates that the research on ML in PdM of railway infrastructure assets is a relatively new technology, since no papers were published before 2014. Additionally, the trend line shows that the research is increasing in popularity. Before 2017, only one paper was published, and the average number of papers has increased since. Another contributor to this argument is that in 2022, as of March, already four papers have been published, which is likely to increase by the end of the year. This increase in publications can be linked to monitoring equipment becoming more available and less expensive, and the improvement of ML algorithms the recent years, as expressed in 3.2 and 4.4.

The interviews provide further support to this argumentation. The first interview described a few cases where they had implemented ML, but technological issues still exist. Current railway infrastructures lack digitalized assets to support ML, and the edge to the cloud processes and systems struggle with the enormous amounts of data spread across systems. These challenges were also found in chapter 4.4. From the second interview, it was clearer that ML in PdM of the railway infrastructure is still developing and will be a more common approach in the future. Although Bane NOR will soon develop an ML method for predicting faults to switches, their current maintenance is more focused on anomaly detection, rather than predicting anomalies.

From what is discussed above, it is definite that the transition to purely ML in PdM of railways infrastructures is being researched and developed. Nevertheless, this transition may take time. Railway agencies need to be willing to invest in ML, both financially and organizationally, since it requires time and provides several changes to the organization and maintenance procedures. In addition, they need to consider the challenges (presented in chapter 4.4), which are further discussed in chapter 6.4.

6.2 Selecting Methods for Inspection and Data

The results showed that vehicles and sensors installed in the infrastructure were the most common inspection methods. Visual inspection was used by only one paper, implying that it may not be suitable as a data source for ML, which is linked to the theory found in chapter 3.2.1. The results highlight that ML in PdM of railway infrastructures should focus on the inspection methods used in the SLR papers. However, as learned from the second SSI, inspection methods for other assets are lacking and should be developed.

Regarding data types, geometry, asset properties, and historical data were most utilized and implied strong argumentations for utilizing these three in ML training datasets. Yet, the geometry data applies to tracks and switches, and no other assets. Adding other data types should be considered since the ML algorithms could perform better with many datasets. For instance, Lee et al. (2018) used tamping data in addition to geometry and asset properties data, to investigate how this data impacted the accuracy of the ML algorithm. However, the degree of heterogeneity can affect the predictability of the algorithms and needs to be assessed.

Asset dimensions consisted of various sizes for tracks and switches, which might imply that there are discussions between the researchers on the correct asset dimensions. Proper asset dimensions must be selected for a sufficient ML algorithm. The results presented in chapter 5.1.3 can assist the selection of other railway organizations and other studies, as it is important to find the optimal balance between the dimension and the amount of data to achieve accurate performance. Limited papers were found on other assets. Hence, few asset dimensions for other assets were determined.

6.3 Selecting the Suitable ML Algorithm

Several works achieved cost savings and good performance accuracy with their respective ML applications. However, selecting the most suitable algorithm for a railway asset is complicated and often involves a training process of several algorithms, as evidenced in chapters 5.1 and 5.2.1. In addition to the purpose of the PdM (RUL or failure prediction), the desired ML type and tasks must be selected with respect to inspection methods, tools, and the railway asset. The theory from the LR and the results from the SLR verifies that several ML algorithms can be utilized in PdM of railway infrastructures. The distribution of SLR papers per algorithm is shown in Figure 21. Corresponding to the findings from Arena et al. (2022) and Xie et al. (2020), supervised, unsupervised, ensemble, and DL methods are most relevant for PdM practices in the railway infrastructure, suggesting that they should be evaluated in the selection process. Although this research covered the most relevant algorithms, there might be other algorithms that are applicable. Contrary to Ouadah et al. (2022), this study found only one paper that utilized the kNN algorithm. Speculations on the reasons might be that kNN has not achieved many successful implementations in the railway industry.

In addition to selecting the algorithm, the process must take the railway asset into consideration, which is connected to developing the foundations of a ML model (4.4). As previously presented in Figure 22, railway tracks and switches are most researched, and works on other assets are few. While this was unanticipated, the results from the SLR and SSIs imply that it is due to the current railway monitoring methods. The railway track and switches have been equipped with more tools for inspection, in contrast to, e.g., catenary masts, illumination devices, drainage elements, and fences, amongst others. Another reason can be that tracks and switches are more critical for the railway infrastructure since faults in these assets, cause disruptions that affect the whole infrastructure and negatively influence customer service. Allah Bukhsh et al. (2019), Cárdenas-Gallo et al. (2017) and Arslan and Tiryaki (2020) argue that their ML method can be applied to other assets. However, not enough research exists to provide solid suggestions with assisting evidence.

Table 3 presented the papers sorted by the aim of PdM and can be used as guidelines for the evaluation of algorithms in the selection process. The following subchapter further discusses the selection of ML algorithms for PdM of railway tracks and switches.

6.3.1 Selecting ML Algorithms for Tracks and Switches

In terms of failure prediction of railway tracks, seven papers were identified that utilized several ML algorithms. Sharma et al. (2018) used RF, and Lopes Gerum et al. (2019) applied RNN and RF, while other works argued against RF. For instance, Lasisi and Attoh-Okine (2018) found that SVM was more effective than RF, and Mercy and Srinivasa Rao (2018) proved that DT, in some cases, performed better than RF. The other papers did not train RF for failure prediction of tracks. Hovad et al. (2021) performed image classification tasks with CNN, whereas Hao et al. (2022) applied an ensemble model with CNN, and argued that the model was better than the CNN alone. Sresakoolchai and Kaewunruen (2022) also trained CNN, amongst many other algorithms for predicting track defects. Their findings show that DNN had a slightly better performance accuracy than CNN, and performed much better than DT and RF. These argumentations might imply that there is an uncertainty of algorithm(s) decisions and enhances the importance of training several algorithms before deciding for one or more algorithms. However, the abovementioned algorithms are a good starting point and correlate to previous research from Ouadah et al. (2022), except for kNN. The kNN algorithm is not considered by any of the papers, which might indicate that it is not suitable for failure prediction of tracks. For image classification tasks, CNN should be examined, corresponding with information from the first interview.

Regarding RUL of railway tracks, five papers were obtained from the SLR. Contrary to failure prediction of tracks, a much lower number of algorithms were trained. Guler (2014) applied ANN for predicting track deterioration, while Khajehei et al. (2022) used LnR in addition to ANN. Lee et al. (2018) found that ANN performed marginally better than SVR. In contrast to the other authors, Vale and Simões (2022) proposed PCA and LgR, and Cárdenas-Gallo et al. (2017) applied SVM and LgR for estimating RUL based on track geometry data. From these papers, argumentations for recommending ANN and LgR in RUL of railway tracks can be made, but the use of PCA, SVM, and LnR, should also be investigated.

A total of six research papers were found for ML in PdM of railway switches, and five papers were related to failure prediction. Najeh et al. (2021) was the only paper concerning RUL of switches and presented a solution consisting of RNN and CNN. Soares et al. (2021) and Vassos et al. (2021) applied the unsupervised learning algorithms, PCA and k-means, for failure prediction of switches. Arslan and Tiryaki (2020) found ANN most effective and Allah Bukhsh et al. (2019) selected GBT and RF as best performers. As mentioned above, Sresakoolchai and Kaewunruen (2022), who predicted failures of switches in addition to rails, found DNN and CNN as the best algorithms. Due to the various selection of algorithms from these papers, it is difficult to establish the most suitable algorithm. However, DL methods can be viewed as a priority for estimating RUL of switches, which is in line with some of the papers on RUL of tracks. For predicting failures of switches, the abovementioned algorithms should be evaluated.

6.4 Challenges with ML in PdM

From what is established in chapters 3.2 and 4.4, ML-related challenges in PdM can be divided into four main categories: foundational, data collection, data quality, and data knowledge.

The foundational challenges are due to the selection of suitable algorithms and the lack of an existing universal ML model for all industries, corresponding to railway infrastructure assets. The inexistence of a ML model that applies to all assets is a main challenge, and chapters 6.3 and 6.3.1 provides further evidence to this statement. Compared to other assets, several ML methods were found for tracks and switches. Therefore, ML in PdM of other assets is another issue. The algorithms presented can be utilized for particular assets with a specific goal, but do not cover all of them. Arslan and Tiryaki (2020) supported this argument by expressing difficulties regarding algorithm selection. Cárdenas-Gallo et al. (2017) stated that their method could be implemented on other railway tracks and other types of asset defects. To the knowledge of this study, this remark remains unsettled because no other research has pursued it. On the contrary, Hao et al. (2022) and Allah Bukhsh et al. (2019) mentioned that data and algorithms could not be settled forever. Certain assets will gradually change over time and require new data and retraining of the ML model. Another challenge is the foundations for data collection. Wang et al. (2020) experienced issues with data deficiency. This correlates to the interview findings. It is crucial to establish the needs of the implementation and assets lack data and support to ML since they are currently not digitalized enough to the desired degree.

Data collection and quality consist of acquisition, processing, and organization issues. Inspection vehicles and sensors installed in the infrastructure are the main data collection methods for ML in PdM of railway infrastructure assets. Chapter 5.1.3 illustrates that for ML algorithms to work, the types and dimensions of the datasets are decisive. On the opposite, it is not clear what the “perfect” asset dimensions are due to the various sizes, and data is often established from several years of inspection. This depends on the desired size or number of assets combined with the amount of data the ML algorithms need to work properly, concerning the heterogeneity issues. According to Lasisi and Attoh-Okine (2018), modern cloud systems have reduced the problem of data storage, but the enormous amounts of data collected still cause issues. The edge-to-cloud systems face data extended to several different platforms. Lee et al. (2018) reported issues concerning the big datasets used for training the algorithms. In addition, Guler (2014) had problems with datasets having large numbers of influencing parameters, and Soares et al. (2021) expressed the issues with selecting the proper parameters. Furthermore, inspection methods and tools gather data that contain errors, and in some cases lead to a reevaluation of the data features and retraining of algorithms, as experienced by Khajehei et al. (2022) and Soares et al. (2021). Accurate ML algorithms require detailed data preprocessing, which consumes time and resources for data cleaning and structuring.

Regarding data quality, imbalanced data is a primary concern and was reported as an issue by both the interviews and Sresakoolchai and Kaewunruen (2022), Lasisi and Attoh-Okine (2018), and Allah Bukhsh et al. (2019). There are different causes for imbalanced data. Lopes Gerum et al. (2019) issued imbalanced data due to dealing with several rail segments, Hovad et al. (2021) image classification algorithm detected multiple and irrelevant objects, and Vale and Simões (2022) had dissimilar measurement data in different time intervals. To deal with imbalanced data issues, feature extraction and dimensionality reduction algorithms such as k-means and PCA, can be implemented to potentially improve the ML method applied in PdM of the railway infrastructure.

The last challenge is related to the knowledge required for implementing ML algorithms. Both the interviews and Arslan and Tiryaki (2020) stated that it is essential to understand the collected data, and this demands expert knowledge. For a successful ML implementation to railway assets, great expertise in ML algorithms must be combined with insights into the assets. Expert knowledge is required for dealing with the priorly stated implementation challenges due to the application complexity of algorithms, which is expressed by Vale and Simões (2022)

These challenges must be assessed before establishing a ML solution, as an unknown issue can potentially cause consequences. Additionally, more solutions to the challenges should be established, preventing companies from possibly losing motivation to implement ML methods.

6.5 Weaknesses

The study has limitations related to the research design, which are presented in chapter 2.4. In addition, the number of interviews is a weakness of this study, as more interviews could have provided additional information and new perspectives on the problem. Another weakness is the number of papers gathered in the SLR. They were dominated by two asset types, tracks and switches. Few papers on other assets were identified, making it difficult to perform proper analysis to provide solid solutions.

Another weakness is that the data analysis is from a general view. A thorough study of the parameters and values of the collected datasets allows for more solid argumentation on what data to utilize for the training and testing of ML algorithms. Additionally, a study that includes ML implementation for assets outside the infrastructure or fault and anomaly detection purposes, can provide more information about the current status of ML implementations in the railway industry. Regarding the presented ML methods in the theory, it could be argued that this research did not study the disadvantages of each method, which could have improved the recommendations for algorithm selections.

7 Conclusion

This chapter concludes the project by addressing the four RQs, followed by the contribution of the thesis and suggestions to further research.

One of the primary purposes of this study was to uncover the progress towards ML-PdM of railway infrastructure assets. The main conclusion is that this maintenance strategy is in its early stages but is progressively developing. The research on the topic was limited, as discovered from the SLR. A total of 20 papers were identified, and the publication years show an increasing trend in popularity, since half of the papers were published in the last few years. This indicates that ML-PdM applications in the railway industry are a relatively new approach, which is also acknowledged from the interviews. This is due to several reasons, and they are connected to the challenges of ML and the lack of inspection methods and tools for certain infrastructure assets. The following paragraphs briefly present the findings concerning the RQs.

RQ1: Which ML algorithms are applied to PdM of railway infrastructure assets?

The results from the SLR and the first interview show that several algorithms can be applied. These are supervised, unsupervised, ensemble, and DL methods. A total of 17 algorithms were identified in the SLR, including algorithms that perform regression, classification, and clustering tasks. These findings contribute to railway organizations and other researchers developing a ML method for PdM of railway infrastructure assets. The author attempted to select a suitable ML algorithm for railway tracks and switches, where a few guidelines were established. However, due to the complexity of the connection between data collection, assets, tasks, and algorithms, the case of determining the "best" algorithm is troublesome.

RQ2: Which railway infrastructure assets are subjected to ML algorithms in PdM?

This research found that the assets subjected to ML algorithms related to PdM were mainly railway tracks and switches, with 57% and 29%, respectively. The remaining 14% were related to three papers regarding railway embankment (slab track and ballast) and gas-insulated switchgear (electrical asset). The reason for the low research on other assets is discovered to be less digitalized inspection methods and tools. In addition, tracks and switches are more critical to the infrastructure, and since ML in PdM is an emerging technology, the research focus might be more directed to these.

RQ3: What inspection methods and tools are utilized for data acquisition for ML in PdM of railway infrastructure assets?

The inspection methods and tools for collecting data to support ML-based PdM of railway infrastructures are inspection vehicles and sensors installed along the track. There are various vehicles and sensors used for data acquisition. Geometry data, historical data, and asset properties data are the types of data usually applied to the ML algorithm. All these consist of several parameters that are gathered from various asset dimensions. For proper algorithm training, there must be a balance between the data amount and asset dimensions, which underlines the complexity involved and the importance of establishing the correct foundations for implementing ML in PdM.

RQ4: What are the main challenges of ML in PdM of railway infrastructure assets?

The results from the LR, SLR, and the SSIs characterized the challenges of railway infrastructure ML in PdM into four main categories. The challenges are related to, and the categories are related to foundations, data collection, data quality, and data knowledge. Each challenge consisted of several issues, and a few suggestions were expressed to mitigate them. These challenges might prevent railway infrastructure organizations from being willing to apply ML in their PdM practices, since the technology is possibly viewed in the light of the challenges and risks, rather than the advantages.

7.1 Contribution

The results of this thesis are valuable to railway organizations and other researchers developing a ML method for PdM of railway infrastructure assets. The thesis provides comprehensive insights into ML methods, explaining which algorithms are suitable for PdM of railway infrastructure assets. In addition, it detailedly describes the utilized data sources, data types, and asset dimensions.

Furthermore, challenges related to the topic have been presented, which justifies the focus areas that organizations and researchers need to consider to achieve a successful implementation.

An important note is that some relevant papers may have been unidentified in the SLR. The search string was broad, but more papers could potentially have been detected in other databases, languages, or with other search strings. These could have impacted the results of the thesis. In addition, the quality of the papers may vary. The selected papers were peer-reviewed and went through citation analysis, hoping to ensure quality. However, one can not be certain that the selected papers are unbiased.

7.2 Further Research

Since ML in PdM in railway infrastructures is a relatively new topic, several areas are available for further research, and the prior paragraphs of this chapter provide suggestions. These are described in the paragraphs below.

The first suggestion is to establish solid argumentations for the most suitable ML algorithm for a railway infrastructure asset. To the author's knowledge, not enough research has been conducted. The research could also focus on creating a universal model for several railway infrastructure assets.

Another research can proceed on findings from this study, and investigate the disadvantages of the ML methods, to draw additional comparisons between algorithms, data types, asset dimensions, and the proposed methods from the papers.

The next suggestion is to study the different data collection methods, parameters, and values in detail to analyze and establish the optimal data foundations for one or more specific railway infrastructure assets.

The last suggestion is to investigate the main challenges described in this thesis, and develop more solutions to mitigate them. This could be very beneficial for the success of ML implementations in PdM, and could greatly support railway organizations and other researchers.

8 References

ADAMS, W. C. 2015. Conducting semi-structured interviews. *Handbook of practical program evaluation,* 4**,** 492-505.

ALLAH BUKHSH, Z., SAEED, A., STIPANOVIC, I. & DOREE, A. G. 2019. Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation research. Part C, Emerging technologies,* 101**,** 35-54.

ALPAYDIN, E. 2016. Machine learning : the new AI. Cambridge, Massachusetts: The MIT Press.

ALSHARIF, M. H., KELECHI, A. H., YAHYA, K. & CHAUDHRY, S. A. 2020. Machine learning algorithms for smart data analysis in internet of things environment: Taxonomies and research trends. *Symmetry (Basel),* 12**,** 88.

ALZUBI, J., NAYYAR, A. & KUMAR, A. Machine Learning from Theory to Algorithms: An Overview. Journal of Physics: Conference Series, 2018.

ARENA, S., FLORIAN, E., ZENNARO, I., ORRÙ, P. F. & SGARBOSSA, F. 2022. A novel decision support system for managing predictive maintenance strategies based on machine learning approaches. *Safety science,* 146**,** 105529.

ARSLAN, B. & TIRYAKI, H. 2020. Prediction of railway switch point failures by artificial intelligence methods. *Turkish Journal of Electrical Engineering and Computer Sciences,* 28**,** 1044-1058.

BALOGLU, O., LATIFI, S. Q. & NAZHA, A. 2021. What is machine learning? *Arch Dis Child Educ Pract Ed***,** edpract-2020-319415.

BANENOR. 2020. *Teknisk Regelverk* [Online]. Norway: BaneNOR. Available: https://trv.banenor.no/wiki/Forside [Accessed 03.10. 2021].

BI, Q., GOODMAN, K. E., KAMINSKY, J. & LESSLER, J. 2019. What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology,* 188**,** 2222-2239.

BUKHSH, Z. A. & STIPANOVIC, I. 2020. Predictive Maintenance for Infrastructure Asset Management. *IT professional,* 22**,** 40-45.

BURKOV, A. 2019. *The Hundred-Page Machine Learning Book*, Kindle Direct.

CÁRDENAS-GALLO, I., SARMIENTO, C. A., MORALES, G. A., BOLIVAR, M. A. & AKHAVAN-TABATABAEI, R. 2017. An ensemble classifier to predict track geometry degradation. *Reliability engineering & system safety,* 161**,** 53-60.

CARVALHO, T. P., SOARES, F. A. A. M. N., VITA, R., FRANCISCO, R. D. P., BASTO, J. P. & ALCALÁ, S. G. S. 2019. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & industrial engineering,* 137**,** 106024.

ÇINAR, Z. M., NUHU, A. A., ZEESHAN, Q., KORHAN, O., ASMAEL, M. & SAFAEI, B. 2020. Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability (Basel, Switzerland),* 12**,** 8211.

CONNOR, P. 2019. *Infrastructure* [Online]. UK: The Railway Technical Website. Available: http://www.railway-technical.com/infrastructure/ [Accessed 04.10 2021].

DALZOCHIO, J., KUNST, R., PIGNATON, E., BINOTTO, A., SANYAL, S., FAVILLA, J. & BARBOSA, J. 2020. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in industry,* 123**,** 103298.

DAVARI, N., VELOSO, B., COSTA, G. D. A., PEREIRA, P. M., RIBEIRO, R. P. & GAMA, J. 2021. A survey on data-driven predictive maintenance for the railway industry. *Sensors (Basel, Switzerland),* 21**,** 5739.

DINDAR, S. & KAEWUNRUEN, S. 2017. Assessment of Turnout-Related Derailments by Various Causes. Cham: Cham: Springer International Publishing.

EL NAQA, I. & MURPHY, M. J. 2015. What is machine learning? *machine learning in radiation oncology.* Springer.

ESPLING, U. & KUMAR, U. 2007. Benchmarking of the Maintenance Process at Banverket (The Swedish National Rail Administration). London: London: Springer London.

GIVEN, L. 2008. The SAGE Encyclopedia of Qualitative Research Methods.

GULER, H. 2014. Prediction of railway track geometry deterioration using artificial neural networks: a case study for Turkish state railways. *Structure and infrastructure engineering,* 10**,** 614-626.

HAO, X., YANG, J., YANG, F., SUN, X., HOU, Y. & WANG, J. 2022. Track geometry estimation from vehicle-body acceleration for high-speed railway using deep learning technique. *Vehicle system dynamics,* ahead-of-print**,** 1-21.

HOVAD, E., WIX, T., KHOMIAKOV, M., VASSOS, G., DA SILVA RODRIGUES, A. F., DE MIGUEL TEJADA, A. & CLEMMENSEN, L. H. 2021. Deep Learning for Automatic Railway Maintenance. Cham: Cham: Springer International Publishing.

HAARMAN, M., MULDERS, M. & VASSILIADIS, C. 2017. Predictive maintenance 4.0: predict the unpredictable. *PwC and Mainnovation,* 4.

JING, G., QIN, X., WANG, H. & DENG, C. 2022. Developments, challenges, and perspectives of railway inspection robots. *Automation in construction,* 138**,** 1.

JO, T. 2021. Machine learning foundations : supervised, unsupervised, and advanced learning. 1st ed. 2021. ed. Cham, Switzerland: Springer.

KALATHAS, I. & PAPOUTSIDAKIS, M. 2021. Predictive maintenance using machine learning and data mining: A pioneer method implemented to greek railways. *Designs,* 5**,** 1-18.

KANG, Z., CATAL, C. & TEKINERDOGAN, B. 2020. Machine learning applications in production lines: A systematic literature review. *Computers & industrial engineering,* 149**,** 106773.

KAUR, H., PANNU, H. & MALHI, A. 2019. A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions. *ACM computing surveys,* 52**,** 1-36.

KHAJEHEI, H., AHMADI, A., SOLEIMANMEIGOUNI, I., HADDADZADE, M., NISSEN, A. & LATIFI JEBELLI, M. J. 2022. Prediction of track geometry degradation using artificial neural network: a case study. *International journal of rail transportation (Online),* 10**,** 24.

KUMAR, U. & GALAR, D. 2020. Transformative Maintenance Technologies and Business Solutions for the Railway Assets. Cham: Cham: Springer International Publishing.

LAMBERTS, S. W. J. 2009. Operations Research Models for Scheduling RailwayInfrastructure Maintenance. Available: http://docplayer.net/17901020-Operations-research-models-for-scheduling-railway-infrastructure-maintenance.html.
LASISI, A. & ATTOH-OKINE, N. 2018. Principal components analysis and track quality index: A machine learning approach. *Transportation research. Part C, Emerging technologies,* 91**,** 230-248.
LEE, J. S., HWANG, S. H., CHOI, I. Y. & KIM, I. K. 2018. Prediction of Track Deterioration Using Maintenance Data and Machine Learning Schemes. *Journal of transportation engineering, Part A,* 144**,** 4018045.
LOPES GERUM, P. C., ALTAY, A. & BAYKAL-GÜRSOY, M. 2019. Data-driven predictive maintenance scheduling policies for railways. *Transportation research. Part C, Emerging technologies,* 107**,** 137-154.
MARHON, S. A., CAMERON, C. J. F., KREMER, S. C., BIANCHINI, M., MAGGINI, M. & JAIN, L. C. 2013. *Recurrent Neural Networks,* Berlin, Heidelberg :.
MATIC, A. 2021. Implementation of Predictive Maintenance in Railway Infrastructure Maintenance Trondheim: NTNU.
MEINDL, B., AYALA, N. F., MENDONÇA, J. & FRANK, A. G. 2021. The four smarts of Industry 4.0: Evolution of ten years of research and future perspectives. *Technological forecasting & social change,* 168**,** 120784.
MERCY, K. G. & SRINIVASA RAO, S. K. A Framework for Rail Surface Defect Prediction Using Machine Learning Algorithms. Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018, 2018. 972-977.
NAJEH, T., LUNDBERG, J. & KERROUCHE, A. 2021. Deep-learning and vibration-based system for wear size estimation of railway switches and crossings. *Sensors (Basel, Switzerland),* 21**,** 5217.
NAKHAEE, M. C., HIEMSTRA, D., STOELINGA, M., NOORT, M. V., COLLART-DUTILLEUL, S., LECOMTE, T. & ROMANOVSKY, A. B. 2019. The Recent Applications of Machine Learning in Rail Track Maintenance: A Survey. *Lecture notes in computer science***,** 91-105.
NÉMETH, A. & FISCHER, S. 2019. LABORATORY TEST RESULTS OF GLUED INSULATED RAIL JOINTS ASSEMBLED WITH TRADITIONAL STEEL AND FIBRE-GLASS REINFORCED RESIN-BONDED FISHPLATES. *Science and Transport Progress. Bulletin of Dnipropetrovsk National University of Railway Transport*.
OUADAH, A., ZEMMOUCHI-GHOMARI, L. & SALHI, N. 2022. Selecting an appropriate supervised machine learning algorithm for predictive maintenance. *International journal of advanced manufacturing technology*.
PATRA, A. P. 2009. *Maintenance decision support models for railway infrastructure using RAMS & LCC analyses*.
PINTELON, L. & PARODI-HERZ, A. 2008. Maintenance: An Evolutionary Perspective. London: London: Springer London.
PURSSELL, E. & MCCRAE, N. 2020. *How to Perform a Systematic Literature Review: A Guide for Healthcare Researchers, Practitioners and Students,* Cham, Cham: Springer International Publishing.
RAMA, D. & ANDREWS, J. D. 2013. A reliability analysis of railway switches. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 227**,** 344-363.

RENGASAMY, D., JAFARI, M., ROTHWELL, B., CHEN, X. & FIGUEREDO, G. P. 2020. Deep learning with dynamically weighted loss function for sensor-based prognostics and health management. *Sensors (Switzerland),* 20.

RIBEIRO, R. P., PEREIRA, P. & GAMA, J. 2016. Sequential anomalies: a study in the Railway Industry. *Machine Learning,* 105**,** 127-153.

SHAI, B.-D. & SHAI, S.-S. 2014. Nearest Neighbor. *Understanding Machine Learning: From Theory to Algorithms.* Cambridge: Cambridge University Press.

SHARMA, S., CUI, Y., HE, Q., MOHAMMADI, R. & LI, Z. 2018. Data-driven optimization of railway maintenance for track geometry. *Transportation research. Part C, Emerging technologies,* 90**,** 34-58.

SIDDAWAY, A. P., WOOD, A. M. & HEDGES, L. V. 2019. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annu Rev Psychol,* 70**,** 747-770.

SOARES, N., AGUIAR, E. P. D., SOUZA, A. C. & GOLIATT, L. 2021. Unsupervised machine learning techniques to prevent faults in railroad switch machines. *International Journal of Critical Infrastructure Protection,* 33.

SRESAKOOLCHAI, J. & KAEWUNRUEN, S. 2022. Railway defect detection based on track geometry using supervised and unsupervised machine learning. *Structural health monitoring*.

STENSTRÖM, C. 2014. *Operation and maintenance performance of rail infrastructure: Model and Methods*.

SUTHAHARAN, S. 2016. Machine Learning Models and Algorithms for Big Data Classification : Thinking with Examples for Effective Learning. 1st ed. 2016. ed. New York, NY: Springer US : Imprint: Springer.

TAN, C. W., WEBB, G., PETITJEAN, F. & REICHL, P. 2017. *Machine learning approaches for tamping effectiveness prediction*.

TRETTEN, P., ILLANKOON, P. & CANDELL, O. 2021. Digitalization of Railway Maintenance: A Situation Awareness Perspective.

TZANAKAKIS, K. 2013. The Railway Track and Its Long Term Behaviour : A Handbook for a Railway Track of High Quality. 1st ed. 2013. ed. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer.

UIC 2016. UIC Railway Application Guide. *Practical implementation of Asset Management through ISO 55001.* Paris: UIC.

VALE, C. & SIMÕES, M. L. 2022. Prediction of Railway Track Condition for Preventive Maintenance by Using a Data-Driven Approach. *Infrastructures (Basel),* 7**,** 34.

VASSOS, G., HOVAD, E., DUROSKA, P., THYREGOD, C., DA SILVA RODRIGUES, A. F. & CLEMMENSEN, L. H. 2021. Labelling the State of Railway Turnouts Based on Repair Records. Cham: Cham: Springer International Publishing.

WANG, Q.-A. & NI, Y.-Q. 2019. Measurement and forecasting of high-speed rail track slab deformation under uncertain SHM data using variational heteroscedastic gaussian process. *Sensors (Basel),* 19**,** 3311.

WANG, Q., BU, S. & HE, Z. 2020. Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment with LSTM-RNN. *IEEE Transactions on Industrial Informatics,* 16**,** 6509-6517.

XIAO, C. & SUN, J. 2021a. Convolutional Neural Networks (CNN). *Introduction to Deep Learning for Healthcare.* Cham: Springer International Publishing.

XIAO, C. & SUN, J. 2021b. Deep Neural Networks (DNN). *Introduction to Deep Learning for Healthcare.* Cham: Springer International Publishing.

XIAO, C. & SUN, J. 2021c. Introduction. *Introduction to Deep Learning for Healthcare.* Cham: Springer International Publishing.

XIAO, C. & SUN, J. 2021d. Recurrent Neural Networks (RNN). *Introduction to Deep Learning for Healthcare.* Cham: Springer International Publishing.

XIE, J., HUANG, J., ZENG, C., JIANG, S.-H. & PODLICH, N. 2020. Systematic literature review on data-driven models for predictive maintenance of railway track: Implications in geotechnical engineering. *Geosciences (Basel),* 10**,** 1-24.

9 Appendix

9.1 Quality Assessment Form

Questions
Were the aim and scope clearly stated?
Was the research design appropriate for addressing the aim and scope?
Was the data collected properly associated with the problem statement?
Was the data analysis performed appropriately? Diversion from initial scope?
Are the findings stated clearly? Are some results missing?
Is the research valuable?

9.2 Interview Guide

Show gratitude for participation
Shortly explain the topic and scope of the project
Anonymity?
What is the status of ML implementations to PdM? - How many projects are completed? - Have they been successful? Why, why not?
Which ML algorithms/methods have you utilized? - Reason for selection?
Which ML implementation challenges have you faced? - How did you solve these?
Which railway infrastructure assets have been supported by ML? - Some more than others?
Do you have related project documents that I can look at?
Thanks for your time

9.3 Table Overview of Selected Papers

Reference	Aim	ML Algorithm(s)	Asset(s)	Data Acquisition Source(s)	Datatypes	Asset dimensions and data size	Main Findings	Database	Citations
Lopes Gerum et al. (2019)	Failure Prediction	RF, RNN, k-means	Rail and track geometry defects	Rail defects and inspection database	Historical data, geometry data, Asset data	Rail data - 2-yr period (2016 - 2017). Segments averaging 17 km in length. 26 000 inspections and 82 000 defects	• Presented a new approach for rail and geometry defects prediction, and an optimal scheduling approach • Used RNN to predict the defects based on integrated defect and inspection data, before a discounted Markov decision process model determined the optimal inspection and maintenance scheduling strategies • New framework for integrating prediction with inspection and maintenance scheduling activities • The framework is effective for defect prediction and for developing long-term maintenance scheduling strategies with real-time track conditions	Oria	45
Allah Bukhsh et al. (2019)	Failure Prediction	DT, RF, GBT	Railway switches	SAP/ERP maintenance request process	Historical data of visual inspection, maintenance records, and condition state	802 switches (2011-2017)	• Utilized tree-based methods for maintenance prediction of railway switches • Used maintenance input data, e.g., switch component, problem reason and/or cause, location, track type, age etc. • GBT performed best for the prediction of maintenance need, while RF was the most accurate prediction model for maintenance activity type and trigger's status • The approach is applicable to discrete types of railway infrastructure assets	Oria	58

Wang et al. (2020)	Failure Prediction	RNN	Gas-insulated switchgear (electrical equipment)	Data from power equipment	Historical sample data	Sample generator from deterioration and failure data. 10^5 samples with 20 observations each	• Proposed new solution for maintenance of power equipment, by combining both PdM and model-based approaches • Designed a sample generator and a maintenance predictor • The maintenance predictor powered by LSTM-RNN brought positive results to predict future maintenance activities based on historical sample data	Oria	29
Lasisi and Attoh-Okine (2018)	Failure Prediction	SVM, RF, LDA, PCA	Railway Track	Manual inspection or inspection vehicles	Track geometry (includes wide gauge, cross level, vertical profile, warp/twist, alignment) data	28 inspection dates of 1 mile of track. Dataset of 31 features collected from a section of US Railway track	• Demonstrated how combined track geometry parameters with track quality indices, simplified the track properties without loss of variability in the data • The most effective technique was discovered to be SVM, which predicts track defects better than track quality indices • Since the defect data was very unbalanced, prediction performance was measured using TPR (true positive rate) and FPR (false positive rate)	Oria	73
Hao et al. (2022)	Failure Prediction	AM, CNN, GRU	Railway track	Inspection trains	Track geometry data (vertical profile, alignment, gauge, cross-level, twist, etc.), vehicle body-acceleration and vehicle speeds	Inspection data collected at 0.25 m increments along 600 km track (300 km for testing). Max management wavelength of track irregularities is 120 m.	• Proposed a model for predicting track irregularities using vehicle-body accelerations from measurement data • Proposed AM-CNN-GRU model to learn shape- and sequential features and focus on the most important features • AM-CNN-GRU performs better than GRU and CNN-GRU models	Oria	0

Hovad et al. (2021)	Failure Prediction	CNN	Rails and joints	Line scan cameras on recording car	Vertically captured images	Each image covers approx. 2.507 m of track and do not overlap with other images. 2155 total images. Ca. 5.4 km in total	• Proposed a computer vision system for automatic detection and prediction of defects on rails and joints through images • The results showed great versatility of the model which can potentially reduce the need for visual inspections • The recall rate for defects was 84 %, rarely missing surface defects	Oria	1
Vassos et al. (2021)	Failure Prediction	PCA, k-means	Switches	Repair records and track recording car	Historical repair data and track turnout geometry data	Data from 166 turnouts with similar geometric design and components	• Predicted a turnout's maintenance need based on two clusters of data • Geometric track measures described one of the clusters to show problematic behavior, while the other cluster was the opposite • No clear boundary exists between the clusters, and it is difficult to assign a specific cluster label successfully	Oria	0
Vale and Simões (2022)	RUL	PCA, LgR	Railway Track	Historical track condition data and inspection vehicle	Track geometry (includes longitudinal level, alignment, gauge, cross level, and twist) data	51 km track, 14 inspections	• Developed a data-driven model based on LgR and PCA, to predict railway track condition • The utilized predictors are longitudinal level and alignment of left and right rail, time interval between inspection actions, and sequential number of track segments • The model is easy to implement and resulted in 91.1% success rate	Oria	0

Guler (2014)	RUL	ANN	Railway track	Vehicle and various devices	Track structure, traffic characteristics, track layout, environmental factors, track geometry, and maintenance and renewal data.	180 km railway track gathered over 2 years	• Modelled railway track degradation with ANN • The study proved that ANN is one of the best ways to predict track geometry degradation linked with a big inherent complexity • Presented a robust model without needing expert knowledge	Oria	89
Tan et al. (2017)	RUL	kNN	Ballast	Multivariate sensors from Instrumented revenue vehicle	Tamping effectiveness data	Three months of journey data. 50 m block of tracks	• The presented method achieved high accuracy in prediction of tamping effectiveness • Tamping effectiveness should improve maintenance efficiency • Achieved accuracy of 68% when predicting tamping effectiveness 12 weeks prior tamping, and 70% accuracy 1 day prior	Oria	4
Cárdenas-Gallo et al. (2017)	RUL	LgR, SVM	Railway track	Track geometry vehicles	Track geometry data	Dataset from four track segments. Unspecified total length	• Created a method for developing an ensemble classifier to predict evolution of track geometry defects • The ensemble classifier outperformed the individual algorithms in every case • Can be implemented to other tracks and other types of defects	Oria	43
Wang and Ni (2019)	RUL	Variational heteroscedastic Gaussian process (Bayesian method)	Track slab	Fiber Bragg Grating sensor	Real-time monitoring data	5 track slabs in laboratory and field test datasets	• Developed, tested, and implemented a novel online SHM system using FBG sensing technology • Experimental results show improvement in terms of regression compared to the state-of-the-art algorithm • The method improves the prediction accuracy for all sensors	Oria	17

Sharma et al. (2018)	Failure Prediction	RF	Railway track	Public data	Track geometry data	50 miles track, 33 months	• Developed a fault prediction maintenance policy for the railway track geometry • The developed policy results in an approximately 10% savings in the total maintenance costs for every 1 mile of track • Predicted the spot geometric defect occurrence probability using RF	Oria	80
Sresakoolchai and Kaewunruen (2022)	Failure Prediction	DNN, CNN, MR, SVM, GB, DT, k-means	rail, rail joint, switch and crossing, and fastener	Track geometry car	Track geometry data, track component defect data, and track profile data	30 km, data collected from 2016-2019	• Developed models with supervised learning algorithms to predict component defects using track geometry • DNN performed with the best accuracy of 94.3%, followed by CNN with 93.8%. Other models were about 50% or lower • Detected defects and categorized into rail, rail joint, switch and crossing, and fastener • Insights of track component defects are investigated using k-means • The method does not add any additional costs for developing the system	Oria	0
Najeh et al. (2021)	RUL	RNN (LSTM), CNN (RESNet)	Switches and crossings	Sensors (accelerometer) on track	Vibration data	6 sensors on the S&C	• Developed a RNN and CNN based framework to design an effective prediction system of S&C wear development • Resulted in acceptable accuracy estimation of wear in the middle section of the S&C • The solution will be implemented by the railway owner to predict the wear evolution and monitor and analyze the condition of S&Cs	Oria	1

Khajehei et al. (2022)	RUL	ANN, LnR	Railway track	Database and system	Track geometry data, asset information, and maintenance history	5-line sections (total of 216km)	• Presented a model for predicting track degradation rate, collecting track geometry data, asset information and maintenance history • Applied ANN which resulted in adequate performance • The most important contributor to the prediction of track geometry degradation rate is maintenance history • The amount of tamping activities highly influences the degradation rate du to its destructiveness	Oria	3
Lee et al. (2018)	RUL	ANN, SVR	Railway Track	ERP system	Track properties (ERP), tamping data, geometry data	Used datasets from external sources	• Studied application of SVR and ANN to predict track degradation with simulation data similar to field conditions • Performed two case studies with different input parameters • ANN performed marginally better than SVR due to the retraining process • Results show that min. 2 yrs of maintenance data are needed to develop a stable prediction • The prediction results can be used in the DSS within the framework of ERP	Scopus	19

Mercy and Srinivasa Rao (2018)	Failure Prediction	DNN, DT, RF	Rail	Track geometry car	Rail geometry data	9 attributes, 228 instances. Unspecified total length.	• Presented three ML methods for prediction of rail surface defects (e.g., unevenness, twist, alignment and gauge are used) • The results showed that DT had the highest accuracy for predicting twist, and unevenness right and left. RF and DT have equal accuracy when predicting alignment left. RF has the best prediction accuracy for alignment right and gauge. • DNN was outperformed by DT and RF	Scopus	5
Soares et al. (2021)	Failure Prediction	PCA, k-means, Gaussian method	Switch machine	Switch machine sensors	Switch parameters data	615 inspections of 6 switch machines with same specifications	• Proposed a predictive maintenance model to prevent failures of switch machines • The model includes feature extraction and selection procedures based on unsupervised ML techniques • The approach showed impressive results and was efficient once it had considered critical operations conducted in the vicinity of the ones classified as faults	Scopus	4
Arslan and Tiryaki (2020)	Failure Prediction	ANN, SVM, naive bayes, DT	Switches	Switch motors	Switch movement, position, status, movement time	Data from six switch points. Consisted of 7 inputs and 1 output from each, and 6000 switch point movements was utilized	• Aimed to predict railway switch point failures by utilizing SVM and ANN • Compared ANN and SVM to other methods such as DT and Naive bayes. ANN and SVM outperformed all of them • ANN was observed to be more appropriate in the implementation of the established model • If inputs were to change, the presented method will correctly determine the outputs and correctly predict failures	Scopus	7

NTNU

Norwegian University of
Science and Technology