Deliverable D1.2

Summary of existing relevant projects and state-of-the-art of AI application in railways

Project acronym:	RAILS
Starting date:	01/12/2019
Duration (in months):	43
Call (part) identifier:	H2020-S2R-OC-IPX-01-2019
Grant agreement no:	881782
Due date of deliverable:	Month 16
Actual submission date:	29 March 2021
Responsible/Author:	Nikola Bešinović, TUDELFT
Dissemination level:	Public
Status:	Issued

Reviewed: no

Document history		
Revision	*Date*	*Description*
1	5 February 2021	First issue for internal review
2	11 February 2021	Second issue for Advisory Board review
3	20 March 2021	Third issue for external review

Report contributors		
Name	Beneficiary Short Name	Details of contribution
Nikola Bešinović	TUDELFT	First version, Second version, Third version
Ruifan Tang	LEEDS	First version, Second version, Third version
Zhiyuan Lin	LEEDS	First version, Second version, Third version
Ronghui Liu	LEEDS	First version
Tianli Tang	LEEDS	First version
Lorenzo De Donato	CINI	First version, Second version, Third version
Valeria Vittorini	CINI	First version, Third version
Ziyulong Wang	TUDELFT	First version
Francesco Flammini	LNU	Internal review
Mauro Pappaterra	LNU	Internal review
Rob Goverde	TUDELFT	Internal review

Advisory Board Reviewers	
Name	*Company or Institution*
Wilco Tielman	ProRail
Stefano Delucchi	Aitek

Funding

This project has received funding from the Shift2Rail Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement no 881782 Rails. The JU receives support from the European Union's Horizon 2020 research and innovation program and the Shift2Rail JU members other than the Union.

Disclaimer

Contents

Executive Summary 5
Abbreviations and acronyms 6
1. Background 8
2. Objective 9
3. Introduction 10
3.1 Definitions 10
3.1.1 AI domains 11
3.1.2 Railway domain 11
3.2 Report outline 13
4. Review of related projects on AI in railways 14
4.1 Methodology 14
4.2 European projects 16
4.2.1 Overview of Shift2Rail programme 16
4.2.2 Shift2Rail projects 16
4.2.2.1 New 2020 S2R starting projects 32
4.2.3 Non-Shift2Rail projects 37
4.3 Non-European Projects 42
4.3.1 United States 42
4.3.2 China 47
4.4 Classification and Discussion 48
4.4.1 Shift2Rail perspective 49
4.4.1.1 Classification 50
4.4.1.2 Discussion 52
4.4.2 Global perspective 53
4.4.2.1 Classification 54
4.4.2.2 Discussion 55

5. State-of-the-art 58
5.1 Methodology 58
5.2 Existing review papers 61
5.2.1 AI in Railway domains & other sectors 61
5.2.2 Findings & Gaps 63
5.3 Overview of selected papers 66
5.3.1 Distribution of papers per journal or conference 66
5.3.2 Number of papers in each single year 67
5.3.3 Distribution of paper per railway subdomain over years 67
5.3.4 Proportion of papers in rail domains 68
5.3.5 Classification of papers in accordance with their objectives 69
5.3.6 Classification of reviewed papers 74
5.4 Paper reviews by railway domains 74
5.4.1 Maintenance and inspection 75
5.4.2 Safety and security 79
5.4.3 Autonomous train driving and train control 80
5.4.4 Traffic planning and management 80
5.4.5 Passenger mobility 83
5.5 Discussion 84
6. Future directions 90
7. Conclusions 96

Executive Summary

This report presents a comprehensive review of research projects and state-of-the-art scientific papers of Artificial Intelligence (AI) in railway industry. It represents the output of RAILS Work Package 1 Tasks 1.2 and 1.3. The focus is on European and overseas projects mainly in United States and China. Specific emphasis is devoted to reviewing projects funded by the European Shift2Rail Joint Undertaking, which represents one of the main funding bodies in EU for railway research and innovation. We address the projects and scientific papers from a holistic railway perspective covering selected areas as defined in Deliverable 1.1, including (1) maintenance and inspection, (2) safety and security, (3) autonomous driving and control, (4) transport planning and management, (5) revenue management, (6) transport policy and (7) passenger mobility. As such, this report makes an initial step towards shaping the role of AI in future railways, and provides an in-depth summary of current focus of AI research connected to rail transport. In addition, the report determines some promising research directions towards further uptake of AI in railways.

The report recognizes that the major research efforts have been put in AI for rail maintenance and inspection, while very limited or no research has been done on AI for rail transport policy and revenue management. The remaining subdomains received mild to moderate attention. The potential of AI applications is evident, but it is also highlighted that AI research in railways is at its early stages. Thus, many open research topics are envisioned, some of which could contribute to fundamental usage of AI in general. Future research can be expected towards, among others, developing advanced combined AI applications, using AI in decision making, dealing with uncertainty and tackling newly rising cybersecurity challenges.

This review provides general guidelines to support railway researchers to assess and understand the usability of AI techniques. It is also meant to support industry stakeholders to promptly determine promising AI domains for given railway problems. The objective of this deliverable is to further contribute to bridging the gap between AI and railway-domain experts.

Abbreviations and acronyms

Abbreviations / Acronyms	Description
ACO	Ant Colony Optimization
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AR	Augmented Reality
ATO	Automatic Train Operation
ATP	Automatic Train Protection
BIM	Building Information Management/Modelling
CBM	Condition-Based Maintenance
CNN	Convolutional Neural Networks
CV	Computer Vision
DBM	Deep Boltzman Machines
DBN	Deep Belief Networks
DL	Deep Learning
DM	Data Mining
EAs	Evolutionary Algorithms
ETCS	European Train Control System
EU	European Union
GA	Genetic Algorithms
GBDT	Gradient Boosting Decision Tree
GP	Genetic Programming
ICT	Information and Communication Technology
IoT	Internet of Things
IP	Innovation Programme
IVG	Intelligent Video Gate
IXL	Interlocking
KPI	Key Performance Indicator
ML	Machine Learning
NLP	Natural Language Processing
OR	Operations Research
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
RAILS	Roadmaps for AI integration in the raiL Sector
RNN	Recurrent Neural Networks
RUL	Remaining Useful Life
S2R	Shift2Rail
SI	Swarm Intelligence
SVM	Support Vector Machine
TMS	Traffic Management System
UAV	Unmanned Aerial Vehicle
US	United States
UTC	University Transportation Centers

1. Background

The present document constitutes the Deliverable D1.2 “Summary of existing relevant projects and state-of-the-art of AI application in railways” of the S2R JU project “Roadmaps for AI integration in the Rail Sector” (RAILS). The project is in the framework of Shift2Rail’s Innovation Programme IPX. As such, RAILS does not focus on a specific domain, nor does directly contribute to specific Technical Demonstrators but contributes to Disruptive Innovation and Exploratory Research in the field of Artificial Intelligence within the Shift2Rail Innovation Programme.

Deliverable D 1.2 describes the work carried out in tasks 1.2 and 1.3 of Work Package WP1 whose objectives are:

- Define taxonomy of AI to enable its application in railway transport
- Determine state-of-the-art of AI techniques in railway transport
- Determine state-of-the-art of Shift2Rail projects
- Identify application areas of AI in railways

The work package identifies essential conditions: it provides specific needs, determined capabilities and gap analysis; it evaluates and selects techniques and methods for picking the right AI technology able to solve open problems or improve performance in railway scenarios. To this aim, a systematic literature review is performed including the study of the available results from recent and ongoing projects. This work package clearly states a vision: it depicts a preliminary shared vision of the future railway systems among partners, it bridges the gap between AI and railway-domain experts by clearly defining the expectations of the potential of emerging technologies, and selects the most suitable AI techniques for their realization. Finally, it defines scope and boundaries of AI-Technology against railway targets, for identifying new opportunities in achieving desired objectives from the development of emerging AI-technologies, i.e. identify application areas of AI across railway domains. The main outcomes of this work package are:

i. A taxonomy of suitable AI techniques to be adopted for railways (Deliverable D1.1);
ii. A map of the current state-of-the-art in railway research from S2R and other relevant projects (this Deliverable);
iii. A set of current and potential application areas (Deliverable D1.3).

The overall objective of the RAILS research project is to investigate the potential of AI approaches in the rail sector and contribute to the definition of roadmaps for future research in next generation signalling systems, operational intelligence, and network management. RAILS addresses the training of PhD students to support the research capacity in A.I. within the rail sector across Europe by involving research institutions in four different countries with a combined background in both computer science and transportation systems.

2. Objective

This deliverable reviews the outcomes and the ongoing progress of relevant railway projects worldwide with the emphasis on AI-focused developments and state-of-the-art research of using AI in railway domain. First, it covers Shift2Rail projects in all IPs to determine known railway challenges that have already being tackled such as SMART [1], MyTrack [2], and IN2SMART [3]. In addition, a review of previous relevant work undertaken at a European level, for example, ON-TIME [4], CAPACITY4RAIL [5], as well as overseas, from the US and China. Second, we review the existing AI applications in railways to determine current research and practice directions. The review includes usage of AI in railway domains as defined in Deliverable D1.1 [6], among others, in maintenance and inspection, safety and security, vehicle automation, signalling and control, traffic planning and management, transport policy, revenue management, and passenger mobility.

This review of projects and scientific papers leads to recognising the current focus of research worldwide and determining most promising state-of-the-art AI approaches for railway system. We also aim to identify existing challenging railway problems in railways that can be solved with AI; existing problems that can be solved more efficiently by the support of AI, and promising new problems that can be tackled in the future by AI.

3. Introduction

It is widely accepted that artificial intelligence (AI) influences our life in every domain. However, in the railway sector, AI is still largely at its infancy stage. There are certain evidences showing that its potential should not be underestimated and suggesting that AI can play an important role such as optimization of complex railway systems, improving safety and security of urban rail networks and customer service [7]. We have seen some promising studies on AI in railways that suggest AI systems to show potential in providing powerful and solving the critical challenges that railways are facing today [7]. Also, certain authors [8] expect that AI will soon become a common tool used throughout the rail industry. Several topics are discussed where AI is supposed to be able to act as a game changer for the railway sector, such as capacity management, life cycle cost, maintenance, reducing error from both humans and computers, high-level automation and auto-adaptive systems.

To understand the current position of AI in railways, this report reviews existing projects and the state-of-the-art research studies to recognise the current focus worldwide and determine the most promising AI algorithms for railway systems. It also aims to identify promising future research directions to facilitate further uptake of AI by the industry.

The research questions to be answered in this report are the following:

1. What are the current focuses of the surveyed research projects?
2. What is the state-of-the-art in the railway research?
3. What are advantages and limitations of applying AI in railways?
4. What are possible directions for future research?

3.1. Definitions

Based on Deliverable D1.1, this section states a comprehensive definition of Artificial Intelligence (AI), a general AI taxonomy (Section 3.1.1) and railway subdomains addressed in the RAILS project (Section 3.1.2). A general perception when discussing the AI is to describe what the extent a man-made machine thinks and acts in a way of human beings. In the era when Artificial Intelligence was not proposed, machines are born to perform complex computations in accordance with instructions or commands entered by humans. The significant advancements brought by AI represented as an artificial agent not only shows abilities to acquire new knowledge, make decisions and solve complex problems, but have a good performance on speech understanding, reasoning, image processing and sentiment analysing. Nevertheless, the scenarios when a specific AI contribution applied in railways are different, since complex business requirements need to be satisfied. It is challenging task to include all the relevant applications in and summarize them with few descriptions. Alternatively, an integrated framework of terminologies can be used to properly accommodate the complexity and heterogeneity of AI in railways.

3.1.1. AI domains

AI itself is a broad domain with various branches where multiple techniques and applications have been well-developed no matter in the perspective of breadth and depth. By considering the definitions given in previous literature and combining them with the definitions of the reference classification of AI in railways, we define AI taxonomy as an integrated structured concept. We frame the complexity and breadth of AI terminology and divide it into the following four main categories: AI techniques, AI research fields, AI applications and AI related disciplines.

AI Techniques AI Techniques represent methods, algorithms and approaches enabling systems to perform tasks commonly associated with intelligent behaviours. We have the follow three sub-domains included in this category: Machine Learning, Evolutionary computing, Logic programming.

AI Research Fields AI Research Fields represent the commonly investigated research areas that rely on the AI techniques introduced above and would not exist without them. We define the following domains as AI research fields: expert systems, data mining, pattern recognition and adversarial search.

AI Applications AI Applications represent cross-domain applications where AI can be used, either solely or in conjunction with other conventional methods, to solve real-world problems as reflected in these domains. The following are considered as AI applications in our literature review such as computer vision & image processing, natural language processing & speech recognition, autonomous systems & robotics, and operations research & scheduling and planning.

AI Related Disciplines AI Related Disciplines represent domains at the same time benefiting from and beneficial to AI. The following are included as related disciplines in our literature review such as big data analytics, digital twins, and augmented reality.

Figure 3.1 [9] gives a Class Diagram illustrating the AI domains in techniques, research fields, applications and related disciplines according to the definitions given above. Note that AI is constantly evolving and new concepts would need to be added as they emerge therefore we do not intend here to develop the most exhaustive taxonomy of AI. Instead, we mainly focus on the mainstream AI domains and also the ones having close interactions with railway domains.

3.1.2. Railway domain

In order to better clarify the existing or potential applications of AI in the railway field, we consider seven subdomains of railways including maintenance and inspection, safety and security, autonomous driving and control, traffic planning and management, revenue management, transport policy and passenger mobility.

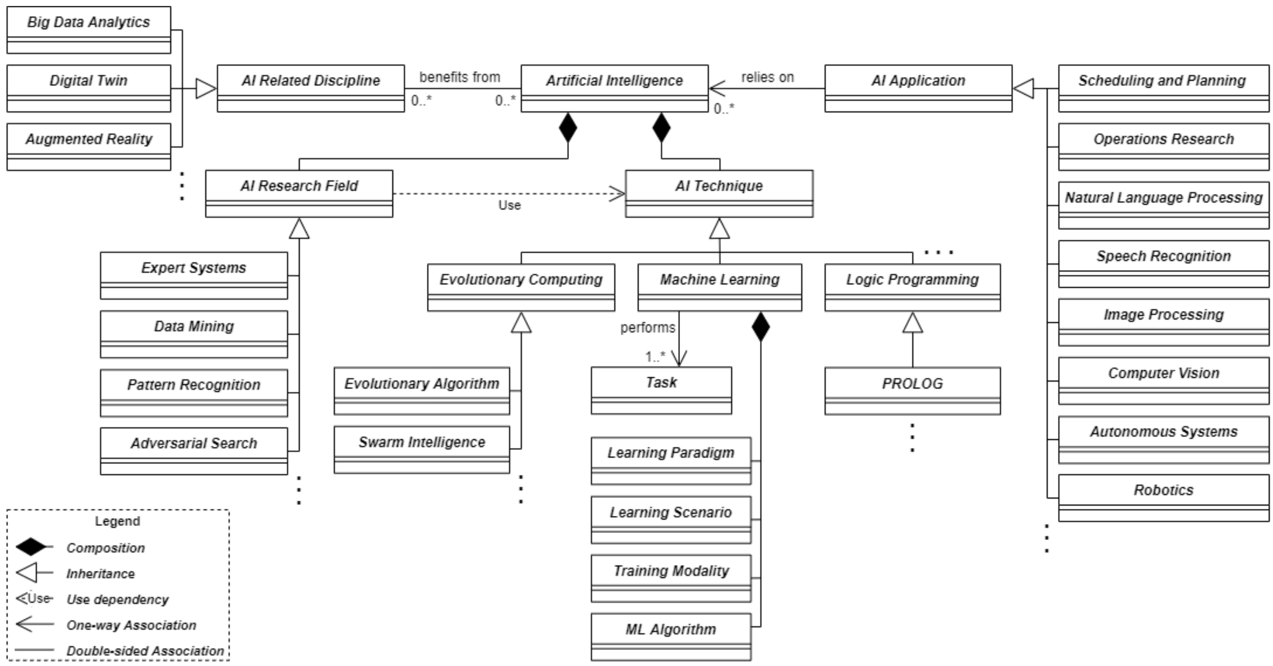

Fig. 3.1. Artificial Intelligence Taxonomy Class Diagram [9]

Maintenance and inspection Railways are made up of complex mechanical and electrical systems and there are hundreds of thousands of moving parts. A railway will not survive for long as a viable operation if it is allowed to deteriorate and become unsafe because of lack of maintenance.

Safety and security Safety and security are of primary concern for any transport system. Travellers expect transportation to be safe.Transport safety is defined as the absence of accidents that may cause injury to persons and/or physical damages.

Autonomous driving and control Autonomous driving and control represent operating trains automatically without any (or with only limited) human intervention. Trains can be classified according to four levels of automation, also known as Grade-of-Automation (GoA). From GoA1, where the presence of the driver is strictly necessary also for basic operations, up to GoA3/4, where the driver acts as a supervisor and advanced ATO functionalities are implemented (e.g. obstacle detection) to allow the train to anonymously run from a station to another also handling emerging situation on its own. GoA2 is the intermediate level that encompasses semi-automatic train operations, and the driver is responsible of many safety-related procedures. During operation, the rolling stocks are constantly monitored by a train control and monitoring system (TCMS) which aids with coupling the trains and ensure safety. The goal is to move towards automatic train control (ATC) functionalities to allow trains to self-adapt and communicate (e.g. virtual coupling) in order to implement autonomous control features.

Traffic planning and management Traffic planning and management includes research and applications tackling traffic state prediction, timetabling, and traffic rescheduling, as well as some more strategic planning decisions such as signalling and station design.

Revenue management Revenue management is the application of disciplined analytics that predict consumer behaviour at the micro-market levels and optimize product availability and price to maximize revenue growth.

Transport policy Transport policy deals with the development of a set of constructs and propositions that are established to achieve specific objectives relating to social, economic and environmental conditions, and the functioning and performance of the transport system.

Passenger mobility Mobility is the ability to move all people safely and affordably between where they live, work, and spend their leisure time. It includes walking, cycling, vehicle sharing, public transportation, and much more. Mobility is the ability to move or be moved freely.

3.2. Report outline

This report is organized as follows. Section 4 gives an overview of AI-related railway projects in Europe and overseas and answers the research question 1. Section 5 surveys state-of-the-art articles in railway domain and answers the research question 2 and 3. Section 6 presents future directions for the research on AI in railway sector and answers the research question 4. Finally, Section 7 provides final remarks.

4. Review of related projects on AI in railways

In this chapter, we give a systematic review on related projects worldwide where artificial intelligence is used in railways to determine known railway challenges, including outcomes and progress, that have already being tackled around the world. This review covers relevant projects at European and international levels, where focused regions and countries include specifically Europe, US, and China. In addition, we put additional focus on reviewing projects within the dedicated EU funding programme for railway sector called Shift2Rail

The efforts are made to cover as many projects as possible as long as their information is publicly available. The used methodology for selecting projects is given in 4.1. We first focus on European projects in 4.2. Specifically, Section 4.2.1 briefly overviews the S2R programme, Section 4.2.2 surveys Shift2Rail projects and Section 4.2.3 covers non-Shift2Rail projects within Europe. In Section 4.3, we survey projects outside Europe. In Section 4.3.1 relevant projects carried out in the United States (US) have been reviewed, then, in Section 4.3.2 we focused on Chinese projects. Lastly, in Section 4.4, we present some summarizing charts and final considerations about our findings.

4.1. Methodology

The major tool for surveying projects in the above countries is the Internet and we survey projects with information and output that are publicly available. For EU projects, *Cordis* [10] is used, which is a well developed and maintained project database and afterwards, we reached out to project webpages, when available, for additional information. For US projects, we used Transportation Research Board's Research in Progress TRB's RiP [11] database, and for Chinese projects – China National Knowledge Infrastructure (*CNKI*) [12]. We focused on reviewing projects in the period from 2011 to 2020, as we expect that AI research received significant funding mainly in the recent years. In addition, to collect Shift2Rail projects Shift2Rail official website [13] was used. Table 4.1 gives more details about the above sources.

Table 4.1: Sources of databases used for project surveying

Source	URL	Region	Remark
Cordis	https://cordis.europa.eu/	EU	
S2R	https://shift2rail.org/	EU	Projects section
RiP	https://rip.trb.org/	US	
CNKI	https://www.cnki.net/	China	

The review was performed according to a fixed methodology which encompasses two main steps:

1. Build one or more queries combining keywords related to AI Techniques/Applications (as form Figure 3.1) and filtering the results by projects. We use key words such "ar-

tificial intelligence", "railway", "projects", "machine learning", "evolutionary computing", "computer vision", and so on,

2. Manually analyse the projects' documents retrieved and filter out the irrelevant projects. We established an exclusion criterion by focusing only on projects which directly involve Artificial Intelligence Techniques, or Applications (when they are effectively AI-related), to achieve at least one of their purposes. However, we also retrieved projects that, being still ongoing, only make assumptions on the possibility to involve AI in their future tasks. Although no deliverables/outcomes are already analysable, in our understanding, these projects are equally noteworthy since they contribute to defining the panorama of the future investigations on AI within the Railway sector.

As an exception to this methodology, we dedicate more attention to the S2R projects, given the relevance of S2R programme as the main pillar of railway research in Europe. We reviewed all projects over the complete duration of the S2R programme, i.e. covering the period from 2015 to 2020. However, we recognised that several relevant projects have started only in December 2020 and thus only with limited information available. Nevertheless, we also briefly refer to these projects to highlight their expected research directions in Section 4.2.2.1 without including them in the further analysis.

After an initial analysis, we observed that the projects' deliverables are not always available (e.g. for ongoing projects) and, in many cases, AI is not explicitly stated as the main focus. Therefore, to assess the S2R projects more systematically/thoroughly, we start the evaluation with all S2R projects. Based on our expert knowledge we indicate manually the projects that are expected to have implemented or will implement some AI-based approaches, even if not specific information about their involvement are available. Given that, the methodology adopted to select S2R projects follows a refinement process slightly different from that described above:

1. At first we analysed the "Objective" and "Project Structure" sections of each project on the Shift2Rail website, as well as their description on Cordis, looking for the same AI-related keywords as in the first step of the methodology mentioned above. Furthermore, since in many cases the focus was not on AI, we also checked for topics we know can be addressed through AI such as "Predictive Maintenance", "Condition Based Maintenance (CBM)", "Smart Planning", "Optimization Strategies", etc. From this first step, we obtained about 40 projects over the 83 (excluding RAILS) S2R projects currently available on the website. Then, we performed further analysis on the projects obtained in this step to understand if they effectively involved AI in at least one of the proposed solutions.
2. Then, we identified the Work Packages and their deliverables (if any) which could contain information about AI and its applicability. If these deliverables were available, then we analyzed them in detail to assess whether AI has been involved and for which purpose; otherwise, we simply focused on the description of the project on Cordis and on the S2R site trying to understand, if not explicitly expressed, how AI could have been applied. This resulted in 28 S2R projects.
3. As the last step, we classified the projects according to the amount of information available. We highlighted the projects for which we found sufficient information with a star ($\star$), and thus for these projects we were able to conduct in depth analysis. It

is important to underline that this is not a division based on the importance of the projects, but a way to better express our results. The star indicates that it is possible to analyse, considering the references, further information on the contributions these projects have made to the integration of AI in the railway sector. We finally obtained 28 S2R projects of which 19 are star projects.

Lastly, to emphasize the AI-related information we retrieved from the reviewed projects (both S2R and not), we generated some tables encompassing roughly the same fields. It is important to clarify that the information reported is strictly related to AI. A given project can face issues belonging to different railway subdomains, but it does not necessarily apply AI in all the proposed solutions. Therefore, we indicate in the tables only such railway subdomains in which AI is directly involved.

4.2. European projects

4.2.1. Overview of Shift2Rail programme

Shift2Rail projects are mainly cataloged in five key Innovation Programmes (IPs) that are:

- *IP1*: future generation passenger trains aiming to achieve Cost-efficient and reliable trains, including high-capacity trains and high-speed trains.
- *IP2*: Advanced Traffic Management and Control, Command and Communication Systems (including automatic train operation, safe satellite positioning, virtual coupling and cyber-security);
- *IP3*: Cost-Efficient and Reliable High-Capacity Infrastructure, including design, construction, operation and maintenance of rail network infrastructure;
- *IP4*: IT Solutions for Attractive Passenger Railway Services.
- *IP5*: Technologies for Sustainable & Attractive European Rail Freight.

These IPs involve different sub-fields in the railway system. The structure of these IPs is given in Figure 4.1.

Nevertheless, it is worth noting that S2R projects can be catalogued also considering two additional areas:

- Innovation Programme X (IPX): Disrupting Innovation Activities. The RAILS project falls within this IP;
- Cross-Cutting Activities (CCA): activities supporting the work conducted within the different above IPs.

4.2.2. Shift2Rail projects

Innovative RUNning gear soluTiOns for new dependable, sustainable, intelligent and comfortable RAIL vehicles (RUN2RAIL)* - IP1

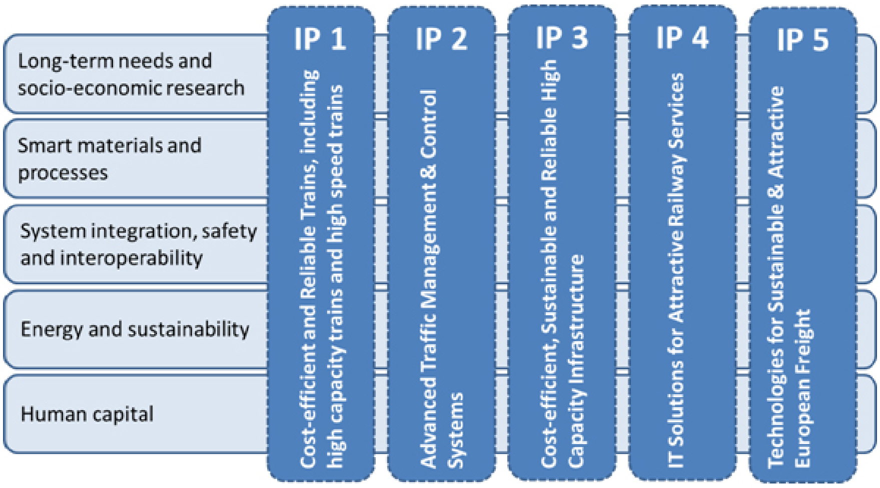

Fig. 4.1. The structure of IPs (Source: https://shift2rail.org/research-development/)

RUN2RAIL [14] aimed to explore new technologies for future running gear in order to design more reliable, lighter, less damaging, more comfortable and less noisy trains. To these aims, the projects went through four main investigative areas achieving, as the final outcome, an ensemble of comprehensive technology concepts and methodologies which represent a starting point for the design of future running-gear generations. Investigated areas and some of the outcomes are listed below:

- Innovative sensors and condition monitoring. It involves a new architecture for on-board condition monitoring, as well as the identification of smart sensors and running gear components with self-diagnosis capability. In this direction, wheelsets, bearings, gearboxes, and suspension components were identified as a case study for condition monitoring proposals. In particular, a Machine Learning approach for fault detection and classification was considered for the creation of a signal analysis toolbox aiming to perform fault diagnosis within bearings and gearboxes.
- Optimised materials and manufacturing technologies. Investigate the possibility to adopt new materials and propose novel production processes and methods (e.g. 3D metal printing) for running gear components.
- Active suspension and mechatronics. An innovative concept related to a single-axle running gear with active suspension and wheelset has been developed aiming to obtain simpler and lighter architectures. Moreover, an approach to determine the fault tolerance of bogies with active suspension has also been developed.
- Noise and vibration. Within the project, a methodology for predicting the transmission of noise and vibration from the running gear to the car-body has been developed. At the same time, an assessment of techniques for noise and vibration reduction has also been carried out.

Activities above also included the examination of case studies to assess the methods and tools proposed within the project in terms of feasibility, cost, and safety.

Cybersecurity in the RAILway sector (CYRAIL)* - IP2

Since the intelligent railway infrastructures easily suffer the cyber-criminals and terrorists, CYRAIL [15] aims to deliver tailored specifications and recommendations for secure modern rail systems design and operation. The main technical objectives of CYRAIL are: i) to select security analysis frameworks capable of assessing the most critical railway services, zones and communications to perform an exhaustive cyber security assessment of the Railway systems; ii) to deliver a taxonomy of threats targeting rail management and control systems capable of classifying, describing and analyse cyber-attack threats; iii) to assess and select innovative rail management systems attack detection techniques; iv) to specify Countermeasures and Mitigation strategies for improved quality levels; v) to describe Resilience Mechanisms for Operational Safety; and vi) to specify Protection Profiles with Evaluation of Assurance Levels.

This project applies an unsupervised learning, anomaly detection, to detect significant deviations in known behaviour. Profiles are derived from monitoring regular activities, network connections, hosts and users over a period time. It can be configured to monitoring many behaviour attributes, such as processor usage and number of failed login attempts. Anomaly detection is very good at detecting previously unknown threats from the regular monitoring data.

X2RAIL projects - IP2

In essence, X2RAIL represent a series of closely related projects, started in 2016, encompassing four S2R projects:

- Start-up activities for Advanced Signalling and Automation Systems (X2RAIL-1) [16];
- Enhancing railway signalling systems based on train satellite positioning, on-board safe train integrity, formal methods approach and standard interfaces, enhancing Traffic Management System functions (X2RAIL-2) [17];
- Advanced Signalling, Automation and Communication System (IP2 and IP5) – Prototyping the future by means of capacity increase, autonomy and flexible communication (X2RAIL-3) [18];
- Advanced signalling and automation system - Completion of activities for enhanced automation systems, train integrity, traffic management evolution and smart object controllers (X2RAIL-4) [19].

In particular, X2RAIL-1 and X2RAIL-4 introduce AI to tackle specific railway challenges. X2RAIL-1, among the other tasks (Adaptable Communication System, ATO over ETCS, Moving Blocks, Zero on-site Testing, and Cybersecurity), a focus on AI is given as an enabling technology to achieve Smart Wayside Object Controllers (SWOC). This is a component that interposes between wayside objects, the Route Management System (e.g. IXL, ATP, etc.), and other SWOCs, intended to manage control, maintenance, and diagnosis data related to the wayside objects. A feasibility study has been carried out analysing the state-of-the-art of wireless data transmission, power supply, maintenance, and diagnosis to assess the SWOC adoption for the realization of a distributed approach to rail automation. The Artificial Intelligence, in the form of Machine Learning, but also related disciplines such as augmented reality, have been identified as exploitable technologies for the diagnosis and maintenance facet.

X2RAIL-4 will mostly focus on finalising the work started within X2RAIL-1 and X2RAIL-2, also providing demonstrators. As reported in the Project Structure, Artificial Intelligence and Operation Research approaches will be evaluated and tested, also considering their combination, in the Traffic Management System scenario given their potentials in the automation and optimization of train movements.

Measuring, monitoring and data handling for railway assets; bridges, tunnels, tracks and safety systems (ASSETS4RAIL)⋆ – IP3

ASSETS4RAIL [20] is an ongoing project aiming to exploit, adapt and test cutting-edge technologies, including Machine Learning and Deep Learning approaches, for railway asset monitoring and maintenance. Its practical purpose is to improve inspection and reduce costs following a twofold approach which focuses on infrastructures (e.g. tunnels, bridges, etc.) in the first workstream, and vehicles in the second one. The keystone of the project is the development of a Building Information Management/Modelling (BIM) platform [21, 22] that will integrate algorithms and will gather information collected by different kinds of sensors. ASSETS4RAIL will results in different exploitable results [23] which can be summarized as follows:

- Validate TRL5[1] new subsurface defect detection scanner and method for cleaning long tunnels drainage pipes;
- Validate TRL6 i) drone for scanning, mapping and inspection of tunnels and bridges, ii) noise and vibration monitoring platform, iii) BIM module for tunnels and bridges with integrated management of monitoring data, iv) software module for fatigue assessment, v) vibration damper for noise reduction in steel railway bridges, vi) wayside stereo Computer Vision hardware, vii) underframe image monitoring, and viii) on-board hardware system for monitoring track geometry;
- A modular measuring system for railway infrastructures (WITT Bridge).

Future Secure and Accessible Rail Stations (FAIR STATIONS)⋆ – IP3

FAIR STATION [24] aimed to design new high-capacity rail station improving users flows, and also providing safety and security. The emphasis has been made on various categories of Persons with Reduced Mobility (PRMs), proposing a new (TRL3) structural design for the Platform Train Interface (PTI) to facilitate independent boarding and alighting of PRMs.

[1] Technology Readiness Level

It also considered a Train Door Access system to detect platform position and couple door access with the gap filler and step compensation system (e.g. a lifting mechanism that allows PRMs to access the train). Beside such structural proposal, for which a 3D virtual model has been developed, the project also focused on a passenger crowd movement analysis tool that incorporates both PRMs and baggage handling. Then, as a final result, the purpose was to integrate the two proposals above also including a security risk assessment and crowd emergency evacuation flow in response to natural accidents or man-induced crimes. Within the FAIR STATION project, the AI has been used as one of the enabling technologies for the Crowd Management Analysis (and videosurveillance), in the form of Machine Learning (specifically Deep Learning) and Computer Vision approaches (e.g. Object Tracking).

INtelligent solutions 2ward the Development of Railway Energy and Asset Management Systems in Europe (IN2DREAMS)* - IP3

In2Dreams [25] has been designed to answer the challenges related to the technology development and demonstrator implementation in order to realise:

1. a non-intrusive Smart Metering sensor network at Railway System level;
2. an open system and interface for data collection, aggregation and analysis in an open source Operational Data Management Platform;
3. a set of User Applications design and specifications to exploit the energy analysis process, as well as other possible improvements such as preventive maintenance.

To address those challenges, IN2DREAMS is structured in two work streams:

- Management of Energy-related Data: remove the current limitations of Railway Energy Management Systems (REMS), allowing them to flexibly adapt to different energy and asset management application requirements;
- Management of Asset-related Data: define IT solutions and methodologies for improving business security, economic sustainability and decision support in the field of data processing for asset management.

AI in this project is used as a tool for data-based operation in railways infrastructure. Neural network-based techniques are used to estimate the voltage on trackside from on-board measurements, predict the power and optimise the driving profiles. AI, together with experience-based model, is used to predict train movements such as running time, dwelling time, delay and penalty costs, and restoration time in large-scale railway networks. Additionally, this project investigates the provisioning of 5G services employing AI techniques to minimise the overall energy consumption of the 5G infrastructure.

Intelligent Innovative Smart Maintenance of Assets by integRated Technologies (IN2SMART)* - IP3

IN2SMART [3] aims to contribute to the overall concept for Intelligent Asset Management based on the following three main interlinked layers: measuring and Monitoring systems

to collect data from the field related to the railway assets status; data management, data mining and data analysis procedures to process data from the field and from other sources; and degradation models and decision-making tools to support maintenance strategies and execution.
In this project, there is much prediction and detection work using AI techniques:

(i) Signalling: the Track Circuits (TCs) false occupancy can be predicted by using TCs status monitoring data, system and maintenance logs and weather. And the switch condition degradation can be modelled by using anomaly detection.

(ii) Earthworks: the events and degradation phenomena of earthwork assets can be modelled by the historical inspection reports along with the weather data to predict the changes of earthwork status.

(iii) Track geometry: the degradation and status of track are predicted by using the track register information and train loadings.

This AI-related work monitors and predicts the anomalies and statues in the rail system including the assets and environment to support the decision-making of maintenance and to assessing the effectiveness of the maintenance interventions performed on these assets.

Intelligent Innovative Smart Maintenance of Assets by integRated Technologies 2 (IN2SMART2) - IP3

IN2SMART 2 [26] represents the second phase of the “Intelligent Innovative Smart Maintenance of Assets by integRated Technologies” (IN2SMART) projects already described in the section above. IN2SMART 2 is an ongoing project and no deliverable, in our understanding, is already available. Nevertheless, it is worth mentioning since it will leverage on the outcomes obtained within the IN2SMART project to implement i) a Dynamic Railway Information Management System, ii) a Railway Integrated Measuring and Monitoring System, and iii) Intelligent Asset Management Strategies. It will also provide on field applications.

Innovative Solutions in Future Stations, Energy Metering and Power Supply (IN2STEMPO) - IP3

IN2STEMPO [27] aims to reduce lifecycle costs, improve reliability and punctuality, whilst increasing capacity, enhancing interoperability and improving the customer experience by:

1. developing a smart railway power grid, in an interconnected and communicated system;
2. achieving a fine mapping of energy flows within the entire railway system, forming the basis of later energy management strategy;
3. improving the customer experience at Railway Stations.

This project plans to apply the neural network-based AI techniques to cope with the data from an Operational Data Management System. This technique is used to measure the energy, environmental and physical parameters, e.g. train power, temperature, acceleration and speed collected from the devices equipped in railway substations, depots and maintenance halls. Besides for the inspection on rail, this project also plans to use AI techniques, such as Populate Digital Twin and Computer Vision, to manage the crown in stations.

Research into enhanced track and switch and crossing system 2 (IN2TRACK2) - IP3

IN2TRACK2 [28] is an ongoing project that aims to extend researches carried out within the IN2TRACK [29] project ('Research into enhanced track and switch and crossing system"). In our understanding, no deliverable/outcomes are already available and analysable for IN2TRACK2. IN2TRACK aimed to pave the way for resilient, consistent, cost-efficient, and high capacity European rail network investigating three railways' technical areas:

- Switch and Crossings. First, core issues and their causes have been identified for switches and crossings. Then, new reliable and available subsystems were presented, which include sensors for an enhanced monitoring phase. Some of them also have self-adjusting and noise & vibration reduction capabilities.
- Tracks. Innovative solutions have been explored in terms of products, processes and procedures aiming to improve tack structures (e.g. new slab track system, sub-ballast layer, etc.). Moreover, the enhancing of inspection and maintenance tasks has been faced. Existing predictive methods for rail and track deterioration have been analysed and their limitations emphasized then different approaches to address such topic have been presented, also identifying in Machine Learning / Artificial Intelligence Technologies two of the key aspects for future developments.
- Bridges and Tunnels. Improved inspection methods have been outlined within this project. In particular, two different approaches/systems have been presented: i) an autonomous image-based tunnel lining inspection system that also aims to improve worker safety, and ii) optical measurement methods, both ground-based and UAV-based, have been analysed to test the applicability of different image-based technologies to create digital twins and detect damages/changes within bridges, also basing on the BIM environment. As for the previous point, also in this direction, Machine Learning has been identified as a key technology to evaluate changes in crack patterns, displacements, and so on, basing on optical measurements. However, this applicability study will be carried out within the second phase of the programme (IN2TRACK 2). Beyond this, new techniques were emphasized to reinforce or replace bridge elements.

Multi-scale Observation and Monitoring of railway Infrastructure Threats (MOMIT)* - IP3

The MOMIT [30] project aimed to validate remote sensing technologies based on UAV and Satellite based solutions for railways infrastructure monitoring. Actually, as the first step, a new Remotely Piloted Aircraft System (RPAS) prototype has been proposed. The aim was to develop monitoring solutions easily integrable within the current monitoring and management workflow, which also reduce costs, time, and increment safety in respect to traditional manual inspection routines. To evaluate the potentials of the remote monitoring technologies, six application cases were considered: i) Ground Movements; ii) Hydraulic Activities (e.g. floods), iii) Natural Hazards (e.g. vegetation growth); iv) Electrical System; v) Civil Engineering Structures; and, vi) Anomalies and Illicit Activities (Safety Monitoring). All these monitoring scenarios have been considered along with the tracks or nearby railways infrastructures, and have been faced through the analysis of RPAS and/or Satellite data. Different algorithms and tools have been developed to, for instance, identify and classify Active Deformation Areas (ADA), extract planar discontinuity sets of rocky areas, and so on, leveraging on various techniques among which also clustering algorithms (e.g. DBSCAN).

Advanced Travel Companion and Tracking Services (ATTRAkTIVE)* - IP4

ATTRAkTIVE [31] expands and further develops solutions of the prior lighthouse project IT2Rail. It aims to improve the attractiveness of rail transport by offering more intuitive and engaging travel experiences to customers while shielding them from the complexity and heterogeneity of services for door to door intermodal journey. This project has two major components: Trip Tracker and Trip Companion. Trip Tracker enables:

(i) collecting planned and real time data for all modes including personal transport which is essential for all following up calculations and assistance;
(ii) prognosticating and predicting travel situations to find solutions for possible interruptions in advance;
(iii) generating events that can be displayed on smart devices that will form the base for subsequent processing;
(iv) analysing personal transport events and events which impact the itinerary of the traveller;
(v) orchestrating and distributing tracking services; and
(vi) providing alternatives if necessary.

Trip Companion enables:

(i) creating and editing novel forms of location based experiences for travellers cost-effectively;
(ii) door-to-door experiencing of the created experiences;
(iii) indoor/outdoor positioning of travellers;
(iv) navigation assistance, especially at interchanges; and
(v) homogenized interface shielding the travellers from the complexity and heterogeneity of inter-modal services.

AI is a prediction tool for the trip information, such as delays, cancellations and route changes, and prognosis unexpected events.

My TRAvel Companion (My-TRAC)* - IP4

My-TRAC project [2] develops a transport service platform for offer the information and guidance to passengers and operators. This system analyses passengers' travelling data, captures passengers' affect and then simulates their choice behaviour. The main objects of My-TRAC are: i) developing a platform improving traveller's experience and the quality of transport services; ii) understanding passenger behaviour and state of mind in order to make suggestions that fit their profile and preferences; iii) obtaining data from the operator and exogenous sources and tailoring the recommendations of road choice with user strategic and dynamic preferences; iv) recommending the market services during the trip based on behavioural patterns and travellers with similar itineraries; v) developing an interface for users with viable access to the information and vi) developing an interface for operators to retrieve real-time and historic data concerning behavioural analytics of passengers.

It uses AI to predict the load of public transport (crowdedness) and the delay, to recognise the pattern of travelling behaviour and to cluster the travellers.

Real time information applications and energy efficient solutions for rail freight (FR8HUB)⋆ – IP5

FR8HUB [32] is a wide-spectrum project that aimed to increase capacity, operation reliability, and energy efficiency, as well as to reduce noise, costs, and emissions in freight trains. It provided solutions for the Condition Based Monitoring of train bogies introducing new kind of low-cost sensors for physical parameters monitoring and ad-hoc algorithms for the estimations of the health status; unfortunately, we were not able to analyse WP2 deliverables (the work package related to the CBM) to check which techniques have been used. Also, the hybridisation of legacy shunting fleets has been investigated to improve the ecological footprint and lifecycle cost of Diesel shunters, as well as the energy saving; in this direction, the development and construction of the Locomotive BR 294 "HELMS" has been carried out. Moreover, as the last contribution, the task of optimize and manage yards and rail networks was also tackled. Actually, it considered the results achieved within the Optimised Real-time Yard and Network Management (OPTIYARD) project [33], another IP5 S2R project, that aimed to design optimized processes for managing marshalling yards and terminals to guarantee operational efficiency. Marshalling yards are complex and their optimal management requires real-time information exchange between yards and the rail network ecosystem to perform yard operations rescheduling at short notice basing on network's perturbations; indeed, within the FR8HUB a Data-Exchange Platform has been defined to improve data exchange and traffic information between infrastructure managers and freight transport stakeholders. Within the OPTIYARD project, a separate module based on an ad-hoc timetable planning algorithm has been implemented to identify probable conflicts and reschedule trains by switching or postponing them; the aim is to minimize the deviation from the original timetable. On the other hand, within the FR8HUB project, the focus is moved on the estimation of delays of all departing trains in a pre-defined timespan; clearly, the departure time is defined from the duration of the yard operation, then the prediction model has to take them into account. However, FR8HUB also developed a hybrid simulation-optimization model to reschedule trains in short term and it analysed a couple of heuristic algorithms

(First Eligible Path, and Maximum Bottleneck Path) to achieve an optimal insertion of a train in the scheduled timetable. The last outcome, always related to the yard-network optimization task, has been the introduction of the Intelligent Video Gate (IVG) system. Installed on the exit/entrance track of the terminal, it relies on cameras, Radio-frequency Identification (RFID) antennas and wheel sensors to collect data concerning freight trains and then:

- through Optical Character Recognition (OCR) engines, it identifies the Intermodal Load Unit (ILU) code, i.e. the identification code of the container, and the UIC (Union Internationale des Chemins de fer - International Union of Railways) code that identifies the type of vehicle and the administration to which the vehicle belongs;
- through Computer Vision approaches, it first recognizes the Placard cutting it out from the entire image (extraction of the Region of Interest - RoI), and then classify the dangerousness of the cargo according to its placard.

Development of Functional Requirements for Sustainable and Attractive European Rail Freight (FR8RAIL)⋆ – IP5

FR8RAIL [34] represents the first phase of a macro-project which is subdivided into three IP5 projects:

- Development of Functional Requirements for Sustainable and Attractive European Rail Freight (FR8RAIL);
- Digitalization and Automation of Freight Rail (FR8RAIL II) [35];
- Smart data-based assets and efficient rail freight operation (FR8RAIL III) [36].

The FR8RAIL purpose is to develop functional requirements for sustainable and attractive European railways. Such an aim has been achieved through various investigation in six different fields:

- Business Analytic, KPIs, top level requirements: identification and analysis of market segments and requirements to reach higher competitiveness for societal needs, as well as the development and refinement of freight rail KPIs in collaboration with IMPACT 2.
- Condition Based and Predictive Maintenance: development of a condition based and predictive maintenance strategy, determining the relevant variables and relative thresholds for condition monitoring of locomotives' and wagons' components (e.g. suspensions, wheels, gearboxes, axles, etc.). In this direction, a state-of-the-art analysis has been performed concerning the condition based monitoring and predictive maintenance methodologies for bogie components already applied which also include some Machine Learning techniques. Unfortunately, we were not able to analyse deliverable 2.3 and 2.4, however, in D2.5 [37] the online monitoring of oil in diesel locomotive has been addressed, identifying, beyond sensors and variables' thresholds, in the deep learning one of the possible techniques for the identification of the patterns which assess the actual component conditions.
- Telematics & Electrification: development of telematics technologies (hardware, software, and algorithms) for different applications such as condition based and predictive maintenance, traffic management, etc. Beyond the validation of a freight train position

solution, based on the development of a Wagon On-Board Unit (wOBU) and a Locomotive On-Board Unit (LOBU), also technologies for wagon monitoring systems have been investigated. Indeed, in this direction, almost all the Machine Learning techniques have been identified as possible data-driven approaches for CBM and have been tested considering the wheel profile condition monitoring scenario (tests have been performed through simulations).

- Running Gear, Core and Extended Market Wagon: the creation of a framework and definition of functional requirements for the development of a new freight running gear, core and Extended Market Wagons in order to increase volume saving both costs and environment.
- Automatic Coupling: definition of technical requirements and a migration strategy for automatic coupler designing and integration.
- High level System Architecture and Integration: performing of alignment activities to ensure an overall picture and enable future integration based on the results/systems obtained in the above areas.

Both FR8RAIL II [35] & FR8RAIL III [36] are still ongoing and aim to further investigate directions already considered in FR8RAIL, developing new relevant technologies to increase the competitiveness of rail freight transport.

FR8RAIL II is facing different challenges, among which: i) New wagon design and automatic coupling solution in order to increase the competitiveness of freight transport by reducing costs, increasing capacity and safety, and enabling CBM features; ii) Enable Condition Based Maintenance and Cargo Monitoring System (CMS) through the implementation of intelligent wagons based on Telematics and Electrification (as seen for FR8RAIL). Beyond the development of the wagon On-Board Unit, the purpose is also to expand the scope defined in FR8RAIL for what concern the CBM. Indeed, in this direction, innovative sensors are being developed for the locomotive bogies that will allow reducing maintenance costs; iii) Develop new improved methods for annual and short-term timetable planning and real-time network management in order to increase capacity and punctuality for both passengers and freight trains and also optimize operational planning at yards/terminals. The aim is also to develop a demonstration to showcase innovation; iv) Define requirements, system concept and demonstration for a Driver Advisory System connected to the Traffic Management System. The concept is to rely on real-time data to evaluate the speed profiles basing on the traffic real conditions to increase service efficiency and energy savings; v) Finalize the development of the last mile propulsion system proposed in FFL4E [38] and FR8HUB, as well as the development of the Distribution Power System investigated within FFL4E.

At the same time, FR8RAIL III is investigating six different areas: i) Condition Based Maintenance, scaling the outcomes and results obtained in FR8RAIL I, aiming, among the others, to improve existing algorithms and patterns increasing the positive rate; ii) Real-Time Network Management, to reduce the gap between timetable planning and operational traffic; iii) Intelligent Video Gate, aiming to optimize a fully operational terminal and the data management to improve the detection of incoming and outgoing assets (also leveraging on software applications for object recognition, data processing, etc.); iv) Extended Market Wagon, aiming to create the final basis for realising the prototype of the EMW leveraging on the work performed within FR8RAIL I and II concerning the market segments and requirements to be served by the new wagon; v) Telematics and Electrification, enabling intelligent wagons for new digitalization and automation functions (e.g. CBM, Automatic Train Configuration, etc.); and, vi) Freight Loco of the Future, to improve the propulsion system of locomotives lowering the life cycle cost (LCC) and the total cost of ownership (TCO) of the traction chain.

INNOvative monitoring and predictive maintenance solutions on lightweight WAGon (INNOWAG)* - IP5

The aim of the INNOWAG project [39] is to develop intelligent cargo monitoring and predictive maintenance solutions integrated on a novel concept of lightweight wagon, which would respond to major challenges in rail freight competitiveness, in relation to the increase of transport capacity, logistic capability and an improved RAMS and lower LCC. It has three main objectives:

1. Cargo condition monitoring: to develop an autonomous self-powered wireless sensor system for cargo tracing and condition monitoring of key parameters for critical types of cargo;
2. Wagon design: to reduce the wagon tare with low mass reduction rates for bogie, buffers and couplers and high mass reduction for braking systems and composite materials implemented on the wagon body;
3. Predictive maintenance: to develop and demonstrate solutions for the predictive maintenance of freight wagons.

The AI is used to detect wheel flat in two specific areas: distinguishing wheel flats with different sizes using axle box accelerations and detecting wheel flats using carbody accelerations by various machine learning algorithms. However, it points out that the occurrence of wheel flat cannot be predicted but the early detection enables an optimised planning and bunding of maintenance work.

Locomotive bOgie Condition mAinTEnance (LOCATE) - IP5

LOCATE [40] aims to replace the traditional maintenance approaches (e.g. preventive, scheduled, etc.) with a predictive one concerning the bogie's mechanical parts. It will then develop condition-based maintenance strategies and program in order to increase availability and reliability of bogies' components also reducing their maintenance costs through, among the others, the application of reliability-based sensors and the development of intelligent and integrated tools for the localization of faulty components, the development of digital twins, and so on.

Smart Automation of Rail Transport (SMART)* - IP5

The main goal of SMART [1] is to increase the quality of rail freight, as well as its effectiveness and capacity, through the contribution to automation of railway cargo haul at European railways. To achieve the main goal, SMART will deliver the following measurable objectives:

1. complete, safe and reliable prototype solution for obstacle detection and initiation of long-distance forward-looking braking;
2. short distance wagon recognition for shunting onto buffers;
3. development of a real-time marshalling yard management system integrated into IT platform available at the market.

Since there are cameras equipped at ahead of trains, images in front of trains are the useful input dataset for the obstacle detection. Using the image process and recognition in AI techniques enables the system to find out the track in images and the obstacle on the track. Besides, AI techniques are used to learn the error in distance calculation between laser scanner and vision sensor to improve the accuracy of distance detection for only camera-equipped trains. As for the real-time marshalling yard management system, AI will be train to give the optimal or near-optimal solution of marshalling operations.

Advanced integrated obstacle and track intrusion detection system for smart automation of rail transport (SMART2) - IP5

The SMART2 [41] project is leveraging on results obtained in SMART. It will focus on developing long-range all-weather obstacle detection (OD) and track intrusion detection (TID) system. The purpose is also to develop two more systems, an advanced SMART2 trackside (TS) and an airborne OD&TID system, and integrate all three of them into a holistic OD&TDI system via interfaces to central Decision Support System (DSS). The DSS will integrate the information coming from the subsystems (e.g. obstacles on the tracks) to suggest possible actions for the train control.

Translation for breaking language barriers in the railway field (TRANSLATE4RAIL) - IPX

TRANSLATE4RAIL [42] is an ongoing IPX project that aims to offer drivers the possibility to interact easily with the infrastructure manager traffic control in a country where they do not understand/speak the local language. To this aim, an IT Tool will be developed within this project which will allow divers and traffic controller to communicate speaking in their native language. The tool will be provided with some pre-defined messages, encompassing various kinds of situations and scenarios; when a user pronounces a sentence in its language, the system will analyse it through a voice recognition mechanism and translate it, then the tool will utter the message in the language of the other party. The project is still in an initial phase, and, in our understanding, no outcome has already been released. Nevertheless, it is quite probable that the applicability of Speech Recognition/Natural Language Processing approaches will be analysed to address this topic.

Global Safety Management Framework for RAIL Operations (GoSAFE RAIL)* - CCA

The GoSAFE RAIL project [43] focuses on the asset safety in the rail sector, and bring A.I. and network microsimulation modelling technologies, together with infrastructure manager and railway undertakings, to develop a decision support tool for supporting infrastructure safety. This project works on the four objects: i) developing the risk assessment methodology by detecting the objects on tracks with monitoring and operational data; ii) planning with micro-simulation models and testing the punctuality prognosis; iii) developing a Decision Support Tool for maintenance and operational plans; and iv) demonstrating the previous works on practical cases.

AI in this project is used to in the following works:

- Detecting the obstacle on track: analysing the image/video from the camera using image recognition and detecting the rockfall from the signal of sensors using pattern recognition.
- Predicting the landslides: incorporating weather conditions with the soil suction data to predict the slope stability. The method used in this work is a three-hidden-layer neural network.
- Classifying fine-grained soils: ANNs are used to predict the fines content, liquid limit and corresponding plasticity index. Using these three outputs, a cohesive soil can be fully classified using either the USCS or the ESCS systems.
- Testing the punctuality prognosis: predicting the total knock-on delays of the whole network rather than the individual delays. RNN is the method to deal with the sequential data in this work.

Indicator Monitoring for a new railway PAradigm in seamlessly integrated Cross modal Transport chains – Phase 2 (IMPACT-2)* – CCA

IMPACT-2 [44] is the second phase of the “Indicator Monitoring for a new railway PAradigm in seamlessly integrated Cross modal Transport chains” (IMPACT-1) [45] project. Both of them are classified as CCA activities. Within the first phase, IMPACT-1, a socio-economic impact analysis for different kinds of trains has been performed. The purposes were to i) evaluate the effects of the new technologies, ii) increase the attractiveness, competitiveness and sustainability of the rail system introducing relevant target and needs, and iii) define System Platform Demonstrators (SPD) and Key Performances Indicators (KPIs) to outline future application use cases and enable the monitoring/assessment of the S2R overall aims (S2R Master Plan). IMPACT-2 is adding more values to the outcomes achieved within the fist phase. Indeed, some SPD use cases have been identified concerning high-speed, regional, and urban electric multiple unit (EMU) trains and rail freight. Within this project, huge importance is given to the Smart Maintenance concepts definition [46], Condition Based Monitoring applications [47] and Data management for Smart Maintenance (it seems to be an ongoing task). The project covers and analyses all the steps for a correct CBM, from data collection (monitoring) and formatting to data analysis and pattern recognition.

Four scenarios have been identified concerning the monitoring "vision": i) vehicle monitor itself, ii) vehicle monitor infrastructure, iii) infrastructure monitor itself, and iv) infrastructure monitor vehicles. Then, the applicability of different kinds of techniques has been already investigated for CBM data analysis and pattern recognition considering the SPD use cases identified involving Linear Regression Models, Recurrent Neural Networks, etc.

Innovative Intelligent Rail (IN2RAIL)* - CCA

In2Rail [48] project is to set the foundations for a resilient, consistent, cost-efficient, high capacity European network by delivering important building blocks that unlock the innovation potential. To achieve the above, a holistic approach will be applied, which contains:

1. Smart Infrastructure: intelligent reliable infrastructure, better system resilience and a reduced need for maintenance, and overall reduction in carbon emissions, noise and vibration, and improved levels of sustainability.
2. Intelligent Mobility Management: a standardised approach to information management and dispatching system enabling an integrated Traffic Management System (TMS); an Information and Communication Technology (ICT) environment supporting all transport operational systems with standardised interfaces and with a plug and play framework for TMS applications; and an advanced asset information system with the ability to 'nowcast' and forecast network asset statuses with the associated uncertainties from heterogeneous data sources.
3. Rail Power Supply and Energy Management: the design of a future AC Rail Power Supply System with minimised energy losses and optimised loads; and the implementation of an efficient energy management system allowing understanding of energy flows within a railway system, a reduction of the energy consumption and cost, optimised asset management and enabling better use of the railway capacity.

AI is employed in this project for the following domains:

1. Delay prediction: The extreme learning machines algorithms is applied in the historical train movement data for predicting train delays. Additionally, the prediction model is improved be including weather information.
2. Asset maintenance: Data related to track geometry, train, load, and weather condition is used to forecast the possible asset malfunctions, e.g. the remaining useful life and confidence level, and therefore to suggest the possible maintenance actions such as speed reduction, closing and track geometry correction. Similarly, the risk of derailment can also be predicted leveraging track, vehicle, and wind data; also, the monitoring data can be used to predict the probability of failure of switches and crossings.
3. Maintenance planning: The project investigates the past maintenance reports, as well as the weather conditions, to estimate the time to restoration for future and urgent maintenance.

Modelling and strategies for the assessment and OPtimisation of Energy USage aspects of rail innovation (OPEUS)* - CCA

The main purposes of the OPEUS project [49] were to evaluate, improve and optimize the energy consumption of rail vehicles, by evaluating and measuring the energy used by the relevant components of the rail system. It aimed to i) define an energy simulation model, ii) define the energy use requirements, and iii) evaluate the energy usage outlook and outline optimization strategies. The main outcome has been an energy simulation tool, developed for electric driven vehicles, to evaluate and calculate the energy consumption of the various vehicle's components. Hence, relying on such tool, possible energy management approaches have been provided in order to implement the optimal speed profile in a form of Driver Assistant System (DAS) or Automatic Train Operation (ATO) aiming to reduce energy consumption. The optimization approaches evaluated to implement this "energy optimal driving strategy" are based on Evolutionary Algorithms that are i) Particle Swarm Optimization (PSO), ii) Grey Wolf Optimizer (GWO), iii) Covariance Matrix Adaption – Evolution Strategy (CMA-ES), and iv) Firefly Algorithm (FFA). Also, a very brief introduction for Generic Algorithm was outlined within the deliverable 4.2 [50] as possible future implementations.

Smart Maintenance and the Rail Traveller Experience (SMaRTE)* – CCA

The SMaRTE [51] project focused on facing two of the S2R macro-objectives which are i) the improving in availability and reliability of railway systems, and ii) the increase in attractiveness. Indeed, the purpose of the project was two-fold: on one side, it addressed the Smart Maintenance applicability, on the other, it analysed some Human Factors providing recommendations for more attractive railways.

The Smart Maintenance aspect of the project was intended to improve the current management systems integrating tools, algorithms, and methods for an improved CBM. Hence, it aimed to reduce the systems' life cycle costs, as well as to improve their availability and reliability, by developing predictive tools and a decision-making support system for optimal maintenance scheduling. After the first investigation on CBM approaches in other transportation sectors (mostly in avionics, in our understanding), a conceptual CBM system was introduced: once data have been collected, they were arranged according to a novel ontology to ensure interoperability, then, prognosis techniques were evaluated to predict the occurrence of system failures. In this direction, Machine Learning techniques (unsupervised methods, regression, and artificial neural networks) were adopted to diagnostic data for a traction and braking system; while, other approaches as statistical models, survival models and Markov Decision Process (MDP), were adopted to support optimal maintenance decision-making concerning wheelset conditions/degradation. It is worth noting that the first facet (traction and braking system) has been investigated in collaboration with the IMPACT-2 project; moreover, as reported in Cordis [52], the most important findings within SMaRTE will be also exploited in the LOCATE project [40] for application in locomotive bogies.

On the other hand, also the users' behaviour has been investigated to analyse psychological factors that deter potential railway users. Beyond the ticket cost, it has been identified that also factors as car parking, cleanliness, frequency of peak services, etc., are reasons for users' dissatisfaction. In this direction, a new Smart Journey Vision has been proposed to also emphasize recommendations for improving users' experience, and then attractiveness.

Table 4.2 summarises the Shift2Rail projects reviewed in this section.

4.2.2.1. New 2020 S2R starting projects

The last Horizon 2020 Shift2Rail JU call proposals 2020 (H2020-S2RJU-2020) involved the initiation of a new 18 projects within the S2R programme. Among these, we have identified the following 8 projects which, as described, will involve or could involve AI to address some issues related to the railway sector:

- **RECET4rail Reliable Energy and Cost Efficient Traction system for Railway (RECET4RAIL) [53] - IP1**. Among the other purposes, RECET4RAIL aims to develop smart maintenance approaches enables by predictive analytics to reduce the system life cycle costs and increase its availability. The project will involve four technical work packages, one of which will focus on "Big Data, Artificial Intelligence (AI) applied to Traction systems smart and predictive maintenance".
- **Development of prescriptive AnalYtics baseD on aRtificial intElligence for iAMS (DAYDREAMS) [54] - IP3**. DAYDREAMS is aiming to integrate and use data and the artificial/human trustworthy intelligence, together with context-driven Human-Machine Interfaces (HMI), to move towards prescriptive Intelligent Asset Management Systems (IAMS). In addition, it also aims to improve decision-making processes alongside the railway sector (e.g., on Traffic Management System, Energy, etc.) by developing multi-objective decision optimization approaches. Lastly, another noteworthy aspect is related to the development of HMIs reinforcing the role of person-in-the-loop and allowing a properly access and visualization of predictions/metrics and models, also aiming to "assess why and how the model predicts something (opening the black-box)".
- **The Next Generation of Railway Transition Zones (IN2ZONE) [55] - IP3**. This project will combine the latest transition solutions with technological advances from other sectors and recent advances in material science to design next generation transition zone solutions which will provide a step-change in track support conditions reducing maintenance interventions. One of the aims is to design new types of sleepers capable of self-correcting vertical track geometry irregularities. In addition, a new monitoring specification for transition zones will be developed by merging datasets from multiple track and vehicle sensors and leveraging edge computing and AI to improve maintenance activities.
- **Smart Tools for Railway work safEty and performAnce iMprovement (STREAM) [56] - IP3**. Aims to develop autonomous and intelligent robots to improve operations and workers safety at worksites.
- **Use-centric rail freight innovation for Single European Railway Area (FR8RAIL IV) [57] - IP5**. Starting form the studies conducted in FR8RAIL III, this project aims to further investigate the topic of freight wagons CBM leveraging AI and the study of intelligent wagons based on Telematics and Electrification (another focus of this and

previous projects).

- **Technologies for the AUtonomous Rail Operation (TAURO) [58] - IPX**. TAURO is aiming to identify, analyse and propose suitable technologies for the future European automated and autonomous rail transport. To achieve this purpose, in our understanding, it will investigate AI to achieve what is defined as "artificial sense" in order to improve the environment perception both in outdoor (e.g. obstacles on tracks, light signal status, etc.) and indoor scenarios. Furthermore, it intends to investigate towards the development of specification and functional architecture for remote driving and control, and towards the identification of technologies that support migration to ATO over ETCS and therefore increase the degree of automation (up to GoA3/4) also in those areas where such systems/standards are not yet implemented. Lastly, it also aims to develop an enhanced Train Control and Monitoring System (TCMS) for autonomous trains in order to, among the others, enhance safety and integrate the train onboard diagnostic with equivalent Digital Twins to improve CBM.
- **System Architecture and Conceptual Data Model For Railway, Common Data Dictionary And Global System Modelling Specifications (LINX4RAIL2) [59] - IPX**. Starting form the results obtained in LINX4RAIL [60], this project will extend the work moving towards a new integrate railway system approach. Moreover, it will also leverage the developments of the TAURO project related to decision making, supervision, remote control and automation all based on sensors and artificial intelligence.
- **Completion of activities for Adaptable Communication, Moving Block, Fail safe Train Localisation (including satellite), Zero on site Testing, Formal Methods and Cyber Security[61] -IP2**. This project aims to the highest readiness level (TRL) taking the results of previous X2Rail-1 and X2Rail-4 projects (including also X2Rail-2 and X2Rail-3, with no AI-related developments). It is expected to include Integrated Technology Demonstrators (ITD) to bring together several S2R IP2 Technology Demonstrators.

Table 4.2: Shift2Rail projects

Project Title	Duration (s: start, e: end)	IP	Railway Subdomain	AI Techniques/ Applications	Contribution to AI Integration	Website
RUN2RAIL*	s: 01/09/2017 e: 30/09/2019	IP1	Maintenance and inspection	Machine Learning	Ensemble of technologies and methodologies for more attractive and reliable trains.	[14]
CYRAIL*	s: 01/10/2016 e: 30/09/2018	IP2	Safety and security	Machine Learning	Protection Profile development	[15]
X2RAIL-1*	s: 01/09/2016 e: 31/12/2020	IP2	Maintenance and Inspection	Machine Learning	Smart Wayside Object Controllers	[16]
X2RAIL-4	s: 01/12/2019 e: 28/02/2023	IP2	Autonomous Driving and Control Maintenance and Inspection Traffic Planning and Management	in progress	in progress	[19]
ASSETS4RAIL*	s: 01/12/2018 e: 31/05/2021	IP3	Maintenance and inspection	Machine Learning	A BIM based infrastructure monitoring data handling [21, 22]	[20]
FAIR STATION*	s: 01/09/2017 e: 31/12/2019	IP3	Safety and security Passenger mobility	Computer Vision	Crowd Flow Management system; Platform Train Interface structural design	[24]
IN2DREAMS*	s: 01/09/2017 e: 31/10/2019	IP3	Autonomous driving and control Maintenance and inspection Traffic planning and management	Machine Learning Computer Vision	Platform for energy, environmental and kinematics monitoring; AI models for fault detection and optimal driving styles	[25]
IN2SMART*	s: 01/09/2016 e: 31/10/2019	IP3	Safety and security Maintenance and inspection	Machine Learning	Decision Support System for failure prevention, operational Asset Management, and maintenance interventions planning	[3]
IN2SMART2	s: 01/12/2019 e: 30/11/2022	IP3	Safety and Security Maintenance and Inspection	in progress	in progress	[26]
IN2STEMPO	s: 01/09/2017 e: 31/08/2022	IP3	Passenger mobility Maintenance and inspection	Machine Learning Computer Vision	in progress	[27]
IN2TRACK2	s: 01/11/2018 e: 30/04/2021	IP3	Maintenance and Inspection	Machine Learning in progress	in progress	[28]
MOMIT*	s: 01/09/2017 e: 31/08/2019	IP3	Maintenance and inspection	Machine Learning	RPAS platform; Tools for RPAS and Satellite data analysis	[30]
ATTRANkTIVE*	s: 01/09/2016 e: 31/05/2019	IP4	Traffic planning and management Passenger mobility	Machine Learning	Trip Tracking system; Travel Companion modular platform	[31]

MY-TRAC*	s: 01/09/2017 e: 30/09/2020	IP4	Traffic planning and management Passenger mobility	Machine Learning	My-TRAC Travel Companion (APP)	[2]
FR8HUB*	s: 01/09/2017 e: 31/08/2020	IP5	Maintenance and inspection Traffic planning and management	Computer Vision	Sensors and algorithms for CBM; Intelligent Video Gate	[32]
FR8RAIL*	s: 01/09/2016 e: 31/08/2019	IP5	Maintenance and inspection	Machine Learning	Technologies for Wagon On-Board Unit, positioning algorithm, wheel condition monitoring (CBM)	[34]
FR8RAIL II	s: 01/05/2018 e: 31/07/2021	IP5	Maintenance and Inspection Traffic Planning and Management	in progress	in progress	[35]
FR8RAIL III	s: 01/09/2019 e: 31/08/2022	IP5	Maintenance and Inspection Traffic Planning and Management	in progress	in progress	[36]
INNOWAG*	s: 01/11/2016 e: 30/04/2019	IP5	Maintenance and inspection	Machine Learning	Data processing methods for bogie components' structural health monitoring; Wizard tool for maintenance policy optimisation	[39]
LOCATE	s: 01/11/2019 e: 31/10/2021	IP5	Maintenance and Inspection	in progress	in progress	[40]
SMART*	s: 01/10/2016 e: 30/09/2019	IP5	Autonomous Driving and Control Safety and security Maintenance and inspection Traffic planning and management	Machine Learning Computer Vision	Autonomous Obstacle Detection System and Real-time Marshalling Yard Management System developing	[1]
SMART2	s: 01/12/2019 e: 31/11/2022	IP5	Autonomous Driving and Control Safety and Security	in progress	in progress	[41]
GoSAFE RAIL*	s: 01/10/2016 e: 30/09/2019	CCA	Autonomous Driving and Control Safety and security Maintenance and inspection Traffic planning and management	Machine Learning Computer Vision	GoSAFE RAIL & Safety Management Framework	[43]
IMPACT-2*	s: 01/09/2017 e: 31/08/2022	CCA	Maintenance and inspection	Machine Learning	Smart Maintenance concepts' definition CBM applications analyses	[44]
IN2RAIL*	s: 01/05/2015 e: 30/04/2019	CCA	Maintenance and inspection Traffic planning and management	Machine Learning	Delay prediction system; Integrated approach to TMS; Optimised real-time traffic management; Asset monitoring tools for predictive maintenance; Nowcasting and forecasting analytics algorithms	[48]

OPEUS*	s: 01/11/2016 e: 31/10/2019	CCA	Autonomous Driving and Control	Evolutionary Computing	Energy simulation tool; Optimization strategies for energy consumption reduction	[49]
SMaRTE*	s: 01/09/2017 e: 31/10/2019	CCA	Maintenance and inspection	Machine Learning	Tools for optimal CBM; Smart Journey Vision	[51]
TRANSLATE4RAIL	s: 01/12/2019 e: 31/11/2021	IPX	Safety and Security Traffic Planning and Management	in progress	in progress	[42]

4.2.3. Non-Shift2Rail projects

Drones4Safety: Inspection Drones for Ensuring Safety in Transport Infrastructures

The aim of the Drones4Safety project [62] is to develop a system of autonomous, self-charging, and collaborative drones that can inspect a big portion of transportation infrastructures in a continuous operation. The collaborative and centralized drone system is used to inspect different sides of the infrastructures by recognising the infrastructure components and discovering eventual faults on assets. On the reliability of the system, Drones4Safety harvests energy from overhead power of rail lines in the proximity of the infrastructure for a longer time; produces a safe operational system resisting harsh electromagnetic environments and circumventing the effects of high-voltage/high-current signals; and monitors and controls remotely the state and location of the drone.
Drones4Safety is an ongoing project which end date has been scheduled in 2023, hence, in our understanding, the particular AI aspects involved have not already been emphasized. Nevertheless, the basic idea is to develop and improve AI algorithms to optimize inspection results onboard of the drones, also leveraging satellite and open maps. The same considerations apply to the outcomes, the overall idea is to present to the transportation operators software services and hardware drone systems.

TRAINSFARE: Smart Tool to Protect Public Transport Revenues, Assets, Passengers and Mobility

TRAINSFARE project [63] aims to solve the problem of fare evasion in public transports and to increment their safety. Actually, two different, but related, projects have been found on Cordis/Trimis under the TRAINSFARE name. The first one, "TRAINSFARE: Transport System with Artificial Intelligence for Safety and Fare Evasion", was a preliminary project with a duration of only 5 months (01/01/2015-31/05/2015) which purpose was to outline a feasibility plan and a clear market strategy to obtain additional resources to implement the project idea. Indeed, in the phase 2 project ("TRAINSFARE: Smart Tool to Protect Public Transport Revenues, Assets, Passengers and Mobility"), an artificial vision system named DETECTOR 1.0 has been developed. It is an automatic real-time video analytics system able to detect various forms of fare evasion such as jumping or sliding under the gates, forcing the gates, entering from the reverse way, or passing closely behind a paying passenger. The system is configurable to send instant alerts to a mobile application on the smartphones or tablets of key personnel. In this way, they are able to intercept and check the offenders on the spot. Also, this system can be used to detect vandalism, unattended baggage, and so on, to increase safety. Lastly, it has been evaluated and is successfully operating in four of FGC's[2] main stations.

The system uses the video from the cameras installed above the gates and applies a deep learning technique for the Computer Vision.

[2] FGC means "Ferrocarrils de la Generalitat de Catalunya"

Transforming Transport (TT)

Transforming Transport (TT) [64] demonstrated that Big Data solutions are technically and economically viable and able to transform transport processes and services. The key enabling Big Data technologies employed by TT to bring about this transformation are predictive data analytics. Predictive analytics is a significant next step from descriptive analytics. This project is bringing about a major paradigm shift in transport to showcase big data impact in the below areas:

- Highways: describing traffic flows and mobility patterns by unsupervised learning, predicting traffic flows by regression models, and detecting intrusion, extrusions and accidents by classification models.
- Rails: predicting the weakness of switch and crossing elements and the degradation of track profiles to suggest the plan of maintenance and decrease the cost by regression models and predicting the evolution of the slopes by classification models.
- Ports: predicting the earliest departure time of a given train and maintenance of cranes by ensemble learning.
- Airports: identifying passengers' arrival time patterns to airports, extracting movement patterns along the airports, detecting transfer passengers with short transferring time and extracting passengers' behavioural features and patterns regarding airport services by k-means clustering.

This project proposes that the big data is the key to solve most of problems in transport systems and AI, mostly in the form of Machine Learning, is the best way to mine the useful information from the massive data.

Runecast A.I. Knowledge Automation (RAIKA)

Project RAIKA [65, 66] works on the disruption of ICT systems. Typically, technical issues can cause problems within different scenarios (e.g. hospitals, banks, railway management systems, etc.); therefore, the RAIKA projects aims to face such threats advancing and involving a new technology, the Runecast Analyzer [67]. RAIKA automatically shoots the trouble in the system by gathering and parsing information from public knowledge bases, indeed, it combines Artificial Intelligence and human expertise to analyse the configuration and logs on the ICT environment and translate them to machine useable data for identifying the known technical issues and problematic pattern before they cause service interruptions.

Safety of Transport Infrastructure on the TEN-T Network (SAFE-10-T)

Within the SAFE-10-T project [68] a wide-spectrum Global Safety Framework has been presented in order to ensure safety and a longer life-cycle for critical infrastructures (e.g. bridges, tunnels, etc.) across road, rail, and inland waterway modes. The purpose was to develop a framework supporting an online Decision Support Tool (DST) concerning the management of infrastructures along the Trans-European Transport Network (TEN-T network). According to the involving of AI, such DST benefits from a Big Data platform that, beyond the possibility to collect different data sources and process large datasets in real-time, it also supports Machine Learning applications to automatically detect and classify damages within tunnel linings.

ZIMASS: Smart Mobile awareness and safety system for workzone invasion

Current technologies applied to road & railway maintenance safety management do not provide adequate warning due to technological limitations that reduce safety system reliability, such as: inaccurate alarms, reduced signal transmission distance, high investment, or do not meet portability & flexibility needs. ZIMASS [69] is based on Computer Vision and AI to create a virtual security zone, recognise risky situations, and alert workers by connected wearable devices to increase work safety and work quality. In order to improve the reliability and accuracy of the system, the detailed work of ZIMASS aims i) to improve hardware design and engineering; ii) to improve software and development of a cloud management platform; iii) to manufacture ZIMASS pilots and verify its performance; iv) to validate and test the system on road maintenance and railways maintenance, also in low-light conditions.

Predictive Maintenance for railway switches. Smart sensor networks on a machine learning analytics platform (ANDROMEDA)

Moved by the delays and accidents that switches' issues can cause, KONUX has developed the ANDROMEDA system [70], a monitoring system which, relying on data coming from IoT sensors, is able to perform predictive maintenance involving AI. The system, in brief, continuously monitors and analyses the key infrastructure's components to allow a maintenance planning based on the prevision of failures. Data are collected by an autonomous and proprietary IoT device able to operate also under extreme environmental conditions; then, a Machine Learning approach has been implemented to i) predict the health conditions in a regression fashion ii) recommend, basing on its results and other comparison analyses about maintenance activities, the optimal time and type of intervention.

Increasing Capacity 4 Rail networks through enhanced infrastructure and optimised operations (Capacity4Rail)

Capacity4Rail [5] is a wide-spectrum project aiming to face the increasing railway traffic and to make the railway system more competitive by augmenting i) the operational capacity, ii) the availability and the performances of the railway systems, and iii) the reliability and resilience to hazards. Practical overall objectives were to:

- investigate automated, intelligent and fully-integrated system for reliable freight operations;
- reduce the time needed to recover from traffic disruption by optimizing operating strategies, enhancing traffic planning, and so on;

- design resilient, reliable and low-maintenance infrastructure and vehicles, also integrating new technologies for non-intrusive infrastructures' monitoring aiming to reduce costs and impact inspection;
- underline recommendations and guidelines for future researches and investments, and perform systems' assessment through technical demonstrations.

Since the vastness of objectives and addressed issues, such a project has achieved different results. Nevertheless, as we have already mentioned, we are interested in emphasizing the AI application in the railway domain that. Beyond a state-of-the-art introduction to Computer Vision approaches for monitoring and inspection purposes [71], in our understanding, in this project AI has been applied in the form of Machine Learning for train delay prediction [72]. The developed algorithm uses a data-driven multivariate regression model which, relying on historical speed and position data, is able to predict the train's future delay. For further clarifications and other results examination, please refer to [73].

Optimal Networks for Train Integration Management across Europe (ON-TIME)

The ON-TIME project [4] aimed to increase available capacity and decrease delays, both for passengers and freight train, by developing new methods and processes relying on Operations Research approaches and also on Evolutionary Algorithms. It contributed to the rail sector enhancement through different innovations within the area of railway planning and operations management which, according to our taxonomy, fall within the "Traffic planning and management" area. Such innovations have been oriented to:

(i) the development of standardised definition and methods for interoperable processes and tools allowing a facilitate approach to cross-border planning and real-time traffic management;
(ii) the development of improved methods for robust and resilient timetable construction allowing long-term planning, real-time planning, and replanning during disruptions. In such direction, the Perturbation Management Module (PMM) has been presented which allows effective real-time traffic management when perturbations occur. In this scenario, the Differential Evolution Junction Rescheduling Model (DEJRM), developed by The University of Birmingham, has been considered since it is able to perform minor perturbation detection and resolution, also relying on Evolutionary Algorithms;
(iii) the development of approaches for the communication of information to drivers and controllers (to present the right information at the right time), as well as the development of an information architecture to support the communication allowing the exchange of train control data between different actors (operators, undertakings, networks, countries).

We summarise the non-Shift2Rail EU projects surveyed in this sub-section in Table 4.3.

Table 4.3: Other EU projects

Project Title	Duration (s: start, e: end)	Founded Programme	Railway Subdomain	AI Techniques/ Applications	Contribution to AI Integration	Website
ANDROMEDA	s: 01/12/2017 e: 30/11/2019	H2020-EU.3.4 H2020-EU.2.1.1 H2020-EU.2.3.1	Maintenance and Inspection	Machine Learning	ANDROMEDA system	[70]
Capacity4Rail	s: 01/10/2013 e: 30/09/2017	FP7-TRANSPORT	Traffic planning and Management	Machine Learning	Tools and roadmaps for enhanced capacity and performances Demonstrations activities	[5]
Drones4Safety	s: 01/06/2020 e: 31/05/2023	H2020-EU.3.4	Maintenance and inspection	in progress	in progress	[62]
ON-TIME	s: 01/10/2011 e: 30/09/2013	FP7-TRANSPORT	Traffic planning and Management	Evolutionary Computing	Improved methods for timetable construction	[4]
RAIKA	s: 01/10/2019 e: 30/09/2021	H2020-EU.3 H2020-EU.2.3 H2020-EU.2.1	Safety and security	Machine learning	Runecast Analyzer	[65–67]
SAFE-10-T	s: 01/05/2017 e: 01/04/2020	H2020-EU.3.4	Maintenance and Inspection Safety and Security	Machine Learning	Global Safety Framework Decision Support Tool	[68]
TRAINSFARE	s: 01/05/2017 e: 31/10/2020	H2020-EU.3.4 H2020-EU.2.1.1 H2020-EU.2.3.1	Safety and security	Computer vision	DETECTOR 1.0 system	[63]
Transforming Transport	s: 01/01/2017 e: 31/07/2019	H2020-EU.2.1.1	Maintenance and inspection	Machine learning	Decision Support Tool for preventive maintenance and optimised intervention scheduling	[64]
ZIMASS	s: 01/08/2019 e: 30/11/2019	H2020-EU.3 H2020-EU.2.3 H2020-EU.2.1	Safety and security Maintenance and inspection	Computer vision	ZIMASS system	[69]

4.3. Non-European Projects

4.3.1. United States

The U.S. projects we reviewed have been obtained by using the Transportation Research Board's Research in Progress TRB's RiP [11] database. It contains different projects funded by the U.S. Department of Transportation and State Departments of Transportation, but also University transportation research.

Almost all the reviewed projects fall within the UTC Program [74] of the U.S. Department of Transportation. This program administers grants to consortia of colleges and universities in order to promote the advancement of the state of the art in the transport sector and to develop the next generation of professionals.

Furthermore, most of the analysed projects are still ongoing, therefore, in our understanding, no outcomes or final report are already available. The information we retrieve are mainly related to the project's introduction available on the RiP's, TRB's or leader organization's website. However, the research directions are explicitly outlined in most of the cases, and the AI is often considered as a central facet.

Application of machine learning techniques toward time-based change in track condition using an onboard sensor in revenue-service rolling stock [75]

This UTC project aims to implement a machine learning based approach to identify both the location and timing of unexpected changes to the track system. Such a study is also motivated by the wealth of historical track data collected by the onboard systems. A framework based on the Matrix Profile concept is presented to deal with the stochastic nature of the data related to the temporal mismatch between time-series in different inspection cycles. The used approach subdivides the data into track segments, then perform a similarity search between the different time-series, and associates locations with higher dissimilarity to potential defects or changes due to maintenance activities.

Artificial Intelligence for Advance Landslide Warning along Railroad Tracks in Pennsylvania and Delaware [76]

The main purpose of this UTC project is to develop an AI-based model to predict landslide along rail tracks. The team, under the Google AI grant which purpose is to improve the prediction of rainfall-induced landslide hazards, will leverage on deep learning techniques to develop tools able to identify landslide events and improve the prediction accuracy of where and when landslides will occur, also identifying the possible impacted areas.

Technology-Driven Train Inspection (Machine Vision) [77]

Different Railinc's projects, as the "Technology-Driven Train Inspection (Machine Vision)" is, originates in committees sponsored by the Association of American Railroads (AAR). Such projects often fall within different programmes identified by Railinc in coordination with the AAR's Railinc Project Support Working Committee and others AAR Committees. The project under examination falls within the "Asset Health Strategic Initiative (AHSI)" programme which is a multi-year, multi-phase programme with the aim to address efficiency, management, maintenance, and inspection challenges within the rail network. "Technology-Driven Train Inspection (Machine Vision)" will design technology-driven standards for train inspection activities taking into account the current detector technologies. At the same time, the project will also determine standards to confirm the appropriate automation levels and specifications, so as to manage the affected areas in the entire railway industry.

Automated Track Geometry Monitoring System [78]

This UTC project will implement a monitoring system to identify track geometry faults during the normal train operations. Data will be collected by smartphones onboard the train, then, these onboard data will be mapped/transformed into track geometry equivalents by using signal processing and machine learning algorithms. Then, the principal purposes of this study are to: i) develop and test a smartphone system to collect track monitoring data; ii) develop a smartphone application able of autonomously collecting data from hi-rail vehicles, geometry cars, locomotives, and so on; iii) develop algorithms to obtain track geometry measures starting from the data collected by the smartphone system, leveraging on signal processing and machine learning algorithms; and iv) produce dissemination materials as journal and conference papers, as well as a final project report, that details the main findings.

Developing a Safety Management System Including Hazardous Materials for Highway-Rail Grade Crossings in Region VII [79]

Accidents at Highway-Rail Grade Crossings (HRGCs) tend to be more severe compared to other non-HRGC locations, this is why, although an ever-increasing focus on improving their design and engineering practices, HRGCs are still of major concern. In such a direction, this UTC project will focus on the development of a prediction model for collisions at HRGCs and an HRGC Safety Management System (SMS). The idea is to build a comprehensive database encompassing crashes, coordinates, inventory, traffic, and freight volume data. Then, prediction models for HRGC crashes will be investigated leveraging both on traditional deterministic models and stochastic machine learning ones, also including the estimation of the risk related to hazardous materials being released in the surrounding areas after the crash. Lastly, an HRGC SMS prototype will be developed to enhance the safety management of HRGCs by assisting managers in taking proactive actions to reduce risk situations.

Estimation of Origin-Destination Matrix and Identification of User Activities Using Public Transit Smart Card Data [80]

The aim of [80] has been to develop a methodology for estimating the passengers' origin-destination leveraging on data related to bus transit smart cards. Such a purpose has been addressed by considering different tasks: detect the direction information and the boarding cluster information (a Cluster Identification Algorithm has been developed to face this second task); extract boarding stop and alighting stop information; generate an origin-destination matrix and analyse users' travel patterns. More information about developed algorithms are available in the final report [81]; it also presents data analysis results in order to help planners to take optimal decisions.

Laser-based Non-destructive Spike Defect Inspection System [82]

The main purpose addressed within this project is the development of low-cost, non-destructive and contact-free inspection system to identify broken spikes. Usually, it is very difficult to inspect and distinguish broken spikes since cracks are typically located underneath the spike head; then, physical examinations are required which can lead the operational safety. Therefore, this project aims to leverage on laser excitation, acoustic analysis, computer vision, pattern recognition and AI to improve accuracy in spike crack detection and enhance railroad track safety.

This project falls within the Rail Safety IDEA (Innovations Deserving Exploratory Analysis) Program [83] which provides founding to promote innovative ideas/approaches aiming to improve railroad safety or performances.

Passenger Flow Modeling and Simulation in Transit Stations [84]

This is a UTC project whose principal purpose is to model and simulate passenger flows leveraging on computer vision and agent-based simulation technologies in order to potentially reduce congestion considering the infrastructure design and possible upgrade.

Principal components analysis and track quality index: A machine learning approach [85]

This UTC study investigates the potential of Principal Components and Machine Learning techniques to express the Track Quality Indices (TQI) differently and predict track defects; the used approach is well described in [86]. At first, the Principal Component Analysis (PCA) was used to extract the Principal Components from the feature space represented by the different possible track geometry parameters (e.g. alignment, gage, surface/profile, cross level, etc.). Then, a two-phase approach was adopted: in the first phase, an investigation of the different ML techniques was carried out to identify the one that works better with track geometry data; in the second phase, the identified ML model (an SVM, in our understanding) was trained and validated by using true and false positive rates. At first, the model was trained by combining Principal Components and traditional TQIs as defect predictors, however, results show that a three-principal component set of features was better at predicting defects than the combination just mentioned.

Video-Sensor Data Fusion for Enhanced Structural Monitoring [87]

This study falls within the UTC program and aims to investigate the applicability of Computer Vision in asset management. The purpose is to improve and promote integrated asset management for condition assessment of infrastructure (e.g. rail bridges), leveraging both on camera-based monitoring and the current many motoring systems (sensor-based) already adopted, through the implementation of a data fusion algorithm capable learning the associations between video measurements and sensor data.

Vibration-based Longitudinal Rail Stress Estimation Exploiting Contactless Measurement and Machine Learning [88]

This study falls within the Safety Rail IDEA program and focuses on the continuous welded rail (CWR) monitoring. Track irregularities can lead to rail thermal buckling which can affect transit safety. Therefore, the project team developed an innovative technology that, relying on acoustic vibration measurements, finite element modelling (FEM), and machine learning, can monitor the rail neutral temperature (RNT) without disrupting the track structure. At first, temperature, RNT, and acoustic vibration data were collected, then, using FEM tools, the behaviours of high-frequency rail track vibration were predicted simulating different mechanical and thermal conditions. Lastly, a neural network was trained to predict RNT leveraging on the identified high-frequency modes. More details on the implementation are available in the project's final report [89].

The projects described above are summarized in table 4.4. In almost all the cases, projects are still active, and, in addition, they have not yet produced any output in our understanding; this issue is reported with the value "in progress" in the table.

Differently from the S2R projects we reviewed in section 4.2.2, US projects are narrower, i.e. they focus on a single, or few, aspects related to the Railway sector. Also, AI Techniques/Applications play a central role within the projects, being one of the strong points of the success of the project. Therefore, although many projects are still "Active" (not completed), their future outcomes (e.g. papers, systems, models, etc.) will be most likely be AI-focused.

Table 4.4: Summary of US Projects

Project Title	Duration (s: start, e: end)	Founded Programme	Raylway Subdomain	AI Techniques/ Applications	Contribution to AI Integration	Website
Application of machine learning techniques toward time-based change in track condition using an onboard sensor in revenue-service rolling stock	s: 01/09/2019 e: 31/08/2021	UTC	Maintenance and Inspection	Machine Learning	in progress	[75]
Artificial Intelligence for Advance Landslide Warning along Railroad Tracks in Pennsylvania and Delaware	s: 01/07/2020 e: 30/06/2022	UTC	Safety and Security	Deep Learning	in progress	[76]
Technology-Driven Train Inspection (Machine Vision)	s: 2019 e: Ongoing	AHSI	Maintenance and Inspection	Computer Vision	in progress	[77]
Automated Track Geometry Monitoring System	s: 04/12/2017 e: 31/07/2022	UTC	Maintenance and Inspection	Machine Learning	in progress	[78]
Developing a Safety Management System Including Hazardous Materials for Highway-Rail Grade Crossings in Region VII	s: 01/09/2020 e: 28/02/2022	UTC	Safety and Security	Machine Learning	in progress	[79]
Estimation of Origin-Destination Matrix and Identification of User Activities Using Public Transit Smart Card Data	s: 03/02/2017 e: 30/09/2018	UTC	Passenger Mobility	Machine Learning	Algorithms for passengers' OD estimation; Report: [81]	[80]
Laser-based Non-destructive Spike Defect Inspection System	s: 13/11/2020 e: ongoing	Rail Safety IDEA	Maintenance and Inspection	Computer Vision	in progress	[82]
Passenger Flow Modeling and Simulation in Transit Stations	s: 01/11/2019 e: 31/10/2020	UTC	Passenger Mobility	Computer Vision	in progress	[84]
Principal components analysis and track quality index: A machine learning approach	s: 01/06/2017 e: 31/05/2021	UTC	Maintenance and Inspection	Machine Learning	PCA for Track Geometry Parameters identification for defect prediction	[85]
Video-Sensor Data Fusion for Enhanced Structural Monitoring	s: 01/09/2020 e: 31/08/2021	UTC	Maintenance and Inspection	Computer Vision	in progress	[87]
Vibration-based Longitudinal Rail Stress Estimation Exploiting Contactless Measurement and Machine Learning	s: April/2019 e: July/2020	Rail Safety IDEA	Maintenance and Inspection	Machine Learning	Innovative Approach for RTN prediction of CWR; Report: [89]	[88]

Beyond the projects described above, the RiP database produced other results related to Augmented Reality, Data Mining and Big Data Analytics, but without indicating the AI technique (if any) implemented. If from one side they are worth mentioning since could be related to AI in some way,it is also important to point out that it depends on the algorithms/implementations adopted. In [90], the research team investigates the applicability of Augmented Reality (AR) as a tool to improve the quality and safety of the structural condition assessment of transportation infrastructures, placing the emphasis on Bridge inspections. Different solutions and software applications for AR have been presented such as Distance Measurement, Area Measurements, connection with a remote database through barcodes, Change Detector, etc. More detailed information is available in the project's final report [91]. On the other hand, in [92], the research team focuses on the realization of a forecasting model based on Big Data Analytics techniques (defining Data Mining as one of those) for maintenance planning and management of the rail infrastructures. It will rely on different kind of data coming from various Class 1 railroads: traffic information, rail type, lubrification, curvature, etc. Therefore, the aim is to modelling large sets of rail wear data to predict replacement point considering the US wear standards.

4.3.2. China

Most Chinese projects do not have a dedicated official website. Therefore, no reference on websites are provided in our report. Instead, where possible, the outcome papers will be cited.

Major Scientific Research Fund Project of China Railway Corporation [93]

The project aims to optimize and solve the problems of route selection and design in the construction of railways in complex regional environments in the past. The project applied the Case-based reasoning technology in the AI field to study the geological route selection in railway construction. It builds the system architecture of Case-based reasoning technology, and discusses the steps of case display, search, key use, optimization, learning and storage in this system. This project focuses on how to display line selection cases through ontology models, and provides algorithms for local similarity and global similarity to search for cases according to the nearest neighbor method. However, this project only provides simple discussion and basic theoretical support, and there is still a long way to go from the implementation and promotion of the system to provide decision-making assistance for engineers. The major results are presented in a journal paper [94].

Hebei Science and Technology Plan Fund Project [95]

The project aims to use the latest AI technology to provide support and optimization for railway turnout diagnostic systems. This project utilizes the excellent function approximation ability, global optimization and best approximation ability of RBF neural network, and uses Auto Encoder-based RBF neural network to model the switch operating current data to discover, explore and learn the switch fault information.

China Railway Corporation Technology Research and Development Project [96]

Based on the principles of practicability, openness, high reliability, scalability, safety, advancement and maturity, the project optimizes the railway real-name entry verification system to enhance passenger travel experience. The project uses AI technology based on face recognition and uses application service clusters to complete the automatic verification of “ticket-voucher-person” consistency, which greatly improves service efficiency. The proposed approach has been used in Nanchang Station and Chengdu East Station.

Sichuan Science and Technology Research and Development Project [97]

The project aims to use AI technology to accurately detect defects in high-speed railway infrastructure in a timely manner, and to take advantage of the advancement of AI technology in this area to overcome the problems caused by the original manual negligence and improve work efficiency. The project mainly uses artificial intelligence technology based on image recognition and a hundred-layer neural network for intelligent detection. Also, the project uses the self-developed NBK-INTARI neural network technology to carry out image enhancement, edge detection, feature extraction and defect features, and associate the analysis results with the database, and then quickly identify the pictures with defects in the actual measurement.

Table 4.5 summarises all the Chinese projects surveyed in Section 4.3.2.

Table 4.5: Summary of Chinese Projects

Project Title	Duration (s: start, e: end)	Railway Subdomains	AI Techniques/ Applications	Contribution to AI Integration	Website
Major Scientific Research Fund Project of China Railway Corporation	s: 2015 e: 30/08/2018	Transport Policy	Machine Learning	Theoretical support for geological route selection decision making system	[93]
Hebei Science and Technology Plan Fund Project	s: 2015 e: 07/03/2017	Maintenance and inspection	Machine Learning	Methods for Railway Turnout Diagnosis	[95]
China Railway Corporation Technology Research and Development Project	s: 2017 e: 10/05/2018	Passenger Mobility Safety and Security	Computer Vision	Real-name entry verification system optimization	[96]
Sichuan Science and Technology Research and Development Project	s: 2018 e: 08/2019	Maintenance and inspection	Computer Vision	NBK-INTARI neural network for defect detection	[97]

4.4. Classification and Discussion

In Chapter 4, we conducted a systematic review of global related projects using artificial intelligence in railways. Overall, Europe is far ahead in terms of the number of projects, especially projects under the Shift2Rail framework. Of all the projects we reviewed, as described in Figure 4.2, European projects accounted for 70%. In addition, the number of projects belonging to the Shift2Rail framework exceeds 50%.

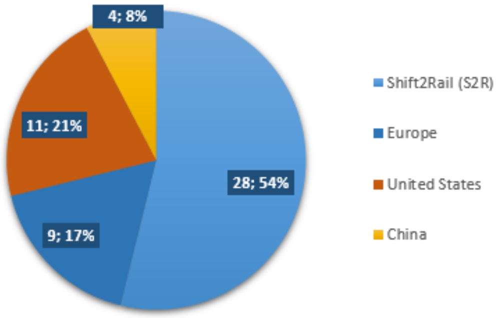

Fig. 4.2. Distribution by Country

Among the 28 Shift2Rail projects, 19 are those we defined as star-projects, i.e. projects for which further information is available regarding their contribution to AI integration in the rail sector. For the other 9 projects, we were not able to indicate the specific AI Techniques/Applications that will be utilised, nevertheless, their contribution to the AI integration in the railway sector will be relevant.

Judging from the duration of the projects we searched, most of the project were conducted in the 2016-2020 period, with some projects starting at an earlier time. Most of the projects reviewed started on the period 2016-2017.

In the following subsections, we have summarized and plotted the main findings of our review by focusing, at first, on Shift2Rail projects (Section 4.4.1) and then providing a holistic (Section 4.4.2) panorama on the current research on Artificial Intelligence in the Railway sector. Finally, Section 4.4.3 highlights some interesting directions for future research based on the existing projects.

4.4.1. Shift2Rail perspective

Most of the reviewed projects under the Shift2Rail framework started around 2016/2017. Among the approximately 83 S2R projects (including RAILS), we found that 19 of them are AI-related and have already produced exploitable outcomes. However, there are other S2R ongoing projects (9) that will probably investigate the AI applicability, therefore they are worth to mention since they will contribute in defining the future panorama of AI in Railways.

In our understanding, S2R projects are broader than the other projects analyzed; many of them addressed issues in more than one Railway Subdomain leveraging various kind of technologies. In many cases, they are not totally focused an Artificial Intelligence and AI is investigated to cope only with a subset of purposes. It follows that retrieving information on AI was not a simple task and in-depth analyses were required to emphasize the true correlation with AI. Moreover, we also found that within a single project different AI Techniques/Applications have been used to face issues in different railway subdomains.

4.4.1.1. Classification

In order to state the current status of AI integration in the Railway Sector, we first considered the 19 star-projects described in section 4.2.2. We drafted three charts to summarize our findings in a statistical way. The pie-chart in Figure 4.3 shows the distribution of S2R star-projects according to the Railway Subdomains (section 3.1.2);the pie-chart in Figure 4.4 reports the distribution of such projects by Innovation Programmes (IPs), already discussed in section 4.2.2; lastly, the bar-chart in Figure 4.5 gives an overview of the number of AI Techniques/Applications used to cope with issues in a specific Railway Subdomain. It is important to emphasize that Computer Vision (CV) applications may not be related to AI techniques. However, in our analysis, we have considered all projects that, in our understanding, implemented a CV approach based on Machine Learning aspects such as Deep Learning. Therefore, although we outlined a clear subdivision among AI Techniques (e.g. Machine Learning) and AI Applications (e.g. Computer Vision), we decided to draft a single chart as, in this case, we will consider only Computer Vision through Machine Learning applications. The same applies to the following overall analysis of all the projects reviewed in section 4.4.2.

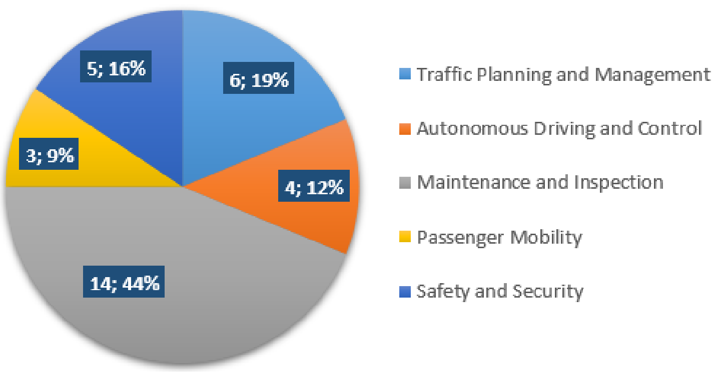

Fig. 4.3. S2R Star-Projects: Distribution of AI within Railways Subdomains

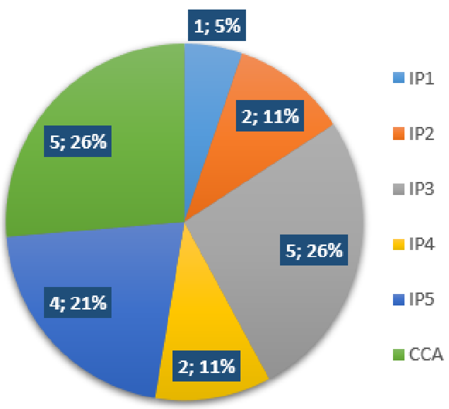

Fig. 4.4. S2R Star-Projects: Distribution of IPs

As expected, the Maintenance and Inspection subdomain was the most investigated, accounting for almost 44% of searches. Similarly, Machine Learning techniques were the

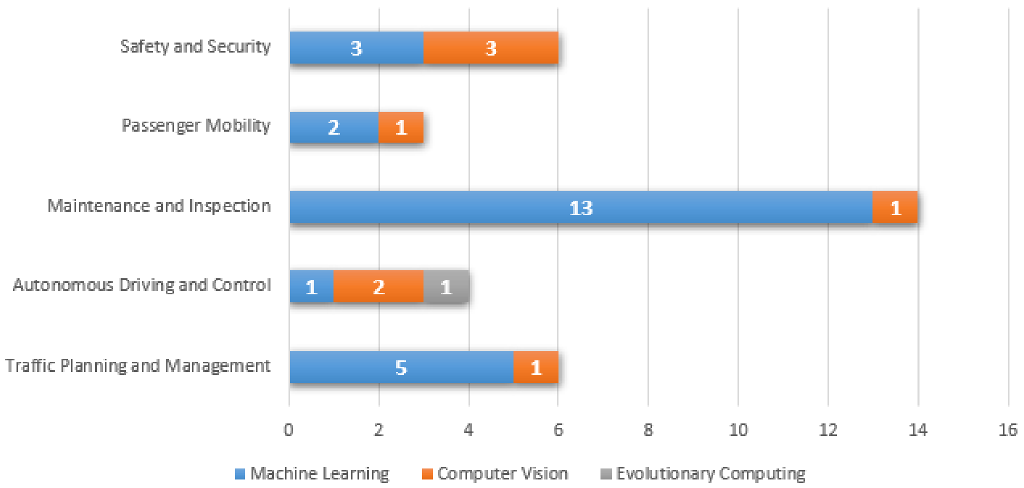

Fig. 4.5. S2R Star-Projects: Distribution of AI Techniques/Applications within Railways Subdomains

most exploited given their versatility and potential to analyse and manage different kinds of data. However, in D1.1 we also defined two more railway subdomains - Traffic Policy and Revenue Management - for which, according to our findings, no step has already been done towards an AI integration.

The charts mentioned above encompasses star projects only. Differently, the pie-chart in Figure 4.6 presents a more comprehensive statistic representing the scenario that will probably be reached when all the ongoing S2R projects will end. The same applies to Figure 4.7, which reports the distribution of all the S2R projects we reviewed according to the IPs. Since we were not able to understand which specific AI Technique/Application will be exploited in these projects due to limited information provided, it was not possible to outline an updated version of the bar-chart presented in Figure 4.5. However, leveraging the projects' descriptions, we can make other considerations to understand which are the research directions that will be analyzed in the near future.

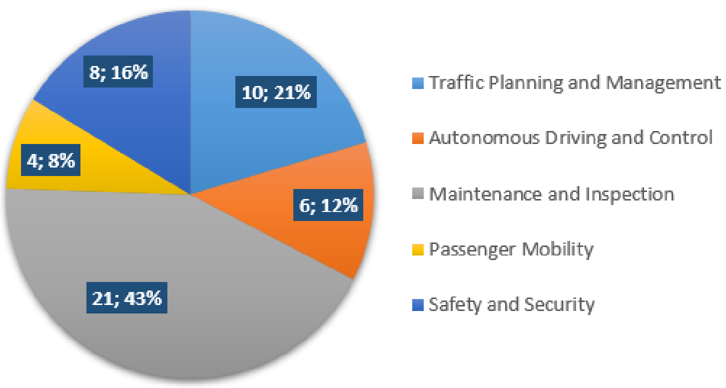

Fig. 4.6. S2R Projects: Distribution of AI in Railways Subdomains

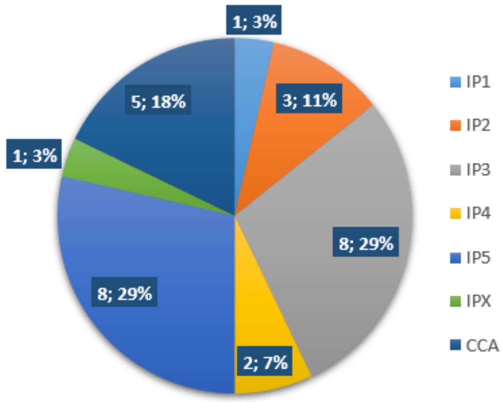

Fig. 4.7. S2R Projects: Distribution of IPs

4.4.1.2. Discussion

As discussed so far, Maintenance and Inspection activities were the most analysed by the star projects and they are also the main topic for ongoing projects. The purpose is clearly to predict eventual failures before they happen to be able to schedule maintenance activities in a data-driven way to optimize costs and ensure safety. Machine Learning has been applied to cope with different issues such as condition-based maintenance for bogies and wagons (FR8RAIL) and traction and braking systems (SMaRTE), fault diagnosis for bogies components (RUN2RAIL), assets malfunctions and failure prediction (e.g. RUL) in switches and crossings (IN2RAIL), anomaly detection at switches as well as track geometry degradation (IN2SMART), wheels defect detection (INNOWAG), and so on. In this direction, X2RAIL-1 investigated the possibility to introduce Smart Wayside Object Controllers to handle both control and maintenance/diagnosis data related to the wayside objects and Route Management Systems thorough the application of AI techniques; this idea is still under examination in X2RAIL-4. Also, other ongoing projects are planning to investigate interesting approaches combining Digital Twins and AI for bogie condition monitoring (LOCATE) and to evaluate changes in crack patterns in tunnels and bridges (IN2TRACK2). Actually, IN2TRACK2 also plans to leverage a Building Information Management/Modelling (BIM) platform; this is also the keystone of the ASSETS4RAIL project that aimed to develop such a platform to integrate algorithms and gather information collected by different kinds of sensors. Beyond structural defects, both IN2TRACK2 and ASSETS4RAIL are also focusing on track inspection and management activities. Lastly, even if we marked them as star-projects, ASSET4RAIL [20] and IMPACT-2 [44] are still ongoing, therefore, they could introduce some innovations in the AI integration for railway asset monitoring and condition based maintenance.

Other inspection-related studies were performed in order to assess the safety of the railway environment. Some approaches were introduced to, among the others, monitor the earth activity and identify active deformation areas through clustering algorithms (MOMIT), predict the changes of earthwork status (IN2SMART), and predict landslides (GoSAFE RAIL).

Some of the projects we already mentioned in this summary have also provided investigations on tasks related to the Safety and Security subdomain: IN2SMART also focused on track circuit false occupancy detection, IN2RAIL applied AI also to predict the risk of derailment, and GoSAFE RAIL also focused on detect obstacles on tracks trough Computer Vision applications. As also investigated by SMART, and under examination in SMART2, systems capable to detect objects on tracks, or any kind of intrusion, can improve the safety on the tracks. This is an interesting topic that, in our understanding, can be also seen as a first step towards the Autonomous Driving and Control. Beyond that, further studies were performed to improve Safety and Security at railway stations leveraging Computer Vision (FAIR STATION) or in rail systems through anomaly detection methods (CYRAIL). Lastly, TRANSLATE4RAIL plans to develop an IT Tool able to allow drivers and traffic controllers to communicate speaking their native language even if they are different (probably based on NLP approaches).

Regarding Traffic Planning and Management, different projects leveraged Machine Learning to tackle tasks such as delay prediction (IN2RAIL, GoSAFE RAIL) and also dwelling time and restoration time (IN2DREAMS), as well as cancellations, route changes, and the prognosis of unexpected events (ATTRAkTIVE). Actually, for the sake of knowledge, IN2DREAMS also focused on the minimization of the overall energy consumption of the 5G infrastructure. Besides, FR8HUB proposed a data exchange platform for exchange traffic information between infrastructure managers and transport stakeholders and implemented an Intelligent Video Gate (IVG), as also FR8RAIL is intended to implement, to recognize wagons at the entrance of marshalling yards to then optimize operations. The involvement of AI to find an optimal or near-optimal solution for marshalling operations has also been investigated in SMART.

Lastly, some steps towards the integration of AI in Autonomous Driving and Control and Passenger Mobility issues have been taken. Concerning Autonomous Driving and Control, beyond SMART, SMART2, and GoSAFE RAIL - which aimed to detect obstacles on the tracks -, IN2DREAMS and OPEUS exploited AI approaches and Evolutionary Algorithms respectively to develop energy optimal driving profiles. Moreover, as an ongoing project, X2RAIL-4 plans to investigate AI and Operations Research techniques to automatize and optimize train movements. Differently, as regards Passenger Mobility, the already mentioned FAIR STATION also focused on crowd management analysis, as well as My-TRAC aimed attention at crowdedness prediction, travellers pattern recognition, and travellers clustering. In this direction, the ongoing project IN2STEMPO is planning to combine Digital Twins and Computer Vision to manage crown in stations, beyond the exploitation of the ML to monitor the train on-board parameters.

4.4.2. Global perspective

Since many of the projects reviewed are still ongoing, in this section we have fused both the current status and the vision of the AI integration in the railway sector at a global level by including all the projects we analysed in this chapter. Hence, we have classified and discussed these projects by providing a global perspective built on a regional comparison between the activities carried out in the different countries. Therefore, we individuated possible future directions of investigation.

4.4.2.1. Classification

The distribution of the investigations in the various Railway Subdomains has not changed drastically in comparison with Shift2Rail. As from Figure 4.8, the Maintenance and Inspection still remains the most investigated subdomain accounting for roughly the 46% of the studies. The next two are Safety and Security and Traffic Planning and Management. However, we also obtained a single result for Transport Policy and, lastly, it is worth noting that in these projects the application of AI has not yet involved the field of Revenue Management.

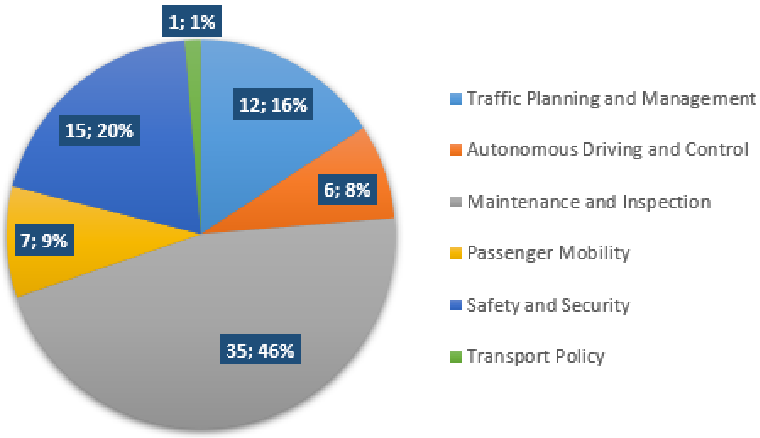

Fig. 4.8. Global Overview: AI in Railway Subdomains

The are strong correlations among different AI Techniques/Applications and railway domains, i.e. certain AI applications are most likely to be applied in specific railway domains. Figure 4.9 shows such a correlation considering an “in progress” value for those projects we were not able to state the specific AI Technique/Application exploited.

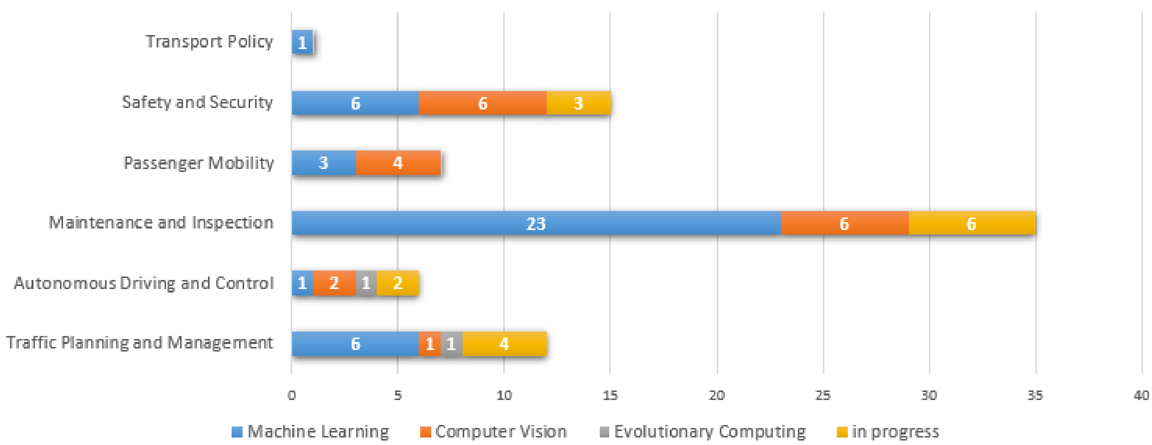

Fig. 4.9. Global Overview: Distribution of AI Techniques/Applications within Railways Subdomains

The Machine Learning techniques are clearly the most implemented, especially to address issues related to Safety and Security, Maintenance and Inspection, and Traffic Planning and Management. On the other hand, Computer Vision applications are the most used for crowd management, i.e. Passenger Mobility purposes, but also highly involved in Safety and Security and Maintenance and Inspection activities. Concerning Autonomous Driving and Control, if we consider that the two *in progress* results are related to SMART2 [41] and X2RAIL-4 [19], the panorama has not changed at all in respect of the one outlined for the Shift2Rail framework.

To better understand differences between S2R projects and the other projects we reviewed, in the next section we will provide a regional-oriented comparison.

4.4.2.2. Discussion

Among the projects reviewed, we found that the public information of European projects is clearer, more specific, and more numerous in comparison to non-European projects. In addition, whether it is in the design and planning of the different stages of the project, or the categories and associations between the projects, the information presented by the European projects is richer and more complete. For example, those projects in the Shift2Rail framework are classified and differentiated according to different Innovation Programmes (IPs). However, China's projects are classified according to the different project funds applied by the sponsors, not according to specific subdomains in the railway. Moreover, none of the Chinese projects have publicly accessible websites. For the US, these projects are basically developed under the University Transportation Centers (UTC) Program of the U.S. Department of Transportation (USDOT) through the collaboration between universities, companies, and other organizations.

Beyond that, there is a more important difference to emphasize. Shift2Rail projects have broader purposes in respect to the projects individuated in the other countries. Shift2Rail projects, and some of the other European ones, focus neither on a single Railway Subdomain nor on the AI applicability. In most cases, they focus on general issues, defined, for instance, by the S2R framework, and try to propose solutions by leveraging on different kinds of technologies where the AI does not have a central role. Indeed, what we indicated as Railway Subdomains in tables 4.2 (Shift2Rail projects) and 4.3 (other European projects), are those subdomains which issues have been addressed through AI, and not all the subdomains investigated by the projects. Differently, US and Chinese projects have a narrower view; they focus on a specific problem and propose an AI-based solution to address it.

For the sake of knowledge, we plotted in Figure 4.10 the distribution of the investigated areas subdivided by regions and in Figure 4.11 the same distribution but contrasting on a more generic level the Shift2Rail projects and those carried out in the other regions.

These charts allow us to easily understand which is the contribution of each region to the AI integration in the Railway Sector. Research conducted within the S2R program exceed 50% in almost all the railway subdomains. As already mentioned, the research focused on Autonomous Driving and Control stops at those already analysed within the S2R program. Similarly, research on Traffic Planning and Management has been carried out only within the European scenario: alongside those S2R, Capacity4Rail relied on ML to predict

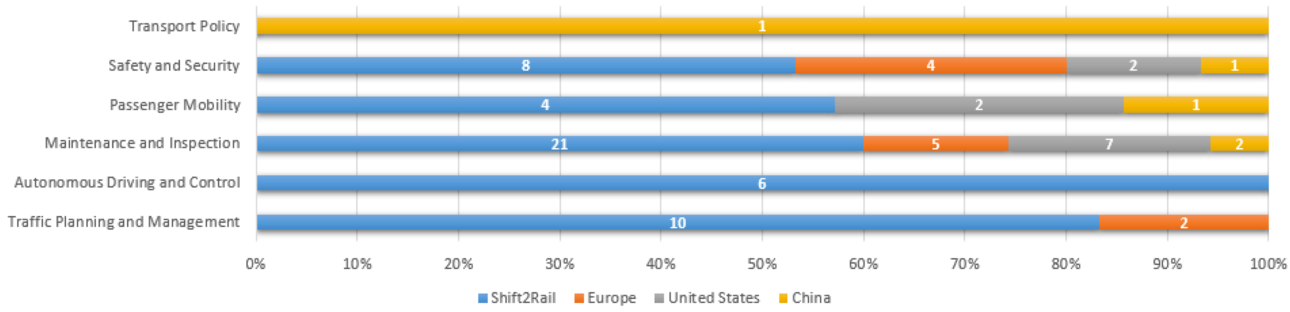

Fig. 4.10. Regional distribution of investigated Railway Subdomains

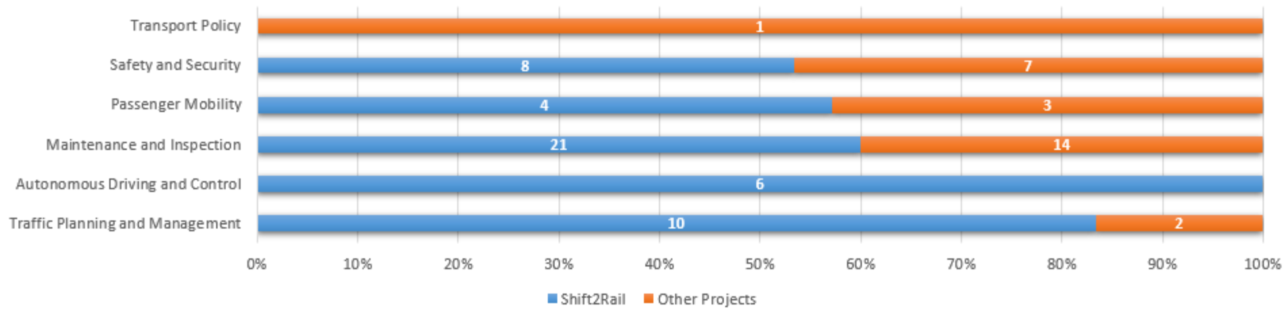

Fig. 4.11. Distribution of investigated Railway Subdomains: Shift2rail vs Other projects

delays, while ON-TIME focused on timetable construction for long-term planning, real-time planning, and replanning during disruptions, also leveraging Evolutionary Algorithms. In addition, the Transport Policy subdomain has been investigated only outside the Europe. Indeed, only in [93], a Chinese project, research was conducted to offer theoretical support to the study of the geological route selection in railway construction through the application of the case-based reasoning technologies in the AI field. Lastly, regarding the subdomains of Maintenance and Inspection, Safety and Security, and Passenger Mobility all the regions provided their own contribution; at the same time, none of them investigated the Revenue Management field.

The field of Maintenance and Inspection has been the most investigated also outside the S2R program, stating that several steps have already been taken in this direction. Searches in China focused on turnout fault diagnosis analysing current data though ML techniques ([95]) and on defect detection in High-Speed railway infrastructures through CV ([97]). In the US, it was developed an innovative technology based on vibration data, finite element modelling, and neural network to monitor and predict the rail neutral temperature of continuous welded rails ([88]). Then, in the European scenario, TT leveraged ML applications to predict the weakness of switch and crossing elements and the degradation of track profiles in order to plan maintenance activities and then reduce costs, ANDROMEDA developed a system to continuously monitor infrastructure's components leveraging data coming from proprietary IoT devices and ML algorithms to predict health conditions and optimal intervention time and type, and SAFE-10-T developed a Decision Support Tool, supported by a Global Safety Framework and based on a Big Data platform, able to detect and classify damages within tunnel linings. Furthermore, in the US, several ongoing projects are are investigating different ML approaches to cope with track defects ([75], [82],[85],[78]), train inspection ([77]), and infrastructure condition assessment ([87]). Particularly, [78] is planning to collect onbord data through smartphones and transform them into track geometry equivalents data leveraging signal processing and machine learning algorithms to identify track geometry fault during the normal train operation; differently, [87] is planning to develop an ML/CV-based system able to operate with both video and sensor data for infrastructure condition assessment. Lastly, as an ongoing project, also the European Drone4Safety is investigating forward the usage of drones and AI to recognise infrastructure components and discover eventual faults on the assets.

Regarding Safety and Security, but also inspection activities, the European project ZIMASS implemented a system based on ML and CV to improve the workers safety during inspection activities; the system is able to create a virtual security zone, recognise risky situations, and alert workers by connected wearable devices. Beyond that, TRAINSFARE proposed a solution based on CV to detect various forms of fare evasion, but also to detect vandalism, unattended baggage, and so on. Then, other projects are investigating Safety and Security tasks both in the European (RAIKA) and American scenario ([76] and [79]). Particularly, RAIKA is planning to analyse ITC systems to identify technical issues and problematic pattern before they cause service interruptions, [76] is focusing on landslide prediction along tracks, and [79] is focusing on landslide prediction along tracks and collision prediction at Highway-Rail Grade Crossings. The latter also includes the estimation of the risk related to hazardous materials and the development of a comprehensive database encompassing crashes, coordinates, inventory, traffic, and freight volume data.

Lastly, beyond those S2R already reviewed, two US projects ([84],[80]) and a Chinese one ([96]) analysed Passenger Mobility tasks. Particularly, [84] modelled and simulated passenger flows leveraging on Computer Vision and agent-based simulation technologies in order to potentially reduce congestion; [80] developed a methodology to estimate the passengers' origin-destination leveraging data related to bus transit smart cards and, among other techniques, cluster algorithms; lastly, [96] exploited an AI-based face recognition solution to optimize the railway real-name entry verification system to enhance passenger travel experience.

5. State-of-the-art

This section reviews a state-of-the-art papers of AI in railway industry. Section 5.1 describes the methodology for selecting relevant papers. Section 5.2 summarises existing review papers and states the existing gaps. Section 5.3 presents an overview and classification of the selected papers. Section 5.4 describes in details the papers based on the railway subdomains. Section 5.5 discusses the main findings and advantages of the state-of-the-art.

5.1. Methodology

To give a comprehensive overview on the state-of-the-art, we started with systematically searching for papers that focus on the application of A.I. in rail systems. Database search was conducted on the well-known database Scopus and Google Scholar was also used as a supplementary source. The searching scope was restricted to academic papers in English, including journal papers and conference proceedings from January 2010 to August 2020. Papers before January 2010 will also be considered only if they play a significant role or are regarded as establishing milestones in their relevant fields. In this section we give a procedure for showing how relevant literature is mined, classified, then archived and applied in this study. This procedure would be introduced in straightforward manner and represented as a step-by-step style.

Table 5.1: Keywords pairs used in initial searching

Rail subdomains	General AI fields
Maintenance and inspection	Expert systems
Safety and security	Data mining
Autonomous driving and control	Pattern recognition
Traffic planning and Management	Adversarial search
Revenue management	Evolutionary computing
Transport policy	Machine learning
Passenger mobility	Operations research and scheduling
	Logic programming
	Natural Language Processing (NLP) & speech recognition
	Computer vision & image processing
	Autonomous systems & robotics

Table 5.2: AI Sub-areas used in following searching

General AI fields	Corresponding AI sub-areas or Algorithms
Expert systems	Rule-based Expert Systems Fuzzy Expert Systems Neuro-Fuzzy Expert Systems
Data mining	Big Data Analytic
Pattern recognition	Pattern recognition
Adversarial search	Deep Reinforcement Learning Agent-based Modelling
Evolutionary computing	Evolution Strategies Genetic Programming (GP) Genetic Algorithms Multi-Objective Evolutionary Algorithms (MOEAs) Swarm Intelligence
Machine learning	Deep Learning Algorithms Deep Neural Networks Convolutional Neural Networks Supervised Learning Unsupervised Learning
Operations research and scheduling	Linear programming Integer programming Nonlinear programming Dynamic programming Constraint programming Network flows and graph theory Scheduling
Logic programming	Abductive logic programming Meta logic programming Constraints logic programming Concurrent logic programming
Natural Language Processing (NLP) & speech recognition	Knowledge extraction Word segmentation Lemmatization Stemming Sentiment analysis
Computer vision & image processing	Object recognition Optical character recognition Movement tracking Image acquisition Feature extraction Image understanding
Autonomous systems & robotics	Robot behavior control Active sensory processing and control Sensor data integration

We initially searched for the terms ‘artificial intelligence’ and ‘rail’ in the title, abstract and keywords. However, A.I. and rail are two big areas. We therefore used the domains in those areas (listed in Table 5.1) and searched for these domains in pairs. Then, we manually checked the title, abstract and keywords and refined the selection by removing all irrelevant papers. Next, the related papers were searched following the forward snowball method, by exploring references and citations of already selected papers. Therefore, a refinement could be made based on full texts and sorting of all papers based on full texts.

1. Broad retrieval to construct a comprehensive paper-set
We first performed one-to-one queries according to the keyword pairs shown in the table 5.1 above. Then, we marked and recorded qualified papers. After querying all of the keyword combinations, a total of 255 papers were marked.

2. Applying 10 year filter on marked papers
On the basis of the first step, we filtered all included articles by setting the filter condition to “papers published after 2010”. The number of all qualified papers is 157.

3. Add a few additional papers
In this step we checked the papers that screened out in the second step, to find out if any of them omitted. And enlarged the paper collection with two extra resources: Significant ones but beyond 10 years limitation; And some articles in Chapter 9 of Deliverable 1.1. Firstly, we added representative papers that were published before 2010. And then, we added the papers that appeared in Chapter 9 of Deliverable 1.1, but were not involved in the first two steps. The total number of all papers obtained in this step is 175.

4. Manually remove irrelevant papers
In the last step, we checked and reviewed all the remaining papers shown in step 3. In the process, we removed some papers related to ‘Operations research and scheduling’ that are unqualified due to the methods they used based on exact optimization approaches without AI components. The number of papers we obtained in this step is 145.

5. Add more relevant papers by retrieving succeeding keywords of AI categories
Steps 1 to 4 largely demonstrate a whole procedure of including/excluding, selecting and assessing the availability of the scoped papers. However, these papers obtained after step 4 cannot fully cover every merging area/direction from the point view of AI professions. Thus we further explore literature database using the subdivided terms shown in Table 5.2 to make sure there is no qualified study neglected. The number of obtained papers is 165 in this step.

6. A systematic inspection towards the obtained paper collection
We conducted a profound investigation with respect to multiple dimensions of gathered articles. The interested aspects include: Journal/Conference, Concerned AI categories, Sub-domains of railways, Objectives, Proposed methods, Contributions, Future directions, Focused areas and available URLs.

   The dimensions listed above could be divided into two groups according to it actually describes what specific aspect of an academic paper. “Metadata Group”, also known as “Data of Data”, mainly describes the property of other data. ‘Journal/Conference’, ‘Concerned AI categories’, ‘sub-domains of railways’ and ‘URL’ can be included in this group due to these four attributes give the information about when/where the paper was published, how we access it online and which AI category and rail sub-domain this paper might be classified into. Another group we would define as “Content data group", which literally contains the dimensions of 'Objectives', 'Proposed methods', 'Contributions', 'Future directions' and 'Focused areas'. The contents of these five fields could be retrieved from the parts of abstract, methodology, solutions, conclusion and keywords in each paper. Nevertheless, Not all the dimensions listed above are explicitly stated in each article. We need to manually search them from each section of text and extract the key points as accurately as possible, and then paraphrase and summarize them into our own words.

7. Manually remove irrelevant papers again
   The final step of the selection process is to remove some unqualified papers based on the result of systematic inspection in step 6. Hence we enlarged our exclusion criteria with the following points.

   Firstly, duplicated ones in the summary of selected paper collection. Secondly, the papers that only pay attention to specific AI techniques or certain railway applications (but fail to incorporate them together). Then, the papers that not focused on the rail sector (some of them that simply mentioned the rails as a possible case study or domain of applicability without going deeper). Lastly, the approaches they proposed were not based on AI (e.g. they applied computer vision through traditional image processing techniques that were not based on AI).

   After applying the exclusive criteria above, we eventually obtained 122 papers to do the systematic review work.

5.2. Existing review papers

The goal of this section is to assess the quality and coverage of existing literature review papers. Peer-reviewed articles involve the topics discussed in section 5.1 and their intersections were investigated for exploring limitations and qualities of these SLRs (Systematic literature reviews). Furthermore, we have introduced the universal adoption of state-of-the-art AI in different business practices in Chapter 10 of Deliverable 1.1 by evaluating a considerable amount supporting materials referring to manufacturing, supply chain management & logistics, aviation and road/public transport. Several review papers relate to these sectors were examined and the key aspects of them are given in a summary tables.

5.2.1. AI in Railway domains & other sectors

We reviewed several recent survey papers related to railway and other sectors. However, these articles are quite different and inadequate to meet our requirements in terms of their perspective height, scope coverage and timeliness.

We started the review work in other sectors by typing in the terms "systematic literature review" or "comprehensive survey" combined with the terminologies we introduced in Chapter 10 of Deliverable 1.1. The results shown in Table 5.3. Subsequently, Table 5.4 shows the results of our further search procedure. In total, 18 individual studies among various railway sub-domains and AI categories have been identified by similar searching strategy. It is easy to noticeable that most of them are published after 2010 and column 'AI aspects/AI categories' covers the AI techniques, AI applications and AI research fields where these papers mainly discussed.

Table 5.3: Review papers in other related sectors

Paper References	focused domain	AI categories	Number of papers Surveyed
Mathew E. (2020)[98]	Road/public transport	Big data Internet of Things	19
Shlain et al. (2020)[99]	Road/public transport	Autonomous systems & robotics	12
Zantalis et al. (2019)[100]	Road/public transport & Aviation	Internet of things Machine learning Big data	74
Patel et al. (2020)[101]	Manufacturing	Autonomous systems & robotics Machine learning	34
Toorajipour et al. (2020)[102]	Supply chain management	Operations research & scheduling and planning	80
Cai et al. (2020)[103]	Supply chain management	Operations research & scheduling and planning	80
Hosseini et al. (2019)[104]	Logistics	Operations research & scheduling and planning	167

Table 5.4: Review papers in Railway domains

Paper References	Railway sub-domains	AI aspect	Number of papers Surveyed
Ghofrani et al. (2018)[105]	Maintenance Traffic planning & management Safety and security	Data mining	171
Zhu et al. (2019)[106]	Maintenance Traffic planning & management	Data mining	161
Y.Jain and Yogesh (2019)[107]	Maintenance and inspection	Data mining	26
Wen et al. (2019)[108]	Traffic planning & management	Data mining	183
Wu et al. (2016)[109]	Traffic planning & management	Evolutionary computing	166
Thilagavathy et al. (2020)[110]	Maintenance and inspection	Machine learning	20
Nakhaee et al. (2019)[111]	Maintenance and inspection	Machine learning	59
Shalini et al. (2017)[112]	Safety and security	Machine learning	15
Bešinović, N. (2020)[113]	Traffic planning & management	Operations research & scheduling and planning	81
Corman et al. (2014)[114]	Traffic planning & management	Operations research & scheduling and planning	82
Scheepmaker et al. (2017)[115]	Traffic planning & management	Operations research & scheduling and planning	115
Schlechte et al. (2014)[116]	Traffic planning & management	Operations research & scheduling and planning	110
Fang et al. (2015)[117]	Traffic planning & management	Operations research & scheduling and planning	129
Liu et al. (2019)[118]	Safety and security	Computer vision & image processing	113
Gibert et al. (2015)[119]	Maintenance and inspection	Autonomous systems & robotics	19
Xie et al. (2020)[120]	Passenger mobility	Machine learning	88
Thaduri et al. (2015)[121]	Maintenance and inspection	Big data analytics & Data mining	36
Alawad et al. (2018)[122]	Safety and security	Machine learning	87

5.2.2. Findings & Gaps

For evaluating and validating the achievements and potentiality of the review studies we collected and assessing them with a widely accepted standard, we introduced a criteria named Database of Abstracts of Reviews of Effects (DARE) [123], which basically contains four assessment questions:

1. Are the inclusion and exclusion criteria described appropriately in review paper?
2. Are the relevant studies properly included in this review?
3. Did this review examined the quality of the investigated papers?
4. Are the primary studies adequately described?

The answers of questions above can be described as follows:

- Yes (Y). The inclusion/exclusion criteria were explicitly stated in this study; Research articles are gathered from multiple electronic library resources; There is no any quality issues to take account; Data have been adequately extracted from primary studies and a critical evaluation on these information has been conducted.

- Partly (P). Some inclusion/exclusion criteria may be implicit; No special searching strategies adopted and digital resources are inadequate; Quality issues founded however addressed effectively; Only summarized knowledge represented without validation.
- No (N). No inclusion/exclusion criteria defined; Papers are collected from limited online resources and the number of included papers is restricted; No quality assessment performed even there are some issues; A total lack of evaluation results towards the investigated papers.

We then assess quality of the papers found in 5.3.1 by using DARE criteria. Through the process of manually inspecting and assigning a rewarded score for each answer of the four quality assessment questions of this criteria: Y = 3, P = 2 and N = 1. We examine each paper independently and compute the total score awarded among these four questions.

Table 5.5: Evaluation of Survey papers in other domains

Paper references	If Inclusion/exclusion criteria stated	If relevant studies included	If quality of papers examined	If primary studies described	Total score
[98]	Y	Y	P	P	10
[99]	N	P	Y	Y	9
[100]	Y	Y	Y	Y	12
[101]	N	Y	P	P	8
[102]	Y	P	Y	Y	11
[103]	Y	Y	P	N	9
[104]	P	Y	Y	Y	11

Table 5.6: Evaluation of Survey papers in Railway domains

Paper references	If Inclusion/exclusion criteria stated	If relevant studies included	If quality of papers examined	If primary studies described	Total score
[105]	Y	Y	Y	Y	12
[106]	P	Y	N	P	8
[107]	N	Y	P	Y	9
[108]	P	Y	P	Y	10
[109]	Y	Y	Y	Y	12
[110]	P	P	P	Y	9
[111]	P	Y	Y	Y	11
[112]	P	N	P	Y	8
[113]	Y	Y	Y	Y	12
[114]	P	P	N	Y	8
[115]	P	Y	Y	Y	11
[116]	P	P	Y	Y	10
[117]	P	Y	Y	Y	11
[118]	Y	Y	P	Y	11
[119]	N	N	Y	P	7
[120]	N	P	Y	Y	9
[121]	N	Y	Y	P	9
[122]	N	Y	N	P	7

The results we obtained from various SLRs shown in Table 5.5 and Table 5.6 respectively. These two tables quantitatively evaluated the quality of chosen review studies in other sectors and Railway domains with a series of assessment questions and manual scoring system.

We would go through the quality of these articles from overall and column-by-column perspectives. The last column in Table 5.6 give the information of that all studies scored more than 6 on the measurements of DARE and there are only 5 papers scored less than 9. Three papers scored 12 ([105], [109] and [113]) and four papers scored 11 ([111], [115], [117] and [118]). In terms of the specific quality criterion, only four of the relevant review studies ([105], [109], [113] and [118]) explicitly described what the inclusion/exclusion criteria are. And relatively half of these papers fail to give a straightforward definition of the quality assessment procedure regarding the papers they investigated. It is shown that most of these scholars have expanded their horizons to a broad range among the relevant studies. These peer-reviewed papers have the best performance on extracting adequate information from primary studies, with sufficient or partly critical evaluation on the original resources. We pay more attentions on those scored higher than others when designing the strategy for reviewing. These desired review works are normally featured with a explicitly described inclusion/exclusion criteria, a broad range of literature resources, an effective quality assessment towards investigated academic papers and an objective description of the primary studies.

According to the quality evaluation towards the reviewed papers, we've discovered the gaps needed to fulfill. First, not only their methodologies are worth to be further explored (e.g. [105][1]), but the analytical strategies and innovative reviewing angles can be referred in our own review procedure (e.g. [113] utilized tables and figures to quantitatively measure distribution of papers per years and scientific journals). Second, we should properly incorporate some inclusion/exclusion criteria and corresponding explanation in this study. A well-stated quality validation procedure towards explored studies is needed as well.

In accordance with the quality evaluation results we stated as before, this study would elevate the past review works with more research values by concentrating on multiple aspects. Firstly, our literature review would be conducted with a higher perspective. In particular, we are focusing on all the connections between the complete railway field and all AI technologies and applications, rather than surveys that only focus on the correlation between specific railway sub-domains and specific AI technologies. For example, [106] [107] [110] and [111] only focuses on the application of big data or machine learning in railway maintenance. By doing so, our literature review can provide better, holistic understanding of the overall development and current status of AI in the railway field.

Secondly, our literature review is broader. As long as it involves the application and technology of AI in the railway field, it will be paid attention to and get involved. These AI technologies and applications covers approximately all subdivisions. At the same time, diverse sub-domains of the railway field will also be considered rather than focusing on one

[1] A structural form which described the formal and general aspects of the studied topics and the major topic categories are considered for further discussion

specific field. Although [105] and [106] involves most sub-domains of the railway, the AI technology is limited to big data.

Finally, our literature review is more time-relevant/up-to-date. The papers we are concerned about are almost published in the last ten years, representing the latest technology and application development status. At the same time, we also selected some representative earlier papers. There are no recent review papers that cover all aspects of AI nor railway still.

5.3. Overview of selected papers

5.3.1. Distribution of papers per journal or conference

We summarise all the reviewed papers by the different journals and conferences where they were firstly published. The results are shown in the table 5.8. Due to the wide variety of all journals listed, we only selected journals that published more than one paper.

Table 5.7: Number of papers in journals and conferences

Journal/Conference	Number of selected papers
Transportation Research Part C: Emerging Technologies	10
IEEE Transactions on Intelligent Transportation Systems	4
Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit	4
Expert Systems with Applications	4
Engineering Applications of Artificial Intelligence	3
Journal of Advanced Transportation	3
IEEE Transactions on Instrumentation and Measurement	3
arXiv	2
IEEE Access	2
IFAC-PapersOnLine	2
Journal of Transportation Engineering, Part A: Systems	2
Public Transport	2
Sensors	2
Other journals	79
Total	122

From this table we can see that transportation-related journals, followed by AI and IEEE publications dominate the list; the number of papers published in *Transport Research Part C: Emerging Technologies* is the greatest, more than doubled the numbers for *IEEE Transactions on Intelligent Transportation Systems*, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* and *Expert Systems with Applications*. And then followed by *Engineering Applications of Artificial Intelligence* and *Journal of Advanced Transportation*, in which 3 studies were selected from 2010 to 2020 respectively. Nevertheless, nearly half of the listed journals contribute limited amount of articles towards our paper collection, with only 2 papers selected from each of these journals during the same time period. In other words, the range of selected papers are not been evenly distributed across all the investigated journals and several representative journals can be identified.

5.3.2. Number of papers in each single year

Now we systematically inspect the selected papers from a quantitative angle by measuring the details of how these studies distribute over the ten-year period (there are only 4 papers beyond the ten-year boundary so we do not display them in the figures of section 5.3.2 and 5.3.3). For illustrating the changes of amount of these papers, we summarized them into Figure 5.1 shown below.

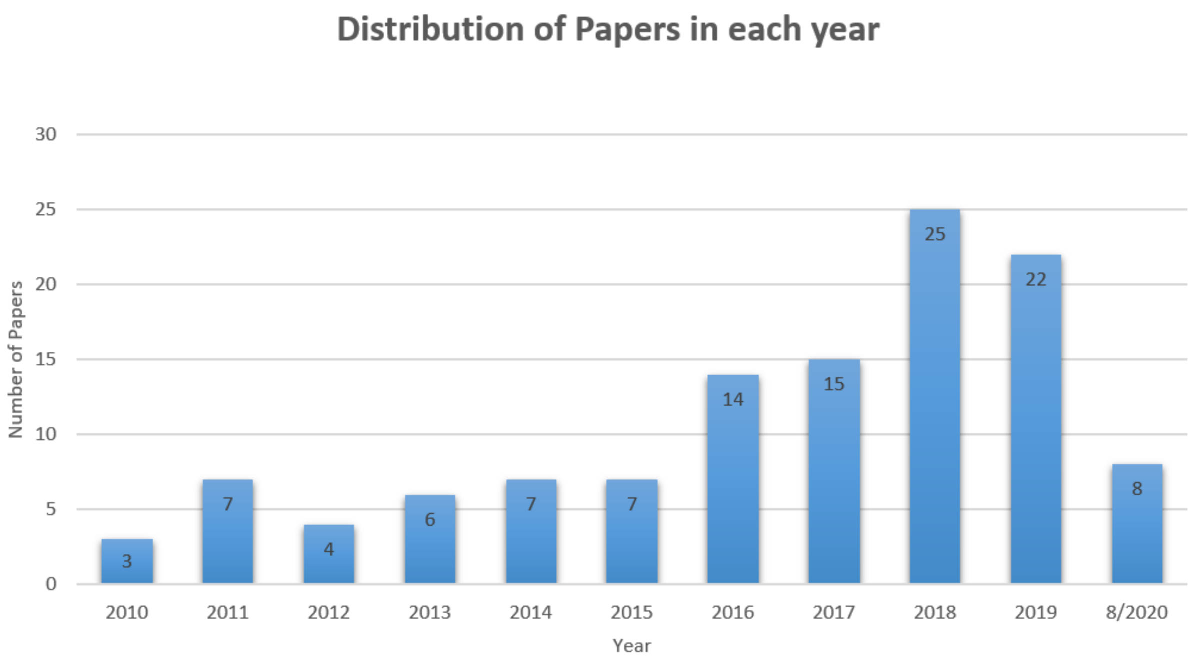

Fig. 5.1. Distribution of Papers for the period 2010-August 2020

The bar chart presents information about the densities of selected papers each year. 2018 is the most flourishing year regarding the number of papers, with 24 articles collected. In the following two years, the figure of 2019 is slightly less than that of 2018 and dropping to 8 in 2020 (disclaimer: 2020 is not complete when paper collecting conducted, only the papers published before August 2020 were shown). On the other hand, the sum of available studies before the end of 2015 was noticeably lower. While it maintained the amount of 7 papers each year between 2014 to 2015. Overall, qualified papers blossom after 2015, before that, it remains a limited level of papers per year.

5.3.3. Distribution of paper per railway subdomain over years

Based on the analysis carried out in subsection 5.3.2, we further summarise all the existing reviewed papers through the different railway subdomains and years of publication. The results are shown in the Figure 5.2 below.

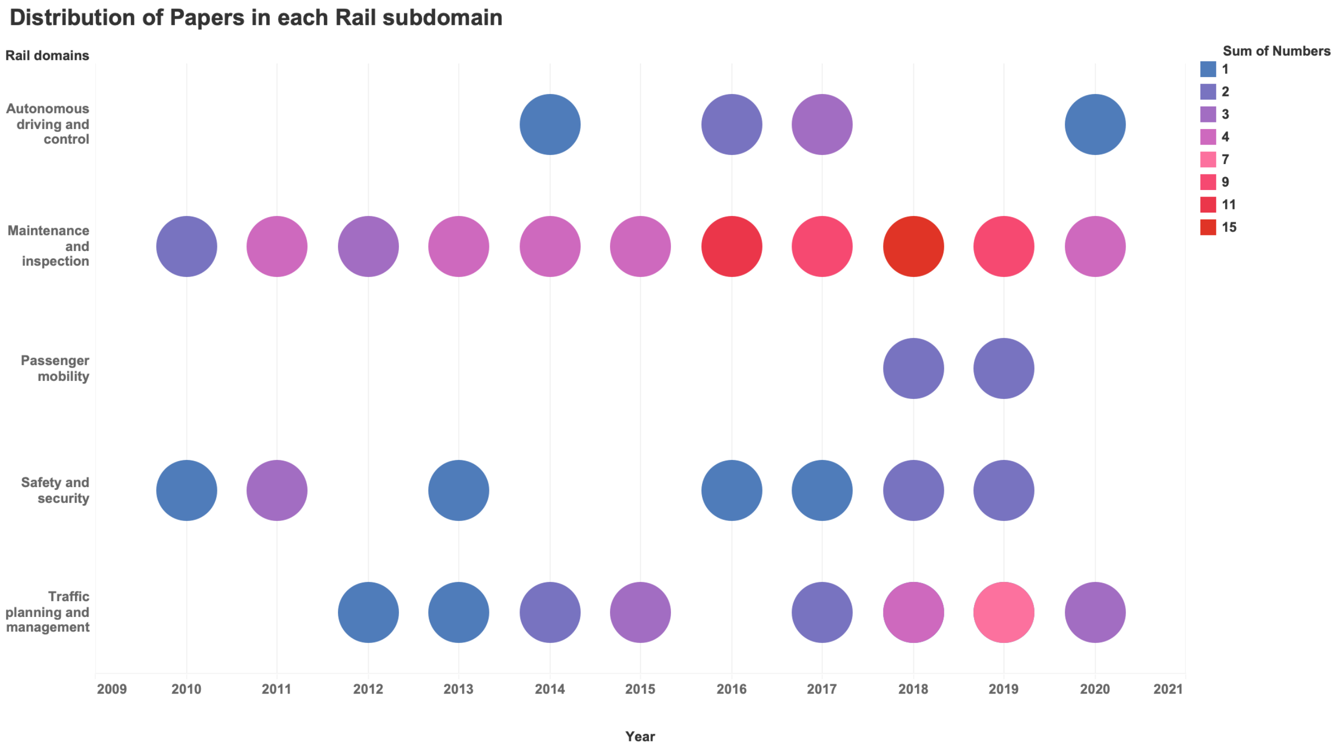

Fig. 5.2. Distribution of papers in each rail sub-domain

This circle chart compares the number of papers in 5 different railway subdomains subject to the time line of 10 years (There is no paper founded in Revenue management between 2010 and 2020). *maintenance and inspection* is the most popular research field over other 5 subdomains for transport investigators throughout the 2010s: despite fluctuation in the first 5 years, the number of chosen papers in this field remains a steady rose and it reaches a new high, peaked at 15 papers in 2018. By the middle of the 2010s, the relevant studies were substantially concentrate on The three domains: *maintenance and inspection*, *safety and security* and *traffic planning and management*. Loads of studies in *autonomous driving and control* and *Passenger mobility* are emerging after 2016. Overall, all subdomains in railway saw an upward trend in the amount of papers, while *maintenance and inspection* and *traffic planning and management* are the two major areas where novel AI techniques largely introduced in.

5.3.4. Proportion of papers in rail domains

According to Figure 5.2 in 5.3.3, the following pie chart 5.3 can be obtained.

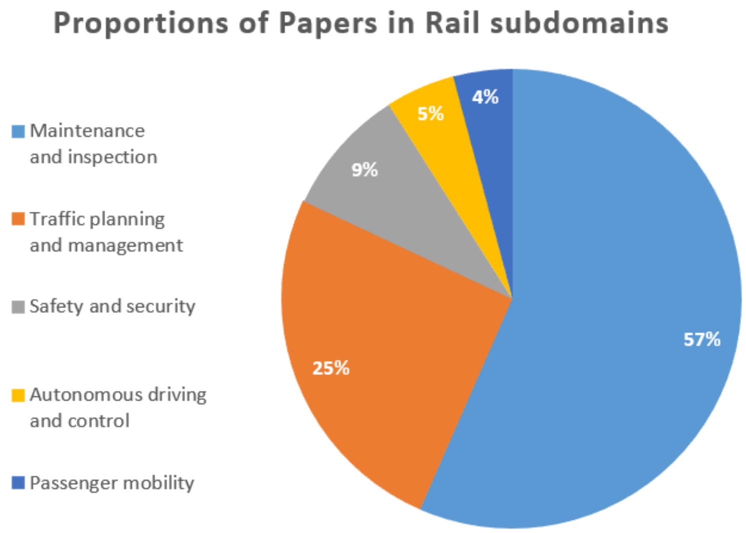

Fig. 5.3. Proportions of papers in rail sub-domains

Figure 5.3 presents the information on how many papers are included in each subdomain of railway. *maintenance and inspection* is the primary component which makes up around 57% of the studies, over twice the amount of papers in *traffic planning and management* and more than six times the papers of papers in *safety and security*. Papers in *autonomous driving and control* and *Passenger mobility* presented a proportion of 5% and 4% respectively. The research paper on the subdomain of *revenue management* and *transport policy* has not been discovered.

5.3.5. Classification of papers in accordance with their objectives

This subsection generated 5 pie charts to investigate what the specific topics or issues are potentially exploratory over the whole railway industries. In the domain of *maintenance and inspection* (See Figure 5.4), defect detection is the most prominent research trends, the proportion of papers which chose these topics is 49%. Researchers were also more concerned with defect prediction and fault diagnosis, with 13% and 12% of articles classified as these two topics respectively.

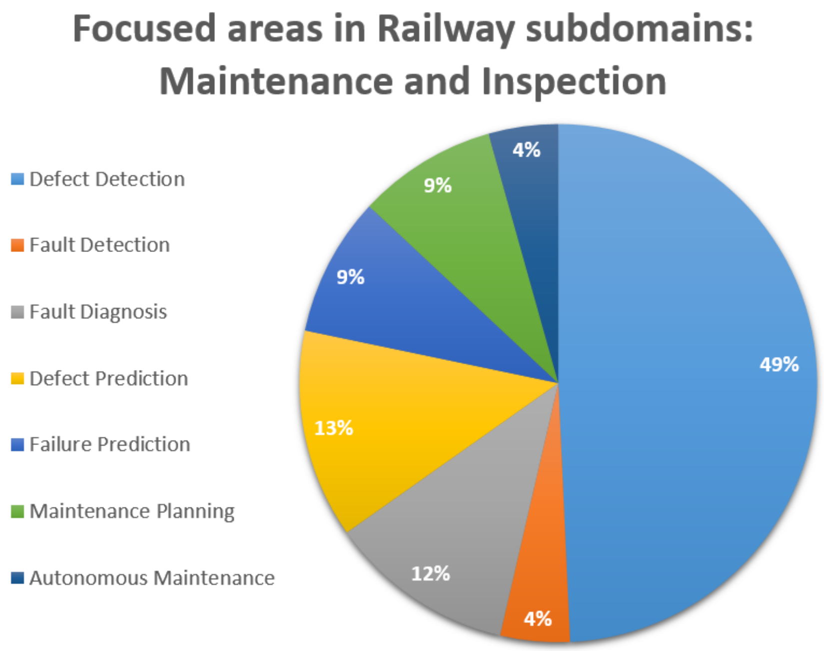

Fig. 5.4. Papers in Maintenance and inspection

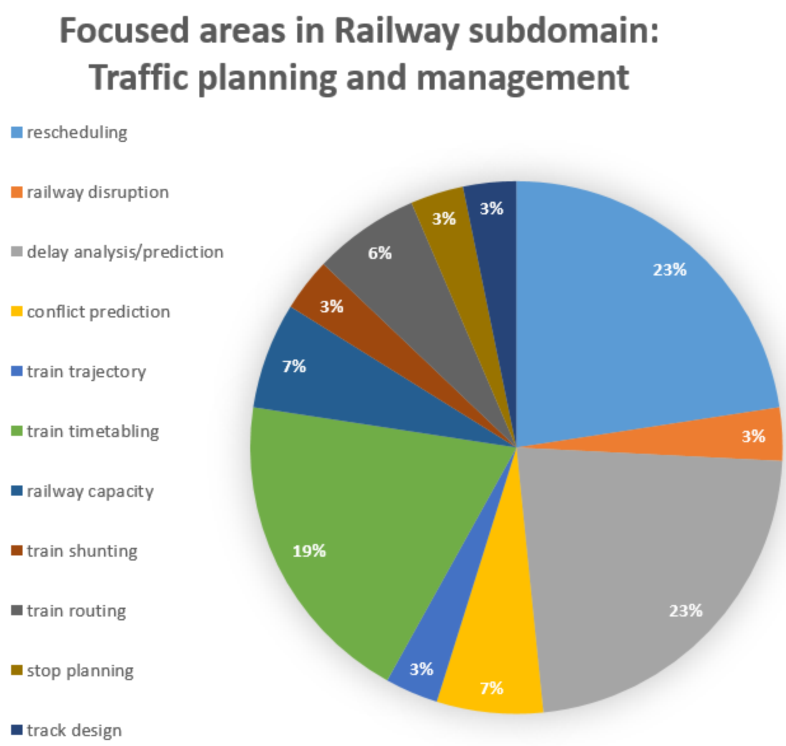

Fig. 5.5. Papers in Traffic planning and management

As for *traffic planning and management*, figure 5.5 shows the percentage of papers which explore rescheduling problems more than tripled the figure for those solving railway capacity tasks. delay analysis/prediction was also a popular research direction and the percentage of papers which paid attention to this holds the same level with rescheduling issues (23% over 23%). Less importantly, 17% of selected papers is project to fall into the group of train timetabling, while the figure was higher than that of railway disruption, conflict prediction and other remaining topics.

Risk management (shown in figure 5.6), is the mainly discussed topic of *safety and security* subdomain, with 28% studies classified in and it more than tripled to the proportion of critical software and anomly detection (9%). While the remaining papers evenly distribute among the topics of railway accidents, railway disruption and collision avoidance with the proportion of 18%.

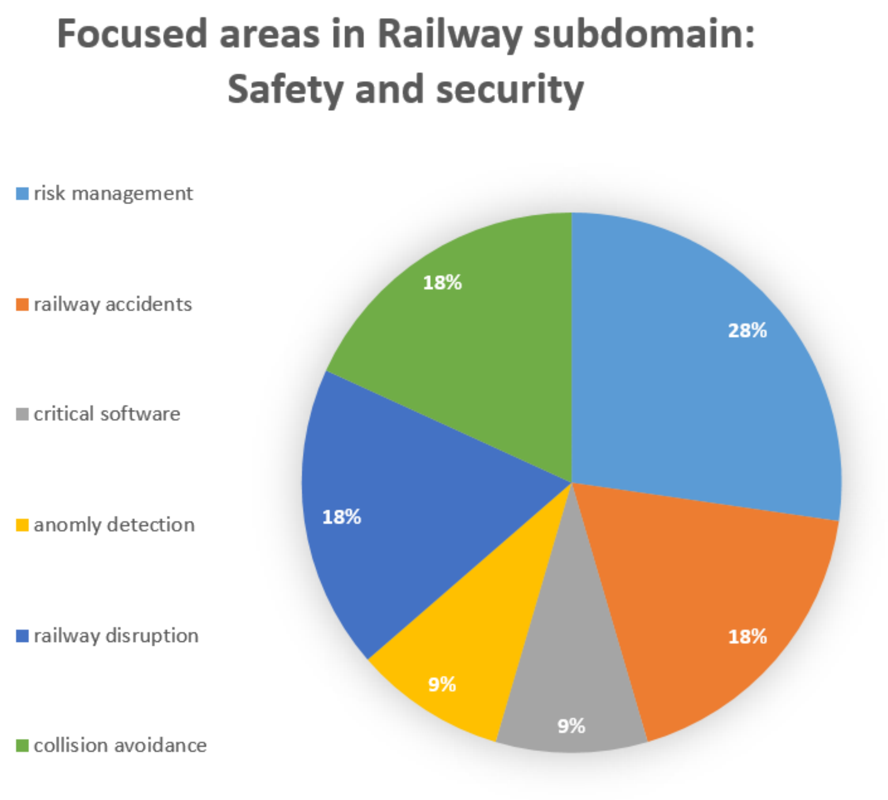

Fig. 5.6. Papers in Safety and security

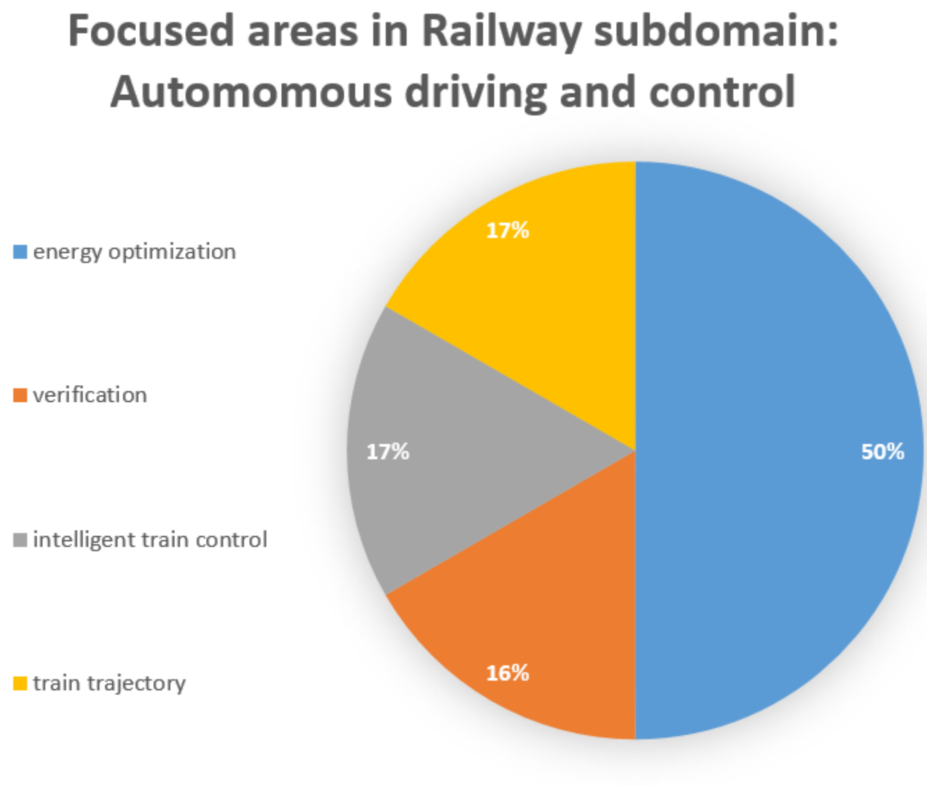

Fig. 5.7. Papers in Autonomous driving and control

Figure 5.7 shows that Energy optimization is the hottest topic among the subdomain of *Autonomous driving and control*, with half of papers are included in this group. The remaining research directions— verification, intelligence train control and train trajectory, hold the same proportion, with around 17% of studies got involved in each of these topics respectively.

There are only one classification among the papers of *Passenger mobility* (See Figure 5.8). Passenger flow prediction seems to be the most discussed issue, making up 100% of the total founded papers.

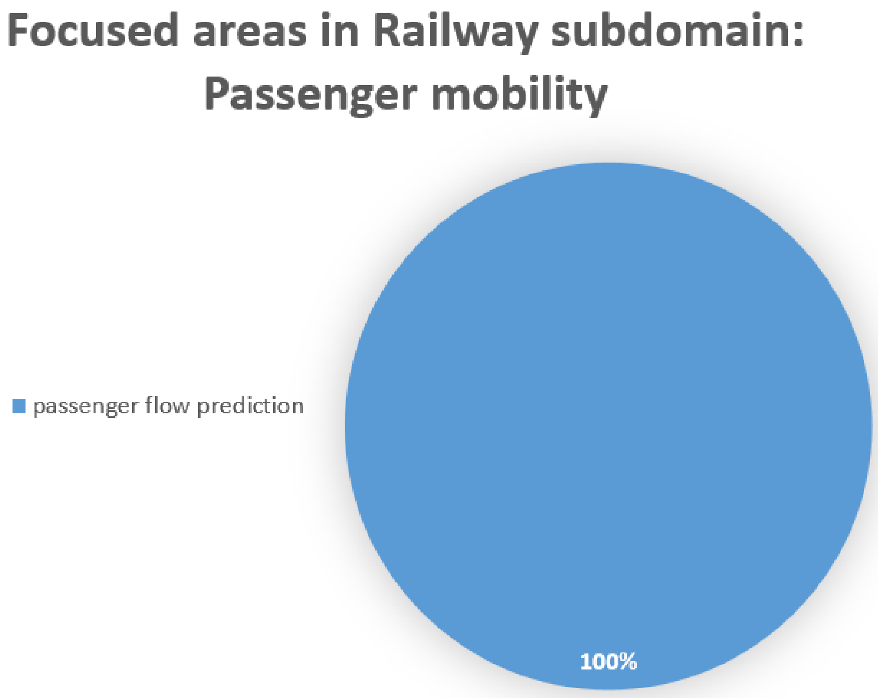

Fig. 5.8. Papers in Passenger mobility

5.3.6. Classification of reviewed papers

Table 5.8: Different rail subdomains investigated in the selected papers

Rail subdomain (frequency)	References
Maintenance and inspection 69 papers (58%)	Eker et al. (2011)[124], Hakan Guler. (2013)[125], Hu et al. (2019)[126], Jamshidi et al. (2017)[127], Jayaswal et al. (2011)[128], Khouzani et al. (2017)[129], Li et al. (2011)[130], Pu et al. (2014)[131], Fumeo et al. (2015)[132], Marsh et al. (2016)[133], Ou et al. (2019)[134], Rabatel et al. (2011)[135], Sammouri et al. (2013)[136], Sharma et al. (2018)[137], Bukhsh et al. (2019)[138], Cherfi et al. (2012)[139], De Bruin et al. (2017)[140], Faghih-Roohi et al. (2016)[141], Famurewa et al. (2017)[142], Firlik et al. (2020)[143], Gao et al. (2018)[144], Giben et al. (2015)[145], Gibert et al. (2017)[146], Fink et al. (2013)[147], Guler et al. (2016)[148], Hajizadeh et al. (2016)[149], Han et al. (2020)[150], Hu et al. (2016)[151], Jamshidi et al. (2018)[152], Jiang et al. (2019)[153], Kang et al. (2019)[154], Krummenacher et al. (2018)[155], Lasisi et al. (2018)[156], Lee et al. (2018)[157], Li et al. (2014)[158], Ritika et al. (2018)[159], Sadeghi et al. (2012)[160], Santur et al. (2016)[161], Santur et al. (2017)[162], Santur et al. (2018)[163], Shang et al. (2018)[164], Shebani et al. (2018)[165], Soukup et al. (2018)[166], Sysyn et al. (2019a)[167], Sysyn et al. (2019b)[168], Tsunashima, H. (2019)[169], Wang et al. (2018)[170], Wei et al. (2019)[171], Xia et al. (2010)[172], Xu et al. (2018)[173], Yin et al. (2016a)[174], Zhang et al. (2020)[175], Yin et al. (2016b)[176], Liu et al. (2016)[177], Sammouri et al. (2014)[178], Ferrari et al. (2018)[179], Lee et al. (2016)[180], Feng et al. (2013)[181], Gibert et al. (2015)[182], Moreno et al. (2020)[183], Santur et al. (2016)[184], Trinh et al. (2012)[185], Moura and Erden (2017)[186], Schlake et al. (2010)[187], Vithanage et al.(2017)[188], Vithanage et al.(2018)[189], Liu et al. (2015)[190], Zhou et al. (2014)[191], Jamshidi et al. (2017)[192]
Safety and security 11 papers (9%)	An et al. (2011)[193], Ćirović and Pamučar. (2013)[194], Gul and Celik. (2018)[195], Li et al. (2011)[196], Alawad et al. (2019)[197], Zhang et al. (2018)[198], Hadj-Mabrouk, H. (2019)[199], Zilko et al. (2016)[200], Sturari et al. (2017)[201], Maire et al. (2010)[202], Wohlfeil, J. (2011)[203]
Autonomous driving and control 6 papers (5%)	Nowakowski et al. (2017)[204], Zhang, D. (2017)[205], Yin et al. (2014)[206], Brenna et al. (2016)[207], Carvajal Carreño et al. (2017)[208], Wang et al. (2020)[209],
Traffic planning and Management 29 papers (24%)	Deng et al. (2018)[210], Alexander,F. (2000)[211], Schaefer,H. (1970)[212], Zhuang et al. (2016)[**?**], Liu et al. (2018)[213], Cerreto et al. (2018)[214], Wang et al. (2019a)[215], Kecman and Goverde. (2015a)[216], Kecman and Goverde. (2015b)[217], Schupbach et al. (2017)[218], Tormos et al. (2008)[219], Barman et al. (2015)[220], Ho et al. (2012)[221], Pu et al. (2019)[222], Wang et al. (2019b)[223], Wang et al. (2019c)[224], Zheng et al. (2014)[225], Khadilkar et al. (2019)[226], Obara et al. (2018)[227], Oneto et al. (2017)[228], Palombarini et al. (2019)[229], Peer et al. (2018)[230], Roost et al. (2020)[231], Salsingikar et al. (2020)[232], Tang et al. (2020)[233], Goverde et al. (2014)[234], Yin et al. (2019)[235], Xue et al. (2019)[236], Bešinović et al. (2013)[237], Prokhorchenko et al. (2019)[238], Barbour et al. (2018)[239]
Passenger mobility 5 papers (4%)	Xu et al. (2004)[240], Gallo et al. (2019)[241], Heydenrijk-Ottens et al. (2018)[242], Liu et al. (2019)[243], Zhu et al. (2018)[244]

5.4. Paper reviews by railway domains

In the following sections, the reviewed papers are presented per specific railway subdomains covering maintenance and inspection, safety and security, traffic planning and management and passenger mobility. Within each subdomain, the papers are further clustered to specific problems/objectives, specific AI techniques/applications have been described, and data used are stated.

5.4.1. Maintenance and inspection

Within maintenance and inspection, we found various applications of AI tackling diverse problems such as defect detection (e.g. Gibert et al. (2017)[146]; Lasisi et al. (2018)[156]; Santur et al. (2018)[163]; Krummenacher et al. (2018)[155]; Han et al. (2020)[150]), fault detection (e.g. Sysyn et al. (2019[168]; Rabatel et al. (2011)[135]), fault diagnosis (e.g. De Bruin et al. (2017)[140]; Yin et al. (2016)[174]), defect prediction (e.g. Sharma et al. (2018)[137]; Shebani et al. (2018)[165]), failure prediction (e.g. Eker et al. (2011)[124]; Fumeo et al. (2015)[132]), maintenance planning (e.g. Guler et al. (2016)[148]; Kumar et al. (2014)[245]), and autonomous maintenance (e.g. Vithanage et al. (2017)[188]).

Defect Detection. Under Defect Detection, we collected papers addressing the identification of physical defects (e.g. cracks, missing items, scratches) in railway tracks (e.g. Gibert et al. (2017)[146], Lasisi et al. (2018)[156], and Santur et al. (2018)[163]), rolling stocks components (e.g. Liu et al. (2016)[177] and Krummenacher et al. (2018)[155]), and catenaries (Han et al. (2020)[150] and Kang et al. (2019)[154]). Beyond these three areas, Ferrari et al. (2018)[179] dealt with the identification of defects within railway signalling requirements documents leveraging NLP applications.

Tracks. We individuated three main track-related subareas: fastening system, track geometry/structure, and rails. Focusing on fastening systems inspection and defect detection, an Adaboost-based approach was implemented both in Trinh et al. (2012)[185], to train multiple classifiers to identify defective anchors and Xia et al. (2010)[172], to detect broken fasteners. In addition, also a Latent Dirichlet Allocation (LaDA) generative statistical model (Feng et al. (2013)[181]). and Deep Neural Networks (Wang et al. (2018)[170], Wei et al. (2019)[171]), were exploited to detect fasteners defects. We also recognized three correlated papers by the same authors proposing a model based on SVMs and Histograms of Oriented Gradients (HOG) features to detect fastening types and defects (Gibert et al. (2015)[182]), a semantic segmentation approach based on fully-convolutional CNN (FCNN) to accurately localize and inspect the condition of railway components using greyscale images (Giben et al. (2015)[145]) and, finally, leveraging the results obtained in these studies, a new multi-task learning framework to detect fasteners type and defects and to identify materials within images (Gibert et al. (2017)[146]).

Regarding track geometry, Neural Networks (Sadeghi et al. (2012)[160]) and SVMs (Tsunashima (2019)[169]) were exploited to detect defects. SVMs were also investigated together with Linear Discriminant Analysis (LDA) and Random Forest basing on Track Quality Indices (TQI) obtained through a Principal Component Analysis (PCA) approach (Lasisi et al. (2018)[156]). Also, the Faster R-CNN framework [246] (a DL-based object detection framework) was exploited to recognize tracks subgrade defects (Xu et al. (2018)[173]).

The last subarea regards rails surface and structural defects. Regarding DL methods for rail surface defect detection, mainly CNN-based architectures have been implemented (Santur et al. (2017)[162], Soukup et al. (2018)[166], Faghih-Roohi et al. (2016)[141], Santur et al. (2018)[163], Jamshidi et al.(2017)[192]), and a transfer learning approach Shang et al.

(2018)[164]). In addition, different traditional ML approaches were exploited such as Linear SVM (Gao et al. (2018)[144]), Random Forest (Santur et al. (2016)[161]), unsupervised approaches in combination with PCA (Famurewa et al. (2017)[142]), and a semi-supervised approach with different sampling methods to cope with imbalanced data (Hajizadeh et al. (2016)[149]) for rail surface defect detection, but also Neural Networks to detect tracks' technical conditions (Firlik et al. (2020)[143]). Lastly, hybrid approaches exist combining multiple methods. Jiang et al. (2019)[153] combined Wavelet Packet Transform (WPT) to decompose the signal of surface defect in different frequency bands, Kernel Principal Component Analysis (KPCA) to reduce the feature space, and SVM to classify rail contact fatigue defects; while, Santur et al. (2016)[184] investigated PCA, KPCA, Singular Value Decomposition (SVD), and Histogram Match (HM) to extract features from video data, then, applied a Random Forest model to detect rail surface defects.

Rolling Stocks. Regarding rolling stocks, different components have been examined such as bolts, angle cocks, and wheels. As ML techniques, SVMs have been exploited to detect angle cock defects (Zhou et al. (2014)[191]), and fastening bolt defects in combination with gradient coded co-occurrence matrix (GCCM) features (Liu et al. (2016)[177]). The same combination was also exploited by Liu et al. (2015)[190] using GCCM features to train, through an Adaboost approach, a cascade detector to identify bogie block key areas and then an SVM to classify defects. From a Deep Learning perspective, authors leveraged CNNs to detect wheels defects (Krummenacher et al. (2018)[155]) and investigated several computer vision approaches for underframe components inspection (Schlake et al. (2010)[187]). Lastly, Posada Moreno et al. (2020)[183] paved the way for a future health estimation framework and decision support system for wagon maintenance implementing a RCNN-based wagon identifier.

Catenaries. Lastly, Han et al. (2020)[150] and Kang et al. (2019)[154] relied on Faster R-CNN to detect defects of catenaries' components. Particularly, the latter also considered a deep multitask learning scenario and merged two networks, a deep material classifier and a deep denoising autoencoder, to define the state of the defects.

Fault Detection and Diagnosis. Fault Detection deals with the identification of faults or anomalies in electrical and similar components, while, Diagnosis aims to identify their causes. Papers under this category, addressed issues related to tracks (Ou et al. (2019)[134], Yin et al. (2016)[176], Cherfi et al. (2012)[139], De Bruin et al. (2017)[140], Sysyn et al. (2019[168], and Lee et al. (2016)[180]), rolling stocks (Li et al. (2011)[130], Jayaswal et al. (2011)[128], Rabatel et al. (2011)[135], and Yin et al. (2016)[174]), and maintenance equipment (Pu et al. (2014)[131]).

Focusing on traditional ML techniques, SVMs were the most involved in railway turnout fault diagnosis supported by a feature space reduction through PCA (Yin et al. (2016)[176]), and both PCA and LDA (Ou et al. (2019)[134]), further, SVM were also combined with mel-frequency cepstrum coefficients (MFCCs) for Railway Point Machines defect detection and diagnosis through audio analysis (Lee et al. (2016)[180]). As other ML techniques: a semi-supervised approach, built on an extension of the EM algorithm, was proposed to learn a statistical model based on Independent Factor Analysis (IFA) for track circuit fault diagnosis (Cherfi et al. (2012)[139]); a model based on Quadratic Discriminant Analysis (QDA) was built for turnout monitoring and fault detection (Sysyn et al. (2019)[168]); and, lastly, back propagation (BP) ANNs were exploited for maintenance equipment diagnosis

(Pu et al. (2014)[131]). In addition, two expert systems based on wavelet transform, ANN and fuzzy rules for rolling bearings fault diagnosis were presented (Jayaswal et al. (2011)[128]); another expert system was used for database establishment for onboard equipment real-time monitoring and diagnosis (Li et al. (2011)[130]); and, roughly for the same purpose, a data mining approach involving a knowledge base built on historical data and a pattern recognition model was implemented to identify anomalies in data coming form train sensors (Rabatel et al. (2011)[135]). Lastly, concerning DL-approaches, a Recurrent Neural Network (RNN) was implemented to measure the spatial and temporal dependencies of faults in track circuits (De Bruin et al. (2017)[140]), while, a Deep Belief Network (DBN)was used for onboard equipment fault diagnosis (Yin et al. (2016)[174]).

Defect Prediction. Under Defect Prediction, researches mainly focused on track-related elements and analysed track geometry (Hu et al. (2016)[151] and Sharma et al. (2018)[137]), track deterioration (Lee et al. (2018)[157]), and rails defects (Jamshidi et al. (2017)[127], Sysyn et al. (2019)[167], Zhang et al. (2020)[175], Li et al. (2014)[158], and Shebani et al. (2018)[165]). In addition, Li et al. (2014)[158] and Shebani et al. (2018)[165] incorporated rolling stocks aspects as well. Beyond these areas, Ritika et al. (2018)[159] leveraged Inception-v3 [247] network to predict vegetation overgrowth and rail buckling.

They addressed these defect prediction tasks exploiting different AI techniques including: ANN for track deterioration (Lee et al. (2018)[157]), SVM for track geometry defects (Hu et al. (2016)[151]), as well as, Decision Trees for both track geometry defects (Sharma et al. (2018)[137]) and rails/wheels wear condition (Li et al. (2014)[158]). Notably, Sharma et al. (2018)[137]) also relied on the Markov Chain and Bernoulli Process to improve the maintenance decision making, while, Li et al. (2014)[158] presented another model to also predict the failure alarm activation caused by hot bearings leveraging Decision Tree and SVM. In addition, the XGBoost algorithm was used to predict the occurrence of broken rails (Zhang et al. (2020)[175]), a regression model, combined with image processing and PCA, was implemented to predict contact fatigue defects (Sysyn et al. (2019)[167]), and a Nonlinear Autoregressive model with exogenous input neural network (NARXNN) was exploited to predict rails and wheels defects (Shebani et al. (2018)[165]). Lastly, an expert system was also built combining fuzzy Takagi–Sugeno interval models to predict squat defects evolution (Jamshidi et al. (2017)[127]).

Failure Prediction. Failure prediction is concerned with forecasting the state of a component that deviates from its nominal behaviour leading to harmful consequences. As for Defect Prediction, papers dealing with Failure Prediction addressed issues mostly related to tracks (Eker et al. (2011)[124] and Hu et al. (2019)[126]) and rolling stocks (Fumeo et al. (2015)[132], Sammouri et al. (2014)[178], and Sammouri et al. (2013)[136]).

These studies mostly aims to estimate the Remaining Useful Life (RUL) of components leveraging different approaches such as the k-means algorithm and Hierarchical Hidden Markov Models (HHMM) for turnouts failures (Eker et al. (2011)[124]), fuzzy neural networks and grey theory for track circuits prognosis (Hu et al. (2019)[126]), and Data Streaming Analysis (DSA) and Support Vector Regression (SVR) for axle bearings RUL estimation (Fumeo et al. (2015)[132]). Also, the same authors in Sammouri et al. (2013)[136] and Sammouri et al. (2014)[178] focused on onboard subsystems failure prediction: in the former, they relied on Null Models-based algorithms and temporal association rules; in the latter, the floating train data were converted in a labeled dataset and analysed through four different ML techniques (K-Nearest Neighbours, Naive Bayes, SVM, and NN). Lastly, a hybrid approach which combined conditional restricted Boltzmann machines (CRBM) and echo state networks (ESN) was proposed to predicting abnormal situation of tilting system of trains based on the analysis results of discrete-event data (Fink et al.(2013)[147]).

Maintenance Planning. The Maintenance Planning category encompasses papers which proposed AI-based approach for decision support systems allowing a dynamic scheduling of tracks' maintenance and inspection activities (Bukhsh et al. (2019)[138], Khouzani et al. (2017)[129], Guler et al. (2016)[148], Guler (2013)[125], Marsh et al. (2016)[133], and Jamshidi et al. (2018)[152]).

Different approaches were proposed leveraging various techniques as expert decision rules (Guler (2013)[125]), Bayesian Networks (Marsh et al. (2016)[133]), and Genetic Algorithms (Khouzani et al. (2017)[129], Guler et al. (2016)[148]) to optimize maintenance activities. In addition, also decision tree, random forest, and gradient boosted tree were investigated to predict whether a maintenance activity, ad its type, must be performed or delayed at switches (Bukhsh et al. (2019)[138]). Lastly, a Decision Support System was built leveraging CNN and fuzzy model combining data coming from multiple sources (cameras and the Axle Box Acceleration (ABA) system) to estimate the rails health conditions (mostly focusing on rail squats defects) and, then, achieve data-driven maintenance planning (Jamshidi et al. (2018)[152]).

Autonomous Maintenance. Papers under Autonomous Maintenance deal with the combination of "Autonomous Systems & Robotic" and "Artificial Intelligence" to obtain intelligent systems able to inspect and maintain railways' components autonomously. The possibility to fuse industrial robots with other techniques/sensors was considered also leveraging Computer Vision and ML to identify plugs in trains' frame and, thus, improve maintenance activities (Vithanage et al. (2017)[188]), furthermore, several supervised machine learning approaches (e.g. Ensemble Bagged Trees (EBGT), Stepwise Linear Regression (SWLR), and rational quadratic Gaussian process (GPRRQ)) were investigated to analyze the feasibility of introducing autonomous robotic systems for automatic train coupler electric head inspection (Vithanage et al. (2018)[189]). Lastly, it was also presented a solution to ensure that the arm of the cab front cleaning robot autonomously adapts to the surface to be cleaned (Moura and Erden (2017)[186]).

Data used. For maintenance and inspection, diverse sets of recorded data were used such as video and images data (e.g. Feng et al. (2013)[181]; Trinh et al. (2012)[185]; Faghih-Roohi et al. (2016)[141]; Wang et al. (2018)[170]; Giben et al. (2015)[145]), and other kinds of records such as data streams by onboard sensors (e.g. Yin et al. (2016)[174]; Fumeo et al. (2015)[132]; Sammouri et al. (2014)[178]; Li et al. (2011)[130]; Rabatel et al. (2011)[135]), audio data (e.g. Lee et al. (2016)[180]), current signals (e.g. Yin et al. (2016)[176], Ou et al. (2019)[134]), vibration signals (e.g. Jayaswal et al. (2011)[128]; Tsunashima (2019)[169]), and so on. In some cases, also combination of multiple data sources were considered; for example, Zhang et al. (2020) [175] analysed track files, traffic information, maintenance history, and prior defect information, while, Jamshidi et al. (2018)[152] combined Axle Box Acceleration (ABA) and video data.

5.4.2. Safety and security

To ease safety & security concerns raised by individual trains even the whole railway system, abundant scholars have investigated the possibilities of introducing AI-based techniques to various complex practical scenarios. For example, risk management (An et al. (2011)[193], Gul and Celik. (2018)[195]), failure/error estimation (Li et al. (2011)[196], Zilko et al. (2016)[200], Sturari et al. (2017)[201]), and railway accidents (Ćirović and Pamučar. (2013)[194], Alawad et al. (2019)[197]), including level crossings accident (Ćirović and Pamučar. (2013)[194]), station accident (Alawad et al. (2019)[197]).

Risk management. A five-phase risk assessment model was proposed using the methods of fuzzy reasoning approach (FRA) and fuzzy analytical hierarchy decision making process (FAHP). More than that, Qualitative measurements were employed by (An et al.(2011)[193]) to describe frequency, severity and probability of consequences for each hazardous event. Based on the result of this study, Gul and Celik.(2018)[195] developed a novel risk assessment approach which combined Fine-Kinney method and fuzzy rule-based expert system, quantitatively revealed risk clusters corresponding control measures. Considering not only the construction expenses & financial cost, but the safety factors, Zhang et al. (2018)[198] utilized the coupled particle swarm optimization (PSO), ant colony optimization (ACO) and A-star algorithm to tackle the problem of hazardous liquid railway network route selection.

Disruption and anomaly estimation. Track circuit is a train detection device that tells whether a block or section of railway tracks is occupied or not. Multiple papers have researched on the maintenance and failure estimation of it. For instance, Zilko et al.(2016)[200] analyzed the factors influencing latency time and repair time respectively and correlations between them thereby a copula Bayesian Network-based prediction model for estimating the length of a track circuit disruption was developed. Li et al.(2011)[196] alternatively applied a knowledge rules-based method to simulate the reasoning & deducting process of track circuit coding for a high-speed railway. Correspondingly, the study of Sturari et al.(2017)[201] presented a novel method that mixed visual analysis and point cloud information to effectively detect those surrounding changes that come from the anomaly in the environment (e.g. a falling tree blocking the track or abnormal railway trawlers). Hadj-Mabrouk (2019)[199] presented a Software Error Effect Analysis (SEEA) framework that comprised of a knowledge-based system, a case-based reasoning system and an assessment model towards the safety of critical software.

Railway accidents. Decision Tree (DT) has been employed by Alawad et al.(2019)[197] to discover hidden patterns & knowledge among hazardous events happened at railway stations and therefore to predict the behavior of affected passengers. Ćirović and Pamučar.(2013)[194] suggested a decision making support system for identifying the qualified candidates among a great amount of level crossings. Each level crossing would be evaluated and assessed by the proposed Adaptive Neural Fuzzy Inference System (ANFIS) and the selected ones would optimized from passively protected to active system-aid protected.

Data used. In the papers that concern Safety & security, multiple sources of statistics have been employed such as expert scoring towards risk-related factors matrix (Gul and Celik.(2018)[195]), historical accident records (Alawad et al.(2019)[197], An et al.(2011)[193]), vision captures of track defects or surroundings (Sturari et al.(2017)[201]).

5.4.3. Autonomous train driving and train control

Autonomous train driving and vehicle control have been attracting more attention to AI in recent years because they aim to enhance safety and enable a more efficient speed profile. The powerful learning and computation capability of AI have led to serve many goals of autonomous driving and control better, especially train energy efficiency (Carvajal Carreño et al. (2017)[208]), as well as verification and validation (Nowakowski et al. (2017)[204]). The problems are present in both metro (e.g. Zhang, D. (2017)[205]) and mainline railways (e.g. Wang et al. (2020)[209]).

Energy-efficient driving. To tackle energy-efficient driving optimization, approaches are developed based on Genetic Algorithms (GA) (Brenna et al. (2016)[207]), Approximate Dynamic Programming (ADP) (Wang et al. (2020)[209]), and Reinforcement Learning (RL) (Yin et al. (2014)[206]). In particular, Wang et al. (2020)[209] uses ADP to learn the costs of energy and time over iterations. Differently, Yin et al. (2014)[206] presents two train control algorithms – an expert system and a reinforcement learning – to operate the train similar to an experienced driver with real-time data to reduce energy consumption whilst maintaining comfort level and punctuality.

Verification and validation. In verification and validation of railway control, AI can be utilized to learn from the expert knowledge without the presence of an expert (Nowakowski et al. (2017)[204]). With respect to big data analysis, Zhang, D. (2017)[205] adopted a model based on fuzzy Resource Description Framework (RDF) model and uncertainty for different high-speed train control systems.

Data used. So far, scholars mostly exploit real-life infrastructure and theoretical engine characteristics to evaluate the performance of algorithms (Brenna et al. (2016)[207], Carvajal Carreño et al. (2017)[208], Wang et al. (2020)[209]). Moreover, studies often draw valuable information from expert knowledge to enable the AL models (Nowakowski et al. (2017)[204], Yin et al. (2014)[206]).

5.4.4. Traffic planning and management

AI has been applied in various traffic planning and management problems such as timetabling, routing, shunting & capacity management (e.g.Khadilkar et al.(2019)[226]; Fragnelli et al.(2014)[248]; Goverde et al.(2014)[234]; Tormos et al.(2008)[219]; Ho et al.(2012)[221]; Bablinski et al.(2016)[249]), traffic analysis and prediction (e.g.Wang et al.(2019)[215]; Oneto et al.(2017)[228]; Kecman and Goverde.(2015)[217]; Bešinović et al.(2013)[237]) and rescheduling & railway disruptions (e.g. Barman et al.(2015)[220]; Wang et al.(2019)[223]; Schaefer, H.(1970)[212]; Deng et al.(2018)[210]; Roost et al.(2020)[231]; Zheng et al.(2014)[225]). In addition, two more papers on strategical planning problems have been found such as stop planning & track design (e.g.Tang et al.(2020)[233]; Pu et al.(2019)[222]). The papers on (general) heuristics or pure mathematical programming are not considered, because they are not conventionally regarded as AI. Instead, the evolutionary programming/heuristics and MP using AI are in the scope.

Strategical Planning. Tang et al.(2020)[233] proposed a Gradient-boosting decision tree (GBDT) model mixed with weather variables and travel historical statistics of individuals for analyzing the origin-destination flow of city shuttle buses and correspondingly predicting the alighting stops for passengers. In order to optimizing the process of railway alignments especially the three-dimensional alignment in mountainous regions, a genetic step-by-step hybrid particle swarm algorithm was constructed by Pu et al.(2019)[222].

Tactical Planning. Several tactical planning problems including timetabling, routing and shunting have been addressed. Given that tactical planning problems are normally designed from the perspective of satisfying the existing constraints and optimizing a multi-criteria objective function, scholars normally aim to generate a feasible timetable which ensure the whole track line (or in a station area/depot) to be conflict-free. Numerous AI-based techniques can contribute to simplify this process, e.g. Bio-inspired algorithms (Tormos et al.(2008)[219]; Ho et al.(2012a)[221]), Reinforcement learning models (Khadilkar et al.(2019)[226], Peer et al.(2018)[230]; Salsingikar et al.(2020)[232]).

We fully inspected the design objectives of these studies and divide them into two groups: train operators-centred (TOs-centred) (e.g. Tormos et al.(2008)[219]; Khadilkar et al.(2019)[226]) and quality of services-centred (QoS-centred) (e.g. Schupbach et al.(2017)[218], Xue et al.(2019)[236]). The first paradigm aims to generate a feasible timetable that specifies the departure and arrival time for all trains such that they are assigned with required resources (e.g. rail infrastructures and facilities). Whilst quality of service-centred models take passengers' service quality into account, it aims to reduce the total travel time or waiting time when transfers are needed.

To minimize train delays, Barman et al.(2015)[220] developed a heuristic model from passengers' scope which combines a set of Fixed Path formulations with a Genetic Algorithm, for selecting a least-time-cost path for each train. A Genetic Algorithm (GA) was introduced by Tormos et al.(2008)[219]. Furthermore, the negotiation process between infrastructure provider and train operators has been modelled as a multi-objective optimization problem according to Ho et al.(2012a)[221] to form a track access rights agreement. A step-by-step approach for a new capacity planning paradigm based on service improvement intention was proposed by Schupbach et al.(2017)[218] and they presented an automatic timetable generation process with genetic algorithm formulations in the context of Swiss Federal Railway. To make use of the wasted capacity at a constant departure frequency, the Genetic Algorithm was adopted by Xue et al.(2019)[236] as well, with the objective of finding the optimal solution in a double-routing optimization model. Wang et al.(2019)[224], developed a continuous multi-objective swarm intelligence algorithm for solving a routing optimization problem and analysing towards simulation results from a quantitative and qualitative point view.

Additionally, Yin et al.(2019)[235] designed a three-phase heuristic algorithm to address a demand-responsive timetabling problem, while a hybrid performance-based timetabling strategy was used in Goverde et al.(2016)[234] where they selected several performance indicators to evaluate constructed timetables.

In timetabling, a reinforcement learning algorithm was developed to allocate track resources to each train and optimize departure/arrival time with the objective of minimizing the total priority-weighted delay (Khadilkar et al.(2019)[226]). Peer et al.(2018)[230] and Salsingikar et al.(2020)[232] applied Deep Reinforcement learning methods for addressing Train Unit Shunting Problem (TUSP) and single-track routing problem respectively. Especially, the former one trained a convolutional neural network using input state representation of shunting yard and gained a better performance than exact operation research method.

Traffic Analysis. Statistics for Large-scale Railway Networks exhibits characteristic of considerable amount and multiple format. Conventional data analysis tools may not perfectly meet the requirements of discovering patterns (from big data) in current railway traffic. Therefore, novel data mining analysis tools (Wang and Zhang.(2019)[215]; Cerreto et al.(2018)[214]; Kecman and Goverde(2015b)[217]), evolutionary-based techniques (e.g. Oneto et al.(2017)[228]) have been introduced to accommodate this challenge both in delay analysis and conflict prediction. Liu et al.(2018)[213] developed a complex three-tier data mining processing framework for analysing train timetable performance measures (i.e. arrival punctuality or running time of the whole line) through a supervised process. Cerreto et al.(2018)[214] employed a data mining technique based on *k-means* clustering to identify substantial delay patterns and summarize primary cause for each clustered group of delay occurrences. Whilst Wang and Zhang.(2019)[215] studied how the factors of weather & scheduled timetable would effect on train delays by proposing a gradient-boosted regression tree model. For accurately predicting running and dwell time and train event times (i.e. and thus expected conflicts), Kecman and Goverde(2015a)[216] and Kecman and Goverde(2015b)[217] developed multiple data-driven approaches such as robust linear regression, tree-based algorithms (regression trees, random forests) and dynamic arc-weighted event graph model. Similarly, Oneto et al.(2017)[228] applied big data analysis tools (i.e. deep/shallow extreme learning machine) to build a data-driven train delay prediction system that considers effects of historical train movements and weather patterns. More than that, A model combined Artificial Neural Networks and MLP approaches to predict the arrival time for freight trains has been proposed by Prokhorchenko et al. (2019)[238]. To estimate arrival times for freight traffic on US railroad, Barbour et al. (2018)[239] proposed a data-driven system to predict times of arrival for individual freight trains based on the property of them, which compared the performance of multiple supervised machine learning models. The study of Zhuang et al.(2016)[250] addressed the gap between traditional approach and novel solution to conflict prediction problems by employing a temporal fuzzy reasoning method. Differently, a simulation-based approach for optimizing the parameters of train dynamics equations and a program for train length estimation were proposed by Bešinović et al.(2013)[237], which aimed to design a reliable train running time model.

Rescheduling and Disruptions. Multiple studies have investigated rescheduling problems for eventual disturbance and disruption of services, and have proposed solutions based on bio-inspired methods (e.g. Wang et al.(2019)[223]) and reinforcement learning (e.g. Obara et al.(2018)[227]; Roost et al.(2020)[231]). The distinctive objectives can be identified as train-oriented and passenger-oriented as well. The former commonly considers reducing differences between the scheduled timetable and actual rescheduled timetable and thus minimising total/primary/knock-on delays of trains (e.g. Wang et al.(2019a)[223]), and the latter – maximization of quality of services for passengers or passengers' satisfaction (e.g. Obara et al.(2018)[227]).

A Genetic Algorithm-based Particle Swarm Optimization (GA-PSO) method was proposed by Wang et al.(2019a)[223] to reduce the sum of secondary delays and the number of trains whose delay overpass a preset threshold. Expert systems/knowledge-based decision support systems have attracted significant dispatchers' attention recently due to its characteristic of lowering computation time dramatically. The proposed models typically use cost functions based on sum of total delays of the train (e.g. Schaefer, H.(1970)[212]; Deng et al.(2018)[210]; Fay, A.(2000)[211]). Regarding RL approaches, Deep Q-network (DQN) method was used both in the studies of Obara et al.(2018)[227] and Palombarini et al. (2019)[229] as an agent to generate instructions to train movements subject to maximizing passengers' satisfaction and minimizing primary delay of trains and Roost et al.(2020)[231] used an model-free Asynchronous Advantage Actor-Critic (A3C) reinforcement learning algorithm (developed by Google Deep Mind, Babaeizadeh et al. (2017)[251]). Differently, Zheng et al.(2014)[225] built a hybrid biogeography-based optimization algorithm combined with differential evolution (DE) to minimize the weighted delivery time in the problem of disaster relief supply operations.

Data used. For traffic planning and management various historical data have been used such as realised traffic movements (Oneto et al.(2017)[228]; Khadilkar et al.(2019)[226]), infrastructure occupation data (Kecman and Goverde(2015a)[216]; Bešinović et al.(2013)[237]; Ho et al.(2012a)[221]; Schupbach et al.(2017)[218]; Goverde et al.(2014)[234]), historical weather records (Wang and Zhang.(2019)[215]; Tang et al.(2020)[233]), existing train scheduled timetables (Deng et al.(2018)[210]; Wang et al.(2019)[223]; Huo et al.(2016)[252]), topology of rail network (Zheng et al.(2014)[225]), accident event data (Fink et al.(2013)[147]).

5.4.5. Passenger mobility

The merits of the railway passenger transportation system drive it to attract an increasing number of passengers, and therefore a reliable prediction is of prime importance to match the passenger demand and vehicle supply, resulting in a higher level of service. AI has been mostly applied to the passenger flow forecasting, both in mainline (e.g. Xu et al. (2004)[240]) and in urban railway (e.g. Gallo et al. (2019)[241]).

Flow prediction. Many machine learning approaches, such as supervised learning paradigms (e.g. Heydenrijk-Ottens et al. (2018)[242] and deep learning architectures (e.g. Liu et al. (2019)[243]) are proposed to improve the prediction accuracy. Particularly, the deep learning method is the hot spot of the application (e.g. Xu et al. (2004)[240] and Zhu et al. (2018)[244]). Specifically, Gallo et al. (2019)[241] predicted metro on-board passenger flows by implementing feed-forward Artificial Neural Networks (ANNs) and selected the best-performing ANN structure for each case study based on the analysis. Similarly, Zhu et al. (2018)[244] performed the same technique to predict the entrance and exit passenger flow, but with a comprehensive influential factors analysis in advance. Differently, Xu et al. (2004)[240] executed ANN to unravel the influence of spatial characteristics in predicting passenger flow. However, feed-forward ANNs have drawbacks like parameter sharing and inefficiency towards time-series data. Hence Liu et al. (2019)[243] applied the Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) and developed an end-to-end passenger flow prediction architecture that integrated the domain knowledge and deep learning. In contrast, Heydenrijk-Ottens et al. (2018)[242] leveraged supervised learning methods to predict and classify the passenger load categories.

Data used. Most commonly historical ridership has been used for the passenger flow prediction (Heydenrijk-Ottens et al. (2018)[242], Zhu et al. (2018)[244], Xu et al. (2004)[240]). This historical ridership data could be derived either from smart card transactions (e.g. Heydenrijk-Ottens et al. (2018)[242]) or tickets (e.g. Xu et al. (2004)[240]). Whereas the long collection time of smart card data could hinder the passenger flow prediction in the short-term, Liu et al. (2019)[243] forecast the passenger flow on-board based on the passenger counts at station turnstiles, which could be retrieved in near real-time.

5.5. Discussion

Maintenance and inspection. According to the findings, researchers have focused their attention greatly on Maintenance and Inspection tasks with about the 58% of the total reviewed papers (as from Figure 5.3). One of the main reasons for this is the general trend to move from corrective or preventive maintenance to predictive one in order to reduce the overall maintenance costs. In essence, given the amount of new data available, the aim is to move towards data-driven methods relying on historical and/or real-time data in order to diagnose faults/defects to predict and avoid possible failures. AI techniques allow maintenance engineers to carry out inspection and diagnostic activities more efficiently in terms of costs, time and accuracy in defect detection, also supporting dynamic activity planning and maintenance automation.

The most involved AI technology in maintenance and inspection is Machine Learning. Traditional ML approaches (e.g. SVM, Decision Tree, Regression algorithms, ANN, etc.) have found great applicability when dealing with tabular (and possibly labelled) datasets (e.g. Sammouri et al. (2014)[178], Sharma et al. (2018)[137]), which, in some cases, can also be obtained starting from images (e.g. Gibert et al. (2015)[182], Liu et al. (2016)[177]), or different kinds of signal like current signals (e.g. Yin et al. (2016)[176]), vibration signals (e.g. Tsunashima (2019)[169]) or audio signals (e.g. Lee et al. (2016)[180]). After pre-processing steps, numerical features are extracted, and tradition ML model builts. Second to that, Deep Learning has also been widely involved, mostly to build Computer Vision applications for Defect Detection and Prediction tasks, where video and images are the most exploited source of data (e.g. Wang et al. (2018)[170], Wei et al. (2019)[171], Santur et al. (2017)[162]). Lastly, as for the traditional ML techniques, DL-based approaches are also used to analyse tabular data for both classification problems (e.g. CNN architectures, Krummenacher et al. (2018)[155]) and temporal dependencies analysis (e.g. RNN architectures, De Bruin et al. (2017)[140]).

The above techniques, together with other Big Data, Data Mining and Pattern Recognition approaches have also been involved in Fault Detection, Fault Diagnosis and Failure Prediction tasks. It is also important to underline the modularity of all the techniques that can be involved in a Data Mining or Pattern Recognition process. There are different techniques for data processing (e.g. Image enhancement for video data), feature selection or dimensionality reduction approaches (e.g. PCA) and data analysis methods to extract knowledge (e.g. ML techniques). It is all about finding the best combination for a given problem; this modularity has been found in the papers we reviewed as many of them have tackled very similar issues but implementing quite different approaches. Given that, it is hard to outline the best approach; it always depends on data exploited and its processing.

However, collecting sufficient and qualitatively good data is not always possible. In fact, in many cases, data augmentation techniques, i.e. retrieving more data from those available through artificial and synthetic transformation, have been used to obtain enough training samples to adequately characterize the models.

Therefore, since data is often the main challenge when dealing with AI-based approaches, as future works most of the authors aim to improve their research by analysing more data (e.g. Eker et al. (2011)[124]), using more sensors (e.g. Liu et al.(2016)[177]), improving data quality (e.g. Tsunashima (2019)[169]), testing their models on on-field data (e.g. Shebani et al. (2018)[165], De Bruin et al. (2017)[140]), adapting the proposed approach to new kind of data (e.g. Trinh et al. (2012)[185], Guler (2013)[125]), considering multiple data sources (e.g. Zhou et al.(2014)[191]), etc. Actually, it is worth mentioning that some steps in combining multiple data sources (e.g. Jamshidi et al. (2018)[152]) and different ML approaches in a multi-task learning scenario (e.g. Gibert et al. (2017)[146]) have already been performed.

At the same time, Expert Systems were exploited in this subfield to predict failures and defects (e.g. Jayaswal et al. (2011)[128], Jamshidi et al. (2017)[127]), and also, beyond other ML-based approaches, to create decision support systems allowing dynamic scheduling of tracks' maintenance and inspection activities (e.g. Guler (2013)[125]). So far, limited work has been done in this direction, then, as further developments, improved rules and more optimized features can be considered, as well as an in-depth integration with other AI technologies and applications. Lastly, it is also important to underline that some initial steps have been taken towards maintenance automation; different approaches have been proposed or investigated in order to combine Autonomous Systems and AI to obtain intelligent autonomous systems able to inspect and maintain railways' components (e.g. Vithanage et al. (2017)[188], Vithanage et al. (2018)[189]).

Traffic Planning and Management. AI innovations in the sub-field of Traffic Planning and Management have been developed a lot, constituting about 24% of the reviewed papers. Although pure mathematical/exact operation research algorithms are popular for those who wanted to find the upper limit of performances of optimization (e.g. Wang et al.(2019a)[223]). Effective AI approaches (i.e. Data mining analysis techniques, reinforcement learning approaches and Expert system/knowledge-based reasoning system) have their advantages to tackle traffic planning and decision making problems—especially those complicated problems that difficult to yield an optimal solution within limited computation time. However, the optimization-based solutions to tackling the procedure for traffic analysis and tactical planning has their own advantages to heuristic approaches (e.g. Genetic Algorithm, Evolutionary Computing and Particle Swarm Optimization).

Conventional machine learning models (e.g. regression trees, reinforcement learning algorithm, Deep Q-leaning method), as an addition to the three significant algorithms with

Bio-inspired/heuristic features recognized above, also widely adopted in some related articles. These algorithms has provided strong support for solving rescheduling problems (e.g. Obara et al.(2018)[227]), timetable design (e.g. Khadilkar et al.(2019)[226]), and train routing (e.g. Salsingikar et al.(2020)[232]). Furthermore, they are effective within big data analytics to identify delay patterns and to estimate delay level for both passenger railway lines and freight network (e.g. Wang and Zhang.(2019)[215], Prokhorchenko et al. (2019)[238] and Barbour et al. (2018)[239]). Using intelligent knowledge-base reasoning and decision making system to support the procedure of train rescheduling (e.g. Schaefer, H.(1970)[212]) would achieve a higher operational efficiency.

The techniques listed above, together with Pattern Recognition and Data Mining approaches were applied to different extents according to the type of target problems. For example, systematic data processing & cleaning frameworks, such as feature engineering, clustering or encoding of time-series data are likely to be adopted against a background of a hybrid large-scale data resources (e.g. the data asset with train performance measures and automatic train supervision data combined). This pattern has been found in the papers we reviewed especially in the cases of delay analysis and conflict prediction (e.g. Liu et al.(2018)[213]). Whilst, a rescheduling problem is all about finding a feasible timetable for train operators after disruptions occurred and this may require a previous experiences in dispatching (e.g. Deng et al.(2018)[210]), expert systems seem to be an effective alternative to manual operations.

All in all, different approaches show their potentials variously in different application scenarios. It is difficult to outline which approach is the most promising one among railway industry even in the domain of traffic planning and management. it largely depends on requirements of processing data and the purpose of studies. However, the applications of machine learning, reinforcement learning are still in their early stage and require further development for satisfying a more complex industrial/business need. Limited work has been done in the direction of natural language processing, computer vision & image processing so far, then, the possibility of incorporating these AI applications into future research strategies should be seriously investigated. Additionally, future research should focus more on model optimization, as well as algorithm improvement with incorporating new indicators or parameters into it, especially the models of machine learning and data mining. For yield a better performance and receive a higher accuracy, using the hybrid model should be considered.

Safety and security. The capability to employ AI as an alternative solution in the sub-domain of safety and security have been partly explored in the reviewed papers. These papers make up approximately 9% of the total number of the reviewed papers.

So far, including expert systems (e.g. fuzzy reasoning system, An et al.(2011)[193]), Bio-inspired methods (e.g. Particle Swarm Optimization, Ant colony optimization) were the genuine attempts on Risk management. The aim of them is typically to quantitatively evaluate the risk score for each physical site and thus invest the limited resources to the most needed areas and reduce general safety cost (e.g. An et al.(2011)[193] and Gul and Celik.(2018)[195]). Hazardous events happened randomly with its own spatial and temporal characteristics, and even it is difficult to estimate the severity of consequences. To quantitatively discover danger clusters seems to be a challenging task for classical Machine Learning algorithms (e.g. supervised learning models) and Data mining. Due to an existing huge variation in the quantity of samples among different level of risks—the amount of small-risk cases are largely surpass that of severe danger conditions. Where the latter ones may not be recognized by a less-trained model. Alternatively, rule-based and case-based reasoning system can be combined in the futures' to work as a more 'experienced' decision maker for allocating safety resources. Computer vision image processing, with its high level of automation, and solid detection accuracy on practical scenarios, has largely engaged with the procedure of discovering, detecting and identifying anomaly in the environment (e.g. Sturari et al.(2017)[201] and Hadj-Mabrouk (2019)[199]). The applications of these techniques generally incorporate cloud information for analysing.

Autonomous driving and control. The combination of AI and autonomous driving and control has shown its potentials with its promising applicability, particularly using reinforcement learning (e.g. Yin et al. (2014)[206]). So far, scholars mostly prove the usefulness of the vehicle control algorithms by using theoretical simulations combining real-life infrastructure and theoretical engine characteristics. However, in real-life train running characteristics are impacted by multiple factors such as wear of rail/wheel contacts, weather conditions, unexpected disruptions or perturbations. Thus, these theoretical simulations (current approaches) still cannot be compared with practical real-life tests, which may raise criticisms towards applicability in real environments. The paper of Wang et al. (2020)[209] using ADP presents an exception as it introduced the stochastic change of traction force and train resistance.

Additionally, the current models typically consider basic signaling systems, while no applications on advanced systems like ETCS, are currently existing. This makes the current models not ready for applications because train operation does not comply with the railway signaling system and respect the speed restrictions and the movement authorities. One of the future avenues of this field could further explore the genetic algorithms and reinforcement learning with a special focus on signaling system-compliant autonomous driving and control to safeguard safety.

Regarding non-vital control systems, only limited research exists. For example, the use of image data has shown its capabilities in controlling the pantograph with the deep learning technique. Besides, the expert system can grant invaluable knowledge without the presence of an expert in certain control fields. However, the number of studies in this non-vital AI-related control domain is still comparably scarce, which implies a future direction to explore the effectiveness of AI.

Passenger mobility. Fifthly, AI innovations have shown their significance in addressing the problem of passenger flow predictions (Gallo et al. (2019)[241], Heydenrijk-Ottens et al. (2018)[242], Liu et al. (2019)[243], Zhu et al. (2018)[244], Xu et al. (2005)[240]). These predictions mainly focus on the short-term prediction tasks. Particularly, scholars are in favor of applying deep learning paradigms to perform the passenger flow predictions (e.g. Gallo et al. (2019)[241], Liu et al. (2019)[243], Zhu et al. (2018)[244]). Compared to other learning paradigms which are normally set as the baseline models (e.g. linear regression), deep learning outperforms than them with a considerable margin by unveiling the non-linear relationship (Liu et al. (2019)[243], Zhu et al. (2018)[244]). Yet, comparing the results across the studies can be difficult at this moment due to several reasons. First, every deep learning architecture needs a well-knitted feature set that is extracted from the input data and this extraction differs from one case study to another due to its influence. This makes the availability of the data, the attributes of the data, and the influence of the data vary between studies. Second, the metrics that the researchers chose to evaluate the precision of the prediction are not entirely similar (e.g. R^2 from Liu et al. (2019)[243] and RMSE from Zhu et al. (2018)[244]), which passes the challenge to interpret the results against each other. This indicates that a set of universal and complete metrics is needed and could be defined. Moreover, one of the directions that could be further explored is to propose an architecture of feature selection and data consideration such that the consistency and convenience of the feature set construction can be guaranteed, for example, spatio-temporal and operation characteristics (Liu et al. (2019)[243]).

Also, historical data (e.g. smart card data retrieved from the Automatic Fare Collection system) has been extensively applied to predict the passenger flow as one of the sound bases. However, this type of data normally is not made (near) real-time which makes the current models unsuitable to perform the short-term prediction tasks because they fail to capture the short-term variations. To this end, incorporating or leveraging other data that could be retrieved (near) in real-time to predict the ridership is another research possibility (Gallo et al. (2019)[241]).

Table 5.9: Overview of AI Algorithms and Models used in specific railway subdomains.

Railway Subdomain		Maintenance and Inspection						Safety and Security			Autonomous Driving and Control		Traffic Planning and Management				Passenger Mobility
Focused Area		Defect Detection	Fault Detection and Diagnosis	Defect Prediction	Failure Prediction	Maintenance Planning	Autonomous Maintenance	Risk Management	Disruption and Anomaly Detection	Railway Accidents	Energy-Efficient Driving	Verification and Validation	Strategical Planning	Tactical Planning	Traffic Analysis	Rescheduling and Disruptions	Passenger Flow Prediction
AI Algorithms/Models: Tree-based	DT			x		x				x					x		
	RF	x				x									x		x
	GBDT					x							x		x		x
	EGBT						x										
Inst.-b	SVM	x	x	x	x												
	SVR				x												
	KNN				x												x
Bayesian	NB				x												
	BN					x			x								
	LaDA	x															
Regres.	LR														x		
	SWLR						x										
	GPRRQ						x										
Neural Networks	MLP																x
	ANN	x	x	x	x												x
	FNN				x					x							
	NARXNN			x													
	CNN	x				x			x								
	RNN		x		x												
	LSTM-RNN																x
	DBN		x														
	RBM				x												
	R-CNN	x															
Cl.	K-means				x										x		
	EM		x														
Other	AdaBoost	x															
	XGBoost			x													
	LDA	x	x														
	QDA		x														
	FRA							x							x		
	FAHP							x									
	KRB								x		x	x					
	ADP										x						
	RL										x			x		x	
	HHMM				x												
EC	GA					x					x			x			
	PSO							x					x			x	
	SI													x			
	ACO							x									

Inst.-b: Instance-based; **Regres.**: Regression; **Cl.**: Clustering; **EC**: Evolutionary Computing
Decision Trees (DT), Random Forest (RF), Gradient Bosting Decision Tree algorithm (GBDT), Ensemble Bagged Trees (EBGT), Support Vector Machine (SVM), Support Vector Regression (SVR), K-Nearest Neighbour (KNN), Naive Bayes (NB), Bayesian Networks (BN), Latent Dirichlet Allocation (LaDA), Linear Regression (LR), Stepwise Linear Regression (SWLR), Rational Quadratic Gaussian Process Regression (GPRRQ), Multi-Layer Perceptron (MLP), Artificial Neural Networks (ANN), Fuzzy Neural Networks (FNN), Nonlinear Autoregressive Network with Exogenous Inputs Neural Network (NARXNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-term Memory Recurrent Neural Network (LSTM-RNN), Deep Belief Network (DBN), Restricted Boltzman Machine (RBM), Region-Based CNN (R-CNN), K-means algorithm (K-means), Expectation-Maximization Algorithm (EM), Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Fuzzy Reasoning Approach (FRA), Fuzzy Analytical Hierarchy decision making Process (FAHP), Knowledge Rules-Based (KRB), Approximate Dynamic Programming (ADP) Reinforcement Learning (RL), Hierarchical Hidden Markov Models (HHMM), Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Swarm Intelligence (SI), Ant Colony Optimization(ACO).

6. Future directions

Based on the review presented in this deliverable, several upcoming research directions relevant to the academic and professional community dealing with AI and railway transport have been identified as follows.

New data sources. Providing quality data has always been one of the main challenges in characterizing AI models. Therefore, collecting new and good quality data narrowed to specific tasks could help to improve existing and new models. To this aim, emerging ad-hoc wireless sensor networks such as the ones based on IoT devices can contribute to ever-increasing real-time streams of data coming from different sources. However, the challenge is to perform proper data management and filtering as not all data is suitable for a given task. In fact, with good quality data we mean that they must express relevant information to tackle the task under examination; large datasets, both in terms of features and entries, do not always lead to better performance. Possibly, publicly providing new standardized datasets specific for the railway sector, could lead to reliable benchmarks and help improve current and new research, such as ImageNet [253] or MS COCO [254] may be for specific computer vision applications. Similarly, in passenger mobility, incorporating or leveraging other data that could be retrieved in (near) real-time to predict the ridership can be addressed in future, e.g. the passenger count at station turnstiles [241]. Also, another case would be the data collected from the trials and real-life experiments, e.g. testing automated train operation at Thameslink in the UK [255] and at the freight-dedicated Betuweroute in the Netherlands [256]. Hitherto, the domain of autonomous driving and vehicle control mainly uses real-life infrastructure and theoretical engine characteristics to evaluate the performance of algorithms. However, the real-life running of vehicles could be impacted by multiple real-life factors, which need to be captured. Therefore, all the data that has been generated during real-world experiments should be valued and well-stored. Another challenge that may be arising is protecting the privacy of companies data (e.g. transport operators or maintenance companies) in open market railway system. To address it, new concepts like Federated Learning (also known as collaborative learning) [257] can become quite useful, to allow processing data locally at each company's environment.

Dealing with limited data. Next to data quality, the quantity is one of the crucial aspects. One way to achieve good performance even on "small" datasets could be using Transfer Learning, which provides transferring knowledge from one domain (or dataset) to another [258]. Otherwise, a sufficient amount of data should be collected for the specific task under examination to allow optimal characterisation of AI models. However, there are many cases in which obtaining certain types of data is not easy, for example accidents or failures of safety-critical systems, or quite impossible when it comes to rare events. In such cases, the most common procedures involve digital models (including digital twins) to generate synthetic data or data augmentation methods to generate them starting from those already available. Many techniques can be exploited to generate new data, some of which are already widely used such as resizing, rotating, and stretching images. However, there are also more powerful methods, such as the Data Augmentation Generative Adversarial Networks (DAGAN) [259], which are able to automatically generate suitable and optimal synthetic data.

Dealing with uncertainty. There are many uncertainties and gaps within machine learning approaches, such as the uncertainties arising from data collection in a specific domain, the variances in a specific type of data, and the imperfection of the models. Those uncertainties and gaps in the data are often difficult to be addressed using traditional techniques, such as a complex non-linear relationship between the variables (e.g. internal and external characteristics of the railway system) and the target (passenger flow) or train trajectory optimization with parametric uncertainty. To this end, AI can leverage the probability to harness and handle these uncertainties and model a relationship between the cause and effect of different real-life scenarios by combining the available data with assumptions (Denoeux et al. (2020) [260]).

Trustworthy and Explainable AI. Trustworthy and explainability, together with ethic regulations, represent some of the main topics that are rising more concerns about AI. As already introduced in [9], according to the guidelines introduced by the AI High-Level Expert Group (HLEG) [261], to achieve trustworthiness AI must be: *lawful*, i.e it must respect all applicable laws, norms, and regulations; *ethical*, i.e. it should respect ethical principles and values; *robust*, from both the technical perspective and taking into account its social environment. Also, these guidelines include four ethical principles on which AI must be based and seven key requirements that AI systems must meet to achieve trustworthiness. These lead us to delineate four main pillars that must be considered when dealing with AI in the railways: the final decisions shall be left to people (human-in-the-loop, human-on-the-loop), it should be fail-safe, the command and responsibility chain should be reconstructable (accountability), and it shall benefit human beings. Also, eXplainable AI (XAI) [262] is becoming ever important as many AI systems result to be too complex to be properly understood by humans; therefore, XAI approaches and methods are necessary to make the reasoning process and the outputs understandable by human operators. Looking at the explainability, there is the necessity to understand the behaviour of the AI systems across the whole railway transport. However, XAI has not received practical attentions in the rail sector. The only contribution can be found in [263], where the authors tackled the problem of discerning different reasons for the occurrence of train delays. In particular, methods from XAI help to classify to which amount the primary and secondary features contribute to a specific prediction of the model. For other domains, a comprehensive review of XAI in various business and industry sectors is given in [264], where case studies are reported in recommendation systems, sales, lending, and fraud detection; while, [265] focuses on Supply Chain Brain discussing XAI issues in that area. These can be used to build on and define important aspects of XAI for railways. In addition, it is also worth mentioning that there are some existing tools and frameworks for assessing the trustworthiness of AI systems (e.g. The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment [266]), deal with the non-explainability of models (e.g. LIME [267]), and improve safety, robustness, and confidentiality of system's outputs/predictions (e.g. SafeML [268])

Ethics of AI. Ethics and regulatory/legislative interventions (e.g. European Parliament's General Data Protection Regulation (GDPR) [269]) tend to be more sensitive concern in some railway subdomains than others, i.e. AI could have different impacts on humans' wellness or data privacy. A major impact can be expected where the safety of people could be directly affected such as the safety and security and the autonomous driving and control subdomains [270]. In addition, minor ethical concerns arise from applying AI to passenger mobility tasks (e.g. flow prediction) or staff/workers activities (e.g. scheduling) as sensitive personal data might be involved (e.g. images/videos in case of computer vision applications). Ethics regulations would be required especially in the safety-critical subdomains as many significant questions arise such as: What is the right decision for an AI system to mitigate the effect of false positives? And, what is the right decision for the same system when an animal or a road vehicle is detected on the track? Lessons on ethical issues of AI can be learnt looking at other sectors such as healthcare (e.g. [271]), robotics (e.g. [272]), and especially road transport (e.g. [273], [274]). In particular, [273] highlights key ethical issues in the use of AI in automated driving; while [274] discusses the dangers of the Moral Machine (MM) experiment in Autonomous Vehicles, alerting against both its uses for normative ends and the whole approach it is built upon to address ethical issues.

Human-in-the-loop Machine Learning. A big progress for intelligent devices is that they are learning more from the human-encoded examples rather than following the instructions of hard-coded pragrammes. Human-in-the-loop machine learning is such a emerging tool that incorporating human opinions into machine learning process, for not only strategically improving the ceiling of performance for machine learning models but also reaching the expected accuracy faster (Monarch R. (2021)[275]). If we regard classical machine learning as a set of algorithms that teach computers how to achieve the desired prediction accuracy within a limited learning procedure, then human-in-the-loop learning is more concerned with how the unstructured data can be properly interpreted and thus more comprehensible data can be available for machines.

Annotation and Active Learning are two representative paradigm for human-computer interactions in Human-in-the-loop Machine Learning (Pandey et al. (2020)[276] . Annotation is the procedure of adding more label information to raw data (e.g. images without descriptions, speeches from the social media or texts with no sentiment judgements) so that it is available to be utilised as a part of training data for machine learning models (Monarch R. (2021)[275]). Active Learning can be seen as a process of determining which data to sample for human annotation (Budd et al. (2019)[277]). Both annotation and algorithms are two equally important and intertwined components for a well-functional machine learning model. In many cases, especially when the characteristics of the data is ever changing over time and places, annotation is commonly used as a complementary technique to improve model performance by labelling more training data (e.g. Sacha et al. (2017) [278]). Human annotation would be regraded as an effective practice to assess the ceiling of model performance because it changes the traditional pattern of adapting new algorithms to a new application scenario, but alternatively focuses on how to transfer efficient labelling strategies to an existing model and hence create more available data.

Sometimes, finding these training items may not be a critical task but how fast we getting the unlabelled data out matters, especially in some cases require swift response speed or in a case of emergency that need to extract information from emerging disasters (extreme weather condition or massive of congestion). Active learning strategies can help with these situation where railway decision makers need to take actions within a limited time. Specifically, Active learning can be used to select the most suitable data for yielding optimal model performance. And then, using iterative feedback to steer models to optimal for a given prediction (e.g. Traffic flow prediction, predictive maintenance actions) and offering understandable ways to interpret and respond to predictions. For example, incorporating different formatted sensory data from multiple on-board sensors and choosing the best ones might be a challenging task without Active Learning in the study of Zhang et al. (2015) [279]).

Computer Vision for safety and security. Computer Vision (CV) includes powerful methods and models to analyse the environment and act in response to its mutation. Both projects and scientific papers have given numerous contributions to the application of computer vision for maintenance and inspection purposes. However, CV applications, from motion tracking to intrusion (video surveillance) and obstacle detection, could also have a great impact on safety, security, and passenger crowd characterization. In this direction, already known hot topics have been identified such as intrusion/obstacle detection on the tracks (e.g. GoAFE RAIL[43], SMART[1], SMART2[41]) and video surveillance or crowd analysis at railway stations (e.g. FAIR STATIONS[24]). However, other interesting ideas emerged from our report, are fare evasion detection in stations (e.g. TRAINSFARE[63]), face recognition for the entry verification system ([96]), and the improvement of workers' safety during inspection activities through IoT devices and computer vision (e.g. ZIMASS[69]). These topics, together with others such as video surveillance of the environment around tracks and level crossings and the prediction of passengers and rolling stocks behaviour, could be interesting applications for further investigations.

Applying NLP to non-structural data. Recently, NLP strategies have been widely utilised to optimize the process of transforming unstructured or hybrid information into a well-formatted one, especially for those from internal reports and social media platform. That is, each incident occurs at the a specific place with its particular reasons. And all the information need to be recorded by accidents logs, integrated through internal communication, and stored with structured natural language. NLP techniques are able to effectively extract essential knowledge from these reports and messages and make it available for the further modelling process.

For example, NLP can be used for maintenance. Originally, historical maintenance records presented unstructured or semi-structured documents, As such, these records can be successfully processed by NLP to determine the most critical components, which can further lead to determining optimised maintenance strategies (Edwards et al. (2017)[280]). For example, Runeson et al. (2007)[281] used NLP to detect duplicate defect reports at Sony Ericsson Mobile Communications. Also, Serna et al.(2018)[282] discussed the potential in applying big data and text mining technologies from social media to help policy makers in transport analysis and policy making, including NLP as a powerful tool for text mining and analysis. This article is about generic transport policy making, and there is no reason that railways, as an important sector of transport, would be excluded from this potential direction. For instance, by analysing the information extracted from train operators' official websites and tweets, and the tweets from customers, policy makers can respond accordingly. In addition, NLP could be exploited in passenger mobility to e.g. monitor and evaluate passenger satisfaction with the operations.

AI in decision making problems. Machine learning and deep learning are not the process of decision-making itself, while they can be incorporated to support the optimization of larger and more complex combinatorial models, such as network planning, crew scheduling, locomotive fleet sizing (e.g. [283], [284]). Further, one of the limitations of bio-inspired approaches (e.g. EA) is that they do not guarantee they will find the optimal solution (decided by the nature of them). That is, the overall quality of solutions cannot be granted, and it has to be benchmarked against other exact optimization methods, such as mathematical-based models. There has been an impressive advance in solving combinatorial optimization problems by mathematical programming and machine learning [285]. This implies that there is a great potential in solving railway planning and scheduling problems using AI-aided optimization approaches given the fast-growing research interests in the theoretical optimization community.

Multi-agent systems & Negotiations.

In particular, self-organizing system is a potential solution for addressing the conflicts between Infrastructure Manager and Train service operators is the core interaction in current railway open market. As a specific application of Distribution Artificial Intelligence, it aims to improve flexibility, capacity and resilience of the railway system as a mobility backbone, to accomplish an efficient and demand-aware urban and interurban rail mobility growth. For example, Cesme and Furth (2014)[286] explored a new paradigm for intelligent traffic signal control — 'self-organizing signals', based on dynamic coordination rules within a group of closely interacted agents and a simulation test was conducted on arterial road corridors in US. The result shows that overall delay has been reduced significantly. It is interesting to further discuss the feasibility of transferring such paradigm to the railway traffic signal system even the rescheduling problems. Additionally, some novel frameworks, such as multi-robotic system (Whitbrook et al. (2021)[287]) may be used to automate and accelerate the business processes both between IM and TSOs, and TSO and TSO where agents (e.g. trains, dispatchers) would properly interact and negotiate with other agents in order to coordinate their actions and strengthen dynamic online rescheduling in real time.

Intelligent Cybersecurity and AI. Like any other industrial sector using automated control systems, railways will have a technical and human challenge to fight against cyber-threats and protect their assets. On the one hand, digital devices and intelligent infrastructures are making railway system a more attractive transport option to passengers and more effective on supporting the decision making process when dealing with optimization problems (e.g. Liu et al. (2018) [213]). On the other hand, the intelligent networking of mobility collects a large amount of data that generated from sensors monitors and mobility apps with an ever-growing manner, and these data may contain massive quantities of sensitive and personal information. Cyber-attacks from hackers, criminal organizations or intelligence services are prone to undermine the security and integrity of these data. More seriously, cyber-manipulation not only poses a risk on the economy of individual/enterprise, but endangers public safety in cases of emergency (Choudhary and Nadim. (2018)[288]). Relevant studies and innovations in the context of AI-aided cybersecurity in railway sector have been launched recent years but it still requires more development (Kour et al. (2019)[289]). At this stage, fast-growing connections between railway system and its customers have been established in different services and applications—such as intelligent passenger information system (Gallo et al. (2019)[241]), predictive maintenance capabilities (Wang et al. (2018)[170]), enhancement of train punctuality(Barman et al. (2015)[220]) and improvement of track capacity (Xue et al. (2019)[236]). However, it is shown that conflicts between railway automation and data safety were not well-explained among the studies mentioned above. There is still a lot of work to be done with a further investigation towards systematically evaluating these relations brought by the rise of big data and the development of cybersecurity concerns. It could help train operators to better decide what cybersecurity capabilities they should pay more attention to as well as to provide better services by taking full advantages of the data. For example, Bayesian Network approaches can be adopted for quantitative threat assessment (Pappaterra and Flammini (2019)[290]). In addition, one can consider extending the conventional risk assessment method from a strategical perspective, in order to incorporate a wider set of procedures related to cybersecurity concerns (Mokalled et al. (2019)[291]).

Combined AI approaches. In many cases, AI-based applications have achieved good results in extracting information from data to detect common features, identify relationship between variables and target, and hence make predictions, pattern recognition and automation procedures, and so on. What is commonly done, as the first step, especially considering the Data Mining processes (data selection, data processing, data transformation, data mining, and model interpretation), is to extract the features and/or reduce the size of the dataset to be analyzed (e.g. through PCA). This data processing played a central role in the characterization of the model. However, what is also interesting, is combining image processing techniques to extract features from video data and then apply traditional ML techniques (e.g. regression models, Sysyn et al. (2019)[167]). Beyond that, multiple AI approaches can also be combined to better cope with a given task. Such approaches can be quite similar, as the series of classifiers implemented by Trinh et al. (2012)[185], or different, as presented by Fink et al.(2013)[147] who combined CRBM and ESN. Combining multiple approaches, where each of them focuses on a specific facet, can lead to a better characterization of the scenario and then to improved performances. Furthermore, deep learning is a black-box technique, which means it lacks explainability. Therefore, the combination of traditional machine learning and deep learning to better investigate the causality and utilize multi-source data could be very useful (Omta et al. (2020) [292]). Additionally, one can build AI models (such as deep learning ones) for automatic train control that learns offline (e.g. using historical train trajectories) and also adapts dynamically online (e.g. due to real-time conditions of engine, delays, weather, traffic) (Wang et al., 2020 [209]).

AI and revenue management. Revenue management is becoming an important topic in the railway industry, and shall gain more attention in the coming years. In addition, AI can be leveraged because of the newly available data. For example, the airline industry is one of the industries on which most studies have been done during the past decades, in order to analyze revenue maximization. AI can be used e.g. for ticket price prediction (Wang et al. (2019) [293]), or seat booking control (Shibab et al. (2019)[294]).

7. Conclusions

This report has provided a comprehensive review of research projects and scientific papers addressing the state-of-the-art of AI in the railway industry. The focus has been on European and overseas projects mainly in the United States and China. Specific emphasis was devoted to reviewing projects funded by the European Shift2Rail (S2R) Joint Undertaking, which represents one of the main funding bodies in EU for railway research and innovation. We addressed projects and papers from a holistic railway perspective, covering subdomains such as maintenance and inspection, safety and security, autonomous driving and control, transport planning and management, revenue management, transport policy and passenger mobility. As such, this report represents a first step towards the adoption of AI in the railway domain by providing an in-depth summary of the current research focus. In addition, we determined some promising research directions to provide further uptake of AI in railways.

We recognized that the majority of the projects worldwide addressing AI in railways is on maintenance and inspection (46%), followed by safety and security (20%), traffic planning and management (16%), passenger mobility (9%) autonomous driving and control (8%), and transport policy (1%, equal 1 project). Finally, no research on AI for railway revenue management has been found. In particular, within S2R projects, the share among the subdomains is similar, except no projects on transport policy exists. Some of the projects we surveyed are wide-spectrum, i.e. they focus on multiple purposes and use AI as one of the different possible enabling technologies to propose solutions; others have a rather narrow and specific focus and address a certain subdomain and/or a particular AI application to a railway problem. The review identified some interesting ongoing projects, e.g. on identifying objects on tracks (e.g. GoAFE RAIL[43], SMART[1], SMART2[41]) which could create multifaceted benefits towards increased safety as well as autonomous driving, on improving security at stations by identifying threats in advance (e.g. FAIR STATIONS[24], TRAINSFARE[63]), and improving workers safety during inspection activities (e.g. ZIMASS[69]).

Similarly to the reviewed projects, the main scientific research efforts published in technical papers[1] have been seen in the subdomain of maintenance and inspection (58%), which is followed by traffic planning and management (24%). The three subdomains, safety and security, autonomous driving and control and passenger mobility received only limited attention (each under 10%), while virtually no research exists on railway revenue management and transport policy based on AI. First, it is expected that AI will generate a strong impact on predicting future trends. For example, in railway passenger mobility, flow predictions have been investigated using AI. Instead, currently predictions take a rather limited portion of research in other subdomains, such as failure prediction in maintenance and inspection and traffic state prediction in real-time traffic management. Second, the use of intelligent systems tends to provide additional support in improving maintenance operations leading to increased safety. Third, AI can also support optimization models to tackle large-scale real-life problems in scheduling, traffic management, maintenance and inspection planning.

[1] Note that part of the research included in those papers got no support from funded projects, thus certain discrepancies may exist.

Fourth, AI applications can use a wide range of data from sensors, visual images/footage, traffic movement logs, which makes the possibilities of AI in railways extremely diverse. Overall, these observations suggest that AI in railways, although being at an early development stage, is attracting an increasing interest, and therefore it can be easily expected that more focus on AI-based research will characterize the future of railway engineering. Thus, many open research topics are envisioned, some of which could contribute to general usage of AI. Future research can be expected towards developing advanced combined AI applications, using AI in decision making, using AI to assist optimization approaches, dealing with uncertainty and tackling newly rising cyber-physical threats.

Compared with the surveyed projects, the RAILS project aims at developing roadmaps for the fast uptake of AI in the railway sector by identifying effective and suitable techniques, by testing AI methods in terms of proof-of-concept, and by assessing impacts towards improving overall performance of the railway system as a *whole*. As such, RAILS provides a platform to explore AI in railways within a holistic, unified and consistent framework, considering various dimensions and facets of the railway system as interrelated, rather than isolated, which is the case in most of the exiting projects. In addition, RAILS will focus on assessing related transport domains to determine promising approaches and on extending the state-of-the-art in several railway subdomains. In particular, building on the results of this report on the state-of-the-art of AI in railways, RAILS will address the specific topics of transport safety and autonomous driving (WP2), maintenance and inspection (WP3) and traffic planning and management (WP4), in reference use-case scenarios.

Bibliography

[1] Smart Automation of Rail Transport (SMART). [Online]. Available: http://www.smartrail-automation-project.net/

[2] My TRAvel Companion (My-TRAC). [Online]. Available: http://www.my-trac.eu/

[3] Intelligent Innovative Smart Maintenance of Assets by integRated Technologies (IN2SMART). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=IN2SMART

[4] Optimal Networks for Train Integration Management across Europe (ON-TIME). [Online]. Available: https://cordis.europa.eu/project/id/285243

[5] Increasing Capacity 4 Rail networks through enhanced infrastructure and optimised operations (CAPACITY4RAIL). [Online]. Available: http://www.capacity4rail.eu/

[6] R. Project, "Definition of a reference taxonomy of ai in railways," Tech. Rep., 2020. [Online]. Available: https://rails-project.eu/wp-content/uploads/sites/73/2020/08/RAILS_D11_v25.pdf

[7] M. Trosino, J. Cunningham, and A. Shaw, "Automated track inspection vehicle and method," March 2002, uS Patent 6,356,299.

[8] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 153–164, Jan. 2017.

[9] RAILS, "Deliverable d1.1 definition of a reference taxonomy of ai in railways," October 2020, https://rails-project.eu/wp-content/uploads/sites/73/2020/08/RAILS_D11_v25.pdf. DOI: 10.13140/RG.2.2.24887.75681.

[10] Cordis. [Online]. Available: https://cordis.europa.eu/EU

[11] Transportation Research Board's Research in Progress (RiP). [Online]. Available: https://rip.trb.org/

[12] China National Knowledge Infrastructure (CNKI). [Online]. Available: https://www.cnki.net

[13] S2R (Shift2Rail). [Online]. Available: https://shift2rail.org/

[14] Innovative RUNning gear soluTiOns for new dependable, sustainable, intelligent and comfortable RAIL vehicles (RUN2RAIL). [Online]. Available: http://www.run2rail.eu/home.aspx

[15] Cybersecurity in the RAILway sector (CYRAIL). [Online]. Available: https://cyrail.eu/

[16] Start-up activities for Advanced Signalling and Automation Systems (X2RAIL-1). [Online]. Available: https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-1

[17] Enhancing railway signalling systems based on train satellite positioning, on-board safe train integrity, formal methods approach and standard interfaces, enhancing Traffic Management System functions (X2RAIL-2). [Online]. Available: https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-2

[18] Advanced Signalling, Automation and Communication System (IP2 and IP5) – Prototyping the future by means of capacity increase, autonomy and flexible

communication (X2RAIL-3). [Online]. Available: https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-3
[19] Advanced signalling and automation system - Completion of activities for enhanced automation systems, train integrity, traffic management evolution and smart object controllers (X2RAIL-4). [Online]. Available: https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-4
[20] Measuring, monitoring and data handling for railway assets; bridges, tunnels, tracks and safety systems (ASSET4RAIL). [Online]. Available: http://www.assets4rail.eu/
[21] L. Stanislav, J. Veljko, and J. Uroš, "Infrastructure monitoring data handling with a hellp of bim."
[22] L. Stanislav, J. Veljko, J. Uroš, and A. Andrej, "Bim based integrated road assessment management."
[23] ASSETS4RAIL, "Deliverable d11.1 - plan for exploitation and dissemination of results (pedr)," 2018.
[24] Future Secure and Accessible Rail Stations (FAIRSTATIONS). [Online]. Available: http://www.fairstations.eu/
[25] INtelligent solutions 2ward the Development of Railway Energy and Asset Management Systems in Europe (IN2DREAMS). [Online]. Available: http://www.in2dreams.eu/
[26] Intelligent Innovative Smart Maintenance of Assets by integRated Technologies 2 (IN2SMART2). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=IN2SMART2
[27] Innovative Solutions in Future Stations, Energy Metering and Power Supply (IN2STEMPO). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=IN2stempo
[28] Research into enhanced track and switch and crossing system 2 (IN2TRACK2). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=IN2TRACK2
[29] Research into enhanced tracks, switches and structures (IN2TRACK). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=IN2TRACK
[30] Multi-scale Observation and Monitoring of railway Infrastructure Threats (MOMIT). [Online]. Available: https://www.momit-project.eu/
[31] Advanced Travel Companion and Tracking Services (ATTRAkTIVE). [Online]. Available: https://projects.shift2rail.org/s2r_ip4_n.aspx?p=ATTRACKTIVE
[32] Real time information applications and energy efficient solutions for rail freight (FR8HUB). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FR8HUB
[33] Optimised Real-time Yard and Network Management (OPTIYARD). [Online]. Available: https://optiyard.eu/
[34] Development of Functional Requirements for Sustainable and Attractive European Rail Freight (FR8RAIL). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FR8RAIL
[35] Digitalization and Automation of Freight Rail (FR8RAIL II). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FR8RAIL%20ii

[36] Smart data-based assets and efficient rail freight operation (FR8RAIL III). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FR8RAIL%20iii
[37] FR8RAIL, “Deliverable d2.5 – defined threshold, pre-warning threshold and opening value,” 2019.
[38] Future Freight Loco for Europe (FFL4E). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FFL4E
[39] INNOvative monitoring and predictive maintenance solutions on lightweight WAGon (INNOWAG). [Online]. Available: http://newrail.org/innowag/
[40] Locomotive bOgie Condition mAinTEnance (LOCATE). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=S2R_LOCATE
[41] Advanced integrated obstacle and track intrusion detection system for smart automation of rail transport Danijela Ristic-durrant (SMART2). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=S2R_SMART2
[42] Translation for breaking language barriers in the railway field (TRANSLATE4RAIL). [Online]. Available: https://translate4rail.eu/
[43] Global Safety Management Framework for RAIL Operations (GoSAFE RAIL). [Online]. Available: http://www.gosaferail.eu/
[44] Indicator Monitoring for a new railway PAradigm in seamlessly integrated Cross modal Transport chains – Phase 2 (IMPACT-2). [Online]. Available: https://cordis.europa.eu/project/id/777513
[45] Indicator Monitoring for a new railway PAradigm in seamlessly integrated Cross modal Transport chains – Phase 1 (IMPACT-1). [Online]. Available: https://cordis.europa.eu/project/id/730816
[46] IMPACT-2, “Deliverable d6.5 - smart maintenance concept,” it is still in a draft/issued status (accessed 2020-11-15).
[47] IMPACT-2, “Deliverable D6.4 - CBM results for rail vehicles,” 2019.
[48] Innovative Intelligent Rail (IN2RAIL). [Online]. Available: http://www.in2rail.eu/
[49] Modelling and strategies for the assessment and OPtimisation of Energy USage aspects of rail innovation (OPEUS). [Online]. Available: http://www.opeus-project.eu/
[50] OPEUS, “Deliverable d 4.2 - driving strategies and energy management for das,” 2019.
[51] Smart Maintenance and the Rail Traveller Experience (SMaRTE). [Online]. Available: http://www.smarte-rail.eu/
[52] SMaRTE investigates innovations in smart rail maintenance and passenger satisfaction. [Online]. Available: https://cordis.europa.eu/article/id/418247-smarte-investigates-innovations-in-smart-rail-maintenance-and-passenger-satisfaction
[53] RECET4rail Reliable Energy and Cost Efficient Traction system for Railway (RECET4RAIL). [Online]. Available: https://projects.shift2rail.org/s2r_ip1_n.aspx?p=S2R_RECET4rail
[54] Development of prescriptive AnalYtics baseD on aRtificial intElligence for iAMS (DAYDREAMS). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=S2R_DAYDREAMS
[55] The Next Generation of Railway Transition Zones (IN2ZONE). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=S2R_IN2ZONE

[56] Smart Tools for Railway work safEty and performAnce iMprovement (STREAM). [Online]. Available: https://projects.shift2rail.org/s2r_ip3_n.aspx?p=S2R_STREAM
[57] Use-centric rail freight innovation for Single European Railway Area (FR8RAIL IV). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FR8RAIL%20iv
[58] Technologies for the AUtonomous Rail Operation (TAURO). [Online]. Available: https://projects.shift2rail.org/s2r_ipx_n.aspx?p=tauro
[59] System Architecture and Conceptual Data Model For Railway, Common Data Dictionary And Global System Modelling Specifications (LINX4RAIL2). [Online]. Available: https://projects.shift2rail.org/s2r_ipx_n.aspx?p=LINX4RAIL2
[60] System architecture and Conceptual Data Model for railway, common data dictionary and global system modelling specifications (LINX4RAIL). [Online]. Available: https://projects.shift2rail.org/s2r_ipx_n.aspx?p=LINX4RAIL
[61] Completion of activities for Adaptable Communication, Moving Block, Fail safe Train Localisation (including satellite), Zero on site Testing, Formal Methods and Cyber Security (X2RAIL-5). [Online]. Available: https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-5#
[62] Inspection Drones for Ensuring Safety in Transport Infrastructures (Drones4Safety). [Online]. Available: https://drones4safety.eu/
[63] TRAINSFARE: Smart Tool to Protect Public Transport Revenues, Assets, Passengers and Mobility. [Online]. Available: https://www.trainsfare.eu/
[64] Transforming transport (tt). [Online]. Available: https://transformingtransport.eu/
[65] A.I. enabled knowledge analysis automation to increase resilience, security and performance of Enterprise ICT systems (RAIKA). [Online]. Available: https://cordis.europa.eu/project/id/874184
[66] Runecast A.I. Knowledge Automation (RAIKA). [Online]. Available: https://ec.europa.eu/eipp/desktop/en/projects/project-11053.html
[67] Runecast. [Online]. Available: https://www.runecast.com/
[68] Safety of Transport Infrastructure on the TEN-T Network (SAFE-10-T). [Online]. Available: https://www.safe10tproject.eu/
[69] ZIMASS: Smart Mobile awareness and safety system for workzone invasion. [Online]. Available: https://cordis.europa.eu/project/id/874184
[70] Predictive Maintenance for railway switches. Smart sensor networks on a machine learning analytics platform (ANDROMEDA). [Online]. Available: https://www.konux.com/
[71] CAPACITY4RAIL, "Deliverable 42.2: Recommendations and guidelines for next generation monitoring and inspection," 2017. [Online]. Available: http://www.capacity4rail.eu/results
[72] ——, "Deliverable 33.2: Recommendations for a European standard for traffic management under large disruptions," 2017. [Online]. Available: http://www.capacity4rail.eu/results
[73] ——, "Final publishable summary report (1 october 2013 to 30 september 2017)," 2017. [Online]. Available: https://cordis.europa.eu/project/id/605650/reporting
[74] U.S. Department of Transportation's (USDOT) University Transportation Centers

(UTC) Program. [Online]. Available: https://www.transportation.gov/content/university-transportation-centers
[75] Application of machine learning techniques toward time-based change in track condition using an onboard sensor in revenue-service rolling stock. [Online]. Available: https://rip.trb.org/View/1604536
[76] Artificial intelligence for advance landslide warning along railroad tracks in pennsylvania and delaware. [Online]. Available: https://r3utc.psu.edu/research/current-research-projects/active-projects-year-2-3/warning-along-railroad-tracks/
[77] Asset Health Strategic Initiative (AHSI) - Technology-Driven Train Inspection. [Online]. Available: https://public.railinc.com/about-railinc/aar-industry-projects
[78] Automated track geometry monitoring system. [Online]. Available: https://www.mountain-plains.org/research/details.php?id=448
[79] Developing a safety management system including hazardous materials for highway-rail grade crossings in region vii. [Online]. Available: https://rip.trb.org/View/1740394
[80] Estimation of Origin-Destination Matrix and Identification of User Activities Using Public Transit Smart Card Data. [Online]. Available: https://rip.trb.org/View/1455346
[81] W. Fan, Z. Chen, and P. Director, "Estimation of origin-destination matrix and identification of user activities using public transit smart card data," *Center for Advanced Multimodal Mobility Solutions and Education*, 2018.
[82] Laser-based non-destructive spike defect inspection system. [Online]. Available: https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=5065
[83] Rail safety idea (innovations deserving exploratory analysis) program. [Online]. Available: http://www.trb.org/IDEAProgram/IDEASafety.aspx
[84] Passenger flow modeling and simulation in transit stations. [Online]. Available: https://cait.rutgers.edu/research/passenger-flow-modeling-and-simulation-in-transit-stations/
[85] Principal components analysis and track quality index: A machine learning approach. [Online]. Available: https://www.unlv.edu/railteam/research/principal-components-analysis
[86] A. Lasisi and N. Attoh-Okine, "Principal components analysis and track quality index: A machine learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 91, pp. 230 – 248, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X18304303
[87] Video-sensor data fusion for enhanced structural monitoring. [Online]. Available: https://r3utc.psu.edu/research/core-research-projects/ciam-cor-r25/
[88] Vibration-based longitudinal rail stress estimation exploiting acoustic measurement and machine learning. [Online]. Available: https://rip.trb.org/View/1627383
[89] J. S. Popovics, M. S. Dersch, and X. Zhu, "Vibration-based longitudinal rail stress estimation exploiting acoustic measurement and machine learning," Tech. Rep., 2020. [Online]. Available: http://www.trb.org/Main/Blurbs/181453.aspx
[90] Augmented reality enhancing the inspections of transportation infrastructure: Research, education, and industry implementation. [Online]. Available: https://rip.trb.org/View/1505456

[91] F. Moreu, C. Lippitt, D. Maharjan, M. Aguero, and X. Yuan, “Augmented reality enhancing the inspections of transportation infrastructure: Research, education, and industry implementation,” Tech. Rep., 2019.

[92] Development and validation of a new generation rail wear model using emerging big-data analytic techniques. [Online]. Available: https://rip.trb.org/View/1475691

[93] Major scientific research fund project of china railway corporation. [Online]. Available: http://gb.oversea.cnki.net/KCMS/detail/detail.aspx?filename=TDBS201807005&dbcode=CJF&dbname=CJFD2018

[94] P. Yan, Y. Liao, W. G. Chen, X. J. Liu, and C. W. Yang, “Application of intelligent image recognition technology for detection in hsr infrastructure,” *CHINA Railway*, pp. 109–112, 11 2019.

[95] Hebei science and technology plan fund project. [Online]. Available: http://gb.oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=WJFZ201705033&DbName=CJFDLAST2017&DbCode=CJFD

[96] China railway corporation technology research and development project. [Online]. Available: http://gb.oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=TLJS201807013&DbName=CJFDLAST2018&DbCode=CJFD

[97] Sichuan science and technology research and development project. [Online]. Available: http://gb.oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=TLZG201911019&DbName=CJFDLAST2019&DbCode=CJFD

[98] E. Mathew, “Intelligent transport systems and its challenges,” in *International Conference on Advanced Intelligent Systems and Informatics*. Springer, 2019, pp. 663–672.

[99] L. A. Shlain, A. Gayathri, and G. Aiswarya, “Modeling transport pattern using artificial intelligence,” in *International Conference on Intelligent Data Communication Technologies and Internet of Things*. Springer, 2019, pp. 554–560.

[100] F. Zantalis, G. Koulouras, S. Karabetsos, and D. Kandris, “A review of machine learning and iot in smart transportation,” *Future Internet*, vol. 11, no. 4, 2019.

[101] A. R. Patel, K. K. Ramaiya, C. V. Bhatia, H. N. Shah, and S. N. Bhavsar, “Artificial intelligence: Prospect in mechanical engineering field—a review,” in *Data Science and Intelligent Applications*, K. Kotecha, V. Piuri, H. N. Shah, and R. Patel, Eds. Singapore: Springer Singapore, 2021, pp. 267–282.

[102] R. Toorajipour, V. Sohrabpour, A. Nazarpour, P. Oghazi, and M. Fischl, “Artificial intelligence in supply chain management: A systematic literature review,” *Journal of Business Research*, vol. 122, pp. 502 – 517, 2021.

[103] Y.-J. Cai and C. K. Lo, “Omni-channel management in the new retailing era: A systematic review and future research agenda,” *International Journal of Production Economics*, vol. 229, p. 107729, 2020.

[104] S. Hosseini, D. Ivanov, and A. Dolgui, “Review of quantitative methods for supply chain resilience analysis,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 285 – 307, 2019.

[105] F. Ghofrani, Q. He, R. M. Goverde, and X. Liu, “Recent applications of big data analytics in railway transportation systems: A survey,” *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 226–246, 2018.

[106] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, “Big data analytics in intelligent trans-

portation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2018.

[107] Y. Jain *et al.*, "A survey on railway assets: A potential domain for big data," in *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, vol. 1. IEEE, 2019, pp. 1–6.

[108] C. Wen, P. Huang, Z. Li, J. Lessan, L. Fu, C. Jiang, and X. Xu, "Train dispatching management with data- driven approaches: A comprehensive review and appraisal," *IEEE Access*, vol. 7, pp. 114 547–114 571, 2019.

[109] Q. Wu, C. Cole, and T. McSweeney, "Applications of particle swarm optimization in the railway domain," *International Journal of Rail Transportation*, vol. 4, no. 3, pp. 167–190, 2016.

[110] N. Thilagavathy, J. Harene, M. Sherine, and T. Shanmugasundari, "Survey on railway wheel defect detection using machine learning."

[111] M. C. Nakhaee, D. Hiemstra, M. Stoelinga, and M. van Noort, "The recent applications of machine learning in rail track maintenance: A survey," in *International Conference on Reliability, Safety, and Security of Railway Systems*. Springer, 2019, pp. 91–105.

[112] S. Shalini and A. A. Narasimham, "A review on rail accidents and predictions using data mining techniques," *International Journal of Research*, vol. 4, pp. 2094–2098, 2017.

[113] N. Bešinović, "Resilience in railway transport systems: a literature review and research agenda," *Transport Reviews*, vol. 40, no. 4, pp. 457–478, 2020.

[114] F. Corman and L. Meng, "A review of online dynamic models and algorithms for railway traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1274–1284, 2015.

[115] G. M. Scheepmaker, R. M. Goverde, and L. G. Kroon, "Review of energy-efficient train control and timetabling," *European Journal of Operational Research*, vol. 257, no. 2, pp. 355 – 376, 2017.

[116] R. Lusby, J. Larsen, M. Ehrgott, and D. Ryan, "Railway track allocation: Models and methods," *OR Spectrum*, vol. 33, pp. 843–883, 10 2011.

[117] W. Fang, S. Yang, and X. Yao, "A survey on problem models and solution approaches to rescheduling in railway networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 2997–3016, 2015.

[118] S. Liu, Q. Wang, and Y. Luo, "A review of applications of visual inspection technology based on image processing in the railway industry," *Transportation Safety and Environment*, vol. 1, no. 3, pp. 185–204, 2019.

[119] X. Gibert, V. M. Patel, and R. Chellappa, "Robust fastener detection for autonomous visual railway track inspection," in *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 694–701.

[120] P. Xie, T. Li, J. Liu, S. Du, X. Yang, and J. Zhang, "Urban flow prediction from spatiotemporal data using machine learning: A survey," *Information Fusion*, vol. 59, pp. 1 – 12, 2020.

[121] A. Thaduri, D. Galar, and U. Kumar, "Railway assets: A potential domain for big data analytics," *Procedia Computer Science*, vol. 53, pp. 457 – 467, 2015, iNNS Conference on Big Data 2015 Program San Francisco, CA, USA 8-10 August 2015.

[122] H. Alawad and S. Kaewunruen, “Wireless sensor networks: Toward smarter railway stations,” *Infrastructures*, vol. 3, no. 3, p. 24, 2018.

[123] U. Dissemination, “The database of abstracts of reviews of effects (dare),” 12 2002.

[124] O. F. Eker, F. Camci, A. Guclu, H. Yilboga, M. Sevkli, and S. Baskan, “A simple state-based prognostic model for railway turnout systems,” *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1718–1726, 2011.

[125] H. Guler, “Decision support system for railway track maintenance and renewal management,” *Journal of Computing in Civil Engineering*, vol. 27, no. 3, pp. 292–306, 2013.

[126] L.-Q. Hu, C.-F. He, Z.-Q. Cai, L. Wen, and T. Ren, “Track circuit fault prediction method based on grey theory and expert system,” *Journal of Visual Communication and Image Representation*, vol. 58, pp. 37–45, 2019.

[127] A. Jamshidi, A. Núñez, R. Dollevoet, and Z. Li, “Robust and predictive fuzzy key performance indicators for condition-based treatment of squats in railway infrastructures,” *Journal of Infrastructure Systems*, vol. 23, no. 3, p. 04017006, 2017.

[128] P. Jayaswal, S. Verma, and A. Wadhwani, “Development of ebp-artificial neural network expert system for rolling element bearing fault diagnosis,” *Journal of Vibration and Control*, vol. 17, no. 8, pp. 1131–1148, 2011.

[129] A. H. E. Khouzani, A. Golroo, and M. Bagheri, “Railway maintenance management using a stochastic geometrical degradation model,” *Journal of Transportation Engineering, Part A: Systems*, vol. 143, no. 1, p. 04016002, 2017.

[130] G. Li, S. Zhang, and Y. Yin, “An expert system for remote monitoring and fault diagnosis of railway locomotives,” in *ICTE 2011*, 2011, pp. 2959–2964.

[131] M. Pu and H. Wang, “Fault diagnosis expert system for large-scale railway maintenance equipment based on bp neural network,” *Computer Modelling & New Technologies*, vol. 18, no. 11, pp. 1068–1072, 2014.

[132] E. Fumeo, L. Oneto, and D. Anguita, “Condition based maintenance in railway transportation systems based on big data streaming analysis.” in *INNS Conference on Big Data*, 2015, pp. 437–446.

[133] W. Marsh, K. Nur, B. Yet, and A. Majumdar, “Using operational data for decision making: a feasibility study in rail maintenance,” in *Safety and Reliability*, vol. 36, no. 1. Taylor & Francis, 2016, pp. 35–47.

[134] D. Ou, R. Xue, and K. Cui, “A data-driven fault diagnosis method for railway turnouts,” *Transportation research record*, vol. 2673, no. 4, pp. 448–457, 2019.

[135] J. Rabatel, S. Bringay, and P. Poncelet, “Anomaly detection in monitoring sensor data for preventive maintenance,” *Expert Systems with Applications*, vol. 38, no. 6, pp. 7003–7015, 2011.

[136] W. Sammouri, E. Côme, L. Oukhellou, P. Aknin, and C. FONLLADOSA, “Floating train data systems for preventive maintenance: A data mining approach,” *Proceedings of 2013 International Conference on Industrial Engineering and Systems Management, IEEE - IESM 2013*, 10 2013.

[137] S. Sharma, Y. Cui, Q. He, R. Mohammadi, and Z. Li, “Data-driven optimization of railway maintenance for track geometry,” *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 34–58, 2018.

[138] Z. Allah Bukhsh, A. Saeed, I. Stipanovic, and A. G. Doree, “Predictive maintenance using tree-based classification techniques: A case of railway switches,” *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 35 – 54, 2019.

[139] Z. Cherfi, L. Oukhellou, E. Côme, T. Denœux, and P. Aknin, “Partially supervised independent factor analysis using soft labels elicited from multiple experts: Application to railway track circuit diagnosis,” *Soft Computing*, vol. 16, 05 2012.

[140] T. de Bruin, K. Verbert, and R. Babuška, “Railway track circuit fault diagnosis using recurrent neural networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 523–533, 2017.

[141] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter, “Deep convolutional neural networks for detection of rail surface defects,” in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 2584–2589.

[142] S. Famurewa, L. Zhang, and M. Asplund, “Maintenance analytics for railway infrastructure decision support,” *Journal of Quality in Maintenance Engineering*, vol. 23, pp. 00–00, 07 2017.

[143] B. Firlik and M. Tabaszewski, “Monitoring of the technical condition of tracks based on machine learning,” *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 234, no. 7, pp. 702–708, 2020.

[144] S. Gao, T. Szugs, and R. Ahlbrink, “Use of combined railway inspection data sources for characterization of rolling contact fatigue,” 06 2018.

[145] X. Giben, V. M. Patel, and R. Chellappa, “Material classification and semantic segmentation of railway track images with deep convolutional neural networks,” in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 621–625.

[146] X. Gibert, V. M. Patel, and R. Chellappa, “Deep multitask learning for railway track inspection,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 153–164, 2017.

[147] O. Fink, E. Zio, and U. Weidmann, “Predicting time series of railway speed restrictions with time-dependent machine learning techniques,” *Expert Systems with Applications*, vol. 40, no. 15, pp. 6033 – 6040, 2013.

[148] H. Guler, “Optimisation of railway track maintenance and renewal works by genetic algorithms,” *Journal of the Croatian Association of Civil Engineers*, vol. 68, pp. 979–993, 12 2016.

[149] S. Hajizadeh, A. Núñez, and D. M. Tax, “Semi-supervised rail defect detection from imbalanced image data,” *IFAC-PapersOnLine*, vol. 49, no. 3, pp. 78 – 83, 2016, 14th IFAC Symposium on Control in Transportation SystemsCTS 2016.

[150] Y. Han, Z. Liu, Y. Lyu, K. Liu, C. Li, and W. Zhang, “Deep learning-based visual ensemble method for high-speed railway catenary clevis fracture detection,” *Neurocomputing*, vol. 396, pp. 556 – 568, 2020.

[151] C. Hu and X. Liu, “Modeling track geometry degradation using support vector machine technique,” 04 2016, p. V001T01A011.

[152] A. Jamshidi, S. Hajizadeh, Z. Su, M. Naeimi, A. Núñez, R. Dollevoet, B. De Schutter, and Z. Li, “A decision support approach for condition-based maintenance of rails based on big data analysis,” *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 185 – 206, 2018.

[153] Y. Jiang, H. Wang, G. Tian, Q. Yi, J. Zhao, and K. Zhen, “Fast classification for rail defect depths using a hybrid intelligent method,” *Optik*, vol. 180, pp. 455 – 468, 2019.

[154] G. Kang, S. Gao, L. Yu, and D. Zhang, “Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2679–2690, 2019.

[155] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, “Wheel defect detection with machine learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1176–1187, 2018.

[156] A. Lasisi and N. Attoh-Okine, “Principal components analysis and track quality index: A machine learning approach,” *Transportation Research Part C: Emerging Technologies*, vol. 91, pp. 230 – 248, 2018.

[157] J. S. Lee, S. H. Hwang, I. Y. Choi, and I. K. Kim, “Prediction of track deterioration using maintenance data and machine learning schemes,” *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 9, p. 04018045, 2018.

[158] H. Li, D. Parikh, Q. He, B. Qian, Z. Li, D. Fang, and A. Hampapur, “Improving rail network velocity: A machine learning approach to predictive maintenance,” *Transportation Research Part C: Emerging Technologies*, vol. 45, pp. 17 – 26, 2014, advances in Computing and Communications and their Impact on Transportation Science and Technologies.

[159] S. Ritika and D. Rao, “Data augmentation of railway images for track inspection,” 2018.

[160] J. Sadeghi and H. Askarinejad, “Application of neural networks in evaluation of railway track quality condition,” *Journal of Mechanical Science and Technology*, vol. 26, 01 2012.

[161] Y. Santur, M. Karaköse, and E. Akin, “Learning based experimental approach for condition monitoring using laser cameras in railway tracks,” *International Journal of Applied Mathematics, Electronics and Computers*, vol. 4, pp. 1–1, 12 2016.

[162] Y. Santur, M. Karaköse, and E. Akin, “A new rail inspection method based on deep learning using laser cameras,” in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017, pp. 1–6.

[163] Y. Santur, M. KARAKÖSE, and E. Akin, “An adaptive fault diagnosis approach using pipeline implementation for railway inspection,” *TURKISH JOURNAL OF ELECTRICAL ENGINEERING COMPUTER SCIENCES*, vol. 26, pp. 987–998, 03 2018.

[164] L. Shang, Q. Yang, J. Wang, S. Li, and W. Lei, “Detection of rail surface defects based on cnn image recognition and classification,” in *2018 20th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2018, pp. 45–51.

[165] A. Shebani and S. Iwnicki, “Prediction of wheel and rail wear under different contact conditions using artificial neural networks,” *Wear*, vol. 406-407, pp. 173 – 184, 2018.

[166] D. Soukup and R. Huber-Mörk, “Convolutional neural networks for steel surface defect detection from photometric stereo images,” in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. McMahan, J. Jerald, H. Zhang, S. M. Drucker, C. Kambhamettu, M. El Choubassi, Z. Deng, and M. Carlson, Eds. Cham: Springer International Publishing, 2014, pp. 668–677.

[167] M. Sysyn, U. Gerber, O. Nabochenko, D. Gruen, and F. Kluge, “Prediction of rail con-

tact fatigue on crossings using image processing and machine learning methods," *Urban Rail Transit*, vol. 5, pp. 123–132, 2019.

[168] M. Sysyn, D. Gruen, U. Gerber, O. Nabochenko, and V. Kovalchuk, "Turnout monitoring with vehicle based inertial measurements of operational trains: A machine learning approach," *Communications - Scientific Letters of the University of Zilina*, vol. 21, pp. 42–48, 02 2019.

[169] H. Tsunashima, "Condition monitoring of railway tracks from car-body vibration using a machine learning technique," *Applied Sciences*, vol. 9, p. 2734, 07 2019.

[170] P. Dai, X. Du, S. Wang, Z. Gu, and Y. Ma, "Rail fastener automatic recognition method in complex background," 08 2018, p. 314.

[171] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, and Y. Li, "Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study," *Engineering Applications of Artificial Intelligence*, vol. 80, pp. 66 – 81, 2019.

[172] Y. Xia, F. Xie, and Z. Jiang, "Broken railway fastener detection based on adaboost algorithm," in *2010 International Conference on Optoelectronics and Image Processing*, vol. 1, 2010, pp. 313–316.

[173] X. Xu, Y. Lei, and F. Yang, "Railway subgrade defect automatic recognition method based on improved faster r-cnn," *Scientific Programming*, vol. 2018, pp. 1–12, 06 2018.

[174] J. Yin and W. Zhao, "Fault diagnosis network design for vehicle on-board equipments of high-speed railway: A deep learning approach," *Engineering Applications of Artificial Intelligence*, vol. 56, pp. 250 – 259, 2016.

[175] Z. Zhang, K. Zhou, and X. Liu, "Broken rail prediction with machine learning-based approach," ser. ASME/IEEE Joint Rail Conference, vol. 2020 Joint Rail Conference, 04 2020, v001T08A014.

[176] F. Zhou, L. Xia, W. Dong, X. Sun, X. Yan, and Q. Zhao, "Fault diagnosis of high-speed railway turnout based on support vector machine," in *2016 IEEE International Conference on Industrial Technology (ICIT)*, 2016, pp. 1539–1544.

[177] L. Liu, F. Zhou, and Y. He, "Automated status inspection of fastening bolts on freight trains using a machine vision approach," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, 12 2015.

[178] W. Sammouri, E. Côme, L. Oukhellou, P. Aknin, and C. Fonlladosa, "Pattern recognition approach for the prediction of infrequent target events in floating train data sequences within a predictive maintenance framework," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, pp. 918–923.

[179] A. Ferrari, G. Gori, B. Rosadini, I. Trotta, S. Bacherini, A. Fantechi, and S. Gnesi, "Detecting requirements defects with nlp patterns: an industrial experience in the railway domain," *Empirical Software Engineering*, vol. 23, 02 2018.

[180] J. Lee, H. Choi, D. Park, Y. Chung, H.-Y. Kim, and S. Yoon, "Fault detection and diagnosis of railway point machines by sound analysis," *Sensors*, vol. 16, p. 549, 04 2016.

[181] I. Aydin, E. Karaköse, M. Karaköse, M. T. Gençoğlu, and E. Akın, "A new computer vision approach for active pantograph control," in *2013 IEEE INISTA*. IEEE, 2013, pp. 1–5.

[182] X. Gibert, V. M. Patel, and R. Chellappa, "Robust fastener detection for autonomous

visual railway track inspection," in *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 694–701.

[183] A. F. Posada Moreno, C. Klein, M. Haßler, D. Pehar, A. F. Solvay, and C. P. Kohlschein, "Cargo Wagon Structural Health Estimation Using Computer Vision," in *[8th Transport Research Arena, TRA2020, 2020-04-27 - 2020-04-30, Helsinki, Finland]*. 8th Transport Research Arena, Helsinki (Finland), 27 Apr 2020 - 30 Apr 2020, Apr 2020, veröffentlicht auf dem Publikationsserver der RWTH Aachen University.

[184] Y. Santur, M. Karaköse, and E. Akin, "Random forest based diagnosis approach for rail fault inspection in railways," 12 2016.

[185] H. Trinh, N. Haas, Y. Li, C. Otto, and S. Pankanti, "Enhanced rail component detection and consolidation for rail track inspection," *Proceedings of IEEE Workshop on Applications of Computer Vision*, pp. 289–295, 01 2012.

[186] J. Moura and M. S. Erden, "Formulation of a control and path planning approach for a cab front cleaning robot," *Procedia CIRP*, vol. 59, pp. 67 – 71, 2017, proceedings of the 5th International Conference in Through-life Engineering Services Cranfield University, 1st and 2nd November 2016.

[187] B. Schlake, J. Edwards, J. Hart, C. Barkan, S. Todorovic, and N. Ahuja, "Machine vision condition monitoring of heavy-axle load railcar structural underframe components," *Proceedings of The Institution of Mechanical Engineers Part F-journal of Rail and Rapid Transit - PROC INST MECH ENG F-J RAIL R*, vol. 1, pp. 1–13, 09 2010.

[188] R. Vithanage, C. Harrison, and A. DeSilva, "A study on automating rolling-stock maintenance in the rail industry using robotics," 01 2017, pp. 278–283.

[189] ——, "Enhance 3d point cloud accuracy through supervised machine learning for automated rolling stock maintenance: A railway sector case study," 08 2018, pp. 242–246.

[190] L. Liu, F. Zhou, and Y. He, "Automated visual inspection system for bogie block key under complex freight train environment," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 1, pp. 2–14, 2016.

[191] F. Zhou, R. Zou, Y. Qiu, and H. Gao, "Automated visual inspection of angle cocks during train operation," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 228, pp. 794–806, 09 2013.

[192] A. Jamshidi, S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, R. Dollevoet, Z. Li, and B. De Schutter, "A big data analysis approach for rail failure risk assessment," *Risk analysis : an official publication of the Society for Risk Analysis*, vol. 37, 05 2017.

[193] M. An, Y. Chen, and C. J. Baker, "A fuzzy reasoning and fuzzy-analytical hierarchy process based approach to the process of railway risk information: A railway risk management system," *Information Sciences*, vol. 181, no. 18, pp. 3946–3966, 2011.

[194] G. Ćirović and D. Pamučar, "Decision support model for prioritizing railway level crossings for safety improvements: Application of the adaptive neuro-fuzzy system," *Expert Systems with Applications*, vol. 40, no. 6, pp. 2208–2223, 2013.

[195] M. Gul and E. Celik, "Fuzzy rule-based fine–kinney risk assessment approach for rail transportation systems," *Human and Ecological Risk Assessment: An International Journal*, vol. 24, no. 7, pp. 1786–1812, 2018.

[196] Y. Li, D. Chen, and X. Yang, "Expert system applied to high-speed railway track circuit

coding and its simulation," in *International Conference on Artificial Intelligence and Computational Intelligence*. Springer, 2011, pp. 517–524.

[197] H. Alawad, S. Kaewunruen, and M. An, "Learning from accidents: Machine learning for safety at railway stations," *IEEE Access*, vol. 8, pp. 633–648, 2020.

[198] H. Zhang, M. Yuan, Y. Liang, B. Wang, W. Zhang, and J. Zheng, "A risk assessment based optimization method for route selection of hazardous liquid railway network," *Safety science*, vol. 110, pp. 217–229, 2018.

[199] H. Hadj-Mabrouk, "Contribution of artificial intelligence and machine learning to the assessment of the safety of critical software used in railway transport," vol. 3, pp. 33–70, 02 2019.

[200] A. A. Zilko, D. Kurowicka, and R. M. Goverde, "Modeling railway disruption lengths with copula bayesian networks," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 350 – 368, 2016.

[201] M. Sturari, M. Paolanti, E. Frontoni, A. Mancini, and P. Zingaretti, "Robotic platform for deep change detection for rail safety and security," in *2017 European Conference on Mobile Robots (ECMR)*, 2017, pp. 1–6.

[202] F. Maire and A. Bigdeli, "Obstacle-free range determination for rail track maintenance vehicles," in *2010 11th International Conference on Control Automation Robotics Vision*, 2010, pp. 2172–2178.

[203] J. Wohlfeil, "Vision based rail track and switch recognition for self-localization of trains in a rail network," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, 2011, pp. 1025–1030.

[204] W. Nowakowski, P. Bojarczak, and Z. Łukasik, "Verification and validation of railway control systems using an expert system," in *Intelligent Transport Systems – From Research and Development to the Market Uptake*, T. Kováčiková, Ľ. Buzna, G. Pourhashem, G. Lugano, Y. Cornet, and N. Lugano, Eds. Cham: Springer International Publishing, 2018, pp. 43–50.

[205] D. Zhang, "High-speed train control system big data analysis based on the fuzzy rdf model and uncertain reasoning," *International Journal of Computers, Communications and Control*, vol. 12, pp. 577–591, 06 2017.

[206] J. Yin, D. Chen, and L. Li, "Intelligent train operation algorithms for subway by expert system and reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2561–2571, 2014.

[207] M. Brenna, F. Foiadelli, and M. Longo, "Application of genetic algorithms for driverless subway train energy optimization," *International Journal of Vehicular Technology*, vol. 2016, pp. 1–14, 02 2016.

[208] W. Carvajal Carreño, "Efficient driving of cbtc ato operated trains," Ph.D. dissertation, KTH Royal Institute of Technology, 2017.

[209] P. Wang, A. Trivella, R. M. Goverde, and F. Corman, "Train trajectory optimization for improved on-time arrival under parametric uncertainty," *Transportation Research Part C: Emerging Technologies*, vol. 119, p. 102680, 2020.

[210] W.-t. DENG, H.-b. ZHAO, and H. HE, "Research on heavy haul railway dispatching system based on fuzzy expert system," *DEStech Transactions on Computer Science and Engineering*, 08 2018.

[211] A. Fay, "A fuzzy knowledge-based system for railway traffic control," *Engineering Applications of Artificial Intelligence*, vol. 13, no. 6, pp. 719–729, 2000.

[212] H. Schaefer and S. Pferdmenges, "An expert system for real-time train dispatching," *WIT Transactions on The Built Environment*, vol. 7, 1970.

[213] F. Liu, R. Xu, W. Fan, and Z. Jiang, "A data analytics approach for train timetable performance measures using automatic train supervision data," *IET Intelligent Transport Systems*, vol. 12, 02 2018.

[214] F. Cerreto, B. F. Nielsen, O. A. Nielsen, and S. S. Harrod, "Application of data clustering to railway delay pattern recognition," *Journal of Advanced Transportation*, vol. 2018, 2018.

[215] P. Wang and Q.-p. Zhang, "Train delay analysis and prediction based on big data fusion," *Transportation Safety and Environment*, vol. 1, no. 1, pp. 79–88, 02 2019.

[216] P. Kecman and R. Goverde, "Predictive modelling of running and dwell times in railway traffic," *Public Transport*, vol. 7, 06 2015.

[217] P. Kecman and R. M. P. Goverde, "Online data-driven adaptive prediction of train event times," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 465–474, 2015.

[218] K. Schüpbach, G. Caimi, and J. Jordi, "Towards automated capacity planning in railways," 11 2017.

[219] P. Tormos, A. Lova, F. Barber, L. Ingolotti, M. Abril, and M. A. Salido, *A Genetic Algorithm for Railway Scheduling Problems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 255–276.

[220] R. Barman, C. Baishya, B. Kharmalki, A. Syiemlieh, K. Pegu, T. Das, and G. Saha, "Automated train scheduling system using genetic algorithm," 09 2015, pp. 28–33.

[221] T. Ho, C. W. Tsang, K. H. Ip, and K. Kwan, "Train service timetabling in railway open markets by particle swarm optimisation," *Expert Systems with Applications*, vol. 39, no. 1, pp. 861–868, 2012.

[222] H. Pu, T. Song, P. Schonfeld, W. Li, H. Zhang, J. Hu, X. Peng, and J. Wang, "Mountain railway alignment optimization using stepwise & hybrid particle swarm optimization incorporating genetic operators," *Applied Soft Computing*, vol. 78, pp. 41–57, 2019.

[223] M. Wang, L. Wang, X. Xu, Y. Qin, and L. Qin, "Genetic algorithm-based particle swarm optimization approach to reschedule high-speed railway timetables: a case study in china," *Journal of Advanced Transportation*, vol. 2019, 2019.

[224] R. Wang, Z. Zhang, L. Wang, and H. Zhang, "Resource-expandable railway freight transportation routing optimization," *IEEE Access*, vol. 7, pp. 161 607–161 620, 2019.

[225] Y.-J. Zheng, H.-F. Ling, H.-H. Shi, H.-S. Chen, and S.-Y. Chen, "Emergency railway wagon scheduling by hybrid biogeography-based optimization," *Computers & Operations Research*, vol. 43, pp. 1–8, 2014.

[226] H. Khadilkar, "A scalable reinforcement learning algorithm for scheduling railway lines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 727–736, 2019.

[227] M. Obara, T. Kashiyama, and Y. Sekimoto, "Deep reinforcement learning approach for train rescheduling utilizing graph theory," 12 2018, pp. 4525–4533.

[228] L. Oneto, E. Fumeo, G. Clerico, R. Canepa, F. Papa, C. Dambra, N. Mazzino, and D. Anguita, "Dynamic delay predictions for large-scale railway networks: Deep and shallow extreme learning machines tuned via thresholdout," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 10, pp. 2754–2767, 2017.

[229] J. A. Palombarini and E. C. Martínez, "Closed-loop rescheduling using deep reinforcement learning," *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 231 – 236, 2019, 12th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems DYCOPS 2019.

[230] E. Peer, V. Menkovski, Y. Zhang, and W. Lee, "Shunting trains with deep reinforcement learning," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 3063–3068.

[231] D. Roost, R. Meier, S. Huschauer, E. Nygren, A. Egli, A. Weiler, and T. Stadelmann, "Improving sample efficiency and multi-agent communication in rl-based train rescheduling," 2020.

[232] S. Salsingikar and N. Rangaraj, "Reinforcement learning for train movement planning at railway stations," in *Adaptive and Learning Agents Workshop 2020*, Auckland, New Zealand.

[233] T. Tang, R. Liu, and C. Choudhury, "Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data," *Sustainable Cities and Society*, vol. 53, p. 101927, 2020.

[234] R. Goverde, N. Bešinović, A. Binder, V. Cacchiani, E. Quaglietta, R. Roberti, and P. Toth, "A three-level framework for performance-based railway timetabling," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 62–83, 06 2016.

[235] Y. Yin, D. li, N. Bešinović, and Z. Cao, "Hybrid demand-driven and cyclic timetabling considering rolling stock circulation for a bidirectional railway line," *Computer-Aided Civil and Infrastructure Engineering*, 09 2018.

[236] Q. Xue, X. Yang, J. Wu, H. Sun, H. Yin, and Y. Qu, "Urban rail timetable optimization to improve operational efficiency with flexible routing plans: A nonlinear integer programming model," *Sustainability*, vol. 11, no. 13, p. 3701, 2019.

[237] N. Bešinović, E. Quaglietta, and R. M. Goverde, "A simulation-based optimization approach for the calibration of dynamic train speed profiles," *Journal of Rail Transport Planning Management*, vol. 3, no. 4, pp. 126 – 136, 2013.

[238] A. Prokhorchenko, A. Panchenko, L. Parkhomenko, G. Nesterenko, M. Muzykin, G. Prokhorchenko, and A. Kolisnyk, "Forecasting the estimated time of arrival for a cargo dispatch delivered by a freight train along a railway section," *Eastern-European Journal of Enterprise Technologies*, vol. 3, pp. 30–38, 06 2019.

[239] W. Barbour, J. C. Martinez Mori, S. Kuppa, and D. B. Work, "Prediction of arrival times of freight traffic on us railroads using support vector regression," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 211–227, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X18307101

[240] Wei Xu, Yong Qin, and Houkuan Huang, "A new method of railway passenger flow forecasting based on spatio-temporal data mining," in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)*, 2004, pp. 402–405.

[241] M. Gallo, G. De Luca, L. D'Acierno, and M. Botte, "Artificial neural networks for forecasting passenger flows on metro lines," *Sensors*, vol. 19, pp. 1–14, 08 2019.

[242] L. Heydenrijk-Ottens, V. Degeler, D. Luo, N. V. Oort, and J. V. Lint, "Supervised learning : Predicting passenger load in public transport," 2018.

[243] Y. Liu, Z. Liu, and R. Jia, "DeepPF: A deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 18 – 34, 2019.

[244] H. Zhu, X. Yang, and Y. Wang, "Prediction of daily entrance and exit passenger flow of rail transit stations by deep learning method," *Journal of Advanced Transportation*, vol. 2018, pp. 1–11, 04 2018.

[245] R. Kour, P. Tretten, and R. Karim, "Emaintenance solution through online data analysis for railway maintenance decision-making," *Journal of Quality in Maintenance Engineering*, vol. 20, 08 2014.

[246] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[247] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[248] V. Fragnelli and S. Sanguineti, "A game theoretic model for re-optimizing a railway timetable," *European Transport Research Review*, vol. 6, pp. 113–125, 06 2013.

[249] K. Bablinski, "A game-based analysis of freight paths allocation with a case study on great britain brighton main line," *Transportation Research Procedia*, vol. 13, pp. 196 – 208, 2016, towards future innovative transport: visions, trends and methods 43rd European Transport Conference Selected Proceedings.

[250] H. Zhuang, L. Feng, C. Wen, Q. Peng, and Q. Tang, "High-speed railway train timetable conflict prediction based on fuzzy temporal knowledge reasoning," *Engineering*, vol. 2, no. 3, pp. 366 – 373, 2016.

[251] M. Babaeizadeh, I. Frosio, S. Tyree, J. Clemons, and J. Kautz, "Reinforcement learning through asynchronous advantage actor-critic on a gpu," 2017.

[252] J. Huo, J. Wu, L. Kang, and B. Wang, "Railway timetable rescheduling based on priority and train order entropy," *Journal of Computing in Civil Engineering*, vol. 30, p. 04016006, 01 2016.

[253] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[254] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[255] AECOM, "Thameslink programme evaluation: Baseline report," https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/689701/thameslink-programme-baseline-evaluation-report.pdf, accessed: 4.2.2021.

[256] ProRail, "ATO verkenning Betuweroute Eindrapportage," https://www.prorail.nl/conte

ntassets/39331d97c0a24281a342bc408e7b05a9/eindrapport_ato_verkenning_betuweroute.pdf, accessed: 4.2.2021.
[257] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.
[258] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
[259] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," 2018.
[260] T. Denœux, D. Dubois, and H. Prade, "Representations of uncertainty in ai: Beyond probability and possibility," in *A Guided Tour of Artificial Intelligence Research*. Springer, 2020, pp. 119–150.
[261] AI HLEG, "Ethics guidelines for trustworthy AI," *B-1049 Brussels*, 2019.
[262] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," 2019.
[263] D. Roeßler, J. Reisch, F. Hauck, and N. Kliewer, "Discerning primary and secondary delays in railway networks using explainable ai," *Transportation Research Procedia*, vol. 52, pp. 171–178, 2021.
[264] K. Gade, S. C. Geyik, K. Kenthapadi, V. Mithal, and A. Taly, "Explainable ai in industry," ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 3203–3204.
[265] N. Duckworth, "AI in supply chain: Six barriers to seeing results," Supply Chain Brain, Tech. Rep., 2019. [Online]. Available: https://www.supplychainbrain.com/blogs/1-think-tank/post/30051-six-barriers-to-getting-results-with-ai-in-supply-chain-management
[266] AI HLEG, "The assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment," *B-1049 Brussels*, 2020.
[267] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.
[268] K. Aslansefat, I. Sorokos, D. Whiting, R. Tavakoli Kolagari, and Y. Papadopoulos, "SafeML: Safety Monitoring of Machine Learning Classifiers through Statistical Difference Measure," *arXiv e-prints*, 2020.
[269] E. Commission. (2016) Data protection: Rules for the protection of personal data inside and outside the eu. [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection_en
[270] M. Niestadt, A. Debyser, D. Scordamaglia, and M. Pape, "Artificial intelligence in transport current and future developments, opportunities and challenges," 2019.
[271] J. Morley, C. C. Machado, C. Burr, J. Cowls, I. Joshi, M. Taddeo, and L. Floridi, "The ethics of AI in health care: A mapping review," *Social Science Medicine*, vol. 260, p. 113172, 2020.
[272] A. Winfield, "Ethical standards in robotics and AI," *Nature Electronics*, 2020.

[273] B. W. Smith, *The Oxford Handbook of Ethics of AI*. Oxford Press, 2020, ch. Ethics of Artificial Intelligence in Transport.

[274] H. Etienne, “When ai ethics goes astray: A case study of autonomous vehicles,” *Social Science Computer Review*, 2020.

[275] R. M. Monarch, *Human-in-the-Loop Machine Learning – Active learning and annotation for human-centered AI*. Manning publications, 2021.

[276] R. Pandey, H. Purohit, C. Castillo, and V. L. Shalin, “Modeling and mitigating human annotation errors to design efficient stream processing systems with human-in-the-loop machine learning,” 2020.

[277] S. Budd, E. C. Robinson, and B. Kainz, “A survey on active learning and human-in-the-loop deep learning for medical image analysis,” 2019.

[278] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim, “What you see is what you can change: Human-centered machine learning by interactive visualization,” *Neurocomputing*, vol. 268, pp. 164–175, 2017, advances in artificial neural networks, machine learning and computational intelligence.

[279] Y. Zhang, H. L. Yang, S. Prasad, E. Pasolli, J. Jung, and M. Crawford, “Ensemble multiple kernel active learning for classification of multisource remote sensing data,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 2, pp. 845–858, 2015.

[280] T. Edwards, A. Bayoumi, and M. L. Eisner, “Internet of things–a complete solution for aviation’s predictive maintenance,” in *Advanced Technologies for Sustainable Systems*. Springer, 2017, pp. 167–177.

[281] P. Runeson, M. Alexandersson, and O. Nyholm, “Detection of duplicate defect reports using natural language processing,” in *29th International Conference on Software Engineering (ICSE’07)*, 2007, pp. 499–510.

[282] A. Serna and S. Gasparovic, “Transport analysis approach based on big data and text mining analysis from social media,” *Transportation research procedia*, vol. 33, pp. 291–298, 2018.

[283] W. B. Powell, B. Bouzaiene-Ayari, C. Lawrence, C. Cheng, S. Das, and R. Fiorillo, “Locomotive planning at norfolk southern: an optimizing simulator using approximate dynamic programming,” *Interfaces*, vol. 44, no. 6, pp. 567–578, 2014.

[284] M. Morabit, G. Desaulniers, and A. Lodi, “Automated track inspection vehicle and method,” May 2020, gERAD, HEC Montreal, Canada.

[285] Y. Bengio, A. Lodi, and A. Prouvost, “Machine learning for combinatorial optimization: a methodological tour d’horizon,” *European Journal of Operational Research*, 2020.

[286] B. Cesme and P. G. Furth, “Self-organizing traffic signals using secondary extension and dynamic coordination,” *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 1–15, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X14002228

[287] A. Whitbrook, Q. Meng, and P. W. H. Chung, “Reliable, distributed scheduling and rescheduling for time-critical, multiagent systems,” *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 732–747, 2018.

[288] N. Choudhary, “The role of safety risk management in the uk rail industry when dealing

with cyber threats the role of safety risk management in the uk rail industry when dealing with cyber threats."

[289] R. Kour, M. Aljumaili, R. Karim, and P. Tretten, "eMaintenance in railways: Issues and challenges in cybersecurity," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 233, no. 10, pp. 1012–1022, 2019.

[290] M. J. Pappaterra and F. Flammini, "A review of intelligent cybersecurity with bayesian networks," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 445–452.

[291] H. Mokalled, C. Pragliola, D. Debertol, E. Meda, and R. Zunino, *A Comprehensive Framework for the Security Risk Management of Cyber-Physical Systems*. Cham: Springer International Publishing, 2019, pp. 49–68.

[292] W. A. Omta, R. G. van Heesbeen, I. Shen, J. de Nobel, D. Robers, L. M. van Der Velden, R. H. Medema, A. P. Siebes, A. J. Feelders, S. Brinkkemper *et al.*, "Combining supervised and unsupervised machine learning methods for phenotypic functional genomics screening," *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, p. 2472555220919345, 2020.

[293] T. Wang, S. Pouyanfar, H. Tian, Y. Tao, M. Alonso, S. Luis, and S.-C. Chen, "A framework for airfare price prediction: A machine learning approach," in *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE, 2019, pp. 200–207.

[294] S. A. M. Shihab, C. Logemann, D.-G. Thomas, and P. Wei, "Towards the next generation airline revenue management: A deep reinforcement learning approach to seat inventory control and overbooking," *arXiv preprint arXiv:1902.06824*, 2019.