Deliverable D 3.4

WP3 Report on identification of future innovation needs and recommendations for improvements

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 40
Actual submission date:	June 30th 2023
Responsible/Author:	Francesco Flammini (LNU)
Dissemination level:	Public
Status:	Issued

Reviewed: yes

Document history		
Revision	*Date*	*Description*
1	June 5^{th} 2023	First issue for internal review
2	June 30^{th} 2023	Second issue after internal review
3	July 29^{th} 2023	Third issue after JU review

Report contributors		
Name	Beneficiary Short Name	Details of contribution
Francesco Flammini	LNU	WP Leader
Lorenzo De Donato	CINI	Contributor
Ruifan Tang	UNIVLEEDS	Contributor
Valeria Vittorini	CINI	Internal Reviewer

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 n. 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

This deliverable contains a critical examination of the work and the results obtained in WP3, also against the current state-of-the-art in railways. From the Proofs-of-Concept (PoCs) and the experience gained by developing the project's tasks, this document reports lessons learned, weaknesses and strengths shown by each exploited technology, technical and implementation recommendations, unaddressed issues, and innovation needs. Specifically, an in-depth analysis is provided of the methodological and experimental PoCs carried out for the two selected case studies proposed in the previous WP3 deliverables, namely, "Smart Maintenance at Level Crossings" and "AI-assisted Rolling Stock Rostering". This analysis, arranged according to the SWOT structure, is going to highlight the main strengths (S) and weaknesses (W) shown by the proposed AI approaches, and identify external opportunities (O) that could support its technical feasibility, as well as some open challenges related to the main threats (T) emerged from each PoC. Starting from this evidence, some specific recommendations are addressed regarding the effectiveness of the proposed AI approaches. Furthermore, general recommendations are drawn, including suggestions for future development, experimentation, and applications for the integration of AI in the rail sector.

Abbreviations and acronyms

Abbreviations / Acronyms	Description
AI	Artificial Intelligence
BAM	Barrier Analysis Module
DL	Deep Learning
DT	Digital Twin
FN	False Negative
FoV	Field of View
FP	False Positive
IoT	Internet of Things
LC	Level Crossing
M&I	Maintenance and Inspection
PoC	Proof-of-Concept
RL	Reinforcement Learning
SOTA	State-of-the-Art
SWOT	Strengths, Weaknesses, Opportunities, and Threats
WB	Warning Bell
WBDM	Warning Bell Detection Module
WL	Warning Light
WP	Work Package

1. Objective

This document aims to draw some conclusions from the work carried out in WP1 and WP3, helping to clearly identify possible future innovations, research directions, and impacts for the European railway sector. These objectives are strictly related to the work addressed in the previous WP3 deliverables, in which methodological and experimental Proofs-of-Concept (PoCs) have been carried out for two selected case studies, namely, “Smart Maintenance at Level Crossings” and “AI-assisted Rolling Stock Rostering”.

On the basis of the results that emerged from the PoCs, the main goal of this deliverable is to provide some recommendations and innovation needs that could support the effective adoption of Artificial Intelligence (AI) in the rail sector. To this aim, the present document focuses on the following objectives:

- A brief overview of the recent advancements in the field of the related PoCs.
- An in-depth analysis of the proposed PoCs, arranged according to the SWOT structure, to highlight the main strengths (S) and weaknesses (W) shown by the proposed AI approaches, and to identify external opportunities (O) and threats (T) which could affect their technical feasibility.
- The identification of some specific recommendations which could support the effectiveness of the proposed AI approaches;
- The drawing of general recommendations regarding the integration of AI for railway maintenance and inspection, mainly encompassing indications for further investigation, including:
  - Further development of approaches, methods, models, technologies, and tools.
  - Further experimentation with additional data, case studies, pilot studies, and scenarios.
  - Applications to other areas and subsystems within the railway transport sector.

The set of recommendations will be used in WP5 to identify migration strategies and roadmaps for AI integration in the rail sector.

2. Introduction

This deliverable reports technical/implementation recommendations and innovation needs that would support future investigations in the context of AI for railway maintenance and inspection.

The recommendations provided by this document can be subdivided into two main macro-categories:

1. Recommendations coming from the critical examination of the Proofs-of-Concept (PoCs) developed within the RAILS WP3, which would be potentially useful to support future development of approaches, methods, models, technologies, and tools in the specific contexts of the PoCs and related areas.
2. General Recommendations, coming from lessons learned while both working at the PoCs and investigating the state-of-the-art of AI in railways, which aim at providing hints about practices and activities that would support the integration of AI across various railway applications.

The remainder of this deliverable is organised as follows. Chapter 3 summarises the findings of WP1 about the state-of-the-art and promising research directions in railway maintenance and inspection, as well as the documents and results produced during the project and addressing the topics investigated in the context of WP3, including the related scientific publications stemming from the project activities. Then, Chapter 3 provides the context and the background of the discussion reported in the present deliverable. Chapter 4 and Chapter 5 respectively address critical examinations of the PoCs on “Smart Maintenance at Level Crossings” and on “AI-assisted Rolling Stock Rostering”. Chapter 4 and Chapter 5 share the same structure: Sections 4.1 and 5.1 discuss a high-level overview of the recent advancements in the context of the corresponding PoCs; Sections 4.2 and 5.2 present a bird-eye view of the investigative approaches; Sections 4.3 and 5.3 propose structured analyses of the implemented approaches in the form of SWOT (strengths, weaknesses, opportunities, and threats) analyses; lastly, Section 4.4 and 5.4 highlight the main recommendations resulted from the lessons learned while working at the PoCs. Then, Chapter 6 discusses the general recommendations and some innovation needs that would be required for the fast take-up of AI in railways. Lastly, Chapter 7 provides some concluding remarks.

3. Background

This section recalls most of the findings from the analyses carried out in the previous phases of the RAILS project with specific emphasis on AI applications for "Maintenance and Inspection" (M&I) activities. Table 3.1 reports all the documents (deliverables and papers) resulting from the aforementioned research activities and specifies their main contributions/results.

Table 3.1: Published Documents discussing AI for Railway M&I Applications.

Focus	Document	Type	Main contribution(s)
Taxonomy	Deliverable D1.1: Definition of a Reference Taxonomy of AI in Railways [1]	PD	1. Delineation of a definition for AI in railway 2. Establishment of a taxonomy of AI in railway 3. Preliminary overview of regulations for AI 4. Identification of Railway Subdomains 5. Preliminary mapping of existing AI applications on Railway Subdomains
	Artificial Intelligence in Railway Transport: Taxonomy, Regulations, and Applications [2]	SP	
State of the Art	Deliverable D1.2: Summary of Existing Relevant Projects and State-of-the-Art of AI Application in Railways [3]	PD	1. Review of projects conducted worldwide (with emphasis on S2R projects) dealing with AI in Railway Subdomains 2. Review of scientific papers dealing with AI in Railway Subdomains 3. Preliminary definition of future direction towards the integration of AI
	A Literature Review of Artificial Intelligence Applications in Railway Systems [4]	SP	Extended review of scientific papers dealing with AI in Railway Subdomains
	A Survey on Audio-Video Based Defect Detection Through Deep Learning in Railway Maintenance [5]	SP	In-depth review of scientific papers dealing with AI for M&I applications exploiting audio-video data
	A Systematic Review of Artificial Intelligence Public Datasets for Railway Applications [6]	SP	In-depth review of publicly available datasets for each Railway Subdomain
	Railway Digital Twins and Artificial Intelligence: Challenges and Design Guidelines [7]	SP	1. Review of scientific papers discussing the combination of DTs and AI for railway M&I applications 2. Preliminary guidelines for the generation of AI-aided DTs for railway maintenance applications
Application Areas	Deliverable 1.3: Application Areas [8]	PD	1. Identification of relevant railway Application Areas for AI together with the main challenges to be tackled for its effective integration basing on: i) The review of projects conducted worldwide and the scientific literature dealing with AI in railways; ii) Suggestions from the Advisory Board; and iii) the results from a comprehensive survey submitted to researchers and practitioners from different organisations operating worldwide 2. Delineation of basic AI usage guidelines to select the most appropriate AI approach by taking into account: i) the goal; ii) the type of available data; and iii) the required responsiveness of the AI system
Transferability	Deliverable 3.1: WP3 Report on Case Studies and Analysis of Transferability from Other Sectors [9]	PD	1. Review of AI-based emerging technologies developed in (transport and non-transport) sectors other than railways 2. Identification of AI approaches that can be transferred to or adapted for railway applications
AI-aided DTs for M&I	Towards AI-Assisted Digital Twins for Smart Railways: Preliminary Guideline and Reference Architecture [10]	SP	1. Extended guidelines for the generation of AI-aided DTs for railway maintenance applications 2. Delineation of a preliminary reference architecture of AI-aided DTs for railway M&I applications
PoCs Development	Deliverable 3.2: WP3 Report on AI approaches and models [11]	PD	Identification of PoCs to be developed together with Research Questions, Methodology, Reference Datasets, AI and ML Models, and Expected Results
	Deliverable 3.3: WP3 Report on experimentation, analysis, and discussion of results [12]	PD	Development of identified PoCs including model/architecture description, data generation, training and validation, and evaluation and discussion of results
	Intelligent detection of warning bells at level crossings through deep transfer learning for smarter railway maintenance [13]	SP	Insights on the development of a DL model for the detection of level crossings' warning bells.
Recommendations	Deliverable 3.4: WP3 Report on identification of future innovation needs and recommendations for improvements [this document]	PD	1. Identification of sectorialized recommendations oriented at supporting AI integration in WP3 PoCs' topic and related areas 2. Identification of general recommendations aiming at supporting AI integration across different railway applications

PD: Project Deliverable; SP: Scientific Paper

Herein, we recall some of the results obtained within the first phase of the project as they impacted the subsequent investigation carried out in WP3. First, a *Taxonomy of AI in Railway* has been introduced and seven *Railway Subdomains* have been identified for the RAILS investigation [1, 2]. Then, in Deliverable D1.2, for each of the subdomains, a review of the State-of-the-Art (SOTA) of AI in railways has been developed by analysing i) research projects conducted worldwide (with a particular focus on S2R projects) and ii) the scientific literature. The latter analysis has been then extended in [4]. The main statistical findings we deducted from the SOTA are reported in Fig. 3.1.

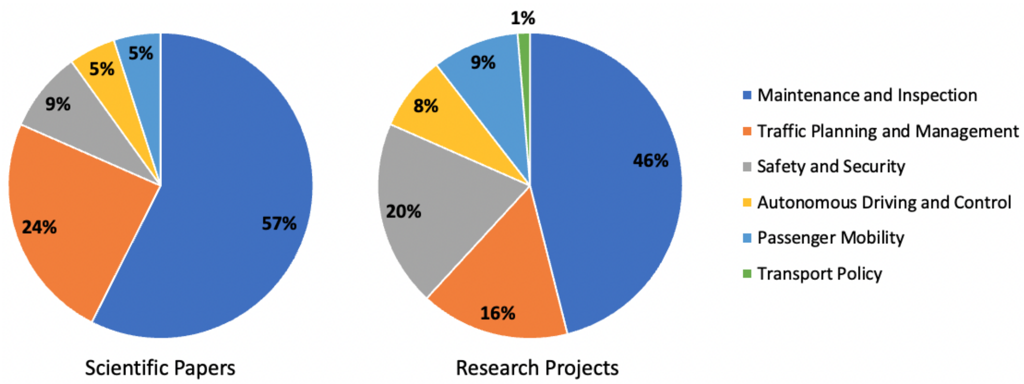


Fig. 3.1. Distribution of Scientific Papers and Research Projects per Railway Subdomain.

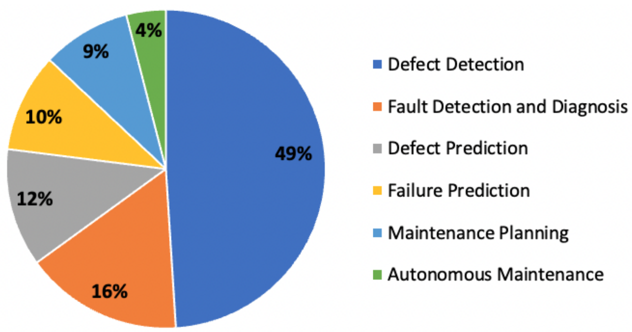


Fig. 3.2. Distribution of Scientific Papers per M&I tasks.

As expected, M&I resulted to be the most investigated area of the railway sector with most of the AI approaches developed/tested to deal with problems related to the following M&I tasks (also reported in Fig. 3.2):

- **Defect Detection**, which deals with the identification of physical defects (e.g. cracks, missing items, scratches) in railway assets.
- **Defect Prediction**, which involves the identification of the status of deterioration of physical components.
- **Fault Detection and Diagnosis**. Detection is related to the identification of faults or anomalies in components, while Diagnosis aims at identifying their causes.
- **Failure Prediction**, which is concerned with forecasting the state of a component that deviates from its nominal behaviour leading to harmful consequences.

- **Maintenance Planning**, which deals with approaches for decision support systems allowing dynamic scheduling of M&I activities through, for example, continuous remote monitoring.
- **Autonomous Maintenance**, which focuses on the design of intelligent systems aiming at autonomously inspecting and maintaining railway assets.

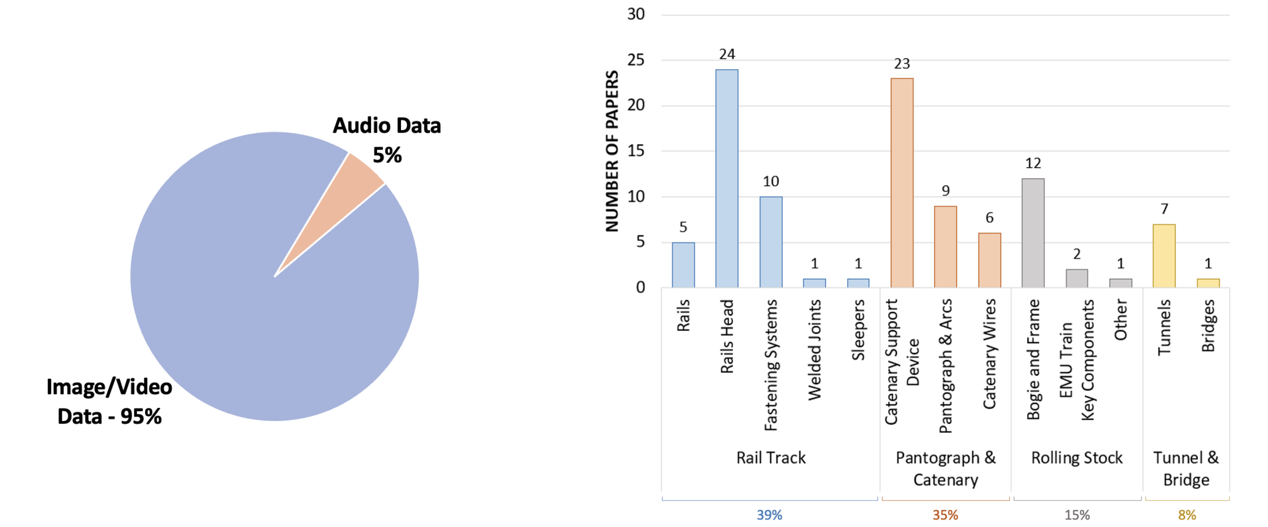
Table 3.2 summarises the main AI approaches that researchers and practitioners have leveraged, subdivided for M&I tasks.

Table 3.2: Mapping AI Algorithms/Models with M&I Tasks.

M&I Task	AI Algorithm / Model																												
	Tree-based				*Inst.-b.*			*Batesian*			*Regression*		*Neural Networks*								*Clus.*		*Others*						
	DT	RF	GBDT	EGBT	SVM	SVR	KNN	NB	BN	LaDA	SWLR	GPRRQ	ANN	FNN	NARXNN	CNN	RNN	DBN	RBM	ODs	K-m	EM	AdaB	XGB	LDA	QDA	PCA	HHMM	GA
Defect Detection		x			x					x			x			x	x			x			x		x		x		
Fault Detection and Diagnosis					x								x				x	x				x			x	x	x		
Defect Prediction	x				x								x		x									x			x		
Failure Prediction					x	x	x	x			x		x	x			x		x									x	
Maintenance Planning	x	x	x						x							x					x								x
Autonomous Maintenance				x																									

Primary Header Acronyms: Instance-based (Inst.-b), Clustering (Clus.).
Secondary Header Acronyms: 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), Stepwise Linear Regression (SWLR), Rational Quadratic Gaussian Process Regression (GPRRQ), 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), Deep Belief Networks (DBN), Restricted Boltzmann Machine (RBM), Object Detectors (ODs), K-means Algorithm (K-m), Expectation-Maximization Algorithm (EM), Adaptive Boosting (AdaB), eXtreme Gradient Boosting (XGB), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Principal Component Analysis (PCA), Hierarchical Hidden Markov Models (HHMM), Genetic Algorithms (GA).

The majority of the studies we analysed within the aforementioned documents focused on the analysis of rail tracks, rolling stock, concrete infrastructures (e.g., tunnels), and catenary systems components. The same findings were also obtained by specifically analysing the SOTA of Audio-Video based Deep Learning (DL) approaches for M&I applications in railways [5]. This work was oriented at defining a picture of the contributions provided by researchers for Surface Defect Detection (SDD), Defect Inspection (DI), and Object Identification (OI) in the context of railway assets M&I, as well as identifying publicly available datasets that could be used as benchmarks for the development of future applications. For the sake of completeness, SDD includes approaches oriented at detecting surface defects (e.g., scratches); DI encompasses solutions aiming at detecting substantial defects like broken/missing objects; lastly, OI refers to applications which are focused on the detection of particular components/sub-components of railway assets. Figures 3.3 and 3.4 report the main statistical findings obtained from this study. Fig. 3.3a reports the distribution of reviewed works basing on the kind of data they leveraged, i.e., video or audio data; Fig. 3.3b shows the distribution of studies in terms of railway assets (and specific components) they focused on; lastly, Fig. 3.4 evidences the most used Deep Neural Networks (DNNs) that researchers exploited catalogued per defect detection task (i.e., SDD, DI, and OI) and typology of the DL approach they implemented (i.e., Classification, Semantic Segmentation, and Object Detection).



(a) Studies Distribution per Data Type (b) Studies Distribution per Railway Asset Component

Fig. 3.3. Statistical Results from [5]

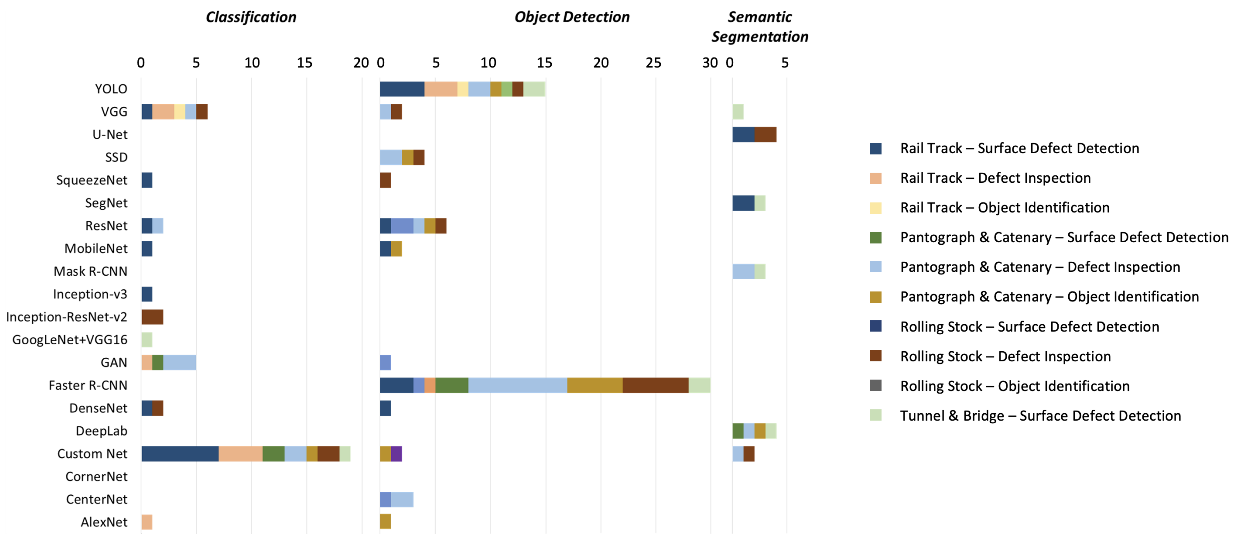


Fig. 3.4. Distribution of used DNNs per DL Approach and Defect Detection Task (from [5]).

The SOTA analysis above served as a basis for further investigations oriented at identifying, in Deliverable D1.3, the main railway Application Areas, i.e., sets of railway applications that could benefit from AI. Then, the findings we obtained in these contexts, together with the analysis of AI approaches proposed within other sectors (Deliverable D3.1), contributed to the identification and development of WP3 PoCs which are detailed in Deliverables 3.2 and 3.3. To conclude, the investigations conducted within the aforementioned deliverables and all the other documents reported in Table 3.1 converged into the definition of the practical recommendations that are discussed in the following chapters of this deliverable.

4. A Critical Examination of the Level Crossing Proof-of-Concept

Level Crossings (LCs) are among the most sensitive railway assets since they represent the junction point between roads and railways. Hence, they rise several concerns in terms of safety (about one-third of railway accidents happen at LCs [14]) and maintainability: guaranteeing their correct functioning should be the first step to ensure safe operability and transit availability. As better detailed in the following Section 4.1, LCs are typically inspected manually, however, new solutions oriented at continuously monitoring LCs have been proposed in recent years. In this context, with this PoC, we aimed at understating to what extent it would have been possible to leverage AI in combination with non-intrusive sensors (cameras and microphones) to evaluate the health status of LCs. Important to highlight, the focus on non-intrusive sensors has been posed given the safety-critical nature of LCs; physically applying intrusive sensors on certified assets could probably lead to re-approval processes that would be expensive and time-consuming.

To conclude, *the aim of this PoC was not to propose the most suitable and effective solution for this specific task, instead, we focused on investigating the opportunities that AI and non-intrusive sensors could introduce in LCs real-time monitoring and predictive maintenance with the final objective of highlighting recommendations for future research in this direction*.

4.1. Recent Advancements on Level Crossings Monitoring

In this section, a brief overview of the evolution of M&I tasks is given, including a few of the latest developments for *Smart Maintenance at Level Crossings*.

Over the past years, M&I activities of railway assets have mainly been performed on the field by following fixed scheduling, and LCs were not an exception [15]. However, this strategy does not allow timely detection of failures that may occur between two consecutive inspections. This is one of the primary motivations that has been leading railway stakeholders towards a paradigm shift from scheduled-based inspections and corrective actions to online monitoring and predictive maintenance. In this context, a few solutions have already been proposed[1,2] to collect and analyse LC data in real-time to check their health status and promptly detect malfunctions.

The main shortcoming of some of these systems regards the usage and installation of intrusive (IoT) sensors, i.e., applied directly on the asset to be monitored, which may lead to two main complications:

- The newest LCs may already be equipped with the latest and adequate sensors, however, there are thousands of LCs deployed decades ago which may be not; therefore, this would lead to a massive, expensive, and time-consuming sensorisation.
- Being safety-critical assets, level crossings are subject to strict standards and regulations including, for example, CEN-CENELEC standards for the assessment of the Safety Integrity Level [16]. Hence, the introduction of intrusive sensors (if not contemplated by design) could lead to expensive and time-consuming re-approval processes.

[1] https://www.smartmotors.org/level-crossing-monitoring

[2] https://www.voestalpine.com/railway-systems/en/products/rxm-rail-crossing-monitoring/

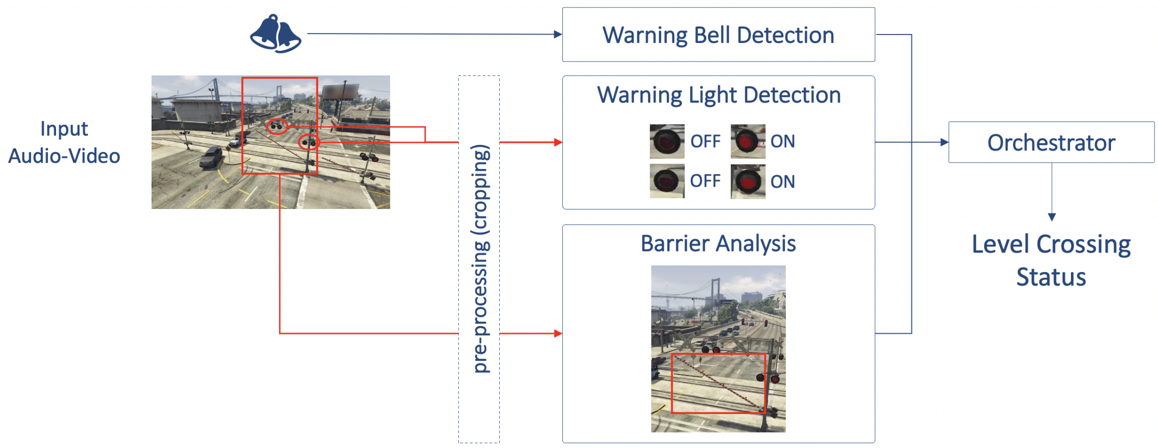


Fig. 4.1. High-Level Architecture for Level Crossing Intelligent Monitoring.
Image extracted from Grand Theft Auto V (GTA V)[3].

4.1.1. Bird-eye View of the PoC Approach

To overcome this issue and propose a cost-effective solution, in Deliverable D3.3 [12], we introduced a multi-modular framework for the intelligent monitoring of LCs based on Deep Learning and non-intrusive sensors like cameras and microphones. Compared to sensors directly installed on LC devices, approaches based on audio-video monitoring are less intrusive since they do not interfere with the system and therefore they have no effect on railway regulations and certification processes. Furthermore, such an approach allows dual use of technology, e.g., by (re-)using surveillance cameras featuring onboard microphones and computing capabilities [17].

The multi-modular architecture encompasses three main modules, each of which monitors a specific LC macro-component – Warning Bells (WBs), Warning Lights (WLs), and Barrier –, and a fourth component that should implement the malfunction detection logic. Important to underline, in our analyses, we considered a fixed camera monitoring the LC, which means that the Field of View (FoV) of the camera does not change over time. This requirement is not so distant from what can be applied in reality and, in our opinion, facilitate the implementation of the modules monitoring the WLs and the barrier. For the sake of completeness, Fig. 4.1 graphically shows the proposed architecture, while a brief description of the aforementioned modules, together with the identification of the DL approaches we investigated or that could be analysed in the future, is given in the following:

- **Warning Bell Detection Module (WBDM).** This module is oriented at detecting whether the WB is ringing. Inspired by VGGish [18], we implemented a Convolutional Neural Network (CNN) which is capable of processing 1-second audio frames so that, for each second, we can understand whether the WB is ringing or not.
- **Warning Light Detection Module (WLDM).** The purpose of this module is to detect WLs. Assuming that the camera is fixed, it would be possible to crop out each WL from the original frame (see Fig. 4.1) and then apply a simple classifier to detect whether the light is turned on or off. Given the simplicity of such classification tasks and the large attention that these kinds of problems have received over the past years, we did not address the implementation of this module since we believe that potentially any

[3] https://www.rockstargames.com/it/gta-v

kind of CNN (or even less complex Machine Learning - ML - approaches) could be easily implemented to cope with WL detection.

- **Barrier Analysis Module (BAM).** This module extracts the motion of the barrier over time. We adopted YOLOv5s[4] to detect the barrier within the frame and then applied a post-processing algorithm to extract the height of the barrier. The motion of the barrier is obtained by plotting the heights over time. In this case, the fixed camera facilitates the detection process since the DNN has to look for the barrier within an image smaller than the original frame.
- **Orchestrator.** This module would gather the outputs from the aforementioned model and, possibly, some additional information (e.g., the trigger event that activates the LC given the incoming train) to implement the malfunction detection logic. For example, focusing on the barrier, the Orchestrator would identify how much the detected motion deviates from the nominal behaviour the barrier should have; also, delays in the activation of the barrier could be identified by counting the time that passes from the aforementioned trigger and the moment where the barrier starts closing.

4.2. A SWOT Analysis of the PoC

In this section, we highlight the inner Strengths (S) and the Weaknesses (W) of the approach we investigated and identify some external Opportunities (O) and Threats (T) that cloud respectively support or challenge its technical implementation. Important to mention, this SWOT analysis is not oriented at defining the market viability of the solution we investigated, instead, it has been used to evaluate the approach in a structured manner. Fig. 4.2 reports all the aspects that we found to be relevant, based on our investigation and tests, arranged according to the SWOT structure.

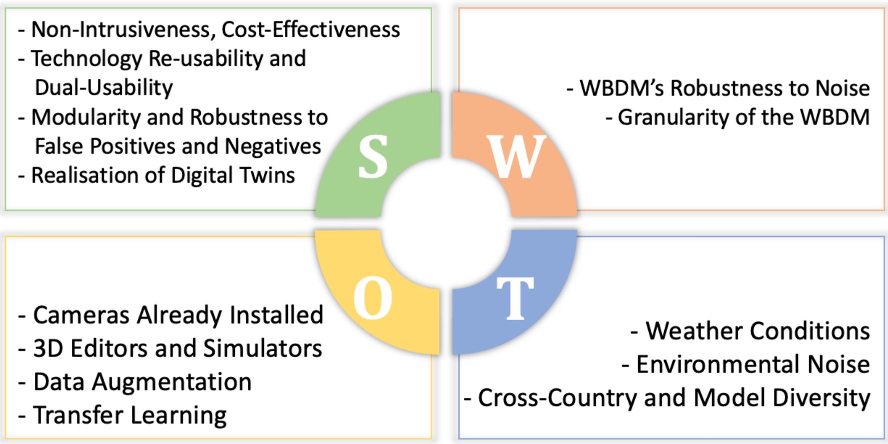


Fig. 4.2. Relevant aspects from the Level Crossing PoC arranged according to the SWOT Structure.

4.2.1. Strengths and Weaknesses

Non-Intrusiveness, Cost-Effectiveness, and Technology Re-usability. The architecture we proposed is centred around the usage of *non-intrusive* and *cost-effective* sensors (i.e., *cameras and microphones*) that, as mentioned, would possibly avoid any

[4] https://github.com/ultralytics/yolov5

re-approval process and railway traffic interruption. The cost-effectiveness is also related to the *technology re-usability* of such sensors: once installed for maintenance purposes, cameras can also be adopted to improve safety at LC as obstacle detection mechanisms (e.g., [19–21]) could be implemented by exploiting the same sensors.

Real-Time Analyses. The DL approaches we adopted have quite low computation times which make the system viable for *real-rime analyses* of the LC status.

Modularity and Applicability. The *modularity* of the architecture makes it suitable for (i.e., applicable to) almost any kind of LC, whether they involve all three components (i.e., WBs, WLs, and barriers) or only a subset of them.

Modularity and Robustness to False Positives/Negatives. Diving deeper into technical aspects, the approach we investigated can also be considered quite *robust to possible False Positives (FPs) and False Negatives (FNs)*. The DL approaches we implemented are not free from such issues, however, the aleatory of the environment and the modularity of the approach allow us to conceptually overcome these problems. For example, by taking into account the WBDM, the occurrence and repetition of FPs (i.e., when a specific sound is wrongly classified as a WB) would unlikely happen at each activation of the WBs. The same goes for FNs (i.e., a WB sound wrongly classified as non-WB). Differently, by taking into account the BAM, FPs and FNs can be easily fixed by applying a corrective algorithm in post-processing as described in [12]. In addition to that, the modularity of the architecture also helps in overcoming possible malfunctioning of the monitoring system itself since it can also implement a *triple modular redundancy behaviour*: if two out of three modules detect that the corresponding components are behaving correctly, while the remaining one detects a malfunction, either the component is actually faulty or the module itself is not behaving properly. A human operator can then remotely check the outcome of the system to understand whether a repair action is needed; alternatively, if the detected malfunction persists over multiple activations of the LC, it would be most likely a symptom of a faulty component rather than a malfunction of the monitoring system. Clearly, further tests would be required to ensure that systems' malfunctions happen as rarely as possible.

Contribution to the Realisation of LC Digital Twins. Another important aspect to consider is the impact that such a system can have on the realisation of *LC Digital Twins (DTs)*. Among the main enablers for the creation of DTs, IoT (Internet of Things) sensors are those that allow the DTs to evolve with their physical counterpart. Cameras and microphones, combined with DL approaches, could act as IoT sensors to extract useful information for the realisation of LC DTs. For example, instead of using intrusive IoT sensors to monitor, e.g., the velocity of the barrier, it would be possible to adopt the BAM we implemented to extract the position of the barrier over time. These data can be sent to the DT (in real-time) so that the virtual barrier behaves as the physical one.

Dual-usability: Improve Maintenance and Safety. A monitoring system of this kind (which relies on audio-video data) *can also bring benefits in terms of safety*. If properly developed, DL approaches can approximate human senses so that, besides analysing the correct functioning of LC components, they can also provide information about their visibility or audibility. For example, in relation to the WBDM, in Deliverable [12], we studied the robustness of the CNN against different types of noises at various intensity levels; the model resulted to be quite sensitive to high-intensity noises and, as a general role, the higher the contribution of the noise the worse the classification performances. From the perspective of maintenance, this may be considered a weakness of the approach, however, it can actually introduce opportunities in terms of safety. Even though the WB is correctly functioning, it is not said that it is clearly audible (e.g., given a disturbing noise) from the approaching road users leading, in conjunction with other possible malfunctions or poor visibility conditions, to potential threats. Therefore, the fact that the DL system we developed is not capable of correctly "hearing" (i.e., classifying) the WB could be a symptom of poor signal audibility also by road users. This suggests the potential *dual-use* of the proposed approach, which should be clearly further investigated in the future to be validated.

WBDM's Robustness to Noise. Despite the misclassifications due to a noisy environment may have some relevance in terms of safety, and despite the modularity of the approach can help to overcome misclassification issues, when it comes to the maintenance task it would be advisable to use systems as robust as possible. Therefore, further improvements would be required to make the WBMS less sensitive to noises.

Granularity of the WBDM. To conclude, at the current stage of development, the WBDM is capable of analysing audio-frames of about 1 second. Probably, a more fine granularity, i.e., the analysis of audio-frames with a smaller duration (in the order of milliseconds), could help to better understand the health status of the WB.

4.2.2. Opportunities and Threats

Cameras Already Installed. Recalling what was mentioned at the beginning of the previous section, cameras may be used for both safety and maintenance purposes. Worth underlining, there may be cases in which *cameras are already installed* for security or safety purposes, therefore, there would be the additional benefit of saving some installation costs.

3D Editors and Real-Life Simulators. As for implementability aspects, one of the main challenges we faced was related to the *lack of available data*. At the PoC level, *real-life simulators* (e.g., GTA5) offers the possibility to collect *synthetic video data* sufficiently approximating the reality in terms of graphics; on the other hand, they do not allow easy customisation of the assets to be monitored (e.g., it is not trivial to replicate faulty LCs). Opportunities in this direction are given by *3D Editors* like Unity[5] or Unreal Engine[6]. By exploiting software of this kind, reality could be properly approximated in terms of both graphics and simulation of the behaviour of the asset. However, the realisation of the asset itself would be quite more complicated than simply recording data from already implemented assets as happens with real-life simulators. In terms of *audio data collection*, opportunities are given by online repositories such as AudioSet[7] (and consequently YouTube[8]) which would allow for the creation of a dataset containing real-world sounds without the necessity of collecting them on the field.

Data Augmentation and Transfer Learning. In addition to what discussed so far, other techniques can be used to increment the amount of collected data such that AI models could be properly trained. We experimented *data augmentation* techniques and *transfer learning* approaches that, as described in [12], allowed us to improve the per-

[5] https://unity.com
[6] https://www.unrealengine.com/en-US
[7] https://research.google.com/audioset/download.html
[8] https://www.youtube.com

formances of the DL architectures we adopted.

However, it is worth mentioning that the datasets, obtained through the collection methods and approaches discussed at this point and the previous one, may not be fully representative of reality. They are extremely useful to develop PoCs, experiment DL models, and define suitable reasoning to address the problem; however, for the evaluation of the real effectiveness of the implemented architecture, real data should be used, possibly collected by directly considering the real asset to be monitored.

Weather Conditions and Environmental Noise. As for possible threats, *weather phenomena* and *environmental noises* might affect the performance of the proposed Intelligent LC Monitoring System. As for the former, in our dataset, we included different weather and light conditions (e.g., rain, snow, day/night light), additionally, through data augmentation, we simulated other phenomena such as sun flares and fog. Although the BAM (the only video-based module we implemented) resulted to be quite robust to these phenomena, and therefore these events should not affect the effectiveness of the DL approach, there is another aspect to consider in real scenarios. As mentioned, we exploited a *fixed camera*; this allowed us to first statically crop some areas of the original frame, and then apply the DL model to perform the detection. There might be some harsh phenomena that could shift the FoV of the camera and, consequently, compromise the functioning of the static cropping. Although these events are quite unlikely to happen, they are worth pointing out because these may imply some corrective actions to re-calibrate the FoV of the camera. On the other hand, concerning *environmental noises*, in the previous section, we have already discussed the effects they may have on the WBDM. From the perspective of maintenance, these phenomena imply that it might not be possible to properly evaluate the functioning of the WB at each activation. However, given the multiple activations of these assets during each day, even though the system would be not able to work perfectly in real-time, correct detections would most likely happen more frequently than the current inspections (performed on weekly or monthly bases).

Cross-Country and Model Diversity. To conclude, it is important to highlight a crucial aspect of the real implementability of the proposed approach. Our analyses highlighted that different countries (and different railway operators) adopt slightly or completely different sounds for the WBs. In addition, also WLs can vary in position, although they are typically red when active and brown/black when turned off. Hence, *despite the architecture we investigated being agnostic in terms of modules and corresponding DL models (i.e., the structure and the reasoning can be easily adapted to any kind of LCs)*, *it would not be possible to build a general-purpose system* capable of properly working with any typology of WLs and WBs. It would be advisable to train (or fine-tune) the DL approaches composing the Intelligent LC Monitoring System on data coming from the specific type of LC that is intended to be monitored. The main module that suffers this cross-country diversity is the WBDM, not only because of the difference among WBs themselves but also because we implemented that module through a three-class classification approach where one of the classes encompasses audios that are similar to WBs (so that the system do not confuse them with WBs). These kinds of audio should be properly selected according to the sound of the WB of the LC that is intended to be monitored. We believe that the approach we investigate would be capable of working with any kind of LC, however, it should be properly calibrated by taking into account the intrinsic characteristics of each type of LC. Detailed considerations are given in [12].

4.2.3. Recommendations from the PoC

Herein, we report all the recommendations that specifically came out from the analyses we conducted in the context of the PoC discussed above.

Combine Audio-Video Data with AI to Generate Digital Twins. A Digital Twin requires to be connected to its physical counterpart to evolve with it over time. IoT sensors are the main enablers to establish this connection but, typically, they are intrusively installed on the asset that should be digitalised. In the case of LCs, we think that it would be possible to pass to the DT about the same information that IoT sensors would pass without being intrusive. For example, instead of monitoring the physical parameters (e.g., the current signal) of the engine that moves the barrier, it is possible to extract the path the barrier traces over time by adopting DL Object Detectors and video data. Malfunctions of the engine would most likely reflect in a non-smoothed behaviour of the barrier; therefore, even though it is not exactly the same information, it would be possible to deduce such malfunctions by monitoring the *visible* behaviour of the barrier. Therefore, the behaviour of the various components can be extracted by means of DL algorithms leveraging audio-video data and passed to the DT so that the digital WBs, WLs, and barriers would behave as their physical counterparts.

3D Editors and Real-Life Simulators for Synthetic Data Generation. Tools of this kind represent suitable solutions to collect synthetic data that sufficiently approximate real phenomena. Real data would always be the best choice, but there are cases in which they cannot be easily collected or it would not be possible to collect all the possible events (whether in terms of weather/light conditions or malfunctions) in a timely manner given, for example, their rarity (e.g., high SIL systems rarely show malfunctions). In these cases, synthetic data could be leveraged to set up a suitable architecture and train and test the potential effectiveness of AI approaches.

Modular Approaches to Improve Efficiency. Almost every Machine Learning approach suffers from intrinsic errors (due to internal mathematical operations and approximations) which makes it extremely hard (nearly impossible in some cases) to obtain ML systems with an accuracy equal to 100%. Also, the prediction capability of a specific ML model depends on the dataset the model has been trained on and it is quite unlikely that a dataset can represent all the possible phenomena that would happen in reality. Therefore, if a dataset is a partial representation of reality, the ML model built on it is capable of processing only a subset of the real events; in other cases, it will most likely fail. By using (when possible) a modular approach, i.e., defining a DL model for each sub-component of the asset that is intended to be monitored, such issues can be somehow mitigated given that it would be quite unlikely that each DL model will fail at the same time. For example, it would be possible to use a single DL model (an object detector to be precise) to detect both the barrier and the status of each WL in a video frame simultaneously. However, in case a frame is characterised by particular features (also due to adversarial attacks) that would fool the network, both the barrier and WLs detection would be compromised. On the other hand, if there are two modules, one that monitors the WLs and one that monitors the barrier, which are trained on two different datasets, the probability that both of them will fail at the same time decreases.

Facilitate AI Tasks. Instead of adopting a single AI model to solve a complex task (e.g., detecting both the WLs and the barrier as in the example above), it is advisable to subdivide it, if possible, into multiple – possibly easier – sub-tasks. Each of them can then be handled by a separate AI model, specifically designed for its resolution. In addition to that, task-specific characteristics can be leveraged to further facilitate the sub-task that AI models should address. As an example, this approach could be adopted within the LC PoC when implementing the WBDM. By leveraging the fixed camera's characteristics, it would be possible to pre-process video frames and statically crop out the WLs from the frames; this would allow for the adoption of a classification approach to identify the WLs' status, as opposed to using object detection methods which are typically more complex (in terms of architectures) and may also be affected by issues like small-scale detection.

Exploit Audio Data for Maintenance Purposes. The analysis of the State-of-the-Art conducted within the first phase of the RAILS project showed a strong deficiency in the use of audio-based techniques in conjunction with Machine Learning and Deep Learning to support maintenance activities in railways but for a few works addressing rail tracks. Based on our experiments with the LC PoC and due to its known potential of warning maintenance operators about defects, the exploitation of audio data through DL is an aspect worthy of further investigation and extensive evaluation.

Exploit AI to Increase Safety at LCs. DL approaches analysing audio and video data could also approximate human behaviour. For example, as discussed above, misclassifications of the WBDM given a heavily noisy environment could be a symptom of the fact that the WB is not properly audible for road users as well. We think this aspect is quite relevant and could be exploited to analyse WBs from the perspective of safety if further investigated.

5. A Critical Examination of the Rolling Stock Proof-of-Concept

Rolling Stock Rostering allows for efficient allocation and utilisation of limited railway resources including, among others, locomotives, wagons, and maintenance facilities. By carefully scheduling the maintenance activities with the consecutive passenger services, the downtime of rolling stock units can be minimised and, at the same time, the maximum number of trains available for services at any given time could be ensured. Effective rostering helps maintain the reliability and availability of rolling stock units, such that the overall costs of identifying and resolving potential issues can be reduced proactively. With a more efficient maintenance schedule, unnecessary maintenance can be avoided but the essential check at a given time point will be guaranteed. By incorporating maintenance activities into the rostering process, railway operators can adhere to safety regulations and industry standards. Regular inspections, repairs, and component replacements help prevent accidents caused by faulty equipment. Furthermore, rostering provides a structured framework for planning and coordinating maintenance activities. It enables better coordination between maintenance teams, depots, and operational staff. By having a well-defined schedule, maintenance tasks can be prioritised, resources can be allocated efficiently, and potential conflicts or overlaps can be minimised. Typically, Rolling Stock Rostering with maintenance activities is implemented through a combination of manual planning and the use of specialised software tools. It is important to note that the implementation of rolling stock rostering with maintenance can vary across railway operators. Some may rely on manual planning processes, while others may utilise advanced techniques or integrated maintenance management systems. The level of automation and sophistication depends on the complexity of the railway network, the size of the rolling stock fleet, and the specific requirements of the operator. With this PoC, we intend to explore to what extent it would have been possible to leverage AI, specifically Reinforcement Learning (RL), to improve the process of rolling stock rostering and maintenance scheduling.

To conclude, the aim of this PoC was not to propose a practical solution within the content of this background. Alternatively, we seek to determine the degree to which reinforcement learning techniques can be introduced in intelligent rolling stock rostering and predictive maintenance scheduling with the ultimate goal of emphasising suggestions for further study in this area.

5.1. Recent Advancements on RL-based Rolling Stock Rostering

This section provides a brief history of rolling stock maintenance and rostering task, as well as some of the most recent advancements in smart maintenance scheduling for locomotives. The majority of maintenance work has been conducted at the workshop site based on a predetermined schedule, without considering any incidents or disruptions that may occur on the railway track. Moreover, it is important to note that maintenance tasks are closely linked to subsequent rostering activities [22]. However, very few current approaches allow for the efficient scheduling of maintenance activities that might arise between consecutive train services. This is one of the fundamental driving forces behind the paradigm shift that has occurred among railway stakeholders, from manually scheduling-based maintenance to flexible scheduling and preventative maintenance [23]. To gather and interpret operational data in real-time to determine their health state for a given time point and schedule the upcoming schedule event proactively. The primary drawback of this framework comes from the application of RL, which may arise with the following concerns:

- RL algorithms typically balance exploration (trying out new actions to learn) and exploitation (utilising learned policies to exploit known good actions). In the context of rolling stock rostering and maintenance scheduling, exploration can lead to disruptions and potential risks if non-optimal actions are implemented. In other words, tremendous consequences would be generated if an RL agent posed an action totally based on what has been learned from the past without considering the actual scenarios. However, striking the right balance between exploration and exploitation can be challenging.
- RL algorithms often operate as black boxes, making it inherently difficult to be interpreted and explicitly understood in the decision-making process. Providing explanations for the generated schedules and maintenance plans is the most significant part for those who do not have sufficient expertise in machine learning models. The lack of interpretability can hinder trust and acceptance among stakeholders, especially when dealing with critical operations.
- Railway operations are subject to various uncertainties, such as changes in demand, disruptions, maintenance delays, and unforeseen events. Incorporating RL into rolling stock rostering and maintenance scheduling must account for these dynamic and uncertain environments, adapting the learned policies to changing conditions.
- Introducing RL into an established railway system may require integration with existing scheduling and maintenance management systems. Ensuring compatibility, data sharing, and seamless coordination between RL algorithms and existing processes can be a complex task.

5.1.1. Bird-eye View of the PoC Approach

The problem of this study can be summarised as follows: given timetables, rolling stock assets, maintenance workshops, and maintenance operations, a rolling stock circulation solution must be calculated with the least amount of expenditure, which is expressed in terms of the number of used rolling stock units, the number of used empty runs, and the number of train movements between platforms at stations and within workshops. Specifically, as shown in Fi.g 5.1, the time given by roster between two train services can be decomposed into: i) waiting time at the passenger station before the train is brought to the workshop; ii) time to provide empty run to the workshop; iii) maintenance processing time window (including recovery time); iv) time to provide empty run to passenger station; and v) waiting time at the passenger station after the train leaves the workshop.

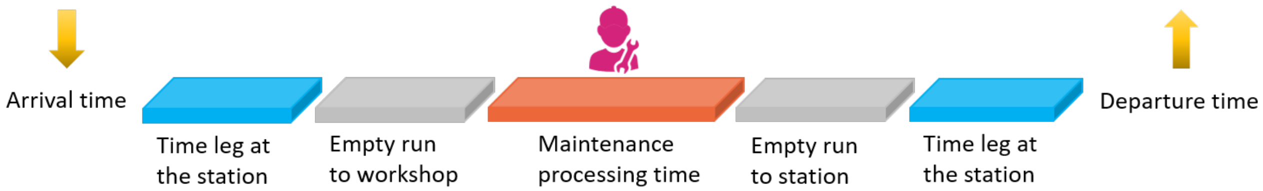


Fig. 5.1. A typical cycle of RS unit rostering with maintenance.

Fig. 5.2 shows a small graph to illustrate the problem formulation. For this particular train service, there are several round nodes with capital labels indicating departure and arrival stations, plus the associated running times. The green (solid) arcs indicate paired services, the purple (solid) arcs represent the empty rides without maintenance, the purple (dotted) arcs denote the empty rides with maintenance tasks, and the green (dotted) arcs represent the maintenance tasks without empty rides. The numerical labels shown besides arcs are the time costs of the empty runs and maintenance tasks (if any). A solution is a Hamiltonian path with one or more maintenance arcs to guarantee the maintenance expiry. In the solution the empty rides (purple arcs) are optional. The objective function of the problem is the minimisation of the number of days included in the roster, i.e., the number of trains required to perform all services in a day.

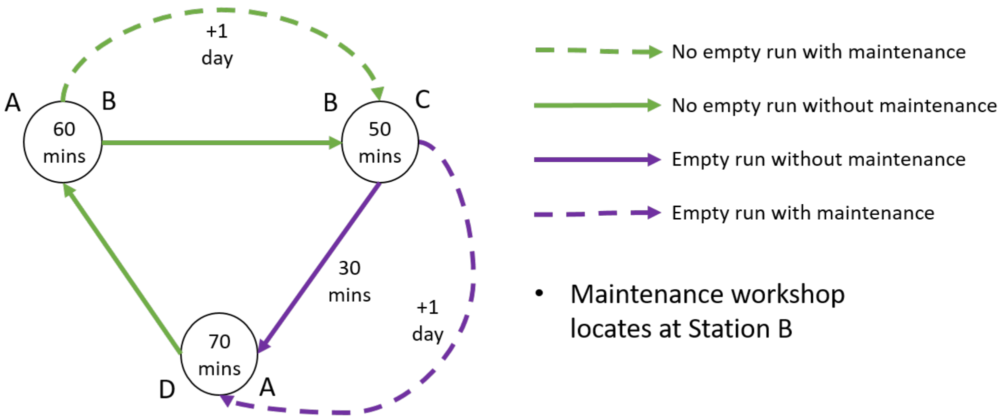


Fig. 5.2. An illustrative example for the problem.

Figure 5.3 gives information about what the proposed Reinforcement learning-based RS rostering/maintenance framework looks like. From the diagram we can easily conclude that this approach intakes the data as input from four different aspects: 1) Rolling stock assets data; 2) Train operational timetables; 3) the features of the considered maintenance workshops; and 4) information of each kind of maintenance activities. The module of 'Rostering' and 'Maintenance' run in parallel with the objectives of minimizing the total costs of rolling stock units. In the component of the Reinforcement Learning operator, several essential components such as Agent, Environment, States, Actions, as well as Rewards have been well-defined.

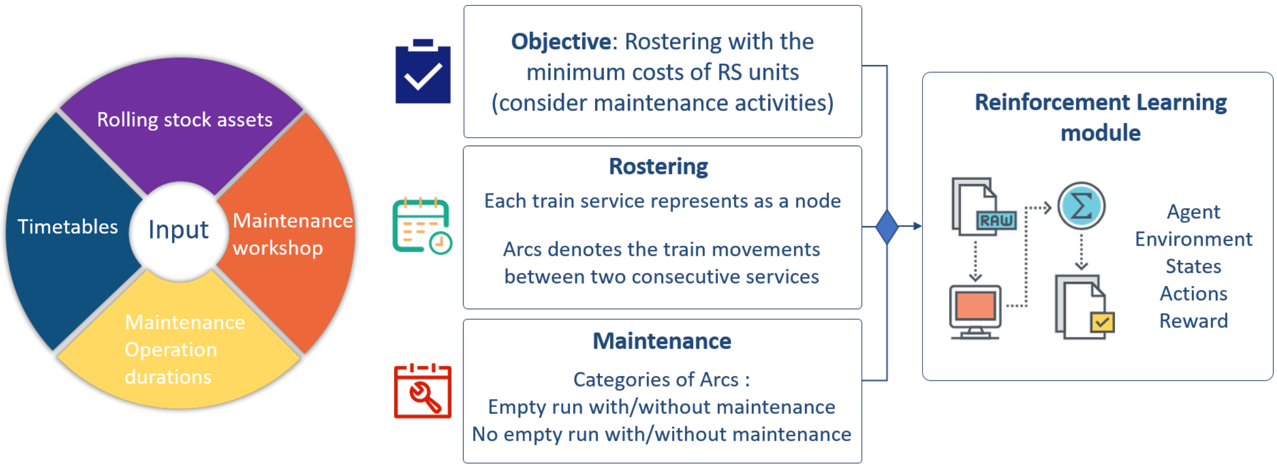


Fig. 5.3. The high-level architecture of RS rostering with maintenance tasks.

5.2. A SWOT Analysis of the PoC

As for the previous case study, in this section, we highlight the Strengths (S) and the Weaknesses (W) of the proposed approach in the context of the Rolling Stock maintenance scheduling and identified some external Opportunities (O) and Threats (T) that specify some support evidences or challenges that caused by technical implementation. This SWOT analysis is not oriented at defining the market viability of the solution we investigated, instead, it has been used to evaluate the approach in a structured manner. Fig. 5.4 reports all the aspects that we found to be relevant, based on our investigations, arranged according to the SWOT structure.

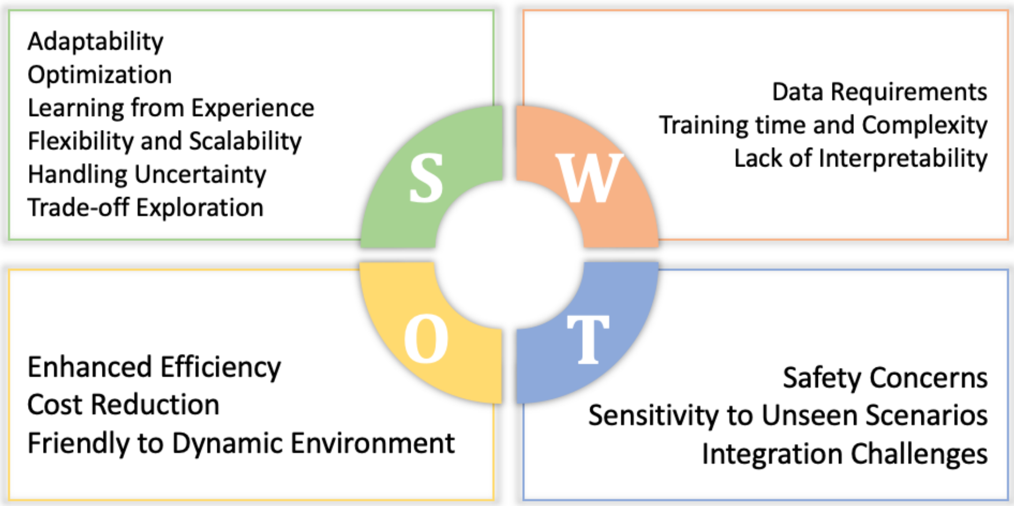


Fig. 5.4. Relevant aspects from the Rolling Stock PoC arranged according to the SWOT Structure.

5.2.1. Strengths and Weaknesses

Adaptability to various policies. RL can be implemented into dynamic and complex scenarios, making it a good fit for the ever-changing nature of railway operations. It can learn and adapt policies based on feedback from the environment, allowing it to adjust to new situations, changing schedules, and evolving maintenance requirements.

All these advantages are built on the usage of flexible interactions between action-reward-state pairs, rather than the hard-coded optimisation constraints

Flexibility and Scalability. RL can handle large-scale problems and can be scaled to encompass multiple trains, stations, maintenance activities, and scheduling constraints. It can consider numerous variables simultaneously, optimising the allocation and circulation of rolling stock units while considering maintenance needs, train schedules, and other relevant factors.

Cooping with Uncertainties. Railway operations involve various uncertainties, such as delays, disruptions, and unforeseen events. RL can handle such uncertainties by incorporating probabilistic models or by learning from historical data, making it capable of dynamically adjusting decisions in response to changing conditions.

Trade-off Exploration and Optimisation. RL algorithms can explore the trade-offs between conflicting objectives, such as minimising time costs while considering maintenance activities. By designing appropriate reward functions, RL can strike a balance between different goals, allowing decision-makers to prioritise different factors based on their preferences and constraints.

Data Availability. RL algorithms typically require substantial amounts of data to effectively learn and generalise optimal policies. In the context of rolling stock rostering/circulation, obtaining high-quality and diverse data may be challenging, especially if historical data is limited or unavailable. Insufficient data can hinder the RL agent's ability to learn accurate models of the environment and make optimal decisions.

Complexity of the Approach. RL algorithms often require significant computational resources and training time to converge to optimal or near-optimal policies. The complexity of the rolling stock rostering/circulation problem, including the consideration of maintenance activities, may amplify the training time and computational requirements. Training RL agents in such complex domains may involve long training periods, making it impractical for real-time decision-making.

Interpretability Issues. RL agents often operate as black boxes, making it challenging to interpret the decision-making process. Understanding why the RL agent selects certain actions or makes specific recommendations can be difficult, which may raise concerns in safety-critical domains like railway operations. Ensuring transparency and interpretability of the learned policies can be important for gaining trust and acceptance from stakeholders.

5.2.2. Opportunities and Threats

Efficiency. RL has the potential to significantly improve the efficiency of rolling stock rostering/circulation by dynamically optimising the allocation and circulation of rolling stock units. By learning from data and interactions with the environment, RL can discover more efficient scheduling strategies, leading to reduced downtime, better utilisation of resources, and improved overall system performance.

Cost Reduction. Optimising rolling stock rostering and maintenance activities can result in cost reductions. By minimising overall time costs, such as reducing delays, improving maintenance planning, and optimising train schedules, operational expenses can be reduced. This can lead to cost savings for railway operators and potentially lower fares for passengers.

Well-respond to Dynamic Environments. Railway operations often face dynamic and unpredictable conditions, such as changing demand patterns, unexpected incidents, or maintenance requirements. RL's ability to adapt to changing circumstances and learn from experience makes it well-suited for addressing these challenges. RL-based solutions can continuously adjust and optimise rolling stock rostering and maintenance activities in response to real-time data and evolving operational conditions.

Safety Concerns. Safety is of utmost importance in railway operations. Implementing RL-based solutions introduces a level of complexity and uncertainty, which can increase the risk of unforeseen or undesirable behaviours. It is crucial to thoroughly validate and test RL models to ensure they do not compromise safety regulations or result in hazardous situations.

Sensitivity to Unseen Scenarios. RL models trained on historical data may not generalise well to future or unseen scenarios. Changes in the operating environment, such as shifts in demand patterns, the introduction of new rolling stock units, or modifications to maintenance schedules, can affect the performance of RL models. Continuous monitoring and adaptation of the RL system are necessary to ensure robustness and responsiveness to evolving conditions.

Integration Challenges. Integrating RL-based solutions into existing railway systems can pose challenges. Existing infrastructure, legacy systems, and operational constraints may need to be considered during the implementation process. Collaborative efforts between domain experts, data scientists, and railway operators are essential to address integration challenges effectively.

5.2.3. Recommendations from the PoC

Hereby, we report all the recommendations that specifically came out from the analyses we conducted in the context of the PoC discussed in this chapter.

Multi-Objective Optimisation. Investigating techniques to handle multiple conflicting objectives simultaneously. Developing RL-based approaches that can optimise rolling stock rostering and circulation considering not only time costs and maintenance activities but also other factors like energy consumption, passenger comfort degree, as well as operational resilience. Multi-objective RL algorithms and methods can help identify trade-offs and Pareto optimal solutions.

Integration with Predictive Maintenance. Consider integrating RL-based rolling stock rostering/circulation with predictive maintenance models. By leveraging predictive maintenance techniques, such as condition monitoring or failure prediction, RL can proactively schedule maintenance activities, minimising unexpected breakdowns and optimising the availability of rolling stock units.

Transfer Learning and Generalisation. Explore techniques for transferring learned policies from one railway system to another or from one time period to another. Investigate the potential of transfer learning and generalisation methods to reduce the training time and resource requirements when deploying RL-based solutions in different contexts or when dealing with similar rolling stock rostering/circulation problems.

6. Recommendations and Innovation Needs

6.1. Share the Knowledge

There is a lot of done and ongoing work on the application of AI in railway maintenance, as already recalled in Section 3. Such work is carried out both in industry and in research institutes and universities. This huge effort requires and produces knowledge and expertise which can feed further research and development. Therefore, one of the most important keys to the fast take-up of AI is the spreading of multi-disciplinary knowledge and expertise with the objective of sharing one or more of the following, respecting the security and privacy policy of each organisation:

- **The current state of practice**. It is very difficult to build a picture of the existing applications and results. As a first step *a shared showcase* could be created, just by collecting public information provided by companies and research institutions about available AI applications for maintenance (but the same could be done for other domains).
- **The gained experience**. It would be useful to share the experience about the use of technologies, AI techniques and models for a given task or application, also sharing information about roads that have *not* brought satisfactory results.
- **The enabling technologies and the resources needed to achieve results**. This will allow the understanding of the computing and hardware requirements for each application and maintenance task and the reached level of digitalization required/achieved (for example IoT devices, GPUs, cloud services, edge and fog computing).
- **The open problems that still deserve further investigation**. Reporting unresolved challenges that require additional investigation in the application of AI techniques and sharing documentation of problems that have yet to be adequately addressed from the perspective of railway companies and operators would greatly motivate the research to produce solutions.

The list above contains only some of the main information about the status of the play and the necessary skills, tools, and resources needed to foster research and innovation in AI-powered maintenance. Of course, software frameworks exist and can be used to support the suggested process of information gathering. Knowledge sharing is a matter of method and policy choices.

6.2. Deal with the Data Challenge

As far as we know, there is no gold solution to overcome the various challenges coming from the lack of suitable data that could be exploited to build AI systems in railways. Here, we point out two different types of intervention that, in our view, could be adopted to cope with this problem. The first delineates a structural approach at the European level, while the latter encompasses some specific technologies and methods that could be adopted to mitigate the problem mentioned above

6.2.1. A Structural Solution to Data Availability

In our vision, the ultimate recommendation to effectively promote the integration of AI and feed innovation in the rail sector is the establishment of a European Framework for AI application, allowing the sharing of the resources needed for scientific and industrial research. This goal is ambitious and difficult to reach in the short to medium term, nonetheless, it could be pursued incrementally. The framework can be built according to a bottom-up approach, starting with more affordable objectives. The main elements needed for the realisation of this vision are reported in the following in the form of recommendations:

Maintain a collection of public datasets in railways. This is a short-term objective. Within the RAILS project, we have already put some effort into collecting references to publicly available railway datasets [5, 7]. However, the objective here is to make these references easily accessible to all practitioners and researchers through, e.g., a simple web page. The main issue would be to keep the information up to date, i.e., guaranteeing the maintenance of the website in the long term.

Build a Common Conceptual Data Model for classes of railway applications. This is a medium-long-term objective. Some effort is already ongoing in this direction in EU-RAIL (e.g., LINX4RAIL [24] and LINX4RAIL-2 [25]). Therefore, the main issue is not the time needed for the full definition of these models, but the time necessary for reaching a consensus and implementing the necessary steps for their concrete adoption.

Develop standards for datasets to be used in AI applications. This is a medium-long-term objective whose fulfilment in part relies upon the existence of the Conceptual Data Model. With the term standard, we here refer to the definition of a set of guidelines and procedures that would allow for the construction of suitable datasets to properly train and test AI models for specific tasks. By *suitable* we mean datasets including all the relevant pieces of information to properly characterise the problem or the events that are intended to be analysed/predicted through the AI model. Aspects that should be considered in this context include, among others: a well-defined data format; data variety (e.g., data collected under different conditions); an adequate number of entries to make the training viable; coherent data annotation and labelling; and the peculiarity of the problem that is intended to be addressed which can facilitate the task that AI models should address (see "Facilitate AI Tasks" in Section 4.2.3 for additional details).

Standards for the creation of datasets would be extremely useful for the following reason. Given the lack of data as better discussed below, when addressing *the same problem*, researchers and practitioners typically collect data from scratch by setting some collection criteria and adopting some specific annotations which may differ from study to study. This means that different researchers may train AI models on datasets that, despite being related to the same problem, may have completely different characteristics. In turn, this lead to the impossibility of comparing AI solutions since it is not said that a model that performs properly on a dataset would work equally well on another dataset. Also, it may be possible that datasets are created without considering some characteristics that could be crucial for the resolution of a specific problem. Standards would help to set common procedures so that, in the future, data would be collected, labelled, and processed according to a set of ad-hoc and reliable roles by all the researchers that are intended to solve a specific problem; in addition, this would also lead to the generation of datasets (and, consequently, AI models) that would be comparable to some extents.

Establish Consortia for data utilisation. This objective can be pursued at different levels and with different goals that span from agreements for (controlled and reliable) data sharing to agreements for the protected usage of proprietary data. In this context, two main directions can be identified:

- *Data Sharing*;
- *Model Sharing*.

Both of them expect different parties (e.g., railway stakeholders and research institutions) to cooperate with the aim of sharing knowledge to pursue the technological advancement of AI in railways.

Data Sharing expects to build a common database collecting railway datasets to be used to train and test AI models. Work in this direction is being carried out within the project TAURO [26].

On the other hand, *Model Sharing* expects that each part trains an AI model on proprietary data which are kept confidential; then, instead of sharing data, they will share the trained AI model. The main technologies and strategies we identified that can be adopted to move towards this direction are, among others:

- *Federated Learning* [27], which is a machine learning technique allowing the training of a model on distributed data, that is through independent sessions each using its own dataset without data exchange.
- *Secure multiparty computation* [28], that is a cryptographic technology that allows privacy-preserving computation. The ultimate goal with respect to data is exactly the same as the one expressed above, i.e., multiple users can jointly work on the same task using their own data while keeping their data private, but this technology works at the cryptographic level.
- *Distributed Ledgers and Blockchain*, which allow recording, sharing, and synchronising transactions in a decentralised way guaranteeing that the stored information is immutable. Blockchain is a type of distributed ledger. A study on the usage of these technologies in railways has already been conducted in EU-RAILS (e.g., the B4CM project [29]).

6.2.2. Technologies and Methods to Cope with Limited Data

There are some approaches that can be exploited to cope with the problem of data availability. In our view, these can be subdivided into four macro-categories:

- **Generating synthetic data from scratch.** These include, among others:
  - The usage of *real-life simulators* (e.g., GTA V) or *3D editors and scenario simulators* (e.g., RoadRunner, Unity, or Unreal Engine).
  - The usage of Digital Twins, which is better discussed in Section 6.6.
- **Generating proper annotations and labels.** Data, whether synthetically generated or real, must be properly annotated according to the task to be solved and the AI model that is intended to be adopted. It is typically performed manually by experts, which results to be extremely time-consuming in most cases. We identified three options to facilitate the labelling process:

- In the case of synthetic data, annotations can be automatically retrieved during the data generation process of. For example, 3D editors could include functionalities to automatically label scenes and objects (e.g., Unity's Perception Package[1]).
- In case of real data, or if the aforementioned option cannot be applied, they could be used labelling software and tools including functionalities that could support and facilitate the labelling process[2,3,4,5].
- As the last option, it could be possible to implement semi-automatic labelling processes – based, e.g., on self-training – to generate labels for larger datasets starting from a small subset of hand-labelled data. An approach of this kind has been adopted in the context of the Obstacle Detection PoC we developed within WP2; Deliverable D2.3 [30] can be visioned for further details.

- **Extending existing datasets.** These include all the Data Augmentation techniques, from the traditional ones (e.g., cropping, stretching, etc.) to those leveraging advanced generative models (e.g., Generative Adversarial Networks - GANs).
- **Leveraging existing knowledge.** These mainly include Transfer Learning approaches oriented at exploiting the knowledge acquired to solve a *source* task to address a *target* task for which only a limited amount of data is available [31].

Artificially generated data are a good solution to the problem of the lack of data. Nonetheless, the methods mentioned above may have some limits. To fully exploit their potentiality in railways, the quality of synthetic data must be evaluated. Therefore, **an experimental campaign on a set of relevant real scenarios should be conducted to assess the usage of these techniques in the railway domain**. The generated datasets should be evaluated against key dimensions, including the utility for the specific task and the correlation between simulated and real events/scenes (i.e., if the simulations recall reality with enough fidelity).

6.3. Promote Public Contests based on Selected Case Studies and Benchmarks

The setup of contests and challenges (to be intended as competitions) may contribute to the fast take-up of AI in railways. Challenges like those proposed on the MS COCO[6], ImageNet[7], and RailSem19[8] datasets have caught the attention of researchers worldwide, who basically compete to achieve the best AI solution to overcome the proposed task(s). The same can be done for other railway problems, by establishing challenges for those that are intended to be addressed so that all the interested researchers and practitioners could participate in their resolution.

One of the main problems is that a challenge requires a shared dataset that all the participants should use to train and validate their AI models. In our view, this can be mitigated by

[1] https://docs.unity3d.com/Packages/com.unity.perception@1.0/manual/index.html
[2] https://it.mathworks.com/help/vision/ug/get-started-with-the-image-labeler.html
[3] https://www.v7labs.com
[4] https://keymakr.com/annotation-tool.php
[5] https://segment-anything.com
[6] https://cocodataset.org/#home
[7] https://www.image-net.org/challenges/LSVRC/index.php
[8] https://wilddash.cc/benchmark/summary_tbl?hc=semantic_rob

generating synthetic datasets as it has been done for some of the PoCs developed within the whole RAILS project. This solution also avoids any problem related to data privacy.

A further aspect that would require much attention is the definition of the challenge itself. First, each challenge should encompass a set of requirements and constraints that should be adequately established by experts; second, the dataset for the challenge should be properly collected or generated to reflect these characteristics. Therefore, the specification of the problem to be addressed, the definition of the guidelines for the realisation of the dataset, the generation/collection of the required data, and the management of the results of the competitions should require a centralised approach (e.g., basing on an agreement between railway stakeholders and/or railway experts)

Despite the latter could be a quite tough problem to overcome, we think that, given the various achievements obtained within the various AI competitions including those mentioned above, these kinds of actions could effectively promote advancements in AI-powered technologies in railways.

6.4. A vision of a European Railway Lab for AI application

A vision of a collaboration space supporting the investigation and the fast take-up of innovative AI-powered applications in rail transportation can be derived from Sections 6.1, 6.2, 6.3. This vision envisages collaboration among railway players, including industries, rail operators, and research communities, entering into alliances (i.e., consortia) to share knowledge, data, and models. Different kinds of activities are possible, aimed at both the general public, to foster investigation and research even outside the consortia, and at the members of the laboratory, who can join one or more alliances depending on the objectives that they pursue and on the type of information that they can share. Both data and model sharing can be achieved, as it is explained in subsection 6.2.1, but the effective implementation of joint activities also relies on the availability of a conceptual data model and standards for dataset construction.

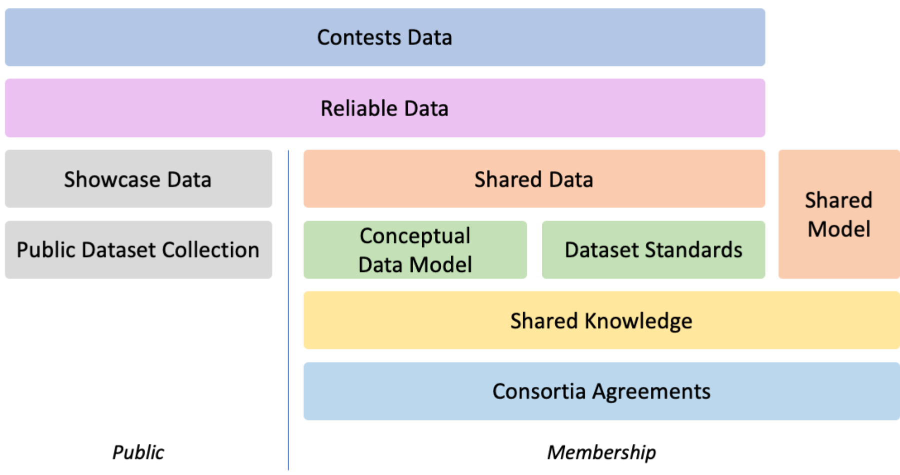


Fig. 6.1. A data-centric view of an AI-lab for joint research in railways

Figure 6.1 provides a high-level picture of this **vision** from the point of view of data. The vision presented here does not propose or introduce a possible reference architecture but rather a collection of crucial elements for Sharing Models and *Proposing Con-*

tests/Challenges. Although many of these elements have been discussed earlier, it is important to emphasize the concept of **"Reliable Data"**. It refers to data that have undergone processing and evaluation to be considered trustworthy. Whether these data come from *Public Dataset Collections* or are provided by a *Consortium*, they must meet certain criteria to be deemed reliable. This includes certification of their origin for security purposes and validation to protect privacy, reduce potential bias, and ensure data quality, accuracy, and fairness.

6.5. Exploit Cloud-Fog-Edge Computing Paradigms

The possibility of collecting and elaborating data at different levels – from IoT devices (e.g., used to monitor the rail system components) to the cloud – may greatly improve the performance and effectiveness of AI solutions for *autonomous maintenance and planning*.

Cloud, Fog, and Edge computing paradigms can be exploited to distribute knowledge and computations to efficiently monitor railway assets and improve the safety of railway systems, especially if combined with Digital Twins as discussed in Section 6.6.

Data collected at the **Edge Level** – i.e., through IoT sensors directly applied on railway assets like trains, LCs, and so on – can be propagated to the **Fog Level**, which is intended to monitor a cluster of railway assets (e.g., those operating on an entire railway line), and to the **Cloud Level**, that would allow for large data storage and high-performance computing, enabling the development of remote large-scale data centres which are provided as services over the Internet. The Cloud Level would then store data from different assets, possibly deployed worldwide, to build a larger knowledge that can be used to efficiently train and test AI predictive algorithms.

The same Cloud-Fog-Edge view works for the distribution of the concerns among AI-aided systems leading to what can be defined as **distributed intelligence**. AI systems deployed at the Edge Level would be in charge of managing compelling events that could affect railway assets (e.g., on-board components fault prediction, on-boards obstacles detection, etc.); those deployed at the Fog Level would monitor clusters of assets and manage their interaction to achieve a comprehensive optimisation of railway systems (e.g., implement Virtual Coupling functionalities); lastly, AI systems deployed at the Cloud Level would exploit a larger knowledge base to, e.g., efficiently identify maintenance actions to be taken and when to take these actions (maintenance planning). However, in our view, despite having great potential for the integration of AI in railways, the paradigms of cloud, fog, and edge computing are quite far to be practically implementable mainly because **they would require a shift from a local corporate view (data and computational resources) to a multi-tenant model**.

Besides maintenance, *distributed intelligence* is a very important aspect for the development of autonomous systems and, in particular, for the implementation of intelligent control; therefore, this point is discussed in more detail in Deliverable D2.4.

6.6. Seize the Opportunities Offered by the Digital Twins Technology

As widely recognised in the literature [7, 10], Digital Twins (DTs) can offer several opportunities when it comes to monitoring and analysing systems in real time, including railway assets. For the sake of knowledge, we refer to a DT as *"an accurate model of a physical entity which is kept alive at run-time and updated with real-time data collected from (IoT) monitoring devices"* [7]. In other words, at each instant in time, a DT is a digital replica of the corresponding physical asset and represents its current state of evolution. Herein we list two of the main opportunities that we think DTs could open for in railways:

Cognitive Digital Twins (CDT) for Predictive Analyses. Found that DTs evolve with the corresponding assets, it would be possible to make copies of DTs (which represent the physical assets at a given instant in time) and stress them with various inputs to study the potential evolution of physical systems *and/or generate data that can be used to train AI models*. CDTs [32], i.e., DTs enhanced with cognitive capabilities making them capable of learning and reasoning, would introduce several opportunities in this direction as this kind of DTs not only evolve with the physical counterpart but would also be able of simulating their behaviour.

Integrating DTs with Distributed Intelligence Levels. In the previous section, we have mentioned some Levels of Intelligence (i.e., Edge, Fog, and Cloud). The integration of DTs within these levels could allow for the implementation of some protection mechanisms in order to promptly manage faulty systems and avoid unpleasant consequences. An example, from a high-level perspective, of how these protection mechanisms could be implemented is given in Fig. 6.2. Assuming it would be possible to generate DTs for both trains and LCs, these can be exploited at the Fog Level to promptly adopt countermeasures in case of failure of the LC. Additional details are given in Deliverable D3.2 [11].

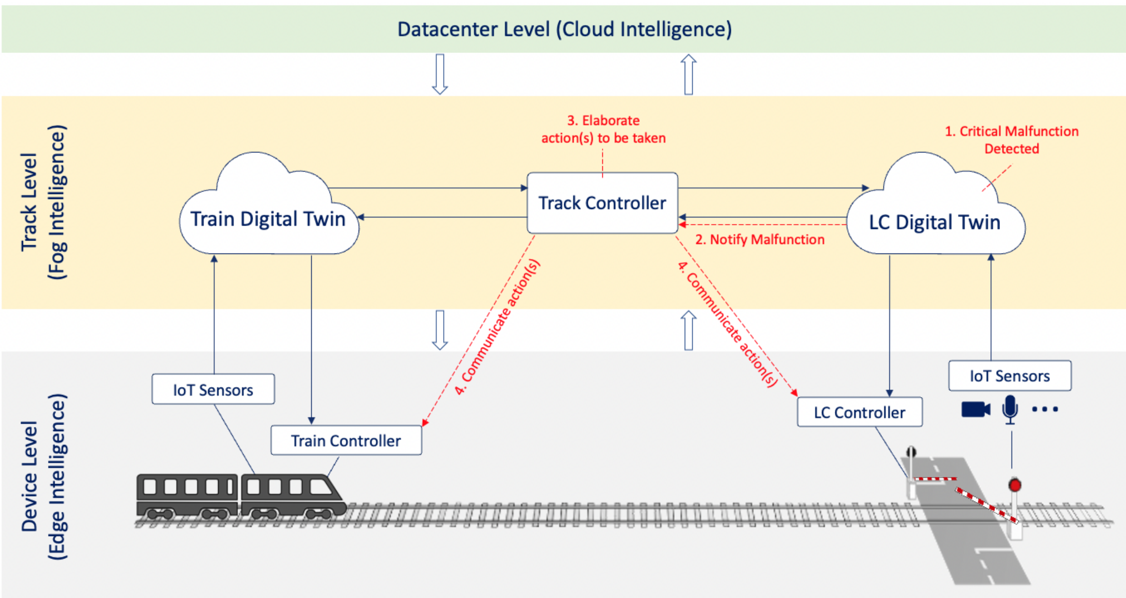


Fig. 6.2. Integrating DTs and Levels of Intelligence to Ensure Protection in Case of LC Failure (excerpted from [11]).

6.7. Investigate Approaches for (Semi-)Automatic Generation of Digital Twins

As discussed above, DTs introduce benefits for the development of AI applications in railways. However, as emerged from some preliminary investigations we performed on this topic, it seems that guidelines for the generation are still missing. We tried to formalise these guidelines [7,10] into some steps that could guide the realisation of a DT starting from the identification of the asset to be digitalised to the implementation of the DT; the guidelines also include insights on the integration of AI functionalities.

In our view, the main gaps to be full-filled regard: i) reference standards for the development of DTs (which is something that has already been under the spotlights in the last years [10]), that would help to underline in detail development requirements and possible evaluation metrics; and ii) procedures that, starting from a set of specifics, could be exploited to (semi-)automatically generate (i.e., synthesise the design) DTs.

7. Conclusions

This document reported the identification of possible future innovation needs and recommendations in the railway industry to enhance maintenance and inspection activities. It addressed a detailed analysis of the two proofs-of-concept (PoCs) proposed in the previous WP3 deliverables, namely, “Smart Maintenance at Level Crossings” and “AI-assisted Rolling Stock Rostering”. The report critically examined the outcomes of the PoCs, discussing recent advancements in their respective fields, and conducting SWOT analyses to identify the main Strengths, Weaknesses, Opportunities, and Threats. Some recommendations emerged that aimed to address the identified challenges in order to enhance the technical and operational feasibility of the proposed AI approaches.

Specifically, as for “Smart Maintenance at Level Crossings”, recommendations are mainly oriented at: i) highlighting the benefits that 3D editors could introduce when it comes to data collection for the fast realisation of PoCs; ii) promoting the exploitation of task-specific characteristics which help to facilitate the tasks that AI model should face; iii) emphasising the contributions that AI and audio-video data can bring in the realisation of Digital Twins; iv) underlining the benefits of using modular approaches; and v) suggesting the possible dual-use of approaches leveraging audio-video data which could be exploited also for safety purposes.

On the other hand, regarding “AI-assisted Rolling Stock Rostering”, recommendations encompass: i) investigating techniques for multi-objective optimisation to simultaneously handle multiple conflicting objectives; ii) integrating predictive maintenance models to improve the scheduling of maintenance activities; and iii) exploiting transfer learning and generalisation methods to reduce the training time and the resources required by Reinforcement Learning models.

Eventually, general recommendations and innovation needs for future developments in railway maintenance and inspection activities have been drawn. These include sharing knowledge among railway stakeholders and research institutions, dealing with data challenges related to the lack of data availability, promoting public challenges based on specific case studies, establishing European laboratories, exploiting cloud-fog-edge computing paradigms to organise AI solutions at different levels of applications, and sizing the opportunities offered by Digital Twins together with the possibility of semi-automatically generate them.

All the recommendations presented in this document will converge into the definition of roadmaps of AI in railways which will be discussed in our next Deliverable 5.3.

Bibliography

[1] RAILS, "Deliverable D1.1: Definition of a reference taxonomy of AI in railways," October 2020. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[2] N. Bešinović, L. De Donato, F. Flammini, R. M. P. Goverde, Z. Lin, R. Liu, S. Marrone, R. Nardone, T. Tang, and V. Vittorini, "Artificial intelligence in railway transport: Taxonomy, regulations, and applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 011–14 024, 2022.

[3] RAILS, "Deliverable D1.2: Summary of existing relevant projects and state-of-the-art of AI application in railways," 2021. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[4] R. Tang, L. De Donato, N. Besinović, F. Flammini, R. M. Goverde, Z. Lin, R. Liu, T. Tang, V. Vittorini, and Z. Wang, "A literature review of artificial intelligence applications in railway systems," *Transportation Research Part C: Emerging Technologies*, vol. 140, p. 103679, 2022.

[5] L. D. Donato, F. Flammini, S. Marrone, C. Mazzariello, R. Nardone, C. Sansone, and V. Vittorini, "A survey on audio-video based defect detection through deep learning in railway maintenance," *IEEE Access*, vol. 10, pp. 65 376–65 400, 2022.

[6] M. J. Pappaterra, F. Flammini, V. Vittorini, and N. Bešinović, "A systematic review of artificial intelligence public datasets for railway applications," *Infrastructures*, vol. 6, no. 10, p. 136, 2021.

[7] R. Dirnfeld, L. De Donato, F. Flammini, M. S. Azari, and V. Vittorini, "Railway digital twins and artificial intelligence: Challenges and design guidelines," in *Dependable Computing–EDCC 2022 Workshops: SERENE, DREAMS, AI4RAILS, Zaragoza, Spain, September 12, 2022, Proceedings*. Springer, 2022, pp. 102–113.

[8] RAILS, "Deliverable 1.3: Application Areas," 2021. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[9] ——, "Deliverable 3.1 – WP3 Report on case studies and analysis of transferability from other sectors (predictive maintenance and defect detection)," 2021. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[10] L. De Donato, R. Dirnfeld, A. Somma, A. De Benedictis, F. Flammini, S. Marrone, M. Saman Azari, and V. Vittorini, "Towards ai-assisted digital twins for smart railways: preliminary guideline and reference architecture," *Journal of Reliable Intelligent Environments*, pp. 1–15, 2023.

[11] RAILS, "Deliverable 3.2 – WP3 Report on AI approaches and models," 2022. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[12] ——, "Deliverable 3.3 – WP3 Report on experimentation, analysis, and discussion of results," 2023. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[13] L. De Donato, S. Marrone, F. Flammini, C. Sansone, V. Vittorini, R. Nardone, C. Mazzariello, and F. Bernaudin, "Intelligent detection of warning bells at level crossings through deep transfer learning for smarter railway maintenance," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106405, 2023.

[14] ERA, "Report on Railway Safety and Interoperability in the EU 2022," 2022. [Online]. Available: https://www.era.europa.eu/content/railway-safety-and-interoperability-2022-report_en
[15] Network Rail, "NR/L2/XNG/19608 ISSUE 8," 2022. [Online]. Available: https://standards.globalspec.com/std/14461410/nr-l2-xng-19608-issue-8
[16] SAFER-LC, "Deliverable D1.3: Needs and requirements for improving level crossing safety," 2018. [Online]. Available: https://safer-lc.eu/deliverables
[17] G. Bocchetti, F. Flammini, C. Pragliola, and A. Pappalardo, "Dependable integrated surveillance systems for the physical security of metro railways," in *2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, 2009, pp. 1–7.
[18] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold *et al.*, "CNN architectures for large-scale audio classification," in *Int. Conf. on Acoustics, Speech and Signal Processing*. IEEE, 2017, pp. 131–135.
[19] SAFER-LC, "Deliverable D3.5: Report on Smart Detection System (SDS)," 2019. [Online]. Available: https://safer-lc.eu/deliverables
[20] M. A. B. Fayyaz, A. C. Alexoulis-Chrysovergis, M. J. Southgate, and C. Johnson, "A review of the technological developments for interlocking at level crossing," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, p. 0954409720941726, 2020.
[21] R. P. Pratama and S. H. Supangkat, "Smart video surveillance system for level crossing: A systematic literature review," in *2021 International Conference on ICT for Smart Society (ICISS)*. IEEE, 2021, pp. 1–5.
[22] G. L. Giacco, A. D'Ariano, and D. Pacciarelli, "Rolling stock rostering optimization under maintenance constraints," *Journal of Intelligent Transportation Systems*, vol. 18, no. 1, pp. 95–105, 2014.
[23] G. L. Giacco, D. Carillo, A. D'Ariano, D. Pacciarelli, and Á. G. Marín, "Short-term rail rolling stock rostering and maintenance scheduling," *Transportation Research Procedia*, vol. 3, pp. 651–659, 2014.
[24] LINX4RAIL, "System architecture and Conceptual Data Model for railway, common data dictionary and global system modelling specifications," 2019-2022, https://projects.shift2rail.org/s2r_ipx_n.aspx?p=LINX4RAIL.
[25] LINX4RAIL-2, "System Architecture and Conceptual Data Model For Railway, Common Data Dictionary And Global System Modelling Specifications," 2020-2023, https://projects.shift2rail.org/s2r_ipx_n.aspx?p=LINX4RAIL2.
[26] TAURO, "Technologies for the AUtonomous Rail Operation," 2020-2023, https://projects.shift2rail.org/s2r_ipx_n.aspx?p=tauro.
[27] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.
[28] Y. Lindell and B. Pinkas, "Secure multiparty computation for privacy-preserving data mining," *Journal of Privacy and Confidentiality*, vol. 1, no. 1, 2009.
[29] B4CM, "Blockchains for condition monitoring: Blockchains as a Distributed Ledger for

Attribution of RCM Data in Rail," 2018-2022, https://projects.shift2rail.org/s2r_ipx_n.aspx?p=B4CM.

[30] RAILS, "Deliverable D2.3 - WP2 Report on experimentation, analysis, and discussion of results," 2023. [Online]. Available: https://rails-project.eu/downloads/deliverables

[31] 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.

[32] X. Zheng, J. Lu, and D. Kiritsis, "The emergence of cognitive digital twin: vision, challenges and opportunities," *International Journal of Production Research*, vol. 60, no. 24, pp. 7610–7632, 2022.