Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.067143

ARTICLE

Expert System Based on Ontology and Interpretable Machine Learning to Assist in the Discovery of Railway Accident Scenarios

Habib Hadj-Mabrouk*

Vice-Presidency Research, Gustave Eiffel University, Marne-la-Vallée, F-77454, France

*Corresponding Author: Habib Hadj-Mabrouk. Email: habib.hadj-mabrouk@univ-eiffel.fr

Received: 26 April 2025; Accepted: 10 June 2025; Published: 30 July 2025

ABSTRACT: A literature review on AI applications in the field of railway safety shows that the implemented approaches mainly concern the operational, maintenance, and feedback phases following railway incidents or accidents. These approaches exploit railway safety data once the transport system has received authorization for commissioning. However, railway standards and regulations require the development of a safety management system (SMS) from the specification and design phases of the railway system. This article proposes a new AI approach for analyzing and assessing safety from the specification and design phases of the railway system with a view to improving the development of the SMS. Unlike some learning methods, the proposed approach, which is dedicated in particular to safety assessment bodies, is based on semi-supervised learning carried out in close collaboration with safety experts who contributed to the development of a database of potential accident scenarios (learning example database) relating to the risk of rail collision. The proposed decision support is based on the use of an expert system whose knowledge base is automatically generated by inductive learning in the form of an association rule (rule base) and whose main objective is to suggest to the safety expert possible hazards not considered during the development of the SMS to complete the initial hazard register.

KEYWORDS: Artificial intelligence; ontology; semi-supervised learning; expert system; association rules; railways; safety; hazard; accident scenarios; classification; assessment

1 Introduction

Machine learning (ML) is an important branch of AI research. Within ML, a distinction is made between supervised learning, semi-supervised learning, and unsupervised learning (Fig. 1). There are several machine learning approaches and algorithms that rely largely on regression, discrimination (or classification), and clustering techniques.

In supervised learning, we find the following main approaches:

- Discrimination or Classification: supervised classification, K-nearest neighbors (KNN), Naive Bayes classifier, Random Forests, Decision trees;
- Regression: simple linear regression, Multiple linear regression, Logistic regression, Logarithmic regression, Support Vector Regression (SVR);
- PAC-Bayesian theory (Naive Bayesian classifier, Bayesian Network (BN), Bayesian Belief Network (BBN);
- Backpropagation (Gradient backpropagation).

On the other hand, unsupervised learning includes several methods and algorithms:

- Grouping: Partitioning method: (K-means), Dynamic swarms (disjoint groups), k-median, Hierarchical Clustering, Support Vector Machines (SVM), Probabilistic clustering;
- Artificial Neural Network (ANN): Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), Deep Neural Network (DNN), etc.;
- Dimension reduction: discriminant Factor Analysis (DFA), Principal Component Analysis (PCA), etc.;
- Association: association rules;
- Optimization: genetic algorithm, etc.;
- Feature extraction (Data Mining): descriptive Tasks, Predictive Tasks.

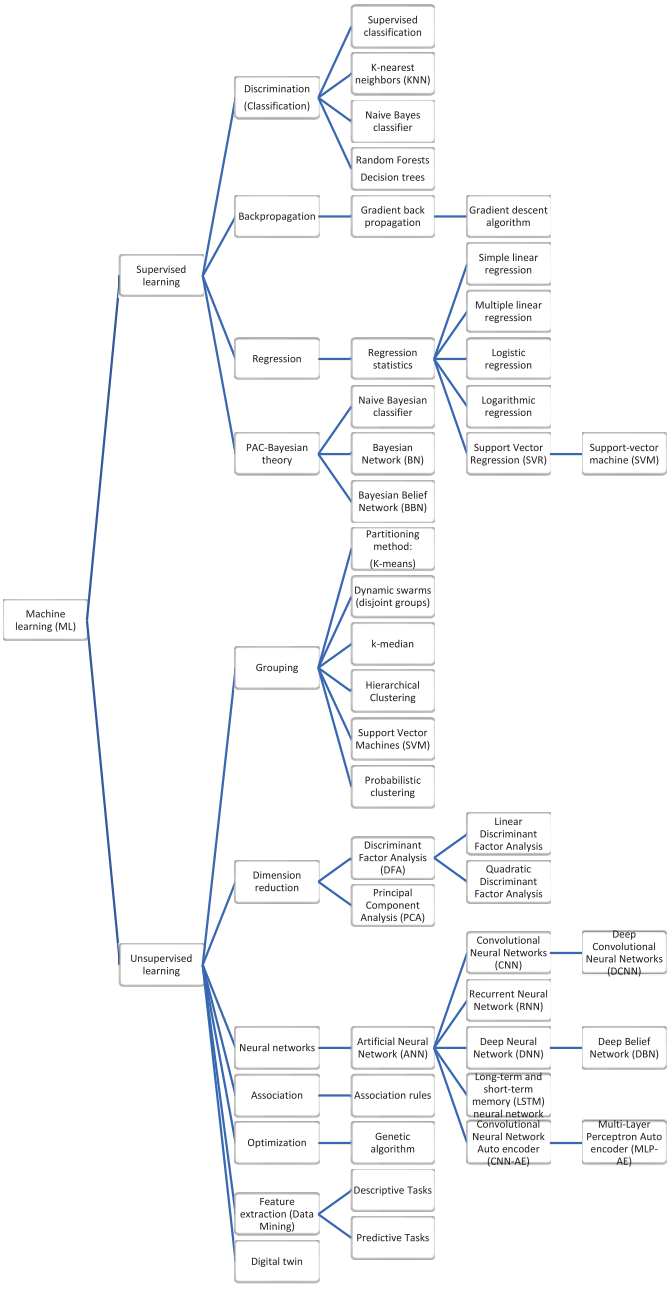

Figure 1: Proposal for a classification of machine learning methods and algorithms

Several review studies have been published in the literature on the application of AI in the transportation sector: Intelligent transportation [1,2]; Public transport [3]; Energy, water, transport, and telecommunications infrastructure sectors [4]; Different modes of transport land (road, rail), maritime and air [5]. The challenges, arguments, and interests of "Big Data" for risk management in rail transport are presented in [6–8]. After giving an overview of "Big Data" technologies in the rail transport sector, Ghofrani et al. [9] presented an interesting survey from 2003 to 2017 on the potential applications of big data analysis in railway systems rail transport. According to Laiton-Bonadiez et al. [10], the areas of application of Industry 4.0 technologies in rail transport relate to the following three branches: (1) Surveillance, (2) Decision and planning, (3) Communication and safety. Dong et al. [11] presented a critical review of recent textual research and their applications in railways which notably concern the analysis of accidents and incidents, sentiment analysis in particular passenger complaints and speech synthesis, detection of technical specifications, fault diagnosis, servicing/maintenance and inspection, accident risk assessment, extraction of safety information, identification of accident causes and finally the identification of maintenance events. An interesting taxonomy of artificial intelligence methods and algorithms as well as their applications in rail transport has been proposed in Bešinović et al. [12]. The applications identified by the authors relate to autonomous train driving and control, maintenance, and inspection in particular fault diagnosis, infrastructure condition monitoring, fault detection and prediction, and mobility of passengers which brings together the prevention and prediction of passenger flows and passenger satisfaction, traffic planning and management, finally safety and safety of transport in particular the analysis of incidents, the safety of stations, the detection of defects, rail disruptions and research into the causes of accidents. Tang et al. [13] also proposed a literature review on the applications of AI in railway transport systems and the applications studied concern autonomous driving and control, revenue management, inspection, passenger mobility, traffic planning and management, transport policy, and safety and security. An interesting study on recent applications of machine learning in railway maintenance was proposed by Chenariyan-Nakhaee et al. [14]. Finally, several studies are increasingly interested in AI applications which are essentially based on ontology and knowledge graphs:

- Ontologies for transportation research [15],
- Ontology-based systems engineering [16],
- Knowledge graphs as tools for explainable machine learning [17],
- Application of ontology and knowledge graphs in rail transport [18],
- Applications of "classical" AI, ontology, and knowledge graphs to European rail transport safety [19].

After this introduction to the methods, algorithms and types of machine learning (ML), the following paragraph presents a quick review of the literature on the applications of AI and ML in rail transport.

2 Proposal for Classification of AI Applications Relating to Rail Transport

Drawing inspiration from the two European directives relating to the development (Directive (EU) 2012/34 [20]) and interoperability (Directive (EU) 2016/797 [21]) of the European railway system, we propose to complete and refine previous studies on applications of AI techniques and in particular machine learning (ML) to rail transport by focusing in particular on the contributions and limits of AI to railway safety. The AI applications studied are classified into two approaches (Fig. 2) [19]:

1. AI approaches related to the "Structural Elements" of the railway system (Table 1):
   - AI approaches related to railway "infrastructure": switch system, rail, ballast, rail track geometry, level crossings, tunnels, etc.
   - AI approaches related to railway "rolling stock": axles, wheels, pantographs, locomotives, wagons, etc.

2. AI approaches related to the "Functional Elements" of the railway system (Table 1):
   - Operation/traffic management,
   - Maintenance,
   - Telematics applications,
   - Investigation into railway accidents and incidents.

Figure 2: Decomposition of the railway system (inspired by Directives (EU) 2016/797 and (EU) 2012/34)

Table 1: Review of AI applications studied and their distribution by type of railway equipment

Railway system	Subsystems	Equipment/ Constituents	Goals	AI methods and algorithm
Structural elements	Infrastructure	Switch system	Diagnosis of faults in the switch system	- Artificial neural networks (ANN), - Case-based reasoning (CBR) - Convolutional Neural Network (CNN), - Deep Convolutional Neural Networks (DCNN) - Deep learning - Expert system, - Finite element method (FEM), - Long-term and short-term memory (LSTM) neural network, - Natural language processing (NLP), - Ontology, - Petri net, - Recurrent Neural Networks, - SHAP: SHapley Additive exPlanations, - Support Vector Regression (SVR), - Vision-based AI.
		Rail	Detection and inspection of rail surface defects	
		Ballast	Railway ballast maintenance	
		Rail track geometry	Inspection and prediction of track failures	
		Level Crossings (LC)	Hazard analysis, monitoring, and violation detection of level crossings (LC)	
		Tunnel	Railway tunnel construction	
		Infrastructure	Ontology-based railway infrastructure topology	
	Rolling stock	Axle	Maintenance of a railway axle	- Case-based reasoning (CBR), - Classification-based learning, - Convolutional Neural Network Autoencoder (CNN-AE), - Decision trees, - Deep Belief Network (DBN), - Digital twin based on learning, - Expert system, - Fuzzy logic, - Multi-layer perceptron auto-encoder (MLP-AE), - Non-negative matrix factorization (NMF), - One-class support vector machine (OC-SVM), - Principal component analysis (PCA), - Short-term Fourier transform (STFT), - Statistical analysis methods.
		Wheels	Diagnosis and detection of defects in wheels	
		Pantograph	Identify hazards related to train traction equipment (faulty sensors, faulty pantographs)	
		Locomotive	Railway locomotive fault diagnosis	
		Wagons	Maintenance and prediction of the risk of derailment of wagons	
Functional elements	Operation: (Route Compatibility)	Infrastructure & Rolling stock	Checking the technical compatibility between the vehicle and the route	- Case-Based Reasoning (CBR), - Expert system, - Fuzzy Petri Net, - K-nearest neighbors, - Knowledge Graphs, - Naive Bayes, - Ontology, - Random forests, - Support vector machines, - Train simulator.
	Operation: (Traffic Planning and Control)	Rolling stock (Locomotive)	Rail traffic planning and control	
	Operation: (Signaling)	ERTMS/ETCS	Ontology for modeling ERTMS/ETCS	
	Operation: (Train driving)	Train	Driving assistance (travel time and fuel consumption)	
	Operation: (Train driving)	Train	Detecting train driver fatigue	
	Maintenance	Wagon	Wagon maintenance	- Decision Trees, - Classification Rules, - Expert system, - Linear Classifier, - Markov Chain Monte Carlo, - Ontology, - Principal Component Analysis (PCA), - Random forests, - Support Vector Machine (SVM), - Support Vector Machine (SVM).
		Metro	Maintenance of automated metro lines	
		Rolling stock	Maintenance and monitoring of rolling stock and track condition (Ontology-Based)	

(Continued)

Table 1 (continued)

Railway system	Subsystems	Equipment/ Constituents	Goals	AI methods and algorithm
	Telematics applications (TA)	TA Traveler TA Freight	- Train delay prevention, - Customizing user-interfaces, - Personalized route finding, - Harmonizing information systems, - Evaluating key performance indicators (KPIs) *(Ontology-Based)*	- Accelerated failure time, - Adaptive boosting, - Case-based reasoning (CBR), - Cosine Similarity, - Extreme gradient boosting (XGBoost) tree, - Gradient boosting decision tree (GBDT), - Graph Theory, - K-nearest neighbor, - ML Regression Models, - Ontology, - Ordinary least squares, - Quantile Regression (QR), - Random forest, - Short-Text Topic Modeling, - Support vector regression (SVR).
	Investigation into railway accidents and incidents	Infrastructure & Rolling stock	- Analysis of railway incident and accident data, - Exploration of hazard causes, - Identification of actors and safety risk factors, - Modeling correlations between accident-related hazards, - Discovery of accident characteristics and particularities, - Prediction of hazards and accident risks, - Prediction of the annual number of injuries, etc.	- Artificial neural networks (ANN), - Association rules (Aprioriand Clementine software), - Case-based reasoning (CBR), - "K-means" classifier (ROST software), - Convolutional Neural Networks (CNN), - Decision tree, - Deep Neural Networks (DNN), - Genetic algorithm (GA), - Graph theory, - Knowledge graphs, - Latent Dirichlet Allocation (LDA), - Latent Semantic Analysis (LSA), - Naïve Bays Classifier, - Natural Language Processing (NLP), - Ontology, - Production rule learning, - Random Forests, - Recurrent Neural Networks (RNN), - Rule-based reasoning (RBR).

AI applications relating to railway *"Infrastructure"* equipment are numerous: Diagnosis of faults in the switch system, Detection and Inspection of Rail Surface Defects, Railway ballast maintenance, Inspection and prediction of track failures, Hazard Analysis, Monitoring and Violation Detection of Level Crossings, Railway tunnel construction, Ontology-based railway infrastructure topology.

Regarding railway *"Rolling stock"*, we can cite the following AI applications: Maintenance of a railway axle, Diagnosis and detection of defects in wheels, identify hazards related to train traction equipment (faulty sensors, faulty pantographs), Railway Locomotive Fault Diagnosis, Maintenance and Prediction of the Risk of Derailment of Wagons.

AI work dedicated to railway *"Operations"* primarily concerns: Checking the technical compatibility between the vehicle and the route, Rail Traffic Planning and Control, Ontology for modeling ERTMS/ETCS, driving assistance (travel time and fuel consumption), and Detecting train driver fatigue.

For rail *"Maintenance"* and upkeep operations, we can cite: Wagon maintenance, Maintenance of automated metro lines, Maintenance and monitoring of rolling stock and track condition (Ontology-Based).

Studies related to *"Telematics Applications"* (Traveler and Freight) are primarily based on the development of ontologies: Train Delay Prevention, Customizing User–Interfaces, Personalized Route Finding, Harmonizing Information Systems, and Evaluating Key Performance Indicators (KPIs).

Finally, AI applications dedicated to railway *"safety"* rely on the analysis of accident and incident data from railway investigation reports. The objectives of these studies are numerous: Exploration of hazard causes, Identification of actors and safety risk factors, Modeling correlations between accident-related hazards, Discovery of accident characteristics and particularities, Prediction of hazards and accident risks, Prediction of the annual number of injuries, etc.

For each subsystem (infrastructure, rolling stock) and each function (Operation, Maintenance, Telematics applications, Investigation into accidents and incidents), Table 1 presents the AI methods and algorithms used. For further information on the above-mentioned works, readers can consult Hadj-Mabrouk's [18,19] recent work on the applications of AI, ontologies, and knowledge graphs in the field of rail transport safety.

3 Limitations of Related Work, Problem Positioning, and Objectives

This brief literature review shows that AI and ML applications are increasingly numerous in the rail transport sector and aim to improve, in particular, the operation, maintenance, and safety tasks of railway subsystems (infrastructure, rolling stock, telematics applications, etc.). This study focuses solely on railway safety, which remains a crucial issue for experts in the field as well as for national safety authorities (NSAs). Approaches closely related to railway safety (Table 1) exploit railway feedback data from accident and incident investigation reports and often use text mining and machine learning techniques to analyze accident data, identify causes, predict accident risks, reveal the presence of informative concepts, identify actors and risk factors, discover relationships between accident factors, distinguish accident characteristics and particularities, classify accident causes, identify significant trends in accident data, or predict the annual number of injuries.

Despite the undeniable value of these approaches, their implementation assumes that the railway system has already received authorization for commissioning (rolling stock) or operation (infrastructure). However, railway safety begins during the specification phase and is omnipresent throughout the system's life cycle. For example, during the specification phase, a preliminary hazard analysis (PHA) must be performed; during the design phase, a functional safety analysis (FSA) must be established; during the production of hardware, a failure mode and effect analysis (FMEA) must be performed, supplemented by a root cause analysis (RCA), etc. Safety must therefore be considered not only during the specification, design, and production phases, but must also be validated by the safety analysis and assessment expert (or organization), controlled by rail operators as part of their safety management systems (SMS), monitored by national safety authorities (NSAs), and finally improved based on investigation reports prepared by the investigation organization.

We propose a new approach to railway safety analysis and assessment that is upstream of the AI approaches presented previously. This approach, based on the concept of "accident scenario" can help safety experts from the specification and design phases of the railway system. It should therefore contribute to the improvement of the safety management system (SMS) by improving the completeness and consistency of the step of identifying potential hazards and accidents that could jeopardize safety and for which prevention and/or protection measures (or barriers) are necessary during the design of the hardware and software equipment of the railway system.

It is also important to emphasize that the basic concepts involved in railway safety, which are used by several machine learning algorithms, suffer from a lack of precision and clarity. This includes the apparent confusion between the terms "hazard", "risk", "accident", "incident", "potential accident", "dangerous event", "dangerous situation", "dangerous element", "risk analysis", "hazard analysis", "risk assessment", "risk management", "risk reduction", "safety", etc. Indeed, the vocabulary used during the development of an AI system, particularly to define classes, descriptors, properties, etc., cannot in any way guarantee the semantics, interoperability, and reusability of safety knowledge. Consequently, the validity of certain approaches arises,

as they are approaches intended for risk management in critical systems such as rail transport. To provide an element of response to this crucial problem, we propose the development of a railway safety ontology allowing the harmonization of basic concepts linked to railway risk management.

In fact, two major problems exist in AI applications: the quality of the data required for learning and the explainability (or interpretability) of the data learned by the learning algorithm. These two major obstacles to machine learning are also well detailed in Bešinović et al. [12], Attoh-Okine [6], Cooray [22], Niu et al. [23], Tamascelli et al. [24], Richardson [25], Xu et al. [26], Rohlfing et al. [27], Longo et al. [28], Thekdi and Aven [29]. To address data and explainability issues, several learning constraints are necessary:

- Domain knowledge is required to process noisy data. The goal is to ensure robustness against "noise" and mitigate the disruptive effects of poorly characterized training examples. Machine learning is particularly sensitive to the relevance of available data. Ensuring this quality relies in particular on the acquisition and use of complementary knowledge to reduce diffuse noise in the examples.
- Formalize data and domain knowledge in the form of ontology to explicitly describe and represent domain knowledge by defining classes, their relationships, and their properties.
- Ensure the representativeness and quality of the training example database. If a learning algorithm allows rules or concepts to be generated from experimental examples, the fact remains that the quality of the knowledge learned depends largely on the quality of the example base (correct, complete, coherent, rich information, sufficient number of examples, and descriptors).
- Perform semi-supervised learning: Semi-supervised learning is halfway between supervised learning, which uses known labeled data, and unsupervised learning, which uses unlabeled data. The use of labeled data, in combination with unlabeled data, in our opinion, allows us to improve the quality of learning and in particular the problem of interpretability and explainability which currently constitutes the "bottleneck" of learning techniques applied in high-risk systems such as railway safety. Indeed, we rarely manage to extract all the data and knowledge from domain experts at the first attempt, but when we present the knowledge learned by the system to the expert, the latter becomes aware of their interest, identifies contradictions, and relevant rules, completes the training examples, possibly corrects the description language of the examples, adjusts the learning parameters, etc. Thus, involving the domain expert in the learning "loop", will certainly help him to better verbalize his know-how and consequently, we contribute not only to the enrichment of domain knowledge but also to the interpretability of the learning models developed, which today constitutes the main objective of explanatory AI.
- Consider the incrementality and stability of the knowledge learned by the system to facilitate its updates.
- Carry out symbolic/numerical learning: The numerical approach focuses on optimizing a global criterion such as entropy or the distance between examples in data analysis. The major drawback of numerical methods lies not only in the impoverishment of the initial data when translating them into numbers, but also in the fact that the semantics of numerical operations sometimes differ from that of the initial symbolic data. In addition, the knowledge generated is often incomprehensible to humans. On the other hand, the objective of symbolic methods is to use knowledge to produce new knowledge, not presented trivially in the initial description of the problem. This new knowledge constitutes an explanation at a higher level than that of observation in traditional data analysis. In the symbolic approach, we no longer ask what is most effective, but what is most meaningful. Indeed, the symbolic approach is capable of explanations because it operates on data in the form of conceptual graphs, semantic networks, ontologies, etc. The use of a digital component is fundamental, even indispensable, to optimize the learning process and deal with complex real-life problems where domain knowledge is often incomplete, non-exhaustive, or noisy. These remarks attest to the interest of the symbolic-digital approach for the creation of effective learning systems integrating the explanatory component.

– Achieve human-centered learning by ensuring interactivity between the safety expert and the learning system: the system must explain its reasoning by producing knowledge that is understandable and interpretable by the expert whose role is to control, complete, and validate this knowledge. The expert can indeed contribute to evaluating and validating the knowledge learned. This approach implies intense interactivity and cooperation between the expert and the learning system. This transparency of the approach requires particular care when creating Human/Machine interfaces.

All of these learning constraints were considered for the development of an expert system based on ontology and interpretable learning to help in the discovery of railway accident scenarios.

4 Methodological Approach for the Acquisition, Modeling, Classification, and Evaluation of Railway Accident Scenarios

During the design and development of a rail transport system, all stakeholders involved use one or more safety methods to identify hazardous elements, hazardous situations (or hazards), their causes, potential accidents, and the severity of the resulting consequences. The main objective is to justify and ensure that the transport system's design architecture is safe and poses no particular risks to users and the environment. As part of this process, safety analysis and assessment experts are required to imagine new potential accident scenarios to enhance the comprehensiveness of safety studies. One of the challenges then lies in identifying abnormal scenarios that could lead to a specific potential accident. This is the fundamental point that motivated this work, the objective of which is to develop a decision-making tool to assist safety experts in their crucial task of analyzing and assessing railway safety. Knowledge of railway accidents and incidents is essentially derived from the contribution of lessons learned and experiences gained resulting from feedback from railway systems. It is therefore appropriate to exploit this historical knowledge to understand and explain the causes and circumstances of accident risks and consequently avoid at least the recurrence of similar accidents by using AI techniques and in particular ontology, machine learning, and expert systems. The objective is to anticipate and prevent the recurrence of risks of similar accidents or incidents and possibly to discover and identify new potential accident scenarios likely to compromise railway safety. Fig. 3 presents the approach adopted for the analysis and evaluation of railway safety based on modeling, capitalization, classification, evaluation, and discovery of potential accident scenarios. The following paragraphs successively present the following steps: Conceptual model of risk management, Ontology of accident scenarios, Knowledge acquisition, Classification of new scenarios, and Deduction of hazardous events.

4.1 Conceptual Model of Railway Risk Management

To eliminate inconsistencies, confusion, and terminological conflicts at the various levels of railway safety analysis and assessment, a conceptual model for railway risk management should be developed to harmonize certain vocabulary used in railway safety analysis and assessment, particularly railway hazard analysis. The goal is to provide common terminology, a uniform vocabulary, precise definitions, terms, concepts, and relationships to support railway hazard studies, particularly during the development of a safety management system (SMS). The established vocabulary is largely based on experience acquired, particularly during expertise and certification of railway systems, as well as on the analysis of a set of railway standards and regulations relating to railway safety and risk management. Several definitions were thus established, including the concepts of hazardous elements, hazard, potential accidents, accidents, near misses (or incidents), damage, risk level, and preventive or protective measures. Fig. 4 presents the main results of this study in the form of a conceptual model semantically articulating all the descriptive parameters involved in railway risk analysis and assessment.

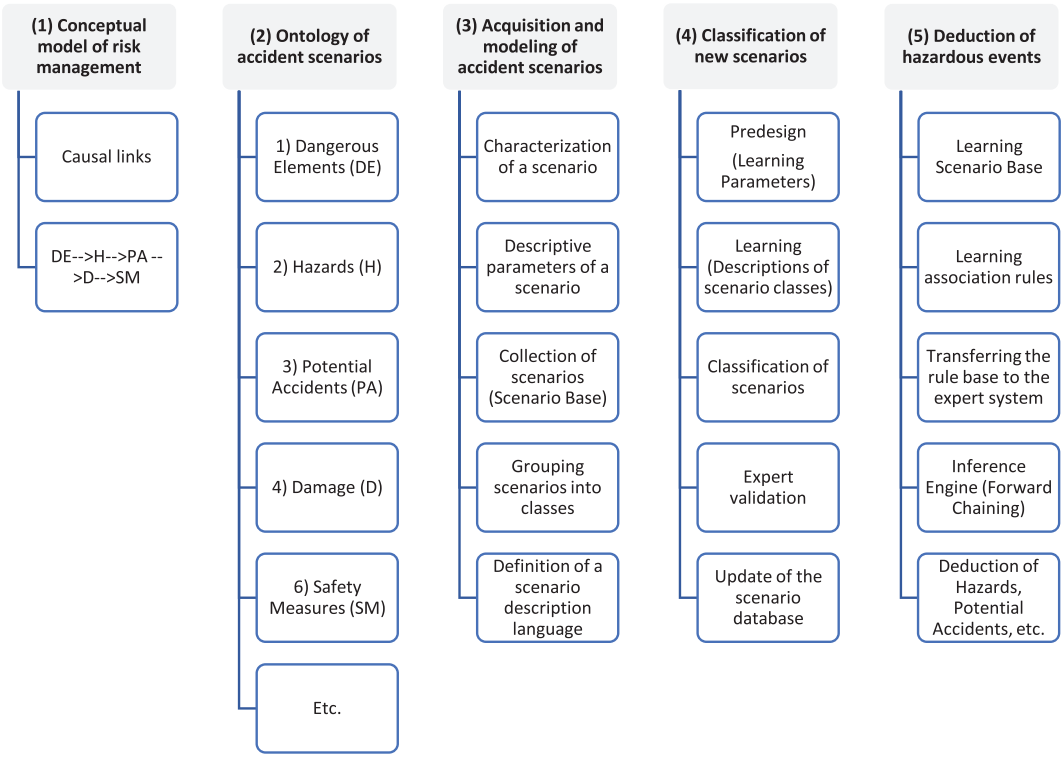

Figure 3: Approach to railway safety analysis and evaluation based on accident scenarios

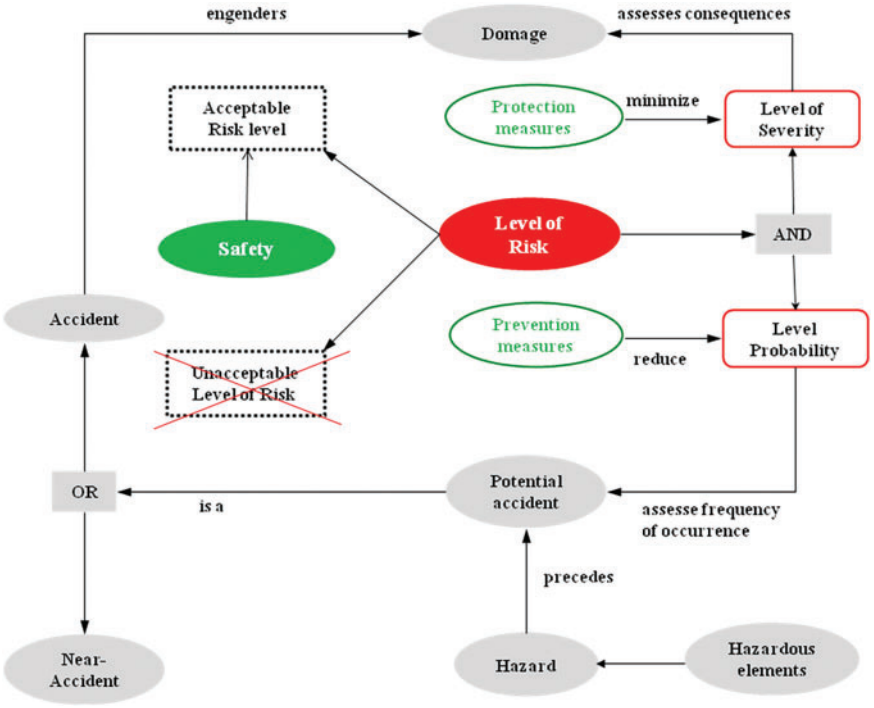

Figure 4: Conceptual model of railway risk management

This conceptual model was subsequently implemented by the software "web protégé" [30]. For example, Fig. 5 intuitively shows the semantic links between "Dangerous Element", "Danger", "Potential Accident", "Accident", "Near Miss (Incident)", "Damage", etc.

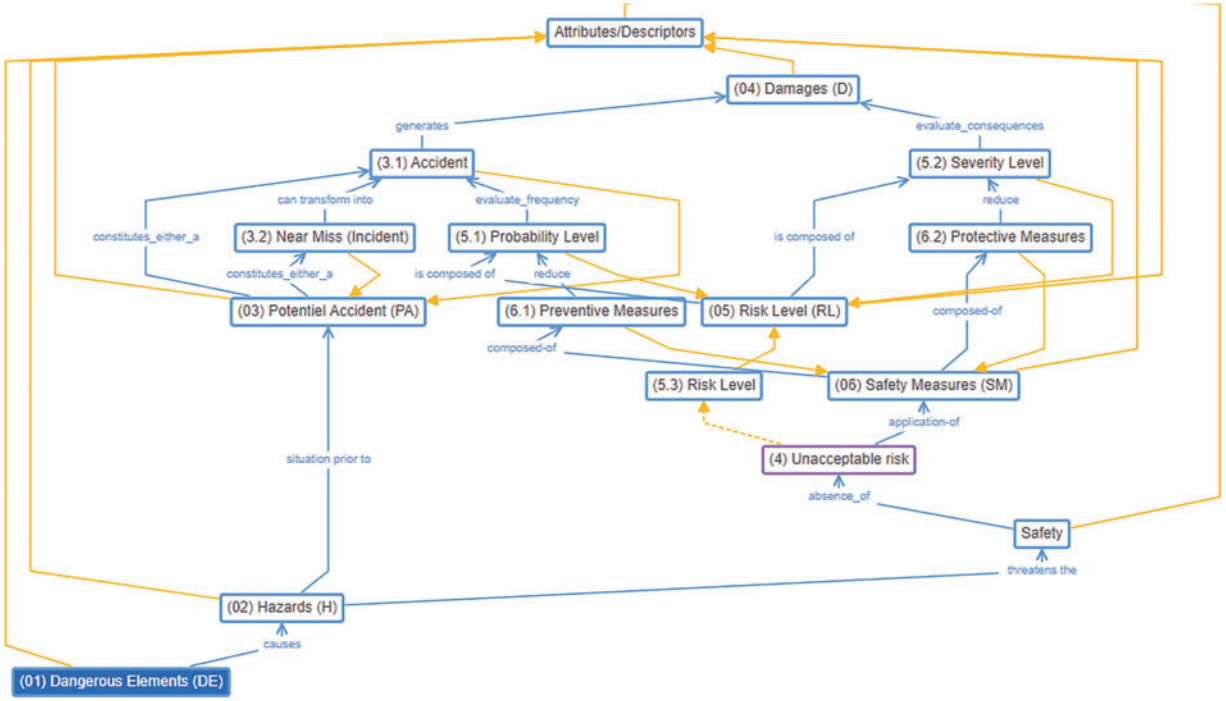

Figure 5: Relationship between "Dangerous Element", "Danger", "Potential Accident" and "Damage"

"Safety" remains the key component of any high-risk industrial system such as rail transport. It is often defined as "the absence of unacceptable risk" or "the state of a hazardous situation presenting a minimum acceptable risk and therefore being free from all unacceptable hazards and risks".

Regardless of the scope of application and regardless of the technical and organizational prevention and/or protection measures in critical systems such as rail transport, absolute safety does not exist. Indeed, even if prevention, control, and monitoring measures, operating and maintenance procedures, barriers, and system protection and/or prevention equipment are essential and mandatory, they cannot guarantee absolute safety. There is always a hazard, a dangerous event, or an uncertain hazardous situation that is more or less predictable, more or less possible, or even more or less probable to which the rail system is exposed. Moreover, this unfortunately still explains the numerous accidents and incidents in all transportation sectors, including the rail sector. This has led to the implementation of a feedback process not only to investigate the causes of these rail accidents and incidents but above all to avoid and prevent the recurrence of such dangerous situations, which often result in damage, sometimes critical, in terms of deaths and serious injuries. Since absolute safety does not exist, the presence of a potential risk must be acknowledged, provided that it is acceptable or tolerable with respect to humans, the system, and the environment.

It is therefore necessary to define an acceptable "risk level" that measures both the occurrence (or frequency or probability) of an adverse event, termed a "potential accident", and the "damage" (or effects or consequences) caused by this accident. The occurrence of this adverse event is generally measured by its "probability of occurrence" over a given period. The consequences or effects can be human, economic,

or environmental in nature. Risk is thus expressed, for example, in monetary units per unit of time, in the number of deaths per unit of time, or in the probability of death per unit of time. In this context, we refer to the definition of quantified safety objectives (or probabilistic safety objectives). Unfortunately, this acceptable conventional safety level is not defined objectively because it depends on the subjective assessment of the system's designers. Thus, the dangerous scenario that no one has considered always exists (for example, the forgotten tool that connects two distant points or the rat that gnaws through insulation, etc.). Designers and safety managers must therefore continue to imagine potential accident scenarios and stimulate the search for solutions before a new transportation system is put into service. Thus, based on the definition of the system's operation, the qualitative analysis consists of studying the system's safety overall by carrying out a risk analysis whose objective is to identify dangerous situations, potential accidents, dangerous elements or equipment as well as the seriousness of the resulting consequences.

4.2 Ontology for the Harmonization of Concepts Involved in Accident Scenarios

An ontology is «a formal and explicit specification of a shared conceptualization that is characterized by high semantic expressiveness required for increased complexity» [31]. It is a form of graphical, formal, precise, and explicit knowledge representation that considers the semantics of the application domain, identifying inconsistencies in the data, and establishing a common vocabulary for better information sharing. These characteristics give ontologies a useful role in knowledge engineering to formalize, structure, represent, capitalize, and reuse the knowledge of a domain with a great power of explainability and interpretability. A clear, structured, and explicit representation of knowledge allows, among other things, to mitigate the problem of bias involved in AI systems (black boxes) particularly deep learning by providing a clear explanation during the decision-making process.

In our context, the "Protégé" tool developed by Stanford University [32] was used to implement the ontology of railway accident scenarios. This ontology editor is free and open source and integrates Semantic Web standards, including the OWL language. In addition to its graphical interfaces, such as "OntoGraf", for visualizing the constructed ontologies (classes and class hierarchies), "Protégé" implements reasoners (inference engines). Several reasoners allow reasoning on the ontologies described in description logic. With the "Protégé" tool, it is possible not only to create and edit ontology in OWL format but also to use an inference engine like "Pellet". Fig. 6 shows an overview of the railway accident scenario ontology developed using the "Protégé" tool. This ontology is subsequently used during the classification, learning, evaluation, and accident scenario generation stages.

4.3 Acquisition and Modeling of Accident Scenarios

4.3.1 Characterization of an Accident Scenario

An accident scenario is an appropriate and orderly sequence of unpredictable events (or a combination of unexpected circumstances) that can lead to an undesirable, even dangerous, situation, and potentially cause an accident such as a train collision or derailment. Each risk of an accident or incident likely to endanger passenger safety or impair the system's ability to perform required safety functions is generally translated by safety experts into an accident scenario. The development of an accident scenario draws in particular on the experience and know-how of safety analysis experts, as well as on historical data, railway investigations, and feedback from rail transport systems already certified and in operation.

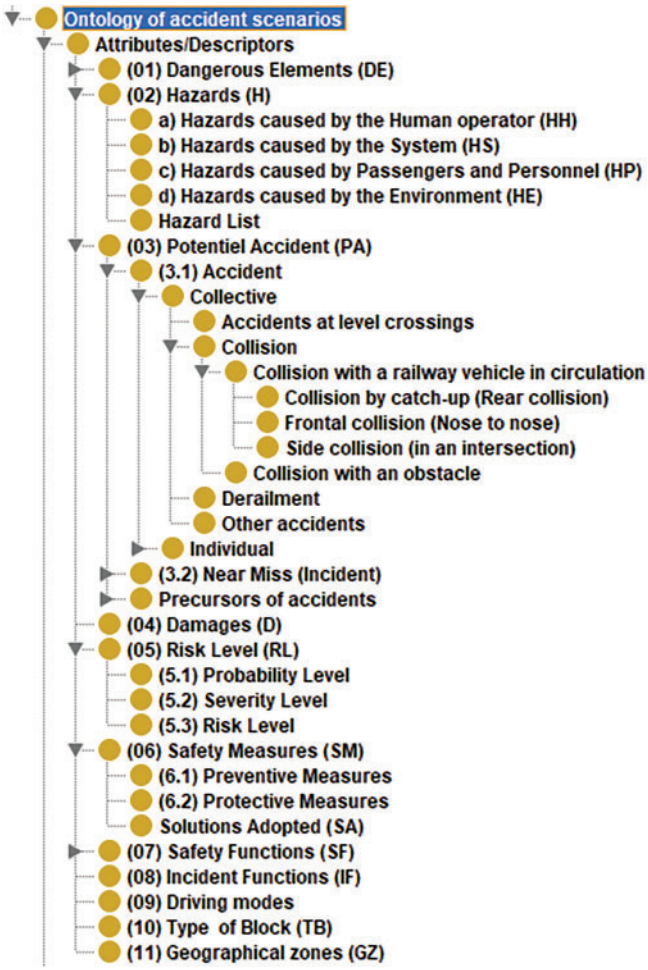

Figure 6: Ontology of railway accident scenarios

Example accident scenario: wrongly storing incompatible routes on two redundant autopilots (APs)

Scenario label: S01: redundancy switching between (Automatic Pilot)

To improve the availability of automated transport systems, it is common practice to duplicate certain equipment, one being active, the other passive, and capable of replacing the first in the event of a failure. The scenario presented here concerns the erroneous storage of incompatible routes on two redundant autopilots (APs) in a terminal:

- A T1 train has turned back and is at the departure station under the control of the active AP (AP A).
- A T2 train arrives following route I1.
- Following a failure, the redundant AP (AP B) has not deleted route I2, even though it has been deleted on the active AP (AP A).
- Consequence: A discrepancy alarm between the two autopilots (APs).
- Following this alarm, the central control center (CCP) could switch the active AP (AP A) to the passive AP (AP B).
- The passive PA (PA B) would allow the T2 train to reverse.
- Potential accident: Collision between trains T1 and T2.

Fig. 7 illustrates the ontological representation of the concepts and events involved in this example scenario.

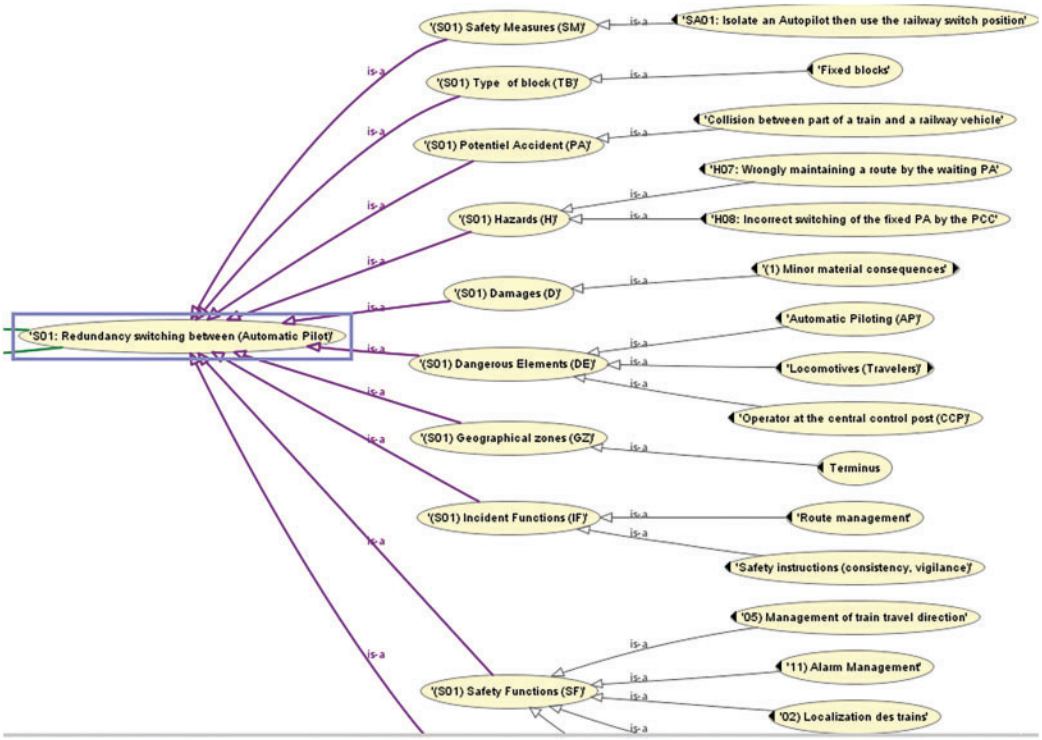

Figure 7: Descriptors involved in the example scenario: "Redundancy switching between (Automatic Pilot)"

4.3.2 Descriptive Parameters of an Accident Scenario

Each accident or incident scenario was formalized and characterized by eleven descriptors such as "Dangerous Element", "Hazards", "Potential Accident", "Safety Function", "Incident Function", etc. (Fig. 8). The scenarios collected so far in the historical scenario database concern the "railway collision" problem and were constructed from several safety files of French rail transport systems: VAL, POMA 2000, MAGGALY, and TVM430 (Nord TGV), and the expertise of safety experts.

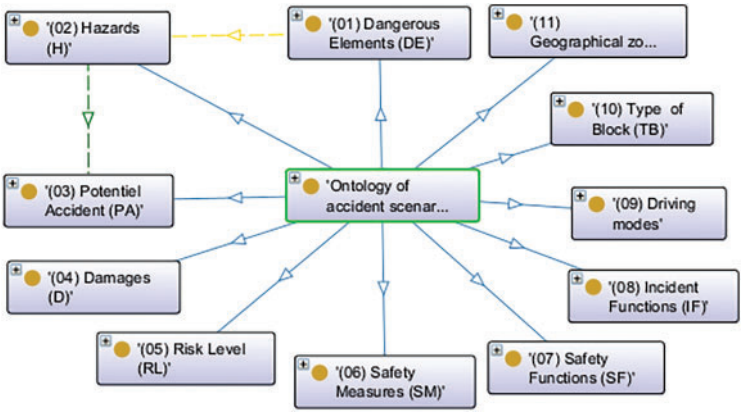

Figure 8: Descriptive parameters of an accident scenario

Figs. 9 to 16 give several examples of values for each descriptor (Attribute) involved in an accident scenario:

- Example of Dangerous Elements (DE): Fig. 9
- Example of Hazards (H): Fig. 10
- Examples of Potential Accidents (PA): Fig. 11
- Example of Damage (D): Fig. 12
- Example of Risk Level (RL): Fig. 13
- Example of Safety Measures (SM): Fig. 14
- Example of Safety Functions (SF) and Incident Functions (IF): Fig. 15
- Example of Driving Modes (DM), Block Types (BT), and Geographic Zones (GZ): Fig. 16

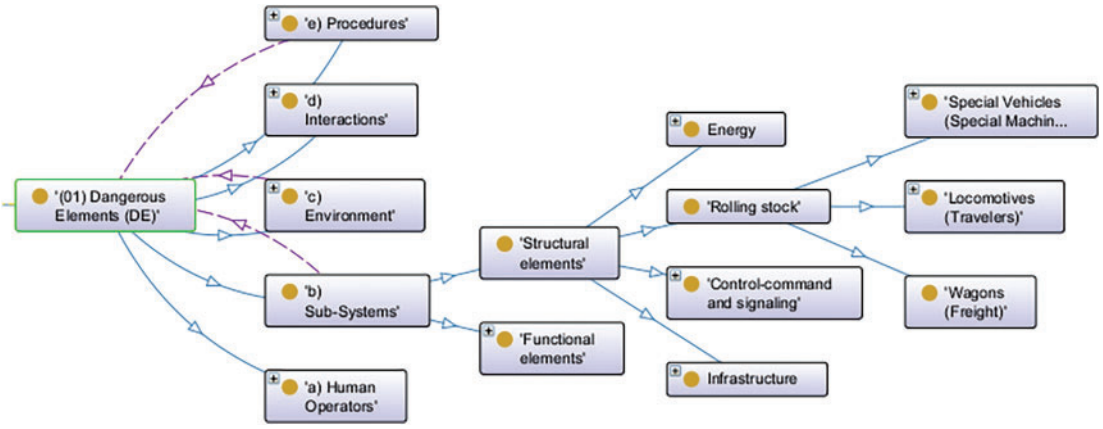

Figure 9: Example of dangerous elements (DE)

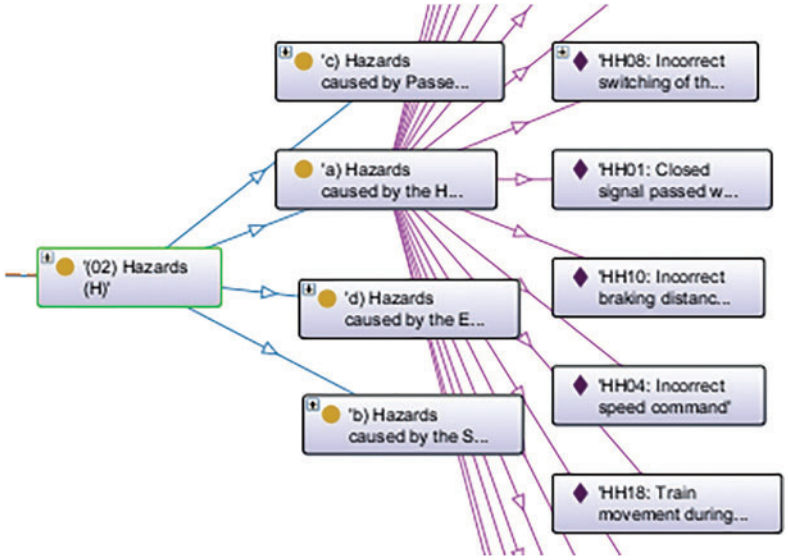

Figure 10: Example of hazards (H)

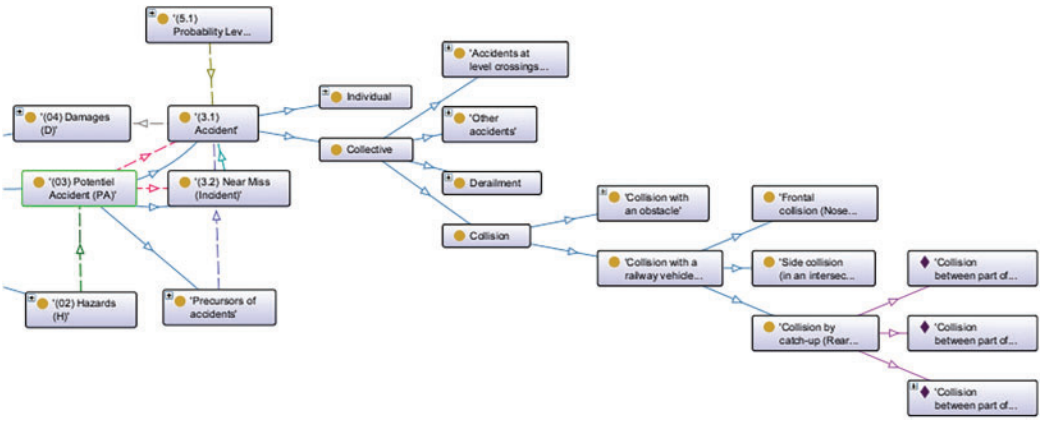

Figure 11: Examples of potential accidents (PA)

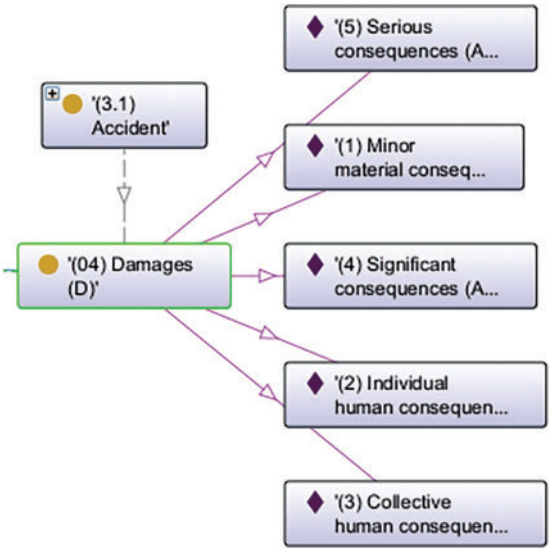

Figure 12: Example of damage (D)

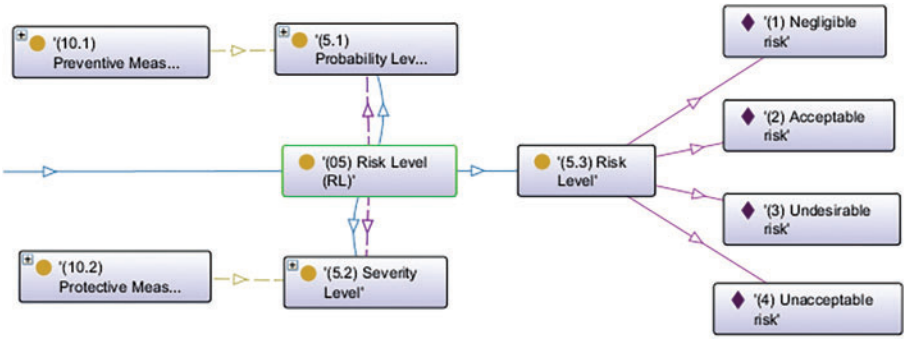

Figure 13: Example of risk level (RL)

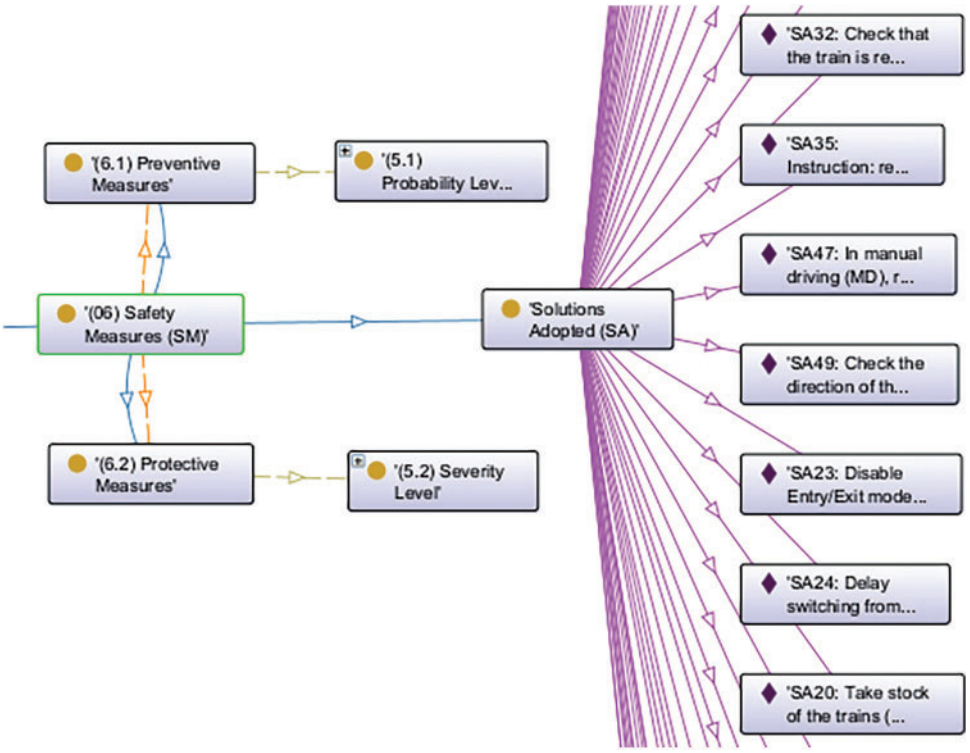

Figure 14: Example of safety measures (SM)

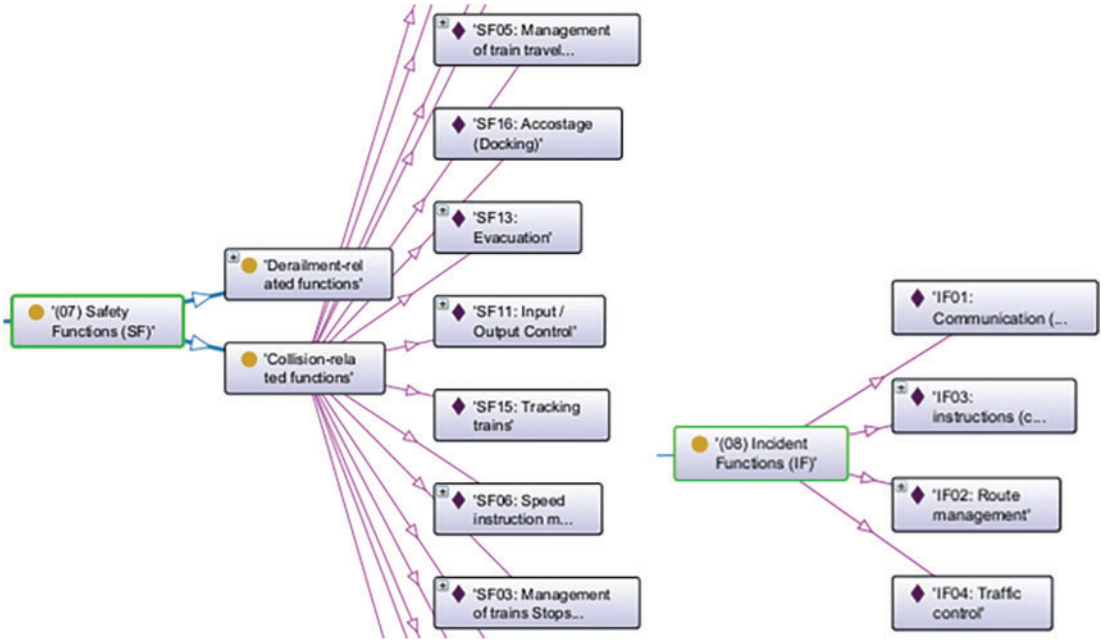

Figure 15: Example of safety functions (SF) and incident functions (IF)

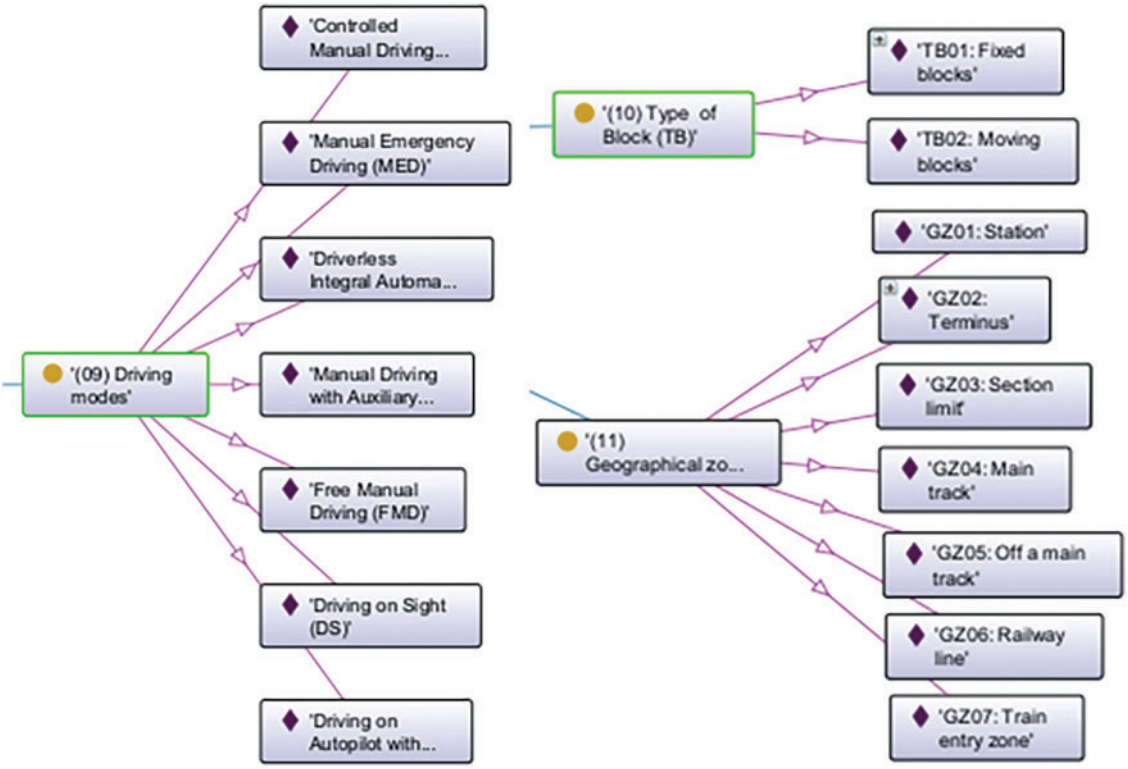

Figure 16: Example of driving modes (DM), Block types (BT), and Geographic zones (GZ)

These descriptive parameters of a scenario (HE, HA, PA, D, RL, SM, SF, IF, DM, BT, GZ) are then used by a learning system to search for empirical regularities between several scenarios and then generate dangerous situations likely to occur in a particular context and which require particular attention from railway safety experts and evaluators.

4.3.3 Collection of Accident and/or Incident Scenarios

After approximately thirty interviews and knowledge-gathering sessions with rail safety experts in France, the analysis of several rail transport system safety files, particularly preliminary hazard analyses (PHAs) and functional safety analyses (FSAs), and the study of rail safety standards and European rail regulations relating to rail safety and risk management, the knowledge acquisition phase led to the inventory of approximately one hundred accident or incident scenarios relating to several risks of collision, derailment, electrocution, etc. To demonstrate the feasibility of the decision support system, this study was deliberately limited to the problem of "rail collisions". Fig. 17 shows an excerpt from the list of accident scenarios, such as redundancy switching problems, entering an occupied block, incorrect initialization, element coupling failure, element order inversion, recording failure after a switch, or crossing a stopping point in manual driving.

4.3.4 Aggregation (Grouping of Accident Scenarios)

The collected scenarios are grouped by the safety experts into around ten scenario classes such as the "Initialization", "Train localization", "Redundancy switching", "Train docking" or "Emergency braking management" class (Fig. 18).

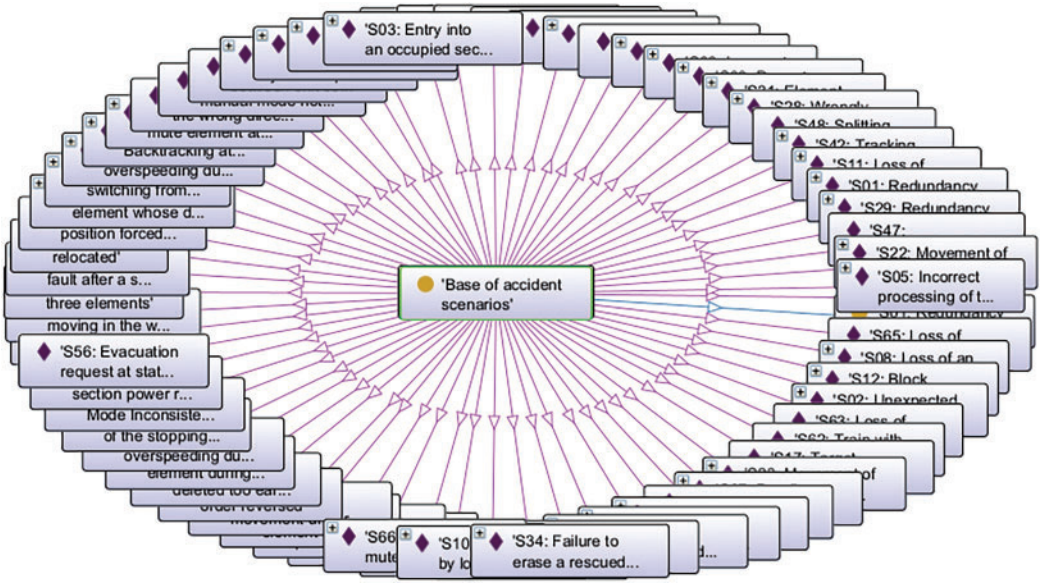

Figure 17: Accident scenario database relating to the risk of "collision"

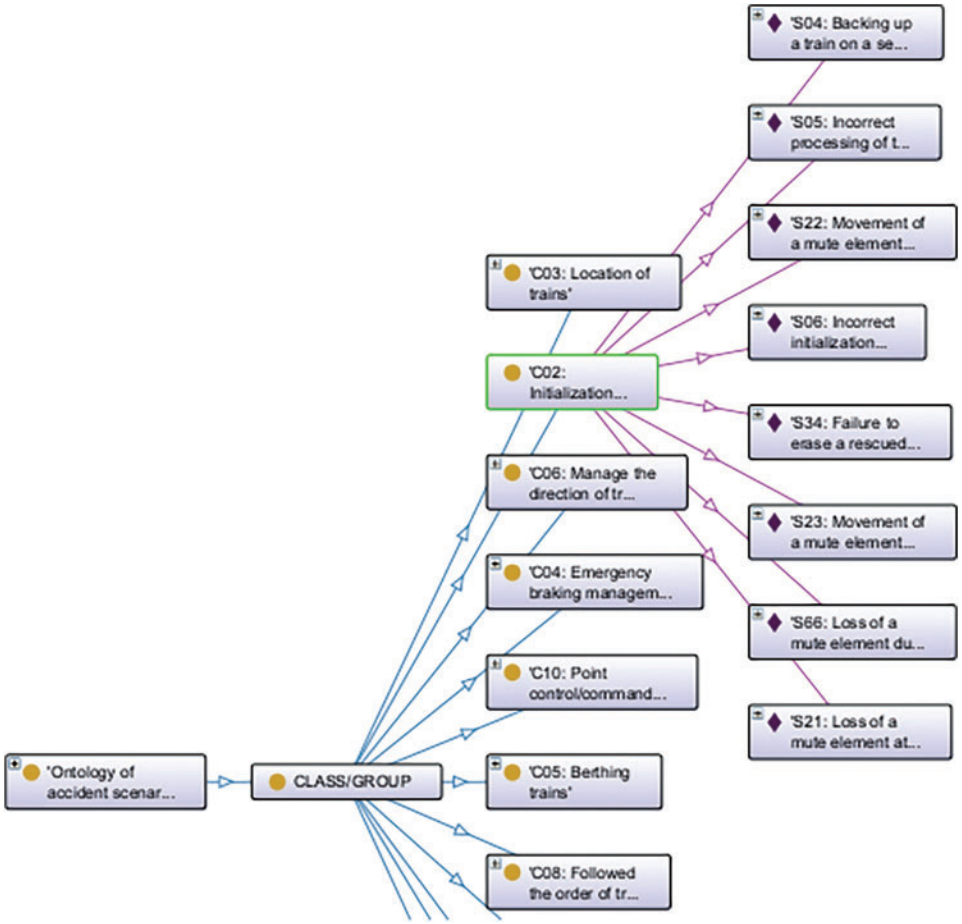

Figure 18: Classes of accident scenarios

4.3.5 Definition of a Language for Describing Scenario Examples

The language used to describe the examples must be perfectly understandable by the domain expert and semantically rich to convey the concepts of the expert assessment. The chosen language, which is directly derived from the formalism used to characterize accident scenarios, is based on a classic representation by descriptors (Attributes/Values). We have defined seven types of descriptors:

- Enumerated descriptor: in the description of an example scenario, this type of descriptor takes a single value from a set of possible values (domain of definition). For example, the attribute "Blocking Principle" (BP) takes either the value "moving block" or "fixed block".
- Multivalued descriptors such as (train tracking AND initialization AND docking). This type of descriptor can take multiple values at once from the set of possible values. It is expressed as a conjunction (&) of values.
- Unknown descriptor that has no value in the description of an example (missing attribute value) for a particular situation.
- "Key" descriptor for a given class of examples. "Key" descriptors are the descriptors deemed by the expert to be the most relevant for characterizing a class of accident scenarios.
- "Minimal" descriptors for describing an example of an accident scenario. Unlike key descriptors that characterize a class of accident scenarios, "minimal" descriptors consist of the essential attributes for characterizing an example of an accident scenario. They define the necessary (but not sufficient) conditions for an example to be admissible. Four "minimal" descriptors are identified during the knowledge acquisition phase to describe an accident scenario: Hazard (H), Potential Accident (PA), Safety Function (SF), and Geographical Zones (GZ). A scenario example is declared complete if: (1) the four "minimal" descriptors defined by the expert are present in the scenario description and (2) among the non-"minimal" (irrelevant) descriptors for describing a scenario example, some have no value; they are considered irrelevant for characterizing an example. Instead of assigning them any possible value, they are assigned the value "unknown". However, missing values are not tolerated for relevant attributes ("minimal" descriptors).
- Probable descriptor such as (GZ = terminus OR line OR section boundary). A "probable" descriptor is a descriptor that can take multiple values, declared probable by the expert, from among the set of possible values forming the definition domain. In the example above, an accident scenario can occur in several Geographical Zones, it can occur at a terminus, a line, or a section boundary. This is a disjunction of values, as opposed to a multi-valued descriptor.
- Comment descriptor: This type of descriptor is associated with a value expressed in the form of a comment. This type of descriptor is used to describe Dangers (D) and Adopted Solutions (AS). For example, H25 = Penetration of a train on a block by recoil, and AS36 = Prohibit I/O mode switching while a mode is in progress.
- Inconsistent Descriptors: The final criterion for preventing "noise" concerns the handling of data inconsistency. This involves extracting from the expert the values of an attribute "Ai" that should not be present in the attribute "Aj" when describing a scenario example. If these values are present in "Aj", then the example is inconsistent and requires verification. For example, during the knowledge acquisition phase with the safety experts, an exhaustive inventory of Safety Functions (SF) was carried out, an example of which is shown below (Fig. 19). From these two taxonomies relating to the SFs successively involved in the potential accident (PA) "collision" and "derailment", we can deduce that SF14 = "Integral authorization driving and high voltage", SF15 = "Tracking trains" and SF16 = "docking" should not be included in the description of an accident scenario relating to PA = "Derailment".

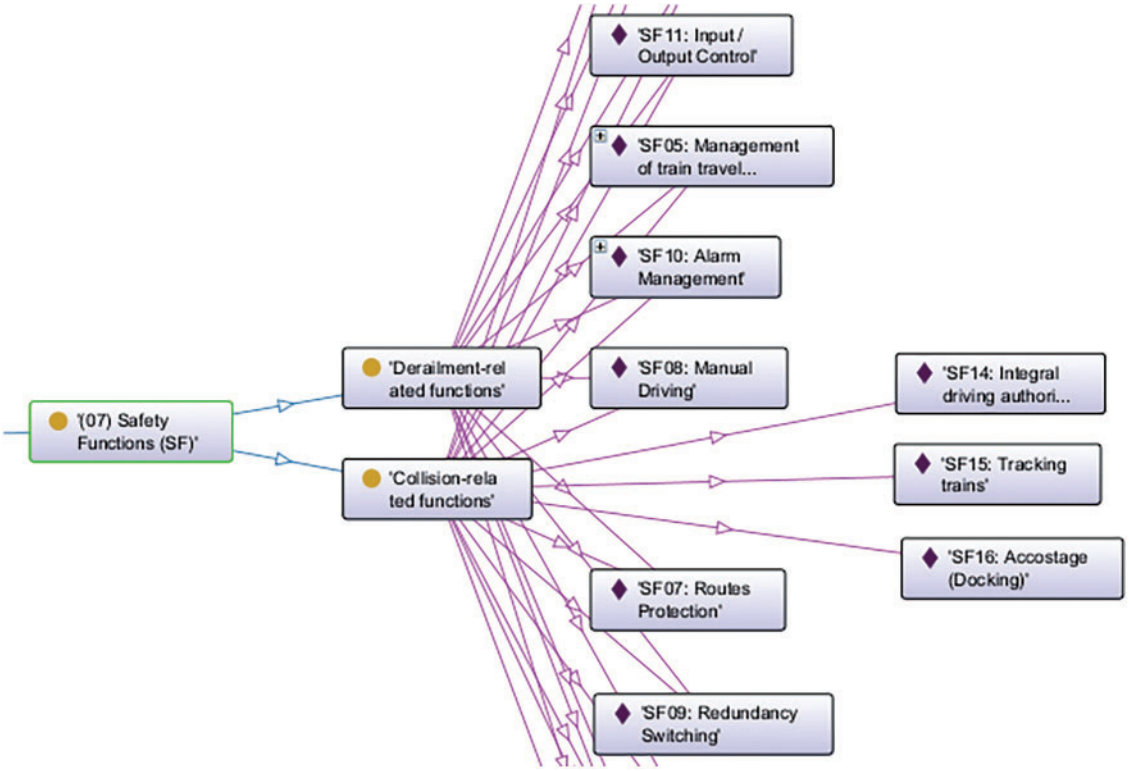

Figure 19: Examples of safety functions (SF) involved in the railway "Collision" and "Derailment"

4.4 Classification of New Scenarios

This study is part of supervised learning since the accident scenarios were grouped by the safety expert into several accident classes. Consequently, the training examples are "labeled" (or classified) and the objective is to predict the membership class of a new scenario. This is called "supervised" learning or "discriminant analysis" which is a statistical technique whose objective is to describe, explain and predict the membership to groups (scenario classes) of a set of observations (scenario examples) from a series of predictive variables (scenario characteristic descriptors). In data analysis, discriminant analysis can be either a "descriptive" technique such as discriminant factor analysis or a "predictive" technique which aims to construct a classification function (or assignment rule) to predict the membership class of an individual from the values taken by the predictive variables. A supervised learning process is structured around two steps: one to determine a model from the labeled data, and a second (testing) step, which consists of predicting the label of a new data item from the previously learned model, often based on a probability of belonging to each of the predetermined classes. This is referred to as "probabilistic" supervised learning. It is therefore necessary to learn to construct a function F such that Y = F(X), Y being one or more results calculated based on input data X (based on accident scenarios). Y can be a continuous quantity (e.g., an electric current or a speed), in which case we refer to regression, or a discrete quantity (e.g., an accident class, collision, or derailment), in which case we refer to automatic classification or supervised classification, which consists of assigning a class or category to each object (or individual) to be classified, based on statistical data. Thus, the learning approach adopted in the context of accident scenarios can be described as a supervised automatic classification method (labeled training data), discriminative, probabilistic, predictive, and discrete, the objective of which is, on the one hand, to discover the relevant characteristics of the accident classes grouped by the safety expert and,

on the other hand, to predict the membership class of the new scenarios whose relevance the expert seeks to assess with regard to the safety of a new railway system.

For reasons of readability, the mathematical calculation method is not presented in this article. For further information on the approach to classifying accident scenarios, the reader can consult the article [33].

Fig. 20 illustrates the architecture of the incremental learning system intended for the classification and capitalization of accident scenario classes linked to the risk of "railway collision". This classification system consists of the following five modules:

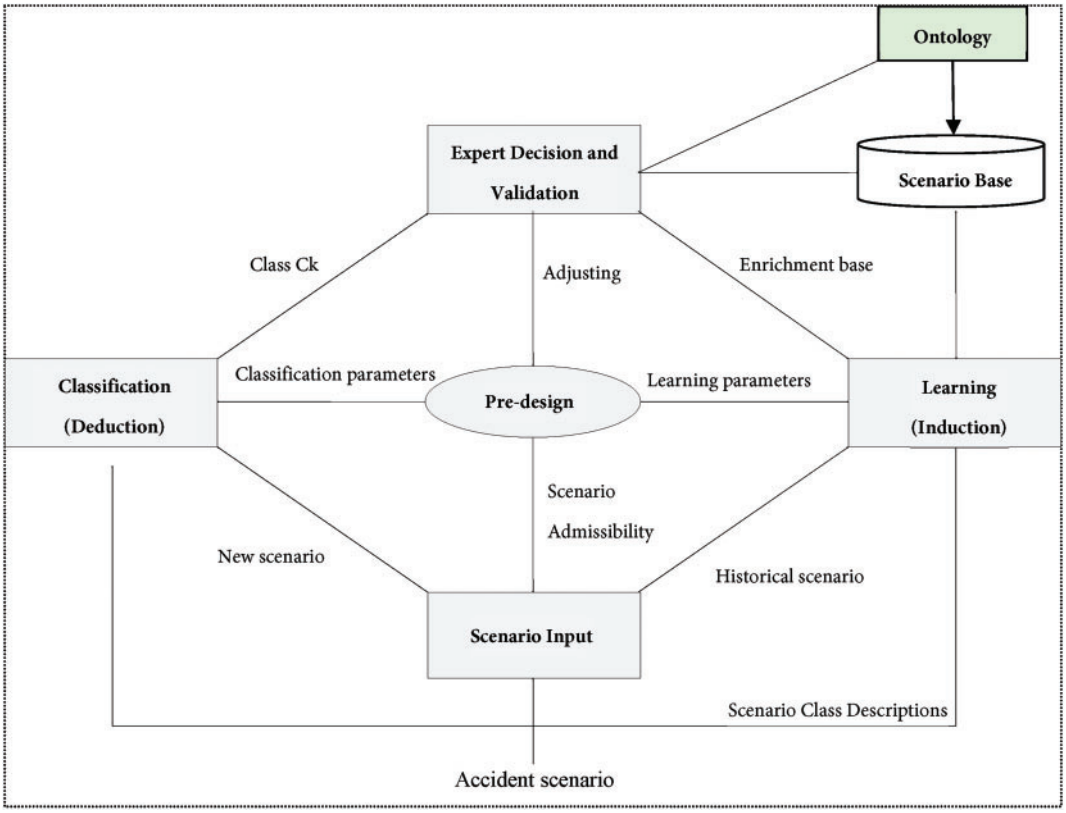

Figure 20: Supervised automatic classification approach for railway accident scenarios [33]

- A module for entering three types of accident scenarios: (1) new scenario to verify its admissibility before processing; (2) known scenario, pre-classified, tested by experts, and whose class Ck is known; and (3) scenario to be classified (new scenario to be classified whose consistency the expert seeks to assess).
- A pre-design module is used to set the various values of the learning parameters and constraints required by the decision support system. During this stage, the user defines the learning parameters (induction, classification, and convergence parameters) and the admissibility constraints of a scenario, which define the conditions necessary for its acceptance by the system. All these parameters primarily influence the relevance and quality of the learned knowledge, as well as the system's convergence speed.
- An induction module (concept learning) for learning the conjunctive descriptions of scenario classes.
- A classification module whose objective is to deduce the class Ck to which a new scenario belongs based on the previously derived class descriptions and with reference to an Adequacy rate.

– A dialog module for system argumentation and expert decision-making. During the argumentation or justification phase, the system keeps track of the deduction phase to construct its explanation. Following this classification decision justification phase, the safety expert decides either to accept the proposed classification, in which case the scenario database will be updated or to reject the classification. In the second case, it is up to the expert to decide how to proceed. For example, they may decide to evaluate the scenario using the expert system based on the learning of association rules which we detail below.

The data and knowledge required to implement these modules are derived from the ontology (developed above) in close collaboration with the user and the safety expert.

4.5 Deduction of Hazardous Events

Let us recall that an accident or incident scenario was modeled by the ontology through eleven descriptive parameters (Fig. 21): (1) Hazardous Elements (HE), (2) Hazards (H), (3) Potential Accidents (PA), (4) Damage (D), (5) Risk Levels (RL), (6) Safety Measures (SM), (7) Safety Functions (SF), (8) Incident Functions (IF), (9) Driving Modes (DM), (10) Type Block (TB) and (11) Geographic Zone (GZ).

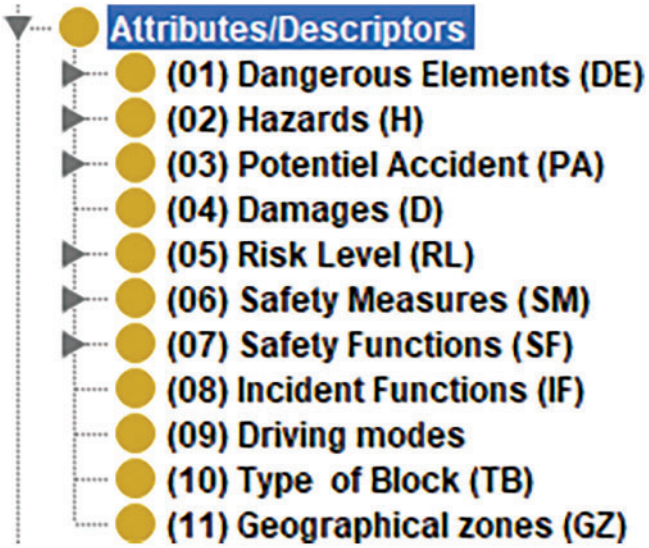

Figure 21: Descriptive parameters (or descriptors) of an accident scenario

Let us also recall that the conceptual model previously developed made it possible to identify the "causal links" between Dangerous Element (DE), Hazards (H), Potential Accident (PA), Damages (D), and Safety Measures (SM) (Fig. 22):

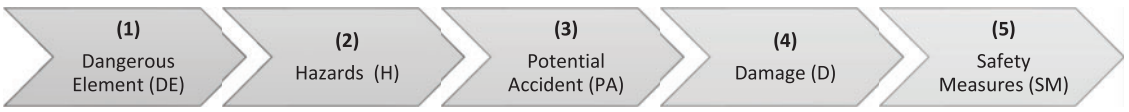

Figure 22: Causal link between the descriptors of a scenario

The following descriptors of a scenario: Risk Levels (RL), Safety Functions (SF), Incident Functions (IF), Driving Modes (DM), Type Block (TB), and Geographic Zone (GZ) will now be called "Static Context" of

the scenario's appearance as opposed to the "Dynamic Context" which describes the evolution and progress of a scenario. Thus, the other descriptors: Dangerous Element (DE), Hazards (H), Potential Accident (PA), Damage (D), and Safety Measures (SM) describe the causal links between the "key" descriptors which evolve dynamically, thus creating a real potential accident scenario (Fig. 23).

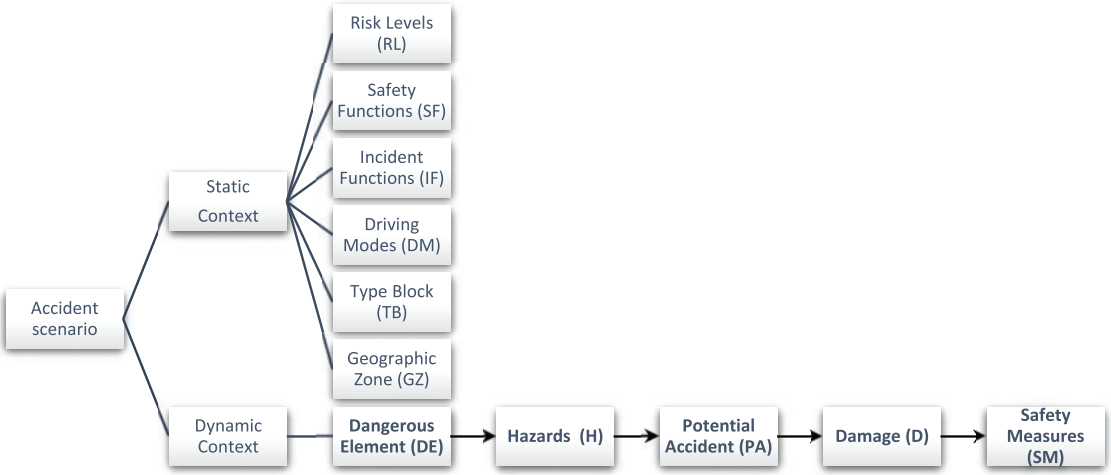

Figure 23: Decomposition of the potential accident scenario into "Static Context" and "Dynamic Context"

This approach allows for the guidance of rules discovered through learning by using a "dynamic context", meaning that the rules must be organized in accordance with the following causal links: Hazardous Elements (HE) Hazards (H) Potential Accidents (PA) Damage (D) Safety Measures (SM).

The use of a learning method that allows the generation of rules oriented from one descriptor to another from a set of examples (or scenarios) is essential. The specification of the properties required by the learning system, as well as the analysis of existing systems, led to the choice of the CHARADE [34] learning system, which not only allows the generation of a structured rule system that can be used by an inference engine but also allows the simultaneous learning of certain logical rules and uncertain rules modulated by a likelihood coefficient. The automatic induction of a rule system, rather than isolated rules, as well as the ability to structure the rules, give CHARADE an undeniable interest. Rule generation in CHARADE is based on the search for and discovery of empirical regularities present in the training set. A regularity corresponds to an observed correlation between descriptors in the training example database: if all the examples in the training set that possess the descriptor d1 also possess the descriptor d2, we can infer that d1 d2 on the training set. To illustrate this principle of rule generation, suppose we have a training set consisting of three scenario examples S1, S2, and S3:

- S1 = DE & H & SF & PA
- S2 = DE & H & PA & GZ
- S3 = DE & H & SF & PA & IF

We can then detect an empirical regularity between the conjunction of descriptors (DE & H) and the descriptor PA. Indeed, all the scenario examples described by DE & H also possess PA in their description. This regularity is obtained using two functions: function C (lattice of descriptors) and function D (lattice of examples). Thus, CoD (DE & H) = C ({S1, S2, S3}) = DE & H & PA. Finally, the rule: DE & H PA is obtained.

The proposed approach to assist in the assessment and discovery of potential accident scenarios is organized around the following four modules, illustrated in Fig. 24:

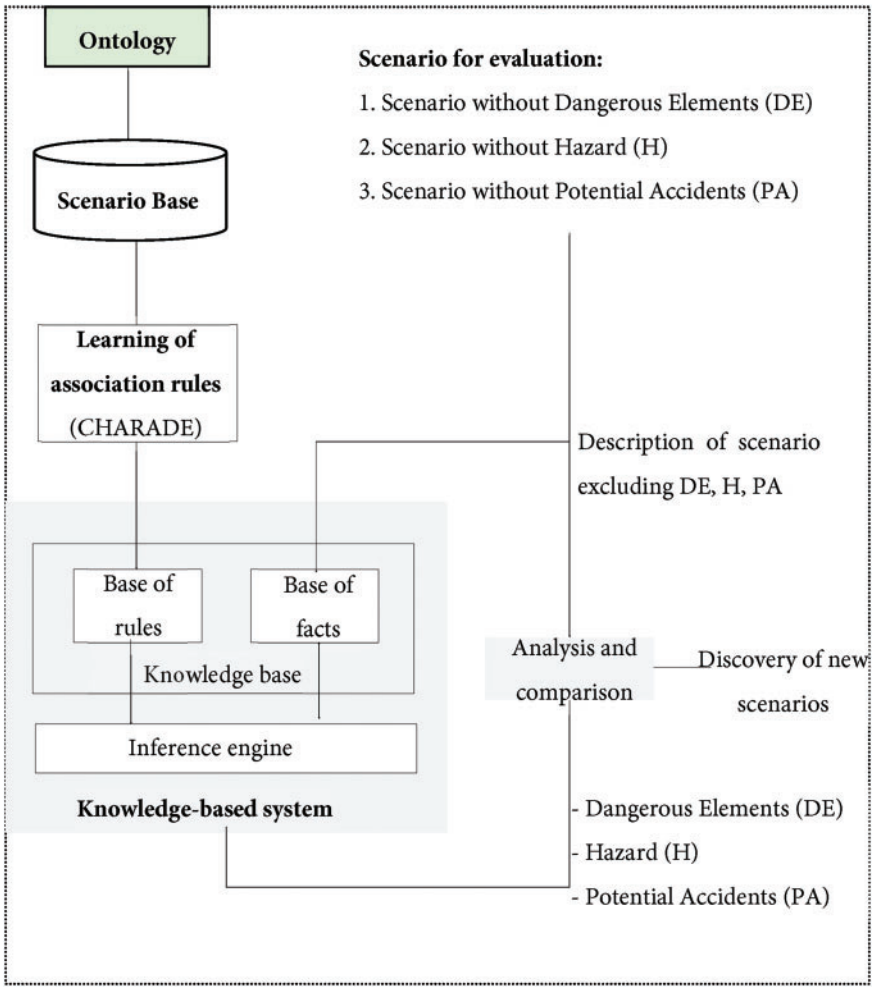

Figure 24: Expert system based on association rule learning to assist in the discovery of accident scenarios [33]

- A scenario database (training examples) derived from the previous classification system and capitalized in the ontology;
- An association rule learning system called CHARADE uses this database of examples to produce recognition functions for Dangerous Elements (DE), Hazards (H), and Potential Accidents (PA);
- A rules translation and transfer module. The rules produced by CHARADE are written using a specific syntax. They are translated to be compatible with the expert system generator, which does not process multi-valued descriptors, but only enumerated or Boolean descriptors. They are then automatically transferred to the expert system's knowledge base;
- A knowledge base for assessing scenarios used by the expert system's inference engine to deduce the DE, H, and PA to be considered in the new scenario proposed by the expert.

Here is an example of rules generated by the Charade system where the rules are oriented towards the "Dangers" descriptor at the end of the rules.

If Hazardous elements = mobile operator,
Incident functions = instructions,
Hazardous elements = operator in CC,
Then Hazards = H11 (Invisible element on the zone of completely automatic driving),
Hazardous elements = AP with redundancy,
Safety functions = train localization,
Geographic zone = terminus.
[0]
If Type of block = fixed block,
Safety functions = initialization,
Incident functions = instructions,
Then Hazards = H10 (Erroneous re-establishment of safety frequency/high voltage)
Safety functions = Full control/High voltage permission
Safety functions = alarm management,
Safety functions = train localization.
[0]
If Safety functions = train localization,
Hazardous elements = AP without redundancy,
Then Hazards = H9
(Entry of a train into an occupied block),
Geographic zone = line,
Type of block = fixed block.
[0]

The evaluation process requires a preliminary phase during which the rules generated by CHARADE are transferred to an expert system in order to build a knowledge base for the evaluation of scenarios. Based on the deductions (Forward and Backward Chaining) performed by the inference engine of the expert system, the objective is to compare in particular the list of Dangerous Element (DE), Dangers (H), and Potential Accident (PA), proposed in a new scenario developed by the safety expert with the list of historical Dangerous Element (DE), Hazards (H) and Potential Accident (PA) stored in the rule base of the expert system in order to stimulate the formulation of dangerous situations not anticipated by the expert during the development of hazard analyses. This evaluation task draws the expert's attention to the events contrary to safety not considered during the specification and design phases of the system and likely to compromise the safety of a new railway system. It can thus promote the generation of new accident scenarios.

For further information on modeling, classification, and evaluation of railway accident scenarios, the reader can consult the works [33,35,36].

This phase of assessing and assisting in the generation of potential accident scenarios is currently ongoing to develop a hybrid reasoning approach. Indeed, while theory advocates an "inductive" approach (consequences/causes) to railway hazard analysis, the procedures applied in practice are mostly "deductive" (causes/consequences). The approach (currently under development) that we recommend explicitly combines the two approaches in order to enhance the quality of the analysis in terms of completeness and consistency. Indeed, analyzing the safety of a complex system requires experts in the field to implement

an iterative analysis process combining “inductive” and “deductive” approaches. Thus, the hybrid approach envisaged is structured around three complementary and iterative stages (Fig. 25).

- From the (PA), the first step determines the list of (D) by “induction” and the list of (H) by “deduction”;
- The previous (PA) is used to find the list of (DE) by “deduction” and, by “induction”, the list of (PA). This is a verification loop allowing the initial list of (PA) to be expanded if necessary;
- From the previous (DE), the third step allows the (PA) to be “induced” which will, in turn, be compared to those from the first step. Generating a new (H) requires repeating the two previous analysis steps.

This is an iterative process allowing for a comprehensive railway risk analysis.

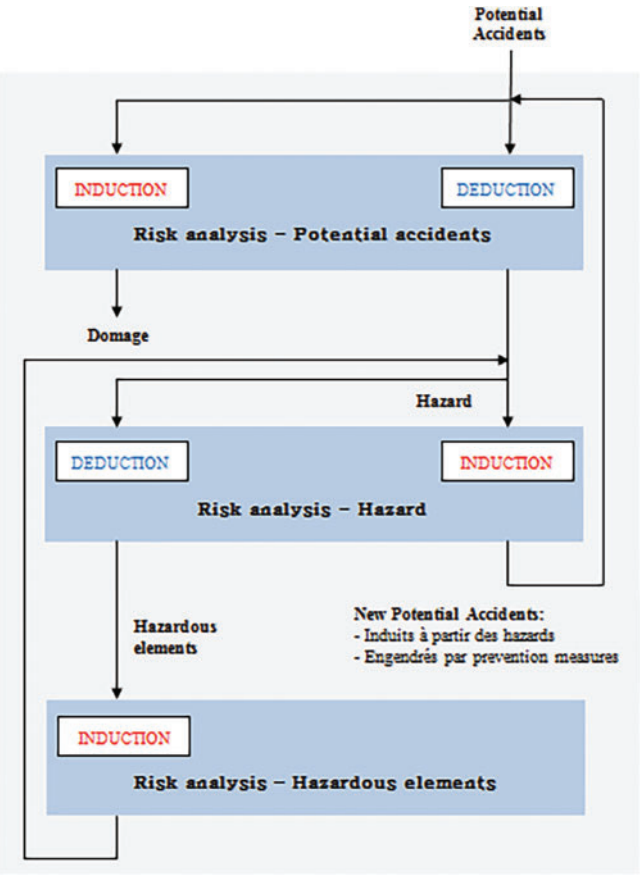

Figure 25: General description of the hazard analysis method

5 Contribution of Ontology to Explainable Artificial Intelligence (IAX)

In recent years, explainability and/or interpretability has become a key factor for the adoption of AI systems. It is now a very active research topic and has seen a resurgence of interest in order not only to gain the trust of the system’s users but also to demonstrate “how” and “why” a learning system, particularly deep learning (based on artificial neural networks), led to such a decision. The objective is to understand the reasoning carried out following a decision as well as the predictions made by the learning algorithms (“black boxes”) in order to make them more intelligible, transparent, and understandable for the user of the AI system. Thus, the ability to explain the reasons for a decision has become an indispensable property of AI

systems. However, to date, there is no clear consensus between the term's "explanation", "interpretation" and "transparency".

Graziani et al. [37] highlight the problems of terminological discrepancies and inconsistencies between the terms "interpretable", "explainable", and "transparency" and propose a "global taxonomy of interpretable AI" to unify this terminology:

- *Interpretability of AI defines AI systems for which it is possible to translate the operating principles and results into a human-understood language without affecting the validity of the system.*
- *Explainable AI defines the branch of AI research that focuses on generating explanations for complex AI systems.*
- *Transparency is used in AI to characterize systems for which the role of internal components, paradigms, and overall behavior is known and can be simulated.*

Confalonieri et al. [38] reviewed the literature on explainable artificial intelligence (XAI), not only reviewing traditional approaches but also approaches currently under development. The authors describe the different notions of explanations in expert systems (explanations as lines of reasoning and explanations as problem-solving activities), in machine learning (local, global, introspective, or counterfactual explanations), in recommender systems (explainable recommendation models and explanation styles), and in neural-symbolic learning and reasoning. According to the authors, explanations generated by ontologies, conceptual networks, or knowledge graphs can support reasoning by implementing several forms of knowledge use, such as abstraction and refinement.

Several other similar works use ontologies for explainability purposes.

To address the explainability problem, Confalonieri et al. [39] fed the classifier with data from a knowledge graph and describe its behavior using rules expressed in the knowledge graph terminology. The latter is perceived by the authors as a structured, common, and understandable representation of a domain based on how humans mentally perceive the world. Indeed, the representation of structured knowledge in the form of ontologies plays an important role in explainable artificial intelligence (XAI), not only to enrich explanations with semantic information but also to support the personalization of the specificity and generality levels of explanations based on user profiles [40]. The study proposed by [41] makes it possible to exploit the properties of ontology to improve explainability. To improve the explainability of cardiac AI, Tsolakis et al. [42] used ontologies to evaluate and ensure the clarity and relevance of explanations. Bellucci et al. [43] combined an ontology-based explainable model with an explanation interface to classify images.

However, to our knowledge, there is currently no work on explainable AI applied to the field of railway safety. Indeed, the proposed article demonstrates that the joint use of ontology, expert systems, and machine learning to improve railway safety from accident scenarios has not yet been studied. The representation of domain knowledge in the form of ontologies undoubtedly promotes the interpretability of the data produced by the learning system. In addition, expert systems (or knowledge-based systems) can be considered interpretable by design because they were developed to assist humans in decision-making.

6 Conclusion

After a brief introduction to the methods, algorithms, and types of machine learning (ML), a rapid literature review on the applications of AI and ML in rail transport was presented. Drawing inspiration from the two European directives relating to the development and interoperability of the European railway system, we have proposed a classification of AI applications related to rail transport, distinguishing between:

(1) AI approaches related to the "Structural Elements" of the railway system: (a) "railway infrastructure" (switches, rail, ballast, track geometry, level crossings, tunnels, etc.), (b) "rolling stock" (axles, wheels, pantographs, locomotives, wagons, etc.)
(2) AI approaches related to the "Functional Elements" of the railway system (operations/traffic management, maintenance, telematics applications).

Particular emphasis is placed on AI approaches related to railway accident and incident investigations to explore the causes of hazards identify actors and safety risk factors, model correlations between accident-related hazards, discover accident characteristics and peculiarities, etc.

Despite the undeniable interest in these approaches, their implementation assumes that the railway system has already received authorization for commissioning.

The proposed approach improves railway safety from the specification and design phases and is based on a set of potential accident scenarios developed from railway system design files as well as the know-how and experience of safety experts.

This approach is essentially based on the development of railway safety ontology to harmonize the fundamental concepts related to railway risk management. This ontology, whose objective is to promote the interpretability of learning data by safety experts, is used during the classification, evaluation, and generation of potential accident scenarios. The primary benefit of the proposed approach is to improve the completeness of hazard analysis, which is essential when developing a railway safety management system (SMS).

To date, the tools developed (scenario classification algorithm, expert system, production rule learning system "Charade", an approach to generating new dangerous situations), are at the model stage, but an initial overall validation by experts has demonstrated the interest of the proposed approaches to improve railway risk management from the system design phase.

The knowledge acquisition and modeling steps involved in accident scenarios represent a laborious effort that required approximately thirty data and knowledge-gathering sessions. To this end, and to demonstrate the feasibility of the proposed approach, we deliberately limited the development to a single accident type: "collision". However, the overall architecture of the decision support system is open and can handle other accident types such as "derailment". The number of scenarios processed to date is approximately one hundred related to train collisions. We are aware that this reduction in the number of scenarios may affect the quality of the training example database and obviously raises the performance issue of the developed learning system. Therefore, it is essential to enrich the training example database with additional scenarios. It should be remembered that this was initially a feasibility study aimed at demonstrating the contribution of AI, particularly machine learning and ontology, to railway risk management.

Acknowledgement: Not applicable.

Funding Statement: Not applicable.

Availability of Data and Materials: Data not available due to data confidentiality restrictions on railway accidents.

Ethics Approval: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest to report regarding the present study.

References

1. Zantalis F, Koulouras G, Karabetsos S, Kandris D. A review of machine learning and IoT in smart transportation. Future Internet. 2019;11(4):94. doi:10.3390/fi11040094.
2. Iyer LS. AI enabled applications towards intelligent transportation. Transp Eng. 2021;5(189):100083. doi:10.1016/j.treng.2021.100083.
3. Jevinger Å, Zhao C, Persson JA, Davidsson P. Artificial intelligence for improving public transport: a mapping study. Public Transp. 2024;16(1):99–158. doi:10.1007/s12469-023-00334-7.
4. McMillan L, Varga L. A review of the use of artificial intelligence methods in infrastructure systems. Eng Appl Artif Intell. 2022;116(1):105472. doi:10.1016/j.engappai.2022.105472.
5. Tselentis DI, Papadimitriou E, van Gelder P. The usefulness of artificial intelligence for safety assessment of different transport modes. Accid Anal Prev. 2023;186(1):107034. doi:10.1016/j.aap.2023.107034.
6. Attoh-Okine N. Big data challenges in railway engineering. In: 2014 IEEE International Conference on Big Data (Big Data); 2014 Oct 27–30; Washington, DC, USA. p. 7–9. doi:10.1109/BigData.2014.7004424.
7. Thaduri A, Galar D, Kumar U. Railway assets: a potential domain for big data analytics. Procedia Comput Sci. 2015;53(1):457–67. doi:10.1016/j.procs.2015.07.323.
8. van Gulijk C, Hughes P, Figueres-Esteban M, El-Rashidy R, Bearfield G. The case for IT transformation and big data for safety risk management on the GB railways. Proc Inst Mech Eng Part O J Risk Reliab. 2018;232(2):151–63. doi:10.1177/1748006x17728210.
9. Ghofrani F, He Q, Goverde RMP, Liu X. Recent applications of big data analytics in railway transportation systems: a survey. Transp Res Part C Emerg Technol. 2018;90(1–2):226–46. doi:10.1016/j.trc.2018.03.010.
10. Laiton-Bonadiez C, Branch-Bedoya JW, Zapata-Cortes J, Paipa-Sanabria E, Arango-Serna M. Industry 4.0 technologies applied to the rail transportation industry: a systematic review. Sensors. 2022;22(7):2491. doi:10.3390/s22072491.
11. Dong K, Romanov I, McLellan C, Esen AF. Recent text-based research and applications in railways: a critical review and future trends. Eng Appl Artif Intell. 2022;116(7):105435. doi:10.1016/j.engappai.2022.105435.
12. Bešinović N, De Donato L, Flammini F, Goverde RMP, Lin Z, Liu R, et al. Artificial intelligence in railway transport: taxonomy, regulations, and applications. IEEE Trans Intell Transp Syst. 2022;23(9):14011–24. doi:10.1109/TITS.2021.3131637.
13. Tang R, De Donato L, BeSinović N, Flammini F, Goverde RMP, Lin Z, et al. A literature review of Artificial Intelligence applications in railway systems. Transp Res Part C Emerg Technol. 2022;140(3):103679. doi:10.1016/j.trc.2022.103679.
14. Chenariyan Nakhaee M, Hiemstra D, Stoelinga M, van Noort M. The recent applications of machine learning in rail track maintenance: a survey. In: Reliability, safety, and security of railway systems. modelling, analysis, verification, and certification. Cham Switzerland: Springer; 2019. p. 91–105. doi:10.1007/978-3-030-18744-6_6.
15. Katsumi M, Fox M. Ontologies for transportation research: a survey. Transp Res Part C Emerg Technol. 2018;89(6):53–82. doi:10.1016/j.trc.2018.01.023.
16. Yang L, Cormican K, Yu M. Ontology-based systems engineering: a state-of-the-art review. Comput Ind. 2019;111:148–71. doi:10.1016/j.compind.2019.05.003.
17. Tiddi I, Schlobach S. Knowledge graphs as tools for explainable machine learning: a survey. Artif Intell. 2022;302(5):103627. doi:10.1016/j.artint.2021.103627.
18. Hadj-Mabrouk H. Literature review on applications of ontologies and knowledge graphs in railway transport safety. In: Book railway transport and engineering. London, UK: IntechOpen; 2024. 57 p. doi:10.5772/intechopen.1006278.
19. Hadj-Mabrouk H. A literature review on the applications of artificial intelligence to European rail transport safety. IET Intell Transp Syst. 2024;18(12):2291–324. doi:10.1049/itr2.12587.
20. Directive 2012/34/EU of the European Parliament and of the Council of 21 November 2012 establishing a single European railway area (recast) [Internet]. [cited 2025 Jun 1]. Available from: http://data.europa.eu/eli/dir/2012/34/oj.

21. Directive (EU) 2016/797 of the European Parliament and of the Council of 11 May 2016 on the interoperability of the rail system within the European Union (recast) [Internet]. [cited 2025 Jun 1]. Available from: http://data.europa.eu/eli/dir/2016/797/oj.
22. Cooray S. The subjectivity of data scientists in machine learning design. J Comput Inf Syst. 2024;64(5):665–82. doi:10.1080/08874417.2023.2240755.
23. Niu Y, Fan Y, Ju X. Critical review on data-driven approaches for learning from accidents: comparative analysis and future research. Saf Sci. 2024;171(1):106381. doi:10.1016/j.ssci.2023.106381.
24. Tamascelli N, Campari A, Parhizkar T, Paltrinieri N. Artificial intelligence for safety and reliability: a descriptive, bibliometric and interpretative review on machine learning. J Loss Prev Process Ind. 2024;90:105343. doi:10.1016/j.jlp.2024.105343.
25. Richardson S. Exposing the many biases in machine learning. Bus Inf Rev. 2022;39(3):82–9. doi:10.1177/02663821221121024.
26. Xu Y, Kohtz S, Boakye J, Gardoni P, Wang P. Physics-informed machine learning for reliability and systems safety applications: state of the art and challenges. Reliab Eng Syst Saf. 2023;230(4):108900. doi:10.1016/j.ress.2022.108900.
27. Rohlfing KJ, Cimiano P, Scharlau I, Matzner T, Buhl HM, Buschmeier H, et al. Explanation as a social practice: toward a conceptual framework for the social design of AI systems. IEEE Trans Cogn Dev Syst. 2021;13(3):717–28. doi:10.1109/tcds.2020.3044366.
28. Longo L, Brcic M, Cabitza F, Choi J, Confalonieri R, Del Ser J, et al. Explainable artificial intelligence (XAI) 2.0: a manifesto of open challenges and interdisciplinary research directions. Inf Fusion. 2024;106(3):102301. doi:10.1016/j.inffus.2024.102301.
29. Thekdi S, Aven T. Understanding explainability and interpretability for risk science applications. Saf Sci. 2024;176(4):106566. doi:10.1016/j.ssci.2024.106566.
30. Webprotege [Internet]. [cited 2025 Jun 1]. Available from: https://webprotege.stanford.edu/.
31. Feilmayr C, Wöß W. An analysis of ontologies and their success factors for application to business. Data Knowl Eng. 2016;101:1–23. doi:10.1016/j.datak.2015.11.003.
32. Protégé [Internet]. [cited 2025 Jun 1]. Available from: https://protege.stanford.edu/.
33. Hadj-Mabrouk H. Contribution of artificial intelligence to risk assessment of railway accidents. Urban Rail Transit. 2019;5(2):104–22. doi:10.1007/s40864-019-0102-3.
34. Ganascia J-G. Agape and Charade: two Symbolic Learning Mechanisms Applied to the Construction of Knowledge Bases [Ph.D. thesis] Paris, France: University of Paris-Sud; 1987 [cited 2025 Jun 1]. Available from: https://api.semanticscholar.org/CorpusID:169643071. (In French).
35. Hadj-Mabrouk H. Contribution of machine learning to rail transport safety. Adv Mach Learn Clean Energy Transp Ind. 2021;2021:277–312. doi:10.52305/SJDR3905.
36. Habib HM. Approach to assist in the discovery of railway accident scenarios based on supervised learning. In: Transportation energy and dynamics. Singapore: Springer Nature; 2023. p. 129–56. doi:10.1007/978-981-99-2150-8_7.
37. Graziani M, Dutkiewicz L, Calvaresi D, Amorim JP, Yordanova K, Vered M, et al. A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences. Artif Intell Rev. 2023;56(4):3473–504. doi:10.1007/s10462-022-10256-8.
38. Confalonieri R, Coba L, Wagner B, Besold TR. A historical perspective of explainable artificial intelligence. Wires Data Min Knowl Discov. 2021;11(1):e1391. doi:10.1002/widm.1391.
39. Liartis J, Dervakos E, Menis-Mastromichalakis O, Chortaras A, Stamou G. Searching for explanations of black-box classifiers in the space of semantic queries. Semant Web. 2024;15(4):1085–126. doi:10.3233/sw-233469.
40. Confalonieri R, Kutz O, Calvanese D, Alonso-Moral JM, Zhou SM. The role of ontologies and knowledge in explainable AI. Semant Web. 2024;15(4):933–6. doi:10.3233/sw-243529.
41. Kosov P, El Kadhi N, Zanni-Merk C, Gardashova L. Semantic-based XAI: leveraging ontology properties to enhance explainability. In: 2024 International Conference on Decision Aid Sciences and Applications (DASA); 2024 Dec 11–12; Manama, Bahrain. p. 1–5. doi:10.1109/DASA63652.2024.10836289.

42. Tsolakis N, Maga-Nteve C, Vrochidis S, Bassiliades N, Meditskos G. Enhancing cardiac AI explainability through ontology-based evaluation. In: 2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA); 2024 Jul 17–19; Chania Crete, Greece. p. 1–4. doi:10.1109/IISA62523.2024.10786663.
43. Bellucci M, Delestre N, Malandain N, Zanni-Merk C. Combining an explainable model based on ontologies with an explanation interface to classify images. Procedia Comput Sci. 2022;207(2):2395–403. doi:10.1016/j.procs.2022.09.298.