

Artificial Intelligence Applications to Enhance Risk Management of Railway Station Operations

By
Hamad Ali Alawad

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

College of Engineering and Physical Sciences
School of Engineering
University of Birmingham
30 September 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

Railway stations are a vital part of the railway system and a critical hub of public transportation. This valuable infrastructure in the heart of cities requires effective risk management processes as well as high-level safety. However, railway stations have faced increasing operational challenges in recent years as a result of the growth in urban areas, high demand, intensive usage, and an increase in train capacity often without any consideration of the size of the existing stations. Consequently, increased risks may exist and lead to train delay, injuries, or fatalities as well as loss with disruption of the whole system and impact on business. Risk management is thus an essential part of the safety system, a system which currently is mostly traditional and does not take advantage of available technologies. The traditional risk process has uncertainties, it takes time, is costly and it depends on experts who are not available all the time. These processes could be developed to cope with future requirements by creating smarter stations and cities, embracing new technologies such as artificial intelligence (AI). It is both wise and necessary to make use of all possible technologies to stop or mitigate the risks and consequences. This can be achieved by improving the risk management process, which will then reflect positively on the quality, reliability, safety, security, as well as cost. This study proposes utilising (AI) methods to improve railway stations' safety and risk management. The methods have powerful analytics and predictive ability. Machine learning (ML) has been proposed using the classification tree, Fuzzy and Nerul network (ANFIS), and deep learning (DL) to identify the risks and people's behaviours in the stations. The information examined is from occurrences at UK stations and ended in casualties. The characteristics of the process and the selected approach are fundamental to selecting the appropriate data. An advanced analytical approach is adopted, using predictions via classification in addition to timely detections. The proposed application presents novel work contributing to the field and aiming to save lives in the future. DL shows a great opportunity to capture many risks in the stations. The outcomes provide a unique contribution where the approach is

applicable and makes the process smarter with less human intervention, whilst also supporting decision-makers in real-time. This will also upgrade the risk management to be effective, reflect positively on people's lives and provide a safe environment for all of society.

ACKNOWLEDGEMENT

First, I thank Allah for facilitating this opportunity of learning.

My most genuine gratitude goes out to my supervisor, Dr Sakdirat Kaewunruen, for his great supervision and guidance, and also my thanks to Professor Min An for all the support along this journey.

A special thanks to my family, especially my parents for encouraging me during the Covid-19 pandemic. Thank you to everyone who helped me succeed and believed in me.

TABLE OF CONTENTS

ABSTRACT i

ACKNOWLEDGEMENT iii

TABLE OF CONTENTS iv

LIST OF FIGURES ix

LIST OF TABLES xiv

CHAPTER 1: OUTLINE AND RESEARCH CONTRIBUTION 1

1.1 Introduction 1

1.2 Research objectives and scope 3

1.3 Research problem 4

1.3.1 The station 5

1.3.2 The process (risk management) 7

1.4 Research contribution 8

1.5 Novelty of the research 11

1.6 The methodology 13

1.6.1 Machine learning 14

1.6.2 Deep learning 15

1.6.3 The study arrangement and interconnection 16

1.6.4 Research datasets 17

1.7 Research achievements 18

1.8 Organisation of the thesis (outline) 20

CHAPTER 2: LITERATURE REVIEW 22

2.1 Introduction 22

2.2 Machine learning (ML) literature 26

2.3 Deep learning (DL) literature 28

2.4 Big data and railway 31

2.5 AI and risk management framework implementations in the railway station 34

2.5.1 Smart railway stations 36

2.6 Outcomes and discussions 37

2.7 Summary 42

CHAPTER 3 RAILWAY AND RISK MANAGEMENT 44

3.1 Introduction 44

3.2 Risk management process 45

3.2.1 Risk management engineering and AI 50

3.2.2 Risk management and railway stations 51

3.3 Risk identification and assessment 54

3.4 Risk analysis in the stations 60

3.5 Fault tree analysis (FTA) 62

3.6 Summary 64

CHAPTER 4 THE IDENTIFICATION OF THE RISK ROOTS 65

4.1 Introduction 65

4.2 Fire and explosion risk 65

4.3 Slip/trip Risk 73

4.4 Train derailment risk 74

4.5 Security system failure risk 80

4.6 Overcrowding risk 84

3.7 Flooding risk 88

4.8 Earthquake risk 91

4.9 Risk of transporting dangerous goods 93

4.10 The disruption effect 95

4.11 Structure failure 96

4.12 Discussion 98

4.13 Summary 102

CHAPTER 5: ARTIFICIAL INTELLIGENCE (AI) METHODOLOGY 104

5.1 Introduction 104

5.2 Artificial Intelligence and risk management 106

5.2.1 Artificial intelligence boosting the innovation capability of railway safety 108

5.2.2 Framework of developing AI for the railway station risk 109

5.3 Machine learning (ML) 110

5.4 Supervised learning 112

5.4.1 Classification 113

5.4.2 Decision tree (DT) 113

5.5 Neural networks 117

5.6 Fuzzy systems 119

5.6.1 Neuro Fuzzy Approach -Adaptive neuro-fuzzy inference system (ANFIS) 121

5.7 Unsupervised learning 123

5.7.1 Clustering 124

5.7.2 Topic modelling 125

5.7.2.1 Latent Dirichlet Allocation (LDA) 125

5.8 Deep Learning (DL) 127

5.9 Summary 131

CHAPTER 6: MACHINE LEARNING FOR THE RISK MANAGEMENT OF A STATION 133

6.1 Introduction 133

6.2 Decision tree classification 134

6.3 A Case study 1 134

6.4 Preparing data 135

6.5 The analysis 137

6.6 Prediction and analysis 141

6.7 Learning and validation 146

6.8 Deep Learning and risk management 151

6.9 Case study 2 152

6.10 Preparing data 155

6.11 The analysis and prediction 158

6.12 The validation 159

6.13 Summary 162

CHAPTER 7: FUZZY AND NEURAL NETWORK FOR OVERCROWDING RISK MANAGEMENT IN RAILWAY STATIONS 163

7.1 Introduction 163

7.2 Adaptive neuro-fuzzy inference system (ANFIS) 163

7.3 Case study1 165

7.4 Preparing data 167

7.5 The analysis 168

7.6 Prediction and analysis 171

7.7 Learning and validation 173

7.8 Chapter summary 180

CHAPTER 8: TEXT MINING FOR RISK MANAGEMENT IN RAILWAY STATIONS 182

8.1 Introduction 182

8.2 Topic model for railway station safety 183

8.3 Preparing data 185

8.4 Model analysis 187

8.5 Clustering and analysing 197

8.6 Discussion 200

8.7 Chapter summary 201

CHAPTER 9: CONCLUSIONS AND RECOMMENDATIONS 203

9.1 How AI preventing accidents 203

9.2 AI and covid-19 204

9.3 ANFIS controller 205

9.4 Topic modelling 206

9.5 Conclusion 207

9.6 Implementation and future work 211

References 212

LIST OF FIGURES

Figure 1. The two views of the research problem. 5

Figure 2. The total number of passenger journeys trend in the UK (in billions) (Office of Rail and Road (ORR), 2019/20. 6

Figure 3. The benefits of applying AI to the risk management process. 11

Figure 4. The components of AI overlapping with safety science. 13

Figure 5. The Research Development framework. 16

Figure 6. The increase in literature in the last decade. 26

Figure 7. The concept of turning smart city technology into smart train stations [14]. 36

Figure 8. In railway stations, the smart city concept model (SCCM) has the potential to be beneficial and applicable [14]. 37

Figure 9. Australian Standard for Risk Management (AS/NZS 4360). 52

Figure 10. The risk management concepts [115] [116]. 52

Figure 11. Risk identification methods [119][120]. 55

Figure 12. Risk analysis techniques [126]. 57

Figure 13 Risk assessment Life cycle (The Hourglass Model) [127]. 59

Figure 14. Station operations accidents 2019/20 in the U.K. [130]. 61

Figure 15. Areas of non-workforce slips, trips and falls in stations (2019/20)[131]. 62

Figure 16. Simple fault tree. 63

Figure 17. Suggested grouping of risks in railway stations..... 64

Figure 18. Reasons for South Korean train fire. 66

Figure 19. FTA of the fire event in a railway station. 68

Figure 20. FTA of the evacuation failure in a railway station. 69

Figure 21. FTA of the protection system failure in a railway station. 70

Figure 22. FTA of the detection system failure in a railway station. 71

Figure 23. FTA of the water supply system failure in a railway station. 71

Figure 24. Slips/trips formed one of the main causes of injury (Office of Rail and Road (ORR), 2017) [149]. 73

Figure 25. FTA of the slip/trip risk in a railway station. 73

Figure 26. FTA of the train derailment risk in a railway station. 75

Figure 27. FTA of the train rear-end collision accidents [157]. 76

Figure 28. FTA for the manual intervention fail sup tree of derailment. 77

Figure 29. FTA for the Collision avoidance system failure sup tree of derailment. 78

Figure 30. FTA of the train security system failure in a railway station. 80

Figure 31. Asset groups in the stations which are vulnerable to terrorist attacks. 83

Figure 32. FTA of the overcrowding risk in a railway station. 86

Figure 33. FTA of the flooding risk in a railway station. 89

Figure 34. FTA of the earthquake risk in a railway station. 92

Figure 35. FTA of the dangerous goods affect risk in a railway station [200]. 94

Figure 36. FTA of the structure failure risk in a railway station. 97

Figure 37. The correlation between the risks in the stations. 99

Figure 38. Railway station risk framework using a complex group of AI algorithms developed for the study.

..... 106

Figure 39. Examples of AI techniques details. 108

Figure 40. Implementation framework layers of AI in risk management railway stations. 109

Figure 41. Designed ML flow for the safety mechanism in the stations. 111

Figure 42. Splitting algorithm of CART. 115

Figure 43. The simple tree shows the example distribution. 116

Figure 44. The process of single neuron with ReLU Function. 118

Figure 45. The style of a neural network with two hidden layers. 119

Figure 46. The components of a Fuzzy Logic System. 120

Figure 47. Fuzzy Inference Process with two inputs, three rule and one output. 121

Figure 48. ANFIS Structure with 4-rule, 2- input scheme. 122

Figure 49. The triangular and generalized bell mf. 122

Figure 50. Proposed process framework for the Intelligent System for Managing Risks (ISFMR) In railway stations. 123

Figure 51. Concept diagram of k-means clustering. 124

Figure 52. The typical formation of PLSA. 126

Figure 53. Description of graphical representation of LDA. 127

Figure 54. Structure of deep learning. 128

Figure 55. An illustration of a convolutional layer process using a 3 x 3 x 1 Kernel (K). 129

Figure 56. An innovative paradigm for using a CNN to control risks in the railway system. 130

Figure 57. The tree-map presents applications of CNN in railway stations with different risk sources. 131

Figure 58. ML Algorithms hierarchy mapping. 131

Figure 59.Some of the AI current cases of the systems and powered by ML. 134

Figure 60. An alluvial diagram of primary examples of accident data shows the accident characteristics, consequence, date, time, and victim age. 136

Figure 61. Scatterplot representing the relationship between the characteristics of the accidents and the passenger's age, as well as the day of the week. 138
Figure 62. Scatterplot representing the relationship between the characteristics of the accidents and the passenger age with the time. 138
Figure 63. The prediction and data distribution from the training dataset are depicted in a decision tree diagram and graphs. 139
Figure 64. The nodes and histograms of the 8-cluster diagram. 140
Figure 65. The rules and the decision tree with an example of a particular node. 142
Figure 66. On the prediction path, comparisons of the input field with several factors (time and the details of the accidents). During PM, the result is displayed for the example accident type T3-S. 143
Figure 67. Comparison of the input fields on day 7 with different parameters (time/weekday). 143
Figure 68. The stages of splitting the dataset. 146
Figure 69. In the confusion matrix, the evaluation outcomes of each class's performance. 147
Figure 70. Graph of two-dimensional parallel coordinates. 148
Figure 71. The plot of the cumulative distribution function. 149
Figure 72. The ROC curve and the area under the curve (AUC) is 0.90. 149
Figure 73. The proposed CNN approach for the next generation of railway station risk management scheme 152
Figure 74. The convergence of FSTs and system aspects. 153
Figure 75. (a) Dataset of passengers who fall/slip or trip in stations (b) Dataset of passengers' images who are almost in a normal situation at stations 156
Figure 76. Different traveller behaviour samples 157
Figure 77 The accuracy of the trained model and the model loss. 158
Figure 78 Examples of passenger action clusters predictions 160
Figure 79 ROC curve of the model 161

Figure 80. The Back Propagation (BP) Algorithm. 164
Figure 81. The fuzzy logic inference method framework for risk control in railway stations. 165
Figure 82. An ideal platform is represented in 3D, with simple train movement and timelines. 166
Figure 83. The relations among the inputs and the risk amount. 171
Figure 84. The findings of a latent class analysis using 2 segments and 25 samples. 171
Figure 85. The expected and actual values for different types of regression learner optimizable models: (a) prediction (Optimizable SVM), (b) prediction (Optimizable Tree), (c) prediction (Optimizable GPR), and (d) prediction (Optimizable GPR) (Optimizable Ensemble). 172
Figure 86. Expected and actual values for various regression learner optimizable models: (a) prediction (Optimizable SVM), (b) prediction (Optimizable Tree), (c) prediction (Optimizable GPR), and (d) prediction (Optimizable GPR) (Optimizable Ensemble). 173
Figure 87. (a) ANFIS model architecture for overcrowding risk level estimation, with the branches indicating the ANFIS model structure in addition to the feed-forward network, and (b) ANFIS network after subtractive clustering. 174
Figure 88. The association between the inputs and the risk levels is depicted as a heat map. 175
Figure 89. Main ANFIS of indexes and risk performance (Matlab's rule viewer) were trained. 176
Figure 90. In the proposed method, the Sugeno-type FIS method is used. 176
Figure 91. The progression of trained fuzzy system errors. 178
Figure 92. Training dataset and FIS output with two inputs and one output. 178
Figure 93. Example of the rule base system's 3D surface plots adapted for data for various mf types: (a) dsigmf, (b) gbellmf. 179
Figure 94. Comparison of the proposed model's predicted and spotted risk levels. 180
Figure 95. The framework of the topic model in risk management in the stations. 184
Figure 96. Data mining cleaning procedure steps (Text Cleaning and Pre-processing). 186
Figure 97. Example of accidents and their detailed attribute correlations. 188

Figure 98. Comparisons of input subjects for Multiple features (the times of day and the time limit, including the age group). 190
Figure 99. Comparisons of input subjects for different factors (the times of day and weekday, as well as the age group) 191
Figure 100. Co-occurrence network. 192
Figure 101. Topic model map with ten number of topics. 193
Figure 102. The topic model result map shows ten topics with examples of the words cloud and the probabilities. 193
Figure 103. Words map example 194
Figure 104. Words map example 194
Figure 105. Example of topic distribution for different context contains different words. 195
Figure 106. 8 cluster maps with examples of word maps and histograms. 197

LIST OF TABLES

Table 1.List and plan of publications by the researcher. 19
Table 2. Reviewed and summarized literature. 24
Table 3. Selected Examples show terms and definitions from different resources [96][97][98][99][100][101]. 47
Table 4. Comparison between quantitative and qualitative risk assessments. 56
Table 5. List of unwanted events in the stations [128] [129]. 60
Table 6. Example lists of the fires in the stations and number of deaths and injuries. 66
Table 7. List of stations that faced terrorism attacks around the world from 2000 to 2016 [165] [166]. 81
Table 8. List of stations that faced overcrowding accidents around the world from 1999 to 2018 [174]. 85
Table 9. Stations flooded around the world from 2001 to 2018 [7]. 90
Table 10. Cases of railway stations affected by earthquake [7][6]. 92
Table 11. Example of calculation presents the Impurity Measures. 116

Table 12. The analysis of the dataset. 139

Table 13. The details of the clusters..... 141

Table 14. The comparisons of input fields on the prediction path with fixed age= 60. 144

Table 15. The DT prediction path example, where the target in this example is the accident type T1-F 145

Table 16. The Keras library packages example installed detailed results after running the model. 154

Table 17. Detailed outcome of the experiment's models 159

Table 18. As a confusion matrix, show the classification performance evaluation results. 161

Table 19. Lu and Han present a case study of measured data on the station platform during peak hours (2016). 168

Table 20. The magnitude of the risk likelihood level..... 169

Table 21. The two indices with the risk matrix and scale. 170

Table 22. Error tests based on a mix of hyperparameter optimization models. 172

Table 23. RMSE results from the training and training datasets for various MFs with Membership Functions Number (4*4). 177

Table 24. The examples from dataset inputs word cloud visualisation. 188

Table 25. Set up configurations for text analytics. 193

Table 26.Topic batch with topic distribution examples. 196

Table 27. The probabilistic description example of the words among the topics. 196

Table 28.The probabilistic description example of the words among the topics. 197

Table 29.Cluster outcomes of the textual data. 198

Table 30. Text example for extracting information from trained topic model. 199

Table 31. Classifier of overall statistics and confusion matrix..... 199

LIST OF ABBREVIATIONS

AI	Artificial intelligence
ANN	Artificial neural network
AS	Australian standard
CBA/RBA	Cost benefit analysis/risk benefit analysis
FTA	Fault tree analysis
HRA	Human reliability analysis
ETA	Event tree analysis
RSSB	Rail Safety and Standards Board
CCTV	Closed-circuit television
SA	Sensibility analysis
FMEA	Failure mode and effect analysis
HAZOP	Hazard and operability study
PTI	The platform-train interface
ML	Machine learning
MLT	Machine learning tool
SMIS	Safety Management Information System
NN	Neural network
ORR	Office of Rail and Road
ANFIS	Adaptive Network-based Fuzzy Inference system
FIS	Fuzzy Logic Inference System
FWI	Fatalities Weighted Injuries
CV	Computer vision
SOM	Self-organising maps
MLP	Multi-layer perceptron
SVM	Support vector machines
MF	Membership function
CCR	Central Control Room
DL	Deep Learning
CNN	Conventional Neural Network
ReLU	Rectified linear unit
kernel	Central Part of An Operating System
BNs	Bayesian network

DM	Delphi method
BD	Big data
PLSA	Probabilistic Latent Semantic Analysis
PAS	Publicly Available Specification
SRA	Society for Risk Analysis
DT	Decision Tree
GDPR	General Data Protection Regulation

CHAPTER 1: OUTLINE AND RESEARCH CONTRIBUTION

1.1 Introduction

A railway system remains a fundamental component of the economy and the first choice for cities in many countries for movement of passengers and the goods internally or across borders. It is also one of the most environmentally friendly systems and minimises traffic because of its enormous capacity to transport millions of passengers and goods with high energy efficiency. Railway stations are a critical part of the rail system, which is an essential type of public transportation network that plays a significant role in people's lives and is required for economic activity and urban growth. Within the UK rail network, there are over 2,500 stations, which support more than a billion trips every year, and there are about 30, 000 train stations in the EU-28 Member States [1]. However, stations have an open boundary structure with the railway line, and it has been shown that many people consider travel by public transport to be somewhat risky and are therefore reluctant to use it. This fear is driven to an extent by factors such as media coverage of relevant accidents and crime rates and regular safety messages that indicate risk rather than safety [2]. Many railway stations are also shopping areas providing food and other services. This carries and creates risks, which need to be managed to assist in keeping passengers, the workforce and the public safe in such environments [3]. Many risks have different data sources and overlap with other risks, which form various responses, relationships and complex contexts. Moreover, large numbers of passengers make some stations "crowded areas", mainly in peak travel times. Some stations are located in critical areas and/or at an underground level where there are limited, narrow areas and where there is the possibility of natural disasters such as floods or earthquakes or other unwanted events like fire or terrorist attacks. All these events can have

a dangerous or disastrous impact on the structure and the people, and consequently on any management, evacuation and rescue effort. The capacity in some stations has outperformed the designed capacity, and when emergencies, accidents, or unwanted events occur, the consequences can be horrific [4][5]. The Network Rail and British Transport Police indicate that there were more than 8,000 incidents in the year of 2017 where people risked their lives at crossings on the rail network—an 11% increase over the previous year.

Moreover, most accidents at stations are due to slips, trips, and falls on stairways, concourses, and platforms [6]. The accidents include not only employees but also a significant number of children, elderly and other members of the public. It is essential to distinguish the characteristics of risks that might be faced in stations and ensure that risk management is sufficient [7]. Operation and management of safety hazards for such an infrastructure is essential, where critical threats can lead to severe injuries or even fatalities for passengers, public, or staff and bring with them economic and social effects. It has been demonstrated that railway stations have characteristics of complex correlations, and the operation and development of stations require integration of multiple entities [8]. It is essential that analysis and management of the risks, the causes, and consequences of potential hazards in stations are considered, including those that arise from interactions between operations and many internal and external factors.

Consequently, managing risk is integral to moving towards continually improving security, quality, and safety in the railway system. Preventing undesirable consequences with focus on risk-based thinking is part of the quality management process (ISO 9001:2015) where risks affect both quality and safety. The ISO standards provide assistance for understanding railway risk management and the assessment techniques as generic documents (ISO31000 / ISO 31010) However, the risk must be identified and captured, also not cover the new technology tools. Moreover, by implementing standards in Railway risk management, the process is still managed by practical reasoning both

naturally and socially depending on experience [9].Risk management and increasing safety levels of the railway are crucial for all, including researchers, operators, government, and society. To conduct practical safety actions and to consider any future expected dangerous occurrences, it is essential first to investigate the risks to gain a full understanding of the dangers that might occur in stations. Then, a framework should be designed to and improve the risk management process to cover the factors that might lead to these critical threats, to offer a modern method of enhancing safety.

1.2 Research objectives and scope

The aim is to apply new technology (AI) for risk management in the railway station. This will lead to a better understanding of AI applications and an examination of their implications for risk management in train stations. In all industries, including risk management, there is currently a rising market in AI investment. Risk management is considered an analytic plan to protect the passengers, staff, public, and assets for the rail industry. This facilitates business continuity with high performance and quality, and prioritisation of resources in terms of safety and security with the ultimate purpose of reducing or mitigating the threat. Many uncertainties and risks correlated with stations have been analysed, and then proposed as the basis for introducing a novel smart approach to risk management in railway stations. From the literature, many different risk assessment methods have been produced to perform specific applications and scenarios. Most of the state-of-the-art risk management methodologies are focused on traditional methods, which are passive systems that do not include technology such as AI. An AI approach will be utilised here to automate the process, because it can monitor in real-time and can consider many different risks at the same time, thus supporting decision-making. Therefore, we aim to improve risk management approaches by utilising AI methods. The scope of the study is to use AI for risk management in the station to explore how we can proactively avoid risks when using the AI. From this, the issues and any challenges of such implementation will be better understood. The nature of the suggested method is a part of computer science, but the

comparisons or improving the algorithms are beyond the research scope. These were the motivations for developing this approach to risk management in this research. The objectives of the research are as follows:

- Exploring risk management methodologies and the ability to apply new methods like AI and any current associated vulnerabilities.
- Testing the use of AI with the accident history for prediction and in-depth analysis.
- Examining the application of AI to identify the risks in a timely manner.

These objectives are to develop a risk management process in order to be smart through identifying the risks, and reflecting on enhancing performance, reliability, health and safety, and punctuality in stations, which will further reflect on the whole industry. Moreover, it is hoped that risk management can be integrated into stations with other sub-systems in the railway to deliver a sustainable digitalised future for the railway industry. To sum up, the following question will be addressed: How do real-time big data and AI technologies promote risk management in railway stations.

1.3 Research problem

The problem has been considered from two angles: the station as critical infrastructure, the vulnerable point of the system, and the risk management process to ensure safe stations. The intensive demand for travelling via rail together with the location create a pressure for station safety; many stations are in the heart of the cities or connected with other critical services such as airports, stadiums, malls, and other public transportation. The intensive usage results in more dynamic operations and raises the importance of risk management in such places. Furthermore, to cope with demand, the capacity of trains is increased, and urban areas have also been expanded.

However, there is not much consideration in stations where some classified as an old infrastructure built a long time ago is not fit with current capacity and has safety security issues, such as narrow

corridors. Also, some stations have many floors with basements and are linked to other transportation services like metro, high-speed rail, malls. Additionally, intensive usage results in more dynamic operations and rise the importance of risk management in such case. The risk management process often depends on experts, and it is time-consuming, costly, and not modernised to capture the risks in a timely manner nor to suit the dynamic environment. Thus, the problem has two aspects, namely the station as a vulnerable critical infrastructure, and making the process (risk management) more effective, smart and benefits from technology, Figure 1 summarises the points that support both sides of the issue (risk management process and the railway station) gathered via the literature [10][11]. Then, proposed improvements or the solution have been planned.



Figure 1. The two views of the research problem.

1.3.1 The station

The nature of stations, which are dynamic, open environments, is a point on the railway line located in significant locations within cities, which give rise to several different possible threats. Risks such as a fall, slip or trip on the railway station is expected. The consequences can be quite serious for older people, families or disabled people. They can also be costly for the industry as the compensation may need to cover a long period of treatment. Furthermore, such accidents lead to panic in the station

environment, causing delayed services and impacting the business reputation. Many people lost their lives in the stations through accidents such as falling in gaps, touching moving trains, or the electric power. Indeed, time is critical in detecting the risks and responding to save lives.

The railway passenger station is a critical point of the network which interfaces with the line, and it is the start and end point of all journeys. Station operations involve many systems that are connected and interdependent. In fact, the location and nature of the station form a critical infrastructure. Many factors must be managed, which can push the limits of human ability in vast stations at peak times with all the potential uncertainties and expected risks. Moreover, as demand for rail travel increases, increased use of the structure will require the addition of more trains on the line. Rail passenger journeys in the UK have more than doubled in the last 20 years (see Figure 2). Consequently, many countries have experienced significant pressures to increase operations and safety in the industry and increase numbers of passengers. This intensive use of the railways raises the probability of accidents, including in stations.

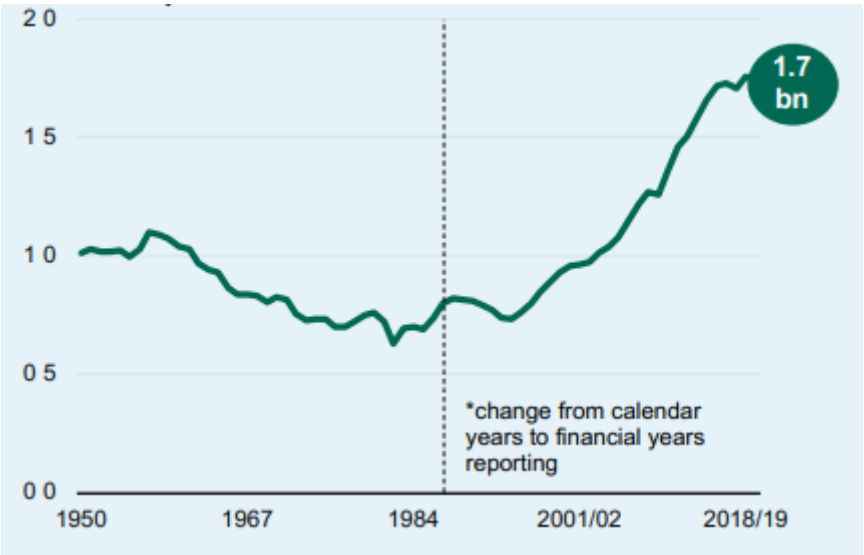


Figure 2. The total number of passenger journeys trend in the UK (in billions) (Office of Rail and Road (ORR), 2019/20.

Any increase in travellers' quality and service requirements, the movement of trains, the flow of passengers, maintenance tasks, human errors, changing weather conditions, human behaviour, infrastructure, together with overlapping tracks and trains and many other sub-systems within the station could all contribute to a massive production of data each minute. Yet, this is all beyond the human ability to analyse and find a potential failure in real-time, and to understand correlations and causation that can be processed quickly to support the decision-maker. Moreover, nowadays with the COVID-19 pandemic, managing the crowd is essential for health and safety more than at any time before. It requires huge efforts and application of technologies such as AI to support and provide safe stations. COVID-19 is a lesson that must be learned for strategic safety plans and future improvement in the stations.

1.3.2 The process (risk management)

There are many reasons why station risk management by experts is problematic, for example, the experts are not available all of the time when they are needed, and they have subjective opinions. From the beginning, it is a deliberate, time consuming process, which requires frequently updated and costly processes. Moreover, the procedure is usually not reliable because it cannot be reproduced, casting uncertainty over the efficacy of the decisions and the accuracy of results. Controlling risk depends on a vast number of factors that ought to be handled consistently and in real-time. Improving safety, security, and risk management, and implementing dynamic monitoring and prevention systems has challenges with the present management methods. Despite the decrease in risk in some periods or some countries, it is unlikely that a traditional way of controlling risks will identify the sheer volume of threats and the possible relevant factors promptly. In prevailing management systems, distribution of the workforce is based on particular career responsibilities; for instance, the station service department and the facilities maintenance staff are generally separated. Likewise, facilities maintenance work is divided into minor professional units (such as rail track, power supply,

communication, fare collection system, and so on). This type of separation is advantageous in the early stages of operations. Nevertheless, processes demand greater collaboration amongst several methods, a fast response rate, broader management regions, and a more efficient and comprehensive management system operating in real-time with a minimum of human errors. Besides, the analysis of the State-of-the-Art in Railway Risk Management emphasises that additional development is required where the amount of time and expertise needed from railway managers to make decisions have been considered. This means reflecting on making decisions, saving time and obtaining better estimates [12].

1.4 Research contribution

More broadly, lifestyle changes and the future of the Cloud and the Internet of Things (IoT) together with smart cities are challenging the industry and require greater steps to manage risks and safety in stations. In fact, in the railway sector, there is a variety and large volume of data sources that are not efficiently invested in, which leads to incomprehensive risk control performance and low usage of prediction rate for the risks [13]. Moreover, for passenger safety, monitoring, securing and decision-making processes to achieve more cost-effective management in railway stations, future smart stations utilising new technology such as wireless networks and development of integrated systems have been proposed as a concept. The analysis of such wireless sensor network (WSNs) applications has shown that such an enormous amount of data from a railway station is new and novel. This concept requires the development of AI methods, such as ML and DL; this is crucial for the railway industry's future and provides the motivation for this research [14]. Increasing technology development and the complexity of the operations demand growth on the railway where currently, extreme weather and other challenges show that the traditional risk process may no longer be sufficient. Thus, dynamic risk management becomes an urgent need and the basis for the future of risk management in railway stations. Also, it has been shown that the safety system in the railway has

uncertainties. These organisation issues should be amenable to effective data analysis methods (Big Data analysis), and there are very complex events with various causes that lead to accidents [15]. In addition, it has been advised that it is worth investing in better usage of data and to use AI applications for analysis and predicting the future for managing the disruption and delays in the railway [16][17][18]. Moreover, there is no deep study of such cases; for example, in China recently, research on such topics shows no systematic risk assessment framework has been created, which has influenced the research and expansion of risk management [19]. Accounting for all these issues suggested that AI will be an effective tool for reducing the global impact of dynamic risks and increasing safety levels in railway stations. All risk management issues are increased if risk management becomes a large, time-sensitive action within critical infrastructure such as railway stations. Moreover, from many discussions with the experts in the area operating in the railway sector, it is remarkable that there are risks of data deficiencies and a lack of a link between the participants who do risk assessments in different departments and safety risk factors. This leads to increased uncertainties and costs and takes up a lot of time. Furthermore, in conventional operations, the decision-maker or the safety managers obtain the information via a paper report which takes time for preparation, or through emails or telephone. In fact, obtaining information in real-time or at least in a short time is challenging, especially for complex dynamic systems such as railway stations. Thus, applying AI in risk management will provide information more quickly and in digital form. Thus, the AI is targeted as a solution, which can cover a variety of inputs and records from predictions in less time and in diverse sections in real-time, as would be the case in railway stations. In this research, it has been suggested that risk management issues can be overcome by applying AI. Utilising AI in risk management will advance the safety, security, and performance of the operation, which will reflect positively on the whole industry. Despite the tools or software employed in the field, which focus on the risk register, historical recording, and reporting, the influence of AI in the risk management

process will produce the automation process, so the risk management method in the stations would benefit greatly from automation. The number of recorded accidents in the station, an increase of the demand, passenger lifestyle, and the movement of technology point towards the necessity for meaningful, novel and active processes rather than a common risk management process. Currently, any disruption or small operational error will introduce widely felt negative impacts and can attract a great deal of public attention if the issue is not properly addressed in a timely manner. The solution for safety and risk management with the challenges to railway station operations can only be resolved by promoting different views and novel actions, such as implementing AI, which may be an unavoidable phase in the development of the rail industry (see Figure 3).

Such a subject fills the gap and presents the work's originality, while the AI method in station risk management is not applied yet. Indeed, this provides a unique novel concept to digitalising the process and coping with future changes. Finally, AI will be used in the methodology of this research, which aims to create a process with the ability to support complicated intellectual jobs (risk management in the railway stations) or exceeding a human level of execution, which is indeed the most common goal of AI.

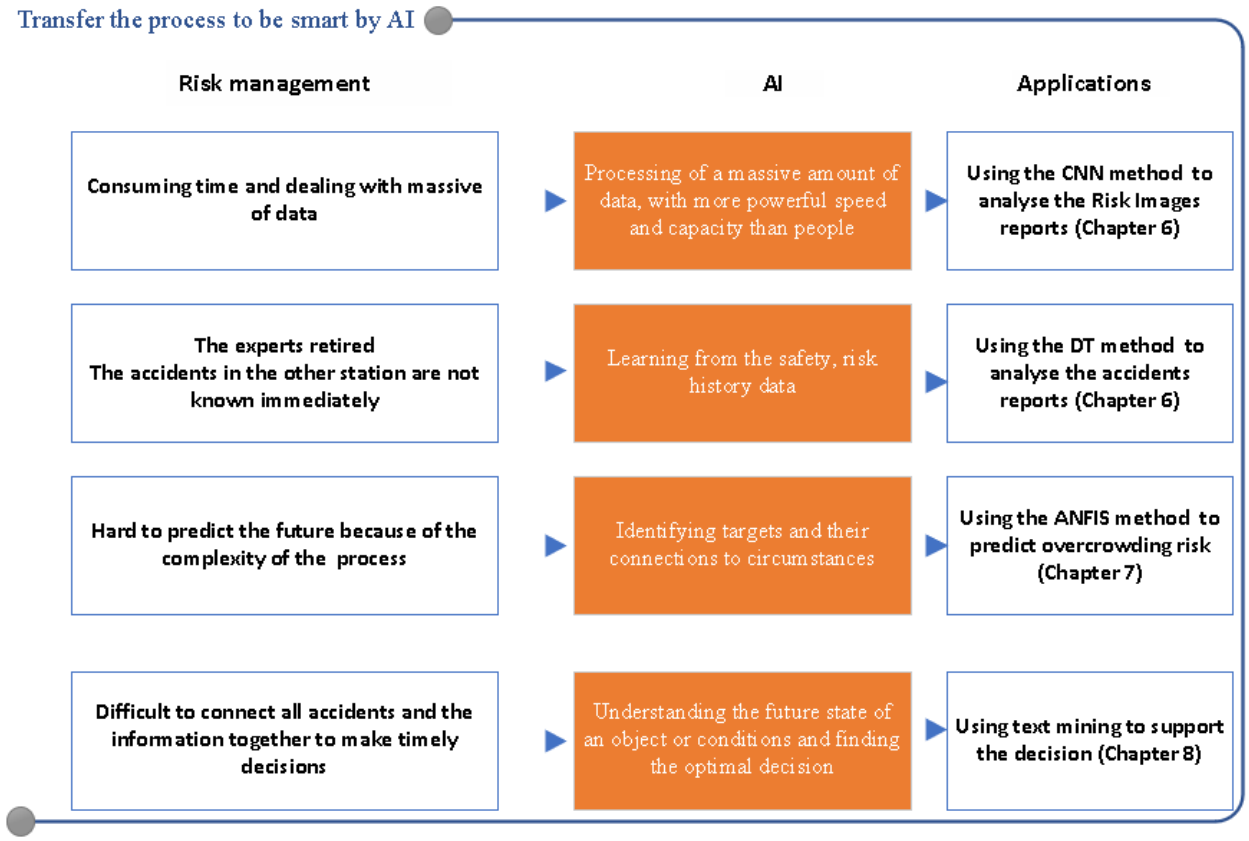


Figure 3. The benefits of applying AI to the risk management process.

1.5 Novelty of the research

The method of AI will lead to new risk management expertise inventions in railway stations. Utilising smart technology such as AI has not yet been thoroughly examined in the literature, indicating that this proposed research is novel. Moreover, changing the process from the conventional to the automated, including less human intervention and capturing knowledge from one generation to the next, is a novel development in safety science. Also, such improvement is supporting the strategic railway plans digitalisation of the industry. However, there may be a shortage of references to support the rationale for the study. Also, it has been shown that BDA theory in the railway has not yet been fully integrated. The application of AI has been proven in other systems such as banking or marketing but is not yet mature for the railways.

Digitalisation has proven to reduce human error, as well as lowering cost in the long term and saving time, which impacts the sustainability and resilience of the systems. This presents the importance and novelty of the proposed application in railway stations.

Moreover, from the IoT and smart city perspectives, this method is a forward step for the future, whereby stations are able to integrate with smart cities. Moreover, the unique value of this study emerges in the following points:

- Risk management modelling of railway stations by utilising AI technologies will advance operational efficiency, reducing downtime, and improving detection capabilities and increasing outcomes.
- AI in railway stations will accommodate accurate and robust obstacle detection and classification and alerts of a variety of safety threats.
- The risk data with AI will analyse behaviour of the people and enhance the efficacy of the stations, risk management process, decision-making and safe operations.
- The power of AI analysis on risk history or safety data will support significant predicting deficiencies in stations.
- This technology will be expected to make the station environment smart, safer and reliable and boost decision-making certainty.
- Applying an innovative technology with the tremendous amounts of risk data and records will enhance and develop risk management science in general and in the railway station specifically as a critical aspect of railway management.

AI and its components have overlapped with data science and computer science. Yet, in our research, a third horizon is adapted to risk management which is a part of safety science with the AI, which

contains ML as a subset of AI and DL as a subset of ML (see Figure 4). Finally, the novel application and approach to risk analysis hopes to highlight how to prevent risk, save people's lives and keep railway stations a safe and secure environment. Indeed, the growth of AI in many aspects of our life and trying to save humans from harm are motivations to apply AI in risk management in railway stations. On this ground, this research is unprecedented and highly original. Finally, the literature review chapter will present how this research differs from other related work.

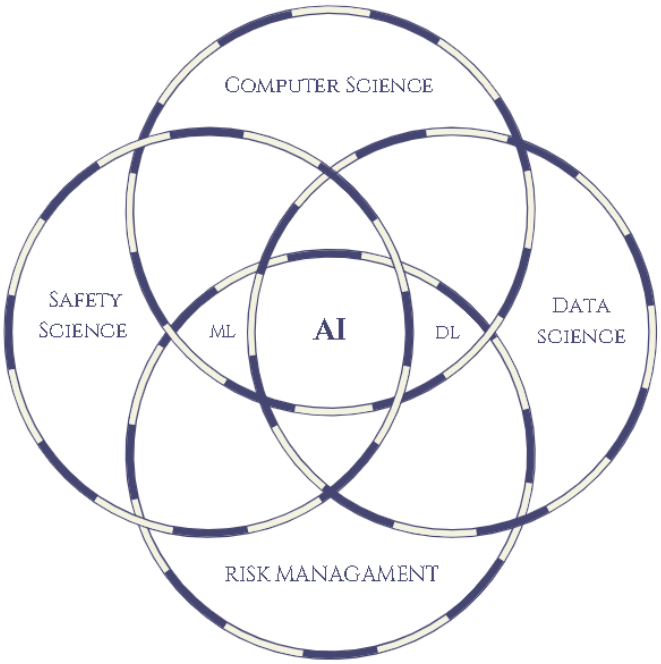


Figure 4. The components of AI overlapping with safety science.

1.6 The methodology

An integrated analysis-design approach is taken to overcome the research problem and for exploring the research subjects. The literature is considered as a way to find the possible AI applications for risks in railway stations research, and the findings in the field have been analysed. Moreover, the FTA has been utilised to analyse the risk in the stations as a selected part of the system. Also, the case studies have been conducted utilising both supervised and unsupervised learning, covering historical

data including pictures, numeric and textual data. The techniques that have been used include DL with CNN, topic analytics, DT, and the ANFIS system. The methods show novelty in terms of the selected analytics for the data sets to develop risk management. Utilising the live images from CCTV to capture the risks in a timely manner is original and opens up the possibility of investing in image data and swapping from passive to active systems to save lives. Moreover, controlling the crowd via an automated process is certainly needed. It will solve many issues in the industry. Considering safety history is great input for the AI, and this will heavily depend on the way of recording, the documentation and investigation into the accidents, as well as the safety policies, which will make the proposed methods very effective.

1.6.1 Machine learning

Supervised and unsupervised ML have been examined, which provide a reliable and beneficial analysis of the history of safety records. In supervised learning, the ML uses data that is labelled (labelled inputs), whereas in unsupervised learning, ML discovers hidden patterns with unlabelled data sets. The applicability of the decision tree to this safety task for railway stations has been examined. Moreover, the ML presents conceivably beneficial classifications and prediction performances and interpretable accident details. The method illustrates the model's validity as well as the data's distributed analysis. It was also used to determine the relevance and significance of the accident conditions chosen. This method has a high and reasonable prediction accuracy as well as a high and acceptable performance. To process safety data in real-time, it is suggested that some of the systems in railway stations be integrated and that automated data gathering be used to extract the most valuable data from a large number of indicators.

1.6.2 Deep learning

Deep learning has been presented as a method for transitioning from traditional risk management to risk management using novel mechanisms; this technique can enhance safety and security across the entire railway industry paradigm. Deep learning in risk management has numerous advantages, including the following [20][21]:

- To better avoid risks, a real-time ability can be gained. Many field subsystems, including maintenance, security, traffic, doors, escalators, and passenger data, can be integrated to produce actions that examine many elements during the operations to assure safeness.
- Machine learning can effectively integrate lessons and experiences into the learning process, which is fundamental in the culture of safety systems (lessons learned).
- It is essential to improve the efficiency of operations in stations, which positively impacts areas related to railway activities.
- Long-term quality improvements can be achieved by reducing the cost and time while enhancing reliability.
- Both passenger and workforce experiences can be improved, which reflect on the overall market image.
- Data gathering can be enhanced to more fully utilise effective connections between assets and people.

The findings of the proposed novel model show that increasing the depth of a deep network can improve accuracy. The use of approaches like computer vision techniques will increase risk detection, enhance safety in railway systems, and ultimately affect the risk management process dramatically.

1.6.3 The study arrangement and interconnection

To fulfil dynamic risk management and increase the knowledge in the railway stations, it considers the analysis of risk factors which is beneficial for applying and exploring the methods, as well as for reaching the objectives of the research. In this study, it explores the knowledge about AI methods, including the literature review and the cases studies with different types of data. From the literature and trialling of the methods, it has been shown that some applications in risk and safety are limited and not applicable for all the process problems. Some challenges emerge because of the unavailability of the data or because of privacy issues and the limitation of the outcomes. For a more comprehensive understanding of the risks, the dangers have been outlined utilising the FTA. Then, the suitable methods of the machine and deep learning are applied to the risks in the stations, and the general phases in the plan of the research are as shown in Figure 5.

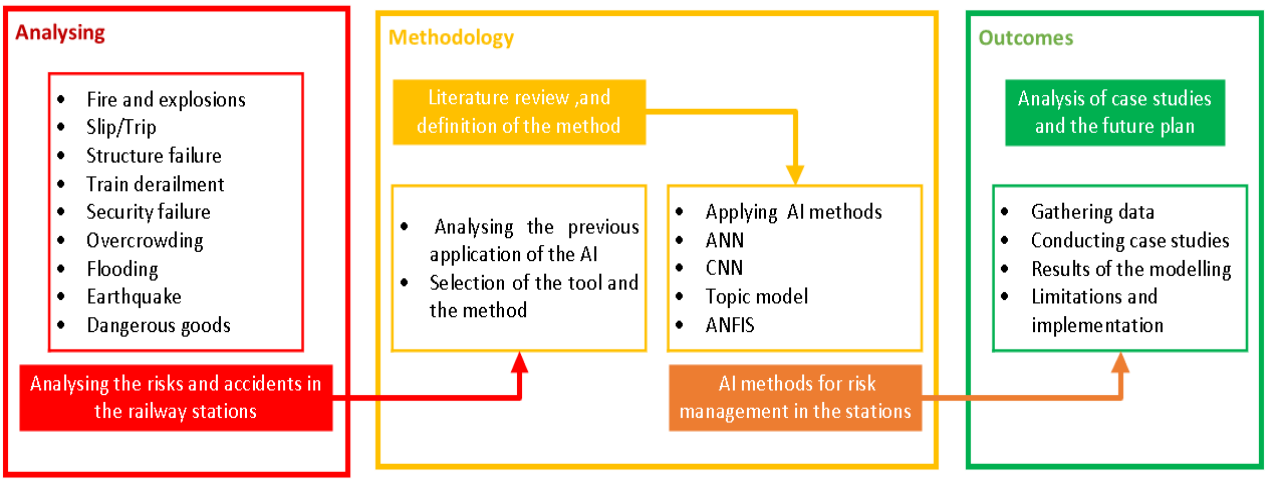


Figure 5. The Research Development framework.

The ML and DL such as DT, also ANFIS and CNN have been applied as a unique approach to be applied in the field and to catch the attention of researchers from subjects such as computer science for more improvement in the future. The CNN is a fully connected neural network mainly utilised in image analysis tasks like image recognition, object detection and segmentation. In this work the risk of falling in the stations has been considered, to be detected from images using CNN images that have

high dimensionality. CNN is very effective in reducing the number of parameters without reducing the quality of models. The diverse methods cope with different datasets, including numerical data from accidents, risk images, and text from safety history data. That variety of technology can capture risk from multi-sources in the future, such as signals from sensors, images from cameras and text from safety and maintenance reports, or social media. The application of AI will provide timely support for decision-makers and improve the safety of the whole transportation system by utilising historical or ongoing real-time data. The use of the proposed, new method can develop the extraction, classification, prediction, and detection of risk better than humans.

1.6.4 Research datasets

The accessibility of data in many fields in an appropriate form has been accelerating and facilitates the implementation of AI while encouraging researchers, in banking and the marketing industry, for example. However, this is not the situation in some other fields, where the data are collected only for statistical analysis or are not available for research. In this work, different case studies are conducted which required exploration of a different form of data. In machine learning, numerical and categorical variables are gathered from the real accidents that occurred in the stations and led to deaths. The designed model is based on data collected from rail reports (RSSB) since 2002. The collected data on accidents that have been reported and published represent a selection of 80 incidents at stations in the UK that have been, or are subject to, an investigation by the UK's national investigation body: the Rail Accident Investigation Branch (RAIB) [22]. Moreover, pictures are used in deep learning, where the input image data of passengers are used and classified into one of two classes: either 'falling' or 'not falling'. The datasets are divided into separate frames with known labels (falling or not falling), which were then used as training data for the classifier. To construct the dataset, different activities and complex falling events from different locations were selected from many open-source sequences, such as falls on stairs or in the gaps between the train and platform. Data are gathered consisting of

both still frames and videos from open sources. Finding such data is challenging both because of privacy concerns and lack of availability for many reasons, for instance, that data is deleted from data centres periodically [20]. Additionally, in the case of text mining, the documents used as textual data are analysed. From the RSSB, the raw data set has accidents registered in the railway stations in the last ten years. The selected data set containing 163 of the accidents cover those incidents occurring at rail stations between 01/01/2000 and 17/04/2020. Finally, the data are analysed and described in the context of each case in the next chapters of the thesis.

1.7 Research achievements

The research accomplishments and outcomes achieved during the doctoral research are related to improving risk management approaches for railway stations via AI. These achievements cover publications (see Table 1), including peer-reviewed journals published in international conferences and a submitted chapter, which have been integrated into the work and referenced. Moreover, during this research period, many discussions and presentations have taken place in relation to the railway industry and technology. The publications included cover cases studies, ML, DL, ANFIS, topic analytics and information such as the concept, methodology, frameworks, and overview of the context. However, some papers have been submitted but not published yet, and the rest are under review for submission and publication later. The experts who have been reviewing the publications appreciate the concept been proposed as a novel and forward step to improve risk management in railway industry research.

Table 1.List and plan of publications by the researcher.

Title	Status	Corresponding chapters
Alawad, Hamad, and Sakdirat Kaewunruen. "Wireless Sensor Networks: Toward Smarter Railway Stations." *Infrastructures* 3, no. 3 (2018): 24.	Published	2,4
H. Alawad, S. Kaewunruen and M. An, "Learning From Accidents: Machine Learning for Safety at Railway Stations," in IEEE Access, vol. 8, pp. 633-648, 2020, doi: 10.1109/ACCESS.2019.2962072.	Published	2,5,6
H. Alawad, S. Kaewunruen and M. An, "A Deep Learning Approach Towards Railway Safety Risk Assessment," in IEEE Access, doi: 10.1109/ACCESS.2020.2997946.	Published	4,6,8
Alawad, H. A. H., Sakdirat Kaewunruen & Min An, 17 Jun 2019, https://research.birmingham.ac.uk/portal/en/publications/risk-management-prediction-for-overcrowding-in-railway-stations-utilising-adaptive-nero-fuzzy-inference-system-anfis(1dcb985c-8c1b-4fab-959f-6a82e7e34d3d).html Utilizing big data for enhancing passenger safety in railway stations, *The 4th World Multidisciplinary Civil Engineering-Architecture-Urban Planning Symposium*. IOP Publishing Ltd.	Published	3,6,8
Alawad, H. A. H., Sakdirat Kaewunruen & Min An, 17 Jun 2019, Risk management prediction for overcrowding in railway stations utilising Adaptive Nero Fuzzy Inference System (ANFIS),*The 4th World Multidisciplinary Civil Engineering-Architecture-Urban Planning Symposium*. IOP Publishing Ltd.	Published	3,5,7
Alawad, H.A. and Kaewunruen, S., 2021. 5G intelligence underpinning railway safety in the COVID-19 era. *Frontiers in Built Environment*, *7*, p.14.	Published	3,4,9
Alawad, H.A.H., Codru, S. and Kaewunruen, S., 2018, May. Complex-system decision framework for managing risks to rail stations at airports from terrorist threats. In *Proceedings of the 5th International Conference on Road and Rail Infrastructure, Croatia, Zadar* (pp. 17-19).	Published	4
Kaewunruen, S., Alawad, H. and Cotruta, S., 2018. A decision framework for managing the risk of terrorist threats at rail stations interconnected with airports. *Safety*, *4*(3), p. 36.	Published	4
Alawad, H. A. H., Sakdirat Kaewunruen & Min An, 2019, Risk management prediction for overcrowding in railway stations utilising Adaptive Nero Fuzzy Inference System (ANFIS)	Published	3,5,7
Alawad, H.A. and Kaewunruen, S., 2021. Kaewunruen, Sakdirat, Hamad Ali H. Alawad, Takashi Omura, and Motoyasu Saito. "Machine learning aided management of motorway facilities using single vehicle accident data." *SAE International Journal of Transportation Safety* (2021)	Published	5
Kaewunruen, Alawad.(2022) "Management of railway stations exposed to a terrorist threat." Best-Practices Handbook on Rail Infrastructure Resilience, Elsevier, Woodhead Publishing, March 01, 2022, pp.1-210.	Submitted	4,5

1.8 Organisation of the thesis (outline)

This thesis is divided into nine chapters. The first one covers the outline and research contribution, and the other chapters involve the following:

Chapter 2: This chapter provides a **literature review** for the emerging AI in the field. The review shows the concept has not been extensively covered. It also considers the application of the method within the scope of the railway industry.

Chapter 3: This chapter displays the formal risk management methods and tools used in the railway industry, which often have parallels with other fields. It evaluates the new technology in risk management approaches and traditional techniques in the rail station.

Chapter 4: This section discusses a risk analysis in a railway station via traditional fault tree analysis (FTA), which sought to discover the causes in a top-down logical fashion based on knowledge of the history of the accident. All possible man-made or natural calamities are covered and discussed with analyses to set comprehensive outcomes and build a base for implementation of the proposed method.

Chapter 5: This chapter introduces the AI methods, which are proposed as a methodology in the research for improving the risk management in the stations, including supervised and unsupervised, and using techniques like decision tree, Neuro-fuzzy and deep learning.

Chapter 6: In this chapter, two case studies are conducted. ML covers a method for risk management in stations using a decision tree classification for case study 1. Also, in this chapter, the DL case study 2 is applied and discussed as a subset of ML, which depends on employing non-linear algorithms to match data.

Chapter 7: As a part of the methodology in this chapter, fuzzy combined with the neural network (ANFIS) will be analysed as case study 3 for managing the overcrowding in the stations.

Chapter 8: The history of safety documents is often in unstructured textual form and yet, it is essential to utilise history to enhance the process of risk management. Textual data are selected as

input in this chapter as part of case study 4, including analysis and examination of unstructured data and the ability to manage risks using records in text form.

Chapter 9: Finally, this chapter sums up the benefits and primary outcomes of utilising AI for advanced risk management in railway stations, and the outcomes from all chapters, showing the application has high performance, exciting results and reveals the unprecedented need for more research in the future.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The railway business, and particularly stations, requires effective risk management that is tailored to its dynamic condition and covers a wide range of potential failure spots. The rising demand for services from clients is an issue that must be factored into future railway safety enhancement. Furthermore, the advancement of technology must be taken into account in the future development of industry safety and risk management, which is a top priority that aims to save lives, assets, and the environment. The technology is enhancing risk management in some fields such as banking, marketing systems and the finance sector. In this section, possible prospective research related to railway risk management that utilises AI technology is presented. In recent years, there has been a growing interest in the application of AI in various fields, particularly transportation in overall and railways in specific. A recent survey on AI in high-speed rail has been carried out, highlighting present applications such as smart planning, intelligent control, and intelligent maintenance [23]. Many aspects in the literature have applied AI for both machine and deep learning. This includes maintenance, control of traffic signals or infrastructure, traffic forecasting, demand predictions, autonomous trains and driver behaviour, diagnosis of equipment, failures of railway tracks/trains, railroad defects, energy, and the line power inspections or infrastructure etc. However, some aspects have been excluded from the literature, as they have been seen as beyond the research scope, which rather attempts to focus on risk management safety in railway stations. Thus, it presents only the topics that discuss risk and safety, as well as those employing AI as a primary method because the work related to the stations seems to be rare or limited. From this literature also, it extracts the motivations and provides a basis for the ideas studied in the research. Therefore, the applications that rely on the domain of the machine and deep learning are covered. Table 2 summarises the work and

presents the methods, the type of input data source applied and the sort of implementations. All the publications in this part were selected manually from Google Scholar and Web Of Science databases using the keywords “AI”, “machine learning,” “deep learning,”, “railway,” “risk management,” and “railway stations,”. The selected work was published between 2010 and 2020, but it is noticeable that the majority of articles were published after 2017, and it is very evident that the literature emerging from the last two years seems to suggest a continuous and rapid development. This forms a fundamental background in understanding the work of others and builds a path of science in the field towards an expansion of the use of AI in risk management in railway stations. The feature of learning in AI is a significant point that leads to enhancing the safety systems. The data of the risks in the station from the lessons learned and from the experts can be captured, which form a knowledge base to nurture the growth of the proposed AI. This base can provide a quick update for the risk decisions knowledge, thus supporting the decision-maker. The outcomes are updated or learned and lead to steady improvements in the safety in the stations. The methods such as machine learning produce results that are more reliable, intuitive, and accurate than those collected from conventional safety evaluation approaches [24]. In the next parts of this analysis, the literature has been divided into subsections (ML, DL, BD) for a more focused exposition.

Table 2. Reviewed and summarized literature.

AI Methods	Technics (algorithm)	Input Data	Field of railway	Comments
DM [25]	ANN, NB, DT and GA	Train derailment incidents	Train traffic accidents	Computer Engineering Science drives the research
BD [26]	PCA	Air pollutants accumulated in the system	Health and safety	BDA
DM [27]	Apriori Algorithm	Log-data of Metro faults operations	Rail operations safety	Data mining
DAE [28]	PSO	Railway traffic accidents recorded at Shenyang Railway Bureaus, Guangzhou Railway Corporation, and Nanchang Railway Bureau from 1999 to 2013	Traffic accident safety classification	Unsupervised deep Learning neural network
DL [13]	LSTM	Scenario-risk-accident chain ontology (SRAC)	To evaluate the risk source-level based on a context	Integrating the Knowledge data and expert rules, two ontologies conducted
DM [29]	ARA	From 1954 to 2014, data of train accidents that occurred in China	Train accidents	Data mining for analysing characteristics of train accidents
DM [30]	GRI, Apriori and CARMA.	Rail accidents from 1996 to 2005 in Iran	Train accidents, analysing 6500 accidents	Accident analysis for development regulations and association rules
BD [31] [32] [33] [13][34]	VA, graph databases model, ontology-based and NLP	On train data recorders (OTDR) from railways in Britain	Data collected from trains (On Train Data Recording (OTDR), signalling location), Close calls, and accident reports	Big Data Risk Analysis (BDRA)
DM [35]	LDA, RF and GB	Accidents from 2001 to 2012, in the U.S.	Rail accident reports	Text mining
natural language process (NLP)[36]	(M-CNN) and (CRF)	183 Chinese railway accidents	Railway accident reports	Text mining and analysis tools for extracting accident risk factors
BD [37]	DCNN and BI	Records from video cameras	Railway Track	Analysing a type of rail surface defect

HMM [38]	HMM	Experimentally real Scenarios	Video-Analysis and Tracking moving objects	Detecting Hazard at Level Crossings for Railway–Road Safety
DL[39]	SAE-CNN	Bolts images from a moving freight train	Train safety	Automated visual inspection of train parts
BD [40]	R-CNN	(CCTV) cameras can be found throughout railroad yards, bridges, grade crossings, and stations	Trespassing Detection	AI–Aided Automated Detection of Railroad Trespassing
ML[41]	NLP	Investigation reports of railway accidents in the UK	The correlation between causes and failures	Classification and prediction
ML[42]	Ensemble methods utilizing decision trees	Data from Centralized Traffic Control (CTC) and meteorological data, from 373-point machines	Recognising risk factors for railway point machine failure	Early detection of point machines failure
AI [43]	SCG	10,778 incident crime from the system of the Chicago Police Department for January of 2013	Crime accidents analysis	Analyse the crime incidents using spatial clustering methods
AI and ML [44][45]	KA, DT &CBR	Case and examples for explanation	Experience feedback from rail systems	Software Error Effect Analysis (SEEA)
ML and DM [46]	Forest of decision trees and gradient boosting	Moscow Railway network-From library and a database of historical incidents	Railway monitoring systems	Automated intelligent ranking of incidents
DL[47]	CNN	Cameras on the top of an inspection train at night-time in China	An object and crack detection for catenary system maintenance and safety	Automatic visual inspection in the catenary systems of high-speed railways
DL[48]	CNN	Images from a device located under a diagnostic train from seven real railway lines	Online template matching-based classification method for detection	Fastener inspection system (VFIS) for Defective Railway Fasteners
ML[49]	BN	Level crossing accidents in France from 1974 to 2014	Level crossing safety	Risk analysis for collision accidents
ML [50]	MLMVN	Time series data	Feedforward NNs, for reliability and degradation prediction	Predicting the level of degradation of railway track turnouts

2.2 Machine learning (ML) literature

The literature is presented but only that which met certain criteria has been included. The conditions that have been considered in the selected literature are:

- The authenticity of the publication
- Written in English
- Within the last 10 years period
- Utilising AI (ML, DL) for safety and risk management.
- The context is situated within the railway sector (see Figure 6).

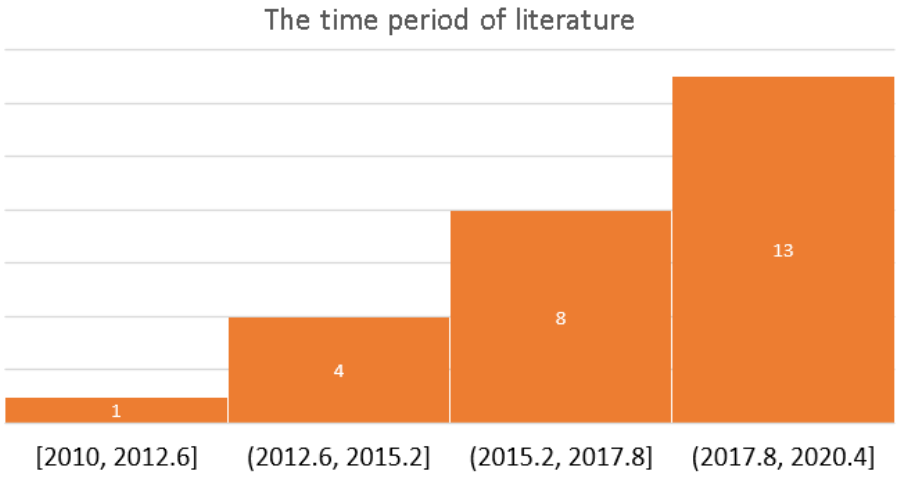


Figure 6. The increase in literature in the last decade.

To enhance safety, ML is suggested in the field as a tool fed by live and past data of safety in the railway to minimize the risks and get a further understanding of the situation and, to analyse relationships and predict railway accidents in the future. This enhances a fundamental safety principle, which is that lessons must be learned.

From a recent structured review, the classical engineering risk assessment has been improved and changed as AI methods are beginning to assist and promote the outcomes [51]. Many accident data

have been used in the literature, such as traffic and train accidents [28] [29], derailment accidents [25], as well as the faults and near miss incidents [27]. From this review, the spreading of methods is noted, where some scholars utilise terms such as data mining [29] [30], deep mining [27] and BD [31][32][26]. A Big Data perspective has been adopted to explore and flag up the method for risk analysis, and three accidents have been analysed. Even though the study only investigates three accidents via expert judgment ratings and proposing a framework for the BD, the investigation proves that BD is a potential methodology for safety and risk analytics [15]. Moreover, the uses of risk management or assessment are interrelated with the safety field and create varied contexts since a lack of descriptions and standards terms are used, requiring thus a strong unifying scientific platform [52] [53]. Unsupervised ML has been proposed, and four classifications of railway traffic accidents have been used (extremely serious, serious, large, and general). The indexes of the death toll, the number of serious injuries, and the direct economic loss were selected as the main indexes for accident ranking [28]. From a data mining view, for railway derailment accident classification algorithms of Artificial Neural Networks, Naïve Bays, and Decision Tree as an ensemble combination are proposed [25]. For predicting accident severity, an ensemble Text Mining methods model (Latent Dirichlet Allocation (LDA)) has been devised for the rail accident reports to analyse the factors contributing to rail accidents [35]. Association rule analysis (ARA) has further been applied to understand the features of train accidents in China from 1954 to 2014. This data mining shows that human error is the most frequent cause of accidents as there are more losses from collisions than derailment. Similarly, other results show the season of summer forms a peak of accidents and that the collision accidents lead to more fatalities than derailment [29]. Moreover, with the same method of Association rule, data mining has been applied for train accidents from 1996 to 2005 on the Iranian Railway to identify patterns of accident causes and development of safety regulations [30]. To determine the risk in the metro operations, deep mining (Apriori algorithm) has been proposed for the fault incidents,

looking at near misses from the Metro Dispatch log, which covers many of the rail operations' abnormal conditions such as vehicles, power supply, communication, maintenance failure and so on [27]. Some scholars were trying to improve the traditional method of risk management with the means of machine learning but had not emphasised machine learning as a method as it was thought that the current risk assessment process could not guarantee systematic and comprehensive consequences. For instance, to cope with the restrictions of subjectivity in the data and to obtain reliable real-time risk and independence of the variables in decision-making operations, decision trees with fuzzy analysis have been proposed to enhance urban rail network safety [54].

The search is growing with time and there is an emphasis on trying to apply new technology in general for safety in combination with AI solutions. This combines computer and data science together with safety and risk sciences, even though they have not yet been proven as effective on a large scale.

2.3 Deep learning (DL) literature

Deep learning (DL) has been proposed as a tool for risk management in railway stations as machine learning in AI. Consequently, in this section, we'll come across various applications that make use of visual data from a variety of sources. Face recognition is becoming an essential part of computer vision. It has a wide range of applications in self-driving cars, health care, medical imaging, automation in manufacturing, video surveillance, robotics, and homecare technologies, to name a few. Improvements in IT have enabled vision sensors to be installed in railway environments. For example, CCTV cameras are widely used and rely on numerous camera sensors; these cameras are intended to avoid and manage safety and accidents in railway environments. Moreover, CNN applications offer improvement and satisfy results dramatically. Due to their high performance and promising results, convolutional neural networks (CNNs) have been utilised in visual computing in many studies, such as crack detections [55][56], and for floor area detection [57]. Additionally, computer vision has been analysed for use in autonomous emergency train stops [58]. Of the many

applications that have been applied to CNN, it presents those that are specifically related to railways. Such studies have been widely reported in the recent literature and use many data sources; they cover management, maintenance, safety and operations [21]. Image-processing approaches for implementing automatic detection have been suggested for monitoring railway infrastructure [37]. Moreover, learning machine algorithms can estimate the exact abnormalities by monitoring rail tracks [59] to perform a risk assessment of rail failure [37], diagnose track circuit faults [60], and provide information for decision support based on big data analysis [61]. For automated risk assessment and detection from the video for analysis of the rail surface defect, deep leering has been applied by a deep convolutional neural network (DCNN) model for image processing [37]. The length of the squats was inputted into the model from the rail tracks of the Dutch railway network and the growth of squats was studied to estimate the risk of failure. Also, from object detection and separation, a real-time Hidden Markov Model (HMM) is used at level crossings by applying a surveillance system coupled with communication systems for safety and security. The model aimed to detect and evaluate abnormal conditions such as people, drivers, and unattended objects at level crossing surroundings [38]. Moreover, for level crossing safety, a Bayesian network model was conducted with an automatic learning approach for predicting accident occurrences in France, which is shown as a useful risk in reasoning analysis to prevent level crossing accidents [49]. Moreover, for railway crossings, risk area detection image pre-processing and linear regression have been applied and have been found to be an effective and corrective method to detect the risk [62]. Also, in the U.S., for a live alert and detection of trespassing and grade-crossing collisions on the railroad, Big Video Data have been used from Closed-circuit television (CCTV) cameras. The Mask R-CNN algorithm has been implemented to recognise and understand such risks in real-time, which proved to automatically differentiate between types of trespassers, as well as compiling BD and mitigating the risk [40]. For detecting obstacles on the track, utilising deep neural network technology for comparing input and reference data to train

frontal view camera pictures was proposed in addition to obstacle detection from a train frontal view camera [63]. Moreover, for train safety and to increase the inspection accuracy, deep learning is proposed, including combining supervised and unsupervised learning using the image for recognition of the key parts and bolts positioning. Furthermore, the plate bolts in a moving freight train are used as a case to identify target parts during the inspection of railway locomotives [39]. For railway passenger safety of Seoul Metro, vision-based object detection algorithms were conducted with stereo and thermal cameras [64][65]. This experiment of vision-based object detection since 2008 has been summarised in conferences and did not present the deep learning method or ability to train clearly, which shows how the process has changed within the last decade, including the sensor and chips technology. In communication-based train control systems (CBTCs) deep-learning has been conducted to predict safety risks [66]. The CNN structure can be used to estimate crowd density at railway stations [67]. From the security side, the method has been used for detecting violent crowd flows [68], and identifying tools such as knives, guns and explosives that are wielded by attackers [69]. The related work from DL literature covers many aspects related to the railway in general and not only the stations, as well as being built from many perspectives. For instance, the view of rail track operation and maintenance is considered, in addition to trains and infrastructure, catenary, overhead and power systems, and other assets aiming to detect failures and/or obstacles. Even so, it have excluded much of the literature that is not related to risk and safety even though the literature that covers the railway stations specifically is very limited. However, some work that can be applied in the station has been mentioned and briefly summarised. A railway system contains a wealth of data, and visual processing technology can play an essential role in the industry's future. However, the data collection seems a barrier for the researchers. The rail track, infrastructure and trains have been explored utilising the deep learning from many aspects including operations, security, and maintenance. Most of the studies using deep learning consider the machines and devices, objects and

failures but the behaviour analysis studies have not been covered. Despite the importance of these components in the railway system, the station safety and the risk management have not been researched to the same extent in the industry. This apparent lack of research shows the need for a full framework to capture the dispersion and guide the future implementation of DL in the industry. Technology such as computer vision will play an essential role in railway system networks and provide practical methods to solve various problems. The wide distances and long tracks in many parts of the world with the growth in complexity provide maintenance and safety, security, and quality concerns. Furthermore, there are cost, schedule, constraints, and reliability, weather issues. Even though the deep learning algorithms have remarkably improved the performance of many fields such as visual recognition and autonomous systems, the concept of deep learning seems to still be growing in the railway industry. For such a method with the videos and image data, it needs a redesign of the subsystems concerning the targeted coverage and the requirements for gathering and analysing such data, which may pose a challenge in the short term.

2.4 Big data and railway

In this section, the BD is the focus, including related programs that have been applied in the railway. BD or BDA, data mining, deep mining, and data analytics which are all focused on the data are terms derived from computer science. However, there is a global shortage of comprehensive databases for risks and accidents in the railway stations. Moreover, it is clear that frameworks and reviews of BDA methods and IoT have not yet been implemented in many railway-related enterprises. IoT and Big Data analytics (IoT BDA) are still in their beginning stages [70]. Likewise, within the IT structure of rail organisations, BD technologies are also still in their infancy [71]. Furthermore, to provide a comprehensive overview of BDA in railway transportation, a review has been done, including operations, maintenance and safety. It has been found that the employment of AI in the railway sector is an emerging field of study. There is a need for improving the risk process in a real-time manner

[72][51] which forms strong evidence of the knowledge that catalyses this research. On the other hand, in terms of BD and databases, there are no professional incident databases in many operational railway stations, if at all. Some works of literature have been showing a move forward into digitalisation, for example, in safety management in the Shanghai metro. The metro operation incident database (MOID) has been presented as a qualitative tool for risk assessment and promoting the safety management level [73]. Such a data centre is the first step in the digitalising concept, which contains details of all incidents that have occurred in the metro operation. However, it can be improved to be associated with gathering risk data in real-time and then utilising AI methods for analytical predictions and monitoring ongoing learning. The BD methods are still not mature yet in the railway field; however, in the near future, there is expected to be a boom in both quality and quantity of data. That will reflect on developing industry-specific huge data analytics methods to make use of the available information and literature for many aspects such as safety, efficiency, decision-making and quality [74]. The BD research often discusses the characteristics of the data (volume, variety, velocity, and veracity) including storage, gathering and extraction tools from sources [75]. Aiming to consider the process of AI, considering ML and DL are a subfield of Computer Science and AI. This gives computers the ability to learn one step ahead of BD, even though they intersect with each other, and both have challenges and need further research [76]. Still, it is beyond the scope of this research. In particular, it provides a smart safety system via advancing risk management in the railway stations from a safety perspective, not developing the AI or the technology itself. In the context of BD techniques, Natural Language Processing (NLP) for text-based railway incident reports, visual analytics and BDA have been used for safety and risk management in the Big Data Risk Analysis (BDRA) programme in the UK [31][32][77]. Moreover, for visualisation tools, a Visual-Analytics System on a geographical map for Railway Safety Management has been created using data analytics to classify incidents (related to run-overs) risk

index along railway lines [78]. The safety performance of Britain's railway has improved dramatically over the last 50 years. However, BDRA program aims to improve it further as an improvement program to risk analysis and safety management led by the Institute of Railway Research, RSSB and the University of Huddersfield. It is based on the BD relevant to safety data as a new approach process aimed at moving from proactive to reactive based on the technology of data analytics, automating, decision support and of an integrated approach to safety and risk assessment. A part of the program (SMIS+) cloud-based on-line system which aims to modernise safety-reporting capabilities has been implemented in close calls, where it has been actively in use by Network Rail and train operating companies. The program in this phase focuses on the close call system and the reporting method via extracting the knowledge in a smart way in real-time, which presents the ability of AI application in safety and risk management. The question raised here is whether all risks can be reported, and whether recorded data can be fully explored. A further future critical step is to design an active system able to capture risks without any human intervention. Yet, it cannot see the whole image for the program because it has not been completed as the published work announced has only implemented Phase 1. No doubt that the use of such data will be beneficial for safety and decision-making. Likewise, this is one of the strong points of this research on the same program which extracts safety lessons and identifies safety-related information more quickly and efficiently, through a visual text analysis of calls records [33] and from multi-lingual text sources [79]. Also, to assess the risk of the trains that pass a red signal without authorisation (SPAD), the graph database was used with the BD technique method applied to a case study integrating multiple data sources [34]. Further, and relating to the sources of the text and analysis, NLP has been used for analysing railway investigation incident reports in the UK, to explore the relationships between causes and failures in accidents [41][80]. Another project in France aimed to develop methods which are assigned to IFSTTAR (French institute of science and technology for transport, development, and networks) by the state to

boost conventional safety analysis, and for developing evaluation techniques. So, AI has been proposed from the point of safety analysis and evaluates the files of the safety of guided transportation systems, tools and methods. Safety files include analysing outcomes of the safety history of accidents, which can be used for assistance to experts, applying knowledge acquisition, generating new scenarios and automatic learning methods [44][45]. Moreover, to improve security and aim to improve the quality of public transportation services, the ANN has been used to predict the high crime risk locations such as bus stops, and rail station locations [43]. Also, in the subway and from indoor air pollutants data for health risk assessment, a principal component analysis (PCA) as a BD method has been utilised to capture the dynamics and periodicity, and to detect the abnormal faults and ensure indoor air quality. In this part, it presents the literature from the BD view and the related programs and at the end of this chapter, some perspectives on this will be offered.

2.5 AI and risk management framework implementations in the railway station

The application of AI in risk management saves time and money with the ability of learning. Moreover, such advanced analysing captures the hidden patterns in the data, for instance, it can understand which time or season specifically has a high frequency of specific types of accidents with any associated factors. The method can provide more than a learning of lessons; for example, it is possible to gain some rules or conditions to enhance the risk management and safety system. Despite the fact that the AI in risk management in the railway industry is not fully mature, from the review of the literature, it is clear that train accidents and track data have been studied to some extent.

However, even though some of the studies have been published in risk and safety-focused journals, there are limitations associated with the adoption of machine learning in risk assessment in the railway. The trains, their components, tracks, signalling, and traffic still need intensive study. Some causes lead to the delay of implementation such as a shortage in fully applied research (complete picture), the lack of skills and lack of realisation of the applications. Also, another notable limitation

of the literature is the inability to collect and process massive data and security issues [81]. In addition, some research has developed a system for adding more passengers and more cars on the tracks; however, the stations do not meet these capacities. Others, from the perspective of operations, maintenance, and safety, have improved the track and trains and linked the subsystems, but the stations do not offer the same level of protection. In addition, some types of accidents in critical points of the system, such as the stations, have not yet been covered. Moreover, study of the risks in all the sub-systems is essential to provide the attributes which are needed to come up with integrated risk management for providing data-driven input. One example is that railway crossings as a risk area are identified in many of the studies, which support the need to conduct more experiments and research on the accessibility aspects. Also, accident reports data are available providing accessible in-text analysis and natural language processing (NLP) as a subset of linguistics, computer science, and AI are concerned with the interactions between computers and the human language. Even though they do not cover the whole system and all accidents in the system reflect the original issue of a lack of data. However, the diversity of languages still opens the gate to study from the researchers. Generally, previous studies have been concerned with the accidents' cause, related roots, emergencies, maintenance, simulations, and traditional analysis in many parts of the system. In contrast, it mainly focused on risk management in railway stations using advanced methods (AI). It might be said that it is only a matter of time before new technology is applied in this field and many others. Yet many studies are utilising the method as an advanced analytical tool but forgetting the ability of training that the AI has. Consequently, the AI method suggested in this research to overcome some of the weaknesses in the field fills the gap and presents the power of the AI for managing risk in railway stations. The application of AI in risk management goes further than advanced data analysis to form part of a smart system that can learn and be trusted as a timely support for decision-makers and as a predictor tool for risks [14].

2.5.1 Smart railway stations

The trend of new technology and application of the smart methods of (AI) seems to be leading the science, including risk management, smart mobility, smart cities, infrastructure, and safety science in the railway industry. Thus, it is possible that the concept of intelligence also applies to railway stations (Figure 7) [14].

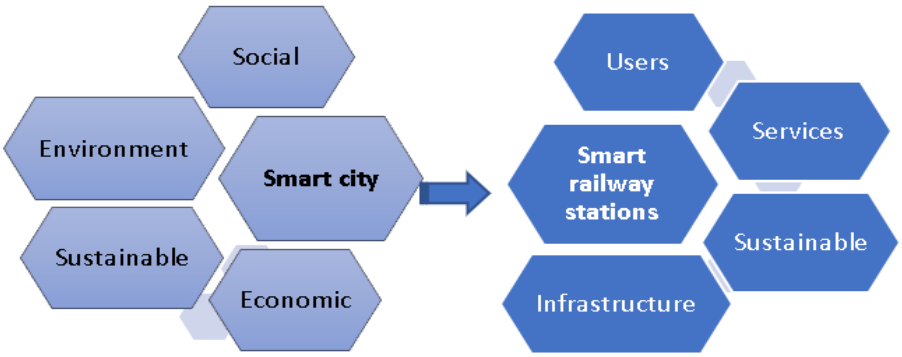


Figure 7. The concept of turning smart city technology into smart train stations [14].

Smart systems, along with smart cities and smart transport, will gain from railway station innovation, data businesses, and services. As more devices join the growing IoT and cloud, the railway industry will have much more incentive to benefits from the advantages of data. Even though the railway sector has unparalleled data owing to BD which can be gathered from the Internet of things, CCTV, ticketing, passengers, the progress may be slower than expected without the technology to turn it into meaningful information. The employment of cloud resources can be integrated with innovative technologies such as AI and smart train stations, both of which are part of the future smart city. For illustration, a standard for Smart cities (PAS 182:2014) defines some concepts (such as location, monitoring, metric, and so on) to try to define a collective system for linking data across organisations in a city, where a city can link a concerted identifiers set for a type of entity, and different

organisations can share information about it, such as accidents or faults in railway stations. Figure 8 illustrates some of the expected potential service technologies being developed for all railway stakeholders [82][83].

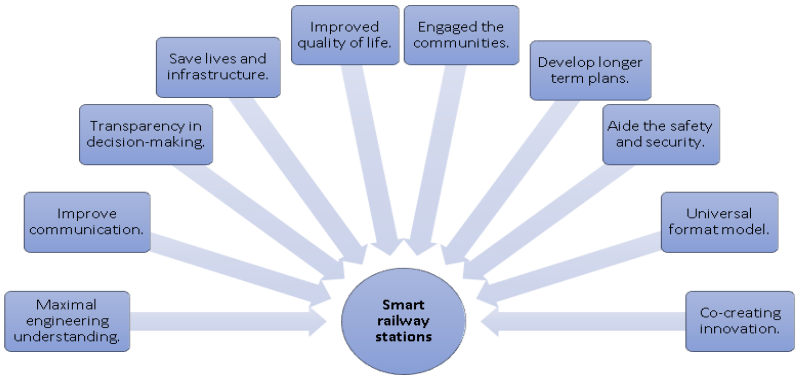


Figure 8. In railway stations, the smart city concept model (SCCM) has the potential to be beneficial and applicable [14].

2.6 Outcomes and discussions

For public and commercial transportation, the future research directions and the applications of operational management are real-time data processing tools and will be in higher demand, which deserves more in-depth explorations [84]. Digital is the trend for transportation, including railways in the UK; for instance, a strategy has been set out for modernisation and digitization in the next 15 years [85]. Yet, the digitization needs more time, research, investment, and employee skills to be a product in railway safety. A significant change must be made in railway safety for continuing improvement. As a new model, the modern socio-technological approach will offer railway stakeholders unprecedented scope, including excellent protection, satisfaction of workers and clients, better reliability, better efficiency, and less waste [86]. The automotive, marketing, business and financial industries are applying the AI for risk assessment, and all exceed railway fields. The application of AI has a variety of perspectives for risk management in the literature, including records, real scenarios, simulation or review and proposed frameworks. Different format of inputs to train the

AI models input have been seen such as textual data, sensor data, numerical data and video data etc. Also, this data input ranges from real-time data gathering to historical data or both. The data are an essential part of the AI research; in some cases, the analysis is data-driven. As railway stations are crowded areas, they are thus a challenge for safety and security, so efforts do not fall exclusively to the state or the stakeholders, but rather relate to the other factors including researchers and investments. The stations have characteristics such as being crowded and complicated, and may have weak management systems and so on. Many systems are located in the stations, with an open structure, which add to the complexity of the railway stations. The control and prevention, and the prediction of unexpected events in the stations are so critical that the new technology needs to be used more frequently and explored with intensive research. Therefore, the railway station will be selected to find suggestions for improvement and employment of technology and to design a framework for risk management. This review has shown that extensive data from the railway industry can be utilised to set up the new technology such as AI. Also, the literature shows there are distinct usages of the terms driven from factors such as the overlapping between safety, computers, and other engineering sciences, as well as the aims and perspectives of the researchers and the shortage of standards. In the next parts, it lists the main points obtained from the literature review that add to the argument for using AI in risk management in railway stations and supporting such a research methodology.

- The increasing use of IoT will reflect on the process of gathering data and benefits the risk inputs; it has been shown that the IoT offers practicability and effectiveness to intelligently realize early-warning risks such as risks of crowds [87].
- AI applications move and integrate with the development of smart systems such as smart cities, stations, smart or digital safety control systems and the concept of intelligent lives or autonomous vehicles systems.

- There has been a considerable rise in recent years and a great deal of development in the computer processors and sensors technology, applications, and reduction in the costs, which must be employed for protection in the railway stations.
- The conventional methods of risk assessment are not perfect for real-time outcomes and do not meet the dynamic conditions of hazards.
- The traditional safety analysis methods are inadequate for complex causal factor analysis and for predictions.
- There is a limitation in reviewing the AI applications in risk management, which enhance more intensive study and increase the importance of such research.
- There is the need for a system that can cope with a long history of accidents and prompt the safety authority to learn from each accident and to be updated in a timely manner with all previous accidents.
- The managers or railway safety officials require a tool that can estimate the failure risk on an earlier phase (risk early-warning), which provides time for actions and reduces the consequences as much as possible.
- Recently, the Advanced sensor network techniques and the passengers or workforce have smart gadgets that allow for collection of large amounts of data, so it is efficient to be used as input for AI models.
- There is a necessity for systems that can automate the process, discover accident characteristics and provide safety authority cognitive lessons with less human subjectivity and intervention.
- Provide a decrease in the direct and indirect costs from accidents in the railway industry.

- In complex situations, the decision-maker may have many inputs or indicators, but they do not easily obtain a result from them, so the prediction is necessary for assistance in such cases to provide decision encouragement.
- The monitoring systems such as CCTV need some amelioration to be efficient, provide online (timely) monitoring support for safety and security controls from different locations, and act and warn in unwanted events rather than being a passive recording system.
- In unwanted events, it is significant to provide quick, timely information from the operations for condition assessments and any incident/accident investigations.
- The needs of a robust system can inspect, detect, and locate the critical failures in all the subsystems, static or moving, and to ensure safety and security railway operations with low human intervention.
- The rise of autonomous transport system research and implementation requires smart sub-systems such as risk management safety systems for the integration process.
- There is a demand for tools to aid the experts who face some difficulties in reasoning and to explain the risks in the large organisations and avoid any possibilities of assessments that are either above or below reality as well as avoiding an incomplete or inconsistent view and any other uncertainties.
- Techniques are needed that can assist in transferring the experts' knowledge from one generation to others and allowing participants the risk analysis nationally or internationally in quick time.
- There is a need to integrate the varied data associated in risk sources with risk knowledge reasoning to obtain timely, precise results.
- There is a need for technology to help safety experts for evaluation, and for analysing historical data which are complicated for the human mind to summarise and remember.

- The traditional risk assessment static methods such as FTA, FMEA, etc. often have uncertainty and inadequate outcomes, so improving risk management is needed.
- The current inspection systems need to be developed to be more accurate, with less human intervention and effort, something which can be provided by the automated risk management approach.
- Looking for methods to avoid delays in operations and disruptions in the system, with high productivity and reliability, that prevents accidents and gather data from multiple sources in real-time.

Furthermore, the projects or research associated with the governments or safety organisations have some similarities. These works seem to be in a closed loop, where the usual outcomes of the project reference the previous works and do not compare the study with other research in the same field. Both BDRA and IFSTTAR programs do not have clear time plans. However, they have the primary raw material, which is the data, and that must give them an advantage in such a research area. Other projects that have been reviewed in the literature do not cover the risk management, which are more generalised for the whole railway system and have views from the data domain, like standardisation and aspects related to reasoning and ontology, such as the reasoning in the Web with Rules and Semantics Project (REWERSE) [88][89]. Likewise, for an ontology and railway data model, there is also the Intelligent Integration of Railway Systems (InteGRail) project [90]. To determine whether the data are large or not, it depends on the available size and of course, that depends on the time and region of interest for the analysis. The programs such as BDRA and IFSTTAR have access to most of the stations in the country, which gives them the priority for applying AI with the accepted size of data. Also, the outcomes of such programs are shown as an entirely unclear base where they were using their software and appear incomplete because the projects are still in progress.

In IFSTTAR programs, they are focused more on using AI as an expert system (Knowledgebase system) and for knowledge acquisition research about the safety as well as for classifying and recording the events of historical scenarios. The track, trains, power, and communications safety have been considered by the scholars and many improvements have been made. Yet, the stations as a point of the system did not get the attention they deserve, even though all passengers begin and end their journey at a station. The station must comply with the full system capacity; it cannot add more trains on the track with no consideration of the stations' safety. Stations are not only a connector to trains, but also a place where the community can stay and meet each other, delivering goods and in many cases, they are critical areas for connecting with other public transportation of the cities. These stations, with complex functions and structure, must take forward steps to benefit from the technology in order to become safer. Finally, the data science concepts methodology is overwhelming the other methods even though the many learning benefits from the AI in the railway are yet to be realised. Thus, it fills the gap and boost the potential of utilising AI to create a machine that can self-learn risk management in the railway stations and overcome some of the key challenges in the industry. In other words, from the literature, it is clear that what is not known about using AI in the risk management of the railway stations is more than what is known, hence the real need for this research.

2.7 Summary

Targeting the station achieves two goals in this work: first, it emphasizes the important role of this element of the system, which has been overlooked in earlier studies, resulting in a knowledge gap; and second, it offers a unique search. Furthermore, the proposed AI in risk management contributes to innovative research and adds to the fields of scientific knowledge. The chapter reviews the literature related to previous applications of AI to railway risk management for improving railway safety, such as ML, DL, BD, and smart stations but the literature is limited. The review highlights some overlap between the sciences and the variety of terms used in such methods of AI. It also

displays a lack of research and data in applying AI in the railway stations, which therefore forms part of the motivation for this study. The rail risk management is not yet intelligent or proactive, and requires intensive research. The literature also emphasises the importance of chosen railway stations to be studied, as a critical part of the system which has not been probed in the literature. Besides, clearly, the concept of transferring risk management to a smart format in the station is totally innovative, and the research will highlight and contribute to this concept. Moreover, the power of learning that AI have can benefit the risk management process for safety systems. Finally, the decision-maker, the dynamic risk, the complexity of the stations, and all the related uncertainties are issues that could be solved using AI as an appropriate tool to develop the performance of the whole process. When selecting the literature one criterion considered was the authenticity of journal articles and conferences. Also, the literature had to be written in English and written in the last decade. The work had to use AI (ML, DL) for safety and risk management and be situated within the railway sector as well as covering possible related literature that mentions the ML or related terms such as self-learning and in safety railway and risk management.

CHAPTER 3 RAILWAY AND RISK MANAGEMENT

3.1 Introduction

Managing risk is a universal requirement, and the railway, as a high-hazard industry, is not exempt as "risk is the effect of the uncertainty on objectives ", so the operational risk management in the station needs to be reliable against the uncertainties, and effective to capture risks (ISO 31000). The risk management methods in the railway industry differ from country to country, depending on the local regulations, methods used, size of the system and the operator, yet they share the same goal. Type of experience, human factors, culture, system modernisation, standards, and available funds are factors that play an essential role in managing risk. Risk management is a key component in the operations of each organisation, including general transport systems and railway stations specifically. Risk management is scattered in many fields, and the techniques are the same, yet many software and tools have been used. The attention from many and different areas of the Risk Management Process has shown the dissimilar but converging meaning [91]. The point of view also differs; for instance, the risk can be positive in project management. The methods start from a superficial level, such as the daily life risk management to the complex risk management in the mega projects. Similarly, risk management in the stations differs from the small to the extremely large stations. Hazards within railway stations may cause serious harm, not only to passengers or employees but also to the public, the business reputation as well as having social or political impact. These risks must be managed to mitigate the consequences and reduce the possibility of accidents, and the industry stakeholders are obliged to carry out risk management to minimise the risk to an acceptable level. This role is recommended by legislation, stockholders, government, passengers, workers, and the public. For

instance, for railway transportation, the EU and national regulations require a reliable system for decision support based on risk management to facilitate complex support and enable the exchange of experience to develop a common knowledge. Currently, however, this ideal has not been fully achieved and in the whole industry accidents are still occurring and different organisations are using very different styles for risk management. Thus, the railway market needs to develop its current risk management systems and provide a system able to aid the risk management, support data sharing and able to use a knowledge base and expertise [92] [93]. Many methods and tools have been applied in risk management, yet a number of fatalities or injuries are still recorded. Moreover, the risk level estimation process has been shown to have a gap regarding the decision support and safety or security measures decreasing identified risk [94]. It has been observed that in traditional risk assessment, some shortcomings have led to poor decisions in response to threats, such as managers being unable to differentiate appropriately between small and large risks. This chapter provides an analysis of current practices used to manage risk processes. Possible gaps within these popular practices are recognised, and research directions are proposed as a way to organise and improve the process. Accidents in stations and elsewhere can be perceived as unexpected/unwanted circumstances, resulting from an unmitigated risk or unidentified hazards. Such hazards are uncertain events or physical substances that might cause harm to people, assets, or processes, whilst risks are classified as the result of the probability and severity of risks occurring. This chapter includes considerations about managing risks and the components of the risk management processes.

3.2 Risk management process

In the common mandatory European risk management process for the rail industry, risk management is defined as “the systematic application of management policies, procedures and practices to the tasks of analysing, evaluating and controlling risks ” (Commission Implementing Regulation (EU) 402/2013)[93].Moreover, the Institute of Risk Management (IRM) states that “risk management

should be a continuous and developing process" and should be able to address all the risks of past, present and future.

Definitions suggest that developing process and risk assessment involves a systematic process, using the past data and addressing, controlling, the risk in the future, which supports the view taken in this research. The risk management process is carried out via risk and safety managers and the employees on the railway organizations as a vital part of public transportation strategy and a part of the critical national infrastructure in modern life, including the stations, which is critical infrastructure, also in many cities around the world. This process aims to keep the risk at an acceptable level and to identify events, or potential events, which could impact the fulfilment of the organization's objectives. So, it is obvious to recognize risk management as a necessary process in railway transportation operations [92][94]. In the railway operations, incidents or threats to safety which can take place are a waste of resources, and their impact extends to the sustainability, quality and time, so eliminating this waste means avoiding those activities through risk management [95]. The general principles applicable to the risk management process in commissioning the implementation of regulations (EU) informed that the risk management process should start from a definition of the system under assessment [93]. Thus, there are Foundations or Key Terms of the Safety Risk Assessment, such as the Hazard, Risk, and risk mitigation. Sundry definitions for risk terms exist due to its wide application in many disciplines (see Table 3 below) [91-96].

Table 3. Selected Examples show terms and definitions from different resources [96][97][98][99][100][101].

Terms	Description	Sources
Hazard	"The real or potential condition that can lead to: •Injury, illness, or death for passenger, public and employees. •Damage to or loss of infrastructure (facilities), equipment, subsystems; or to the environment''.	FTA
	"A risk source where the potential consequences are related to harm.''	SRA
	"Anything with the potential to cause injury or ill health.''	H&SA
Risk	"Risk is the consequence of the activity and associated uncertainties or the possibility of an unfortunate occurrence.''	SRA
	"The possibility of something bad happening.''	Cambridge Dictionary
	"The effect of uncertainty on objectives.''	ISO 31000: (2018)
	"Combination of the probability of occurrence of harm and the severity of that harm.''	BS EN 16704-1:2016
	"The chance that someone will be harmed by the hazard.''	H&SA
Risk analysis	"Risk analysis is a process that is used to understand the nature, sources, and causes of the risks that you have identified and to estimate the level of Risk. ''	ISO 31000 (2018)
	"A systematic process to comprehend the nature of risk and to express the risk, with the available knowledge, which is often also understood in a broader way to include risk assessment, risk characterization, risk communication, risk management, and policy relating to risk, in the context of risks of concern to individuals, to public and private sector organizations, and to society at a local, regional, national, or global level.''	SRA
Consequence	"The outcome of an event affecting objectives can be specific or uncertain and can have positive or negative direct or indirect effects on goals, can be expressed qualitatively or quantitatively, and can escalate through cascading and cumulative impact.''	ISO 31000 (2018)
Risk assessment	A process that is made up of three separate processes: • Risk identification, • Risk Analysis, and Risk evaluation	ISO 31000 (2018)
	"An overall process is comprising a risk analysis and a risk evaluation" (ISO/IEC Guide 51:2014).	BS EN 16704-1:2016
	"A structured approach to assess risks associated with the maintenance of vehicles, including those directly arising from operational processes and the activities of other organisations or persons, and to identify the appropriate risk control measures.''	RSD 2016/798
Accident	"The unplanned, uncontrolled event giving rise to death, ill health, injury or other loss to persons or damage to the material."	BS EN 16704-1:2016

Risk management	“Systematic application of management policies, procedures and practices to the tasks of analysing, evaluating and controlling risk’’	BS EN 50126-1:2017
	“Activities to handle risks, such as prevention, mitigation, adaptation or sharing.’’	SRA
	“A written document that records a three-level process: 1- Identifying the hazards in the workplace(s) under your control. 2- Assessing the risks presented by these hazards. 3- Putting control measures in place to reduce the risk of these hazards causing harm’’.	H&SA
	“Coordinated activities to direct and control an organization with regard to risk.’’	ISO 31000: (2018)
Uncertainty	“Not knowing the right value of a quantity or the future outcomes of activity, that can be for a person or a group.’’	SRA
Black swan	In relation to risk: “An unknown unknown” (rare event)	[102]

There is an overlap between the terms from many perspectives, such as risk management defined as a part of the risk analysis (SRA) or risk analysing as a part of the risk assessment, as (ISO 31000) presents. The area of risk assessment been shown to have a shortage in the application of standardized terminologies, resulting in various contextual interpretations of risk and safety [53]. Likewise, regarding the concept of risk, it has been shown that different fields have different definitions, and it can be defined as the collection of both the opportunity of a specific dangerous event (probability) and its outcomes or its consequences after the apparition of this dangerous event as an equation of the risk = The possibility of the event occurring × The impact consequence of events. The risk has some aspects and characteristics, including the objectivity, uncertainty, difficulty to affirm risk and rejecting the risk when it is not fully disclosed, as the risk cannot be dismissed and the accidents are unpredictable [103]. Risk as a scientific field requires universally understood terms and authoritative definitions, and the Society for Risk Analysis (SRA) has created a glossary to respond to this need [52]. Moreover, it has been noted that the current standards and regulations for safety management of railway systems have some weaknesses and are not fully suitable for critical infrastructure needs and provide only limited assets protection. Some points support that they need more of a description

of methods and principles to improve the safety of railway systems, as currently they only focus on functional safety, and they consider only the primary impacts on human lives. Moreover, they did not consider other assets which affect people's lives and health such as integral risks concepts [104] [105]. Railway operations have experienced quantitative and qualitative changes, and the management of operations is subject to challenges from both inside and outside the industry, for instance, because of the rise in public demand, safety and security of the operational service capacity, and improvements to operations and infrastructure. In fact, the old method of operation management processes in rail networks cannot satisfy the standard developed operation management, so actions must be taken for the process, including risk and emergency systems [106]. Although the Safety Risk Model (SRM) developed in 1999 for the railway is system-wide and well-structured, clearly labelled and generally logically correct, it has also been noticed that there are still shortcomings related to how it is used. Utilising accident data over 5 or 10 years may not identify all the risks of unwanted events that may occur every 20 or 30 years or more. Moreover, the consequences estimate has little data available which can vary considerably in quality and be subject to great uncertainty [107].

The traditional risk assessment depends on causal analysis of the events, also related to the evaluation of the probabilities, which has a strong inability to deal with complexity. Moreover, these approaches have limitations in analysing, as it has been assumed that the components of the system have linear interactions, which prove the need for convenient methods to cover such limitations [52]. Risk management has been defined as a continuous management model with the goal to recognise, analyse, and evaluate possible hazards in a system or linked to an activity, and to distinguish and present risk monitoring measures to eliminate or decrease possible injuries to people, the environment, or other assets [108], it is noticed that the AI has not been considered in previous risk management guidance or standards. Moreover, the black swan concept of an event of risk explains the event has three characteristics. First, that it is an outlier, then that it has severe consequences, and finally, it is

predictable. Thus, the technology of AI supports the proposed application of AI in risk management which has the powerful ability to predict [109].

3.2.1 Risk management engineering and AI

It is believed that risk is prevalent, so maintaining it at an acceptable level is essential and must be controlled all the time. AI provides excellent opportunities for aiding humans to achieve these goals of risk management. Using AI techniques for analytics, decisions support, and engineering risk is rapidly growing. The developing technology in CPUs, GPUs, internet and IoT and with the data importance raised and available in some sectors leads to an increase in the use of AI as well as many applications used in, for example, the engineering and risk fields. Also, these applications are promising multidisciplinary research fields which are influenced by the technology era revolution, and the design of more complex systems, in line with the increase of natural threats, engineering, economic, and technical [110]. Even though there are more than 30 methods for engineering risk assessment, the current techniques of risk assessment have a limitation in executing real-time risk. Also, the scale range of the classical approach is narrow and does not consider any dynamic changes in the process. Thus, the growing demand for real-time risk assessment is expected to increase the use of machine learning [51]. This concept of application has been spread in many areas, such as the NASA Space Shuttle where risk can be predicted based on the principles of neural networks [111]. Moreover, a neural network has been applied for software risk analysis and debugging to identify the location of faults [112], and for robust wireless networks, the Fuzzy systems have been used [113]. In addition, some sectors have used some applications in risk management, such as in the energy sector, Cyber-Threat, Nuclear power, medical and health, infrastructure, maintenance and financing. There is a diversity of applications in many sectors, and the scholars proved the power of AI to solve the issues and show significant improvements. Recently, AI methods such as ML and DL have been utilised in the risk evaluation of the railway industry [114]. Commonly in the railway industry, the

safety management system (SMS) is often the platform for the safety data, risk level estimation process, review, reporting, analysis, and monitoring in the railway stations. However, to be able to deal with the AI process, these platforms or software must be advanced to meet the massive data and to be able to utilise AI methods.

3.2.2 Risk management and railway stations

Currently, there is an investment in the railway industry, for instance, the railway in England has the High Speed 2 project, where the stations will connect to new projects and will obtain redesign or refurbishment, thus increasing the network and increasing the vulnerability and that requires a high level of resilience and safety, and more cooperation between all stakeholders for strategic risk management future. Railway stations are complex open environments that interface with the line and are a critical point for the system, economy, and society. Many passenger accidents have occurred at stations, and it is imperative to ensure that passengers in the station surroundings are safe and secure, with active risk management processes in real-time. Risk management aims to reveal failures, accidents, or weaknesses in methods within a structured approach so that timely actions are taken to avoid, diminish, or transfer risk, and to reduce the probability of influence of risk (Risk Management Standard AS/NZS 4360, 1999), see Figure 9.

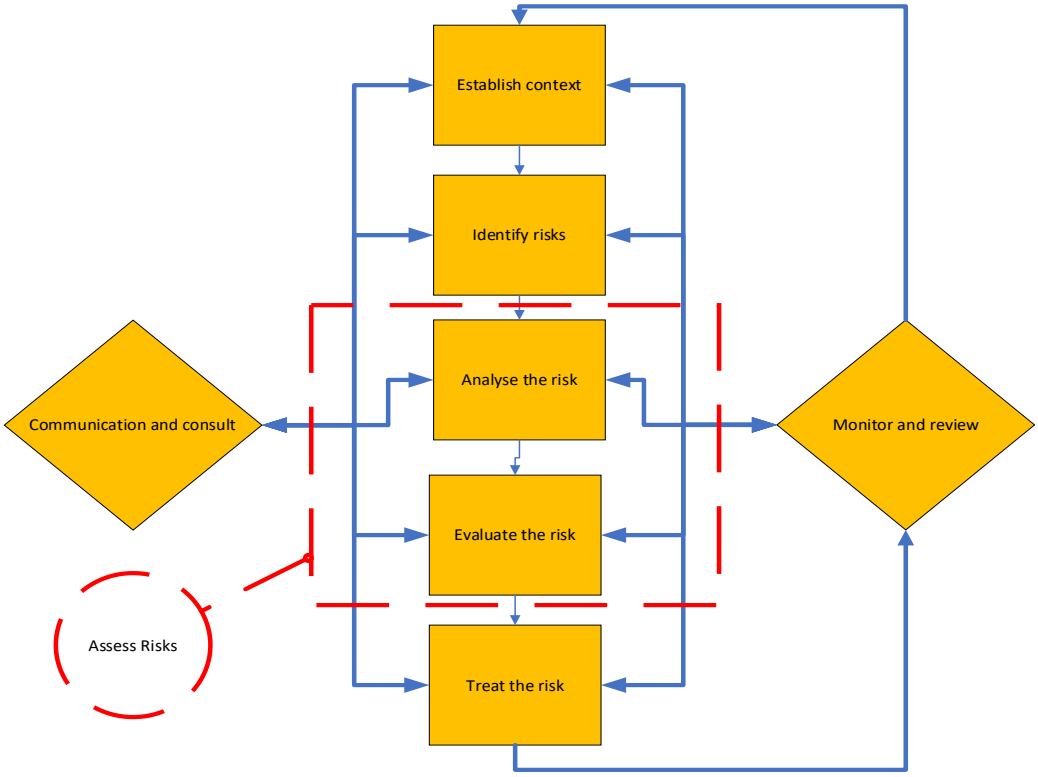


Figure 9. Australian Standard for Risk Management (AS/NZS 4360).

Furthermore, British Standard BS4884-3:1996 (BSI 1996), which is identical to European standards IEC 300-3-9:1995, provides guidelines for risk analysis and definitions to explain the concepts of risk management, as shown in Figure 10 [115] [116].



Figure 10. The risk management concepts [115] [116].

Many factors play a role in station risks, such as human factors. For instance, people fall into the gaps or on escalators; they can slip or trip, they can become trapped in the doors, receive an electric shock. Every day, we see passengers running on stairs, facing overcrowding on platforms during alighting and unloading and so on. The industry operator has a responsibility of care and a duty to put in place any possible risk management monitoring, by gathering data and implementing measures to diminish the risks to travellers. In this research, AI is applied in risk management for stations focusing on the dangers connected with station operations, including the following concerns:

- Safe operation of the whole station to improve passenger safety experience and the business image by delivering dynamic smart risk management.
- Predictions of the primary causes of accidents which occur at stations, then mitigate the risks utilising live history data and external information.
- Station design and layout, including sub-systems in the stations such as signage and flooring, CCTV, and fire alarms to be integrated for gathering real-time data, for a novel method of risk management.
- Taking precautionary measures and supporting decision-makers to alleviate hazards caused by unexpected factors, such as managing overcrowding in real-time.
- Effective risk management with a quick response must have the ability to deal with multi-risks that possibly occur in stations and with their interactions at the same time.
- Reduce human intervention in the system and extend gathering the data that can be an indicator of the risks.

Many station risks need to be managed, including the consequences of accidents. Moreover, full-coverage safety management includes security management and the management of critical sub-systems, as follows:

- Train and station operations
- Passenger flow and traffic
- Equipment and ticketing
- Fire alarms and other subsystems such as HVAC and facilities
- External environment and social security

In fact, a parallel process framework for Utilizing BD for Enhancing Passenger Safety in Railway Stations has been conducted [110]. This framework can support the decision and make a comparison of results between the classical risk management results and risk management with the ML. The conventional process takes the outcomes from the likelihood and the consequences, and the technology of AI such as ML/DL can be compared to the historical and real-time data. The BD inputs in the framework [110] will afford data-driven insights to decrease accidents, improve safety, quality, support the decisions and manage risk in the stations via AI technologies. This framework enhances the automated process, and it is applicable as the first phase in implementing such technology so that both approaches can be executed independently and in co-ordination with each other [117].

3.3 Risk identification and assessment

Detecting, recognizing, and recording are the steps to risk identification which aid the organisations to identify risks spot that affect the achievement of their objectives. Extensive description is significant because a risk that is not recognised at this step will not be covered in the risk analysis stage [118].

Depending on the Commission Regulation (EU) No. 402/2013, there is a common safety method for risk evaluation and assessment (CSM RA), a risk management process and an independent assessment that are all required as parts of an iterative process starting from identification as a part of system definition [93]. For a system targeted assessment, the process starts with identifying the risks,

then a demonstration of the compliance safety measures, and the requirements linked to identifying the risks. The actions or steps end when the compliance with the system and with all the safety measures necessary to accept the risks connected to the identified hazards are observed. This regulation in a common safety method (CSM) stands as a framework for European rail industry risk management and does not set specific tools or techniques to be applied.

Risk identification is the process of exploring (determining) risks in a context to ascertain what could go wrong in either the station or its operations at any given time during the daily processes. Causes of risk and possible outcomes need to be identified before they can be acted upon. Experts in their own field have common methods of identifying a risk situation level. However, these methods present many disadvantages, such as uncertainties, subjectivity from one to another, costliness, and bring time-consuming. The different identification methods are illustrated in Figure 11[119] [120].

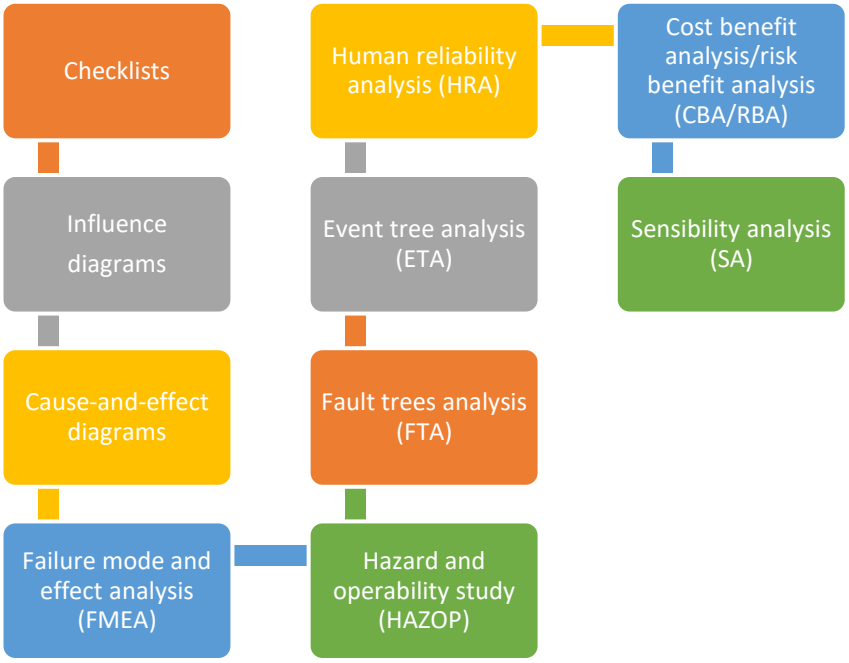


Figure 11. Risk identification methods [119][120].

Depending on the identification of risk, characteristics will be assessed to evaluate whether the risk event is worth additional analysis. Subsequently, measurement metrics (risk probability and risk consequence) are defined, for an estimate of risk level and risk analysis, which leads to risk mitigation or avoidance control measurements. To measure risk level, the likelihood (or probability) and severity of a risk event require definition (risk assessment function). The likelihood designates the possibility of a risk event happening, and the severity or consequence designates the influence of the risk event. The risk level is the result of risk likelihood and severity. In nature, risk quantities are either quantitative or qualitative, as indicated in Table 4 [121][122][123]. Qualitative techniques are analysed by descriptive arguments (low to high), or identified on a predefined range while the quantitative models involve [124][125]:

- Statistical analysis, e.g., regression, least squares, path analysis
- AI methods such as Expert Systems
- Probability theory

Table 4. Comparison between quantitative and qualitative risk assessments.

Risk quantities	Required for Analysis	Impediments
Quantitative	Historical data through statistical analysis	• Data are hard to obtain • Limited to a very small area • Not always accessible • The data are not always in the form wanted • Complex calculations • Hard to perform without an automatic tool • Time-consuming
Qualitative	Utilising experts' judgments	• Depends on attitudes and beliefs • Subjective approach • Influenced by individual bias, decisions, and expertise • It is not an accurate measure • A cost-benefit analysis is not implemented • Insufficient differentiation between major risks • Results depend on the quality of the risk management team

For risk analysis methods, many methods in the literature currently employed for analysis can also be utilised for risk analysis in railway stations, such as:

- Probability and impact grids
- Estimation of system reliability
- Fault tree analysis
- Event tree analysis

It has been observed that the development of urban rail transportation operations creates a demand to transfer from a quantitative to qualitative-focused management. Thus, conventional operational management techniques cannot meet the needs of network operations management. Therefore, it is necessary to improve and study management methods to strengthen safety and risk monitoring [7]. Station risk analysis can be challenging because there are several types of risk and techniques to measure them. Most recognised methods are used across the risk assessment process, as presented in Figure 12.

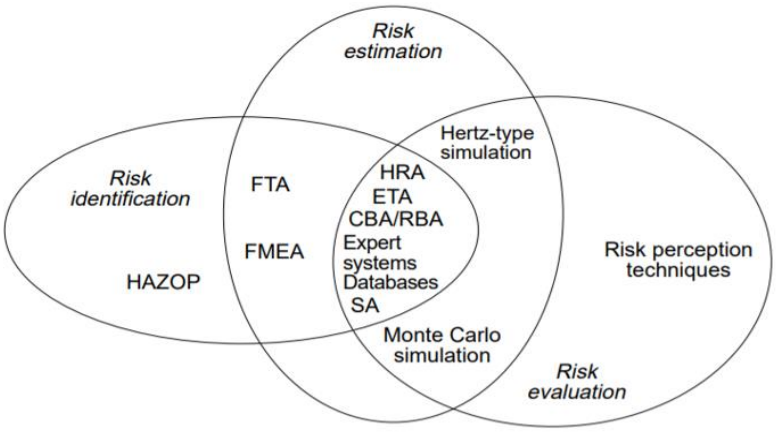


Figure 12. Risk analysis techniques [126].

Moreover, risk management in railway stations is complicated around the world since each country has its own views on what constitutes a risk, and other factors such as how to evaluate it, the safety culture, economics, national regulations and standards, and what hazards are considered. The risk management process should promote an understanding of the nature of these risks and contribute an explanation for:

- Identifying; whether adequate risk control measures are in action.
- Identifying; whether any additional control measures are reasonably practicable.

The risk assessment process can be present as several tasks such as Risk Analysis. The BSI Standard (EN 50126-2:2017) presents the life cycle of risk assessment including Risk analysis and Risk evaluation as part of the safety. The process merges with hazard control and has been described via Hourglass Model considering both the railway duty holder and the operator (see Figure 13). The model targets a technical system from design and to implementation and support documentation. The model shows the risk analysis as part of risk assessment, and as standard it does not mention the details of applications, and connection between subsystems details. The model links the assessment strongly with the history of accidents which is no easy task and needs time as well as facing challenges such as archive process and experts' retirement and delay caused by research of a long history.

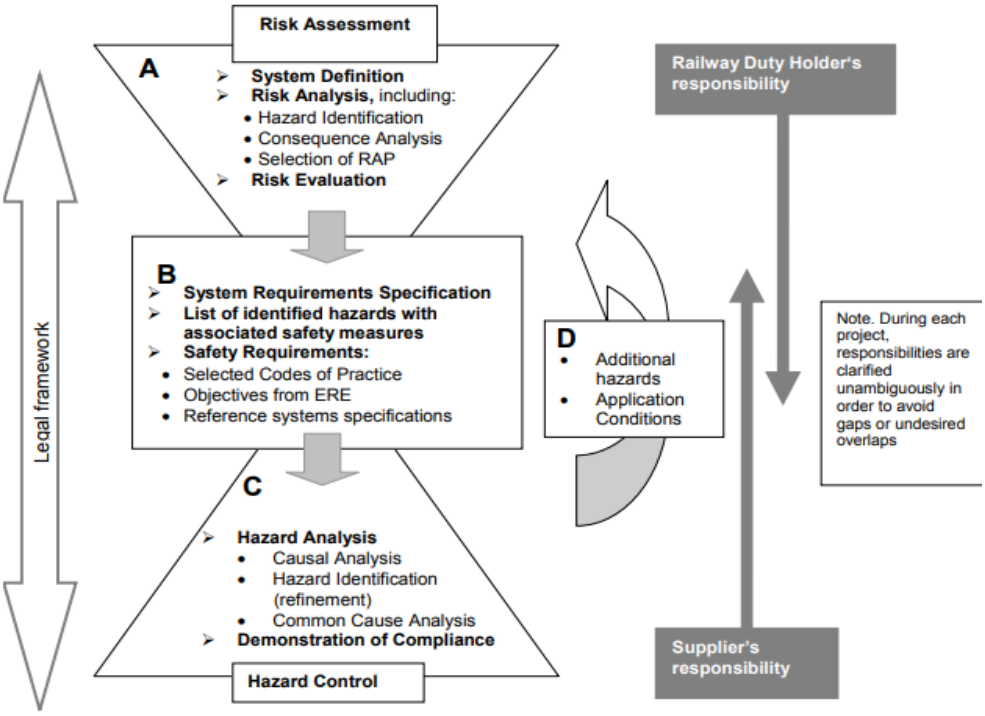


Figure 13 Risk assessment Life cycle (The Hourglass Model) [127].

Also, another model is Bow Tie (BT) which is very popular and has proven to be valuable in analysing past accidents and signifying improvements to avoid further reoccurrence of undesired events which have also demonstrated unwanted event prediction. Moreover, it has been shown that it can be used for both qualitative and quantitative. However, a typical BT has a static structure limiting its ability to capture risks in real-time and monitoring, and this consequently delays the probability of updating, which is a fundamental determinant in dynamic risk management. The BT evaluation methodology can be created to integrate BD into the risk management process. The BT analysis has been enhanced through associating related Industry Data Taxonomy sources to each of the factors of the BT. These data sources can include many aspects, for instance, live monitoring, social media, maintenance, and safety management and so on. BD analysis can be beneficial to the risk assessment, support the decision-makers in real-time, and reduce human errors. This method can be fully integrated into passenger data and the business model for the railway station. The present BT approach is updated to

be a dynamic structure, and this includes the BD, which will reflect on the occurrence and probability of accident consequences and an acceptable safety level. The process will occur in real-time by reliable models and internal or external information in the station and will be periodically updated using indicators of the risks [117].

3.4 Risk analysis in the stations

After the risk has been identified, the probability of this risk to occur and the consequences of the event will be evaluated and rated. In this part, comparative accidents and the numbers will be presented. Unwanted events that can lead to major hazards have been grouped, as shown in Table 5. These illustrate the variety of events that demonstrate how they can be interconnected, making the station a high-risk, critical part of the system. Moreover, a 2019/20 statistical report from (RSSB) (Annual Health and Safety Report 2019/20) informs that most passenger injuries come from accidents in stations. Most major injuries are the result of slips, trips and falls (see Figure 14), of which there were nearly 200 in 2019/2020, including other risks such as Trap-and-drag by trains.

Table 5. List of unwanted events in the stations [128] [129].

Unwanted Events	Details
Platform-train interface accidents	Including falling in the gap, from the train, trapped by trains, and suicide.
Falling	This includes slip or trip of either passengers or workforce in the station areas, such as escalator or on stairs.
Overcrowding	Including stampede, failure during evacuation for emergency.
Structural failure	Any incident associated with civil infrastructure. Such as collapse/failure in the station boundary.
Crossings and trespass	Including suicides.
Power failure	Any incident linked with a major system-wide power loss, which will affect trains and stations.
Explosion	This emerges from either deliberate or accidental ignition, pressure build-up of flammable substances, or bombs.
Fire	This can occur in the station, train, shops, escalators, or other machines, or other locations in the station's boundaries.
Natural disasters	Such as flooding, extreme weather and earthquake.

Toxic goods	Including release goods from a freight train in the station due to a collision or train accident.
Security accidents	Including terrorist attack, misuse, and criminal actions.
Railway staff risks	Including track workers, exposure to moving trains falls, assault from passengers, and high voltage electricity injuries.
Train accidents	This includes detriments, collisions in the station border.

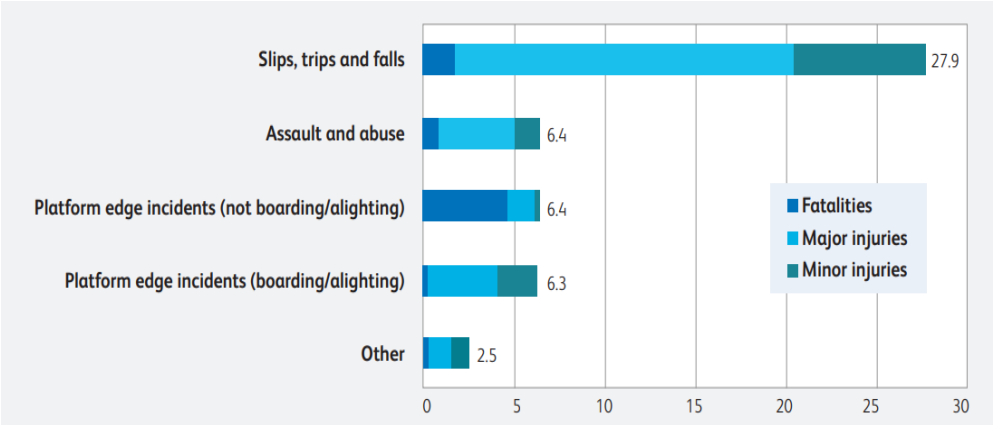


Figure 14. Station operations accidents 2019/20 in the U.K. [130].

Based on the RSSB for eight years ago there has been a focus of resource and expertise on a platform-train interface (PTI) in the stations, which was being exposed to a rise in risk and was becoming more likely to cause a fatality. Also, from the RSSB reports, there is a result that shows that most passenger catastrophe accidents occur as a result of accidents at the platform edge with no train present. These accidents in the station are not exclusively related to passengers and the public but also include the workforce. The crucial risks for all sets result from slips, trips and falls, assaults and abuse and the platform-train interface. The accidents can occur in many locations in the bordered stations, and the falls, slips and trips occur on stairs more than in any other station areas (See Figure 15).

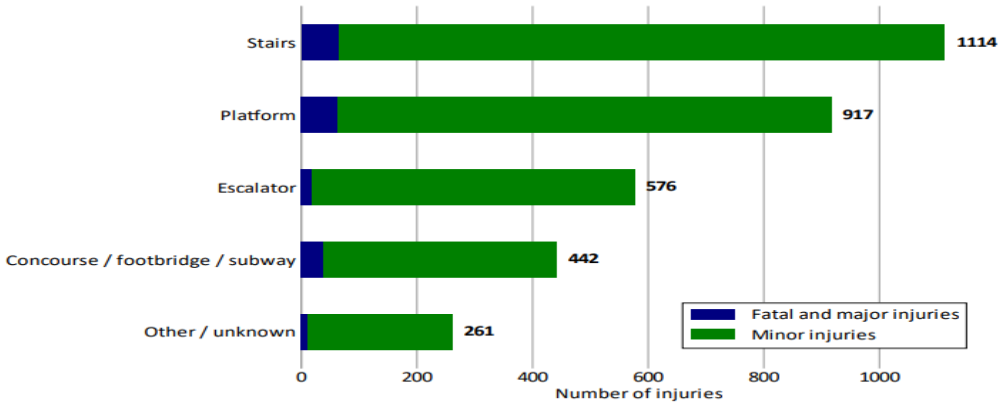


Figure 15. Areas of non-workforce slips, trips and falls in stations (2019/20)[131].

The passenger risk profile includes accidents occurring for a wide range of reasons and many types of risks, for instance slips, trips and falls, crime, bumping into objects or people, these involve train accidents, such as accidents while boarding or alighting. The above lines confirm that various accident types still occur in railway station transport. Also, this presents the importance of risk management in the stations to save more lives and assets, so effective risk management with technology such as that which is proposed in this work, aims at zero injuries and is definitely required. This contribution to the method will reflect on safety of the system and improvement in the industry.

3.5 Fault tree analysis (FTA)

FTA is a logical and top-down technique for failure (cause events) analysis, including their chosen risk accident interactions. In general, the common FTA structure design divides the system based on functional modules or mechanical construction. FTA uses a simple logic presentation to assist researchers in easily comprehending the analysis outcomes and analysing failure through a visual tree. It is a tool widely used for classifying and understanding the roots and grouping the causes that lead to an unsafe event, showing the paths expecting to lead to the top event (possible scenarios). Moreover, FTA is the method of reliability assessment of systems, risk assessment, and safety analysis. It is possible to perform qualitative and quantitative analyses using FTA. The quantitative

outcomes of FTA require professional knowledge from experts to assess and define the failure based on the accident history, analysing the top event based on its probability. However, in this part of the research, this is not the case, as the understanding of the risks and association analysis are targeted as a base for the application of AI methods for risk processes. FTA has analyses (positive and negative) such as maintenance tree, events tree analysis, and Functional Hazard Analysis (FHA). The model uses gates OR (output occurs if one input occurs), and AND (output occurs if all inputs occur). For the negative events, the approach describes things that could go wrong, and for a positive event, for example, the things that must work together in operations. The analysis aims to determine how the top event can be caused by specific or combined lower-level events. The roots of the top failure are linked through logic gates (i.e., AND gates and OR gates). For instance, a simple fault tree is shown in Figure 16 [132][133][134][135].

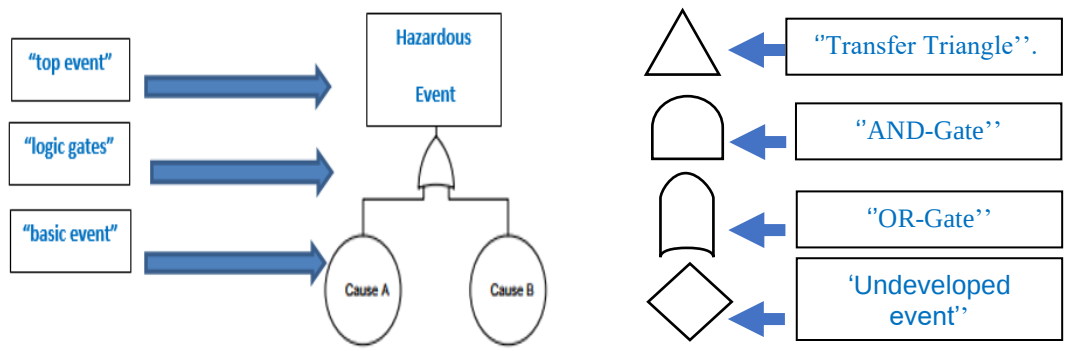


Figure 16. Simple fault tree.

To categorise the risks in the railway station, it has been suggested that the risks be grouped as shown below in Figure 17. In the next few parts of this chapter, using the FTA the existing information relating to the risks from all sources posed to railway stations will be analysed.

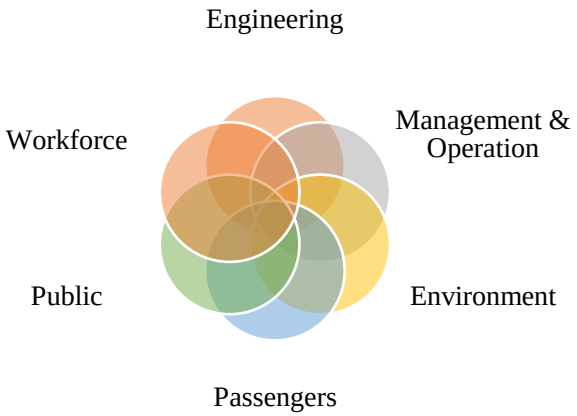


Figure 17. Suggested grouping of risks in railway stations.

3.6 Summary

Clearly, it is concluded that managing risk for an open system with massive flow and many functions is not an easy task, and this is above human capabilities. The famous method of evaluating the risk comes from understanding the probability of occurrence and severity (consequences), which has a percentage of uncertainty and takes time and includes other challenges such as the subjective judgments of humans and the cost. Depending on ISO "risk is the effect of uncertainty on objectives", so the uncertainty needs to be addressed not only by the traditional methods but via applying the new technology. Obviously, risk management has overlapping terms, and many risk management methods do not cover the use of technology methods like AI in the process. Nowadays, with the future of taxi drones and digital lifestyle, and IoT, the risk management process needs careful attention and should be developed by the authorities and operators. The risk management methods with the traditional evaluation are implemented, which are similar for many fields, including the railway. Yet, despite the development and the increased understanding of the risk the process, still there are many areas in need of development and that need to be improved to cope with the new era of technology. The analysis of the Railway Risk Management presents the reasons why additional research and development are required, as no standard designates AI as a method for risk management.

CHAPTER 4 THE IDENTIFICATION OF THE RISK ROOTS

4.1 Introduction

The risk management methods assist in minimising the event of accidents in the system, including stations. However, the main railway stations can be very congested, causing large passenger flows and creating risky places, which may become vulnerable to many unwanted events like natural disasters where the safety of passengers and manpower are threatened. In fact, even though risk management shows some uncertainties, it is critical to carry out a detailed analysis of the issues that may occur at railway stations. In this part, as a part of the methodology, the fault tree analysis (FTA) as a traditional tool for analysing different types of risks and for an overview of the risks in such infrastructures of the system. This will present the risks that might occur, their roots and the correlation between all of them and then the chosen risk will be analysed utilising the suggested method (AI) later in this work. The hazard groups have been recognised and analysed in the following part including dangerous goods and disruption effects

4.2 Fire and explosion risk

Fire is proven to be one of the major risks for railway stations because of the complexity and geometry of some stations and because of other factors or challenges, such as the availability of passengers, evacuation measures, and effectiveness of the fire and safety systems. As soon as a fire arises in a railway station, the most harmful issue is not only the flames and high temperatures but also contaminated gas. Large amounts of smoke tend to spread rapidly to the station and possibly to the platform along the area, depending on the fire growth and the location where the fire started. Once a

fire happens, the effects can be destructive in terms of lives lost, injuries, and damage to the structure, environment, and business continuity [136]. On 18 February 2003, for instance, a subway train was set on fire with fuel by a mentally ill patient at Jungangno Station in Daegu, South Korea. The fire rapidly spread to all six vehicles of the train within 2 minutes due to the extremely flammable inner material of the train, destroying two trains and producing substantial fatalities including 192 dead and 148 injured. The reasons are listed below in Figure 18. A further instance is the 1987 fire at London's King's Cross Station in the United Kingdom leading to 31 deaths and 100 injuries. From the literature, the fire cases in stations over the world are presented in Table 7 [137][138], which show how the industry became remarkably safer from fires, reducing the risks in the last century.

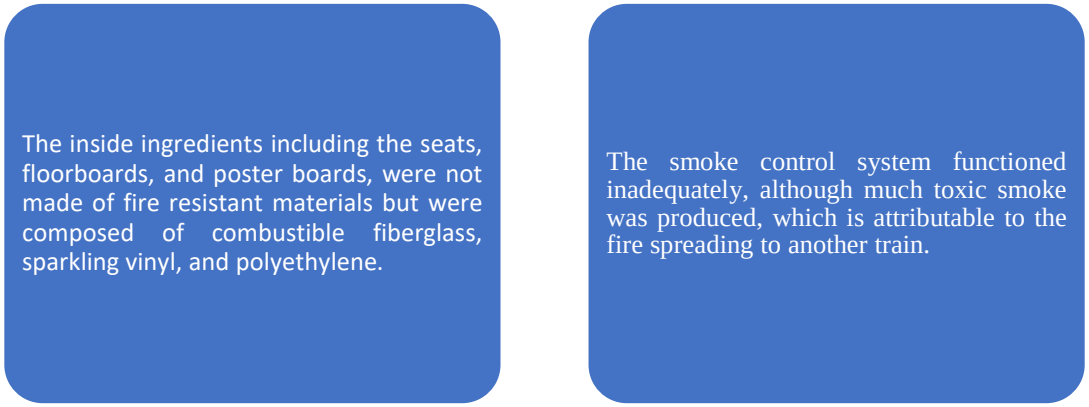


Figure 18. Reasons for South Korean train fire.

Table 6. Example lists of the fires in the stations and number of deaths and injuries.

Year	City	Cause	Deaths	Injuries
1969	Beijing	electric locomotive failure	6	200
1973	Paris	Arson	2	0
1978	Cologne	cigarette butts	0	8
1979	San Francisco	short circuit	1	56
1979	Paris	short circuit of car	0	26
1979	Philadelphia	transformer failure	0	178
1980	New York	cigarette butts	0	4
1980	Hamburg	seat catch fire	0	4
1981	Moscow	circuit failure	7	0
1982	New York	transmission equipment failure	0	86
1982	London	short circuit	0	15
1983	Nagoya	rectifier failure	3	3

1983	Munich	circuit failure	0	7
1984	London	warehouses catch fire	0	18
1987	London	technical failure	32	100
1991	Zurich	short circuit of subway locomotive	0	58
1991	Berlin	Arson	0	18
1995	Baku	caused by electrical malfunction	289	270
1995	Daegu	gas pipeline leakage	103	230
2003	Daegu	arson in the subway train	192	148
2004	Hong Kong	arson in the car	0	14
2013	Moscow	short circuit	0	52
2019	Cairo	train collided, and its fuel tank exploded	25	40

To recognise the most critical causes of the undesired events in a system failure, FTA is widely used as a logical analysis method to deliver logical, practical connections among subsystems of a system by flow chart [136]. The FTA flow chart analysis of the fire event in the railway station is shown in Figure 19 below where the station fire incident involved fires in both public and non-public areas of the station in occupied areas and rentals [139][140][141][142].

There are four sub-hazardous events that could contribute to the train fire, including the following:

- Explosion fire,
- Electrical sparks,
- Open fire, and
- Train fire, which is defined as an event linking the fire on any part of the train both inside and outside.

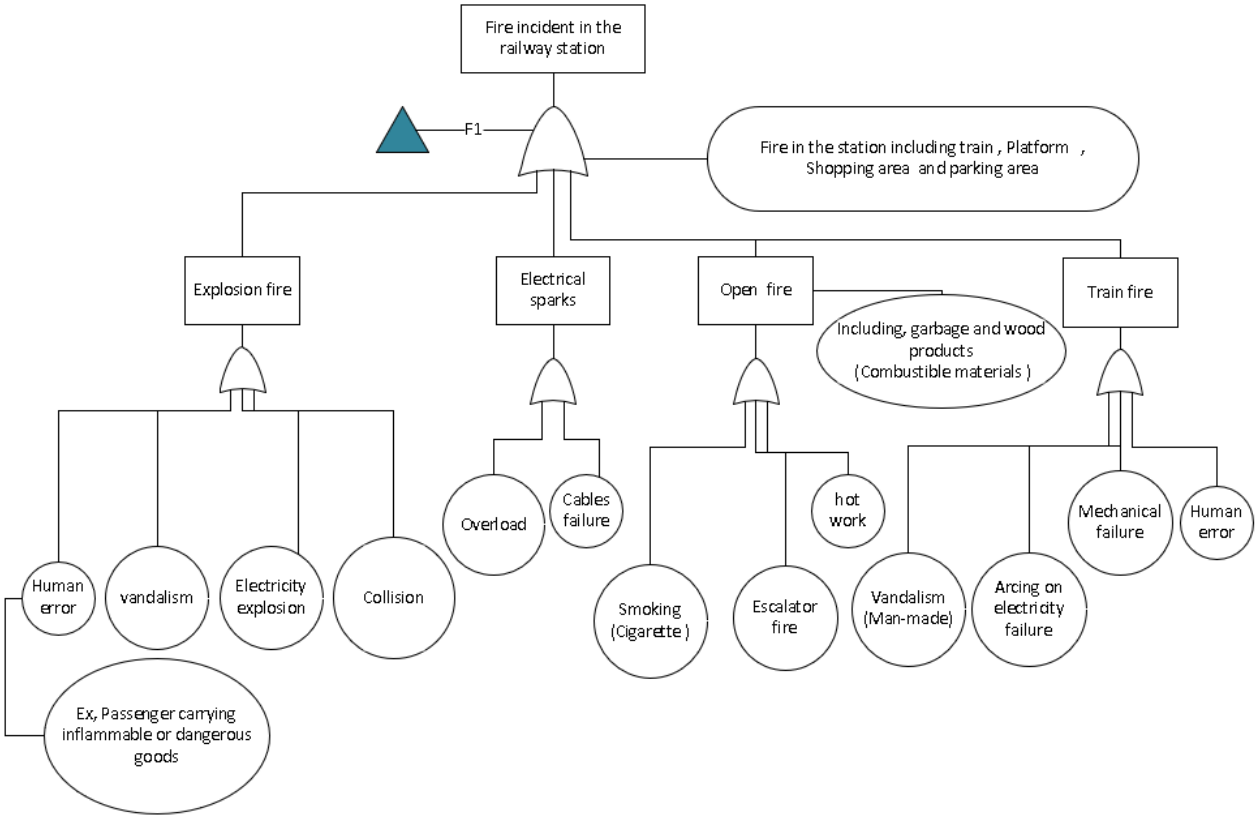


Figure 19. FTA of the fire event in a railway station.

The train station has some systems for emergency suppression in a fire event, and this system is an essential part of the fire safety plan. The suppression, evacuation, and smoke control systems are effective and risky at the same time in cases of failure. Full or partial system failure may increase the consequences of any expected incident, and the FTA for the systems tries to identify the root of failure, as shown in Figures 20-23 below, covering evacuation and failure of the system [141][143][144].

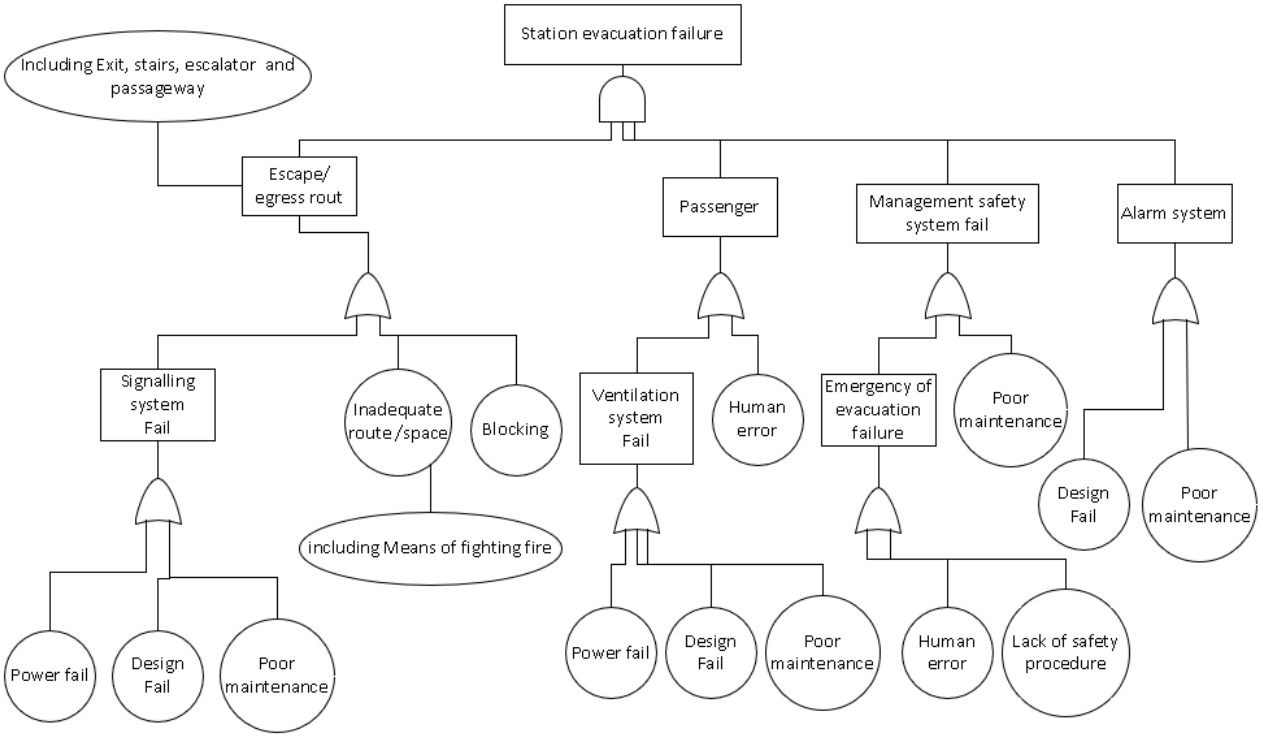


Figure 20. FTA of the evacuation failure in a railway station.

The safe evacuation is an essential response for the fire risk and for emergency procedures, which protect the passengers, workforce, public and provide access for fire fighters and emergency agencies to carry out their assistance like fire rescue services. When a fire occurs at the station in locations such as the platform or on the trains or in the shopping area, it is expected that the people will panic. In this situation, there are many factors that need to be controlled such as the passenger flow, gates, escalators, elevators, alarm systems, smoke control, fire systems and the other emergency measures must be activated. Also, other actions need to be taken related to the trains, such as stopping departures or those arriving or others that will be entering the station, as well as considering the train schedule and the consequences to other stations and the whole line.

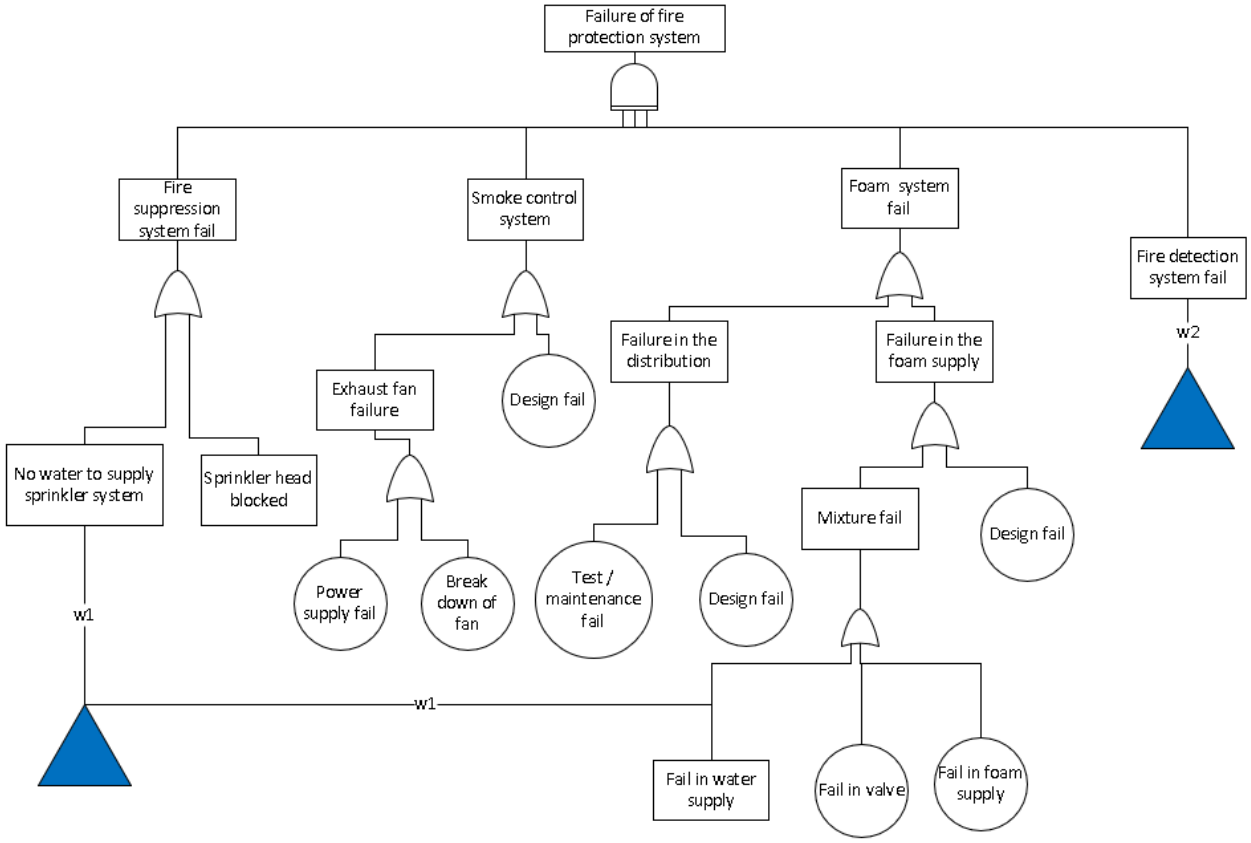


Figure 21. FTA of the protection system failure in a railway station.

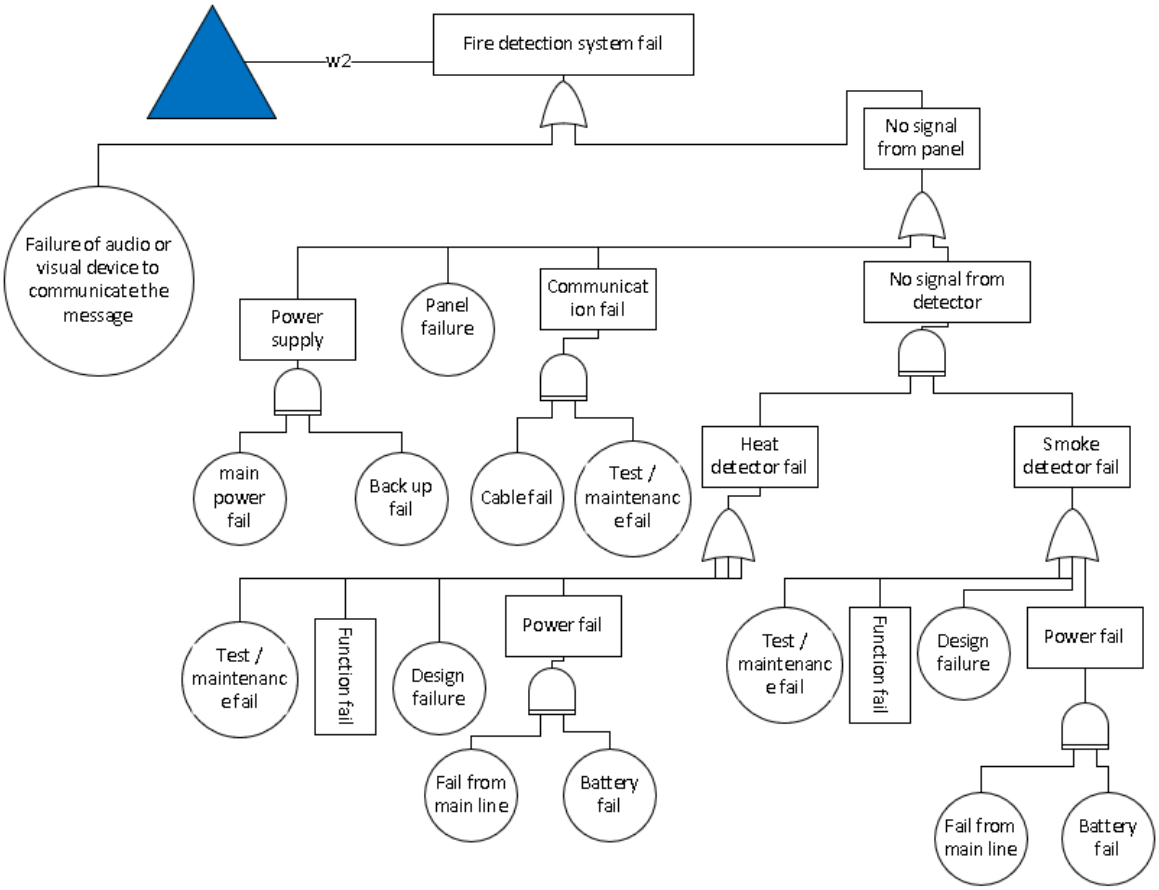


Figure 22. FTA of the detection system failure in a railway station.

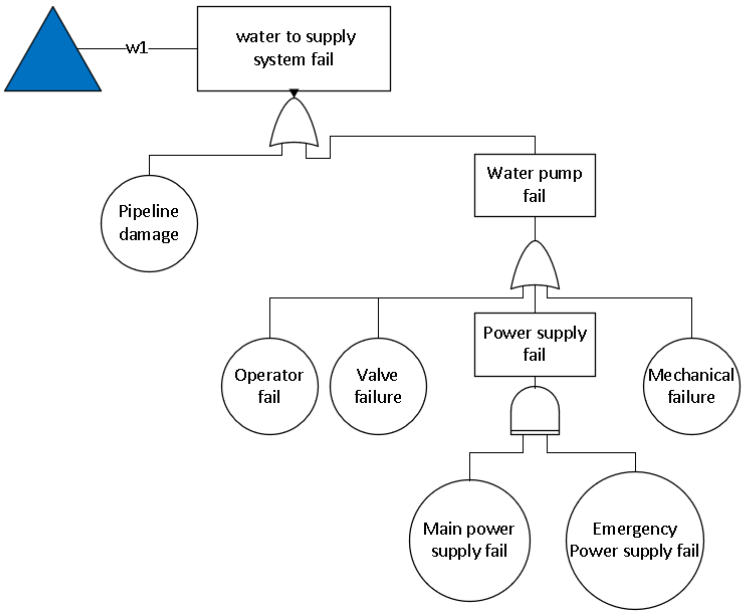


Figure 23. FTA of the water supply system failure in a railway station.

The fires in the stations have been determined by three causes: electrical failure, faulty mechanical equipment, and vandalism [145][7]. The outcome of uncontrolled fires is significant, and linked with other risks, as people may suffocate from the smoke, stampedes, earthquake, security, train delay on the line, impacting the adjacent areas and the public transport linked to the station.

The fire risk and the consequences lead to a disaster in the station, thus, the detection of the risk in the early stages is significant as an active strategy and monitoring the fire systems and equipment are also important as passive fire protection systems. However, human error will occur again in inspections of the systems and the delay of detection. This analysis shows the complexity of the systems in the station and how many aspects need to be active all the time in fighting such risk. Many aspects in the system need to be tested and checked by specialists in the fire and maintenance departments. This risk is linked with other threats such as security and evacuation failure, that make managing the risk in the stations challenging and hard to control in the larger stations. The high-density design challenge forms complex and hard responses to fire and explosion risk. However, the traditional fire risk analysis methods have encountered difficulties in search and analysis and can only use discrete variables for static analysis. Even though neural networks have been suggested for mapping fires, wildland, forest fires and supports and planning, these are all still influenced by human factors where experts' opinions are concerned and not widely used. The results from neural networks have limitations and cannot be generalized [146]. Such a risk is so complex in terms of managing the stations and the decision-maker which relies on human intervention, and this is all expected to improve with more technology implementation, as suggested in this work. Nevertheless, this risk needs a scenario and experiments, gathering data from a related system like detection, fire system and then integration to acting and warning systems in the stations which resulted in this not being chosen as a method in this work. Finally, there is the necessity for more research, real-time detection and monitoring the station systems with minimum human errors and interventions.

4.3 Slip/trip Risk

Falls due to slipping are statistically the main cause of accidents on crossways in built environments and railway stations. From the biggest and busiest 20 railway stations slip, trip, and fall events account for the largest portion of recorded accidents (Network Rail, 2020) [147]. Moreover, RSSB yearly Safety accomplishment Report 2017/18 announced that the number of major injuries caused by such risk in stations rose from 154 events in 2016/17 to 202 in 2017/18 [148]. Slips/trips formed one of the main causes of injury (Figure 24), and information about all risks of slips/trips at the railway station is described using FTA in Figure 25.

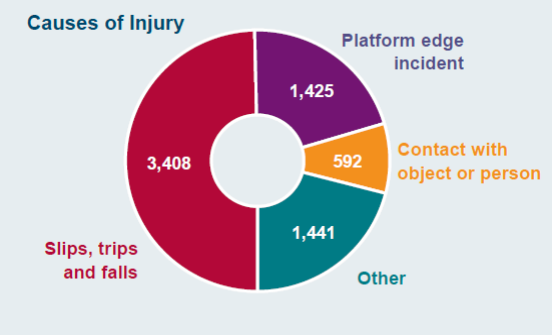


Figure 24. Slips/trips formed one of the main causes of injury (Office of Rail and Road (ORR), 2017) [149].

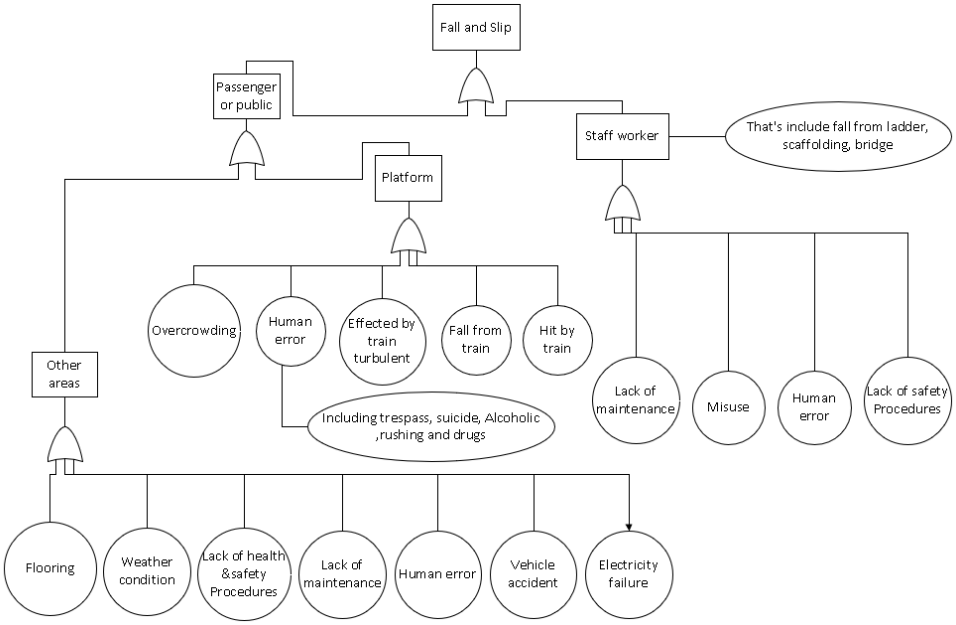


Figure 25. FTA of the slip/trip risk in a railway station.

This risk is common in railway stations, also overlapping with aspects such as crowd management, station design and other risks, which present the difficulty of maintaining timely control in the case of huge railway stations. Escalators, stairs, platforms, the boarding and alighting areas are common locations where the accidents occur, and stairs/escalators, platforms are all high-risk areas [150]. Many measures have been implemented, such as signage, step markings and slip resistance floor, handrails, and warning sounds, yet the accidents still are recorded. Many factors can play a role in such complex risks as overcrowding, passenger flow, luggage, lighting, weather conditions, overcapacity, and passenger behaviour.

Moreover, many accidents occur when the network is overcrowded or almost deserted. However, the platform's incidents– train–tracks interface have been insufficiently investigated, and existing studies do not provide strong conclusions about the actual contribution nor the efficiency of preventive measures. Consequently, the development of a systematic process, including the details of the accidents and data history, would provide a beneficial framework for improving research on safety and a powerful assistance tool for managing accidents at both national and international levels [151]. Finally, even with a challenge to data, this risk has been selected to be analysed utilising the suggested concept of AI.

4.4 Train derailment risk

Train derailment is one of the major accidents in the rail transit safety system, which is an event where the train comes off the rails due to an unexpected event. Moreover, a derailment can occur in many diverse locations on the line, such as at the station, and it is likely to lead to a collision with other trains, passengers, or the platform structure [152]. Main events linked to the top event of the FTA are aggressive signal, natural disasters, and engineering structure damage. The sub-events are shown in Figures 26.

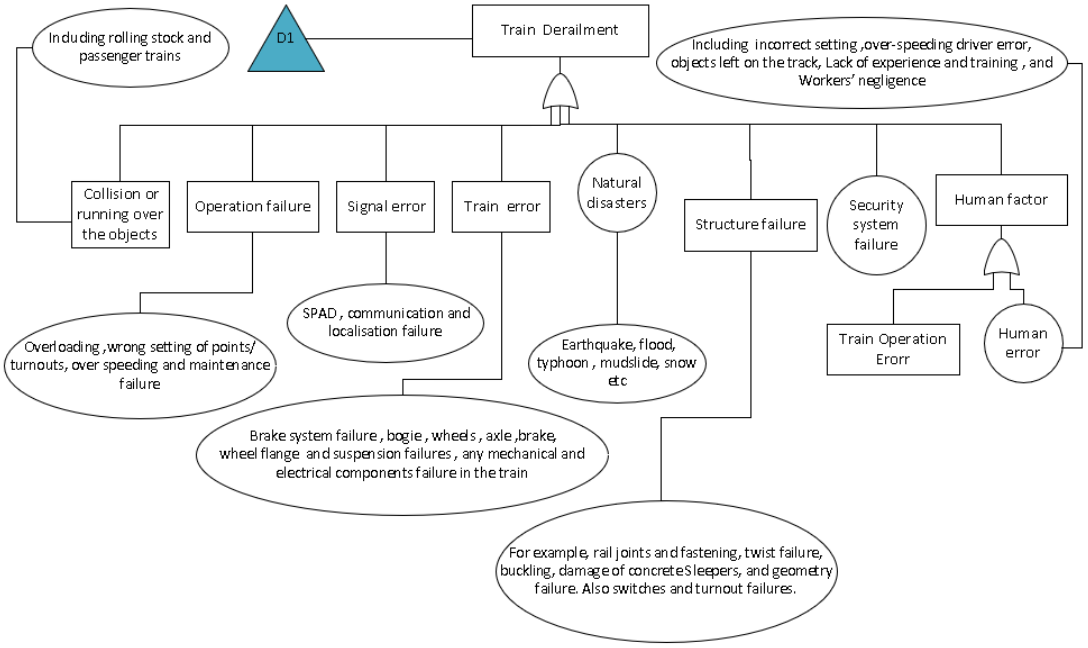


Figure 26. FTA of the train derailment risk in a railway station.

Terrible consequences might be expected with derailment accidents if they occur at a peak hour when the trains and the platforms are crowded, and there could be even more loss of life and injuries with extensive damage to the infrastructure causing a delay to the operations in that station and on the whole line. One example of this was in the UK, the Potters Bar station train derailment in 2002, which led to 7 people losing their lives and over 70 people being injured, some seriously.

So, it is significant to predict risks and manage them effectively for such safety-critical components in the systems which may lead to the circumstances of the derailment. Many systems involved in this risk such as maintenance, signal or the weather effects as well as human error are still present in each risk in the stations. Derailment can be extremely hazardous, and many scenarios can be expected as a result of various reasons where the outcomes may worsen if the derailment leads to falling on the adjacent track, rear-end collisions or colliding with people or cars, or other trains. All of these events lead to passenger casualties, injury and fatality, as well as fire, damaged infrastructure, and disturbances on the whole line. In general, the causes must be related to one or more of the following

aspects: human, machine including infrastructure, environment and management including operation and maintenance.

In the last few years, the planners have been trying to add more passenger trains on the line and freight to meet the increased demand, therefore increasing all risks. In fact, the analysis of the train derailment causes conducted in the literature are taken from many different perspectives and in many countries, and two derailments are never the same [153][154][155][156].

The FTA of the train rear-ends collision accident has been conducted which may lead to the derailment [157] (See Figures 27-29).

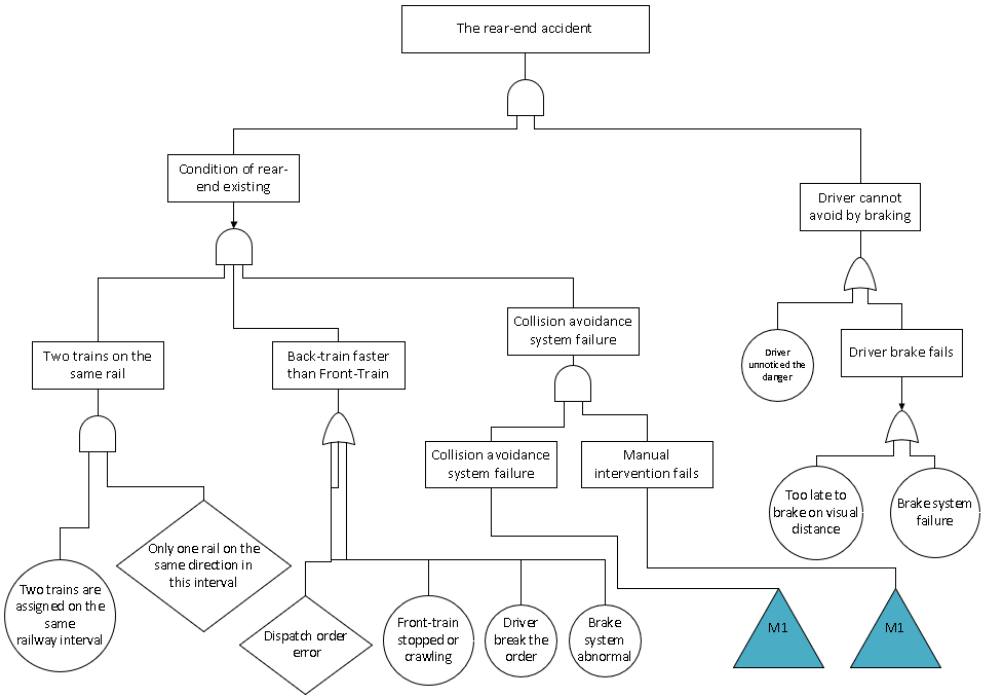


Figure 27. FTA of the train rear-end collision accidents [157].

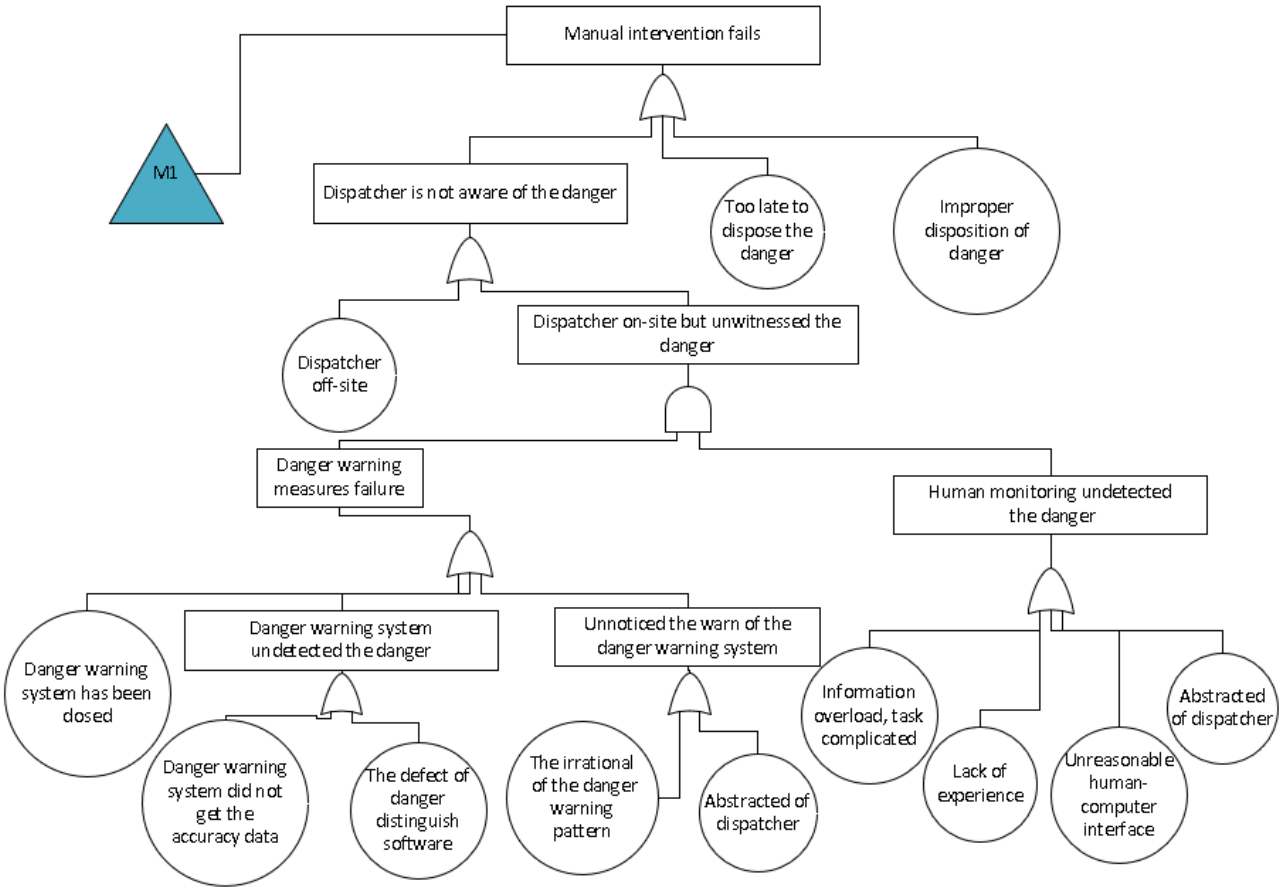


Figure 28. FTA for the manual intervention fail sup tree of derailment.

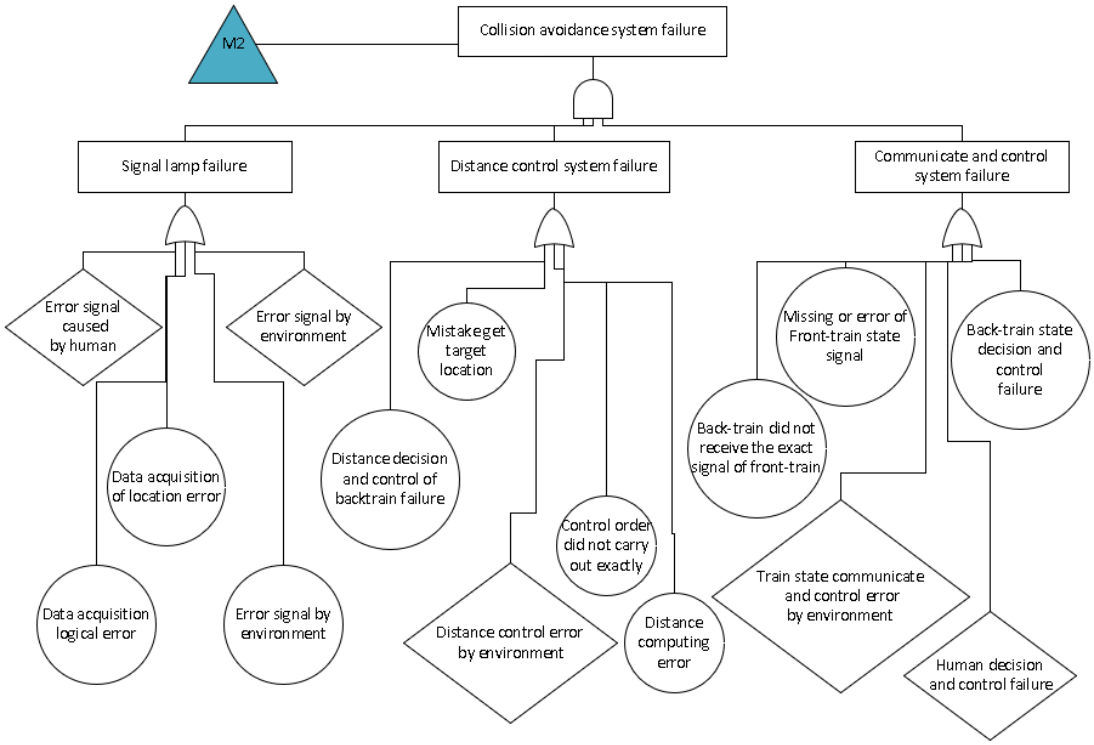


Figure 29. FTA for the Collision avoidance system failure sup tree of derailment.

Moreover, the most prevalent accidents cause group of Freight Train Derailments by Track Type are listed below, which shows distribution differs by accident type, track type, and speed:

- Defective welds or rails
- Defective welds or rails
- Bearing or wheels failure
- Train handling
- Wide gauge
- Obstructions
- Train interaction
- The axle or journal defects (train)

Also, Track and equipment have been highlighted as primary causes, and the use of switches and switching rules have a substantial effect on derailment rate and the speed factor which affect train

accident rate and accident consequences [158]. Exceeding speed limits on the curves will lead to derailment, and any objects on the track will be high risk whereas the speed is low in the stations, such a risk inside the station is rare.

In many cases, paths have been proposed and they can be overlapping, so for instance, the derailment of a train can be caused by a collision and vice versa. In general, some causes can be described as human factor (driver-maintenance -operation) error, external cause (weather - cross-level), or machine failures such as machine of the train and its components, track and components, or communication, and signal system. Others have indirect impacts such as leaves on the lines, which can reduce the grip between the train wheels and the track and make distances of braking longer. However, this risk often occurred on the lines out of the stations and has various contexts beyond the research scope, for example, the objects on the track may lead to derailments, and these objects can be from many causes such as maintenance, the track, the wheels, previous trains, natural effects, structure, collapses or other aspects related to the human factor and security system (vandalism) and so on.

From the analysis of train derailment cause and outcome and implications for regulation, innovative safety interventions are recommended to improve risk controls [159].

The future improvement in the field with technology such as autonomous inspection, Sensing Technology and Railway and ML applications are expected to play an important role in the reduction of accidents and enhancing safety [160][161]. Moreover, technology enhances the safety of the railway stations by enabling reduced derailment accidents and lessening the chance of human error which might be caused by negligence in the operations. Furthermore, advanced analytical video surveillance can cover a range of risks, such as collisions, derailments and intrusions from adjacent areas into unauthorised station locations such as a track [14][162][20]. Even with the current lack of AI application in the field, AI methods have been proven to support the modelling, store and analytics

data and assessment and transfer of knowledge about railway safety and such accident hazards of derailment [44]. This risk leads in general to accidents out of the stations, which had intensive analysis in literature but which meant it was not selected for examining AI in this particular study. Finally, the timely data from trains, infrastructure, and environments with the details of the components' condition will be essential to designing an active protection system with predictions with low human intervention.

4.5 Security system failure risk

Increasing the security at the railway station is essential. The failure in the security system increases with illegal or fake passengers, and the failure of security at the railway station increases the attack risk according to the FTA for a security system failure risk, as shown in Figure 30. This expected failure of the security system will also increase overcrowding in the station or create a huge mass over the platform [163].

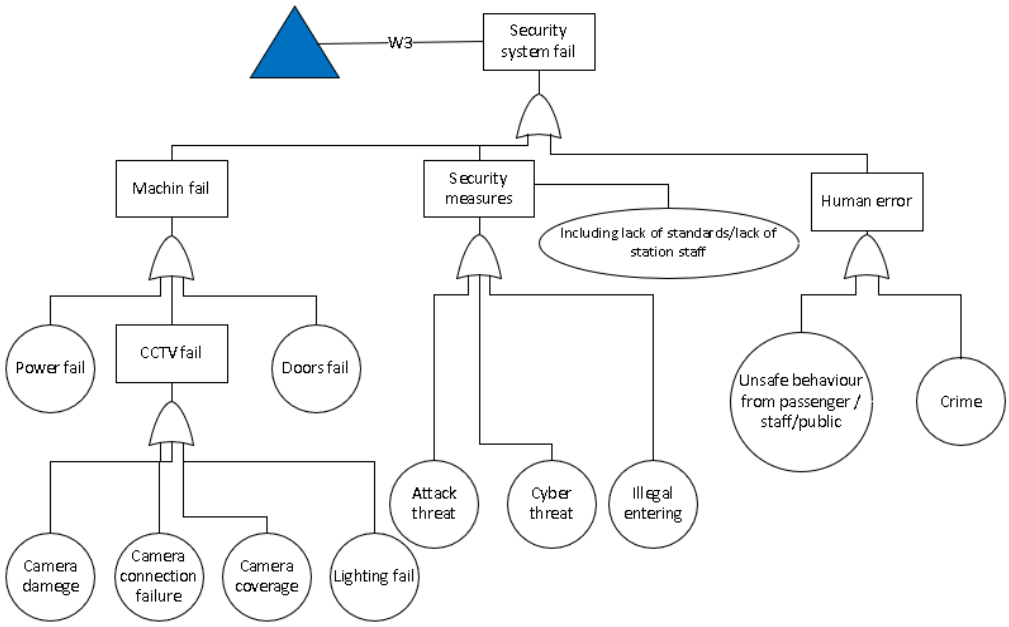


Figure 30. FTA of the train security system failure in a railway station.

A station is a busy location having restricted contact with the outside world, and its social, economic, and political importance makes it a target for attack threats (terror). Terrorism is defined as violence such as explosions, shootings, hostage-taking, or intimidation with the goal of posing a threat to the country and instilling fear. Many people can be in danger as a consequence of such an attack in a critical structure such as railway stations. Moreover, the impact of attack threats are not limited to just the station but extend to the business and society and the whole rail system. There is no doubt that intelligence agency can be an early protective barrier also along with detection and monitoring at an early stage, which is crucial for minimum loss and for evacuation and rescue services [164]. The terrorist attacks against the railway industry include trains, stations and tracks, but in this research is focused on the stations, using a list of the stations around the world that have suffered from terrorist attacks, as shown in Table 8, and which have been extracted from many databases [165] [166]. The analysis of scholars showed that terrorist attacks were committed with explosive devices and toxic substances. A development in protection is needed for early detection to ensure complete inspections with strategies for protection against terrorist attacks in the stations.

Table 7. List of stations that faced terrorism attacks around the world from 2000 to 2016 [165] [166].

Date	Country	Attack	Fatalities	Injuries
29.03.2010	Russia	Bomb	47	102
22.04.2010	Thailand	Bomb	0	3
07.05.2010	Russia	Bomb	1	7
02.02.2010	Russia	Bomb	0	1
19.11.2008	Pakistan	Bomb	0	1
11.03.2009	Ukraine	Bomb	0	7
16.10.2007	Thailand	Bomb	0	1
10.05.2008	India	Assault	3	0
27.06.2008	India	Bomb	0	8
13.09.2008	India	Bomb	6	12
22.01.2007	India	Bomb	0	3
02.10.2004	India	Bomb	12	40
07.01.2005	Thailand	Bomb	0	3
07.07.2005	UK	Bomb	56	700
19.02.2006	India	Bomb	0	25
06.07.2000	Russia	Bomb	2	1

06.10.2000	Russia	Bomb	1	6
25.12.2000	Pakistan	Bomb	0	3
05.02.2001	Russia	Bomb	0	20
07.08.2001	India	Assault	12	30
09.09.2001	Israel	Bomb	3	20
31.08.2004	Russia	Bomb	11	50
03.02.1998	Slovakia	Bomb	0	3
27.02.1998	India	Bomb	5	20
28.06.1999	Russia	Bomb	0	11
01.07.1999	Russia	Bomb	0	11
22.03.2016	Belgium	Bomb	14	70
11.04.2011	Belarus	Bomb	15	204
27.07.2000	Germany	Bomb	0	10
04.09.2001	Canada	Bomb	0	45
11.03.2004	Spain	Bomb	192	2050
18.01.2016	Germany	stabbed	1	5
01.11.2015	Turkey	Bomb	0	5
01.05.2014	India	Bomb	1	14
06.02.2004	Russia	Bomb	41	120
13.03.2003	India	Bomb	10	70
27.01.2003	India	Bomb	1	28
30.11.2000	Philippines	Bomb	22	120

The railway stations have been designated as a soft target because they are accessible, contain a high concentration of people, and have a minimal level of security, making them an appealing target, particularly for terrorists. The following asset groupings were identified as being exposed and so vulnerable to terrorist attacks in the stations, with each asset group further subdivided into subsystems or components (see Figure 31 below).[167][168].Protecting the railway transport system which is a critical infrastructure physical asset is imperative [169].

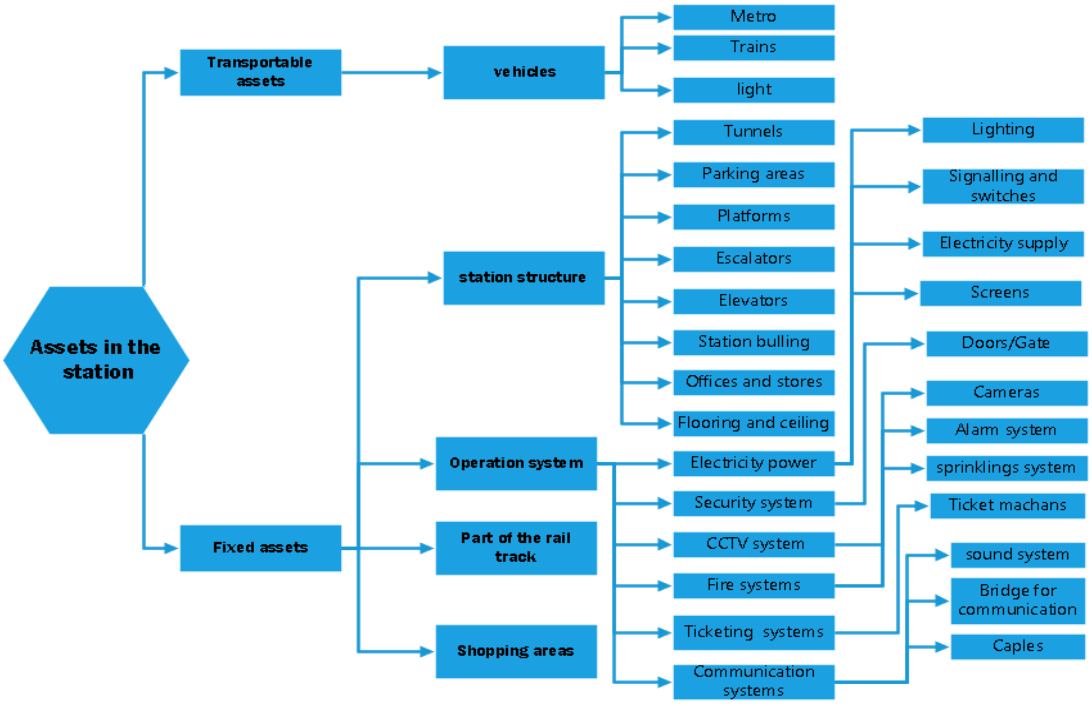


Figure 31. Asset groups in the stations which are vulnerable to terrorist attacks.

In fact, the security is overlapping with other aspects, such as cybersecurity, terrorist attacks, and vandalism and other issues such as violence or burglary. However, technology can provide an early warning and protect the data, assets, operation, and people at an early stage. From the new technology aspects, AI, UAV, and wireless sensor networks (WSNs) applications have been proposed for use in the railway including station systems, and counting advanced WSNs, which will enhance security, and monitor railway assets to achieve more cost-effective management, as well as the development of integrated systems [170][171][20]. With this growing usage of technology, AI can protect the railway operation in many aspects such as cybersecurity, privacy, authentication, mitigating attacks, and dependability, monitoring and managing the system in real-time [169]. AI/ML deliver safe, functional real-time, systematic, cost-effective, and predictive cyber risk data analytics [172].

A security threat such as terrorist risks can be disastrous and impact the infrastructure such as railway stations in the airports which could cause a serious loss of lives and deterioration of the economy and

infrastructure. Management of risk to critical infrastructure is key for prosperity in our modern society, and to a reduction of the probability of attacks and the potential losses following an attack [10]. Moreover, to protect the passengers, staff, public and assets, installation of security technologies and more security measures has been recommended [173]. It is important to note that modelling techniques should be constantly improved and should evolve in parallel with increasing the complexity of the systems and the intelligence of attackers. In future, smart technology will play an essential role in security improvements, providing higher security and scalability, and aiding decision-makers in managing risks to critical infrastructure systems [169][11]. Finally, it believes the nature of the railway makes the system exposed, and the intelligence agency and detection technology are critical for securing the stations, otherwise, the station will become similar to airports.

4.6 Overcrowding risk

It has been noted that in some countries there is a growing public travel market and the passengers raised their requirements for higher-quality service which is leading to great challenges to operations and management. Many factors play an essential role in the passenger movement from, in and to the stations such as the accessibility, the culture, location, the time and so on. These factors can affect the safety of the operations in the stations. In the daily operation of stations, this means the stations are frequently crowded, and this is more often the case in the transfer stations are involved with multi-rail transit lines [150].Basically, operations require more robust management in order to face the challenges at such a critical point (Railway station) of the line which can have a widespread influence on the entire system. Furthermore, the heightened affair among public means that small operations and un-wanted events can incur a great deal of public consciousness, and the negative impressions can be amplified by the social networks if the matter is not appropriately addressed in real time. During un-wanted or unusual events in overcrowded areas like the railway station, when an earthquake or extreme incidents occur, people may fall into panic and then attempt to leave from their

locations. One of the consequences of that accident or any un-wanted event leads to stampede accidents which are shown as difficult for managing such flows of the crowd. Examples of the stations around the world suffering from overcrowding and its consequences can be seen in Table 9 below [174].

Table 8. List of stations that faced overcrowding accidents around the world from 1999 to 2018 [174].

Year	City	Cause	Impact
1999	Japan	The number far exceeded the station's capacity	76 Japanese died.
1943	UK	A woman fell down the stairs	Crowded and falling with an estimated 173 dead.
1999	Belarus	Slipping and falling on the wet pavement, trampling the others.	54 dead, 150 injured.
2004	India	Human stampede	5 dead and 11 injured.
2007	India	Stampede	At least 14 ladies were killed (crushed), and another 40 were injured.
2010	India	Stampede	2 dead and 8 injured.
2013	India	Stampede	36 people died and 39 injured.
2017	India	Stampede	Hundreds were hurt, and at least 22 individuals were killed.
2018	India	Earthquake caused panic	At least 58 people were injured.

In fact, some stations have small space, narrow tunnels, limited capacity of escalators and exit doors, and thus the chance of a stampede accident is increased with the possibility of economic loss and serious injuries and fatalities [106][175][176]. Moreover, it has been noted that, the main kinds of accident occurring in the metro system are the crowding and stampede. Accordingly, conducting risk analysis of the vulnerability of railway accidents is essential and will limit overcrowding and stampedes occurring in the stations [177]. This risk is linked with the emergency response, from other risks such as terrorist attacks and violent incidents. Obviously, to avoid this risk or eliminate it, it is vital to raise the awareness and design of an active reliable management system in real time to support decision-making [174] [87]. The FTA for overcrowding risk is shown in Figure 32.

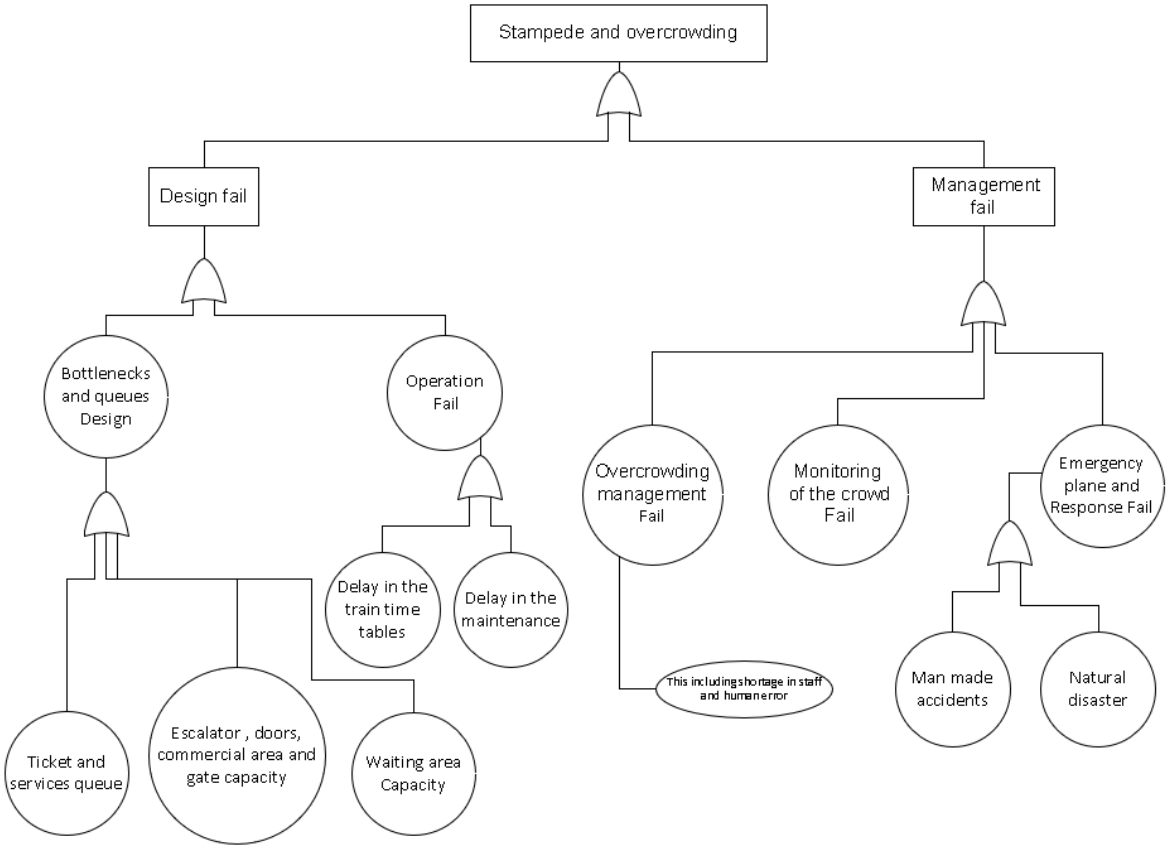


Figure 32. FTA of the overcrowding risk in a railway station.
https://www.sciencedirect.com/science/article/pii/S0925231219306678

In daily operations of some stations, crowded is always, and it is higher in the transfer stations that occupy multi-rail transit lines. The physical environment of the railway stations seems to have a crucial effect on rail passengers' perceptions of safety and security. The crowding is often linked with the thought of panic, crime, risk or may induce feelings of discomfort. This can influence rail passenger behaviour, leading to a reduction in ridership and forming the social conditions of railway stations as risky areas. Smith and Clarke (2000) [178] highlight the connection between crowding, criminal activity and the rate of use of transport. Teague and Leith (2008) [179] further describe a connection between crime and crowding.

Crowding can make it easier for opportunistic offending (for example, pick-pocketing) and less privacy (Uzzell & Brown 2007) [180]. Gershon et al. (2005) [181] point out that the panic of crime can change 'ridership', and so stations can become less utilised, therefore adding more risk to the

people. In an Australian study, Macintyre and Homel (1997) [182] noticed a rise in violence in conditions where crowding happens.

It was discovered that cases of crime were correlated with the design, leading to poor pedestrian flow patterns. Thus, they propose that the intermediate variable between people's experience and their violent behaviour is stress associated with it being crowded. The analysis of the overcrowding illustrates the need to identify and map out the factors of crowding and provide a measurement to expectations and responses to crowding conditions in the stations, and that passengers' reactions to crowding can be mitigated by timely and effective crowding management [183].

This risk is common in many systems worldwide; it is expected to lead to falls, crime, an unhealthy environment, passenger discomfort, and delays. With the growth of passenger and service capacity, it is expected to continue with overcrowding and safety concerns increasing. Even though the number of trains increases, the station design has not been considered to comply with the capacity. In fact, sometimes, the overcrowding is not expected, for instance, group arrivals for events or festivals. Moreover, full trains mean no more passengers can get on the trains, leading to crowding in the platforms and standing near the edge of the platform in risky positions. The platform design has highlighted the impact of the accidents, such as the curved shape or non-standardization locations. Also, delays or cancelations of trains for any reason lead to accumulating passengers and stations becoming more crowed. The corridors, lifts, escalators, stairs are means for passenger flow to the platforms which are still open and allow more and more passengers to come on to the platforms even if they are already crowded. The employees working in the station cannot usually cover the whole station all of the time. Thus, utilising technology is important in managing such overcrowding throughout day.

On a daily basis, improved design and management of passenger flow significantly reduce risk [184]. Manging overcrowding is an essential daily operation in the stations and has become even more so a topic for attention since the outbreak of COVID-19 [185]. In order to defend against COVID-19, there should be as few people as possible in congested areas or as little human interaction as possible. As a result, technology and AI can play a key role in ensuring healthy and safe railway stations, which complements the smart railway station idea [186]. Finally, managing overcrowding through using technology is significant where the early detection is critical to prevent the consequences, take actions, warning decision-makers and acting in a timely manner.

3.7 Flooding risk

In the early stage of design years ago, climate change and the risk of floods were not considered in the stations' design, which are, in fact, vulnerable to rainstorms, flooding and extreme weather and increasingly so through global warming. However, most of the railway stations are located at underground level and may be placed within Environment Flood Zones. Water can enter the stations and tunnels via station entrances (such as emergency entrances) and tunnel outlets or from the failure in some systems such as flooding from sewers or artificial sources. In the United Kingdom, for instance, it has been noticed that many places on the London Underground stations are at high risk of flooding, and almost a third of stations on the London underground have been classified as being at risk of flooding. The design and the capacity of many sewers systems may not meet the capacity of the expected floods nowadays, as London’s sewers were built during the 19th century. The global changing of weather creates extreme climates such as heavy rain or rainstorms, and this is in addition to the rapid urbanization of some developing countries, raising the possibility of the stations being flooded. With so much water in the station and tunnels, such dangers will result in the station's shutdown or, highly likely, the entire line is out of operation. Also, other risks such as fires may lead to water in the stations or water from fire-fighting activities. The potential impact of flood risks to the

station is extended to the construction, people, Emergency Response and Evacuation Planning, cutting the internet and the energy power, economic losses and cancellations and delays on the system operations. The flood which leads to disruptions in a railway system can have critical outcomes, such as direct damage and indirect loss [187]. From the total losses due to all-natural hazards in the globe, the flood forms at least one-third, which is described as one of the most important natural risks [188]. In China, which is a country at risk of flooding, the statistics show that for ten years from 2000, the medium spending on flood-related maintenance in the railway system was about $15million per year, and the indirect loss caused by railway disruption was about $37.5 million per year [189]. Moreover, this risk raises the electric shock risk and leads to damaging the infrastructure in the station which needs time to reopen and start safe operations and increases the cost of maintenance. The list of flood catastrophes at rail stations in various cities is shown in Table 10 below, and FTA for flooding risks, as shown in Figure 33.

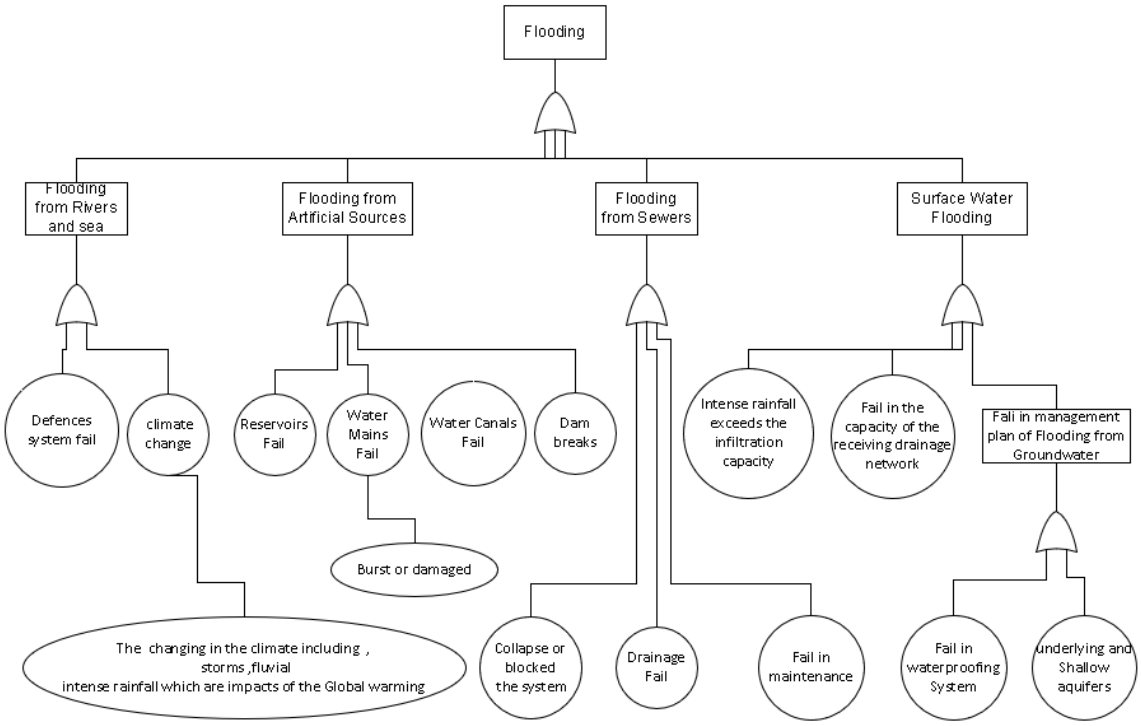


Figure 33. FTA of the flooding risk in a railway station.

Table 9. Stations flooded around the world from 2001 to 2018 [7].

Year	City	Cause	Impact
2001	Taipei	Typhoon	The water crossed into stations and shopping space through the tunnel.
2002	Prague	Rainfall	Around 20 km of lines and 19 stations were flooded.
2004	New York	Hurricane	Water in stations, many lines were out of service.
2007	New York	Rainstorm	The subway system massively flooded.
2008	Beijing	Rainstorm	Line 2 was closed.
2011	Beijing	Rainstorm	A wire was covered in rainwater, creating a short circuit.
2011	Nanjing	Rainstorm	Water flew into the tunnel.
2012	Beijing	Rainstorm	Water caused disruption to work temporarily.
2012	New York	Rainstorm	Several subway tunnels were flooded.
2014	Busan	Hurricane	Several sections of line1, line2 and Line4 were flooded.
2014	New York	Rainstorm	Some subway stations flooded.
2016	Guangzhou	Rainstorm	Through the entrance, the water flowed into Changban station.
2016	Wuhan	Rainstorm	Many stations of Line 4 were flooded.
2018	New York	Heavy rain	Line re-routed and trains skipping the station.

The potential severity of the flood impact depends on many factors such as the flood source, the station location and design, and the early detection with advanced management for such risk. Flood risk needs to be systematically estimated where the magnitude of the water and consequences of such hazards are predicted, and the endangered assets and people, and how vulnerable railway stations and people are given certain hazard conditions [190]. Furthermore, for the future it is expected that global warming will lead to more floods impacting the railway system including stations, for example, London [191], Malaysia–Singapore [192] and China [189]. With accurate forecasting in the weather, it is possible to understand the weather early. However, the need for precise data between organisations requires improvement. The resilience of the infrastructure and global warming are challenges for the railway station infrastructure. However, the data-driven methods, including BD, sensors, IoT, satellite technologies, and machine learning algorithms, have a potential opportunity for predictions, and flood disaster response, facilitating a process of evidence-based decision-making and re-evaluating flooding risk consequences, which cover rail, and other key infrastructure [193].

4.8 Earthquake risk

The station facilities are an integral part of the infrastructure of modern society, and it is a part of cities' economy and transportation, including over ground and underground, highways, material storage, and sewage systems and water transport. Historically, some underground structures have experienced significant damage in large earthquakes, including the 1995 Kobe, Japan earthquake, the 1999 Chi-Chi, Taiwan earthquake and the 1999 Kocaeli, Turkey earthquake. The station structure could be severely damaged in a strong earthquake and may lead to major damage to the assets and the people, where some stations have a very complicated design including many floors and tunnels which make the evacuation and rescue service a difficult mission. Some countries have frequent earthquakes that are widely distributed, the risk will increase, and the earthquake will damage more infrastructure in that area which has very serious weak cementation and a special sensitivity to water. Thus, the seismic response analysis of the stations is of great significance [194][195][196]. Moreover, from the historical earthquake events some lead to damage to the structure that completely collapsed due to the earthquake, which formed a turning point in that underground structures have been considered to be relatively safer compared to structures on the ground. For example, in Kobe in 1995, one of the most serious cases of damage was found in the underground structures as six underground stations out of the total 21 stations in the Kobe area were seriously damaged by the earthquake. Furthermore, some regions of running tunnels connecting with the damaged stations and the tunnel section railway sustained damage. The Daiichi Station was the most severely damaged. In the station, more than half of the centre columns totally failed, and the ceiling slab bent and fell down [197]. If there are no prepared plans for such risks (earthquake) which may suddenly occur, the consequence is more casualties and people falling into panic. Typically, more minor earthquakes may only force the trains not to operate, however, the large quake could create fallen columns, damage infrastructure and even the failure of the whole station. Moreover, it is expected there are no indicators for such

risk. Table 10 shows some examples of earthquake-affected systems, and the FTA for earthquake risk as shown in Figure 34.

Table 10. Cases of railway stations affected by earthquake [7][6].

Year	Magnitude	country	Consequence
1985	8.1	Mexico	All lines of the metro were out of service.
1995	7.3	Japan	Many stations damaged, about 3km of tunnel collapsed.
2003	7	Japan	Sendai Subway was out of service.
2010	8.8	Chile	Some damage at entrance to the station.
2014	5.9	Taiwan	Taipei MRT was out of service.
2015	5.6	Japan	Tokyo subway was out of service.
2015	7.8	Afghanistan	Delhi subway out of service, some people trapped.

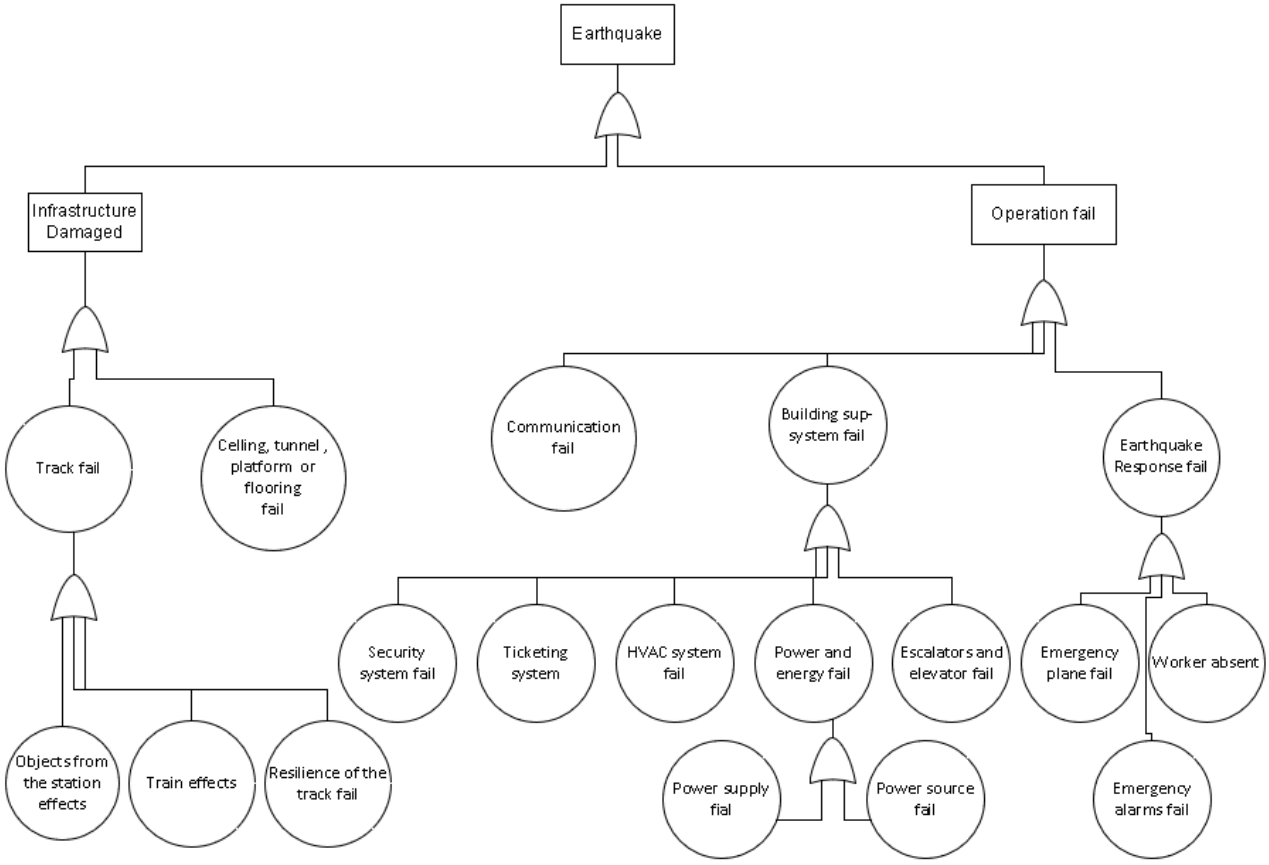


Figure 34. FTA of the earthquake risk in a railway station.

As some buildings and infrastructures can be exposed to natural hazards such as severe earthquakes, the smart sensor network is effective for such risk monitoring. Also, it is essential to get predictions before an earthquake, and gain benefits from the (AI) and (IoT) techniques to enhance the seismic performance of structures [14][198][199]. Comprehensive contingencies for earthquakes of different scales are required so that individuals can be better prepared in the event of an earthquake, and operators must take actions for people, assets, and operations. In addition, information regarding how to survive an event must be disseminated to the general public in a number of congested areas such as platforms, trains, elevators and tunnels, in high-level l districts with this risk. Finally, early warning from the responsible authority to the decision-maker in the station is a key safety measure as it is prediction.

4.9 Risk of transporting dangerous goods

Globally, thousands of tonnes of hazardous materials are transported by train. Railways are frequently used for the carrying of dangerous commodities such as toxic compounds, chemicals, petrol, liquid gases, nuclear waste, and other dangerous goods. However, if these trains collide in the stations, it will be a disaster, causing accidents such as fires, explosions, and hazardous and environmental repercussions, so such a scenario can lead to other risk accidents (see Figure 35). Even with the low proportion for this risk, it requires a number of measures to consider, like train structure design, emergency plan and evacuation, freight train maintenance, routing away from crowded stations, and ensuring the safe transport of dangerous goods via the stations. However, if trains carrying dangerous commodities are kept in stations for any reason, this risk goes up.

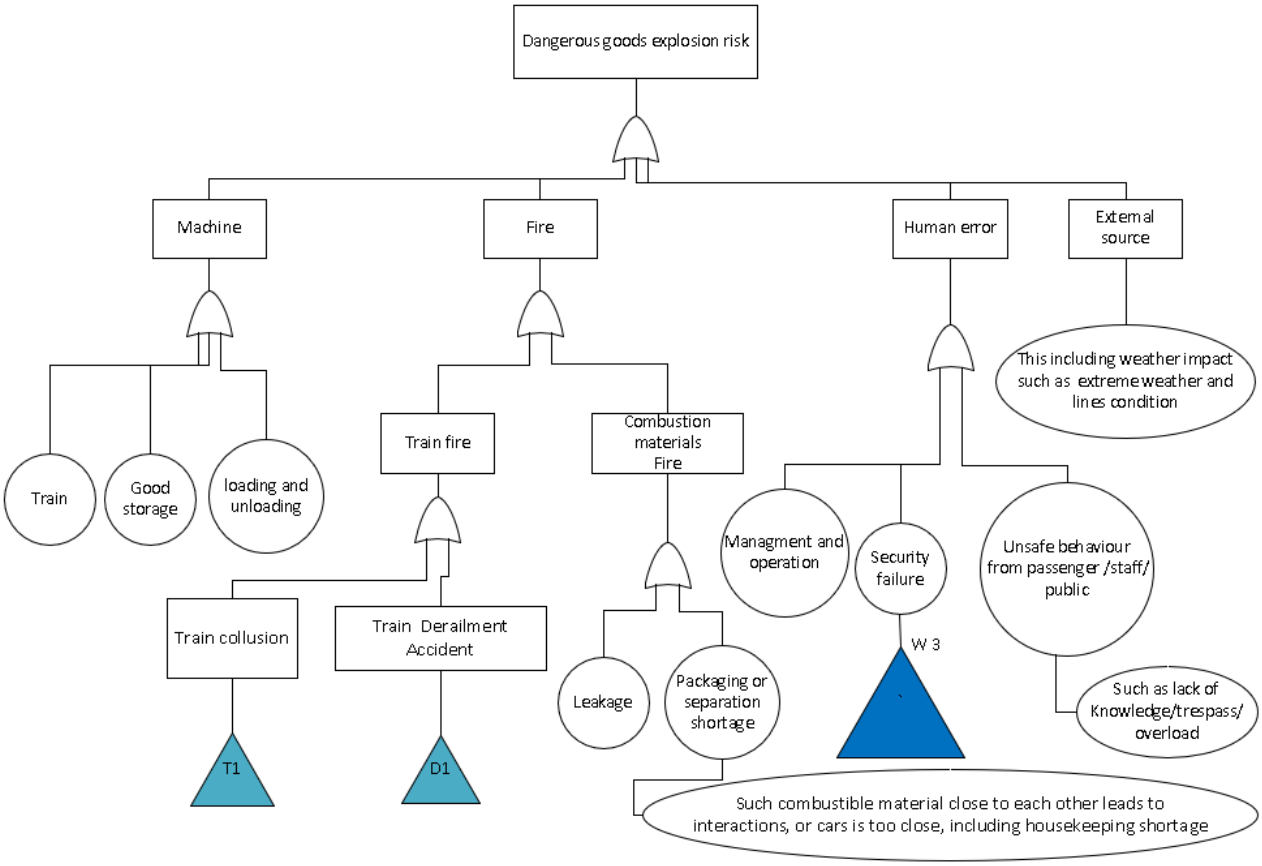


Figure 35. FTA of the dangerous goods affect risk in a railway station [200].

Rapid economic development, the demand for dangerous goods via rail (e.g., natural gas, radioactive materials, and flammable and explosive materials), which may form risk, raise the necessary importance of effective risk management method. Moreover, the process of the risk process in such a case has issues including analysing the correlation between the risk and cause factors that are dynamic and nonlinear interactions, and calculating the weight of the risk; thus, traditional methods cannot handle this confusion and uncertainties [201]. Thus, for operational processes active monitoring and real-time method transportation of dangerous cargoes have been proposed [202]. Moreover, from transport of dangerous goods risk control strategies analysis, it has been recommended that machine learning (ML) and AI be employed to reduce the number of incidents and enhancing the transport safety of dangerous goods [203].

4.10 The disruption effect

This part includes the analysis of the risk to present the consequence of the risks and the importance of disruption management. The analysis of the factors that cause the incidents may aid the decision makers to moderate the negative outcomes (i.e., delay, disruption) emerging from incidents in the stations such as operation errors, power failure and signal failure. Transport disruption can be defined as an event that leads to significant re-planning of a transport service (Pender et al. 2012) [204]. To ensure operational continuity and customer satisfaction in the stations, it is significant to do rail disruption management. The application of disruption management would be expected to diminish the influence of the un-wanted events on the station, such as the other trains on the line which may become stranded and manage the wider service disruption that may occur as a result of the accidents. Disruption can lead sometimes to psychological trauma for railway staff, passengers, and members of the emergency services. If a station is disruptive, passenger flow between nearby stations will be affected, and geographically isolated stations are more vulnerable. Moreover, rail connections between these stations will negatively influence the whole system [205]. The incidents in the station may lead to disruption and then financial costs, stopping the traffic and business reputation [206]. Moreover, the natural incidents like the flood cause disruptive events in the railway and are a link with other vulnerabilities [189]. There are many characteristics around incidents that create a variety of rail disruptions, and resulting from that, a group of events can cause a complete blockage to the service, such as fatality accidents and overhead line failure. These characteristics reflect concerns around management and predict the incidents' influence.

Unwanted events such as train failures or overhead line de-wires due to extreme weather are a daily occurrence on the railways. Accidents such as these may take hours to determine and cause vital delays and disappointment to passengers and freight customers [207].

4.11 Structure failure

The nature of large station infrastructure and the layout is complex which may contain many levels, tunnels, level crossings, bridges, long side platforms, parts of the track, gates, power system rooms, and many others. Thus, the railway stations are complex. Some stations that are more critical may integrate the city transport centre, an airports terminal, and commercial malls. The nature of station construction (multilayers, crowed clusters and openness) leads to various kinds of risks [208]. The railway station building has facilities, and much of the materials are likely affected by many factors. Also, some railway stations were built a long time ago, and materials were used, which did not fit the current standards and demands [209].

The capability of railway structures to handle issues arising from events such as fire or collision is vital to the safety of passengers on trains and in the stations and of the pedestrians or staff using the structures. Failure in the railway infrastructure impacts the system operations; thus, it needs early condition monitoring to detect deterioration [210]. Moreover, it is critical for employees to learn the infrastructure risks and evacuation principles fully in the railway stations [211]. In the FTA in Figure 36, four suggested elementary events affect the structure: extreme heat, collision, poor material, and corrosion. These damage the structure and affect other systems.

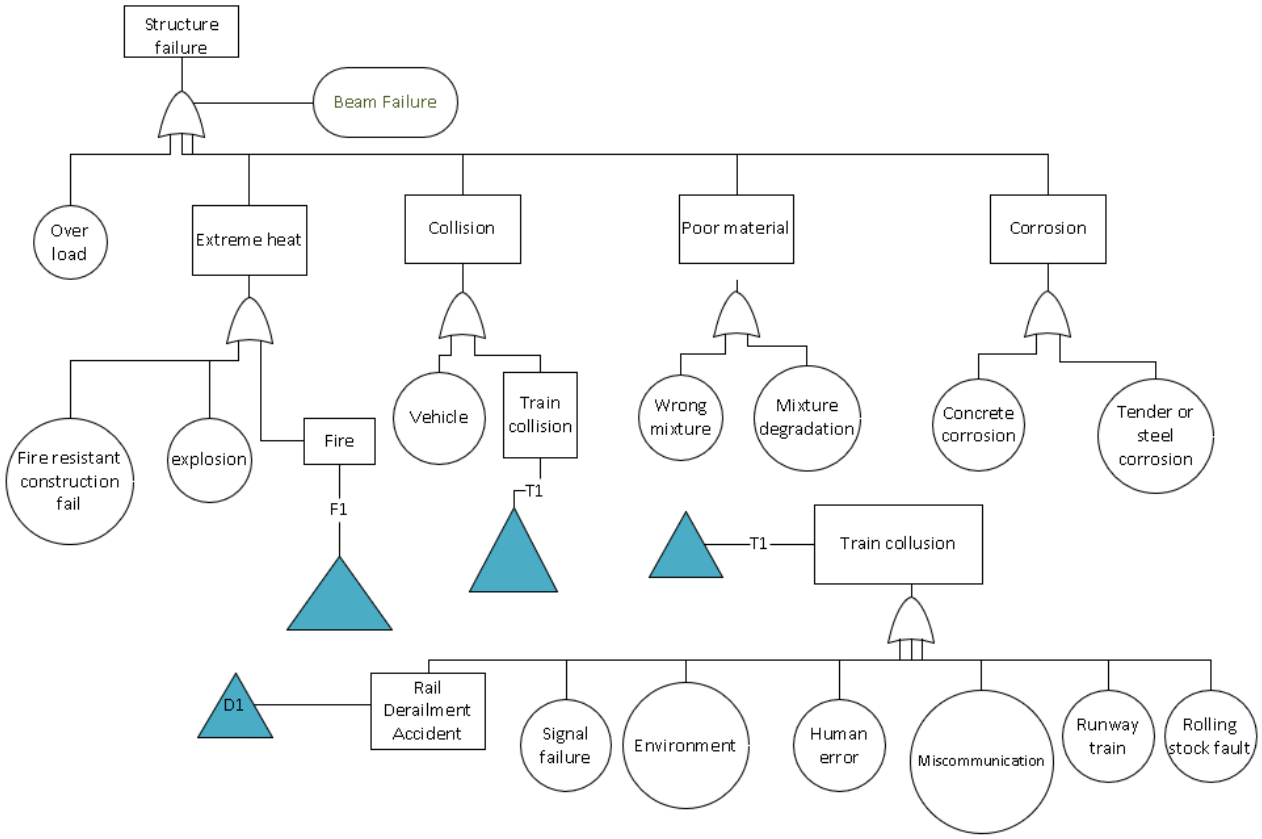


Figure 36. FTA of the structure failure risk in a railway station.

Many risk consequences are expected to lead to structure failures, such as fire, collision, and derailment, and the risks increase with refurbishment and maintenance work. The station has a part of track as an infrastructure, and has the probability of failure. The station structure is of more concern in this analysed and has been highlighted in the derailment risk. The resilience of assets management is important for the health of the assets. For the large stations with huge parts of the structure, it is hard to maintain each part in such a dynamic operation.

Global warming made the world weather change, so some areas become hotter and others cooler. Thus, effective monitoring is essential to protect the station's structure. For example, the heavy snow, and strong wind may affect the station roof or cause other parts to collapse, while the heavy rain may breach the capacity of the drainage system. The outcomes of such incidents may lead to deaths, injuries, delay, panic and direct and indirect costs.

The severe weather, location, design, and the pressure of usage have raised the challenges for stations structure, which requires developing the traditional monitoring method to be effective, smarter, and for the new projects to design stations that are more sustainable and resilient.

4.12 Discussion

An investigation of risks associated with the station's railway risk has been conducted. This section of the report contains the results of a suggested risk assessment in the railway station based on an FTA, which is a common approach for hazard analysis and risk assessment in the railway safety industry. The FTA clarifies the basic events and facilitates more comprehensive risk mitigation. Moreover, FTA has long been used to assess risks in many industries such as nuclear power plants and other critical facilities. FTA has limitations such as understanding the safety concerns of the overall system, dealing with a socio-technical system, analysing the complex system and non-linear dependencies, which are difficult to correctly form the causation chain or to identify the root causal factors. In fact, FTA is deductive, designed to deal with Boolean relationships [212][213].

The analysis of these risks in this Chapter on a large scale is essential for a more comprehensive understanding of the dangers of operations in the railway stations that present the absence of utilising a method that can capture the patterns in the interactions of complex data. The linked risks mean that risk leads to the chain of outcomes or to the others risks, for example, failure in the power supply may result in signal malfunction and then derailment or collision of the train and that may then cause a fire and resulting in congestion and stampedes of passengers. Evacuation is connected to overcrowding, security, fires, and in fact, many connections among risks can be obtained and worse situations might occur in the case where there is more than one risk (see Figure 37).

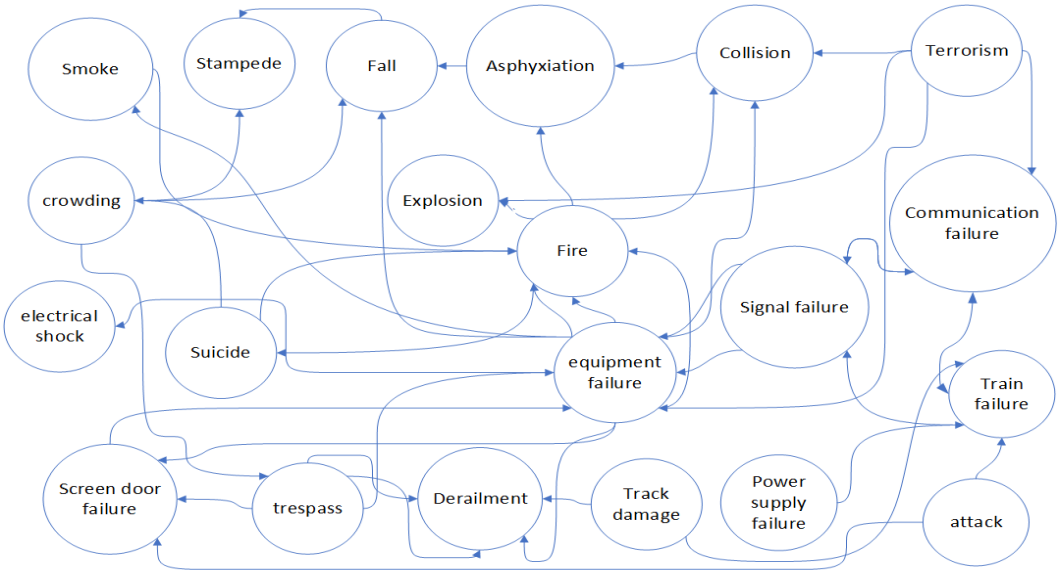


Figure 37. The correlation between the risks in the stations.

A variety of measures can be used to control risks like signage, sound warnings, CCTV, training, and cooperation with police and security services. However, the number of accidents in the railway stations are still increasing.

In fire risks, removing all flammable materials from the trains and stations or nearby is essential, as well as all sources of ignitions or naked flame in order to comply with standards and regulations.

Many risks presented in the station require effective management, such as slips, trips and falls, fires and explosions, boarding and alighting incidents at the interface between the train and the platform, derailments, etc. The security risk can lead to dilemmas; in fact, the management of vast numbers of passengers to allow safe free movement in the stations are not an easy task in the case of terrorism. However, the advancement of technology, such as integration between the police and the railway and other agencies will improve security.

Some of the control measures are passive systems that can be used after the accidents happen; this presents the necessity for a more effective and productive method than conventional methods. For

natural risk (earthquake, flooding, rainstorm, waterlogging and so on), prediction, system resistance and robustness are essential components to mitigate such risks. The research on prevention and control of meteorological disasters has gradually evolved from engineering prevention measures to new technology, including advanced disaster information management technology [214].

From the FTA in this chapter, it has been shown that the relation between the risks and their consequences is complex and potentially linked.

The error of the signal may lead to derailments or collisions, and then infrastructure damage in the stations may cause fire and explosions that can lead to falling in the station because of the panic. The outcomes may be even worse, and this scenario of an accident will cause much disruption and delay and loss in cost, business image and passenger satisfaction. Hence, it is important to do research into the risks in such critical points of the system.

Clearly, it is necessary to improve safety risk management with a new model to implement an effective risk management system in the railway station for strengthening safety, risk identification and performing dynamic risk monitoring and effective prevention systems in real-time. Moreover, it is required to improve control, where risk and emergency measures are formed in advance. For railway safety, the improvements in technical design are one of the largest factors to reductions in risks [184].

From lessons learned from a case study of the Eurostar incident of 18–19 December 2009, it has been recommended to design tools to assist in real time, shared decision support, and requiring real-time support in railway operations, thus, any unwanted event in such areas must be considered for both aspects of the proactive and the recovery [215]. Moreover, from analyses of previous accidents, the passenger actions produce the greatest number of injuries and deaths, and riders are the principal source of risk, top events, and injuries [184]. The workforce in the station are some of the people who

may be harmed by risks, and the Annual Statistical report 2018-19 shows an increase in workforce injuries by 2.5 % compared to the previous year and, sadly, the majority are station staff [216]. Furthermore, as a result of the growing scale of urban rail transit lines, it has been concluded that traditional operation management methods cannot satisfy network operation management requirements [106]. From experience, many safety measures have been applied for protection against risks in the stations especially in new systems, which, of course, lead to a reduction in accidents. However, the target of zero accidents has not been reached and needs more research and investment. There are numerous types of information that can be valuable and be learned from the accidents, but these need to be saved in a way to make them accessible and applicable from one generation to another. These lessons must also be shared as an accident that occurs in the area may be repeated in other stations with another operator. Moreover, modern life produces more risks and challenges that require an appropriate approach. Additionally, the safety, security, and emergency warning need to be developed to perform dynamic tracking, monitoring, decision-making, gathering data and identity management quickly and in real-time to mitigate the consequence of any unwanted event. In other words, applying new technology in the railway is a slowly developing sector where innovation is limited [217]. However, the technology utilising CCTV footage is expected to increase, which is important for safety, security, uses for prevention strategies, and capturing behaviour and locations in real time effectively with action from automated systems [186][218].

Due to the COVID-19 pandemic some health safety issues, such as the overcrowding and air ventilation, will be highlighted more in the near future. In fact, the ventilation system in the station has a complex design and a lower ground structure proves the importance of real-time data examining the air quality monitoring for assessing stations' environmental risks. The environments are extra complex in stations that are above ground level structures due to various pollution sources requiring control. These need to be considered early, and adjusted in real-time according to changing regional

climate and environmental conditions. However, there is a shortage of long-term-effect research and technology applications [219], where technologies and data analytics can be more effective for management, energy-saving and health safety in the stations' environment [220].

The developed model will provide a generic model of the safety risk on railways, which will raise the industry's knowledge of the railway stations' risks and is a starting point for effective railway system safety management innovation. There is no doubt that technological improvement will be expected to meet the station's requirements to be safe and secure in the near future to cope with the current uncertainties and challenges.

4.13 Summary

The analysis of risks is carried out in this part and the FTA tool analysis helps identify the causes and consequences to apply control measures based on the history of the accidents. The analysis results in this study indicate that the authority cannot observe the solution and evaluate the systems each second of its operation. Also, the risk considered in the station can be divided into two categories while some other risks are rare in the stations:

- Man-made risks: like fires, fall in gap/slip and trip, human errors, electric shock, stamped and security issues.
- Natural risks: earthquakes, floods, typhoons, extreme weather, epidemic.

The overcrowding risk management is critical and is connected with other risks like evacuation plans from fire explosions and may lead to a fall /slip, panic, and trip, as well as stranded passengers and stampedes. The places and machines such as platforms and moving trains, stairs, escalators, train doors and power equipment are important points in the station and must be managed via active technology reducing the dependency on human assistants which are not always there. From the analysis, it is concluded that human error appears as the root cause for many risks, which encourages

the proposed concept of reducing human interventions through applying AI. Obviously, the stations are complex and have interdependent subsystems; this reinforces and becomes ever more critical to improving risk management in the stations. Even with the shortage of knowledge of applying technology in risk management, it is expected there will be an acceleration of investment in technology because of COVID-19. Clearly, the analysis shows that there is a tendency to emphasise the research from the last two years to apply the new technology in the industry and there is an expected acceleration because of COVID-19.

The railway's safety and risk system are critical, has high connectivity risks, and is dynamic, which requires an effective response, proactive actions, requiring ongoing and timely control. The mainline has more obvious consequence risks such as derailment or collision and related factors, where the trains are expected to run at high speed and full of passengers, yet the focus of this research is the stations. Due to this, there is a need for innovative methods for risk management for the stations, to guarantee the operation safety. The increase in passengers and cities expands the lead to overcrowding and risks like slip/trip and fall as high-risk in many stations which are hard to manage. Global warming is considered a universal issue that leads to severe weather impacts in many fields and railways are no exception. Technology such as wireless sensors will provide a smart method for monitoring complex and huge areas such as the stations, with low human intervention (smart railway station). However, such large amounts of data require advanced analytics methods such as AI. The technology can benefit safety and risk systems in railway stations and provide great coverage in such a complex system, capturing any change in the system and supporting the decision-maker in a timely manner, preventing expected overlapping consequences in the case where many risks occur. Finally, the railway stations are ideal for advanced systems for a wide range of risks.

CHAPTER 5: ARTIFICIAL INTELLIGENCE (AI) METHODOLOGY

5.1 Introduction

AI is one of the current science and engineering disciplines, and began in earnest soon after World War II [221]. Historically, the term Artificial intelligence dates back to 1956, coined by John McCarthy in the first academic conference on the topic. However, the concept of machines that can think came earlier; for instance, in 1943, there were investigations regarding the construction of artificial neural networks that could compute logical functions [222]. A system was proposed that amplifies people's own knowledge and understanding in Vannevar Bush's seminal work As We May Think [223]. After that, it was published that machines have the ability to do intelligent things and are able to simulate human beings such as playing chess, as suggested by Alan Turing [224][225]. AI development carried on from the knowledge-based systems between the 1960s and 1970s, further improving neural networks in the 1980s, and real-world problems in the 1990s. However, currently, the improvements are expanding to cover a wide range of approaches including robotics, computer vision and the autonomous industry [226]. The applications cover a wide range of areas such as banking, health care and autonomous machines. AI applications are remarkably powerful and have attracted considerable scientific attention from almost all fields within the field of computer science. The AI concept indicates that machines work and react like humans. AI is described as how computers simulate intelligently, including learning, understanding symbolic information in context, associative and reasoning [227].

The data and the power of computers have powered applications for such technology. The data and business problems that need to be solved play an important role in appropriate AI model selection. The AI has enabled many practical forms, such as autonomic driving, preventive healthcare, e-commerce recommendation applications, and hopefully better safety environments through using AI applications in railway station safety in this study. ML as a subsystem of AI is systematically utilising algorithms to process the data, learn from them, and then make decisions or predictions in real-time with less human intervention. So, the system will be trained via BD to learn how to perform the task, while the traditional software with coding has routines with a specific set of instructions to accomplish a particular task. In fact, abundant data is essential for the models to be intelligent and reach the level that acts like a human. Through preparation data using categorized datasets, several machine learning algorithms have been proposed and used for mining meaningful information from the data. ML algorithms have variations of how to classify, but according to their purpose, algorithms are classified into three major categories, supervised, unsupervised and reinforcement learning. Unsupervised learning requires the ability to identify patterns in streams of inputs where supervision requires labels for the attributes [228]. Moreover, AI methods can be roughly split into two classes: a classic such as expert system and a computational intelligence like deep learning, ANN and fuzzy logic systems [227].

Applying AI in risk management needs to consider the data first. Secondly, the appropriate method and then the expected output need to be considered, and a framework should be developed to utilise all the predicted safety and risk management data (see Figure 38). For example, for advanced analyses, the correlations should be found between the attributes to predict the features of risks captured from the history of the accidents, so data is identified, and supervised ML is the method. Many stations need CCTV to be installed around the station to cover many important areas in the stations and to use the evidence for crimes; thus, such data can be used effectively, and the appropriate

method for such data is DL. The history of safety and risk management is often textual and has knowledge that can be used in the future; however, this data is unstructured, so unsupervised method is selected for this mission.

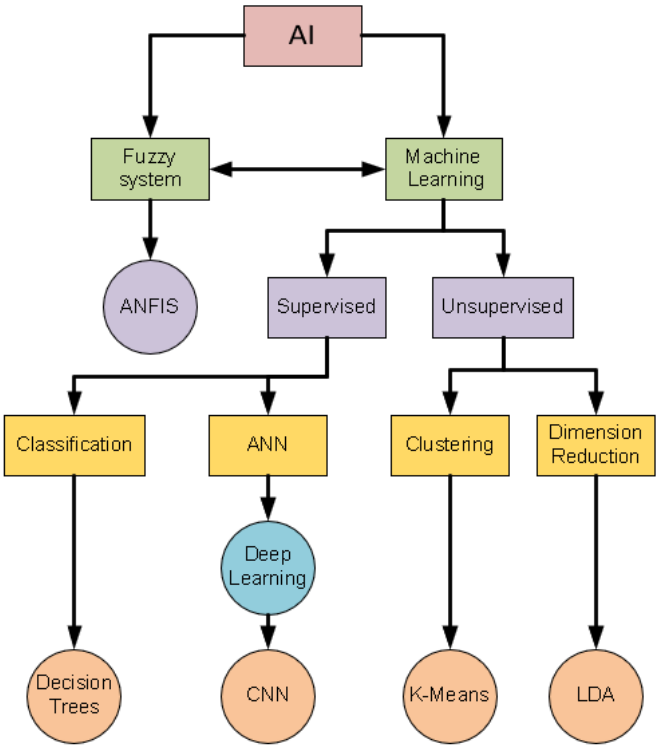


Figure 38. Railway station risk framework using a complex group of AI algorithms developed for the study.

5.2 Artificial Intelligence and risk management

AI is a part of our daily lives, as many people have smartphones, using the recommendation services, online shopping and some apps which learn and adapt based on our user preferences. Currently, numerous industries intend to promote performance via advanced technologies and inspiring transformative initiatives utilising AI, for instance, prevention of fraud, medical diagnosing, cyber security, health conditions, self-driving cars and drones. Furthermore, it has been noted that AI or advanced technology has the intelligence to address complicated conditions that are identified as

ambiguous and have the ability to transfer systems to be automated systems with less human intervention and errors. The domain of process of risk management in the railway stations is a perfect fit for advanced technology and indeed needed, as risk issues/events involve unwanted events, uncertain conditions, dynamic and complex operations, ambiguous events and risks which might happen simultaneously. Moreover, the process did not get much improvement and did not utilise technology for improvement in the field. With the common risk management decision models such as complex computer simulations or empirical experiments usually it is not possible to obtain the risk decisions in real-time. Thus, to enable real-time decisions, it requires a method to notify risk-based decisions to predict the likelihood risk scenarios earlier. The traditional risk management has many drawbacks and is incapable of handling the BD volume. With the extension of the railway network operations management, safety and security concerns, effectiveness in economic and social matters, and the increase in demand, the existing management systems could not satisfy nor barely meet the new requirements of operations. One main reason for this shortage is that the rail system's service coverage sector has undergone great changes and presently comprises a wide range of urban areas and the suburbs [106]. Thus, it is expected that cognitive computing will play a significant role in risk management, which can be applied to train machines to understand and identify risk. In fact, this method has already been established to grow the business, support decisions and boost performance. Some of the AI crucial methods, which are relevant for managing risk research, are shown in Figure 39:

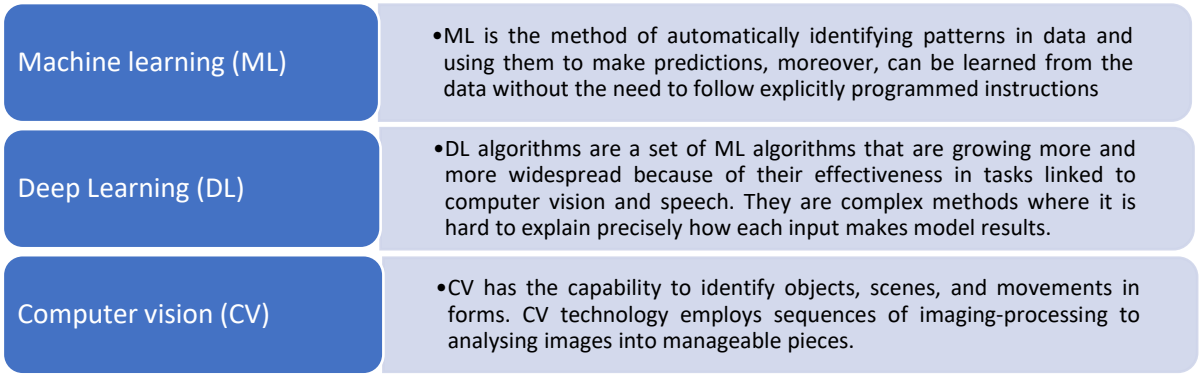


Figure 39. Examples of AI techniques details.

For automated risk analysis as a core process of safety management and effective measures, assembling a database Is required. This should contain all classifications of accidents, scenarios, expert knowledge and indicators as well as automated safety risk analysis when inputting some signs, so that automated retrieval the case can be retrieved with the results of the similarity between them [229]. Moreover, in case of an extensive amount of internal and external data, it is advantageously employed by the railway stations to take a ‘’proactive’’ stance depending on the predictions in risk management and in real-time not only for retrieval but to build learned systems.

5.2.1 Artificial intelligence boosting the innovation capability of railway safety

The railway system has a massive amount of raw data; however, the source of data needs to be identified, gathered and cleaned to be beneficial for exploring. AI can streamline the way it interacts with these large sets of safety data and can reduce the burden of seeking large sets of data. In fact, in many cases, analysing such a volume of data and complexity is beyond the analytical capability of humans. Consequently, AI contains great potential for boosting productivity and supporting systems to use resources efficiently, including reducing time, cost, and human intervention. The AI can assist in diminishing human error, managing the risks and doing common tasks in more efficient ways. The data from the history of safety, including accident reports, risk descriptions or risk assessments etc.,

or from the passengers, can be investigated using this method to promote people management in the stations, to predict their journeys, expecting or recognising unusual patterns of behaviour and preventing possible risks. The technology interaction with the data of risk can enhance active reactions and cover a wide range of risk indicators and improve integrating systems.

5.2.2 Framework of developing AI for the railway station risk

The risks network in the stations and their relationships are complex and difficult to represent, and require effort from all the departments and levels. So, it is essential for strategic planning to compose a framework that describes the trial to afford a holistic strategy to manage the risks in stations and utilise AI, producing direction to the auditors, risk managers, compliance and safety community. The AI approach in the railway station framework is outlined to improve the delivery of AI with the safe management of AI and showing applicability. This suggested framework is utilised as a general guide for professionals to use when confronted with the increasing use of AI in railway stations across different levels of applications. The proposed framework has classified the use of AI for risks into sixteen regions, as set out in Figure 40 below:

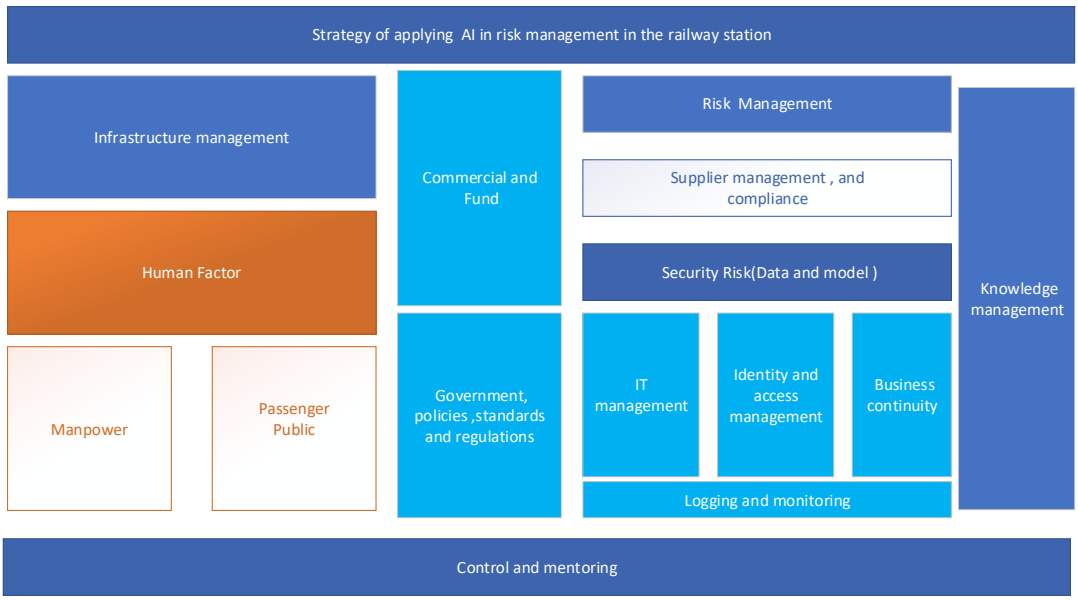


Figure 40. Implementation framework layers of AI in risk management railway stations.

5.3 Machine learning (ML)

The data are key to feeding the model to generate outcomes with high success rates. There are various applications utilising AI technologies such as ML and BD in many industries, such as medical, banking, and marketing; however, few technologies are being used in railways and transportation. Information on safety, risk monitoring, or emergency rescue by supervision in the rail system has not yet been entirely generated due to a lack of integrated systems for rail transportation safety and risk management, as well as delayed implementations of technology. Therefore, there is a lack of ability to utilize significant amounts of data in the railway industry to explain the relationships between operational factors, and safety and risk monitoring, especially in railway stations. Thus, more research is required to validate this relationship, which is the goal of this research and of designing benchmarks for the expected level of safety and security performance in railway stations. In addition, in future work, the obtained data will allow for its validity to be evaluated in a case study of the proposed framework. Massive dataset resources are captured for analysis, including the history of accidents locally and internationally. The concept of a smart city and smart stations represent an advanced level of this technology. Intelligently gathered information like weather conditions and crime can be associated in real-time with the railway data centre and used to predict scenarios and consequences of risks. This knowledge discovery from the predictive model will actively aid decision-makers, save time and enhance risk management, safety and security at stations, therein leading to high-performance predictions. Availability of indicators for risks leads to training the models and gaining high-performance prediction. However, gathering data for each risk is challenging, so a framework is designed in Figure 41 to illustrate such a case. This structure presents the overlapping in railway stations between safety, security and operations which is a dynamic process and requires certain technology [21].

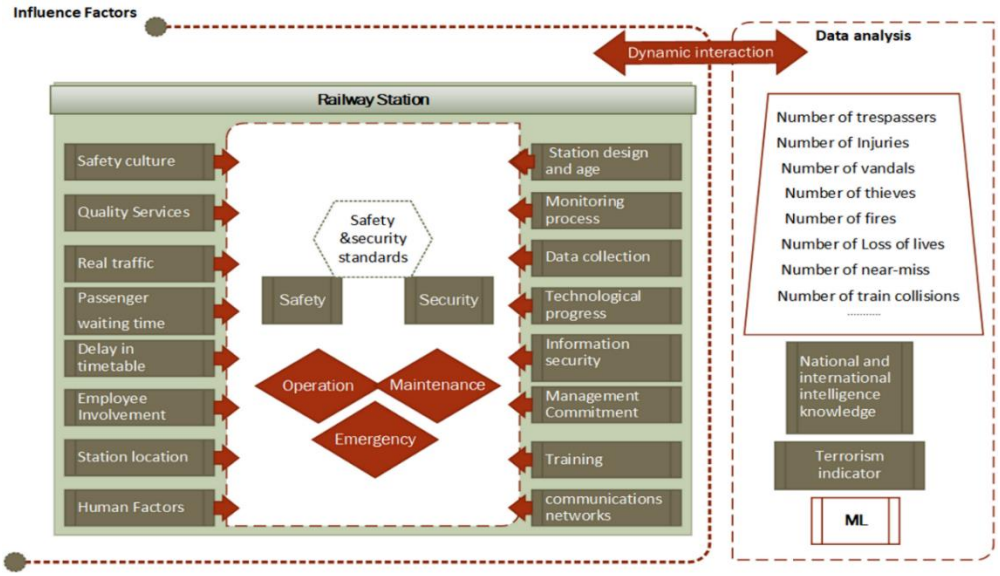


Figure 41. Designed ML flow for the safety mechanism in the stations.

In the case of overcrowding in the stations, to quantify or predict the relation between the overcrowding risk level it uses output of interest $\boldsymbol{Y_i}$ (response or dependent variable) which is referred to as the label, and the related features $\boldsymbol{X_i}$ (inputs available or features).

This creates a dataset with thousands of observations regarding the overcrowding levels in the station, periods (peak-super peak-not peak), day of the week, time of year (seasons), locations, and other related predictors (features). One way of understanding the different categories of ML is in relation to:

$$Y = f(X) + \varepsilon \qquad \text{Equation 1}$$

where the $f(\boldsymbol{X})$ is the unknown model function that represents the information that the feature vector X (predictor variable) provides about the label Y (response variable) and ε is random error. Thus, ML assesses this function with $\hat{\boldsymbol{f}}$. Moreover, the learned model from data can be expressed as:

$$\hat{Y} = \hat{f}(X) \qquad \text{Equation 2}$$

where the $\widehat{\boldsymbol{Y}}$ is the predicted answer variable given the predictors X, when the dataset includes paired features and labels, estimating $\hat{\boldsymbol{f}}$ is supervised learning. However, when the dataset is unlabelled, there is no $\boldsymbol{Y_i}$, and estimating $\hat{\boldsymbol{f}}$ is unsupervised learning. Unsupervised learning has a dataset with only input variables, knowledge discovery and patterns without a guide to supervise the learning. In the case of the dataset that has partial labels but not all (some $\boldsymbol{X_i}$ have no $\boldsymbol{Y_i}$), then estimating $\hat{\boldsymbol{f}}$ is semi-supervised learning. Reinforcement learning is exploring for X and searching for the optimal $\boldsymbol{X}^*$ given a desired output Y [230]. More details regarding the class of supervised learning related to the methods that are targeted to be examined are displayed in the next lines.

5.4 Supervised learning

This subcategory of learning produces inferential or prediction outputs, relying on the labelled dataset inputs. Estimating a function outlines an input vector to an output given a dataset of paired observations:

$$\hat{\boldsymbol{f}}: X \longrightarrow Y$$

This estimation has the ability to infer the association between the predictors and output as well as predicting the responses for observed features not previously seen from the model. The response variable Y is a quantitative or qualitative variable so two main sub-classes are under the supervised learning including regression and classification applications as in the next section. The approach starts from the dataset that has labelled features to find a way to classify. The method explores the patterns and correlations to advance the analytics process. For example, millions of images of passengers at risk, such as from a fall or slip, trip or other motion picture, capture the risks such as overcrowding or insecurity behaviour. Each risk can create a ML application that distinguishes one behaviour from another by labelling this data about types of risks. Supervised learning usually begins with an established set of data and a certain understanding of how it is classified. When the label is

continuous, it is called a regression which aims to understand the correlation between variables. So, for estimating $\hat{f}$, for which the response variable Y is quantitative, the model is regression estimating. Moreover, when data come from a finite set of values, it is known as classification [230][231], which is covered in the next section.

5.4.1 Classification

For estimating $\hat{f}$ for which the response variable is qualitative, the model is classification estimating, given a labelled dataset. The outcomes as the learned knowledge from the model can lead to predicting the classification probability of the dataset not previously seen before [230]. The trained model can classify data into classes. It has wide applications such as speech recognition, medical imaging like whether a tumour is small, medium, or large, and clarifying whether emails are genuine or spam. This category can be applied using many classifier algorithms such as decision trees and deep learning.

5.4.2 Decision tree (DT)

The DT algorithms have the power to illustrate and outline the consequences of a decision using the branching structure where each node describes possible results of the decision path. Each node has the likelihood of the outcome as a percentage depending on the data from the fathers of the node in the tree. For example, the tree for analysing the accidents in the stations, the passenger age, time, and the risk type can describe the outcomes of the accidents like highly possible to death or less possible to injury. So, DT will improve risk management by accident groups and the relations between the passenger features who committed risks in the accidents. Moreover, understanding the context for such historical accidents via descriptive analytics is essential for assessing the current safety level and for future strategic plans. In addition, it can be used to explore hidden patterns while the data cover a wide range of locations and different periods as well as the areas exhibiting specific risk and

seasons and how the implementation of safety measures can have an impact. There are a number of variants of DTs such as classification and regression trees (CART), chi-squared automatic interaction detection (CHAID), and iterative dichotomiser ID3, C4.5, and C5.0. CART is a DT algorithm that produces binary classification or regression trees, relying on whether the target variable is categorical or numeric and exploring the hidden patterns or rules located in the data and suitable for small amounts of data sets. One of the methods adopted from ML and important is the random forest tree (ensemble) of classification or regression trees (CART) (Breiman et al., 1984) [232], which uses the Gini Index as a splitting function [233]. Moreover, it has been shown that it is substantially more scalable and can address multiple data types simultaneously. The model stops growing when they have exhausted their ability to fit the training data better. Each tree node attempts to split the data most optimally so that the classification splits maximize the information gain [21], using methods to quantify the splitting criteria like Gini impurity and information gain. The tree can split the data many times into classes in accordance with some exogenous rule; for instance, an accident history record can be computed subject to time, age or location and the outcome of the accident in whether it led to death or injuries. This is assuming that data for training the model has variable matrix X with M number of variables $\boldsymbol{x_j}$ and N observations and considering that class vector Y consistd of N observations with a total amount of K classes. The tree structure as tp is a parent node and tl, tr - respectively left and right child nodes of the parental node tp (divided into two segments), assuming for a variable $\boldsymbol{x_j}$, the best splitting value is x_j^R (see Figure 42 below).

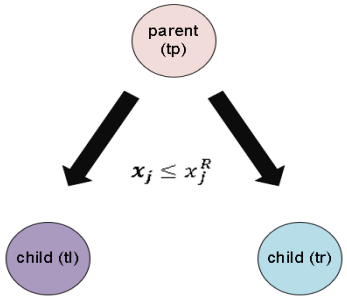


Figure 42. Splitting algorithm of CART.

Aiming to classify data sets for purer (quality) subsets, purity can be reached through criteria such as Gini Index and Entropy. To find the best split, Gini Index is a criterion used in the CART inducer and as a splitting function (impurity measure), which also has a bias that favours attributes that have more outcome values during attribute selection. Gini Index tries to split the attribute into two branches regardless of the attribute type. Gini Index is also based on impurity calculations; therefore, the impurity of the training data is measured, which measures the differences between the probability distributions of the target attribute's values. If the training data "D" has subsets $C1, C2, \ldots C_i$ where m is the number of classes the training dataset has and p_i is the probability that the data instance belongs to class C_i, then the index can be formed as:

Gini Index $=G\ (D) = \ 1 - \sum_{i=1}^{m} p_i^2$ Equation 3

The node with zero impurity is the ideal. Moreover, the entropy and the classification error are considered as e measures for the splitting criteria. The entropy measure or the information required to classify a random data instance where the data instances contained in the node is denoted with D, and it can be expressed as:

Entropy $H(D) = -\sum_{i=1}^{m}(p_i(\log_2 p_i))$ Equation 4

This measure increases if more equally organized points are in a specific space, if the overall entropy is 0, and then it points out that the node is totally pure and a leaf node can be formed. Otherwise, it continues calculations since the node is impure. Also, the Classification Error can be computed as:

$$\text{Classification error } (t) = 1 - max[p_i] \qquad \text{Equation 5}$$

For illustration, for instance, an attribute X has three values and two classes (M and F) as shown in Figure 43, and the calculation of the measures is presented in Table 11.

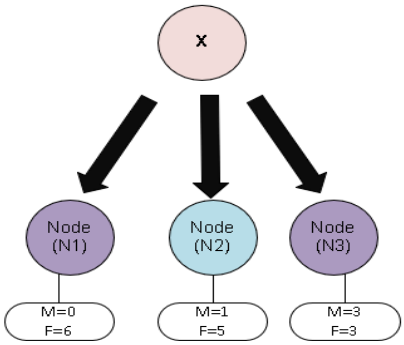


Figure 43. The simple tree shows the example distribution.

Table 11. Example of calculation presents the Impurity Measures.

The Node	The measures		
	Gini Index	Entropy	Classification error
N1	$1-(0/6)^2-(6/6)^2=0$	$-(0/6)\log_2(0/6) -(6/6)\log_2(6/6) =0$	1-Max [0/6,6/6] =0
N2	$1-(1/6)^2-(5/6)^2=0.278$	$-(1/6)\log_2(1/6) -(5/6)\log_2(5/6) =0.65$	1-Max [1/6,5/6]=0.167
N3	$1-(3/6)^2-(3/6)^2=0.5$	$-(3/6)\log_2(3/6) -(3/6)\log_2(3/6) =1$	1-Max [0/6,3/6] =0.5

It is observed that when Gini index has a zero mean there is no more split required for node 1, and the distribution is uniform in node 3 where a fraction of records belong to one of the two classes and the entropy equals one where all the records belong to the same class. The differences between Gini

measure and entropy are not large, as Gini makes a purer separation, but entropy tends to make more balanced subsets. In fact, CART shows advantages such as clearly managing both numerical and categorical variables and it has a power to identify the most significant variables and handle outliers [234][235][236].

5.5 Neural networks

This method has been described as a system that can transmit information from the cause to the effect via neurons or a tool to create a correlation. It has been applied in many problems such as non-linear, dynamic, pattern classification, clustering and optimization [237]. Applying a caricature of the process the human brain processes information, which contains highly interconnected units (neurons or nodes) with weights which are varied to minimise the error between ANN output and the given output. The basic network is structured by an input layer, an output layer, and any number of hidden layers. Assuming Neuron x has m inputs, and one output and g the non-linear activation function [238] can be formulated as:

$y\ (x) = g \sum_{i=0}^{m} w_i\ x_i$+bias (b) Equation 6

where $(w_0, \ldots, w_n)$ are the input weights, where the weighted sum of the output from every neuron is the input to the next neuron. The bias is a constant value that gives the ability for the line to be effectively transposed (shifting the activation function to either right or left) in the space as analogous to the role of a constant in a linear function [239]. The input is then modified within the neuron by utilising a function called an activation function, which aims to define the output of that node, for instance, ReLU function which is more commonly used (Figure 44). Every neuron is a mathematical processing system (computing units) which, connected with all other neurons, is created to learn the correlation between the input characteristics and the output. The updating of the weight values repeats leading to learning during backpropagation, which will guide the inputs to the main output which is

compared with the observed point, based on the label value [240]. The opposite direction order from output to the input layer is called a Backpropagation Algorithm. Backpropagation is based on repeated application of the error used for gradient descent, and an error is computed and operated for updating the weights.

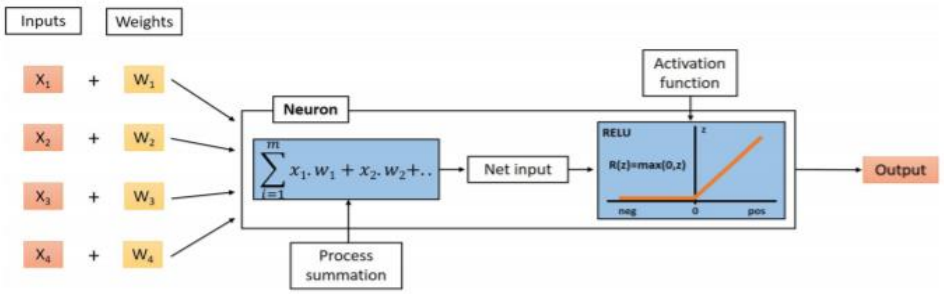


Figure 44. The process of single neuron with ReLU Function.

ANNs require examples to learn, and a few prior assumptions [241]. By adjusting the connection weights, the learning may be supervised or unsupervised. Supervised learning supplies the ANN with the correct output for every input pattern. However, unsupervised learning supplies the ANN with several input patterns. The ANN then investigates the associations between the patterns and learns to classify the input [242] [243]. The ANN consists of three or more layers: an input layer, one or many hidden layers, and an output layer, the structure of an ANN, as in Figure 45 below.

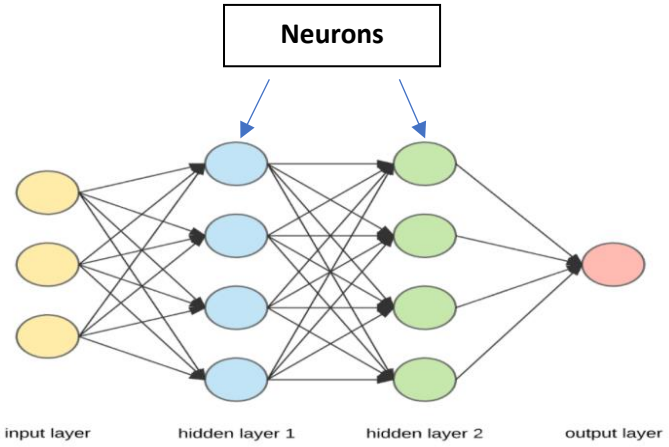


Figure 45. The style of a neural network with two hidden layers.

5.6 Fuzzy systems

The fuzzy term refers to things that are not clear and the fuzzy set theory was introduced by Zadeh in 1965 [244]. In classical set theory, an element is a member of a set representing an absolute value such as 0 or 1, but fuzzy set membership takes any value between 0 and 1, which offers more flexibility and considers uncertainties, so Fuzzy systems (FS) use fuzzy sets to deal with imprecise data and vague statements, which are used to describe fuzziness via its membership function (fuzzy sets). Moreover, this membership function has been noticed as influential characteristics, such as utilizing occurrences rather than knowledge to resolve real-life dilemmas [245]. Fuzzy logic is a system that assists many fields including engineering and control systems to do things better and expand on choices. It is nonlinear mapping of an input dataset to scalar output data. The system contains four main parts: a fuzzifier, a set of rules, an inference engine and a defuzzifier. Figure 46 presents the components of a fuzzy system [246].

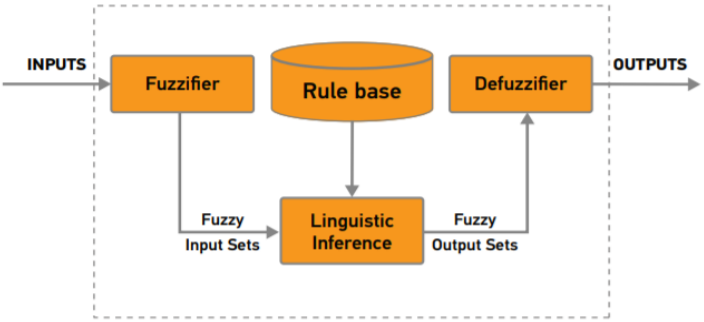


Figure 46. The components of a Fuzzy Logic System.

Fuzzification (fuzzifier) can be defined as the conversion of crisp sets to a fuzzy set as a step to convert inputs. For example, it can measure or index risks that are collected in stations as crisp sets and passed into the model or the control system for processing and to convert inputs, linked to actions or warning systems. Triangular, Gaussian, and trapezoidal membership function MFs are the most utilized in the fuzzification process. The inference engine assists in defining the level of match between fuzzy input and the rules; the rate of the match sets the rules to be implemented, and the outcomes are a set of fuzzy decisions. Then, the control action is taken depending on combined rules which have been applied. The logical operators, such as "AND," "OR," and "NOT" define how the fuzzy variables are combined [247] [248]. The end is defuzzification (defuzzifier), converting the fuzzy member into crisp data [245].

Rule Base contains all the rules and the if-then conditions which have been designed by expert and decision-makers, for example, the overcrowding level threshold in specific platform depending on the standard and the area measured. The process of expressing the mapping from a given data to an output using fuzzy logic is Fuzzy inference—this includes implementing Membership Functions and the rules (see Figure 47). For example, the fuzzy inference process for the overcrowding risk level, with the tow-input, one-output, and three-rule can be used as:

•Rule 1: If the passenger capacity on the platform is below the standard level and no passenger is alighting from the train, then the risk is low.

•Rule 2: If the passenger capacity on the platform is below the standard level and few passengers are alighting from the train, then the risk is average.

•Rule 3: If the passenger capacity is more the standard level or the same, and few passengers are alighting from the train, then the risk is high.

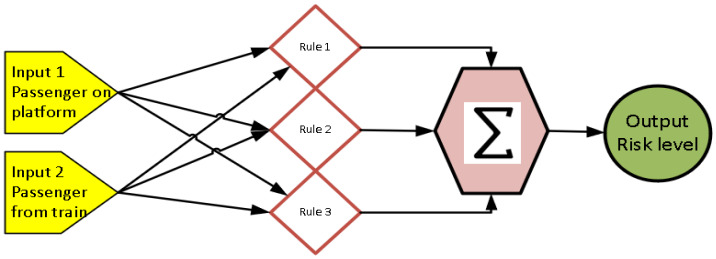


Figure 47. Fuzzy Inference Process with two inputs, three rule and one output.

5.6.1 Neuro Fuzzy Approach -Adaptive neuro-fuzzy inference system (ANFIS)

ANFIS is a combination of the computational power of ANN and the high-level reasoning capability of a fuzzy inference system (FIS) [249]. FIS is the process of expressing the mapping from a given input to an output using fuzzy logic. In other words, by inserting the FIS into the framework of adaptive networks, we receive the ANFIS architecture [250][251] as hybrid intelligent systems. It is a multilayer feed-forward network that employs an ANN and FLS to map the relation between inputs into an output, with the power of identifying patterns, ability to adapt to cope with a dynamic environment, to determine the optimal distribution of membership functions, and solution of function approximation problems [252][253]. It is an FLS implemented in the structure of an adaptive ANN. Furthermore, ANFIS can construct a network realisation of IF/THEN rules [250]. The method utilises supervised learning on a learning algorithm. It consists of five layers (fuzzy, product, normalization,

defuzzy, and output) for the first-order Takagi–Sugeno [254]. For two ANFIS inputs (X_1, X_2), they are transformed into fuzzy inputs in Layer 1 (the adaptive node or fuzzification layer), as shown in Figure 48. Then, the process generates membership functions (MFs) for each of the inputs. There are many MF types; for instance, triangular MF and generalized bell MF (see Figure 49).

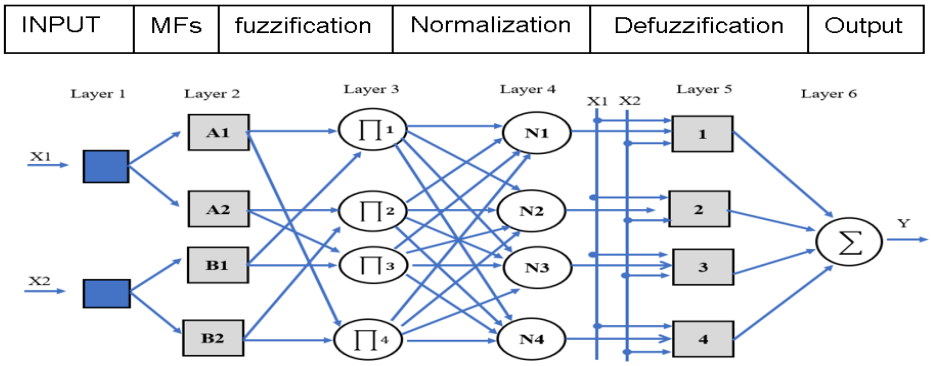


Figure 48. ANFIS Structure with 4-rule, 2- input scheme.

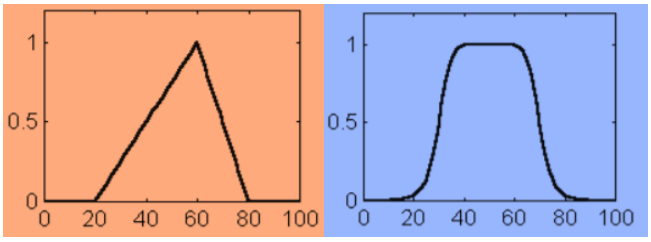


Figure 49. The triangular and generalized bell mf.

The output is predicted through an adaptive network, which feeds forward multi-layers of ANN with adaptive nodes. Learning rules are specified for the parameters of the adaptive node and the adjustment of the parameter depends on error values [255][256]. This method is conducted for risk prediction on station risk levels, and as it has been indicated that the model can be utilised to assess risk during a small computational time, this can be applied to many risk factors with different indicators in the railway industry risk controller [257] (see Figure 50).

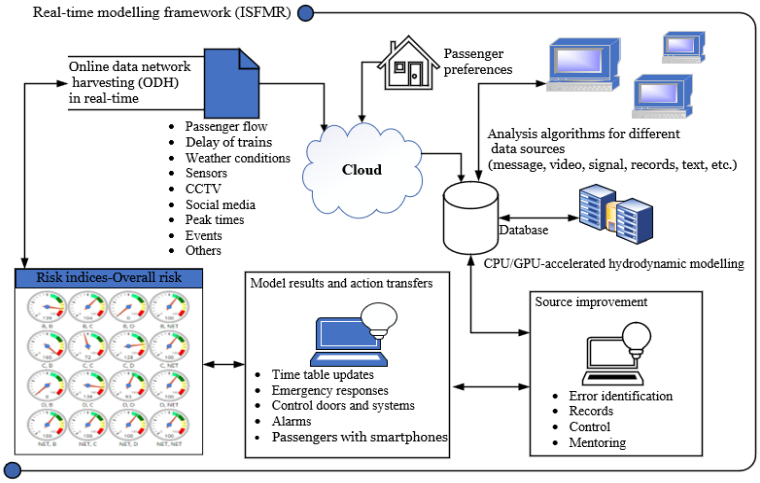


Figure 50. Proposed process framework for the Intelligent System for Managing Risks (ISFMR) In railway stations.

This system has two steps, as initially initial fuzzy model is formed, then the neural network is used. The initial step is fuzzy rules extracted from the input-output data set. The next step is to fine-tune the rules to train the network and produce the final ANFIS model of the system [258].

5.7 Unsupervised learning

In this approach, the dataset is processed with input variables $\boldsymbol{X_i}$, and does not specify the response $\boldsymbol{Y_i}$ output. The aim is to examine the features and explore the patterns, and the highly popular style of unsupervised ML is unsupervised classification (clustering). Unsupervised learning is best suited when the data are unstructured, such as the safety history documents or accident reports. This advanced analysis will help classify the data based on the patterns or clustering and associations and begin to understand the meaning. While there is no labelling the data, the iterative process occurs fully without human intervention. For example, collecting huge amounts of data about a specific risk or accidents can help practitioners gain insights into the failure patterns and consequences, design indicators and safety measures, and relate them to risk management outcomes. Taking action to design a risk assessment covering the whole system is time-consuming with uncertainties and costs

of experts who may not be available all the time. Hence, an unsupervised learning approach can help determine outcomes more instantly than a supervised learning style [231].

5.7.1 Clustering

This unsupervised straightforward method is grouping objects together (cluster) that have similar parameters [231]. As one of the most common pattern recognitions, data mining and analysis methods aim to gather objects which are related to each other or similar in one group and other groups for the different objects. K-means is considered one of the most used clustering algorithms (Partitional clustering approach), which is very popular and used in a variety of applications, and K expresses the number of clusters, where every cluster is connected with centroid (centre point) (see Figure 51).

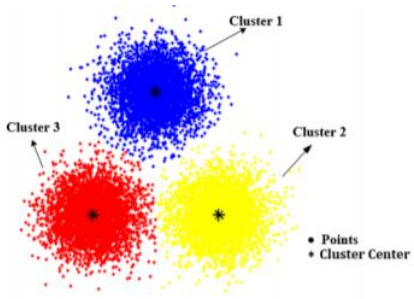


Figure 51. Concept diagram of k-means clustering.

For example, to find the Euclidean distance between two points in space (y_1, x_1) and $(y_2 - x_2)$, the following can be expressed:

$$d\,(x,y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} \qquad \text{Equation 7}$$

The algorithm is created through the following steps. Firstly, a randomly select k point from the data is set as the initial centroids. Then, for each point, the Euclidean distance is computed with each centroid and allocated to the closest cluster. Next, the centroids are updated by calculating the mean of the cluster. Finally, the two procedures are repeated (compute the Euclidean distance and the mean)

until the standard measure of the function begins to converge. The standard measure function, the summation of the square of the Euclidean distance between each cluster and its centroid can be expressed as:

$$E = \sum_{K=1}^{K} \sum_{i=1}^{n_k} \left\| x_i^k - c_k \right\|^2 \quad \text{Equation 8}$$

where K is the number of clusters, c_k is the centroid of the Kth cluster, n_k is the number of the Kth cluster, x_i^k is the ith point of the Kth cluster [259] [260].

5.7.2 Topic modelling

Natural language processing (NLP) has been applied in many fields such as business to understand customers, marketing and sentiment analysis, and machine translation like Google Translate. As an unsupervised method, the method has the ability to analyse unstructured text and capture the meaning. This can be documents, records or calls. This forms an opportunity for safety and risk management to analyse and review current records or safety textual history. The large amount of textual data from the safety history needs to be realised as a source of transfer knowledge, to make informed decisions, and a source for learning systems. Moreover, the technology overcomes the challenges of reading, understanding, organizing and exploring this data, like unstructured format, time, cost which cannot be manual. This history covers a huge range of sources such as accident reports, investigations, near misses, safety reviews, records and more, including news and social media. In the field, there are many ways to do topic modelling, yet in this work, the case study will come over the Latent Dirichlet Allocation (LDA) [261].

5.7.2.1 Latent Dirichlet Allocation (LDA)

LDA is a probabilistic three-level hierarchical Bayesian model for collections of discrete data such as text [261], generating topics based on word frequency from a given set of documents, developed

by Prof. David M. Blei in 2003. LDA develops the technique from the category of topic models and the Probabilistic Latent Semantic Analysis (PLSA) [262].

As a generative system, the data from such processes includes latent/hidden variables which is a joint probability distribution over both the observed and hidden random variables. The process is executed via the hidden variables (topic structure) given the observed variables (the words of the documents) as the conditional distribution (posterior distribution) [263]. For analysis of the Latent variable model, the data can be expressed in terms of 3 sets of variables:

- Documents: $d \in D = \{d_1, \ldots\ldots, d_N\}$ where the N is the size of corpus.
- Words: $w \in W = \{w_1, \ldots\ldots, w_M\}$ where the M is the frequency of distinct words from the corpus.
- Topics (latent / hidden variables): $z \in Z = \{z_1, \ldots\ldots, z_K\}$ where the K is the number of topics to be specified a priori.

The latent topic links the documents and words, the document can be presented via the conditional probabilities (Conditional independence) of concepts weighted by $P(z|d)$ and each word can be expressed as a topic with word probability $P(w|z)$ (see Figure 52) [262].

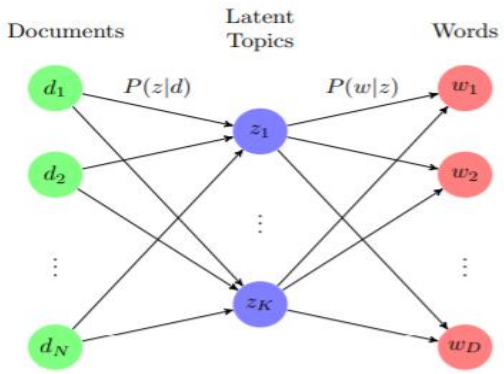


Figure 52. The typical formation of PLSA.

The model can be described with the notation presented in the plate diagram or the graphical model shown in Figure 53, which are means of explaining the probabilistic theories behind LDA mode.

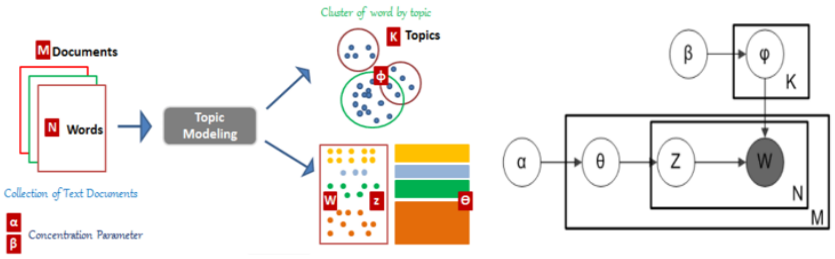


Figure 53. Description of graphical representation of LDA.

Where K, N, M are a number of topics, the words in a document, and the number of documents to analyse, respectively, **α** is Per-document topic distribution, β is Per topic word distribution, the W is word, Z is topic, and **Θ is** Topic distribution for a document (topic probability per document), which is generated by a Dirichlet distribution parameterized by α. The φ**(k)** is word distribution for topic k (Topic probability per word), which is generated by a Dirichlet distribution parameterized by β. For example, exploring safety documents such as accident reports or risk review documents, these documents have many topics such as accident details and safety measures, and each topic has many related words such as fall, slip /trip related to the accidents topic, and these words have a different probability. In other words, LDA clusters the documents into topics and depending on the word's distribution [264][265].

5.8 Deep Learning (DL)

DL is a novel branch from ML that can be considered an evolution of ANN. This learning processes the classification via the hierarchical level of ANN, which attempts to mimic the process of the human mind, and has several connected layers. This approach, which has many hidden layers, is learnt from a compound of unsupervised and supervised algorithms. When there is more than one hidden layer

in an ANN, the network is sometimes called deep learning, and one of the fundamental applications in this learning case is computer vision, including a variety of purposes, for example, risk predictions, autonomous vehicles and medical images. DL has the ability to act in real-time from the captured images in the surroundings and then algorithms interpret the data to help the system make decisions.

DL depends on employing nonlinear algorithms to match data. This technique is used in a variety of ways, but they all have certain commonalities, such as how each layer receives the output from the previous layer as inputs, as shown in Figure 54 [266].

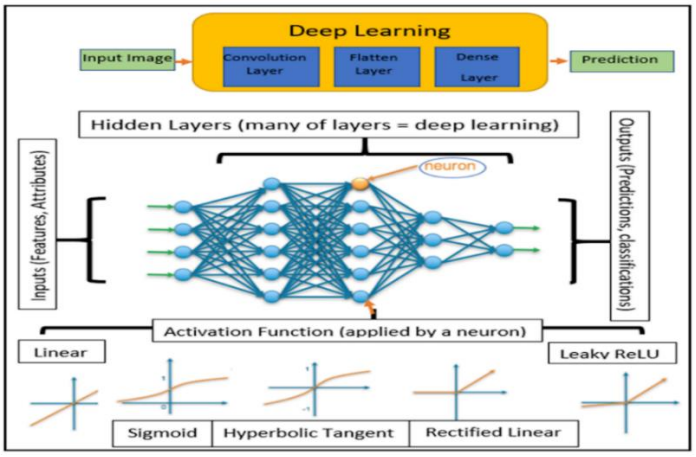


Figure 54. Structure of deep learning.

The deep neural network has a powerful method for a high level of intelligence learning process, and the term ‘deep’ refers to the fact that several layers of processing convert the input data (for example images) into valuable predictions for making decisions. ANN filters transform the data before passing it to the next layer and to the next in order to learn from data in an iterative approach. The ANN may consist of thousands or even millions of simple processing nodes that are densely interconnected [267][231]. DL or hierarchical learning began in 2006 within machine learning as a novel research method for speech recognition, which is neural networks with multiple layers and convolutional neural network CNN is the most representative model of such methods. The concept has breakthrough

results in image analysis, feature extraction, classification, and detection of objects. The following is a description of the convolution operation:

$$s\,(t) = \int x(a)w(t-a)da \qquad \text{Equation 9}$$

where x and w is the kernel, which is an adaptive filter that the network learns [268]. Each layer of CNN is identified as a feature map dealing with input as a 3D matrix of pixel intensities for different colour channels (e.g., Red-Green-Blue channels RGB). This feature map has different types of transformations like convolution operations (See Figure 55) or pooling such as Max Pooling. The process is nonlinear, the layer takes the output of the previous layer as input and the structure connects all layers to filter and capture the useful information.

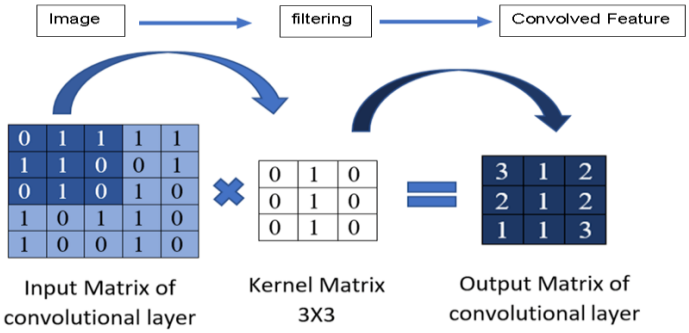


Figure 55. An illustration of a convolutional layer process using a 3 x 3 x 1 Kernel (K).

For example, the convolution operation learns the weights matrix from previous neurons and takes a nonlinear function (such as sigmoid, ReLU) to obtain final responses, and for pooling operations, and to summarise the responses of the previous layer into one value to produce a more robust feature description [269][270]. This method is conducted (CNN) for a case study in the next chapters. To analyse BD that railway systems will collect to build risk-recognition, the framework can be designed as shown in Figure 56. The availability of massive data including temporal data, videos or images and data fusion, with utilised CNN provides novel suitable solutions for safety and security risks.

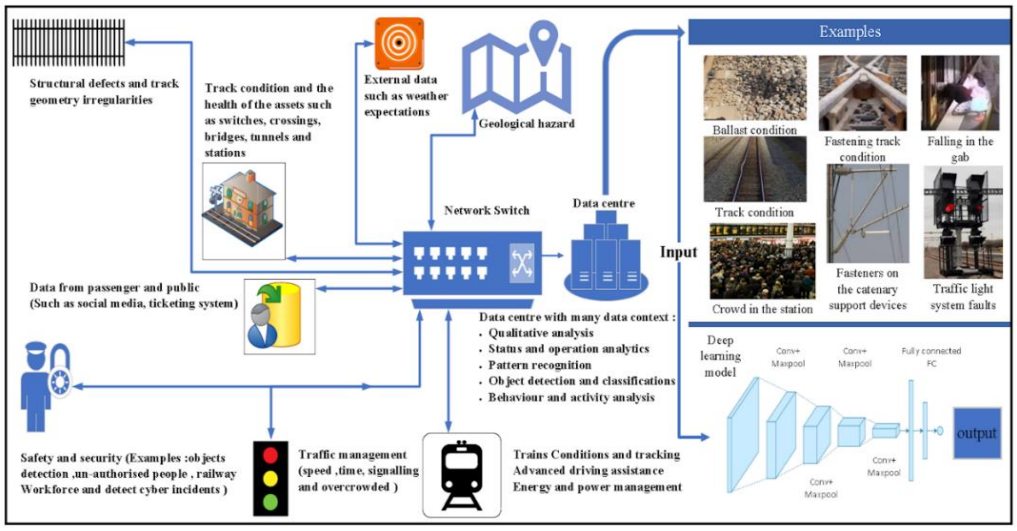


Figure 56. An innovative paradigm for using a CNN to control risks in the railway system.

By using advanced detection algorithms and identifying risks in the early stages, surveillance footage will play a variety of roles in industry risk management and protection, including suicide, traffic flow, illegal activity, trespassers, smoke, and fires [271]. Advanced techniques can help rescue workers and decision-makers by detecting objects and conducting video analytics. These identification strategies are expected to assist in the development of emergency response plans and communication schemes, all of which are important in minimising the risks associated with emergency incidents in railway stations. Additionally, the latest technologies can help with interventions to ensure passenger egress and transportation at vital station locations for emergency responders, such as tunnels and access points. Furthermore, sophisticated computational video monitoring can cover a wide range of threats, including accidents, derailments, and intrusions from adjacent areas into unauthorised station locations, such as a track, while simultaneously handling other subsystems in real-time with minimal manpower and high performance (see Figure 57).

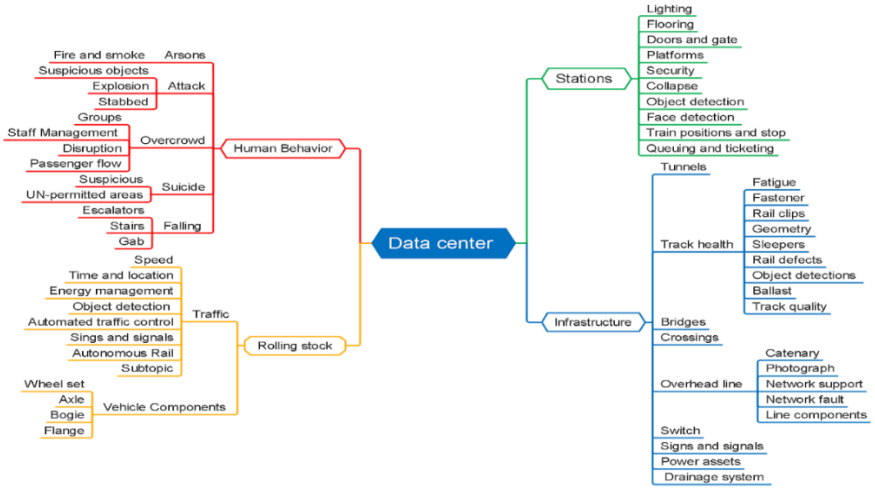


Figure 57. The tree-map presents applications of CNN in railway stations with different risk sources.

5.9 Summary

AI has been analysed in this chapter as well as the selected aspects within the scope of this research. Even though the range of AI is wide, AI applications, including ML, is the trend in both the business and academic sector, supported via progress in the computer industry and data availability. Railway safety, including risk management, will be attractive for AI methods in the near future when data are more available and considerable. The nature of the field problem and the data type is the key for the technology selection, yet some method can be used for the same data type which provides more comparable knowledge in the future (see Figure 58).

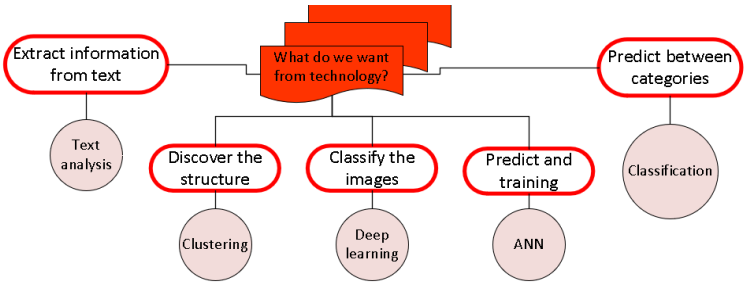


Figure 58. ML Algorithms hierarchy mapping.

Data is the key for the models, so attention has been placed on utilising real data with high quality to reflect the purpose of the work. Evaluating and verifying the models is essential, so a train/test split is the plan to verify the models, so some data sets not seen by the model (new data) will be used to test the model to evaluate the predictions and avoid any overfitting. For example, 80% of the data are for training and 20% of the data are for testing. The training data set is used to train the model, which is the concept of AI (to learn). Moreover, it is considered that the model performance should be evaluated with different measures such as accuracy, loss, and confusion matrix.

The ANN alone cannot illustrate the full risk process; thus, it will be combined with fuzzy to analyse the overcrowding risk, which can be automatically configured through learning. In safety-critical applications, the ANN is still observed as a black-box. However, it knows at least the inputs and expected outputs, which lack satisfactory analytical means, and are not moulding suitable safety arguments [272]. The standard fuzzy logic system requires human intervention to determine both the fuzzy set and MFs [273]. The accident's historical data are complicated, and the decision tree will be a fitting method as information is gained for advanced analytics. DT are widely used and have the position that can easily be expressed as rules, and they are not black-box models [274].The LDA method has many benefits, such as building an expert system from safety history, reducing the impact of human resources retirement, and providing the new authority with reliable knowledge in time from a tremendous range of experiences in many conditions and locations and fitting with the big unstructured risk safety textual data.

The DL method such as computer vision has distinct advantages over other types and is suitable for utilising images and being applied for fall and slip/trip risks assessment in this study. Clearly, for the future, there are opportunities for engaging AI technology in safety, risk management, and advancing smart stations, depending on the data type and the model performance, including the cost and the learning rate.

CHAPTER 6: MACHINE LEARNING FOR THE RISK MANAGEMENT OF A STATION

6.1 Introduction

Nowadays, we are currently in a period of technological revolution, where technology such as ML have been used in many fields, for instance, in banking, business marketing, intelligent energy grids, driverless cars and healthcare and many more (see Figure 59). AI presents a huge opportunity for providing safe and secure environments for humans, saving lives from disease or from all the risks around them in their daily lives. In the railway process and safety, applying this technology will contribute to the railway system as a vital part of the public transportation services. In fact, the governments, business, stockholders, and the scholars are expanding the programs of digitalisation, research into AI and BD analysis, and this is where the science is heading in the railway industry where it expects it will play a leading role in establishing the appropriate use of these powerful technologies. This improvement will reflect on the economy and lead to establishing smart systems. Thus, it is a timely and relevant piece of work researching the application of ML in the risk and management of railway systems. This chapter examines a case analysis of the supervising method ML via Decision Tree (DT) and DL via CNN, as advanced analysis of accidents in the railway stations.

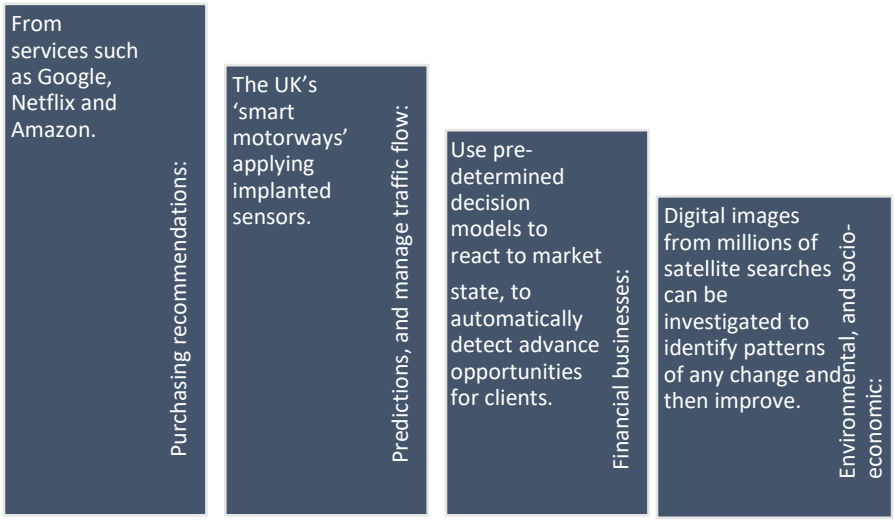


Figure 59.Some of the AI current cases of the systems and powered by ML.

6.2 Decision tree classification

A DT is a determination support tool that applies a tree-like pattern of decisions and their likely outcomes [275]. DTs have useful information, and branches are used to make a branching decision, where the decisions then lead to a given prediction. There are many possible ML approaches to safety analysis. In this case, it trains a DT to classify the accidents and the patterns that occurred in these accidents in the stations [276]. This model is applied to a wide variety of data, and it is preferable because its structured rules are simple to follow and understand. The two general types of DTs are classification (where the class variable is discrete) and regression (where the class variable is continuous) [277][278]. After the datasets are uploaded, a DT model is designed and visualised. The DT for the predictive model provides a visualisation of the prediction case, and presents excellent accuracy with great training efficiency [279].

6.3 A Case study 1

In this chapter, a case study is conducted aiming to expound the potential for applying supervised machine learning in the railway industry. This involves the application of learning a set of rules from

instances (examples in a training set). The importance of this study is in its explanation of the potentials of the machine learning to be used in improving services, management and particularly safety in the station environments. This designed model for predicting the safety and supporting the decision makers will be based on data collection from RSSB reports since 2002 with no missing values. The collected data of accidents that have been reported and published are a selection of 80 incidents at the stations in the UK that have been or are subjected to an investigation by the UK's National Investigation Body, the Rail Accident Investigation Branch [RAIB][280] .

6.4 Preparing data

The first step in the supervised learning method is collecting the dataset and finding the attributes that are the most informative. The next step is preparing the data. After this, the classifier model is selected, and to calculate a classifier's accuracy, it splits the dataset for training the model and evaluations. Utilising AI methods requires massive amounts of data to gain the benefits of such technology and data preparation is a fundamental stage of data analysis. The accidents used are only those which lead to fatalities within the station's boundaries; this gives the research greater precision and indicates the worst-case scenarios. From investigation reports, any data that may not provide all the specifics of the accidents or which do not lead to deaths or which occurred outside of the station's boundaries such as accidents on the line have been excluded. Moreover, the data that did provide information about the passengers or the details of the accident were omitted to ensure that there were no missing attributes. The total number of accidents was 80 (see Figure 60).

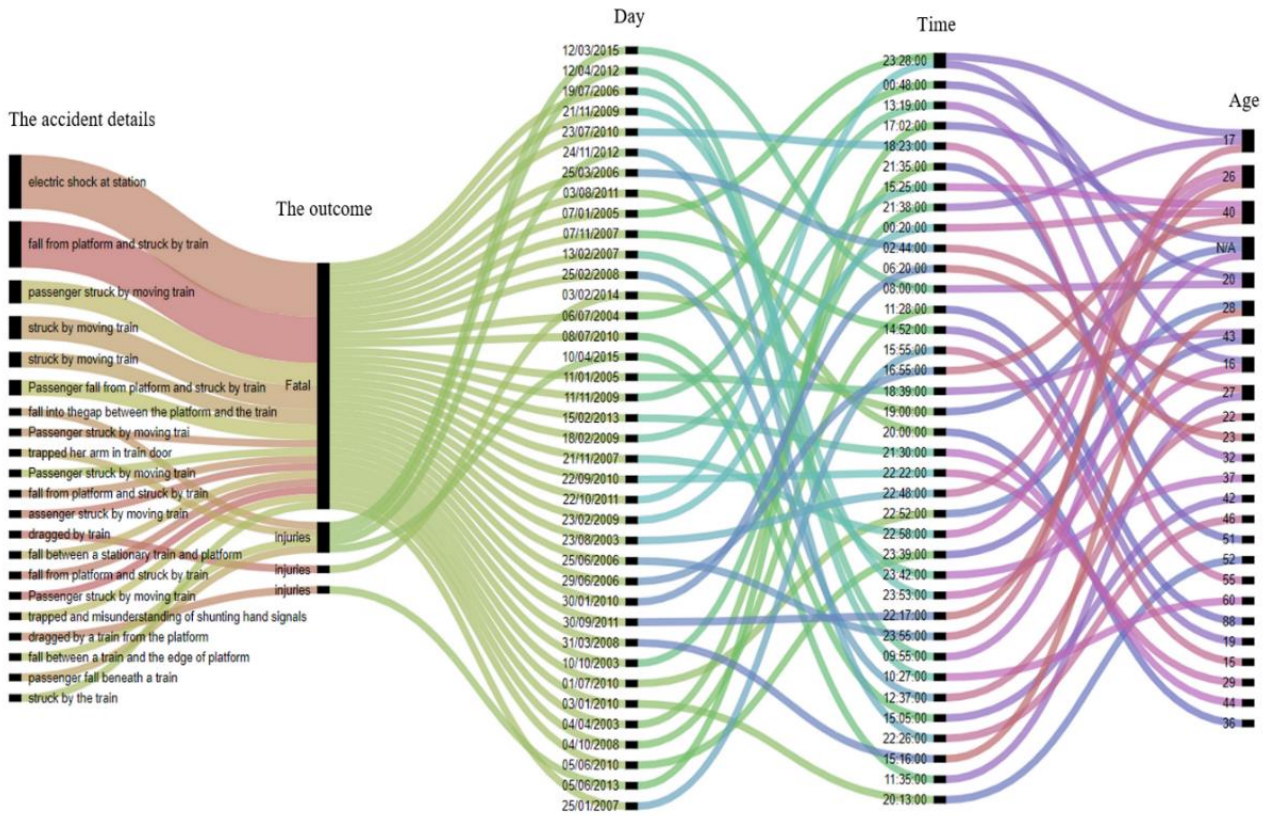


Figure 60. An alluvial diagram of primary examples of accident data shows the accident characteristics, consequence, date, time, and victim age.

Some steps for preparing data include the following:

- Generalization: For example, the date of the accident field in the accident documents, which consists of the year, month, and the day, is amended to contain the specific day of the week (Saturday (D-1) etc.) and particular time such as AM or PM.
- Designing highlights: From the cause of each accident, for example, falling from the platform and being struck by the train (T1-F), electrical shock (T2-E), or being struck by a moving train (T3-S), a distinct feature is created.
- Transforming data: The set of values is consistent with a new set of feature values. For example, the day of the accident, age and gender (Female (G-F) and Male (G-M)) of the person are converted into discrete values.

- Reducing or removing redundant features: Features that are inappropriate for this study, such as accidents occurring out of the stations.

The determinants in scenarios of events, such as passenger age and gender, as well as the day of the event and the exact time have been extracted. After the preparing and cleaning of the data during data exploration, the number of accidents is reduced to 71 accidents (instances), with five variables, including accidents leading to fatalities and only within the train station boundaries. The data set is rigorous and robust, depending on real-life accidents in the stations and from the official investigation authorities. Moreover, the data set only covers accidents that had deadly consequences, providing precise essential outcomes to save lives.

Each accident related to stations and the information from railway industries generally does not have many details except for certain reports on the web. Considering existing data and their value, it works with a small database, and attempts to make these data more useful; thus, non-relevant information is removed [281][282].

6.5 The analysis

A dataset of 71 accidents is used in this analysis. This dataset contains the attributes of age, sex of the passengers, the day of the week and the time of the event as well as the cause of death. The attribute matrix was applied as the input of the DT model, and the time was targeted as a predictor. The process of analysing and utilising ML as the method proposed in this case study is used to learn from the accidents and thereby benefit from this technology in the field. There are more selections to model and predict utilising other predictors and many options for inputs. Following the data cleaning step, some DTs are used in this work for various predictors in the method and for proving the power of ML in analysing safety data. The visualization enables us to analyse and visualise descriptive

statistics of the dataset (see Figures 61 and 62). This also shows the distribution of the passengers' age, time and details for each accident.

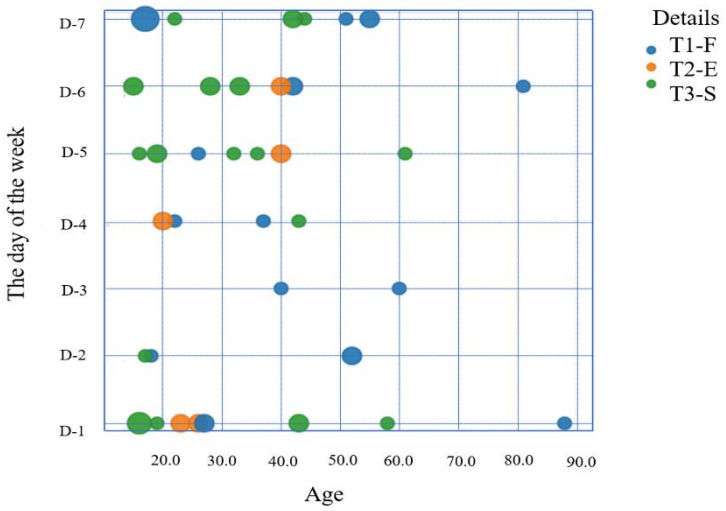


Figure 61. Scatterplot representing the relationship between the characteristics of the accidents and the passenger's age, as well as the day of the week.

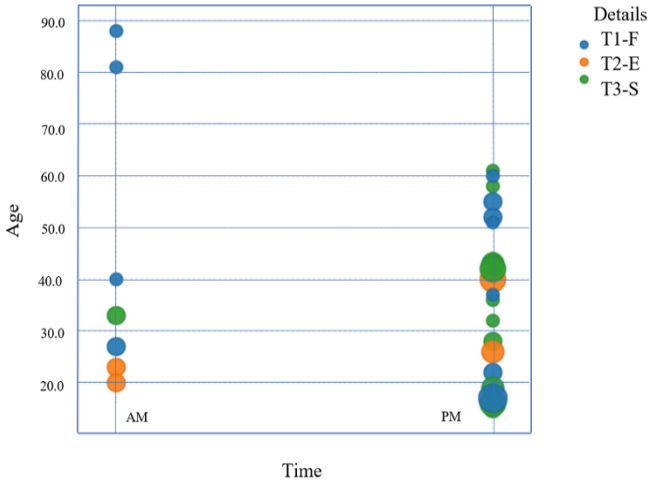


Figure 62. Scatterplot representing the relationship between the characteristics of the accidents and the passenger age with the time.

The selected method is the dynamic methods used to analyse the datasets. Thus, it sets the attributes of the accidents as our target; consequently, any predictor from the dataset can be used. However, the accuracy and prediction path will vary from one attribute to another. Despite this, this application attempts to outline the data to explore the interests for safety data analysis and to demonstrate the suggested method. An example of the results of the DT is shown in Figure 63. The example indicates

how important each factor in the prediction of accidents is, where the day of the week is the most important factor, followed by age (see Table 12).

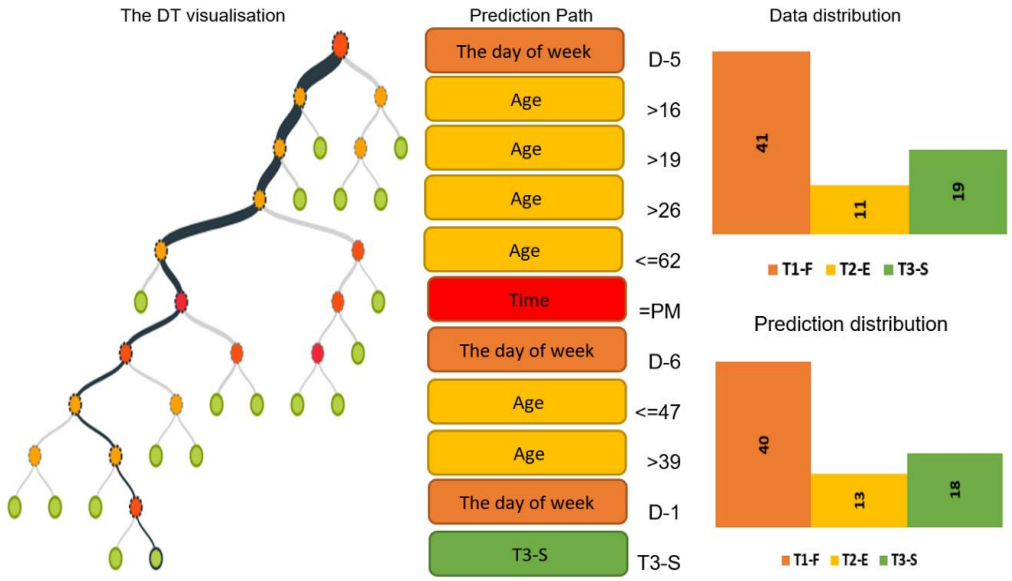


Figure 63. The prediction and data distribution from the training dataset are depicted in a decision tree diagram and graphs.

Table 12. The analysis of the dataset.

The input factors (attributes)	**The analysis**	**The importance of the field %**
Day of the week	Saturday=D1, Sunday=D2 etc.	52.8
Passenger details	Mean age =36.14 (child, elderly, disabled, teenager male or female (G-M, F)	41.41
The details of the accidents	T1-F: Fall from platform and struck T2-S: Struck by moving train T3-E: Electrical shock in the station	3.35
Time	AM, PM	2.48

For further exposition of ML techniques in such cases, a clustering method is applied because intelligent methods used to present and extract data patterns of interest are searched, and it is shown

that ML is a powerful analysis method for safety and risk management in railway stations. To analyse the unsupervised dataset, ML is chosen with the K-means algorithm (canonical clustering), where the number of clusters is eight. However, the remainder of the work is supervised ML. Utilising cluster analysis involves separating datasets into subsets of instances (clusters) and finding similarities (see Figure 64 and Table 13). The clusters are placed closer to one another if they are more similar and farther away if they are very dissimilar, where, for example, cluster number 5 is a long distance from the other clusters. The size of a cluster, presented as a circle, is proportional to the number of instances in that group; thus, the largest cluster has 14 cases, and the smallest cluster has only 3 cases. To process the learning data, 8 clusters are utilised. The K-means algorithm from data mining starts with the first group of randomly selected centroids, which are used as the initial points for every cluster, and then performs iterative (repetitive) calculations to optimise the positions of the centroids.

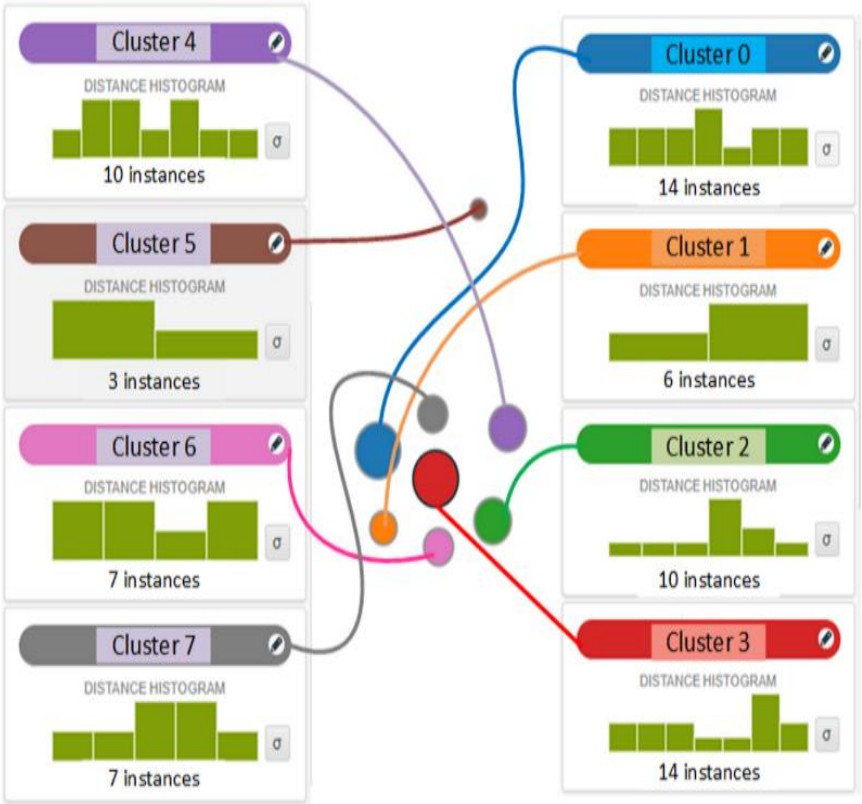


Figure 64. The nodes and histograms of the 8-cluster diagram.

Table 13. The details of the clusters

Centroid name	Instances	Minimum intercentroid distance	Mean intercentroid distance	Maximum Intercentroid Distance	Distance sum squares	Distance standard deviation	Distance sum	Distance median	Distance maximum	Distance minimum	Distance variance	Distance mean
Cluster 0	14.00	0.34	0.51	0.98	0.71	0.08	2.96	0.25	0.29	0.08	0.01	0.21
Cluster 1	6.00	0.43	0.54	0.90	0.25	0.13	1.00	0.25	0.25	0.00	0.02	0.17
Cluster 2	10.00	0.36	0.46	0.65	0.57	0.08	2.28	0.25	0.32	0.04	0.01	0.23
Cluster 3	14.00	0.30	0.47	0.89	0.78	0.12	2.87	0.27	0.31	0.01	0.01	0.21
Cluster 4	10.00	0.35	0.52	0.64	0.59	0.11	2.17	0.25	0.41	0.03	0.01	0.22
Cluster 5	3.00	0.53	0.81	0.98	0.13	0.19	0.42	0.03	0.35	0.03	0.03	0.14
Cluster 6	7.00	0.30	0.43	0.73	0.43	0.13	1.51	0.25	0.35	0.04	0.02	0.22
Cluster 7	7.00	0.36	0.54	0.98	0.52	0.04	1.88	0.25	0.35	0.25	0.00	0.27

6.6 Prediction and analysis

The DT model is utilised for predicting the future values of passenger attributes based on reregistered values. In this case study, the target passenger characteristics can be changed from one characteristic to another which show the flexibility of the method. This results in a unified framework that can perform analyses of variable data using the ML algorithms. This DT shows the prediction path, where the strength of the path in the tree is indicated by bold branch paths. The time attribute has been selected as a predictor (see Figure 65).

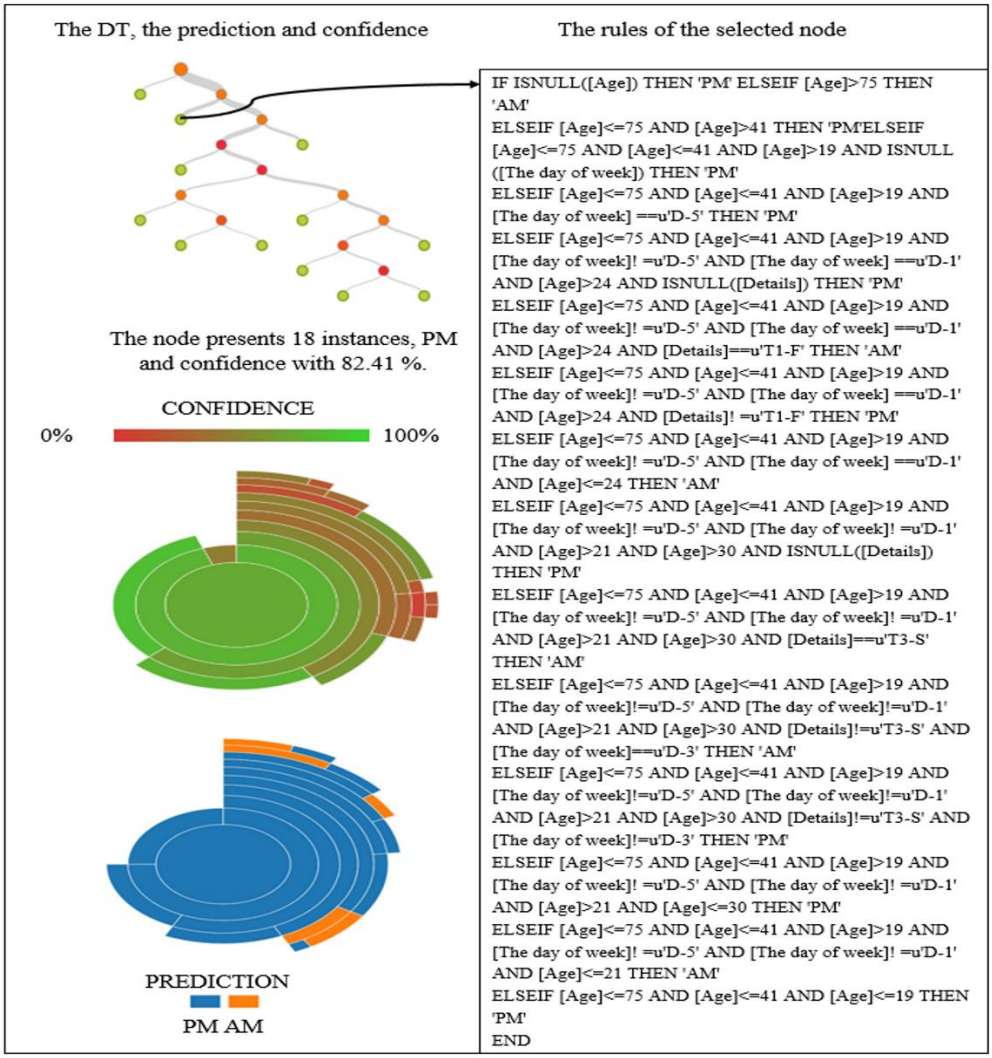


Figure 65. The rules and the decision tree with an example of a particular node.

Obtaining more details of the prediction path and input data changes of the input fields has been an interesting process. After midday, the prediction shows more accidents for an older passenger at the end of the week. The time represents a critical point as an input field affecting the prediction. There is a slight influence of the day in the forecast, which may refer to other factors not involved in this case study (see Figures 66 and 67) (see Table 14).

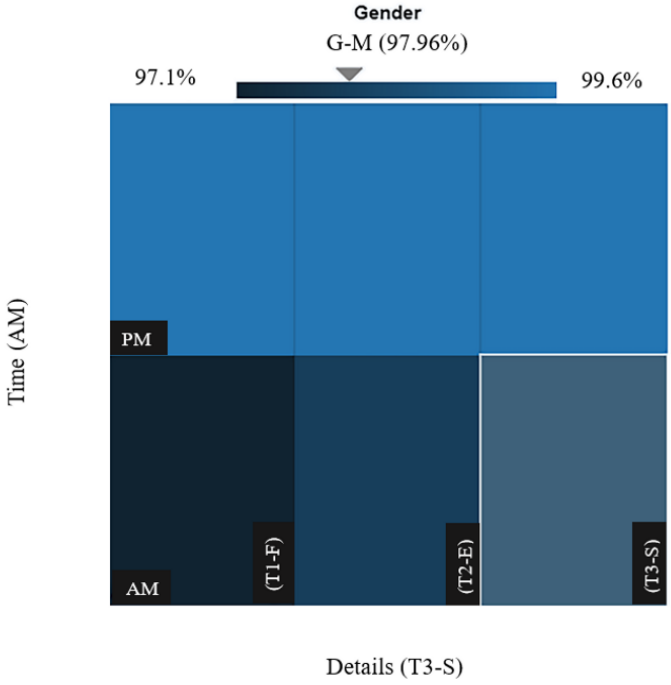


Figure 66. On the prediction path, comparisons of the input field with several factors (time and the details of the accidents). During PM, the result is displayed for the example accident type T3-S.

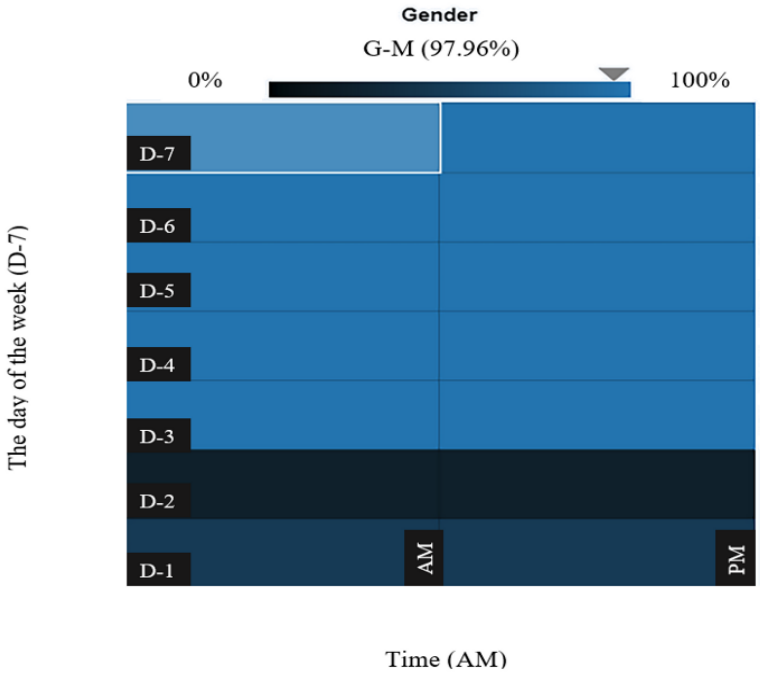
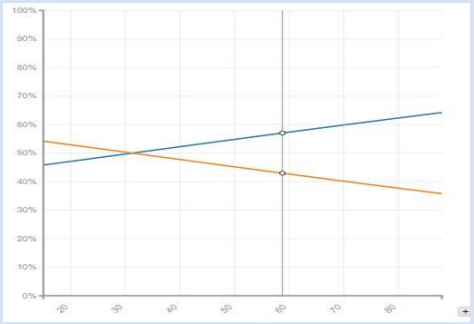
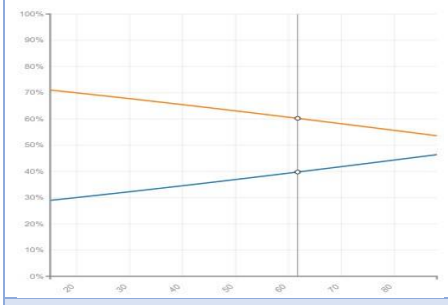
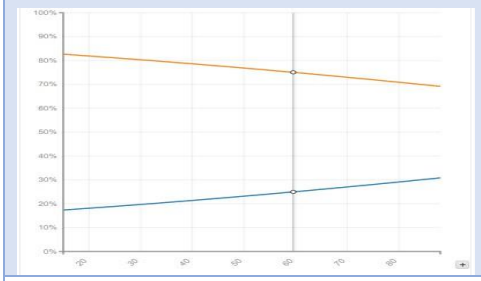
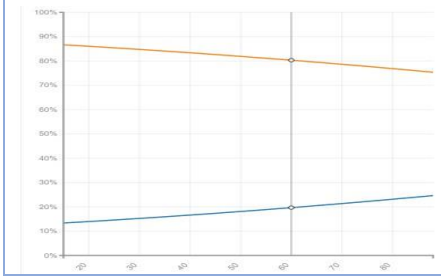


Figure 67. Comparison of the input fields on day 7 with different parameters (time/weekday).

Table 14. The comparisons of input fields on the prediction path with fixed age= 60.

The graph 1D (y= probability X=Age)	The prediction and the input fields
	Day =D-1, Gender=G-F 
	Day =D-1, Gender=G-M 
	Day =D-2, Gender=G-F 
	Day =D-7 , Gender=G-F 

The time is indicated showing that PM experienced more accidents than AM. The large ratio of accidents occurring for males is seen in two important age groups: the young and the old. Many factors, such as intoxication, must be considered; however, in this research, it is attempted to apply ML rather than a deep analysis of the causes of the accident. The DT has been applied, and it shows how the prediction path predicts the target instances depending on the selected data type and the

prediction targets, for instance, the accident type T1-F has been targeted to present the numerical data of the prediction path from the DT (see Table 15).

Table 15. The DT prediction path example, where the target in this example is the accident type T1-F

Node Output	Node Instances	Node Support	Node Confidence	Prediction	Prediction Confidence
T1-F	6	0.107	0.609	T1-F	0.609
T1-F	3	0.053	0.207	T3-S	0.609
T3-S	6	0.107	0.609	T3-S	0.609
T1-F	10	0.178	0.595	T1-F	0.595
T2-E	3	0.053	0.207	T3-S	0.510
T3-S	4	0.071	0.510	T3-S	0.510
T1-F	1	0.017	0.206	T2-E	0.510
T2-E	4	0.071	0.510	T2-E	0.510
T1-F	2	0.035	0.342	T3-S	0.452
T3-S	9	0.160	0.452	T3-S	0.452
T3-S	8	0.142	0.409	T3-S	0.409
T1-F	2	0.035	0.342	T2-E	0.375
T2-E	5	0.089	0.375	T2-E	0.375
T3-S	7	0.125	0.358	T3-S	0.358
T2-E	4	0.071	0.510	T3-S	0.353
T3-S	11	0.196	0.353	T3-S	0.353
T1-F	2	0.035	0.094	T1-F	0.342
T1-F	2	0.035	0.342	T1-F	0.342
T1-F	2	0.035	0.342	T3-S	0.342
T3-S	2	0.035	0.342	T3-S	0.342
T2-E	2	0.035	0.342	T1-F	0.342
T1-F	2	0.035	0.342	T1-F	0.342
T1-F	2	0.035	0.342	T1-F	0.342
T1-F	1	0.017	0.206	T3-S	0.342
T3-S	2	0.035	0.342	T3-S	0.342
T3-S	46	0.821	0.302	T3-S	0.302
T1-F	4	0.071	0.300	T1-F	0.300
T1-F	6	0.107	0.299	T1-F	0.299
T1-F	56	1	0.291	T1-F	0.291
T2-E	7	0.125	0.250	T2-E	0.250
T2-E	7	0.125	0.250	T2-E	0.250
T2-E	13	0.232	0.232	T2-E	0.232
T3-S	35	0.625	0.231	T3-S	0.231
T3-S	3	0.053	0.207	T3-S	0.207
T3-S	21	0.375	0.207	T3-S	0.207
T1-F	28	0.5	0.207	T1-F	0.207
T3-S	4	0.071	0.510	T1-F	0.150
T1-F	4	0.071	0.150	T1-F	0.150
T1-F	4	0.071	0.150	T1-F	0.150

6.7 Learning and validation

Evaluating the model to ensure that it produces reliable predictions is significant. In this section, it is aiming to obtain an overview of the model's predictive performance and create a framework for comparing models with different configurations or different algorithms to classify the models with the best predictive performance. The model is built on a subset of the data, termed the training data, and they are applied to predict new data that are not part of this training subset. This useful model has been shown to be well balanced in terms of avoiding both overfitting and underfitting. The steps begin with a training/testing data split by choosing subsets for generating the 80%/20% split of the dataset. The former can be applied to train the model and the latter to test it; thus, 15 accidents are randomly selected for testing, and the remaining 56 are used for training the model (Figure 68).

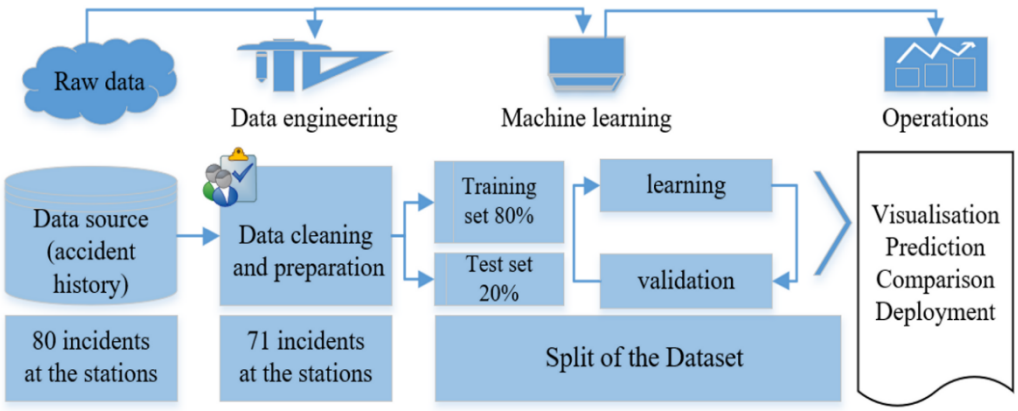


Figure 68. The stages of splitting the dataset.

The predictor selection allows you to assess and evaluate the models' performance in order to build a new DT model with a modified confidence value. It is now possible to compare and assess the DT models. In our prediction model, the prediction model is a classifier of the occurrences between the passenger qualities, which are dependent on the accident parameter history. A binary classification (two-class prediction) was chosen, in the case of fatal accidents in the stations (to determine whether the accident occurred during PM (0) or AM (1)). The outcomes of the prediction are labelled as either

positive or negative. If the prediction is positive and the actual value is also positive, then it is called a true positive (TP); with the same concepts, false positives (FP), true negatives (TN), and false negatives (FN) are realised. The four outcomes can be formulated as a 2 x 2 contingency table or confusion matrix, as shown below (see Figure 69).

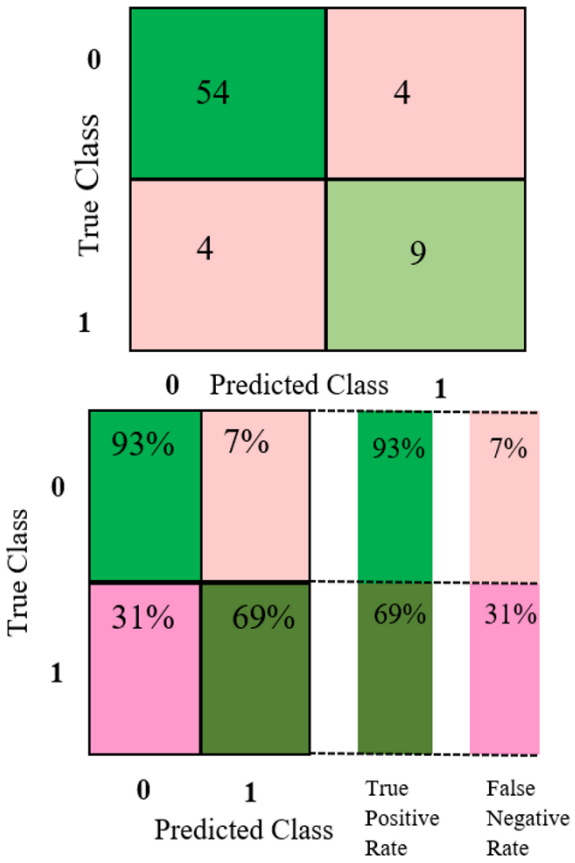


Figure 69. In the confusion matrix, the evaluation outcomes of each class's performance.

The positive class was chosen as PM in applying this evaluation, then, some statistical measures, such as accuracy (88.7%), which is the degree of association for two binary variables. The accuracy is the product of correct predictions over the total number of instances that have been evaluated. For a further investigation of features and visualisation of the prediction traits or correlation features of the

results connecting to the accident patterns and safety predictors, see the parallel coordinate plot (Figure 70).

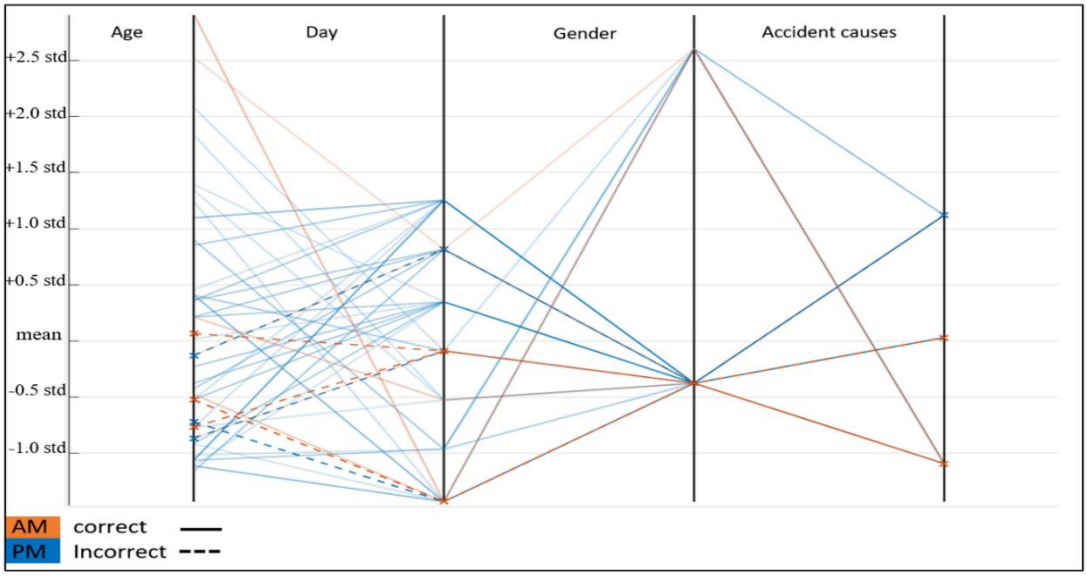


Figure 70. Graph of two-dimensional parallel coordinates.

In the model, the precision percentage and the recall rate indicate that the model had few false positives and negatives; hence, the model was more correct than incorrect when deciding whether the passengers involved in the accidents were there during the AM vs. PM. The ROC curve was formed by plotting the true positive (TP) upon the false positive (FP) (fall-out) at different threshold settings. The true-positive is similarly known as recall or probability of detection in machine learning. The ROC evaluates the sensibility as a function of fall-out, and generally, if the probability distributions for both detection and false are known, the curve can be created by plotting the cumulative distribution function of the false probability (FP) on the x-axis versus function of the detection probability (TP) in the y-axis (see Figure 71) [283].

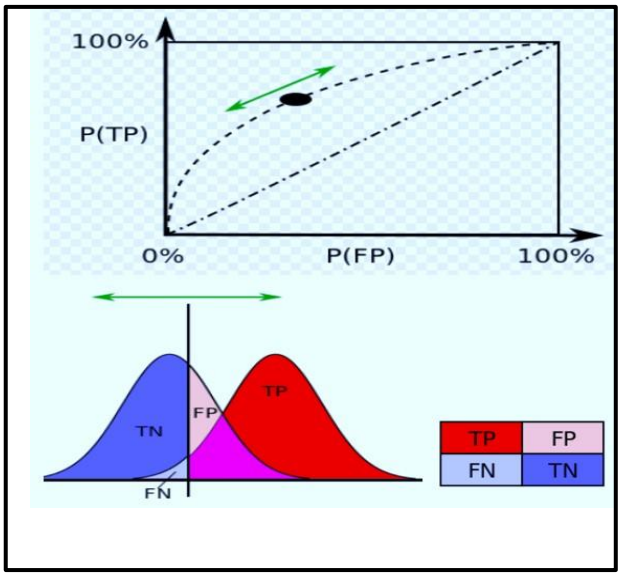


Figure 71. The plot of the cumulative distribution function.

The area under the curve (AUC) was measured under the ROC curve using this evaluation model with positive class PM and a threshold of between 50% and 80% of the data (training data) vs. 20% of the data (testing dataset). The decision tree achieves higher AUC values of 0.90, which indicates an improved classifier performance (Figure 72).

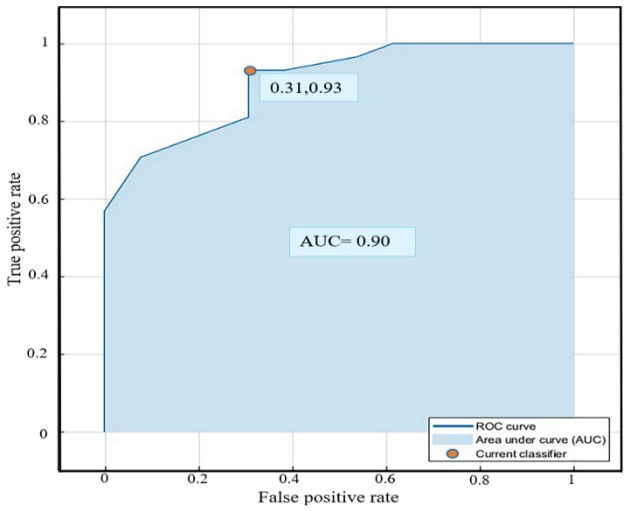


Figure 72. The ROC curve and the area under the curve (AUC) is 0.90.

It has been noted that the platform is a significant area of the station where many accidents occur, and the train interfacing with the passengers is a key aspect of the selected accidents. Some factors, such as time, where PM saw more accidents than AM, are clear. A high ratio of accidents occurring for

males has been seen in two important age groups: the young and the old. Many factors, such as intoxication, must be considered; however, in the research, it is attempting to apply ML rather than any deep analysis of the causes of the accidents. Several factors need to be involved in understanding the entire image of the accidents that have not been available in many open-source datasets. Despite accessibility to the details of the accidents presenting challenges, analysing the history of accidents via ML opens a new possibility for performing locally or internationally and presenting the root cause of the incidents and the correlations between many factors in different systems, as well as overcoming the uncertainty. Yet, more in-depth research is required. The railway industry can choose any safety dataset that has been recorded to teach an ML application with a range of analytical methods. Additionally, they can select safety datasets for analysis and validate other such user behaviours or ticketing systems to determine any correlation and thus design predictors. Furthermore, this method leads to rethinking automated data gathering to extract the most useful data and integrate systems. The model has the ability for processing safety data in real-time which delivers a strong expert system with wide sharing knowledge. The results present acceptable values, and depending on the selected data type and the prediction targets, the appropriate algorithms can be chosen. The classification of supervised ML has been applied and presented, therein showing high performance; some of the objectives of the model are as follows:

1 - Providing information that may demand that future railway stations perform in-depth analysis and classification and consider how they can obtain automated safety, therein being integrated with other developments or advanced techniques.

2 - Determining any possible shorting in current safety systems or frameworks and then improving the comprehensiveness of any sophisticated technology.

3 - Prediction of risk or consequences based on official recorded safety data.

In addition, the method affords new windows to the cloud, IoT, smart stations and smart cities. This method can be used in real-time to present the situation in the station in a timely manner. The technology allows the field to be automated and the process to become smarter. Many systems have the chance to be integrated to support risk management, such as fire and alarm systems, automated fare collection (AFC) systems, and any outer systems such as police and other agencies, along with safety record systems from other stations [14][117]. Finally, the intelligent analytical technique employed in this example yielded more useful knowledge on train station safety, which will be valuable in the future when building risk management for rail stations all over the world within the wide scope of BD.

6.8 Deep Learning and risk management

Face recognition is currently a key component of computer vision, with various applications in autonomous vehicles, human-computer interactions, CCTV cameras, robotics, health care, medical imaging, and homecare technology. Vision sensors can be used in railway environments due to technological advancements. CCTV cameras, for example, are widely used and rely on a variety of camera sensors; these cameras are designed to prevent and maintain safety and accidents in railway environments. DL has been proposed as a method for risk management in railway stations as a subset of machine learning in AI. As a result, it has resolved some risks in this case by combining vision data from several points in the system, including both still frames and motion video. Station CCTV applications can be upgraded and used to capture video frames and gather data representing human actions and motion; the resulting data contains spatial and temporal information from a variety of locations within the station, like platforms. Then, a deep learning technique is used, which is primarily based on a collection of algorithms that attempt to model high-level features in the data, so unsafe acts can be detected. Multiple frames feature datasets are used to train the model. For an illustration of the novel concept and for a more comprehensive application of the risk management and how the

process changed, the conventional risk management method moved to a CNN model process with two processes in parallel, thus improving the automated process, as presented in Figure 73.

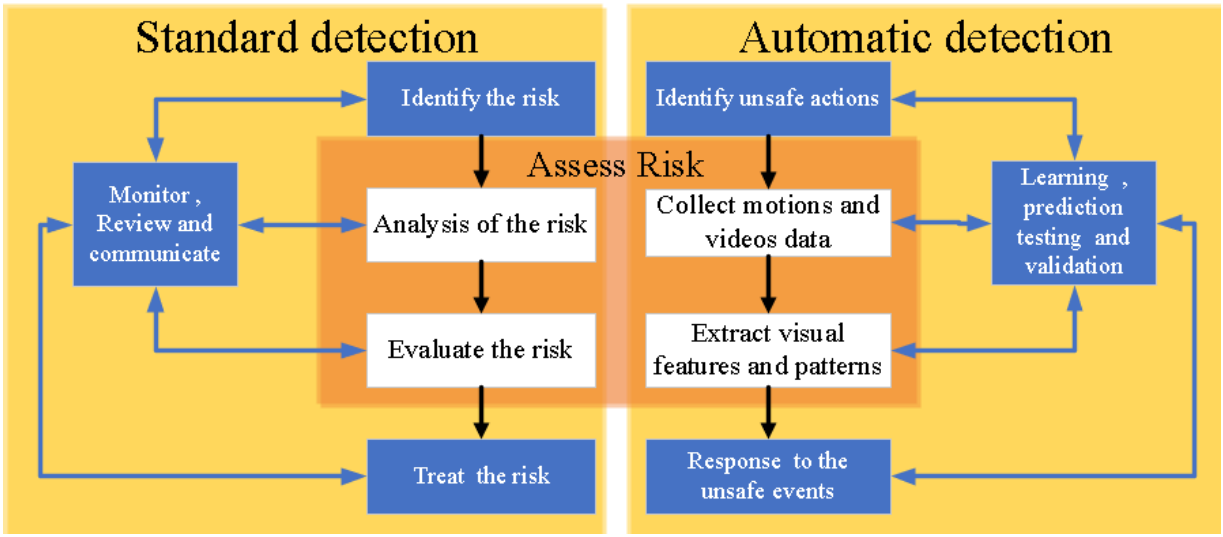


Figure 73. The proposed CNN approach for the next generation of railway station risk management scheme.

6.9 Case study 2

In this study of case study 2, DL has been used as a convolutional neural network (CNN) to detect passenger falls, slips and trips (FST).

To move ahead with the proposed use of CNN to detect falls, the system must be able to identify passengers' complex movements and address several passengers' falls at once. When a fall is detected, the mechanism is expected to notify railway station control and provide pertinent information, such as the time and location.

Such research focuses on how to use technologies like CNN to BDA collected by railway systems in order to create risk-recognition frameworks—in the case of FSTs—risk in railway stations, as shown in Figure 74.

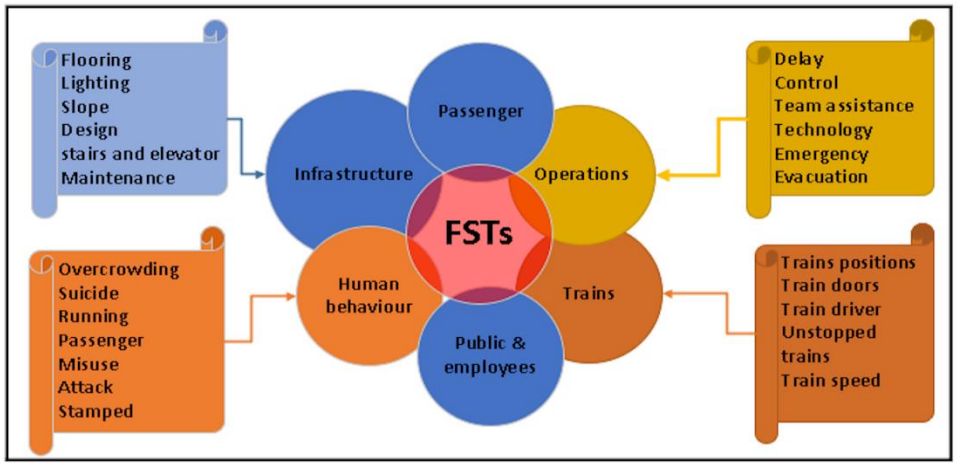


Figure 74. The convergence of FSTs and system aspects.

FTSs are common in stations, and their causes include human factors such as people running, and also factors such as floor damage (due to wet or muddy conditions), a lack of lighting, and poor steps design. Overcrowding, stampedes, security and emergency evacuation are all risks associated with FTS. Passengers may fall into the gap between the train and the platform, onto the track, or even under trains, and such accidents can escape notice and even go unheard by the train driver or station staff, which also contributes to overcrowding and panic. A team on the platform might miss a passenger stuck in the doors, people standing too close to the train, or children—all of whom are at risk. Station CCTV cameras can record a large amount of data, which is usually archived for a period of time before being removed. In fact, with the operating system that are used, it is incredibly difficult to identify risks and manage an emergency situation as soon as it occurs. CCTV management systems (in control rooms) are passive: they have a limited capacity to keep stations secure, and the recorded data can only be used as evidence by the police. As a result, current results depend on workers whose jobs involve constant viewing of video screens or supervision on site, which is extremely costly and challenging at all times and at each point in the station. In such situations, human error is likely to be high, and camera positions may not completely cover all station areas. As a result, it is critical to

effectively keep a close eye on the threats and other factors affecting passengers in the station via an automated approach and to issue a timely alert in the event of an emergency. The Keras library is used in this analysis, which is an open-source neural network library written in Python that allows for fast prototyping. It also keeps the CNN up to date and runs smoothly on both CPUs and GPUs. Keras can use raw images as inputs to the CNN model, which extracts features. Table 16 provides an overview of the experimental setup.

Table 16. The Keras library packages example installed detailed results after running the model.

Layer (type)	Output Shape	Param #
Conv2d_1 (conv2D)	(None,62,62,64)	1792
Max_pooling2d_1 (MaxPooling2)	(None,31,31,64)	0
Conv2d_1 (conv2D)	(None,29,29,128)	73856
Max_pooling2d_1 (MaxPooling2)	(None,14,14,128)	0
Flatten_1(Flatten)	(None,25088)	0
Dense_1(Dense)	(None,128)	3211392
Dense_2(Dense)	(None,64)	8256
Dense_3(Dense)	(None,32)	2080
Dropout_1(dropout)	(None,32)	0
Dense_4(Dense)	(None,32)	1056
Dropout_2(dropout)	(None,32)	0
Dense_5(Dense)	(None,1)	33
Total params: 3,298,465 _Trainable params: 3,298,465 Non-trainable params :0		

Available processors are used to run the system in this part of the study; however, for large data, more efficient CPUs and GPUs are recommended. The sequential model form was chosen, and it builds a model layer by layer, which is the easiest way to construct a model in Keras. then chose Conv2D layers with 64 nodes each to deal with input images as 2D matrices, and for the convolution kernel, a 3 X3 filter matrix was used. The main design blocks of a CNN framework are convolutional layers, which learn the features that are appropriate for distinguishing between a ‘falling' image and a ‘not falling' image. Each convolutional layer uses a collection of kernels to apply a convolution operation to the outputs of the layers before it.

6.10 Preparing data

In this analysis, the CNN goal is to take image data from capturing the passengers in a station and categorise each image into one of two categories: dropping or not falling. The dataset was split into different frames with established labels (falling or not falling), which served as training data for the classifier. Different activities and complex dropping events from various locations, such as falls on stairs or in the gap between the train and platform, were selected from several open-source sequences to create the dataset, as shown in Figure 75.



Figure 75. (a) Dataset of passengers who fall/slip or trip in stations (b) Dataset of passengers' images who are almost in a normal situation at stations.

The model is executed with an Intel Core i7 processor, 64 GB of RAM, and an NVIDIA GeForce MX 150 graphics card. Data from open sources have been collected in the form of still frames and images. In fact, finding such data is difficult due to privacy issues and a lack of availability for a range of causes, including the fact that data is regularly removed from data centres and the difficulty of finding, privacy and collecting such data. In such case for labelling, data be grouped, categorised; the images depict some potentially dangerous passenger behaviours (see Figure 76), and the videos and images collected span several countries. The data raise serious questions that should be taken into account in future station design and operations and have many aspects such as:

- People standing in potentially dangerous positions between the train and the platform and near moving trains;
- Some people trespass into track areas and are located in restricted areas.

- Some passengers have pressed the train track, and may other clothes get entangled in train doors; also, some others cross the tracks to take shortcuts between places.
- Elderly people trip and fall on escalators, and other travellers and other passengers misuse the escalators.
- Children and those susceptible to fainting are falling into the gap.
- Technological and lifestyle effects appear, such as taking selfies close to the moving trains.
- Wheelchairs that fall down escalators and staircases.
- Passengers who are located in prohibited areas.

However, for training deep learning models, the minimal data available for each dataset are insufficient. As a result, it supplemented the collected data with Charfi et al. [284]'s Le2i dataset, which includes a large number of falling positions, and the dataset was split into two parts: training validate and test, which used 80 % of the dataset for training, 10% for validation, and the remaining 10% for model testing in this report.

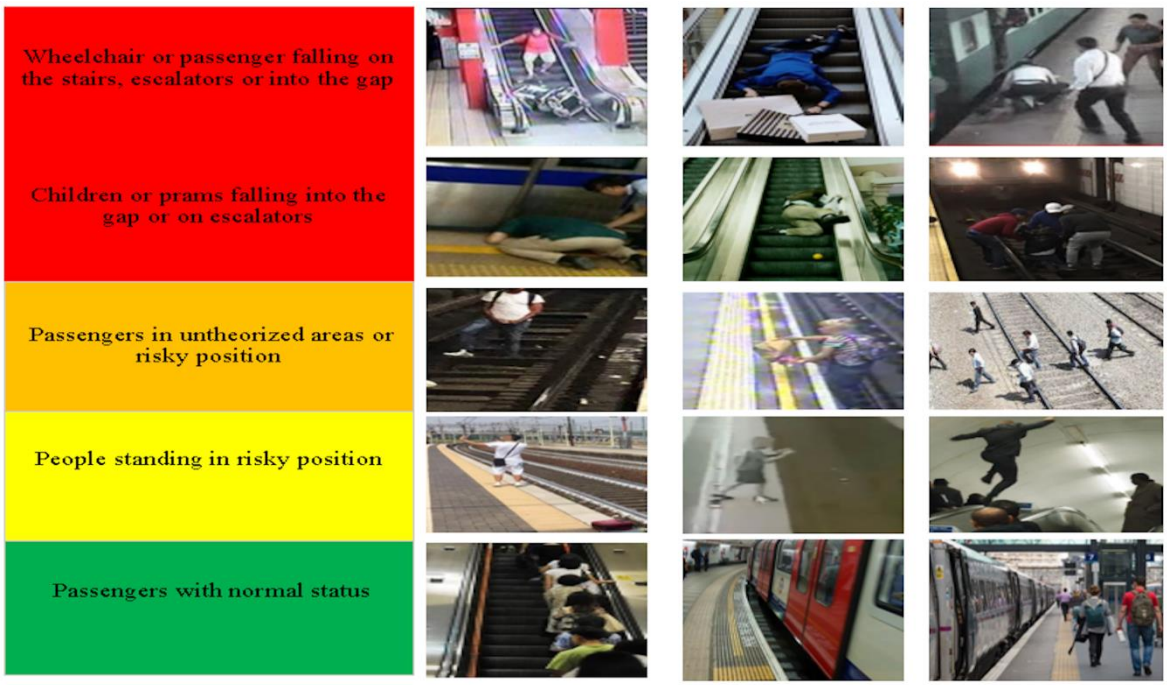


Figure 76. Different traveller behaviour samples.

6.11 The analysis and prediction

More than the data limitations, the data (both images and videos) obtained from the site need extensive cleaning. Many limitations are imposed by the variety of sources; photographs of low quality and obstructed vision (to the point that the targets are not visible) must be avoided. After being educated, the model is used to classify the risk of falling using a test dataset of photos that have not been seen before, having divided the data into three sets and using images that were chosen at random (training, testing and prediction).

The CNN training results in an improvement in accuracy for each model training iteration; thus, as the error decreases, the model output validation data reach an appropriate level. After training, the prediction ability of the model was evaluated on the test sample (see Figure 77).

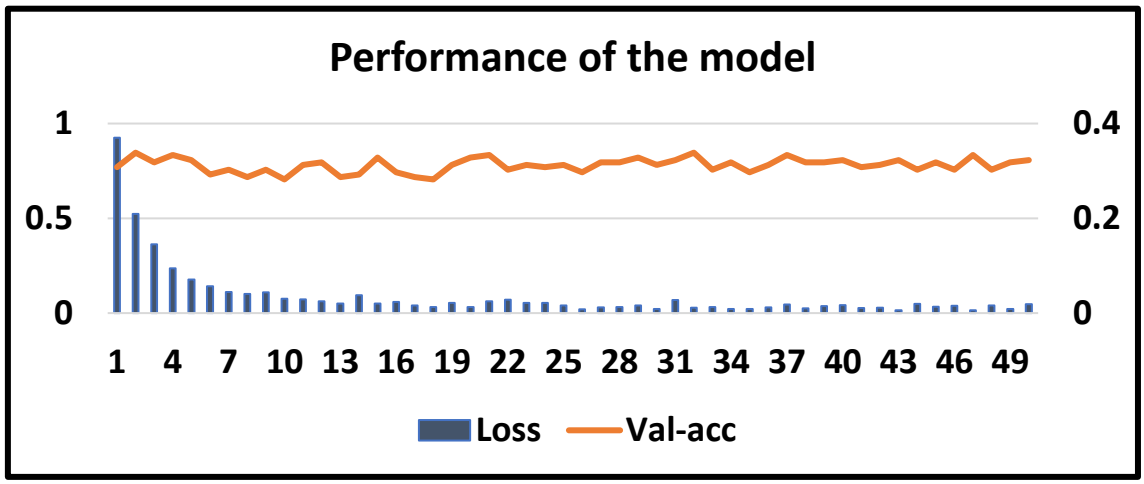


Figure 77 The accuracy of the trained model and the model loss.

There are several advantages of deep learning in risk management, including the following:

- Gaining a real-time capacity to help reduce risks is possible.

- Several field subsystems, such as maintenance, security, traffic, and passenger models, can be combined to shape actions that consider multiple factors.
- Machine learning can successfully incorporate lessons and knowledge into the learning process, which is important for safety systems.
- The efficiency of operations in stations and other railway-related areas can be increased.
- Both passenger and employee interactions can be enhanced, which will boost the market's overall picture.
- Data collection can be improved to take advantage of more efficient interactions between assets and people.

6.12 The validation

Performance measures can be estimated from the qualified model production during the testing phase. For this analysis, it used indicators like accuracy, precision, recall, and F1-score, as well as the receiver operating characteristic curve (ROC).

The model concentrates on finding the possibility of falling in predictions. As a result, it selected data that depict the frequency of falls and their behaviour, as well as people in potentially dangerous positions. Different dataset cases are analysed in order to measure danger and determine the best groups. Table 17 and Figure 78 show two cases: falling and not falling (case 1), as well as three categories: dropping, not falling (normal or safe station), and dangerous behaviour (case 2).

Table 17. Detailed outcome of the experiment's models

Model ID	Number of classes	Classes labels	Accuracy	Number of Parameters	Dataset
CNN-0	3	Fall Not fall	81.90	3,301,635	1782

		Risky position			
CNN-1		Fall Not fall Risky position	71.71	3,301,635	1826
CNN-2		Passenger Fall Risky behaviour Safe station	77.90	3,301,635	1781
CNN-3	2	Fall Not fall	80.77	3,303,681	2820
CNN-4			**82.2**	3,298,465	1980
CNN-5			75.00	3,298,465	2885

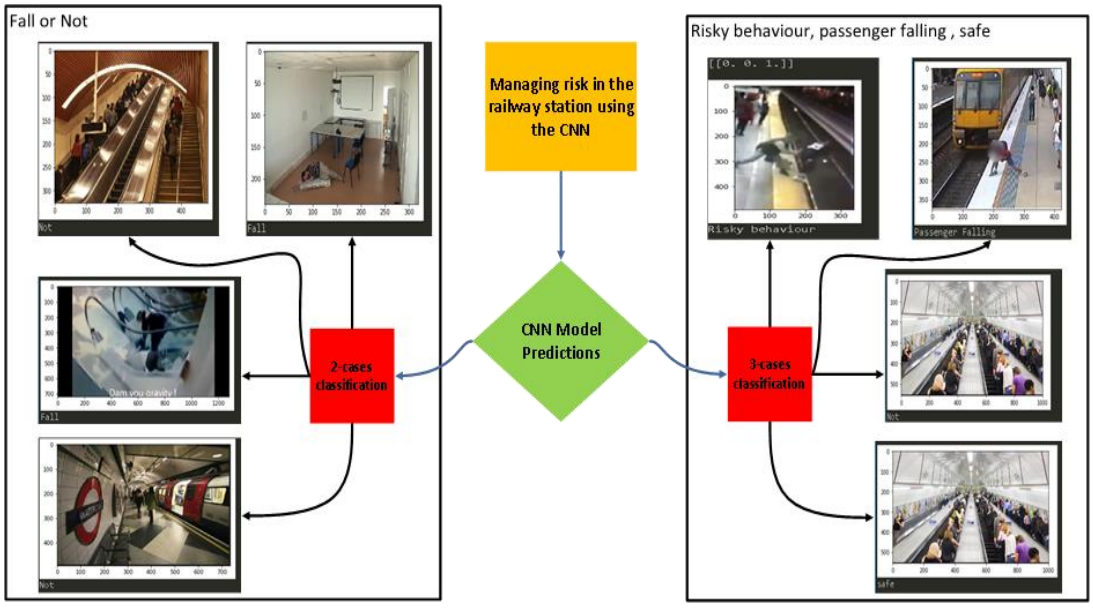


Figure 78 Examples of passenger action clusters predictions

The prediction model uses a two-class prediction (case1) to classify passenger behaviour which shows positive results. A true positive prediction is one in which the prediction is positive, and the ground-truth value is also positive (TP). False-positive (FP), true negative (TN), and false-negative (FN)

values can all be determined in the same way. As shown in Table 18, these four values can be represented as a 2×2 contingency table, also known as an uncertainty matrix [21].

Table 18. As a confusion matrix, show the classification performance evaluation results.

Confusion Matrix: [[26 3] [11 33]]		The original value	
		True	False
Predictive value	True	TP	FP
	False	FN	TN

AUC=82.33	Precision	recall	F1-score
Fall	0.70	0.90	0.79
Not Fall	0.92	0.75	0.83

The ROC is a metric that reveals the relationship between continuous variables by reflecting their sensitivity and specificity. Figure 79 depicts the ROC curve, which presents success in making (fall) predictions. Due to a shortage of risk-class photos in accident cases at stations, the risk and non-risk classes have an unequal number of images. As a result, when using the available images to train a deep learning model, it is difficult to model safety vs risk.

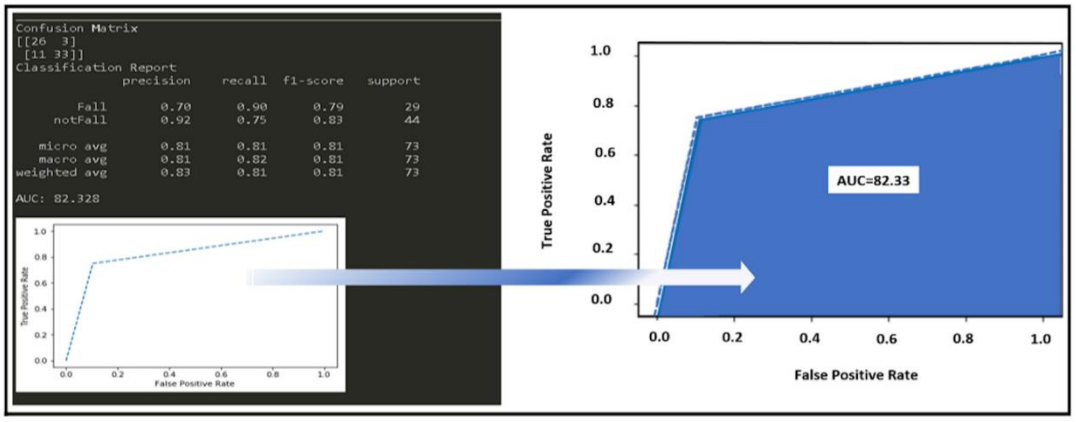


Figure 79 ROC curve of the model.

6.13 Summary

ML has examined two case studies in this chapter, proving the value of transforming risk management from a traditional approach to one that incorporates cutting-edge technology. Firstly, using DT provides a robust and beneficial analysis of the history of safety records. The applicability of the DT has been demonstrated to this safety task for railway stations. Moreover, ML presents conceivably beneficial classifications and prediction performances and interpretable accident details.

The data have been selected to be of high-level quality. The proposed novel application via DL offers original concepts and outcomes via assessing the risk from the images; the test process proves the robustness of the results. That will open a new era of risk management in many fields including the railway industry, as it can save time, cost, and protect lives by providing safe railway stations. The method demonstrated the validity of the model and the distributed analysis of the data. Additionally, it was employed to determine the relevance and importance of the chosen accident conditions. This method achieves good prediction accuracy and high and acceptable performance to be able to process safety data in real-time. Secondly, based on a CNN, applying DL algorithms fosters the identification of unexpected events in railway stations. This early detection would reduce the risk, and allow for faster responses in emergency or evacuation situations. Compiling additional datasets with training examples would undoubtedly increase accuracy and cover a broader range of station threats. It is essential to integrate some of the systems in the railway stations and possibly use the automated gathering of the data to extract the most useful data from many indicators. Finally, this technology (DL) based on computer vision will dominate and play an important role in the field and direct the risk and safety to a high and reliable level in the future.

CHAPTER 7: FUZZY AND NEURAL NETWORK FOR OVERCROWDING RISK MANAGEMENT IN RAILWAY STATIONS

7.1 Introduction

Without doubt, many threats affect railway services and operations, and neglecting them would significantly reduce the success rate of such operations in stations and across the whole system. Overcrowding, for example, causes an unsafe atmosphere, falls/slips or trips, and can result in violence and stampedes; it also increases security vulnerability and evacuation problems, resulting in more disruption and negative feedback. As a result, if planners, decision-makers, safety community, including risk managers, are capable of assessing and analysing the risk of rail stations in a timely and accurate manner, operational risk management will become more confident and effective and provide a higher level of services and safer station environments. In such a dynamic area of risk management, the stations have been seen as complex. So, in order to investigate AI in such threats, an effective method like ANFIS, which is a mixture of ANN and Fuzzy, assists decision-makers in controlling risks such as overcrowding, which is a common risk in many stations around the world. Moreover, the pandemic (COVID-19) proves the importance of the proposed method utilising technology to manage the crowd, as controlling overcrowding is significant and connected to other risks such as evacuation, delays, and quality of services.

7.2 Adaptive neuro-fuzzy inference system (ANFIS)

ANFIS is a mixture of the computational power of ANN and the high-level reasoning capabilities of a FIS. The parameters of the adaptive node have learning rules, and the adjustment of the parameter is based on error values [249][255][256]. The ANFIS model is used to solve parameter recognition

problems. This is accomplished using a hybrid-learning rule that combines back-propagation, gradient descent, and the least-squares approach (see Figure 80).

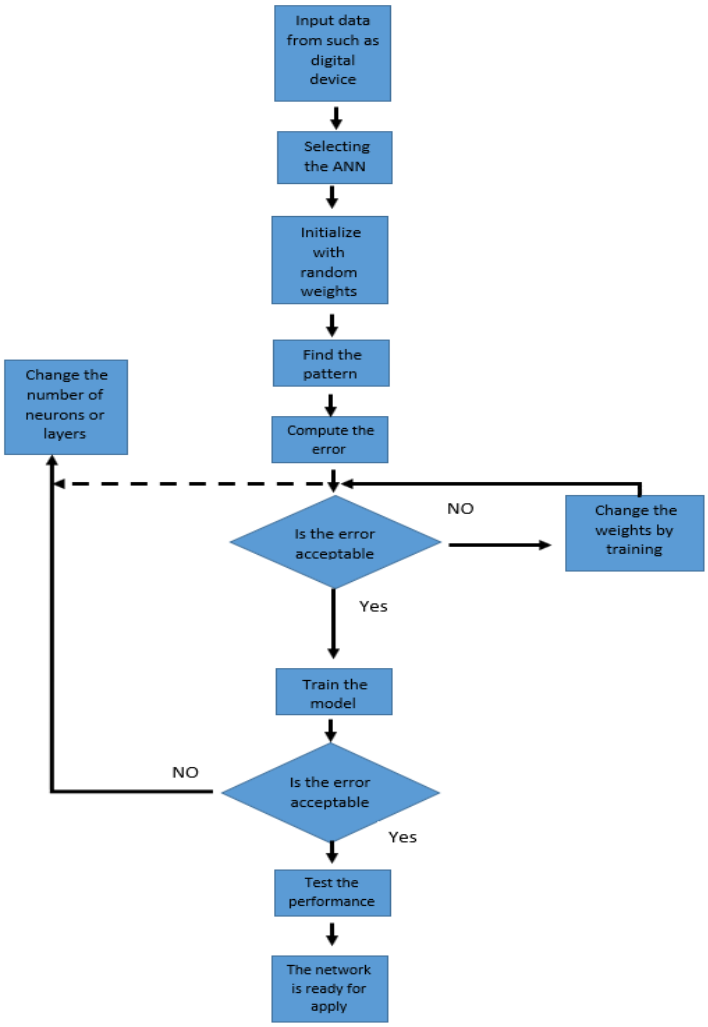


Figure 80. The Back Propagation (BP) Algorithm.

It is a multilayer feed-forward network that uses an ANN and a FLS to map inputs into outputs, as shown in Figure 81. As a result, it is FLS built into the framework of an adaptive ANN. ANFIS can also build a network implementation of IF/THEN rules [250] . The ANFIS combines (neural networks and fuzzy inference systems) modelling and controlling. It has some benefits, including adaptation capability and the ability to capture the nonlinear structure of a process. The model has both numerical and linguistic knowledge and rapid learning capacity, with more transparency, and fewer

memorization errors compared to the ANN [285][286]. Moreover, the model has a FIS in the construction of a predictive tool and an adaptive network that maps a provided input data set to its corresponding output data set. The model has been used for problems involving prior knowledge and provides a method for interpretation [287].

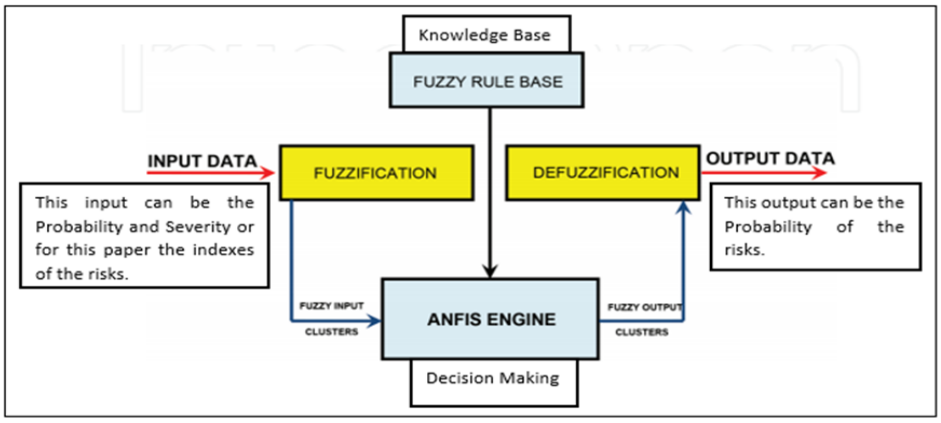


Figure 81. The fuzzy logic inference method framework for risk control in railway stations.

7.3 Case study1

For predicting the station overcrowding risk level, two parameters (retention rate index and transfer efficiency index) were chosen as inputs. So, need the transfer time (in seconds) and the transfer distance to measure the station transfer efficiency, which is an important metric (the station state). In an ideal scenario, passenger traffic on the platform is a periodically replicated phase linked to the train departures and arrivals. The departure and arrival of trains at the platform has been assumed as a prompt action that completes fast. Assume that a train designated (jn-1) is leaving from the platform (i) with some of the specified passengers on board. The number of passengers still on the platform has been reduced to the bare minimum. The passengers will then begin to congregate again until they hit the full density level just before the next train arrives (jn). This cycle is continued, and depending on train schedules and boarding, the commuter size of the sample can increase or decrease See Figure 82.

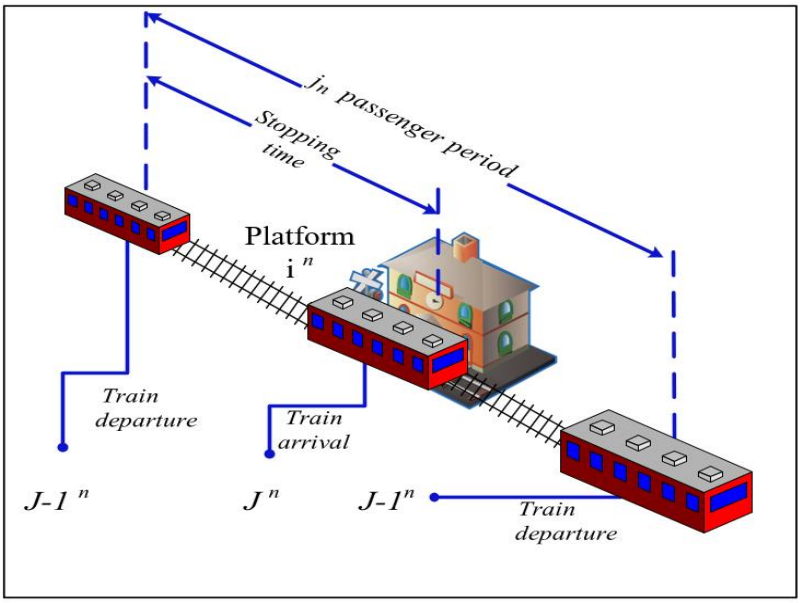


Figure 82. An ideal platform is represented in 3D, with simple train movement and timelines.

The amount of equipment in the channel that leads to the platform (such as stairs and escalators) has been suggested to have to influence the transfer efficiency index in the station channel position. The transfer efficiency η_T can be determined using the following equation if v is the walking velocity and (d) is the distance between equipment:

$$\eta_T = \frac{\sum_{k=1}^{n} \frac{d_k}{v_k}}{T_{to}}$$ Equation 10

The retention rate, which is the second index, measures the number of people that remain on the platform waiting for the second train due to a lack of room on the current train. Depending on the area of the targeted location and the standard applied to measure how many people can stand in one square metre, the density of these passengers reaches a maximum level to be the overcrowding threshold. Passengers on the platform are divided into three groups based on train departure and arrival times, assuming the process is completed rapidly:

Train departure j^{n-1}

Train arrival j^{n}

Train departure j^n

As a result, the following formula can be used to measure the terminal's retention rate η_R as an indicator for the proportion of stranded passengers:

$$\eta_R = \frac{N^J_{r\,i}}{N^J_{w\,i}} \qquad \text{Equation 11}$$

For the ith platform in jth passenger period, the $N^J_{r\,i}$ are the passengers who wait on the platform and cannot get in the current train, so they planned to wait for the next one. For every specific terminal, the number of stranded passengers is restrained by the total waiting for passengers $N^J_{w\,i}$ and boarding passengers, so the stranded passengers can be calculated by the maximum number of passengers remain after the train occupied. Even with the presents formulas, yet the indices not calculated, the parameters (retention rate index and transfer efficiency index) been calculated in literature before, and in this case study, the outcomes data are used as data set for the application. The experiments to capture such information need observation in a real platform which is not an easy task and need much authorisation from operators and differ from country to country depending on the policies. However, to make the process automated and smart it can be integrated with another system such as CCTV and sensors in the trains to count the data in real-time and for long period[185][288].

7.4 Preparing data

The volume of passengers waiting on the platform can be evaluated based on the platform's size and other factors like human factors (e.g., passenger belongings, seasonality, schools' effect or events), then the scheme for overcrowding can be evaluated, and risk level determined to execute the ANFIS model. As a parameter set, existing data from the Chinese metro system (see Table 19) been used. Part of these data will be used to train the prototype, and the rest will be used to test the model. This available data based on on-hand data collected in real-time during peak hours, actual data from the

literature was obtained from the GuoMao station of the Beijing metro in China [288]. Congestion in the station is the risk case, and it will be assessed using the indexes.

Table 19. Lu and Han present a case study of measured data on the station platform during peak hours (2016).

Time	Transfer Passenger %	Train capacity %	Stranded passengers %	Waiting passengers %	Stranded Ratio %	Transfer efficiency
7:31:30	0.77	100	0.00	68.29	0	1.62
7:33:30	0.65	100	0.00	57.62	0	1.62
7:35:30	0.78	100	0.00	68.90	0	1.62
7:37:30	0.78	100	0.00	69.21	0	1.45
7:39:30	0.76	100	0.00	67.07	0	1.45
7:41:30	0.84	0.83	0.40	74.70	0.11	1.45
7:43:30	0.88	0.83	0.55	85.98	0.13	1.28
7:45:30	0.82	0.83	0.31	83.84	0.07	1.28
7:47:30	0.75	0.83	0.00	72.56	0	1.28
7:49:30	0.93	0.83	0.80	82.62	0.19	1.55
7:51:30	0.83	0.83	0.32	89.02	0.07	1.55
7:53:30	0.87	0.83	0.49	82.93	0.12	1.55
7:55:30	0.86	0.83	0.45	85.06	0.10	1.67
7:57:30	0.89	0.83	0.62	87.80	0.14	1.67
7:59:30	0.98	0.83	1.00	98.78	0.20	1.67
8:01:30	0.90	100	0.63	99.09	0.13	1.22
8:03:30	0.90	100	0.65	92.07	0.14	1.22
8:05:30	0.87	100	0.51	89.63	0.11	1.22
8:07:30	0.93	100	0.80	92.68	0.17	1.47
8:09:30	0.95	100	0.88	100.00	0.17	1.47
8:11:30	0.72	100	0.00	81.40	0	1.47
8:13:30	0.92	0.83	0.75	81.71	0.18	1.61
8:15:30	0.93	0.83	0.80	97.56	0.16	1.61
8:17:30	0.96	100	0.25	100	0.05	1.61
8:19:30	1.00	100	0.42	93.29	0.09	1.26
8:21:30	1.00	100	0.43	96.65	0.09	1.26
8:23:30	0.81	100	0.00	79.88	0	1.26
8:25:30	0.93	100	0.12	82.62	0.03	1.21
8:27:30	0.94	100	0.18	85.98	0.04	1.21
8:29:30	0.94	100	0.15	86.59	0.04	1.21

7.5 The analysis

Instead of the risk likelihood, the model proposes a risk scale, which has been built in the form of a 4-choice range (very unlikely, unlikely, likely, very likely). Congestion in the station has been chosen as a dangerous scenario. The probability of congestion is divided into sublevels that fluctuate between the index impact groups, and the following indexes will be used to test this:

- Passengers stranded (Platform)
- Transfer efficiency (Channel), with each level grouped according to its risk level.

The probability of congestion can be divided into several different levels, which will fluctuate depending on the index impact levels. There are four levels of risk, each with their own set of criteria, and their own colour scheme for easy identification. The function dimension of the inputs must be transformed to be based on the likelihood of risk level in the suggested form because the data are linked to the ratio (stranded) and transfer efficiency in the station. The covariance matrix is then obtained, and the risk level's patterns and colours are noted, as seen in Tables 20 and 21. For each of the model's inputs, the scale retains the same set of values [288][289][290].

Table 20. The magnitude of the risk likelihood level.

Result	**Description (Retention rate of the platform)**	**Description (Transfer efficiency)**	**Risk levels** **Indexes**
Very unlikely	Full capacity and have chance to ride without being stranded	Passengers can move easily	D
Unlikely	A few stand and wait for the next train	A few lines before the lift and stairs	C
Likely	Some are waiting for the next train	Some lines before the lift and stairs	B
Very likely	Highly crowded and no chance to ride	Very slow movement	A

Table 21. The two indices with the risk matrix and scale.

		Index 1				Index 2				Risk levels of indexes 1 & 2	Risk values index 1	Ratio Range of index 2	The scale factor (R)) of index 2	Risk value level
		A	B	C	D	A	B	C	D					
Index 1	A									D	R≤0	1.3<R ≤1	32.5 ≤R ≤0	4
	B													
	C									C	0<R≤15	1.7< R≤1.3	42.5 < R < 32.5	3
	D													
Index 2	A									B	15< R<50	2<R ≤1.7	50< R≤ 42.5	2
	B													
	C									A	R>=50	R>=2	R>=50	1
	D													

The retention rate was calculated using data obtained during the peak hours of 07:30–08:30, with a super peak at 07:45–08:15. Furthermore, the loading factor (1.05–1.02) was used in the calculations to obtain the stranded ratio. The index for channel transfer efficiency was then determined, considering variables such as walking velocity (1.5 m/s) and passing time (107 s).

The overall transfer time was 152 seconds, including the time it took to travel the distance to the platform. According to data on transfer performance, after 08:00, congestion remained constant and passenger movement was smooth. As seen in Figures 83 and 84, the level of congestion risk fluctuated between very likely and unlikely.

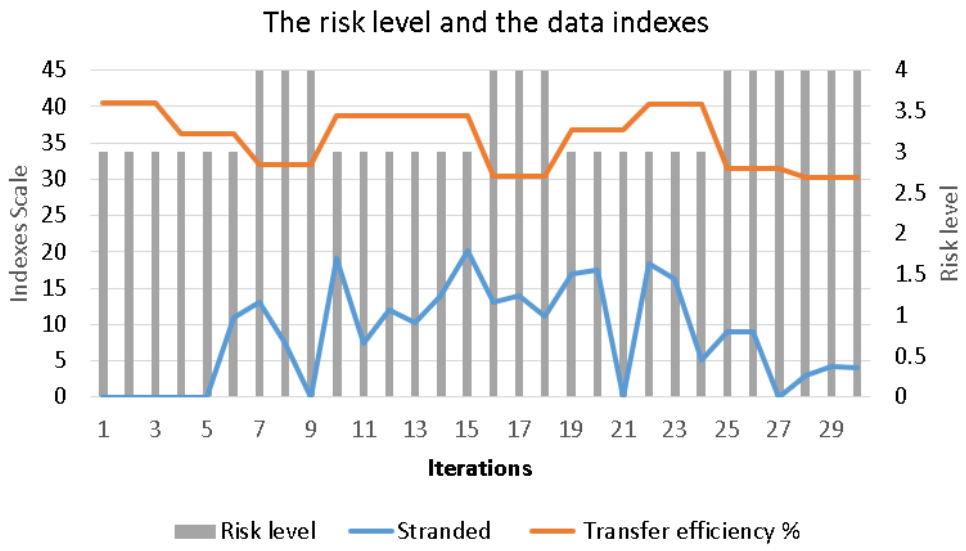


Figure 83. The relations among the inputs and the risk amount.

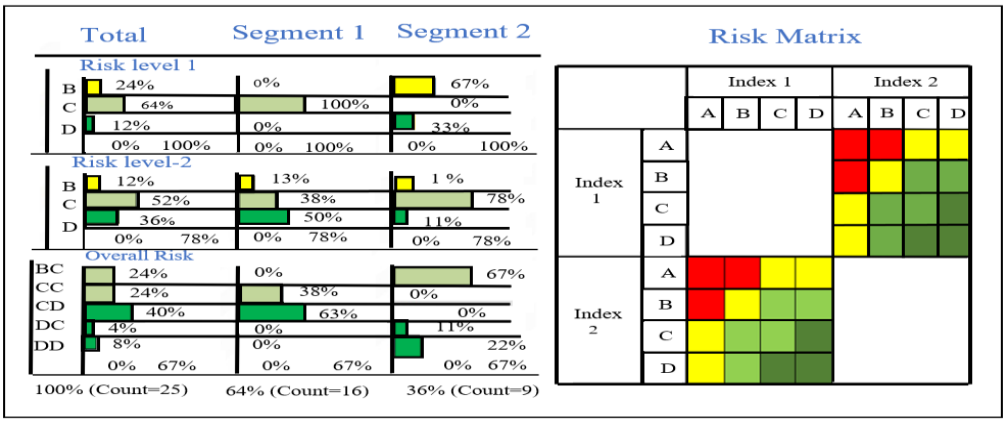


Figure 84. The findings of a latent class analysis using 2 segments and 25 samples.

7.6 Prediction and analysis

It used supervised machine learning to train regression data to assess data for further study and exploration of the data, as suggested by the ISFMR system to advance dynamic risk management in railway stations. A comparison of regression learner optimizable models was performed, and cross-validation was used to protect against overfitting. Furthermore, the model with optimised hyperparameters was used to reduce the model mean squared error (MSE) (Best point

hyperparameters). One of the most important statistical approaches for analysing the relationship between predictor variables and responses is correlation analysis.

Regression trees and Gaussian process regression model (GPR) support vector machines (SVM), and ensembles of regression trees are four different hyperparameter optimization models that have been used. The best fit by the minimum MSE is seen in the GPR analysis (see Table 22). Figure 85 shows the MSE and the value of optimization results for each model at each iteration, whilst Figure 86 displays the Minimum MSE Plot, which updates as the optimization runs.

Table 22. Error tests based on a mix of hyperparameter optimization models.

Model Type	**RMSE**	**R^2**	**MSE**	**MAE**
Optimizable SVM	0.013656	0.99	0.0001864	0.01128
Optimizable Tree	0.063478	0.82	0.0040295	0.051834
Optimizable GPR	0.00647	1	0.00004194	0.0054809
Optimizable Ensemble	0.02726	0.97	0.0007431	0.020145

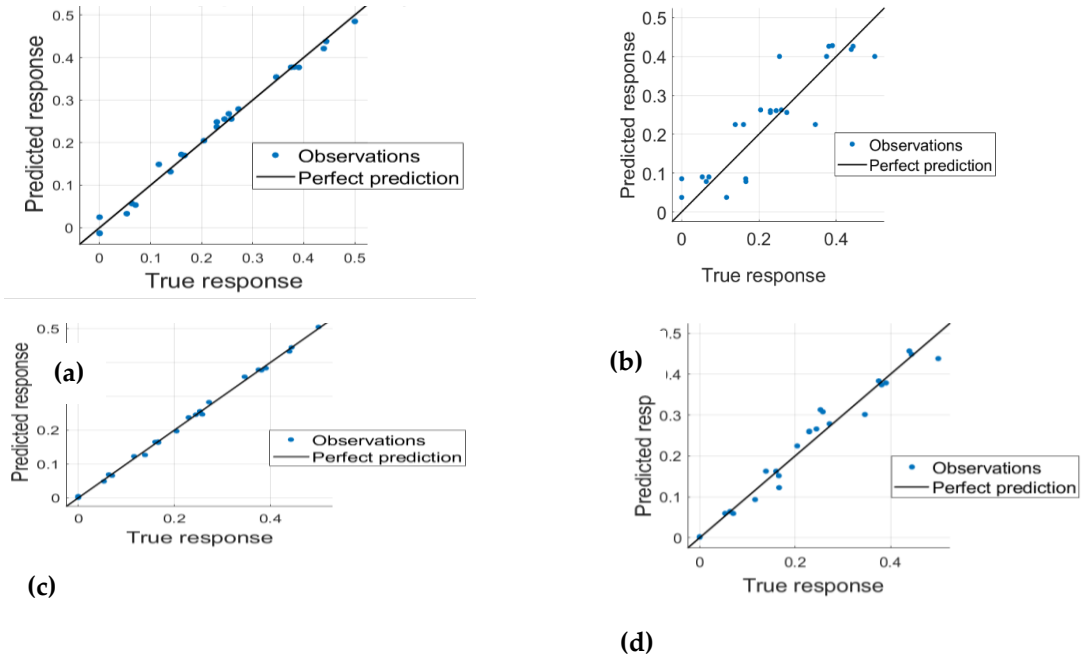


Figure 85. The expected and actual values for different types of regression learner optimizable models: (a) prediction (Optimizable SVM), (b) prediction (Optimizable Tree), (c) prediction (Optimizable GPR), and (d) prediction (Optimizable GPR) (Optimizable Ensemble).

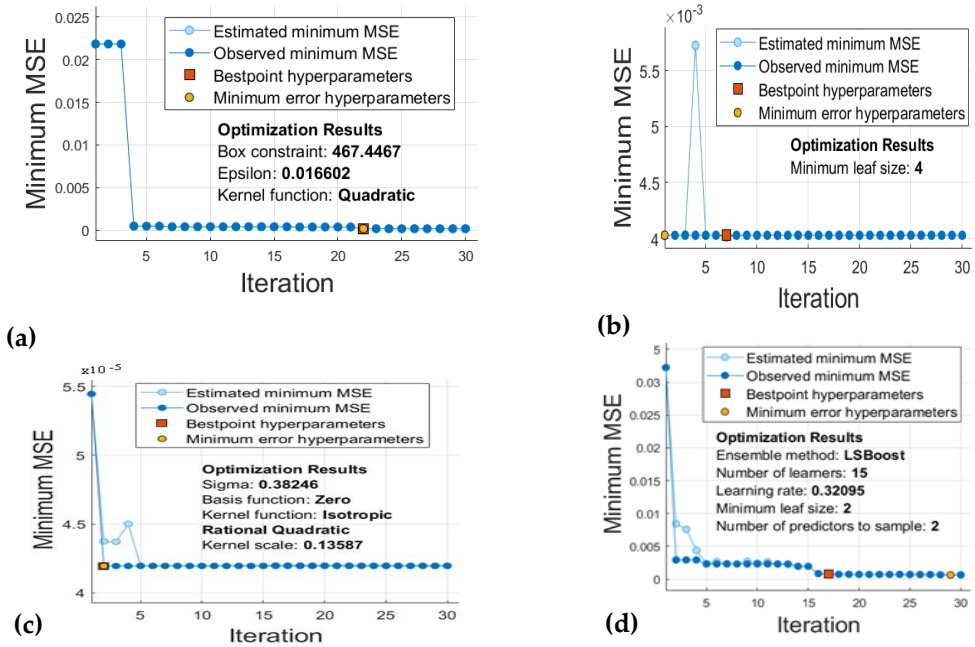


Figure 86. Expected and actual values for various regression learner optimizable models: (a) prediction (Optimizable SVM), (b) prediction (Optimizable Tree), (c) prediction (Optimizable GPR), and (d) prediction (Optimizable GPR) (Optimizable Ensemble).

7.7 Learning and validation

Sugeno-type fuzzy inference systems have been modelled, designed, trained, and tested using the Neuro-Fuzzy Designer. The system was trained using the pre-calculated indices and risk matrices. As shown in Figure 90-A, it chooses a partitioning system in which grid partitioning is the default scheme for FIS structure generation with ANFIS. Datasets were used to train the model using the chosen data, and the residue was used to test and validate the system. Additionally, subtractive clustering was used for rule generation, with the degree of impact, squash factor, acceptance ratio, and rejection ratio set to 0.5, 1.25, 0.5, and 0.15, respectively, as seen in Figure 87-B. The number of MFs allocated to each input in the ANFIS model will vary.

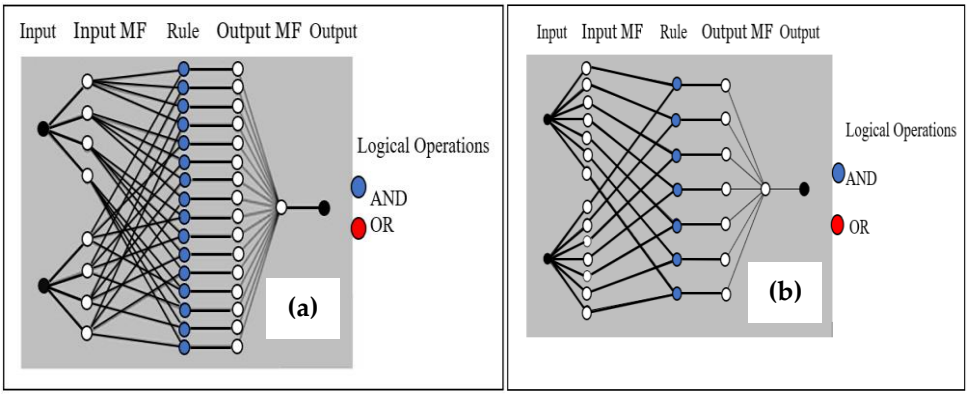


Figure 87. (a) ANFIS model architecture for overcrowding risk level estimation, with the branches indicating the ANFIS model structure in addition to the feed-forward network, and (b) ANFIS network after subtractive clustering.

In addition, the risk indices were calculated in two parts. Figure 88 displays the latent class estimation results assuming that the indices and total risk are homogeneous. This emphasises the importance of platform design and train schedule and load punctuality in meeting the waiting passenger numbers. A low-risk level implies a secure platform with smooth passenger flow; however, other platforms or locations nearby could be overcrowded.

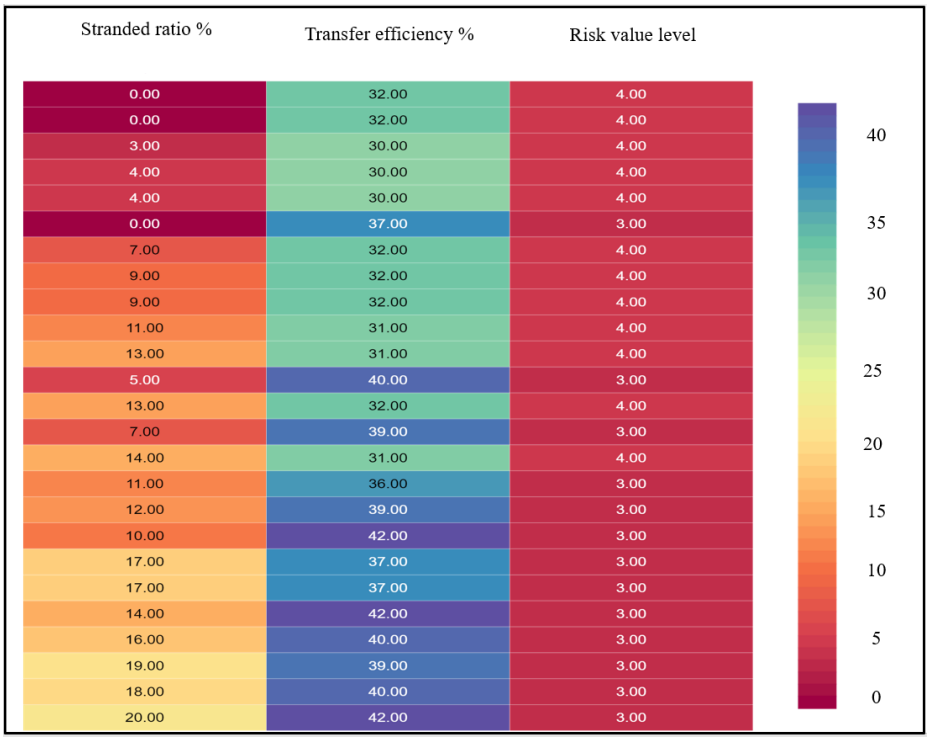


Figure 88. The association between the inputs and the risk levels is depicted as a heat map.

Figure 89 shows the trained FIS, which contains 16 rules (clusters). There are 16 Gaussian curve MFs built through each input. Sets of processed data were used to execute 150 cycles of learning during ANFIS preparation. Several fuzzy sets were assigned to the output data using the Sugeno-type FIS method to consider the relationship between the input variables (risk indices) and the output variable (risk level), as seen in Figure 90.

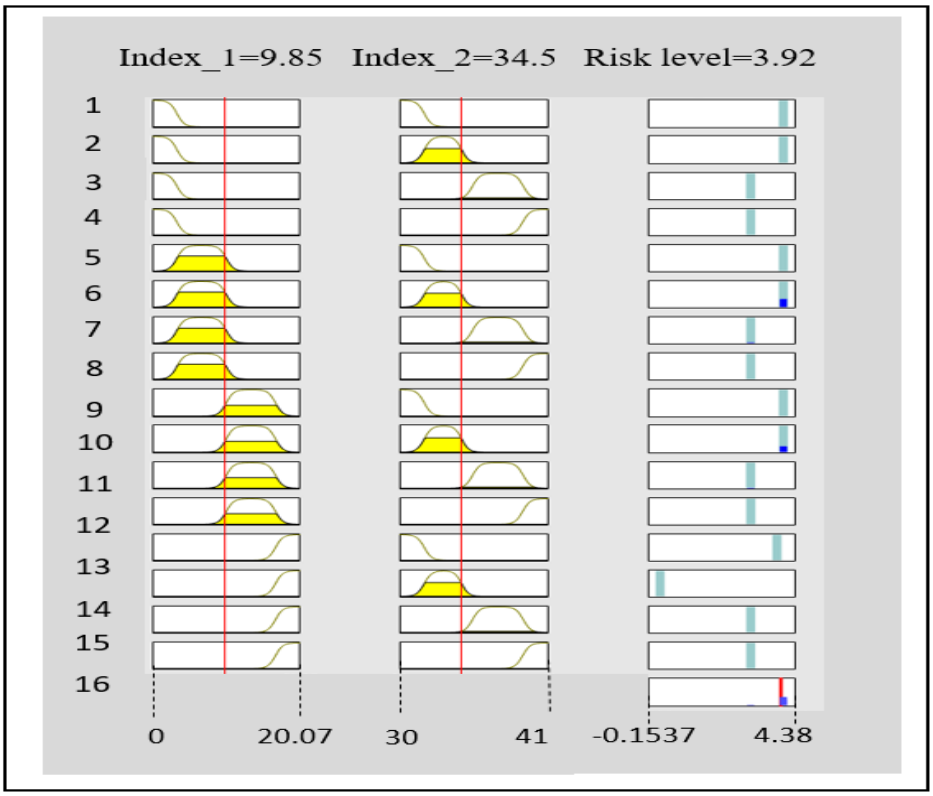


Figure 89. Main ANFIS of indexes and risk performance (Matlab's rule viewer) were trained.

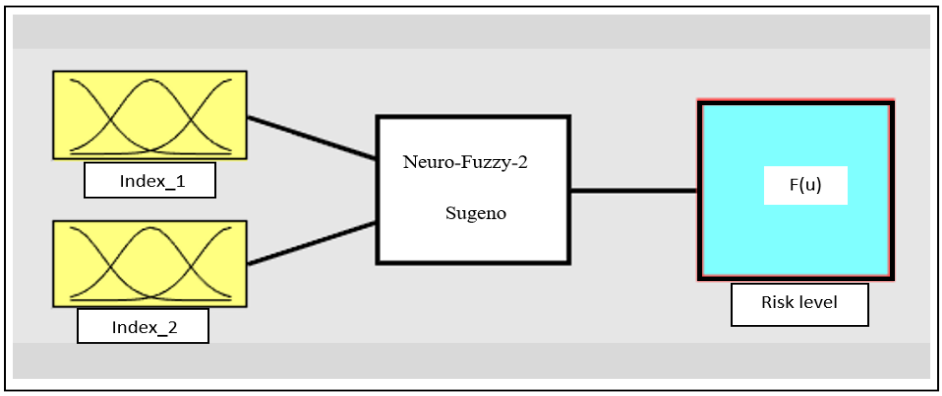


Figure 90. In the proposed method, the Sugeno-type FIS method is used.

For a comparison of RMSE values for the training and testing process, the experiments covered different MFs. Also, the mode train FIS uses the optimizer methods as a backpropagation alone or hybrid which is a combination of backpropagation and least-squares-type method to allow the fuzzy learning from data. While the number of output is only one, grid partition method is used in

establishing the rule base relationship between the input and output variable. During training with 150 epochs, the MF of (dsigmf) and (psigmf) perform the most effectively with minimum RMSE under linear MF type output with training hybrid method (see Table 23).

Table 23. RMSE results from the training and training datasets for various MFs with Membership Functions Number (4*4).

ANFIS info	MF	Number of nonlinear parameters	Training RMSE		Testing RMSE	
			Back Propagation	Hybrid	Back Propagation	Hybrid
Grid partition Method	Trimf	24	0.073364	0.000119	0.214272	0.234563
	Trapmf	32	0.064496	0.000061	0.152923	0.771647
	gbellmf	24	0.063835	0.000001	0.100628	0.103621
	gaussmf	16	0.063897	0.000001	0.176282	0.079462
	Gauss2mf	32	0.074480	0.000001	0.136584	0.87736
	Pimf	32	0.058077	0.000061	0.19069	3.89485
	Dsigmf	32	0.071541	0.000000	0.117291	0.726872
	Psigmf	32	0.039306	0.000000	0.717619	0.720104
Number of linear parameters				48		
Number of Epochs				150		
Number of nodes				53		

In terms of error, types of membership functions, and output process, the ANFIS models have shown substantial performance variations when compared to the evaluation criteria. The ANFIS models tend to be sensitive in various MFs and data sets based on all of the RMSE values. The ANFIS architecture is influenced by the types of MF inputs, data, and MF types of outputs. Figure 91 depicts a graph of the training error. The model has been set up to train steps of 150 epochs, error diminishes, and error fluctuations to reach a steady state. Clearly, the limited amount of training data is a big factor in results. Additionally, the expected ANFIS example does not necessitate time-consuming iterative training.

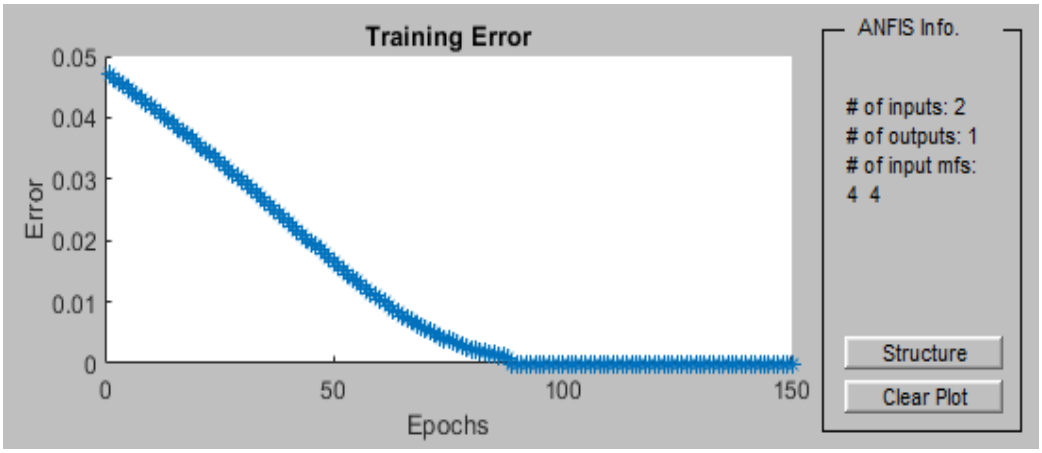


Figure 91. The progression of trained fuzzy system errors.

Figure 92 depicts the trained fuzzy system's error pattern, as well as an illustration of the RMSE increasing with training and the ANFIS paradigm's performance. This demonstrates that the expected short-term risk level for both the training data and the FIS performance is very close to the actual results. The surface plot (the three-dimensional (3D) surface) shows that evaluating the two indices has a major effect on the predicted overcrowding prediction (see Figure 93).

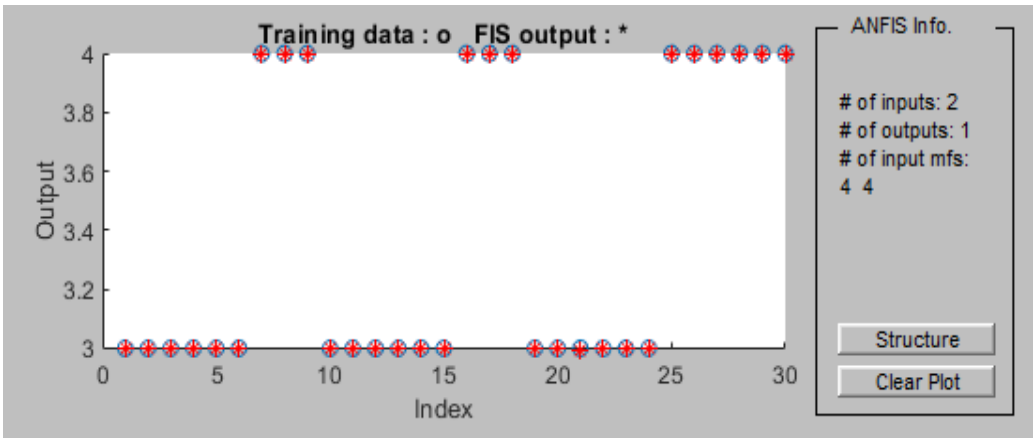


Figure 92. Training dataset and FIS output with two inputs and one output.

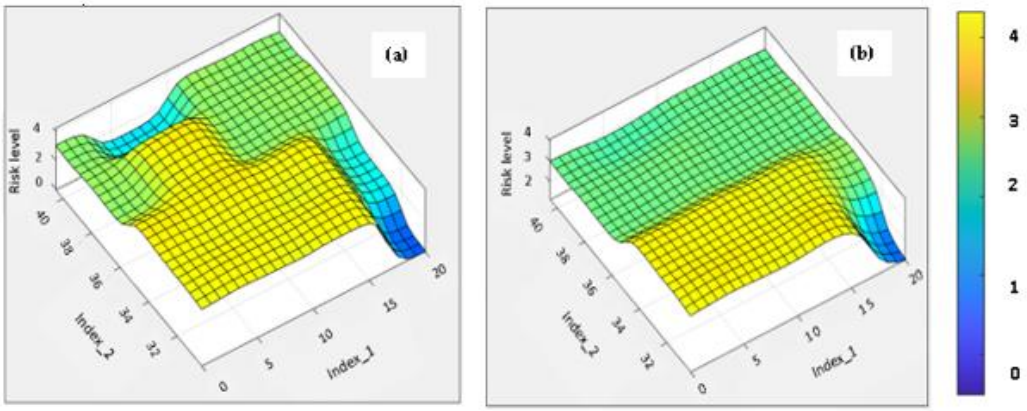


Figure 93. Example of the rule base system's 3D surface plots adapted for data for various mf types: (a) dsigmf, (b) gbellmf.

In general, the 3D diagram indicates that the risk level increases as the index likelihood increases, which is a reasonable inference. The same technique may be used for other risk indices. This was implemented with the aid of data over a short period of time in order to come to conclusions regarding the method's efficacy.

The results show the surface and counters of risky index levels, as well as the effect of the input index on risk level results. Different scenarios are considered when choosing the best MF, and the scenario with the smallest error is acknowledged. Depending on the MFs chosen, the response of the 3D surface of risk level versus indices shows different minimal errors. The high flow from both indices results in a high risk of overcrowding, and increased density will result from a rise in the number of passengers on the platform and reflect on other areas of the station, which is predictable; however, the approach demonstrates that the mechanism can be automated, and the machine can be learned from overcrowding risk level history. The ANFIS assessment results are compared to the observed data in Figure 94, which shows the ANFIS assessment results.

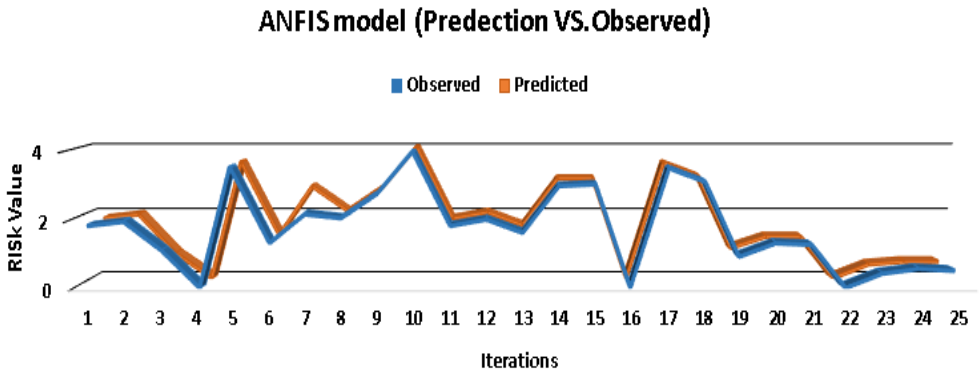


Figure 94. Comparison of the proposed model's predicted and spotted risk levels.

The results of the above studies demonstrate that the designed ANFIS model can accurately predict risk and reflect statistical features. The complex risk can be handled in real-time in the proposed model, which aids decision-making and represents the pace at which overcrowding risks are managed. It can predict the flow safety output using input parameters like the platform's transfer efficiency and retention rate.

7.8 Chapter summary

This part of the research presents the ANN with the Fuzzy applications (ANFIS), which also provides a good source of reference information relating to both (ANN-FUZZY) and their applications in risk management in the stations. The controlling risk of crowding has been conducting utilising unique concepts cope with the data using the novel signs to capture the risk level and provide health safety areas inside the stations. The application can be generalised covers many systems including HSR and metros or other public transportation. The input data can be integrated with other systems such as radar, sensors, automated gate, and the train capacity. Harvesting data is no easy task, yet the Data have been collected from real experiments which can be generalised and expanded to gather all possible indexes in different areas in the stations.

The model could identify disturbances in huge amounts of data automatically, yield the model for future development and generalise, and most importantly, gain the ability to learn. For instance, many places can be controlled so the integration systems will then cover all the platforms and the whole station crowding areas. With powerful training and historical data, the technology can be utilised to heighten analysis in the future and enhance accuracy and reduce human error. The strategy is more practical and cost-effective for rapid risk management. It may be used to either improve risk management in stations or cooperate with existing risk management systems that support decision-making. Gathering risk data in a railway station is challenging and costly because it is not always readily available, and there is a shortage of data. The selected models seem to be powerful for predicting the risks and aiding the decision-making for risk management in the case of overcrowding in railway stations. The proposed new method of railway station risk management model using ANIFS can deal with risk data index and information in real-time. It is shown that the ANFIS appears to be more reasonable and appropriate because of its smoothness in calculating the risk levels. This dynamic model of risk management can define risk levels and aid the decision-makers by producing convenient and reliable results based on real-time data. In practice, the prediction ANFIS model can be a decision smart support system for managing risks which can be captured through indicators or indexes to create rules depending on the inputs. The model has the flexibility to create a greater amount of input for future research.

CHAPTER 8: TEXT MINING FOR RISK MANAGEMENT IN RAILWAY STATIONS

8.1 Introduction

Text mining (also referred to as text analytics) is an automatic process which aims to identify textual patterns and extract valuable information and trends within unstructured data using ML from texts such as documents or reports. With the increase of technology, amounts of data increase each second, including personal data such as text data on social media or the information being captured while passengers get online tickets as well as the safety data of investigations in the accidents, which all form big data to enhance the power of AI. In practice, the availability of data has challenges and are not affordable for the researchers. This proves most modern of the concept and raises the issue of shortage in regulations. Using such data may conflict with individuals' privacy and rights (The General Data Protection Regulation (GDPR in the EU), so, utilising big data for safety analytics purposes must be considered in the legislation and regulation with a clear definition of scientific research, which also aims to protect the people; otherwise, privacy will challenge the application of the new technology. Moreover, the data set is a part of collected information for accident reports which are ‘lawful’. That makes it GDPR compliant. In addition, there is a legitimate interest as one of the GDPR 'lawful' processing. In addition, the history of real accidents data by the Rail Safety and Standards Board (RSSB) has been practised in this chapter which is offer benefits regarding accuracy, reliability, and robustness.

To provide or invest in any control safety measures at stations, it is crucial to consider the risks associated with railway incidents and identify the many factors related to accidents through a comprehensive understanding of the root cause of accidents for a wide range of stations and long

history and from all types of data. In some cases, data are available, but it is hard to achieve such a goal, while such information is written in large documents in textual form. In this chapter, Latent Dirichlet Allocation (LDA) is optimised for fatality accidents in the railway stations from textual data gathered from RSSB, in the UK railway station. This case describes using the ML topic method for systematic accident characteristics to enhance safety and risk management in the stations, providing advanced analysis. This part has presented the power of ML as a sub-method of AI using text data. In general, the history of accidents or safety (including the reports or documents or other records) provides the valuable raw material for risk management, safety, and the future, which can be used to predict and analyse the risks that might occur in stations, using a wide cognitive range with computers that can reach a high level of data processing far better than any human for such a huge set of data in the future. The documents cover the preferences of passengers, activities, and safety records data, so it is worth analysing text from safety documents to provide essential data for an expert smart system considering the privacy and local legalisations

8.2 Topic model for railway station safety

Safety history is a rich source of knowledge discovery and risk management analysis. For instance, investigation reports after accidents by a responsible authority or expert person are one of the most popular sources of safety information, and evaluation and analysis of accident causes and the consequence of the risk can be very effective for analysing behaviour, hidden risks, or causes so hidden risk causes and lessons can be learned. In fact, the language and the structure of records are important factors for using the tool to analyse documents and design a topic model.

The text data has many source forms, including social media, emails, and call recording. Such data exist in a raw and unstructured status that require transfer and cleaning as part of topic modelling to capture the necessary information. A framework based on textual source data using AI algorithms to

build a tag recordation system from safety documents has been suggested; Figure 95 shows a simple idea based on the LDA algorithm for building the process.

Such a method has the ability to explore and digest the complete history; it has the power to track, navigate through time to reveal how specific events have changed and can be adapted to many kinds of data. Moreover, to enable automation and digitisation concepts, currently, more texts are available online. The human does not have the timely ability to read, analyse, explore massive textual data, and study how it is all connected.

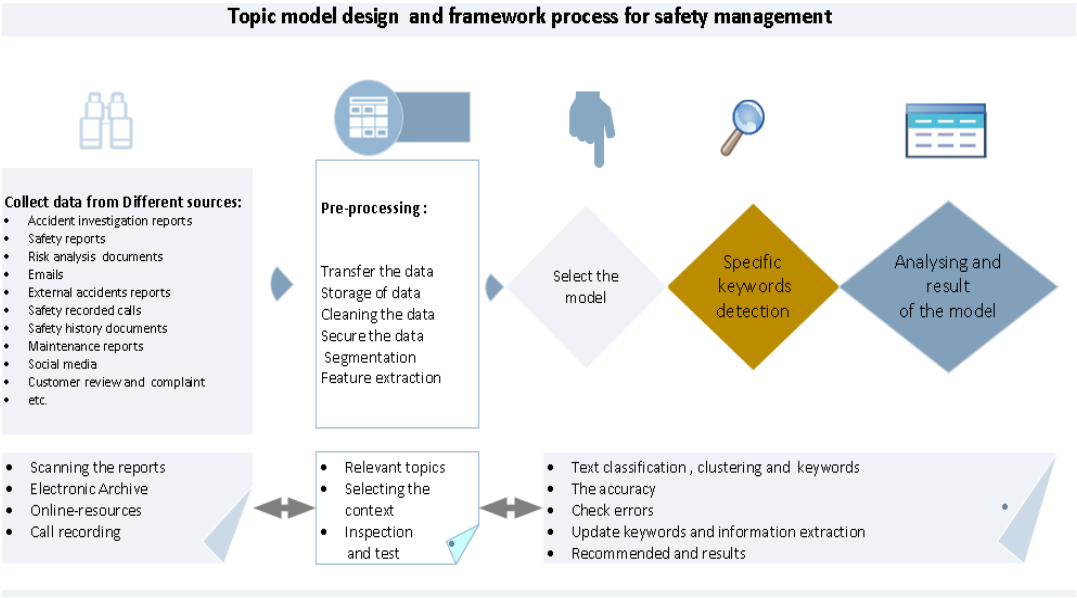


Figure 95. The framework of the topic model in risk management in the stations.

Still, lessons must be learned to prevent repeat accidents in the stations, and a huge efforts have been made in the field for controlling the issues while recommendations from investigations have been yielded for high safety level. Many reports and or documents have been recorded and presented to begin risk assessments and the accident investigation reports (typically narratives) from different organisations are indispensable. Regardless of whether the text data is structured or not, the question is whether the text data is adequate to capture all of the details of an incident and can the human brain analyse this huge amount of information? Many challenges have been expected, such as the shortage

of experts and the context in the documents, which may have nonstandard terms. Also, from generation to generation, the experience must be delivered from experts in an effective way which can be granted through design expert systems relying on smart methods to extract, learn and provide a solution from all related textual contexts. Many challenges can be decreased by the intelligent use of new methods via extracting word-level meaning from safety reports and automating and analysing this as a part of the process [291]. Also, data availability is one of the challenges facing scientific research. The adoption of the General Data Protection Regulation (GDPR) raises a number of issues for scientific research, particularly data-driven research. In fact, the implementation of the General Data Protection Regulation (GDPR) raises a series of challenges for scientific research, in particular for research that is dependent on data [292].

8.3 Preparing data

The textual data contain some key information that can be used as incident causes, the time, description of the accidents, location, seasons, the station situation, for example, crowded or not, the age of the victim and their condition, and many more. However, not all of this information is always recorded. Also, the safety process has regular reviews of much of the significant information and the risk assessment or evaluation, including safety measures from experts. From the RSSB, the raw data set has more than 2250 accidents registered in the railway stations in the last ten years.

However, the data are divided into datasets for more clarification and focus of the research scope; for example, the suicide dataset has been excluded from this work data set.

For more clarification of the data used, the time of accidents has been divided into the Parts of the Day for more mining to capture accurate times. The Morning from 5 AM to 12 PM (Noon), Afternoon from 12 pm (midday) to 5 pm (17:00), the Evening from 5 pm (17:00) to 9 pm (21:00) and the Night from 9 pm to 4 am.

The selected data set containing 163 of the accidents covers those incidents occurring at rail stations between 01/01/2000 and 17/04/2020. The following information is available for each accident and the below text data:

- The day of the week and time the fatality occurred
- Some information on the cause of the fatality
- The age range of the deceased
- The site type (physical environment) where the fatality occurred.

As a part of the data mining, the raw data needed to be processed to extract the knowledge; thus, some steps for that were applied (see Figure 96 below).

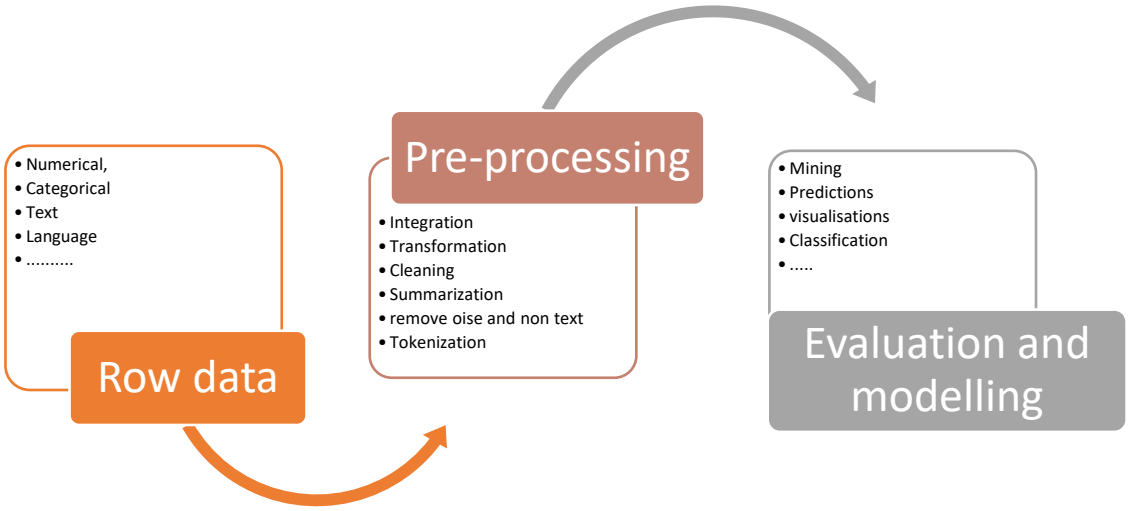


Figure 96. Data mining cleaning procedure steps (Text Cleaning and Pre-processing).

For text cleaning and pre-processing, many documents have additional words like misspelling, punctuation, stop words, slang, and others which affect the algorithms and topic model results performance. Some of the techniques have been noted in the pre-processing text data, converting the context to formal language and removing any noise as follows:

- Tokenization which breaks the context of text into meaningful elements called tokens, aiming for investigation of the words in a sentence by data, which gets split into parts [293] [294].

- Stop and noise words which are the words that do not form a key in the classification algorithms, for example (a, after, about etc.,), so they need to be removed [295].
- Capitalization where the words or abbreviation is written in capital letters, so this is converted into lower case to help account for such exceptions [296] .

For Text Cleaning and Pre-processing, many documents have additional words with misspelling, stop words, slang, and others, which affect the algorithms and topic model results performance [297]. Thus, for improving the quality of the data set and the model performance, filtering configuration is used, which allows for selecting the fields that are considered in the modelling or not, where some information that is in the unprocessed data can lead to fuzzy overview, noise and or has missing values. Fortunately, in this data set, there were no missing values. For topic model configuration, the non-language characters, Non-dictionary, and Numeric digits are excluded from the analyses. Also, uninformative words such as at, on, and or are removed. Moreover, the data used complies with the GDPR crucially as data can be processed only for the specific purpose identified and notified and for no further reason, and no personal information should be used such as names or Identification/Identity/Identifier.

8.4 Model analysis

From the illustration visualisation of the data set the times of night, evening, morning and afternoon have captured more accidents respectively. However, the day of the accident is not easy to identify; also, the adults seem to be involved in the accidents more than children and elderly people. The lineside is the location that involves most accidents, and then the stairs/bridge and escalators (see Figure 97).

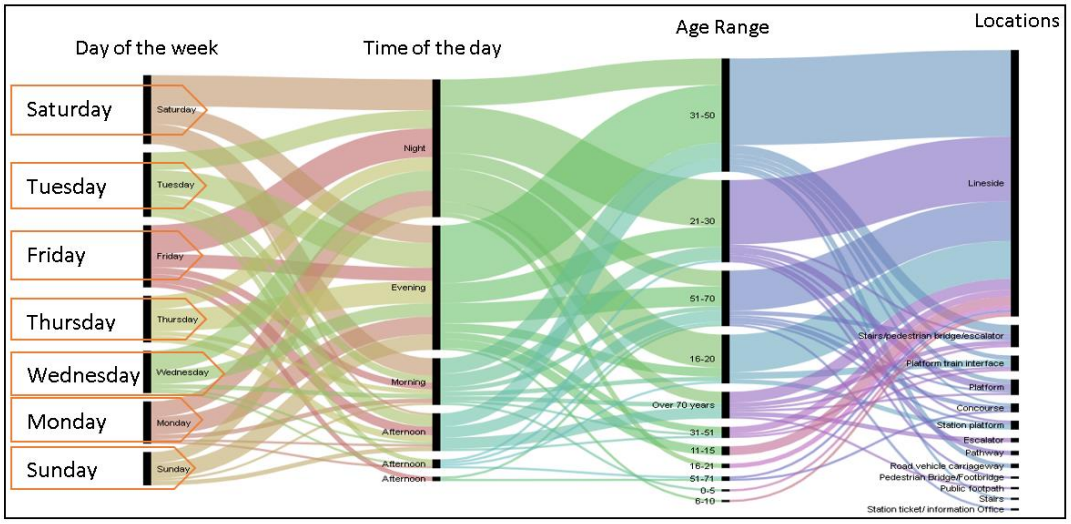
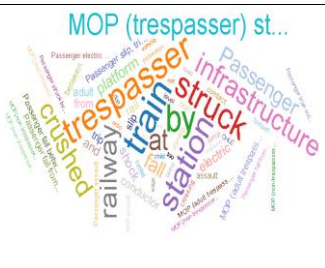


Figure 97. Example of accidents and their detailed attribute correlations.

The dataset that has been used in this work has key text traits, and information such as the day of the week and time of day that the accident occurred, including the hazardous event description and precursor description. Moreover, the Age of the victim and the Site type with data of Physical environment information is noted (see the word cloud in Table 24). The word cloud provides a quick overview, which is a key visualisation since it summaries a bulk of the text. Generally, a variety of accidents occurred, such as falling off train platforms or in the gap, being struck by trains, and electrical shock. These accidents can occur in any place in the stations, including stairs and when alighting from or boarding the train or also when there is no train at the station.

Table 24. The examples from dataset inputs word cloud visualisation.

Day and Time	Physical environment	Hazardous Information
		

Saturday is often linked to the accidents in the stations which is a day usually not as crowded as the workdays, but may be when the public are going out more as they are not working. The details and reasons behind that need more investigation, such as the factor of deficiency of assistance staff and intoxication impact. The night and evening times record more accidents than the morning and afternoons, while the evening is the safest time. In fact, this may be related to the volume of passengers, so the required data needs to be linked to each station and the number of passengers for more investigation.

This also highlights factors that need to be redefined, such as the light condition, the seasonality and weather effect. More information can be gained such as the trespass accidents, which one of the brightness words in the cloud links to the station and the contact with vehicles of the train, which presents the importance of isolating passengers from the train at a safe distance. In another view, this raises the query of overcrowding risk effects which may force the people to be close to the trains and track—moreover, the infrastructure word appeared, which raises the other queries related to factors like the station's age, layout, weather impact and the influence of intensive usage.

The lineside in the platforms is the hotspot place that interacts between the human and the machine (Platform Train Interface (PTI)) and is linked with accidents such as fall /slip and trip, electric shock and so on. The threatening consequences of accidents requesting more engineering solutions such as detecting and preventing the risks in early-stages and stopping trains in accident situations to improve passenger safety in the stations [117][20].

When looking at the time index of the day as (AM & PM), where the range of the time Afternoon, Morning and Night produce more solid information, the children and elderly passengers are more often involved in accidents in the morning. Nevertheless, the Night-time captured most accidents (See Figure 98).

Moreover, these factors overlap with the days of the week, as some accidents occurred on Saturday morning, while it seems that the Afternoon is safe throughout the week (See Figure 99). Many valuable illustrations can be found from the details of data, yet more data dimensions need to be gathered in the future, proving the importance of the data in such cases of safety analytics and providing reliable intelligent systems.

Clearly, if this approach is considered in the future for safety, documenting, and reporting systems, then the system of reporting and gathering data will be improved and become more valuable for advanced analytics such as AI methods.

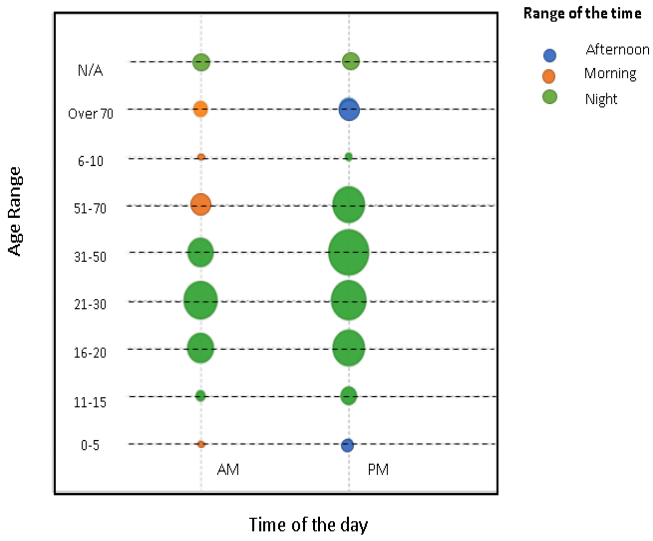


Figure 98. Comparisons of input subjects for Multiple features (the times of day and the time limit, including the age group).

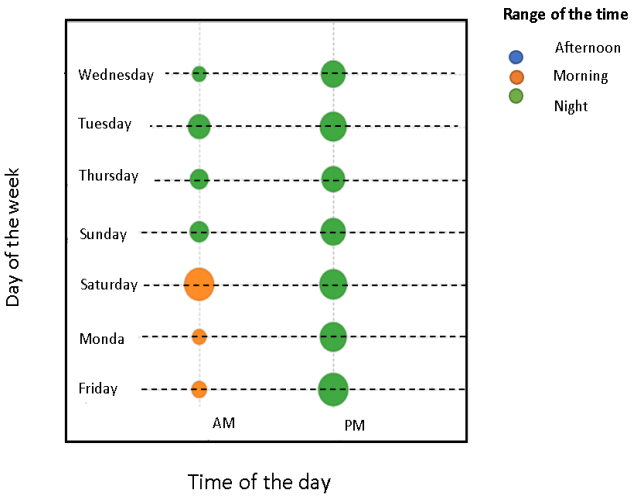


Figure 99. Comparisons of input subjects for different factors (the times of day and weekday, as well as the age group).

Moreover, it is expected that passenger behaviour is primary towards inside accidents, as passengers on the platform tend to walk or stand near the platform edges to avoid crowded areas, and there are others who run to catch the trains or stand too close in order to get on the train before others, and this can be coupled with slow responses of moving trains, or little time to react. From frequency view of the words, the risks related to fall/slip and trip, struck/crashed, train and platform with the passenger, are noticed as risk joint words from the details of the accident occurrence (see Figure 100 below).

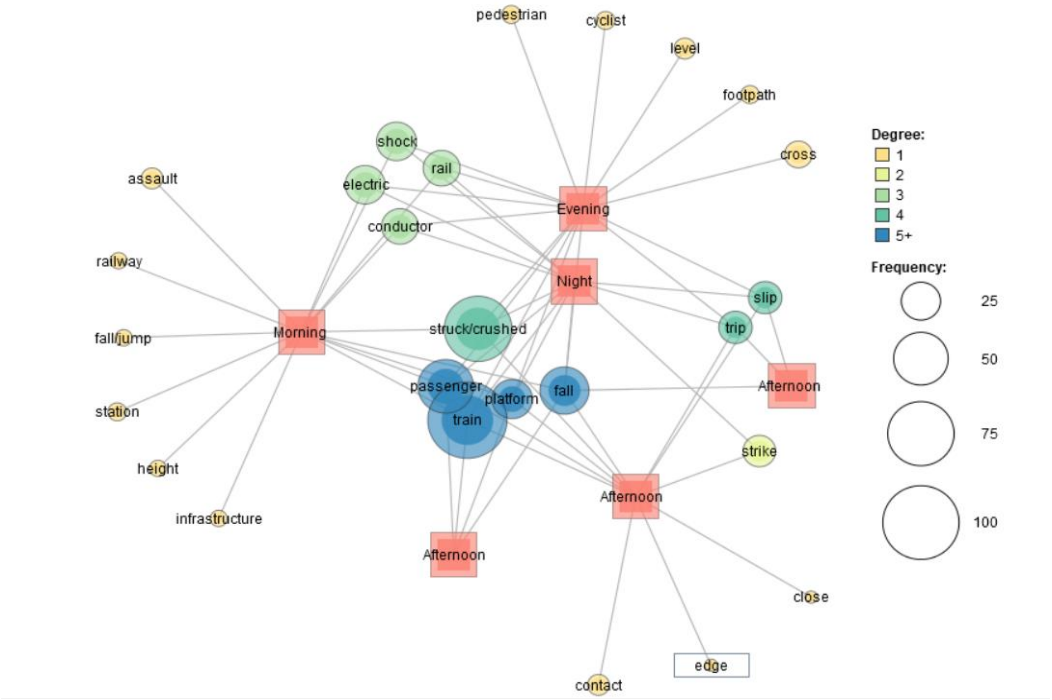


Figure 100. Co-occurrence network.

Even though the available data does not provide a deep understanding of the causes, the information was not analysed independently nor correlated with all the input attributes, and the factors related to the outcomes show the importance of the details of data for each unwanted event for future data gathering.

For example, in a fall in accident cases, the position (status) that passengers had at the time of falling and the position (forward position) that land in is key factor impact also impact the rail track as sharp and solid on the seriousness of the accident's results, and the details of the lights, flooring and the platform slope and the infrastructure status.

To detect the relevant topics within the text and learn from the topics underlying a collection of documents, the Latent Dirichlet Allocation (LDA) algorithm is implemented, which has some configurations (see Table 25). The nodes have been created in the topic map, and illustrate each topic via word probability with different sizes and colours ((see Figure 101 & 102). This collection of

models is powerful for visualisation, and each circle presents a topic, the size notes how common the topic is in the data, and the topics that are close together are semantically linked in the data.

Table 25. Set up configurations for text analytics.

The model confguration	Details
Number of topics	10
Number of top terms	10
Languge	English
Tokenize	Active for all text
Stemming	Active for all text
Filter terms	Active (Non-dictionary +Non-lnaguage words)
Sampling	100%(163) accidents-deterministic)

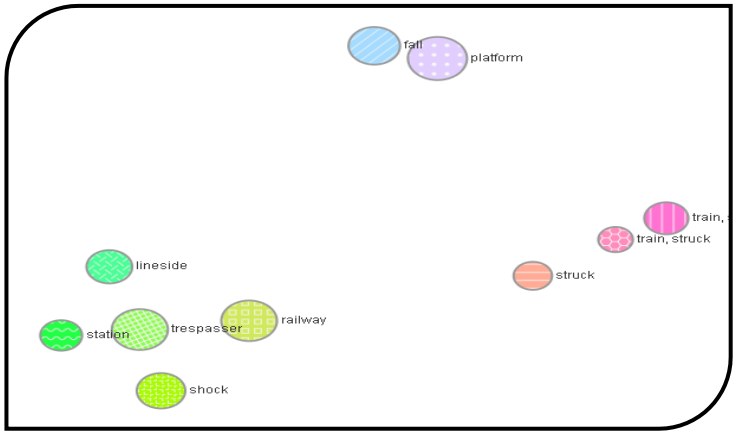


Figure 101. Topic model map with ten number of topics.

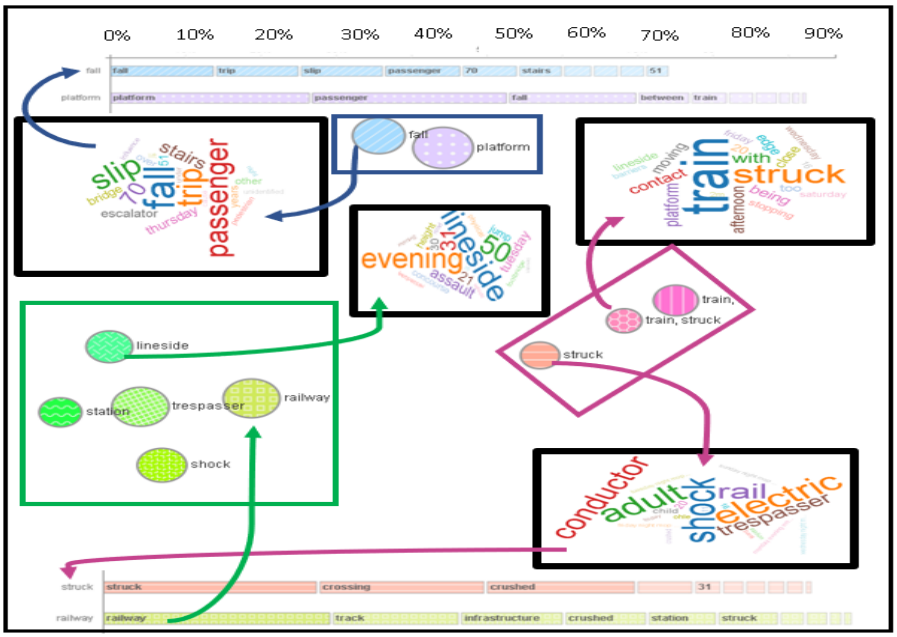


Figure 102. The topic model result map shows ten topics with examples of the words cloud and the probabilities.

In selecting the passenger as a topic specifically, the word maps and the distribution show that this topic is linked to falling accidents in the platform during the departing and arriving, also occurring more during the nights. Falling occurred in many locations such as escalators and bridges and was linked with elderly passengers, which opens up the importance of accessibility in the stations to most vulnerable people like the disabled, aged and families, also the age of mature passengers (31-50) held more dangers in the platforms in the stations. Electrical shock is linked with conductors and small age groups, which raise the importance of standards review for safety related to electrical equipment in the stations. The contact of the passengers and the train in the platform while the train moving is shows an associated cause between topics of train and moving (See figure 103 below).



Figure 103. Words map example.

Additionally, there is a correlation between the fall and slip/trip as a group, and many locations such as stairs, escalators and bridges, and the edge in the lineside, as well as the contact caused with trains (see Figure 104 below).



Figure 104. Words map example.

Such analytics presents some mysteries, which required detailed reporting to be answered in the future by the safety community, such as if this occurred because the passenger is running to catch the train or other hidden reasons such as narrow linesides, slippery floor, floor damage, or congestion. These analytics illustrate the importance of the details of accident reporting and the key factors that can be gained from such a method to provide safety measures. Each topic node has many probabilities of the words. More details can be gained from the topic distribution; for example, by selecting the words, the words most associated with the specific topic will be captured; as long as the words are changing, the topic distribution is updated (see Figure 105).

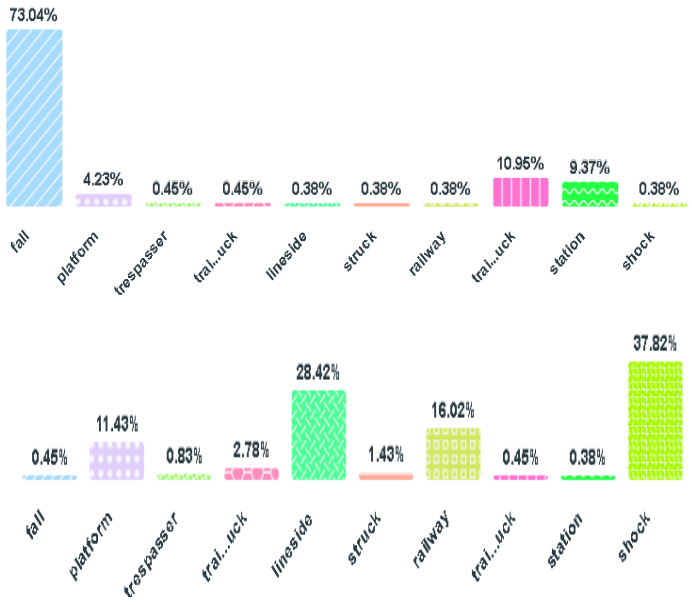


Figure 105. Example of topic distribution for different context contains different words.

The development of data analytics can be with advanced techniques such as linking the outcomes with the dictionaries with the meaning and providing suggestions for the designer and analyst as an expert system or providing a recommendation for the safety employees as a tool of an expert system.

Also, the topics modification and each topic have the distribution of the common words (see Table 26).

Table 26.Topic batch with topic distribution examples.

The text input information examples						
Day of week ▶Time of Day ▶Hazardous Event Description ▶Age Band ▶ Physical environment						
Sunday Evening Passenger slip, trip or fall in a station, his age 31-50, stairs/pedestrian						
Monday Night, MOP (adult trespasser) electric shock (conductor rail) ,21-30 Pathway						
Wednesday Night, MOP (trespasser) struck/crushed by train while on railway infrastructure at a station ,21-30 Lineside						
Thursday Evening, Passenger slip, trip or fall in a station ,31-50 Stairs/pedestrian bridge/escalator						
The predictions probabilities						
struck	railway	fall	platform	trespasser	train, struck	lineside
0.00458	0.00534	0.68626	0.04275	0.00534	0.03588	0.1458
0.00604	0.00378	0.00604	0.03248	0.03172	0.00378	0.03927
0.13585	0.17585	0.00377	0.00377	0.23094	0.01962	0.03472
0.02417	0.00378	0.75378	0.04532	0.00604	0.00453	0.11178

Applying a Batch topic distribution with some sentences that contain information leads to predicting the properties related to the inputs (see Table 27 & 28 below).

Table 27. The probabilistic description example of the words among the topics.

Day of week	Time of Day	Hazardous Event description	Sub Hazardous Event description	Precursor description	Age band	Physical environment
Thursday	Evening	Passenger slip, trip or fall in a station	Passenger slip, trip or fall (stairs)	Passenger slip, trip or fall on stairs due to ill health	31-50	Stairs/pedestrian bridge/escalator
Tuesday	Afternoon	Passengers fall between train and platform	Passengers fall between moving train and platform	Passengers fall between moving train and platform - train arriving at station	51-70	Lineside
Monday	Night	MOP (adult trespasser) electric shock (conductor rail)	MOP (adult trespasser) electric shock (conductor rail) at a station	MOP (adult trespasser) electric shock (conductor rail) in station	16-20	Lineside
Saturday	Morning	MOP (adult trespasser) electric shock (conductor rail)	MOP (adult trespasser) electric shock (conductor rail) at a station	MOP (adult trespasser) electric shock (conductor rail) in station	21-30	Lineside
Saturday	Night	Passengers fall between train and platform	Passengers fall between moving train and platform	Passengers fall between moving train and platform - train departing station	16-20	Platform train interface

Table 28.The probabilistic description example of the words among the topics.

fall	platform	trespasser	train, struck	lineside	struck	railway	train, struck	station	shock
1	0.04532	0.00604	0.00453	0.11178	0.02417	0.00378	0.00453	0.04	0.01
0	0.55363	0.00378	0.19109	0.0574	0.00529	0.0068	0.03852	0.039	0
0	0.02946	0.05891	0.00604	0.0136	0.00831	0.00604	0.04834	0.088	0.74
0	0.00378	0.05136	0.00529	0.09215	0.00755	0.03021	0.01284	0.144	0.65
0	0.63435	0.00382	0.13359	0.00382	0.00611	0.01145	0.13206	0.056	0.01

8.5 Clustering and analysing

For further analysis, the cluster has been conducted in which the data set can be used. In this case, the G-Means are considered to find the best number of cluster group with a critical value (5); contrarily, K-Means can be used for a specific number of clusters (K). 8 clusters are used in the cluster algorithm to show the largest problem, and the correlations in the topic (see Figure 106).

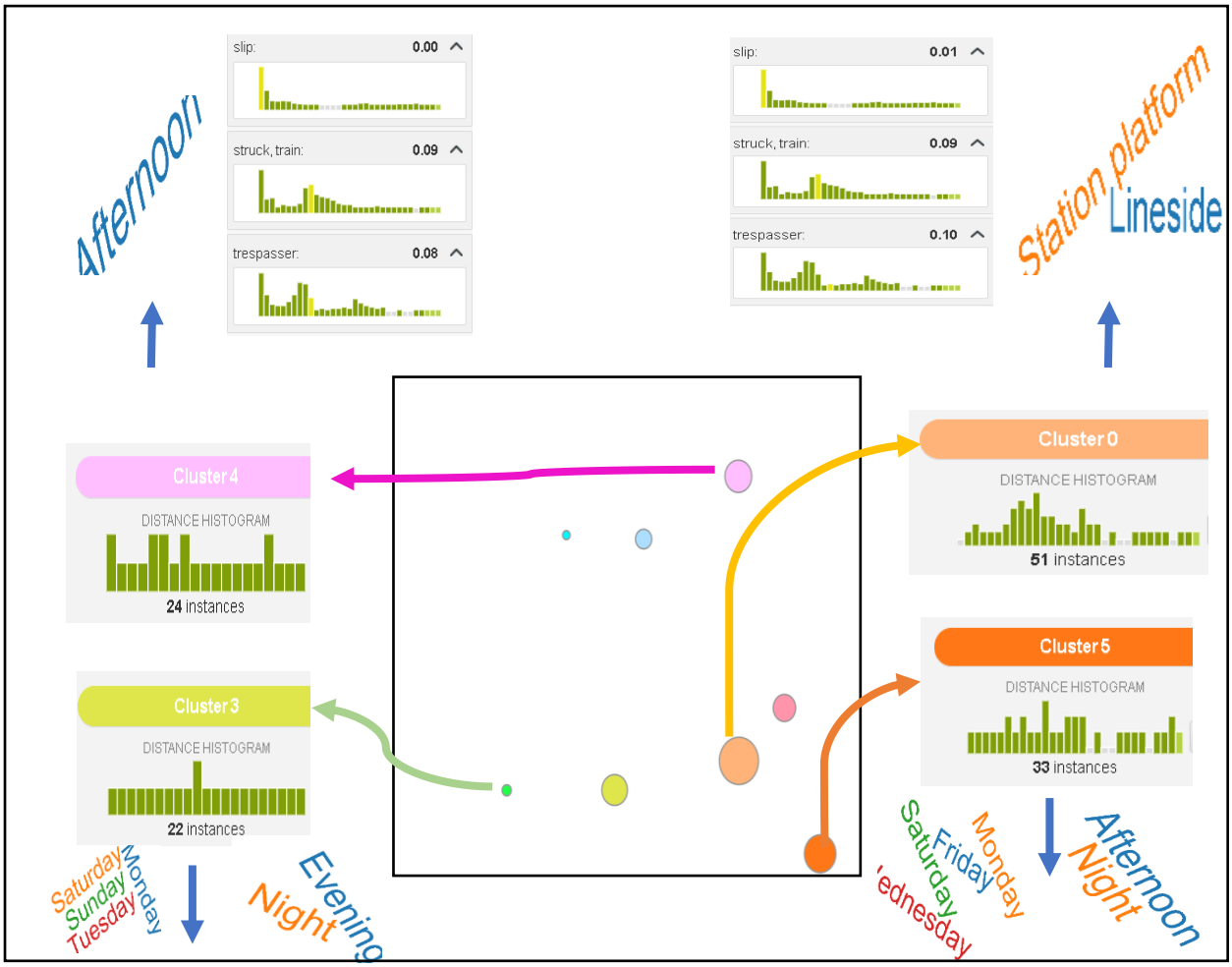


Figure 106. 8 cluster maps with examples of word maps and histograms.

8 clusters appear as the most common correlations in the topic (the clusters with the maximum total of cases). The second cluster reveals the elderly passengers were involved in some accidents at the time when the train is moving. The next cluster shows the electricity risk more specific to the conductor's parts in the platforms. These accidents required details to obtain more safety measures. Analyses can provide causes roots, correlations and any hidden patterns, for example, electric shocks, including contact with Overhead Line Equipment (OHLE) or Overhead Contact Line (OCL), which may occur accidentally by carrying long objects (selfie stick/conductive materials) or vandalism and trespassing. Such details will improve the standards and obligation to add more protections for the public, passengers and workforce at the PTI. All probability of the instances fields is shown in Table 29 below, which can be used as workflow in future projects and generalised. A guide system of the context can be converted to be numerical and cover a huge range of the stations.

Table 29.Cluster outcomes of the textual data.

Centroid_name	Age band	Physical environment	Day of week	Time of Day
Global	31-50	['; Platform', 'Platform']	['Saturday']	['Morning']
Cluster 0	31-50	['Lineside', 'Station platform']	['Friday', 'Thursday', 'Wednesday']	['Evening', 'Morning', 'Night']
Cluster 1	31-50	['Lineside']	['Monday', 'Saturday', 'Tuesday']	['Evening']
Cluster 2	31-50	['Road', 'Road vehicle carriageway', 'carriageway', 'vehicle']	['Tuesday']	['Evening']
Cluster 3	21-30	['Lineside', 'Other', 'Other; Platform']	['Monday', 'Saturday', 'Sunday', 'Tuesday']	['Evening', 'Night']
Cluster 4	31-50	['; Platform', 'Lineside', 'Platform', 'Platform train interface', 'Station', 'Station platform', 'train']	Tuesday']	['Afternoon']
Cluster 5	31-50	['; Platform', 'Other', 'Other; Platform', 'Platform', 'Platform train interface', 'Station', 'Station platform', 'train']	['Friday', 'Monday', 'Saturday', 'Wednesday']	['Afternoon', 'Night']
Cluster 6	31-50	['Concourse', 'Footbridge', 'Platform train interface', 'Stairs', 'Stairs/pedestrian bridge/escalator', 'bridge', 'escalator', 'pedestrian', 'train']	['Saturday', 'Sunday', 'Tuesday']	['Evening', 'Morning']
Cluster 7	Over 70 years	['Concourse', 'Stairs', 'Stairs/pedestrian bridge/escalator', 'Station', 'bridge', 'escalator', 'pedestrian']	['Friday', 'Thursday']	['Afternoon']

For testing, topic evaluation by a short text, which is a text describing an accident, has been used to extract the information, which shows the ability to capture information from the textual content. This depends on the tags (Labelled Itemset) being used for the training model, Labels like the day, the time, event description and location (see Table 30).

Table 30. Text example for extracting information from trained topic model.

Text for test and tags	The result in line
location The day The Time consequences The Age Event description consequances-2 Sunday Night; the passenger has been struck / then crushed by train; struck while crossing tracks at a station the age of the victim were estimated in the range 16-20 Lineside. The accident occurred on the 19th of Sep at the Lineside of the platform, and the first cause was that a passenger is struck and then crushed by a train; the pieces of evidence from According to eyewitnesses on the platform says it was happening while crossing tracks at a station.	Sunday THE DAY Night THE TIME ; the passenger has been struck / then crushed by train; struck while crossing tracks EVENT DESCRIPTION at a station the age of the victim were estimated in the range 16-20 THE AGE Lineside LOCATION . The accident occurred on 19th of Sep at the Lineside of the platform, and the first cause was that a passenger is struck and then crushed by a train; the pieces of evidence from According to eyewitnesses on the platform says it was happening while crossing tracks EVENT DESCRIPTION at a station.

For the overall trained model, statistics measures present excellent outcomes for all the tags, as shown in Table 31 below. The analysis presents the power of the approach to be an expert system and guide the reporting process, also noting any hidden root causes of the accidents.

Table 31. Classifier of overall statistics and confusion matrix.

TN (0)	FP (78)	Precision	Recall	Accuracy
FN (111)	TP 827 (predicts the right tag)	827/ (827+78) = 0.91	827/ (827+111) = 0.88	827/ (827+(78+111) = 0.8

Some data are accurately detected such as the day, location, ages and time of the occurrence of the accident. Others, such as details of the accidents, show less accuracy to detect, but these types of data

can be captured by providing more training data, that will in turn provide more balance for all tags and reflect on accuracy.

8.6 Discussion

Applying such a method shows the ability and the power of the new technology in the safety of the railway industry which has not been used widely to enhance safety and risk management. The systematic reducing of accidents in the station is beneficial for the public and stockholder, sustainability, and the safety community. Topic moulding is a proactive approach, where the input can be analysed in real-time, and actions can be taken. Moreover, multi-input can be used in parallel with supporting decision-makers. This approach reduces the dependence on experts, where they are costly and not available all the time and reduce the impact of manpower knowledge retirement. The development is key for many aspects which reflect on the quality, reliability, and satisfaction for both workers and passengers. The increase of technology such as AI and IoT with the growth of data requires more investment and research of such methods. Even with the limitation of data and delayed application in the field, text from data is analysed via text modelling which opens the novel approach of applying technology in the safety system. Such types of data are essential as they contain the history of safety such as risk evaluation, accident reports and periodic safety analyse documents, and can also use live text from media or calls. Moreover, social media has a textual source of information that can be harvested and analysed as it is related to safety, security, and quality.

The method provides support for the safety authority of the decision-maker from many sides, including improving the service, quick response rates and advanced analysis. Also, the system can build an expert system in a specific area such as the railway stations and learning from the new documents, and can cover an entire country's stations as well as pre-training the system from many sources. In the digitalisation concept, the integration between the source of data is possible, which forms a smart safety system in the railway industry. The LDA provides a statistical model with the

ability to learn and this method does not only deal with the huge data in real-time automatically. It can also provide an expert system and decision support for the safety authority for the researcher. This concept corresponds with the future rail digitalization and the BD revolution. The fixability, effectiveness, and accuracy raise the importance of gathering more data in the station with all the possible details. The accessibility, privacy, skills and the IT infrastructure are some hurdles for the short term, but which it is expected the industry will overcome in the medium and long term.

8.7 Chapter summary

Obtaining the data is a dilemma for many reasons, including the availability, form of data and the volume which often is not suitable for applying AI. This is in addition to rules such as GDPR which are not widely understood. Thus, the safety community must present the challenges and prove that they can demonstrate compliance. The compliance includes showing that all processing takes place in accordance with the data protection principles, and there is a lawful basis for the processing, and it should process only for the specific purpose that is identified. Moreover, security, anonymising personal data, also sharing with any third parties must be documented as we did with RSSB. The massive textual data was an issue due to space for storage, manage and use, without practical benefits from the old data. However, with machine learning (topic models), this vast information is necessary and enabled the examination of risks down to the last detail, managing them in a smart, efficient way and making it easy to use. Topic models have an important role in several disciplines, like marketing, safety, and risk management in railway stations for text mining. In topic modelling, a topic is a list of words that occurs in statistically significant methods. This approach demonstrates the power of both unsupervised ML for unstructured data and supervisor ML for tagged data topic modelling in promoting risk management, safety, accident investigation, and restructuring of accident recording and documentation at the industry-based level. The wide range of available data (for example, from all the stations in the UK or Europe) enables coverage of a wide range of protection issues in stations.

The description of the root causes of an accident by the suggested model has shown that the platform (lineside) is the hotspot within stations concerning accidents. The outcomes reveal that station accidents occur because of some main causes like falls, trespass, and electric shock, and that being struck by a train leads to fatalities. Moreover, young and elderly passengers are more exposed to risks. Managing staffing in the stations is also essential in terms of times and locations. Controls to ensure that passengers are monitored at critical times will be a significant and essential measure to reduce fatalities, and also analysing the standards such as the electric equipment safety standards. In addition, more assistance on the station platform for some age groups is crucial, as well as accessibility through bridges and escalators. Therefore, passenger data in real-time is an important point for managing people in stations. Applying such technology will enhance the integration of many systems for gathering data and predicting risks promptly like maintenance, ticketing system and door to door transportation systems. The results are helpful for safety, emergency, and risk management. Finally, the introduced methodology of AI (topic models) can aid the railway industry in risk management and future safety by employing the recorded largescale safety data.

CHAPTER 9: CONCLUSIONS AND RECOMMENDATIONS

9.1 How AI preventing accidents

Stations are a significant point of the railway and urban areas and have direct and indirect effects economically and socially. They face challenges because of growth, which requires the development of high safety levels in such critical infrastructure. Accidents in the railway stations have a series of consequences, including political, economic, social and business reputation. Tackling risk in open, crowded spaces such as stations in routine or emergency operations is not an easy task and has uncertainties; thus, investing in technology is necessary to assist in such cases. However, technology such as ML requires Big Data, and the industry is not mature yet to employ such a concept, for instance, the data necessary understanding and storage with skills to convert it into the required format as per the requirements from the technology. The industry has BD yet has not been appropriated for analysis via AI, leading to concerns in extraction and use. The study conducted in this work proves the uniqueness of the models developed and the ability to utilise many data sources to enhance the risk management in the stations. Moreover, the importance of restructuring the data in the industry should be considered in applying AI in the near future. Preventing accidents from happening in the railway industry, even in other fields, depends on advanced preventives, providing BD containing indicators that can be predictors and warning for the big risks and major consequences. The risk has two sides, namely occurrence and severity. Early prediction of occurrence leads to preventing the risk or then at least reducing the consequences, which reflects low risk overall. For example, object detection via ML may show there is a person falling in the gap, and the decision from the control room could prevent them from being run over by a train if the action is taken at the right time, and even when it has not been realised or seen by other staff on the platform. Thus, ML can be an active

safety measure in such cases. Real-time data from online such as CCTV can be used, ticket systems and sensors, providing advanced visual analyses, which is reflected in passenger services and assists safety or risk management operations, either in routine operations or in the event of the occurrence of an unexpected event. The method is more practical and cost-effective in real-world circumstances (affording, for instance, CCTV cameras are automated observed 24 hours a day, seven days a week to identify potential threats). In fact, the Ml has a variety of techniques that can cover a wide range of data and even unstructured data. The abundant pictures that can be analysed via ML, even if they are not fully considered via the field, are expected revolutionize risk management as a proactive system of current safety and security vulnerabilities. Moreover, the machine in the stations can predict the time of the failure, for example, failure in power, escalators, lights or elevators, doors, gates and so on, which can be designed to predict and explore patterns associated with machine failures. In fact, AI is an opportunity for the analysis and timely response to ensure safety and quality in railway stations. For example, a passenger mentioning via social media overcrowding in their account on the platform needs to check the traffic levels in the location in a timely manner to prevent stampedes and passengers accumulating or suffocation. Moreover, ML applications as innovative concepts in risk management can provide a powerful system for collecting data from many stations, which may reach a thousand in a country, and prevent and duplicate risk occurrences, which fit an array of essential concepts of safety (lessons must be learned). Also, the implementation of technology can overcome many challenges such as the subjective assessment from experts, and response time, sharing the knowledge, digitalisation system, smart stations, emergency response, simultaneous risk occurrence, and timely decision support.

9.2 AI and covid-19

COVID-19 has resulted in rapid changes that have been challenging in many areas, and the rail industry is not an exception. Businesses are rethinking their expenses and recognising the significance

of AI to drive revenue and efficiency during uncertain times. However, currently, they are worrying more about how to get AI methods into production faster and ensure their performance over time. COVID-19 accelerated AI and ML initiatives and have a priority list in many organizations 2021; this has been reported from many trends such as:

- Year-on-year, the preference and funds for AI are growing significantly.
- The automation and customer experience are activities focused on by organisations.
- COVID-19 has prompted a reconsideration of how to utilize technology, not only for the railway sector but also for all industries[186].

49% of organizations rank AI as a concern, while the organizations that utilise AI save money and spend a shorter time than those that produce their private solutions[298]. Moreover, since COVID-19 has spread, it has been affecting protection, security, and health concepts and strategies, so prediction has become more important than it has ever been.

9.3 ANFIS controller

The ANFIS approach has a range of key characteristics that distinguish it from other risk management systems and methods, such as the ones that follow:

- Station monitoring to detect congestion before it becomes a problem.
- Flexibility in response to the risk of congestion.
- In response to the risk of overcrowding, a comparative learning technique of station resources was developed.
- With the potential of combining devices and intelligent stations, it can be a more intelligent model.
- Real-time data from online devices such as CCTV can be used, ticket system and sensors.

- This is on a limited scale but can be extended to more than just levels of congestion risk to cover other risks such as air contamination.
- Using an ANFIS controller to manage risk such as overcrowding has implications on several levels in the station. To assess the state, the incoming data is compared in real-time to the threshold level, allowing the decision-maker to react to issues sooner. The train positions and the traffic can be set up to transport passengers in the event of an unusually large number of passengers based on the platform's live situation. Passenger routes or marketing plans at the station may be updated based on the size and number of passengers, as well as station redesign or maintenance planning. Finally, the proposed novel model has the ability to generate superior safety performance estimations.

9.4 Topic modelling

A such as safety and risk management in railway stations for text mining, topic models play an important role. A subject is a list of terms that appear in statistically relevant methods of topic modelling. A script can be a voice recording of an investigation report or risk evaluations and safety measures. This method demonstrates the utility of unsupervised machine learning subject modelling for risk control, safety accident investigation, accident recording and reporting at the industry level. Since it is analysing, exploring hidden patterns, and dealing with several problems, therefore it is useful for protection via a strong base of historical evidence.

The data can be a checklist of risk documents or other safety records and possible review and analysed from other departments such as maintenance. Moreover, it can be captured systematically from textual social media any related safety or risk concerns. All that resources are essential for safety and will be valuable for new employees and future improvements for safety in the railway station and deal with many challenges such as:

- Text data from a diversity of viewpoints and in a difference of unstructured formats.
- Ability to spot safety and danger kyes in data, as well as deal with missing values.
- Intelligent labelling, clustering, centroids, sampling, and coordinates
- Record relationships, causes, and more to rate risks and related data.
- Risk's prioritisation and evaluated safety measures**.**

It is necessary to extend research exploring BD topics, including the diversity of station locations, scale, safety cultures, standards and other variables. Finally, although this application improves protection, it also emphasises the value of data in text form. It recommended that to rethinking the way, data is collected to be more robust and detailed, which means redesigning the reporting system to be smarter using such approach and documentation and risk management textual data in the railway.

9.5 Conclusion

It has been proven that AI has many strong competitive advantages in many fields; therefore, it is suggested that AI as a novel unique concept be used as the main part of the methodology to solve and develop the risk management in the railway stations. Moreover, as the railway stations are multi-use facilities, the development and modernisation in operations safety will be beneficial for the economy and society, and will reflect on the other public transport, and all types of rails that might be connecting the stations like high-speed rail (HSR). In this research, these advantages have been explored and the values it adds through its implementation are listed here:

- Strong ability for dealing with the huge history or current data from various sources. This includes the safety reports, including the investigations or regular accidents reports (structured or not). The data can be gathered as signals, human behaviours or times, or from spatial geographic location information and any other form of data such as social media and so on. It can then be integrated as indicators for predictions. The AI provides a quick and easy way to

extract the data and clustering, extract patterns and classification accuracy. AI is an ideal tool that can meet one of the significant safety concepts (lessons learned). Moreover, AI is not only for prediction, but it can also be a smart descriptive tool such as clustering and association rules analysis, which are key for capturing data structures and relationships.

- Acts better than human and data analysis in real-time, which creates automated protection from risks. This satisfies the authorities and clients and provides high reliability to the system, which reflects on the quality and business cost.
- The implementation of AI leads to improvements in many layers associated with the concept, such as smart city, smart transportation, and intelligent cloud. This development will result in organisations via advanced skills in employees and smart IT infrastructure.
- The coverage nationally and internationally, the information and knowledge can be referenced in time and act quickly for safety and security where the system learns and provides the information from different events and locations.
- The significant feature of AI is continuous learning (Self-learning) which is key in any safety system and provides ongoing improvements.
- The improvements in safety and risk management utilising AI reflect on the overall operations in the stations via increasing the level of efficiency in decision-making and avoiding the consequence of unwanted events such as that which may cause trains delays or disrupt the network.
- The knowledge from AI (ML-DL) can be beneficial for the young employees and support people who do not have sufficient information about risks analysis and safety systems management by grasping information and easily starting analyses with confidence and fewer errors.

- Applying AI provides two hallmarks for risk management: first, the ability of learning; and second, the transfer of the way of risk management from reactive to proactive.
- Applying AI will reflect on information resources management (internal and external) and lead to integrating various aspects of the railway station relevant systems for highly efficient data sharing and interconnection for a smart environment and railway intelligent transportation system.
- The AI allows and leads to developing the data collection and sharing them (exchange processes) for local operators in the same country or in the global network (global stations risk management), that will build strong assess of the risks in stations, which can be generalised for the whole railway system.
- Image processing is confirmed to be a promising technology for improving station safety, risk management, and improving operations; for instance, identifying risks early will be reducing dwell times and disruption and reducing the number of operators at stations.
- The proposed CNN model will automatically determine the risky behaviours (such as falling off the platform) with a high level of accuracy.
- The findings can be used to provide reliable, fast outcomes to planners, managers, and decision-makers, enabling them to discover how and where to carry out procedures in a safer manner and prove any new safety precautions have been examined or any redesign of operations or layout inside the station.
- Supervised learning provides an advanced analytics tool for the history of the accidents and risk registry with labelled training datasets to explore the attributes and hidden associations and predict from data not seen before. In contrast, unsupervised learning is effective for unstructured data such as the text, so the ML is powerful and can fit the various forms of data to enhance risk management from different resources.

- It is high in necessary to improve and clarify the regulations and rules to be cover utilising Big Data including data through CCTV for the purpose of safety, risk management and research; for instance, In GDPR, personal data has a comprehensive definition and make it as an excuse not to provide data even that not hold any names or ID and no revealing racial or ethnic origin and so on, clearly this is open up a new window from the specialists' for more research.

From such innovative application of AI in risk management, the challenge of sharing information, privacy, and the skills needed is realised. Moreover, regarding the other technology applications, security must be considered, as well as the data type, quality and gathering process in a way that will be consistent and reliable. There is no doubt that applying AI will raise the needs of data and will lead to a rise in the gathering data process to increase the points coverage for more data from various viewpoints related to safety and risk management in the stations. This covers a wide range of data, including images, signals from machines and external data such as weather conditions. The gathered combinations of data enhance safety and risk management via advanced analytics and grasping the associated factors and any hidden patterns.

Furthermore, the applications can be expanded to numerous safety and risk management systems in train stations, such as maintenance, fire, smoke control and more. Other aspects of the models, such as emergency planning, people counts, train positioning, and security, are worth looking into further. Data availability and quality, however, remain a concern because this technology relies largely on large volumes of high-quality data. Although the technique has not matured in the rail, the coming years of risk management will be a meaningful atmosphere for AI applications with the accessibility of Big Data (BD), saving more lives, and improving safety in the stations and in the transport sector overall.

9.6 Implementation and future work

Regarding implementation and applying a strategic approach for AI in railway system safety (risk management) (stations), many aspects must be considered, not all of which are available but which are dependent on each other:

- The technology, including the cloud, storage, algorithms and so on (available)
- The BD (Value, veracity, variety, velocity, volume) (not available)
- The collaboration between research and industry (needs more)
- The strategic plans, IT infrastructure and the fund (needs more)
- The standards and regulations (need more)
- Ongoing research and applications (need more)

For the future of safety in railways for real-time risk management, it is essential to have new technology skills and infrastructure as well as an entire process for transferring to AI solutions where computer and data science will be combined with safety and risk sciences.

The stations provide a wealth of research opportunities related to safety and risks yet so far, all the risks have not been covered, and this particular study has been limited in scope. Thus, in the future, data should be sought to cover other risks and to discover the potential for applying AI. The range of AI methods is varied and can be investigated to present more comparisons while applications can be used for risk management in the stations and comparisons. Also, the safe and secure integration with external systems in the direction of smart railway stations has been considered as an aspect that should be researched in the future. In the event of overcrowding, a variety of varied and extended training and testing data sets would be helpful for doing comprehensive research over a long time to cover all the complicated changes in crowds and all distinct danger scenarios in order to increase system efficiency and exploring many areas in the station such entrances, ticketing areas.

References

[1] European Railway Agency, *Report on Railway Safety and Interoperability in the EU*. 2018.

[2] T. Litman, "A new transit safety narrative," *J. Public Transp.*, vol. 17, no. 4, pp. 114–135, 2014, doi: 10.5038/2375-0901.17.4.7.

[3] RSSB, "Station safety-A guide to RSSB research in Station safety," *RSSB*. [Online]. Available: http://www.rssb.co.uk. [Accessed: 24-Jun-2020].

[4] C. Shi, M. Zhong, X. Nong, L. He, J. Shi, and G. Feng, "Modeling and safety strategy of passenger evacuation in a metro station in China," *Saf. Sci.*, vol. 50, no. 5, pp. 1319–1332, Jun. 2012, doi: 10.1016/j.ssci.2010.07.017.

[5] L. He, Q. Liang, and S. Fang, "Challenges and Innovative Solutions in Urban Rail Transit Network Operations and Management: China's Guangzhou Metro Experience," *Urban Rail Transit*, vol. 2, no. 1, pp. 33–45, Mar. 2016, doi: 10.1007/s40864-016-0036-y.

[6] NetworkRail, "Number of people risking their lives on the railway hits 10-year-high." [Online]. Available: https://www.networkrailmediacentre.co.uk/news/number-of-people-risking-their-lives-on-the-railway-hits-10-year-high. [Accessed: 24-Jun-2020].

[7] H. Yu, Y. Wang, P. Qiu, and J. Chen, "Analysis of natural and man-made accidents happened in subway stations and trains: based on statistics of accident cases," *MATEC Web Conf.*, vol. 272, p. 01031, 2019, doi: 10.1051/matecconf/201927201031.

[8] S. Zemp, M. Stauffacher, D. J. Lang, and R. W. Scholz, "Classifying railway stations for strategic transport and land use planning: Context matters!," *J. Transp. Geogr.*, vol. 19, no. 4, pp. 670–679, 2011, doi: 10.1016/j.jtrangeo.2010.08.008.

[9] Å. Boholm, "On the organizational practice of expert-based risk management: A case of railway

planning," *Risk Manag.*, vol. 12, no. 4, pp. 235–255, Oct. 2010, doi: 10.1057/rm.2010.4.

[10] H. Alawad, S. Codru, and S. Kaewunruen, "Complex-system decision framework for managing risks to rail stations at airports from terrorist threats," in *Road and Rail Infrastructure V*, 2018, vol. 5, pp. 855–861, doi: 10.5592/co/cetra.2018.665.

[11] S. Kaewunruen, H. Alawad, and S. Cotruta, "A decision framework for managing the risk of terrorist threats at rail stations interconnected with airports," *Safety*, vol. 4, no. 3, 2018, doi: 10.3390/safety4030036.

[12] C. Martani, N. Papathanasioud, and B. T. Adey, "A Review of the State-of-the-Art in Railway Risk Management," *Int. J. Railw.*, vol. 10, no. 1, pp. 5–11, 2017, doi: 10.7782/ijr.2017.10.1.005.

[13] T. Cao, W. Mu, J. Gou, and L. Peng, "A Study of Risk Relevance Reasoning Based on a Context Ontology of Railway Accidents," *Risk Anal.*, vol. 40, no. 8, pp. 1589–1611, Aug. 2020, doi: 10.1111/risa.13506.

[14] H. Alawad and S. Kaewunruen, "Wireless sensor networks: Toward smarter railway stations," *Infrastructures*, vol. 3, no. 3. GB, 2018, doi: 10.3390/infrastructures3030024.

[15] H. J. Parkinson and G. Bamford, "The potential for using big data analytics to predict safety risks by analysing rail accidents," in *Civil-Comp Proceedings*, 2016, vol. 110.

[16] N. Bešinović, "Resilience in railway transport systems: a literature review and research agenda," *Transp. Rev.*, vol. 40, no. 4, pp. 457–478, 2020, doi: 10.1080/01441647.2020.1728419.

[17] Z. Chen and Y. Wang, "Impacts of severe weather events on high-speed rail and aviation delays," *Transp. Res. Part D Transp. Environ.*, vol. 69, pp. 168–183, Apr. 2019, doi: 10.1016/j.trd.2019.01.030.

[18] A. A. Zilko, D. Kurowicka, and R. M. P. Goverde, "Modeling railway disruption lengths with Copula Bayesian Networks," *Transp. Res. Part C Emerg. Technol.*, vol. 68, pp. 350–368, Jul. 2016, doi:

10.1016/j.trc.2016.04.018.

[19] T. Xia, X. Zhu, L. Wang, Y. Ye, and W. Liu, “Railway Passenger and Freight Transportation Operation Risk Identification and Evaluation Technology,” in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 768, no. 5, p. 052016, doi: 10.1088/1757-899X/768/5/052016.

[20] H. Alawad, S. Kaewunruen, and M. An, “A Deep Learning Approach towards Railway Safety Risk Assessment,” *IEEE Access*, vol. 8. Institute of Electrical and Electronics Engineers Inc., pp. 102811–102832, 2020, doi: 10.1109/ACCESS.2020.2997946.

[21] H. Alawad, S. Kaewunruen, and M. An, “Learning From Accidents: Machine Learning for Safety at Railway Stations,” *IEEE Access*, 2020, doi: 10.1109/ACCESS.2019.2962072.

[22] H. Alawad, S. Kaewunruen, and M. An, “Learning from Accidents: Machine Learning for Safety at Railway Stations,” *IEEE Access*, vol. 8, pp. 633–648, 2020, doi: 10.1109/ACCESS.2019.2962072.

[23] M. Yin and X. Cheng, “A review on artificial intelligence in high-speed rail,” *Transp. Saf. Environ.*, vol. 0, no. 0, pp. 1–13, 2020, doi: 10.1093/tse/tdaa022.

[24] T. Peng, C. Li, and X. Zhou, “Application of machine learning to laboratory safety management assessment,” *Saf. Sci.*, vol. 120, pp. 263–267, 2019, doi: 10.1016/j.ssci.2019.07.007.

[25] S. Kaeeni, M. Khalilian, and J. Mohammadzadeh, “Derailment accident risk assessment based on ensemble classification method,” *Saf. Sci.*, vol. 110, pp. 3–10, Dec. 2018, doi: 10.1016/j.ssci.2017.11.006.

[26] H. Shi, M. J. Kim, S. C. Lee, S. H. Pyo, I. J. Esfahani, and C. K. Yoo, “Localized indoor air quality monitoring for indoor pollutants’ healthy risk assessment using sub-principal component analysis driven model and engineering big data,” *Korean J. Chem. Eng.*, vol. 32, no. 10, pp. 1960–1969, 2015, doi: 10.1007/s11814-015-0042-x.

[27] X. Ding, X. Yang, H. Hu, and Z. Liu, “The safety management of urban rail transit based on operation fault log,” *Saf. Sci.*, vol. 94, pp. 10–16, Apr. 2017, doi: 10.1016/j.ssci.2016.12.015.

[28] F. Feng, W. Li, and Q. Jiang, “Railway traffic accident forecast based on an optimized deep auto-encoder,” *Promet - Traffic - Traffico*, vol. 30, no. 4, pp. 379–394, Aug. 2018, doi: 10.7307/ptt.v30i4.2568.

[29] D. Chen, C. Xu, and S. Ni, “Data mining on Chinese train accidents to derive associated rules,” *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 231, no. 2, pp. 239–252, 2015, doi: 10.1177/0954409715624724.

[30] A. Mirabadi and S. Sharifian, “Application of association rules in Iranian Railways (RAI) accident data analysis,” *Saf. Sci.*, vol. 48, no. 10, pp. 1427–1435, Dec. 2010, doi: 10.1016/j.ssci.2010.06.006.

[31] C. van Gulijk, P. Hughes, M. Figueres-Esteban, R. El-Rashidy, and G. Bearfield, “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.*, vol. 232, no. 2, pp. 151–163, Apr. 2018, doi: 10.1177/1748006X17728210.

[32] M. Figueres-Esteban, P. Hughes, and C. van Gulijk, “Using visual analytics to make sense of railway Close Calls,” *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 231, no. 10, pp. 1107–1114, Nov. 2017, doi: 10.1177/0954409716676221.

[33] M. Figueres-Esteban, P. Hughes, and C. van Gulijk, “Visual analytics for text-based railway incident reports,” *Saf. Sci.*, vol. 89, pp. 72–76, Nov. 2016, doi: 10.1016/j.ssci.2016.05.009.

[34] R. A. H. E. L. Rashidy, P. Hughes, M. Figueres-Esteban, C. Harrison, and C. Van Gulijk, “A big data modeling approach with graph databases for SPAD risk,” *Saf. Sci.*, vol. 110, pp. 75–79, Dec. 2018, doi: 10.1016/j.ssci.2017.11.019.

[35] D. E. Brown, “Text Mining the Contributors to Rail Accidents,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 346–355, Feb. 2016, doi: 10.1109/TITS.2015.2472580.

[36] L. Hua, W. Zheng, and S. Gao, “Extraction and Analysis of Risk Factors from Chinese Railway Accident Reports,” in *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*, 2019, pp. 869–874, doi: 10.1109/ITSC.2019.8917094.

[37] A. Jamshidi *et al.*, “A Big Data Analysis Approach for Rail Failure Risk Assessment,” *Risk Anal.*, vol. 37, no. 8, pp. 1495–1507, Aug. 2017, doi: 10.1111/risa.12836.

[38] H. Salmane, L. Khoudour, and Y. Ruichek, “A Video-Analysis-Based Railway-Road Safety System for Detecting Hazard Situations at Level Crossings,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 596–609, Apr. 2015, doi: 10.1109/TITS.2014.2331347.

[39] F. Zhou, Y. Song, L. Liu, and D. Zheng, “Automated visual inspection of target parts for train safety based on deep learning,” in *IET Intelligent Transport Systems*, 2018, vol. 12, no. 6, pp. 550–555, doi: 10.1049/iet-its.2016.0338.

[40] A. Zaman, B. Ren, and X. Liu, “Artificial Intelligence-Aided Automated Detection of Railroad Trespassing,” *Transp. Res. Rec.*, vol. 2673, no. 7, pp. 25–37, Jul. 2019, doi: 10.1177/0361198119846468.

[41] K. N. Syeda, S. N. Shirazi, S. Asad, A. Naqvi, and H. J. Parkinson, “Exploiting Natural Language Processing for Analysing Railway Incident Reports,” pp. 1–18.

[42] I. Doboszewski, S. Fossier, and C. Marsala, “Data Driven Detection of Railway Point Machines Failures,” in *2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019*, 2019, pp. 1233–1240, doi: 10.1109/SSCI44817.2019.9002844.

[43] G. N. Kouziokas, “The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment,” in *Transportation Research Procedia*, 2017, vol. 24, pp. 467–473, doi: 10.1016/j.trpro.2017.05.083.

[44] H. Hadj-mabrouk, “Transportation safety assessment methodology based on artificial intelligence

techniques," *Int. J. Comput. Sci. Commun.*, vol. 1, no. 1, pp. 22–30, 2016.

[45] H. Hadj-Mabrouk, "Contribution of artificial intelligence and machine learning to the assessment of the safety of critical software used in railway transport," *AIMS Electron. Electr. Eng.*, vol. 3, no. 1, pp. 33–70, 2019, doi: 10.3934/electreng.2019.1.33.

[46] E. Bikov, P. Boyko, E. Sokolov, and D. Yarotsky, "Railway incident ranking with machine learning," in *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, 2017, vol. 2017-Decem, pp. 601–606, doi: 10.1109/ICMLA.2017.00-95.

[47] Y. Han, Z. Liu, Y. Lyu, K. Liu, C. Li, and W. Zhang, "Deep learning-based visual ensemble method for high-speed railway catenary clevis fracture detection," *Neurocomputing*, vol. 396, pp. 556–568, Jul. 2020, doi: 10.1016/j.neucom.2018.10.107.

[48] J. Liu *et al.*, "Learning Visual Similarity for Inspecting Defective Railway Fasteners," *IEEE Sens. J.*, vol. 19, no. 16, pp. 6844–6857, Aug. 2019, doi: 10.1109/JSEN.2019.2911015.

[49] C. Liang, M. Ghazel, O. Cazier, and L. Bouillaut, "Advanced model-based risk reasoning on automatic railway level crossings," *Saf. Sci.*, vol. 124, Apr. 2020, doi: 10.1016/j.ssci.2019.104592.

[50] O. Fink, E. Zio, and U. Weidmann, "Predicting component reliability and level of degradation with complex-valued neural networks," *Reliab. Eng. Syst. Saf.*, vol. 121, pp. 198–206, 2014, doi: 10.1016/j.ress.2013.08.004.

[51] J. Hegde and B. Rokseth, "Applications of machine learning methods for engineering risk assessment – A review," *Safety Science*, vol. 122. 2020, doi: 10.1016/j.ssci.2019.09.015.

[52] T. Aven, "Risk assessment and risk management: Review of recent advances on their foundation," *European Journal of Operational Research*, vol. 253, no. 1. Elsevier B.V., pp. 1–13, 16-Aug-2016, doi: 10.1016/j.ejor.2015.12.023.

[53] M. Rausand, *Risk assessment: theory, methods, and applications*. John Wiley & Sons, Inc., 2013.

[54] F. Fang, H. Wang, K. Chen, and F. Khan, "Risk analysis of Chongqing urban rail transit network," *J. Loss Prev. Process Ind.*, vol. 66, pp. 950–4230, Jul. 2020, doi: 10.1016/j.jlp.2020.104182.

[55] R. V. Sánchez *et al.*, "Evaluation of time and frequency condition indicators from vibration signals for crack detection in railway axles," *Appl. Sci.*, vol. 10, no. 12, p. 4367, Jun. 2020, doi: 10.3390/app10124367.

[56] J. Lu *et al.*, "SCueU-Net: Efficient Damage Detection Method for Railway Rail," *IEEE Access*, vol. 8, pp. 125109–125120, 2020, doi: 10.1109/ACCESS.2020.3007603.

[57] C. J. Lin, Y. C. Li, and C. L. Lee, "Using fully convolutional networks for floor area detection," *Sensors Mater.*, vol. 32, no. 1, pp. 159–170, 2020, doi: 10.18494/SAM.2020.2577.

[58] M. Etxeberria-Garcia, M. Labayen, M. Zamalloa, and N. Arana-Arexolaleiba, "Application of Computer Vision and Deep Learning in the railway domain for autonomous train stop operation," in *Proceedings of the 2020 IEEE/SICE International Symposium on System Integration, SII 2020*, 2020, pp. 943–948, doi: 10.1109/SII46433.2020.9026246.

[59] A. K. Singh, A. Swarup, A. Agarwal, and D. Singh, "Vision based rail track extraction and monitoring through drone imagery," *ICT Express*, vol. 5, no. 4, pp. 250–255, Dec. 2019, doi: 10.1016/j.icte.2017.11.010.

[60] T. De Bruin, K. Verbert, and R. Babuska, "Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 3, pp. 523–533, Mar. 2017, doi: 10.1109/TNNLS.2016.2551940.

[61] A. Jamshidi *et al.*, "A decision support approach for condition-based maintenance of rails based on big data analysis," *Transp. Res. Part C Emerg. Technol.*, vol. 95, pp. 185–206, Oct. 2018, doi: 10.1016/j.trc.2018.07.007.

[62] W. Y. Chen, M. Wang, and Z. X. Fu, “Railway crossing risk area detection using linear regression and terrain drop compensation techniques,” *Sensors (Switzerland)*, vol. 14, no. 6, pp. 10578–10597, Jun. 2014, doi: 10.3390/s140610578.

[63] H. Mukojima *et al.*, “Moving camera background-subtraction for obstacle detection on railway tracks,” in *Proceedings - International Conference on Image Processing, ICIP*, 2016, vol. 2016-Augus, pp. 3967–3971, doi: 10.1109/ICIP.2016.7533104.

[64] S. C. Oh and H. Lee, “Performance analysis of vision based monitoring system for passenger’s safety on railway platform,” in *ICCAS 2010 - International Conference on Control, Automation and Systems*, 2010, pp. 1867–1870.

[65] S. C. Oh, G. D. Kim, W. T. Jeong, and Y. T. Park, “Vision-based object detection for passenger’s safety in railway platform,” in *2008 International Conference on Control, Automation and Systems, ICCAS 2008*, 2008, pp. 2134–2137, doi: 10.1109/ICCAS.2008.4694449.

[66] J. Liu, L. Qian, Y. Zhang, J. Han, and J. Sun, “Towards Safety-Risk Prediction of CBTC Systems with Deep Learning and Formal Methods,” *IEEE Access*, vol. 8, pp. 16618–16626, 2020, doi: 10.1109/ACCESS.2020.2967634.

[67] Y. Wang, S. Hu, G. Wang, C. Chen, and Z. Pan, “Multi-scale dilated convolution of convolutional neural network for crowd counting,” *Multimed. Tools Appl.*, vol. 79, no. 1–2, pp. 1057–1073, 2020, doi: 10.1007/s11042-019-08208-6.

[68] S. A. Sumon, M. T. Shahria, M. R. Goni, N. Hasan, A. M. Almarufuzzaman, and R. M. Rahman, “Violent Crowd Flow Detection Using Deep Learning,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11431 LNAI, pp. 613–625, doi: 10.1007/978-3-030-14799-0_53.

[69] S. Akcay and T. Breckon, “Towards Automatic Threat Detection: A Survey of Advances of Deep

Learning within X-ray Security Imaging," 2020.

[70] T. Saheb, "Big data analytics in the context of internet of things and the emergence of real-time systems: A systematic literature review," *International Journal of High Performance Systems Architecture*, vol. 8, no. 1–2. pp. 34–50, 2018, doi: 10.1504/IJHPSA.2018.094143.

[71] P. McMahon, T. Zhang, and R. Dwight, "Requirements for Big Data adoption for Railway asset Management," *IEEE Access*, vol. 8, pp. 15543–15564, 2020, doi: 10.1109/aCCESS.2020.2967436.

[72] F. Ghofrani, Q. He, R. M. P. Goverde, and X. Liu, "Recent applications of big data analytics in railway transportation systems: A survey," *Transp. Res. Part C Emerg. Technol.*, vol. 90, pp. 226–246, May 2018, doi: 10.1016/j.trc.2018.03.010.

[73] X. Zhang, Y. Deng, Q. Li, M. Skitmore, and Z. Zhou, "An incident database for improving metro safety: The case of shanghai," *Saf. Sci.*, vol. 84, pp. 88–96, Apr. 2016, doi: 10.1016/j.ssci.2015.11.023.

[74] N. Attoh-Okine, "Big data challenges in railway engineering," in *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, 2015, pp. 7–9, doi: 10.1109/BigData.2014.7004424.

[75] H. Shu, "Big data analytics: six techniques," *Geo-Spatial Inf. Sci.*, vol. 19, no. 2, pp. 119–128, 2016, doi: 10.1080/10095020.2016.1182307.

[76] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine Learning with Big Data: Challenges and Approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017, doi: 10.1109/ACCESS.2017.2696365.

[77] M. Figueres-Esteban, P. Hughes, and C. Van Gulijk, "The role of data visualization in railway Big Data Risk Analysis," in *Safety and Reliability of Complex Engineered Systems - Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015*, 2015, pp. 2877–2882, doi: 10.1201/b19094-377.

[78] W. P. Lira *et al.*, “A visual-analytics system for railway safety management,” *IEEE Comput. Graph. Appl.*, vol. 34, no. 5, pp. 52–57, Sep. 2014, doi: 10.1109/MCG.2014.79.

[79] P. Hughes, R. Robinson, M. Figueres-Esteban, and C. van Gulijk, “Extracting safety information from multi-lingual accident reports using an ontology-based approach,” *Saf. Sci.*, vol. 118, pp. 288–297, Oct. 2019, doi: 10.1016/j.ssci.2019.05.029.

[80] K. N. Syeda, S. N. Shirazi, S. A. A. Naqvi, H. J. Parkinson, and G. Bamford, “Big Data and Natural Language Processing for Analysing Railway Safety,” in *Human Performance Technology*, 2019, pp. 781–809.

[81] F. Ghofrani, Q. He, R. M. P. Goverde, and X. Liu, “Recent applications of big data analytics in railway transportation systems: A survey,” *Transp. Res. Part C Emerg. Technol.*, vol. 90, pp. 226–246, May 2018, doi: 10.1016/j.trc.2018.03.010.

[82] British Standards Institute, “PAS 182:2014 Smart city concept model – Guide to establishing a model for data interoperability,” 2014.

[83] RebecoSAM AG, “Transforming infrastructure through smarter information Annual,” 2018.

[84] O. M. Araz, T. M. Choi, D. L. Olson, and F. S. Salman, “Role of Analytics for Operational Risk Management in the Era of Big Data,” in *Decision Sciences*, 2020, p. deci.12451, doi: 10.1111/deci.12451.

[85] Digital Railway Group, “Digital Railway Strategy,” London, 2018.

[86] H. J. Parkinson and G. J. Bamford, “Big data and the virtuous circle of railway digitization,” in *Advances in Intelligent Systems and Computing*, 2017, vol. 529, pp. 314–322, doi: 10.1007/978-3-319-47898-2_32.

[87] K. Xie, Y. Mei, P. Gui, and Y. Liu, “Early-warning analysis of crowd stampede in metro station

commercial area based on internet of things," *Multimed. Tools Appl.*, vol. 78, no. 21, pp. 30141–30157, Nov. 2019, doi: 10.1007/s11042-018-6982-5.

[88] "A1-D4 Ontology of Transportation Networks Project title: Reasoning on the Web with Rules and Semantics Project acronym: REWERSE Project number."

[89] C. Morris, J. Easton, and C. Roberts, "Applications of linked data in the rail domain," in *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, 2014, pp. 35–41, doi: 10.1109/BigData.2014.7004429.

[90] "InteGRail Intelligent Integration of Railway Systems Instrument: Integrated Project Thematic Priority: Priority 6.2-Sustainable Surface Transport InteGrail-Publishable Final Activity Report."

[91] A. Berrado, "A Framework for Risk Management in Railway Sector: Application to Road-Rail Level Crossings," *Open Transp. J.*, vol. 5, no. 1, pp. 34–44, 2011, doi: 10.2174/1874447801105010034.

[92] W. Górka *et al.*, "Cloud decision support system for risk management in railway transportation," in *ICSOFT 2019 - Proceedings of the 14th International Conference on Software Technologies*, 2019, pp. 475–482, doi: 10.5220/0007837904750482.

[93] The European Commission, "Commission Implementing Regulation (EU) No 402/2013 of 30 April 2013 on the common safety method for risk evaluation and assessment and repealing Regulation (EC) No 352/2009," *Off. J. Eur. Union*, vol. 56, no. L 121, 2013.

[94] J. Bagiński, B. Flisiuk, W. Górka, D. Rogowski, and T. Stęclik, "Multi-criteria decision analysis in the railway risk management process," in *Communications in Computer and Information Science*, 2019, vol. 1018, pp. 126–138, doi: 10.1007/978-3-030-19093-4_10.

[95] E. Nedeliakova, M. Hudakova, M. Masar, L. Lizbetinova, R. Stasiak-Betlejewska, and P. Šulko, "Sustainability of railway undertaking services with lean philosophy in risk management-Case study," *Sustain.*, vol. 12, no. 13, Jul. 2020, doi: 10.3390/su12135298.

[96] F. Transit Administration, “Sample Safety Risk Assessment Matrices for Rail Transit Agencies,” 2019.

[97] T. Aven *et al.*, “Society for Risk Analysis Glossary Expert group members: Society for Risk Analysis Glossary SOCIETY FOR RISK ANALYSIS GLOSSARY.” 2018.

[98] Health and Safety Authority, “Risk Assessments and Safety Statements A guide to Our Vision healthy, safe and productive lives A Guide to Risk Assessments and Safety Statements,” *Heal. Saf. Auth.*, 2016.

[99] ISO 31000:2018, “ISO 31000:2018(en), Risk management — Guidelines,” *ISO*. 2018.

[100] BS EN 50126-2:2017, “50126 BSI Standards Publication Railway Applications - The Specification and Demonstration of Reliability , Availability , Maintainability and Safety (RAMS) National foreword,” no. Part 1, 2017.

[101] The European Parliament and the Council, “Directive (EU) 2016/798 of the European Parliament and of the Council of 11 May 2016 on railway safety,” *Off. J. Eur. Union*, vol. 59, no. L 138, 2016.

[102] N. Taleb, *The black swan: The impact of the highly improbable*, vol. 2. London: The Random House Publishing Group, 2017.

[103] X. Lu, H. Luyang, and Z. Ma, “Research on risk management of military action,” in *ACM International Conference Proceeding Series*, 2018, pp. 119–124, doi: 10.1145/3277139.3277174.

[104] T. Kertis and D. Procházková, “ASSETS OF MODEL METRO STATION AND THEIR CRITICALITY,” *Acta Polytech. CTU Proc.*, vol. 5, p. 29, Sep. 2016, doi: 10.14311/app.2016.5.0029.

[105] T. Kertis and D. Prochazkova, “Tools for risk management of model metro station,” in *2016 Smart Cities Symposium Prague, SCSP 2016*, 2016, doi: 10.1109/SCSP.2016.7501019.

[106] L. He, Q. Liang, and S. Fang, “Challenges and Innovative Solutions in Urban Rail Transit Network Operations and Management: China’s Guangzhou Metro Experience,” *Urban Rail Transit*, vol. 2, no.

1, pp. 33–45, Mar. 2016, doi: 10.1007/s40864-016-0036-y.

[107] S. B. Turner Deborah Keeley Michael Glossop MEng George Brownless Shane Turner, “Review of Railway Safety’s Safety Risk Model HSL/2002/06,” 2002.

[108] M. Rausand, “Risk Assessment: Theory, Methods, and Applications,” *undefined*, 2011.

[109] T. Aven, “On the meaning of a black swan in a risk context,” *Saf. Sci.*, vol. 57, pp. 44–51, Aug. 2013, doi: 10.1016/j.ssci.2013.01.016.

[110] D. Wu, D. L. Olson, and A. Dolgui, “Artificial intelligence in engineering risk analytics,” *Engineering Applications of Artificial Intelligence*, vol. 65. pp. 433–435, 2017, doi: 10.1016/j.engappai.2017.09.001.

[111] N. Schneidewind, “Software risk analysis,” *Int. J. Reliab. Qual. Saf. Eng.*, vol. 16, no. 2, pp. 117–136, Apr. 2009, doi: 10.1142/S0218539309003320.

[112] W. E. Wong and Y. Qi, “BP neural network-based effective fault localization,” in *International Journal of Software Engineering and Knowledge Engineering*, 2009, vol. 19, no. 4, pp. 573–597, doi: 10.1142/S021819400900426X.

[113] J. Luo, X. Liu, and M. Fan, “A trust model based on fuzzy recommendation for mobile ad-hoc networks,” *Comput. Networks*, vol. 53, no. 14, pp. 2396–2407, 2009, doi: 10.1016/j.comnet.2009.04.008.

[114] E. K. Chalouhi, I. Gonzalez, C. Gentile, and R. Karoumi, “Vibration-based SHM of railway bridges using machine learning: The influence of temperature on the health prediction,” in *Lecture Notes in Civil Engineering*, vol. 5, Springer, 2018, pp. 200–211.

[115] REDRUTH TOWN COUNCIL, “RISK MANAGEMENT STRATEGY & POLICY.” [Online]. Available: http://www.redruth-tc.gov.uk/_UserFiles/Files/Risk Management Strategy Policy 17.pdf. [Accessed:

24-Mar-2022].

[116] University of London, “Introduction to Risk Analysis and Risk Assessment and Risk Assessment,” London, Oct. 2008.

[117] H. Alawad, S. Kaewunruen, and A. Min, “Utilizing Big Data for Enhancing Passenger Safety in Railway Stations,” in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 603, no. 5, doi: 10.1088/1757-899X/603/5/052031.

[118] N. Awang, L. Narayana Samy, N. Maarop, P. Magalingam, N. Hafizah binti Hassan, and D. Wong Hooi Ten, “Risk Assessment using Big Data Analytics Approach: A Review,” *Open Int. J. Informatics*, vol. 6, no. 4, p. 1, 2018.

[119] M. Leimeister and A. Kolios, “A review of reliability-based methods for risk analysis and their application in the offshore wind industry,” *Renewable and Sustainable Energy Reviews*, vol. 91. Pergamon, pp. 1065–1076, 01-Aug-2018, doi: 10.1016/j.rser.2018.04.004.

[120] M. Glossop MEng, A. Ioannides BEng, and J. Gould, “REVIEW OF HAZARD IDENTIFICATION TECHNIQUES,” 2000.

[121] D. Baccarini and R. Archer, “The risk ranking of projects: A methodology,” *Int. J. Proj. Manag.*, vol. 19, no. 3, pp. 139–145, 2001, doi: 10.1016/S0263-7863(99)00074-5.

[122] F. D. Patterson and K. Neailey, “A risk register database system to aid the management of project risk,” *Int. J. Proj. Manag.*, vol. 20, no. 5, pp. 365–374, 2002, doi: 10.1016/S0263-7863(01)00040-0.

[123] J. Pyra and J. Trask, “Risk Management Post Analysis: Gauging the Success of a Simple Strategy in a Complex Project,” *Proj. Manag. J.*, vol. 33, no. 2, pp. 41–48, 2002, doi: 10.1177/875697280203300206.

[124] B. Rodolfo, M. C. Mason, and F. Nassivera, “Green Marketing and Renewable Energy: Evidence on

Motivations and Behaviour in the Aquacultural Market," *Chinese Bus. Rev.*, vol. 10, no. 12, 2011, doi: 10.17265/1537-1506/2011.12.003.

[125] S. P. Smith and M. D. Harrison, "Measuring reuse in hazard analysis," in *Reliability Engineering and System Safety*, 2005, vol. 89, no. 1, pp. 93–104, doi: 10.1016/j.ress.2004.08.010.

[126] D. White, "Application of systems thinking to risk management: A review of the literature," *Management Decision*, vol. 33, no. 10. MCB UP Ltd, pp. 35–45, 01-Dec-1995, doi: 10.1108/EUM0000000003918.

[127] En50126, "Railway applications – The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)," *Part 1 Basic Requir. generic Process*, vol. 3, 1999.

[128] International Railway Safety Council, "Common Risks Managed by the Railway Safety - IRSC." [Online]. Available: https://international-railway-safety-council.com/common-risks-managed-railway-industry/. [Accessed: 09-Apr-2022].

[129] RSSB, "Stations, Platforms and Trains." [Online]. Available: https://www.rssb.co.uk/what-we-do/key-industry-topics/platform-train-interface/stations-platforms-and-trains. [Accessed: 09-Apr-2022].

[130] RSSB, "Annual Health and Safety Report 19/2020," 2020.

[131] RSSB, "Station Operations 2019/20," 2020.

[132] K. Qian, L. Yu, and S. Gao, "Fault Tree Construction Model Based on Association Analysis for Railway Overhead Contact System," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 96–105, Nov. 2020, doi: 10.2991/ijcis.d.201105.002.

[133] Q. Zhang, Y. Zhuang, Y. Wei, H. Jiang, and H. Yang, "Railway Safety Risk Assessment and Control Optimization Method Based on FTA-FPN: A Case Study of Chinese High-Speed Railway Station," *J.*

Adv. Transp., vol. 2020, 2020, doi: 10.1155/2020/3158468.

[134] W. A. Wolfe, "FAULT TREE ANALYSIS.," *At. Energy Canada Limited, AECL*, no. 6172, pp. 59–99, 1978, doi: 10.4018/978-1-4666-9717-1.les2.

[135] J. Bäckman, "RAILWAY SAFETY-Risks and Economics Traffic and Transport Planning."

[136] L. Shi, J. Shuai, and K. Xu, "Fuzzy fault tree assessment based on improved AHP for fire and explosion accidents for steel oil storage tanks," *J. Hazard. Mater.*, vol. 278, pp. 529–538, Aug. 2014, doi: 10.1016/j.jhazmat.2014.06.034.

[137] K. Moodie, "The King's Cross fire: damage assessment and overview of the technical investigation," *Fire Saf. J.*, vol. 18, no. 1, pp. 13–33, Jan. 1992, doi: 10.1016/0379-7112(92)90045-E.

[138] M. Tsukahara, Y. Koshiba, and H. Ohtani, "Effectiveness of downward evacuation in a large-scale subway fire using Fire Dynamics Simulator," *Tunn. Undergr. Sp. Technol.*, vol. 26, no. 4, pp. 573–581, Jul. 2011, doi: 10.1016/j.tust.2011.02.002.

[139] B. Dai and T. Wang, "Risk assessment based on accident theory in urban railway transportation," in *Proceedings - 2010 International Conference on Intelligent System Design and Engineering Application, ISDEA 2010*, 2010, vol. 2, pp. 318–320, doi: 10.1109/ISDEA.2010.374.

[140] W. Chen, Y. Zhang, M. T. Khasawneh, and Z. Geng, "Risk analysis on Beijing metro operation initiated by human factors," *J. Transp. Saf. Secur.*, vol. 11, no. 6, pp. 683–699, 2019, doi: 10.1080/19439962.2018.1470586.

[141] S. Kabir, "An overview of fault tree analysis and its application in model based dependability analysis," *Expert Systems with Applications*, vol. 77. Elsevier Ltd, pp. 114–135, 01-Jul-2017, doi: 10.1016/j.eswa.2017.01.058.

[142] Y. Sun and N. Liu, "Casualties Analysis in Fire Accident of Passenger Transport Station Based on FTA,"

in *Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020*, 2020, pp. 3995–3998, doi: 10.1109/CCDC49329.2020.9163938.

[143] Y. Guan, J. Zhao, T. Shi, and P. Zhu, “Fault tree analysis of fire and explosion accidents for dual fuel (diesel/natural gas) ship engine rooms,” *J. Mar. Sci. Appl.*, vol. 15, no. 3, pp. 331–335, 2016, doi: 10.1007/s11804-016-1366-6.

[144] N. Qu, N. Cheng, Z. Li, and J. Li, “Failure risk assessment of fire alarm system,” in *Proceedings - 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer, MEC 2013*, 2013, pp. 250–253, doi: 10.1109/MEC.2013.6885079.

[145] X. Li and H. K. Lo, “Energy minimization in dynamic train scheduling and control for metro rail operations,” *Transp. Res. Part B Methodol.*, vol. 70, pp. 269–284, Dec. 2014, doi: 10.1016/j.trb.2014.09.009.

[146] J. Wu, Z. Hu, J. Chen, and Z. Li, “Risk assessment of underground subway stations to fire disasters using Bayesian network,” *Sustain.*, vol. 10, no. 10, p. 3810, Oct. 2018, doi: 10.3390/su10103810.

[147] “Passenger safety (stations) - Network Rail.” [Online]. Available: https://www.networkrail.co.uk/who-we-are/how-we-work/performance/safety-performance/passenger-safety-stations/. [Accessed: 25-Jan-2021].

[148] “Contribute | SPARK,” *Mitigation of Slips, Trips and Falls in Stations (S327)*. [Online]. Available: https://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=25889. [Accessed: 25-Jan-2021].

[149] ORR, “Rail Safety Statistics | Office of Rail and Road,” 2021. [Online]. Available: https://www.orr.gov.uk/media/16177. [Accessed: 22-Jan-2021].

[150] H. Xu, L. Jiao, S. Chen, M. Deng, and N. Shen, “An innovative approach to determining high-risk nodes in a complex urban rail transit station: A perspective of promoting urban sustainability,” *Sustain.*, vol. 10, no. 7, p. 2456, Jul. 2018, doi: 10.3390/su10072456.

[151] C. Poirier, S. Adelé, and J. M. Burkhardt, “Individual accidents at the interface between platform, train and tracks (PT2I) in the subway: a literature review,” *Cogn. Technol. Work*, vol. 1, p. 3, Feb. 2020, doi: 10.1007/s10111-020-00624-1.

[152] W. Ming-Yan, W. Hong, and Z.-G. Liu, “Reach on Fault Tree Analysis of Train Derailment in Urban Rail Transit,” *Int. J. Bus. Soc. Sci.*, vol. 5, no. 8, pp. 128–135, 2014.

[153] D. Brabie and E. Andersson, “An overview of some high-speed train derailments: Means of minimizing consequences based on empirical observations,” *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 222, no. 4. pp. 441–463, 2008, doi: 10.1243/09544097JRRT149.

[154] N. Pavlović, M. Kostadinović, M. Marković, M. Ivić, and M. Kosijer, “ANALYSIS OF THE TRAIN DERAILMENT CAUSES USING FAULT TREE ANALYSIS,” *Horizons*, 2016, doi: 10.20544/HORIZONS.B.03.1.16.P52.

[155] J. C. Jong, Y.-C. C. Lai, C. C. Young, and Y. F. Chen, “Application of Fault Tree Analysis and Swiss Cheese Model to the Overspeed Derailment of Puyuma Train in Yilan, Taiwan,” *Transp. Res. Rec.*, vol. 2674, no. 5, pp. 33–46, May 2020, doi: 10.1177/0361198120914887.

[156] J. Lim, Y. Choi, K. Lee, Y. Jang, N. Lim, and J. Kong, “A Study on Cause Analysis of Derailment Accident Using Failure Tree Analysis,” *J. Korean Soc. Hazard Mitig.*, vol. 18, no. 2, pp. 19–26, Feb. 2018, doi: 10.9798/kosham.2018.18.2.19.

[157] Y. F. Li, J. Mi, H. Z. Huang, S. P. Zhu, and N. Xiao, “Fault tree analysis of train rear-end collision accident considering common cause failure,” *Eksploat. i Niezawodn.*, vol. 15, no. 4, pp. 403–408, 2013.

[158] X. Liu, M. R. Saat, and C. P. L. Barkan, “Analysis of causes of major train derailment and their effect on accident rates,” *Transp. Res. Rec.*, vol. 2289, no. 2289, pp. 154–163, 2012, doi: 10.3141/2289-20.

[159] M. A. Britton, S. Asnaashari, and G. J. M. Read, “Analysis of train derailment cause and outcome in Victoria, Australia, between 2007 and 2013: Implications for regulation,” *J. Transp. Saf. Secur.*, vol. 9, no. 1, pp. 45–63, 2017, doi: 10.1080/19439962.2015.1088906.

[160] D. ming Zhang, J. zhang Zhang, H. wei Huang, C. chong Qi, and C. yu Chang, “Machine learning-based prediction of soil compression modulus with application of 1D settlement,” *J. Zhejiang Univ. Sci. A*, vol. 21, no. 6, pp. 430–444, 2020, doi: 10.1631/jzus.A1900515.

[161] M. Chenariyan Nakhaee, D. Hiemstra, M. Stoelinga, and M. van Noort, “The Recent Applications of Machine Learning in Rail Track Maintenance: A Survey,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11495 LNCS, pp. 91–105, doi: 10.1007/978-3-030-18744-6_6.

[162] A. M. Rakoczy, S. T. Wilk, and M. C. Jones, “Security and Safety of Rail Transit Tunnels,” *Transp. Res. Rec.*, vol. 2673, no. 1, pp. 92–101, 2019, doi: 10.1177/0361198118822819.

[163] R. Tanwar, A. K. Nazari, V. Deep, and N. Garg, “Railway Reservation Verification by Aadhar Card,” in *Procedia Computer Science*, 2016, vol. 85, pp. 970–975, doi: 10.1016/j.procs.2016.05.289.

[164] *The Routledge Handbook of Terrorism Research*. 2011.

[165] NATIONAL SECURITY RESEARCH DIVISION, “About RDWTI | RAND.” [Online]. Available: https://www.rand.org/nsrd/projects/terrorism-incidents/about.html. [Accessed: 29-Jan-2021].

[166] A. V. Shvetsov and S. V. Shvetsova, “Research of a Problem of Terrorist Attacks in the Metro (Subway, U-Bahn, Underground, MRT, Rapid Transit, Metrorail),” *Eur. J. Secur. Res.*, vol. 2, no. 2, pp. 131–145, Nov. 2017, doi: 10.1007/s41125-017-0019-3.

[167] R. Jankura, “Assessment of Fenestration Security as Part of Railway Infrastructure,” in *Transportation Research Procedia*, 2021, vol. 55, pp. 1555–1562, doi: 10.1016/j.trpro.2021.07.187.

[168] A. Loukaitou-Sideris, B. D. Taylor, and C. N. Y. Fink, “Rail transit security in an international context: Lessons from four cities,” *Urban Aff. Rev.*, vol. 41, no. 6, pp. 727–748, 2006, doi: 10.1177/1078087406287581.

[169] G. Hatzivasilis *et al.*, “SPD-safe: Secure administration of railway intelligent transportation systems,” *Electron.*, vol. 10, no. 1, pp. 1–26, Jan. 2021, doi: 10.3390/electronics10010092.

[170] H. Alawad and S. Kaewunruen, “Wireless Sensor Networks: Toward Smarter Railway Stations,” *Infrastructures*, vol. 3, no. 3, p. 24, Jul. 2018, doi: 10.3390/infrastructures3030024.

[171] M. Banić, A. Miltenović, M. Pavlović, and I. Ćirić, “Intelligent machine vision based railway infrastructure inspection and monitoring using UAV,” *Facta Univ. Ser. Mech. Eng.*, vol. 17, no. 3, pp. 357–364, Dec. 2019, doi: 10.22190/FUME190507041B.

[172] P. Radanliev *et al.*, “Design of a dynamic and self-adapting system, supported with artificial intelligence, machine learning and real-time intelligence for predictive cyber risk analytics in extreme environments – cyber risk in the colonisation of Mars,” *Saf. Extrem. Environ.*, pp. 1–12, Feb. 2021, doi: 10.1007/s42797-021-00025-1.

[173] E. Matsika, C. O’Neill, U. Battista, M. Khosravi, A. de S. Laporte, and E. Munoz, “Development of Risk Assessment Specifications for Analysing Terrorist Attacks Vulnerability on Metro and Light Rail Systems,” in *Transportation Research Procedia*, 2016, vol. 14, pp. 1345–1354, doi: 10.1016/j.trpro.2016.05.207.

[174] F. T. Illiyas, S. K. Mani, A. P. Pradeepkumar, and K. Mohan, “Human stampedes during religious festivals: A comparative review of mass gathering emergencies in India,” *International Journal of Disaster Risk Reduction*, vol. 5. Elsevier, pp. 10–18, 01-Sep-2013, doi: 10.1016/j.ijdrr.2013.09.003.

[175] A. W. Evans, “Fatal train accidents on Europe’s railways: 1980-2009,” *Accid. Anal. Prev.*, vol. 43, no. 1, pp. 391–401, Jan. 2011, doi: 10.1016/j.aap.2010.09.009.

[176] “Service Reliability and Urban Public Transport Design Niels van Oort.”

[177] C. Martella, J. Li, C. Conrado, and A. Vermeeren, “On current crowd management practices and the need for increased situation awareness, prediction, and intervention,” *Saf. Sci.*, vol. 91, pp. 381–393, Jan. 2017, doi: 10.1016/j.ssci.2016.09.006.

[178] M. J. Smith and R. V. Clarke, “Crime and Public Transport,” *Crime and Justice*, vol. 27, pp. 169–233, Jan. 2000, doi: 10.1086/652200.

[179] C. Teague and D. Leith, “Men of Steel or Plastic Cops : The use of ethnography as a transformative agent,” *ECU Publ. Pre. 2011*, Jan. 2008.

[180] D. Uzzell and J. Brown, “Conceptual progress in understanding fear of crime in railway stations Heritage and the Psychology of the Past View project Psychology and Sustainable Development View project,” 2007.

[181] R. R. M. Gershon, K. A. Qureshi, M. A. Barrera, M. J. Erwin, and F. Goldsmith, “Health and safety hazards associated with subways: A review,” *Journal of Urban Health*, vol. 82, no. 1. pp. 10–20, 2005, doi: 10.1093/jurban/jti004.

[182] S. Macintyre and R. Homel, “Danger on the dance floor: A study of interior design, crowding and aggression in nightclubs,” *Crime Prev. Stud.*, vol. 7, pp. 91–113, 1997.

[183] K. Thompson, L. Hirsch, S. Mueller, and Sophia Rainbird, “A socio-economic study of carriage and platform crowding in the Australian railway industry: Final Report,” Brisbane, 2011.

[184] M. Kyriakidis, R. Hirsch, and A. Majumdar, “Metro railway safety: An analysis of accident precursors,” *Saf. Sci.*, vol. 50, no. 7, pp. 1535–1548, 2012, doi: 10.1016/j.ssci.2012.03.004.

[185] H. Alawad, M. An, and S. Kaewunruen, “Utilizing an adaptive neuro-fuzzy inference system (ANFIS) for overcrowding level risk assessment in railway stations,” *Appl. Sci.*, vol. 10, no. 15, p. 5156, Jul.

2020, doi: 10.3390/app10155156.

[186] H. Alawad and S. Kaewunruen, “5G Intelligence Underpinning Railway Safety in the COVID-19 Era,” *Front. Built Environ.*, vol. 7, p. 14, Feb. 2021, doi: 10.3389/fbuil.2021.639753.

[187] M. MacChi, M. Garetti, D. Centrone, L. Fumagalli, and G. Piero Pavirani, “Maintenance management of railway infrastructures based on reliability analysis,” *Reliab. Eng. Syst. Saf.*, vol. 104, pp. 71–83, Aug. 2012, doi: 10.1016/j.ress.2012.03.017.

[188] T. Loster, “Flood Trends and Global Change,” *China*, pp. 1–7, 1970.

[189] L. Hong, M. Ouyang, S. Peeta, X. He, and Y. Yan, “Vulnerability assessment and mitigation for the Chinese railway system under floods,” *Reliab. Eng. Syst. Saf.*, vol. 137, pp. 58–68, May 2015, doi: 10.1016/j.ress.2014.12.013.

[190] J. C. J. H. Aerts *et al.*, “Integrating human behaviour dynamics into flood disaster risk assessment,” *Nat. Clim. Chang.*, vol. 8, no. 3, pp. 193–199, Mar. 2018, doi: 10.1038/s41558-018-0085-1.

[191] B. P. Arkell and G. J. C. Darch, “Impact of climate change on London’s transport network,” *Proc. Inst. Civ. Eng. Munic. Eng.*, vol. 159, no. 4, pp. 231–237, 2006, doi: 10.1680/muen.2006.159.4.231.

[192] S. L. Binti Sa’adin, S. Kaewunruen, and D. Jaroszweski, “Heavy rainfall and flood vulnerability of Singapore-Malaysia high speed rail system,” *Aust. J. Civ. Eng.*, vol. 14, no. 2, pp. 123–131, Oct. 2016, doi: 10.1080/14488353.2017.1336895.

[193] A. G. Rumson and S. H. Hallett, “Innovations in the use of data facilitating insurance as a resilience mechanism for coastal flood risk,” *Sci. Total Environ.*, vol. 661, pp. 598–612, Apr. 2019, doi: 10.1016/j.scitotenv.2019.01.114.

[194] H. Zhuang, Z. Hu, X. Wang, and G. Chen, “Seismic responses of a large underground structure in liquefied soils by FEM numerical modelling,” *Bull. Earthq. Eng.*, vol. 13, no. 12, pp. 3645–3668, Dec.

2015, doi: 10.1007/s10518-015-9790-6.

[195] W. Ke, S. He, and G. Zhang, “3D DYNAMIC TIME-HISTORY RESPONSE ANALYSIS OF AN ISLAND PLATFORM METRO STATION IN LOESS AREA,” *Stavební Obz. - Civ. Eng. J.*, vol. 27, no. 4, pp. 613–634, 2018, doi: 10.14311/cej.2018.04.0049.

[196] Y. M. A. Hashash, J. J. Hook, B. Schmidt, and J. I-Chiang Yao, “Seismic design and analysis of underground structures,” *Tunn. Undergr. Sp. Technol.*, vol. 16, no. 4, pp. 247–293, Oct. 2001, doi: 10.1016/S0886-7798(01)00051-7.

[197] T. Iwatate, Y. Kobayashi, H. Kusu, and K. Rin, “Investigation and Shaking Table Tests of Subway Structures of the Hyogoken-Nanbu Earthquake,” in *Proceedings 12th World Conference on Earthquake Engineering*, 2000, pp. 1–8.

[198] N. Kurata, B. F. Spencer, and M. Ruiz-Sandoval, “Risk monitoring of buildings with wireless sensor networks,” in *Structural Control and Health Monitoring*, 2005, vol. 12, no. 3–4, pp. 315–327, doi: 10.1002/stc.73.

[199] P. Jiao and A. H. Alavi, “Artificial intelligence in seismology: Advent, performance and future trends,” *Geosci. Front.*, vol. 11, no. 3, pp. 739–744, May 2020, doi: 10.1016/j.gsf.2019.10.004.

[200] A. V. Gheorghe, J. Birchmeier, D. Vamanu, I. Papazoglou, and W. Kröger, “Comprehensive risk assessment for rail transportation of dangerous goods: A validated platform for decision support,” *Reliab. Eng. Syst. Saf.*, vol. 88, no. 3, pp. 247–272, Jun. 2005, doi: 10.1016/j.ress.2004.07.017.

[201] W. Huang and Y. Zhang, “Railway Dangerous Goods Transportation System Risk Assessment: An Approach Combining FMEA With Pessimistic–Optimistic Fuzzy Information Axiom Considering Acceptable Risk Coefficient,” *IEEE Trans. Reliab.*, 2020, doi: 10.1109/TR.2020.2973431.

[202] O. Lavrukhin *et al.*, “Forming an automated technology to actively monitor the transportation of dangerous cargoes by railroad,” *Eastern-European J. Enterp. Technol.*, vol. 3, no. 3–105, pp. 78–85,

2020, doi: 10.15587/1729-4061.2020.205862.

[203] F. Sattari, R. Macciotta, and L. Lefsrud, “Process Safety Approach to Identify Opportunities for Enhancing Rail Transport Safety in Canada,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2675, no. 1, pp. 49–66, Jan. 2021, doi: 10.1177/0361198120957008.

[204] B. Pender, G. Currie, A. Delbosc, and N. Shiwakoti, “Planning for the unplanned: An international review of current approaches to service disruption management of railways,” in *Australasian Transport Research Forum, ATRF 2012 - Proceedings*, 2014.

[205] S. S. Chopra, T. Dillon, M. M. Bilec, and V. Khanna, “A network-based framework for assessing infrastructure resilience: A case study of the London metro system,” *J. R. Soc. Interface*, vol. 13, no. 118, 2016, doi: 10.1098/rsif.2016.0113.

[206] J. Weng, Y. Zheng, X. Yan, and Q. Meng, “Development of a subway operation incident delay model using accelerated failure time approaches,” *Accid. Anal. Prev.*, vol. 73, pp. 12–19, Dec. 2014, doi: 10.1016/j.aap.2014.07.029.

[207] D. Golightly and N. Dadashi, “The characteristics of railway service disruption: implications for disruption management,” *Ergonomics*, vol. 60, no. 3, pp. 307–320, Mar. 2017, doi: 10.1080/00140139.2016.1173231.

[208] H. Xu and B. Xue, “Key indicators for the resilience of complex urban public spaces,” *J. Build. Eng.*, vol. 12, pp. 306–313, 2017, doi: 10.1016/j.jobe.2017.06.018.

[209] L. Pelà, A. Benedetti, A. Aprile, and E. Mangoni, “Seismic assessment of the Milano Centrale railway station,” *Int. J. Archit. Herit.*, vol. 7, no. 6, pp. 609–627, Nov. 2013, doi: 10.1080/15583058.2012.654894.

[210] V. J. Hodge, S. O’Keefe, M. Weeks, and A. Moulds, “Wireless sensor networks for condition monitoring in the railway industry: A survey,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp.

1088–1106, Jun. 2015, doi: 10.1109/TITS.2014.2366512.

[211] “Guidance | NFCC CPO.” [Online]. Available: https://www.ukfrs.com/guidance/search/complexity-rail-infrastructure. [Accessed: 26-Jan-2021].

[212] Y. Zhang, X. Ge, W. Yang, and J. Guo, “Analysing railway safety with systems thinking,” in *ICRT 2017: Railway Development, Operations, and Maintenance - Proceedings of the 1st International Conference on Rail Transportation 2017*, 2018, vol. 2017-July, pp. 464–475, doi: 10.1061/9780784481257.047.

[213] K. Li and S. Wang, “A network accident causation model for monitoring railway safety,” *Saf. Sci.*, vol. 109, pp. 398–402, Nov. 2018, doi: 10.1016/j.ssci.2018.06.008.

[214] Y. Wang, Y. Liang, and H. Sun, “A Regret Theory-Based Decision-Making Method for Urban Rail Transit in Emergency Response of Rainstorm Disaster,” *J. Adv. Transp.*, vol. 2020, 2020, doi: 10.1155/2020/3235429.

[215] L. Collis, F. Schmid, and A. Tobias, “Managing incidents in a complex system: A railway case study,” *Cognition, Technology and Work*, vol. 16, no. 2. Springer-Verlag London Ltd, pp. 171–185, 07-Mar-2014, doi: 10.1007/s10111-013-0255-x.

[216] ORR, “Rail Safety Statistics 2018-19 Annual Statistical Release,” LONDON, 2019.

[217] M. T. Ahi and K. Yildiz, “Determining performance criteria of railway market: A case study,” *J. Open Innov. Technol. Mark. Complex.*, vol. 4, no. 1, p. 6, Feb. 2018, doi: 10.1186/s40852-018-0084-5.

[218] V. Ceccato, D. J. Wiebe, K. Vrotsou, U. Nyberg, and A. Grundberg, “The situational conditions of suicide in transit environments: An analysis using CCTV footage,” *J. Transp. Heal.*, vol. 20, p. 100976, Mar. 2021, doi: 10.1016/j.jth.2020.100976.

[219] Y. Wen, J. Leng, X. Shen, G. Han, L. Sun, and F. Yu, “Environmental and health effects of ventilation in

subway stations: A literature review," *International Journal of Environmental Research and Public Health*, vol. 17, no. 3. MDPI AG, p. 1084, 08-Feb-2020, doi: 10.3390/ijerph17031084.

[220] Y. Wen, J. Leng, F. Yu, and C. W. Yu, "Integrated design for underground space environment control of subway stations with atriums using piston ventilation," *Indoor Built Environ.*, vol. 29, no. 9, pp. 1300–1315, 2020, doi: 10.1177/1420326X20941349.

[221] N. P. Russell Stuart, *Artificial Intelligence: A Modern Approach*, vol. 9, no. 2. 1995.

[222] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943, doi: 10.1007/BF02478259.

[223] V. Bush, "As we may think," *The Atlantic Monthly.*, vol. 58, no. 3. p. 7, 2015, doi: 10.1145/2728204.

[224] J. P. Bowen, "The Impact of Alan Turing: Formal Methods and Beyond," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11430 LNCS, pp. 202–235, doi: 10.1007/978-3-030-17601-3_5.

[225] Y. G. Smith Chris, McGuire Brian, Huang Ting, "The history of artificial intelligence," *Univ. Washingt.*, vol. 27, pp. 1–26, 2006, doi: 10.4044/joma.132.144.

[226] A. SADEK, *Artificial Intelligence in Transportation: Information for Application*. 2007.

[227] Z. X. Guo and W. K. Wong, "Fundamentals of artificial intelligence techniques for apparel management applications," in *Optimizing Decision Making in the Apparel Supply Chain Using Artificial Intelligence (AI): From Production to Retail*, 2013, pp. 13–40.

[228] R. A. Ariyaluran Habeeb, F. Nasaruddin, A. Gani, I. A. Targio Hashem, E. Ahmed, and M. Imran, "Real-time big data processing for anomaly detection: A Survey," *International Journal of Information Management*, vol. 45. Elsevier Ltd, pp. 289–307, 01-Apr-2019, doi: 10.1016/j.ijinfomgt.2018.08.006.

[229] Y. Lu, Q. Li, and W. Xiao, "Case-based reasoning for automated safety risk analysis on subway

operation: Case representation and retrieval," *Saf. Sci.*, vol. 57, pp. 75–81, Aug. 2013, doi: 10.1016/j.ssci.2013.01.020.

[230] Z. Xu and J. H. Saleh, "Machine learning for reliability engineering and safety applications: Review of current status and future opportunities," *arXiv*. arXiv, p. 107530, 18-Aug-2020, doi: 10.1016/j.ress.2021.107530.

[231] D. K. Judith Hurwitz, *Machine Learning, IBM Limited Edition*, vol. 35, no. 5. 2018.

[232] C. Strobl, J. Malley, and G. Tutz, "An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests," *Psychol. Methods*, vol. 14, no. 4, pp. 323–348, 2009, doi: 10.1037/a0016973.

[233] M. Steinbach, G. Karypis, and V. Kumar., "4 Classification: Basic Concepts, Decision Trees, and Model Evaluation," 2000.

[234] S. Singh and P. Gupta, "Comparative Study ID3, CART and C4.5 Decision Tree Algorithm: A Survey," *Smart Innov. Syst. Technol.*, vol. 27, no. 27, pp. 97–103, 2014.

[235] B. Baykara, "Impact of Evaluation Methods on Decision Tree Accuracy," no. April, p. 72, 2015.

[236] Explorium, "Decision Trees: The Complete Guide to Decision Tree Classifier l Explorium." [Online]. Available: https://www.explorium.ai/blog/the-complete-guide-to-decision-trees/. [Accessed: 02-Apr-2021].

[237] T. Marwala, *Artificial Intelligence Techniques for Rational Decision Making*. 2014.

[238] S. Nissen, "Implementation of a Fast Artificial Neural Network Library (FANN)," *Training*, vol. 31, p. 92, 2003.

[239] T. M. Neuron, "3 Artificial Neural Network Structures," pp. 90–127, 1992.

[240] A. I. Georgevici and M. Terblanche, "Neural networks and deep learning: a brief introduction,"

Intensive Care Medicine, vol. 45, no. 5. pp. 712–714, 2019, doi: 10.1007/s00134-019-05537-w.

[241] D. M. Rodvold, D. G. McLeod, J. M. Brandt, P. B. Snow, and G. P. Murphy, "Introduction to artificial neural networks for physicians: Taking the lid off the black box," *Prostate*, vol. 46, no. 1, pp. 39–44, Jan. 2001, doi: 10.1002/1097-0045(200101)46:1<39::AID-PROS1006>3.0.CO;2-M.

[242] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3. pp. 31–44, Mar-1996, doi: 10.1109/2.485891.

[243] S. H. Chen, A. J. Jakeman, and J. P. Norton, "Artificial Intelligence techniques: An introduction to their use for modelling environmental systems," *Math. Comput. Simul.*, vol. 78, no. 2–3, pp. 379–400, Jul. 2008, doi: 10.1016/j.matcom.2008.01.028.

[244] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.

[245] D. Uzun Ozsahin, B. Uzun, I. Ozsahin, M. T. Mustapha, and M. S. Musa, "Fuzzy logic in medicine," in *Biomedical Signal Processing and Artificial Intelligence in Healthcare*, Elsevier, 2020, pp. 153–182.

[246] T. Floor and O. Greenwich, "WorldQuant Perspectives Reshaping the World with Fuzzy Logic Reshaping the World with Fuzzy Logic," 2018.

[247] E. Kayacan and M. A. Khanesar, "Fundamentals of Type-2 Fuzzy Logic Theory," in *Fuzzy Neural Networks for Real Time Control Applications*, Elsevier, 2016, pp. 25–35.

[248] Guru99, "Fuzzy Logic Tutorial: What is, Architecture, Application, Example." [Online]. Available: https://www.guru99.com/what-is-fuzzy-logic.html#4. [Accessed: 16-Mar-2021].

[249] M. Negnevitsky, *Artificial Intelligence - A Guide to Intelligent Systems*, vol. 110, no. 5. 2002.

[250] J. S. R. Jang, "ANFIS : Adap tive-Ne twork-Based Fuzzy Inference System. IEEE Transactions on Systems," *Man Cybern.*, vol. 23, no. 3, pp. 665–685, 1993.

[251] S. Mitra, Y. Hayashi, and S. Member, “Neuro – Fuzzy Rule Generation : Survey in,” *IEEE Trans. Neural Networks*, vol. 11, no. 3, pp. 748–768, 2000.

[252] M. Buragohain and C. Mahanta, “A novel approach for ANFIS modelling based on full factorial design,” *Appl. Soft Comput. J.*, vol. 8, no. 1, pp. 609–625, Jan. 2008, doi: 10.1016/j.asoc.2007.03.010.

[253] U. Çaydaş, A. Hasçalik, and S. Ekici, “An adaptive neuro-fuzzy inference system (ANFIS) model for wire-EDM,” *Expert Syst. Appl.*, vol. 36, no. 3 PART 2, pp. 6135–6139, Apr. 2009, doi: 10.1016/j.eswa.2008.07.019.

[254] T. Takagi and M. Sugeno, “Fuzzy Identification of Systems and Its Applications to Modeling and Control,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, no. 1, pp. 116–132, 1985, doi: 10.1109/TSMC.1985.6313399.

[255] K. V. Shihabudheen and G. N. Pillai, “Recent advances in neuro-fuzzy system: A survey,” *Knowledge-Based Syst.*, vol. 152, pp. 136–162, 2018, doi: 10.1016/j.knosys.2018.04.014.

[256] A. Şamandar, “A model of adaptive neural-based fuzzy inference system (ANFIS) for prediction of friction coefficient in open channel flow,” *Sci. Res. Essays*, vol. 6, no. 5, pp. 1020–1027, 2011, doi: 10.5897/SRE10.558.

[257] Q. Zhou, Y. D. Wong, H. S. Loh, and K. F. Yuen, “ANFIS model for assessing near-miss risk during tanker shipping voyages,” *Marit. Policy Manag.*, vol. 46, no. 4, pp. 377–393, May 2019, doi: 10.1080/03088839.2019.1569765.

[258] M. Buragohain, “Adaptive Network based Fuzzy Inference System (ANFIS) as a Tool for System Identification with Special Emphasis on Training Data Minimization,” 2009.

[259] J. Zhang, W. Chen, M. Gao, and G. Shen, “K-means-clustering-based fiber nonlinearity equalization techniques for 64-QAM coherent optical communication system,” *Opt. Express*, vol. 25, no. 22, p. 27570, Oct. 2017, doi: 10.1364/oe.25.027570.

[260] R. A. Haraty, M. Dimishkieh, and M. Masud, “An enhanced k-means clustering algorithm for pattern discovery in healthcare data,” *Int. J. Distrib. Sens. Networks*, vol. 2015, 2015, doi: 10.1155/2015/615740.

[261] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, “Latent Dirichlet Allocation,” *J. Mach. Learn. Res. 3*, vol. 3, no. 4–5, pp. 993–1022, 2003, doi: 10.1016/B978-0-12-411519-4.00006-9.

[262] D. O. Oneat¸˘, “Probabilistic Latent Semantic Analysis.”

[263] C. LaBonne, B. Burke, and M. Whitman, “Role of MAP kinase in mesoderm induction and axial patterning during Xenopus development,” *Development*, vol. 121, no. 5, pp. 1475–1486, 1995, doi: 10.1016/0168-9525(96)81381-3.

[264] Q. Yan, “6.1 Latent Dirichlet Allocation | Notes for ‘Text Mining with R: A Tidy Approach.’” [Online]. Available: https://bookdown.org/Maxine/tidy-text-mining/latent-dirichlet-allocation.html. [Accessed: 25-Mar-2021].

[265] A. Naskar, “Latent Dirichlet Allocation for Beginners: A high level overview - ThinkInfi.” [Online]. Available: https://thinkinfi.com/latent-dirichlet-allocation-for-beginners-a-high-level-overview/. [Accessed: 25-Mar-2021].

[266] Stacey Ronaghan, “Deep Learning: Overview of Neurons and Activation Functions,” *https://srnghn.medium.com/deep-learning-overview-of-neurons-and-activation-functions-1d98286cf1e4*, 2018. [Online]. Available: https://srnghn.medium.com/deep-learning-overview-of-neurons-and-activation-functions-1d98286cf1e4. [Accessed: 23-Apr-2021].

[267] M. Rigla, G. García-Sáez, B. Pons, and M. E. Hernando, “Artificial Intelligence Methodologies and Their Application to Diabetes,” *J. Diabetes Sci. Technol.*, vol. 12, no. 2, pp. 303–310, Mar. 2018, doi: 10.1177/1932296817710475.

[268] J. Thrane, D. Zibar, and H. L. Christiansen, “Model-aided deep learning method for path loss

prediction in mobile communication systems at 2.6 GHz," *IEEE Access*, vol. 8, pp. 7925–7936, 2020, doi: 10.1109/ACCESS.2020.2964103.

[269] R. Vargas, A. Mosavi, and L. Ruiz, "DEEP LEARNING: A REVIEW," 2017.

[270] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11. pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.

[271] A. M. Rakoczy, S. T. Wilk, and M. C. Jones, "Security and Safety of Rail Transit Tunnels," *Transp. Res. Rec.*, vol. 2673, no. 1, pp. 92–101, Jan. 2019, doi: 10.1177/0361198118822819.

[272] Z. Kurd and T. Kelly, "Establishing safety criteria for artificial neural networks," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2003, vol. 2773 PART, pp. 163–169, doi: 10.1007/978-3-540-45224-9_24.

[273] W. Wen, J. Callahan, and M. Napolitano, "Towards Developing Verifiable Neural Network Controller," *Lab. Morgantown, Wv West Virginia Univ.*, 1996.

[274] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *J. Biomed. Inform.*, vol. 35, no. 5–6, pp. 352–359, Oct. 2002, doi: 10.1016/S1532-0464(03)00034-0.

[275] H. Liu, Z. Liu, and D. Liu, "Application of machine learning methods in maritime safety information classification," in *Proceedings - 2018 10th International Conference on Advanced Computational Intelligence, ICACI 2018*, 2018, pp. 735–740, doi: 10.1109/ICACI.2018.8377552.

[276] M. Chen, A. X. Zheng, J. Lloyd, M. I. Jordan, and E. Brewer, "Failure diagnosis using decision trees," in *Proceedings - International Conference on Autonomic Computing*, 2004, pp. 36–43, doi: 10.1109/ICAC.2004.1301345.

[277] M. Fredrikson, S. Jha, and T. Ristenpart, “Model inversion attacks that exploit confidence information and basic countermeasures,” in *Proceedings of the ACM Conference on Computer and Communications Security*, 2015, vol. 2015-Octob, pp. 1322–1333, doi: 10.1145/2810103.2813677.

[278] Y. Yuan and M. J. Shaw, “Induction of fuzzy decision trees,” *Fuzzy Sets Syst.*, vol. 69, no. 2, pp. 125–139, Jan. 1995, doi: 10.1016/0165-0114(94)00229-Z.

[279] Z. Xu and J. H. Saleh, “Machine learning for reliability engineering and safety applications: Review of current status and future opportunities,” *Reliability Engineering and System Safety*, vol. 211. 2021, doi: 10.1016/j.ress.2021.107530.

[280] GOV/UK, “Rail Accident Investigation Branch reports,” *UK Government*, 2014. [Online]. Available: https://www.gov.uk/government/organisations/rail-accident-investigation-branch. [Accessed: 13-Apr-2021].

[281] R. Houari, A. Bounceur, M. T. Kechadi, A. K. Tari, and R. Euler, “Dimensionality reduction in data mining: A Copula approach,” *Expert Syst. Appl.*, vol. 64, pp. 247–260, Dec. 2016, doi: 10.1016/j.eswa.2016.07.041.

[282] M. S. Hajakbari and B. Minaei-Bidgoli, “A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran’s Ministry of Labor data,” *J. Loss Prev. Process Ind.*, vol. 32, pp. 443–453, Oct. 2014, doi: 10.1016/j.jlp.2014.10.013.

[283] The MathWorks, “Detector Performance Analysis Using ROC Curves - MATLAB,” *http://www.mathworks.com*, 2018. [Online]. Available: https://www.mathworks.com/help/phased/ug/detector-performance-analysis-using-roc-curves.html;jsessionid=3398ea457b801ca02095fdf15ebc. [Accessed: 15-Apr-2021].

[284] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki, “Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification,”

J. Electron. Imaging, vol. 22, no. 4, p. 041106, Jul. 2013, doi: 10.1117/1.jei.22.4.041106.

[285] M. Şahin and R. Erol, "A Comparative Study of Neural Networks and ANFIS for Forecasting Attendance Rate of Soccer Games," *Math. Comput. Appl.*, vol. 22, no. 4, p. 43, Nov. 2017, doi: 10.3390/mca22040043.

[286] P. Srisaeng, G. S. Baxter, and G. Wild, "An adaptive neuro-fuzzy inference system for forecasting Australia's domestic low cost carrier passenger demand," *Aviation*, vol. 19, no. 3, pp. 150–163, Jul. 2015, doi: 10.3846/16487788.2015.1104806.

[287] Z. Pezeshki, · Sayyed, and M. Mazinani, "Comparison of artificial neural networks, fuzzy logic and neuro fuzzy for predicting optimization of building thermal consumption: a survey," *Artif. Intell. Rev.*, vol. 52, pp. 495–525, 2019, doi: 10.1007/s10462-018-9630-6.

[288] K. Lu and B. Han, "Congestion Risk Evaluation and Precaution of Passenger Flow in Metro Stations," *Open Civ. Eng. J.*, vol. 10, no. 1, pp. 93–104, Mar. 2016, doi: 10.2174/1874149501610010093.

[289] S. Ma, J. Weng, C. Wang, D. Alivanistos, and P. Lin, "Bus passenger flow congestion risk evaluation model based on the Pressure-State-Response framework: A case study in Beijing," *Artic. Sci. Prog. Sci. Prog.*, vol. 103, no. 1, pp. 1–14, 2020, doi: 10.1177/0036850419883567.

[290] Z. Gu, T. Osaragi, and W. Lu, "Simulating pedestrians' spatio-temporal distribution in underground spaces," *Sustain. Cities Soc.*, vol. 48, no. August 2018, p. 101552, 2019, doi: 10.1016/j.scs.2019.101552.

[291] A. Chanen, "Deep learning for extracting word-level meaning from safety report narratives," in *ICNS 2016: Securing an Integrated CNS System to Meet Future Challenges*, 2016, doi: 10.1109/ICNSURV.2016.7486358.

[292] E. Vayena *et al.*, *How the General Data Protection Regulation changes the rules for scientific research*, no. July. 2019.

[293] R. Polig, K. Atasu, and C. Hagleitner, “Token-based dictionary pattern matching for text analytics,” in *2013 23rd International Conference on Field Programmable Logic and Applications, FPL 2013 - Proceedings*, 2013, doi: 10.1109/FPL.2013.6645535.

[294] T. Verma, R. Renu, and D. Gaur, “Tokenization and Filtering Process in RapidMiner,” *Int. J. Appl. Inf. Syst.*, vol. 7, no. 2, pp. 16–18, 2014, doi: 10.5120/ijais14-451139.

[295] H. Saif, M. Fernandez, Y. He, and H. Alani, “On stopwords, filtering and data sparsity for sentiment analysis of twitter,” in *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 2014, pp. 810–817.

[296] M. K. Dalal and M. A. Zaveri, “Automatic Text Classification: A Technical Review,” *Int. J. Comput. Appl.*, vol. 28, no. 2, pp. 37–40, 2011, doi: 10.5120/3358-4633.

[297] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text classification algorithms: A survey,” *Information (Switzerland)*, vol. 10, no. 4. MDPI AG, p. 150, 23-Apr-2019, doi: 10.3390/info10040150.

[298] Algorithmia, “2021 Enterprise Trends in Machine Learning,” 2020. [Online]. Available: tinyurl.com/ml-trends-web. [Accessed: 09-Apr-2022].