

**PROVIDING DECISION SUPPORT FOR TRANSPORT
INFRASTRUCTURE MAINTENANCE PLANNING:**

THROUGH APPLICATION OF MULTI-CRITERIA AND MACHINE LEARNING METHODS

Zaharah Allah Bukhsh

PROVIDING DECISION SUPPORT FOR TRANSPORT INFRASTRUCTURE MAINTENANCE PLANNING:

THROUGH APPLICATION OF MULTI-CRITERIA AND MACHINE LEARNING METHODS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. T.T.M. Palstra
on account of the decision of the Doctorate Board
to be publicly defended
on Thursday the 12th of September 2019 at 12:45 hours

by

Zaharah Allah Bukhsh
born on the 16th of November 1992
in Bahawalnagar, Pakistan

This dissertation has been approved by:

prof. dr. ir. A. G. Doree University of Twente, Promoter

Dr. I. Stipanovic University of Twente, Assistant Promoter

Cover design: Zaharah Allah Bukhsh

Cover credits: macrovector-official from Freepik and thenounproject.com

Printed by: Ipskamp

ISBN: 978-90-365-4822-9

DOI: 10.3990/1.9789036548229

© 2019 The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

GRADUATION COMMITTEE:

Chairman	Prof. dr. G. P. M. R. Dewulf, University of Twente
Supervisor	prof.dr. ir. A. G. Doree, University of Twente
Co-supervisor	Dr. I. Stipanovic, University of Twente,
Members	Prof. dr. A. Ceric, University of Zagreb Prof K. G. Gavin, Delft University of Technology Prof. dr. ir. L. Volker , University of Twente Prof. dr. ir. T. Tinga, University of Twente Dr. ir. M. van Sinderen, University of Twente

Summary

Functional and serviceable transport infrastructure presents one of the essential predispositions for the economic growth of a country. The importance of maintaining transport infrastructure is increasingly recognized as we witness the aging of infrastructure, an increase in the frequency of extreme weather events, expanding performance demands, and shrinking financial resources. Under these circumstances, transportation agencies are facing competing demands to optimally spend the limited budget and satisfy various performance requirements related to reliability of assets, safety of users, availability of the network and impact on the environment. The multiple performance requirements of infrastructure give rise to several decision-making dilemmas.

Typically infrastructure managers analyze the condition data, estimate the future performance of assets, and decide on the maintenance actions implicitly based on their technical knowledge, past experiences, and judgments. However, due to varying personnel knowledge, cognition capacity and experiences, the intuition-based decision-making suffers from inconsistency (e.g., different decisions for the same scenarios), and distortion (e.g., over and under emphasis of specific attributes, such as cost and condition states). This results in the decisions which are difficult to follow, justify, and reproduce in the future. The implicit decision-making by asset managers can also be attributed to several related factors such as the single-objective optimization methods, poor integration of qualitative data and preferences of experts, distributed asset management systems, and data accessibility challenges.

Aligned within the focus of two European projects, namely DESTination RAIL and COST ACTION TU1406, the objective of this research is to improve the decision-making process of maintenance planning by developing applied decision support methods and predictive models to aid transport infrastructure managers. The developed data-driven decision support methods firstly enabled the optimal maintenance planning of assets over the multi-year period, and secondly used the data from asset

management systems for predictive modeling of unseen future events. The proposed approaches explicate the implicit reasoning of experts and pave a way forwards towards evidence-based asset maintenance solutions. The methodological developments of the thesis are highlighted below:

1. A multi-attribute utility theory (MAUT) method is developed to accommodate multiple objectives of maintenance for all the assets of the network. The proposed methodology also introduces a procedure to quantify the objectives in the form of performance indicators. Additionally, the model transforms the subjective preferences of a decision-maker into objective values in the form of utility functions, and performs trade-offs among multiple performance attributes. The resulting prioritization of assets directs the investment decisions of maintenance for assets of the network. (*Chapter 2 and 3*)
2. The MAUT method is further extended into a holistic computational framework that aims to find the best time to maintain an asset under the budget and performance constraints. The framework seeks to develop an optimal multi-year maintenance plan by synthesizing the type of maintenance treatment; estimating the future performance of assets by Markov chain processes, and utilizing the genetic algorithm for optimization. The proposed approach enables asset managers to simulate various maintenance planning scenarios under different budget and performance requirements. (*Chapter 4*)
3. With the objective to make the maintenance planning procedure smarter by using the asset management data, the predictive models are developed using tree-based (machine learning) classification techniques. The classifiers accurately identify correct maintenance notifications and treatment type through modeling the historical data of unplanned maintenance triggers of railway switches. Besides, the feature importance analysis of predictive models shows the essential data attributes and also reveals intrinsic decision logic. (*Chapter 5*)
4. The large amount of historical data deemed as big data is processed for the predictive maintenance of road bridges by developing a deep neural networks with entity embeddings. These models can extract insights and have shown to learn complex non-linear features from the inspections and damages data of the bridges. In addition to discrete models, a unified model utilizing the multi-task neural network is developed which jointly learns to solve multiple tasks through utilizing shared embedding and task-specific layers. The introduced deep models can be used for the transfer learning to gain performance improvements on tasks in a low-data regime. (*Chapter 6*)

This paper-based thesis addresses the challenges of maintenance planning by proposing multi-criteria methods and machine learning models. The proposed multi-criteria methods reduce the preferences of experts into objective data, establish the ranking of discrete assets and create multi-year maintenance plans to facilitate asset managers in deciding which assets to maintain, when to maintain them and what are the consequences of delaying maintenance in terms of budget and performance of assets. The developed predictive models learn from the historical asset management data and facilitate in maintenance planning through predicting the (future) condition states, risk levels, need of maintenance for assets.

This research has made progress towards more consistent, explicit, and evidence-based maintenance planning approaches, which makes the decision processes concrete, transparent, and reproducible. The suggested methods specifically concentrated on providing support to infrastructure managers; therefore, the usefulness of the proposed approaches are validated on the real datasets of highway bridges and railway switches. Moreover, where it was possible, the digital tool and code are provided to motivate the implementation of the methods in practice. Finally, these methods eliminate the gap between the appropriate use of historical data and implicit judgment-driven decision-making of experts and pave a way forward towards data-driven resources efficient asset management practices.

Contents

Summary	vii
List of Figures	xv
List of Tables	xvii
Record of published papers	xix
1 Introduction	1
1.1 An overview of transport infrastructure maintenance	1
1.2 Theoretical research background	6
1.3 Problem statement	12
1.4 Research questions	14
1.5 Research design	15
1.6 Outline of the thesis	17
1.7 References	22
2 Evaluation and Application of AHP, MAUT & ELECTRE for Infrastructure management	29
2.1 Introduction	30
2.2 MCDA methods in decision-making of infrastructure maintenance . . .	31
2.3 Evaluation scale of multi-criteria decision analysis (MCDA) methods .	32
2.4 Case study	33
2.5 Results and Discussion	41
2.6 Conclusions	44
2.7 References	46
3 Network level bridges maintenance planning using Multi-Attribute Utility Theory	49
3.1 Introduction	50
3.2 Multi-Criteria Decision Analysis for Maintenance Planning	52
3.3 Object to Network-Level Maintenance Planning	57

3.4	Bridges Maintenance Planning	59
3.5	Multi-attribute Utility Analysis for maintenance planning - Case Study	65
3.6	Discussion and Conclusions	75
3.7	References	77
4	Multi-year maintenance planning framework using MAUT & genetic algorithms	81
4.1	Introduction	82
4.2	Methodology of Multi-year Maintenance Planning Framework	84
4.3	Case Study	95
4.4	Framework Implementation and Numerical Results	100
4.5	Discussion and Conclusions	105
4.6	References	107
5	Predictive maintenance using tree-based classification techniques: A case of railway switches	111
5.1	Introduction	112
5.2	Overview of maintenance requests and data sources	115
5.3	Tree-based classification algorithms	121
5.4	Maintenance classification	125
5.5	Results	132
5.6	Interpretability	138
5.7	Discussion	144
5.8	Conclusions	145
5.9	References	147
6	Predictive maintenance planning of bridges using deep neural networks	151
6.1	Introduction	152
6.2	Related work	155
6.3	Problem domain and data description	156
6.4	Methodology for prediction of maintenance related tasks	165
6.5	Results	173
6.6	Interpretability of models	181
6.7	Discussion	184
6.8	Conclusions	186
6.9	References	188
7	Complementary Works	193
7.1	Research tasks performed in DESTination RAIL project	193
7.2	Research tasks performed in COST ACTION TU1406	198

7.3	Industry Collaborations	201
7.4	Summary	203
8	Conclusions	205
8.1	Overview of research activities and outcomes	206
8.2	Summary of appended chapters	208
8.3	Theoretical contributions	215
8.4	Practical contributions	218
8.5	Recommendations	219
8.6	Critical reflection and future research agenda	221
8.7	Closing remarks	224
8.8	References	226
A	Robustness of the multi-attribute utility model for bridge maintenance planning	229
	List of Acronyms	241
	Acknowledgement	242
	About the Author	245

List of Figures

1.1	Key steps of condition-based maintenance (Jardine, Lin, & Banjevic, 2006)	3
1.2	Factors affecting the infrastructure maintenance	5
1.3	Design of this research	16
1.4	Outline of this research	19
1.5	Timeline of this PhD.	21
2.1	Lottery setup to discern the CE of owner cost	38
3.1	Lottery setup to discern the CE of bridge condition index	67
3.2	Utility plot of condition index	68
3.3	Lottery setup to discern the CE of owner cost	68
3.4	Utility plot of owner cost	69
3.5	Lottery setup to discern the CE of user delay cost	70
3.6	Utility plot of user delay cost	70
3.7	Lottery setup to discern the CE of environmental cost	70
3.8	Utility plot of environmental cost	71
3.9	Ranking of the bridges for maintenance planning	74
4.1	A flowchart of multi-year maintenance planning framework (MMPF). . .	85
4.2	Flowchart of simple genetic algorithm.	94
4.3	Overview of BCI in case study data.	99
4.4	Range of computed maintenance cost.	99
4.5	Range of computed user delay cost.	100
4.6	Top 40 ranked bridges by MAUT.	102
4.7	All the maintenance plans generated by GAs.	103
5.1	Maintenance Request Process (MRP).	115
5.2	Condition states of switches with respect to age.	118
5.3	Ratio of switches and notifications generated with respect to age ranges. .	119
5.4	Number of notifications generated in each year.	119
5.5	Most frequently identified problems.	120
5.6	Confusion matrices of maintenance need prediction on held-out test set. .	133

5.7	Confusion matrices of maintenance need prediction with SCV.	134
5.8	Confusion matrices of maintenance type prediction with held-out test. .	135
5.9	Confusion matrices of maintenance type prediction with SCV.	136
5.10	Confusion matrices of maintenance trigger’s prediction with held-out. . .	137
5.11	Confusion matrices of maintenance trigger’s status prediction with SCV.	138
5.12	Features importance of a RF model for maintenance need prediction. . .	140
5.13	Features importance of a RF model for maintenance type prediction. . .	140
5.14	Features importance of a RF model for maintenance triggers’ status. . . .	141
5.15	LIME explanation of maintenance need prediction model.	142
5.16	LIME explanation of maintenance treatment type prediction model. . . .	142
5.17	LIME explanation of maintenance triggers prediction model.	143
6.1	Process of Inspection to Maintenance Advice (IMA).	157
6.2	Age range of the bridges in years.	161
6.3	Frequently identified damage causes on elements & components of bridges.	162
6.4	Frequently identified damages types on elements & components of bridges.	162
6.5	Representation of condition state classes in the overall data set.	163
6.6	Representation of risk level classes in the overall data set.	163
6.7	Representation of maintenance advice classes in the overall data set. . .	164
6.8	Conceptual representation of tree-based learning techniques.	167
6.9	Architecture of Neural Network with Entity Embeddings (NN-EE).	169
6.10	Multi-task neural networks with entity embedding and two shared layers.	170
6.11	Confusion matrices of condition state prediction with SRS on under- sampled test set.	175
6.12	Confusion matrices of risk level prediction with SRS on under-sampled test set.	177
6.13	Confusion matrices of maintenance advice prediction with SRS on under- sampled test set.	179
6.14	Confusion matrices of all prediction tasks with SRS on test set by MTL- NN(cw) model.	180
6.15	Instance-level explanation of NN-EE(cw) model for condition state pre- diction using the LIME framework	182
6.16	Instance-level explanation of NN-EE(cw) model for risk level prediction using the LIME framework.	183
6.17	Instance-level explanation of NN-EE(cw) model for the maintenance advice prediction using the LIME framework	183
7.1	An example of an information model	195
7.2	Proof of concept of Decision Support Tool (DST) with a case study	197

List of Tables

2.1	Data of twenty-two bridges	33
2.2	Pairwise comparison of criteria to derive their relative importance	35
2.3	Value of Single Utility Function of Attributes	38
2.4	ELECTRE III thresholds values	41
2.5	Ranking of twenty-two bridges computed by AHP, MAUT and ELECTRE III	42
3.1	Steps to apply MAUT on Maintenance Planning Problem	55
3.2	Performance goals quantified by performance indicators and quantification process	58
3.3	Weighted score of bridge elements	62
3.4	Environmental effect categories and shadow prices (TNO-MEP, 2004)	65
3.5	Relative rating of attributes (k)	72
3.6	Ranking of bridges based on utility (minimization) function	74
4.1	Heuristics defining maintenance treatments based on condition state ranges (ERA-NET ROAD, 2012, p. D03).	88
4.2	Weighted score of bridge components (Allah Bukhsh, Stipanovic, Klanker, O'Connor, & Doree, 2018).	89
4.3	Transition probability matrices of each bridge component elicited from inspection data of case study.	97
4.4	System level transition probability matrix of bridges.	98
4.5	Most optimal maintenance plan of 28 bridges out of total 869 using MMPF.	104
5.1	Description of dataset	117
5.2	Class representation for maintenance need prediction.	127
5.3	Undersampling the major classes for maintenance activity prediction.	128
5.4	Undersampling the major classes for maintenance trigger's status prediction.	128
5.5	Confusion matrix for binary classification problem.	130
5.6	Results of maintenance need prediction with held-out test set.	133
5.7	Results of maintenance need prediction with SCV.	134

5.8	Result of maintenance type prediction with held-out test set.	135
5.9	Results of maintenance type prediction with SCV.	136
5.10	Result of maintenance trigger's status prediction with held-out test. . . .	137
5.11	Results of maintenance trigger's status prediction with SCV.	138
6.1	Example of decomposition of road network assets (van der Velde, Klatter, & Bakker, 2013)	156
6.2	Selected features set for developing predictive models	161
6.3	Confusion matrix for binary classification problem.	172
6.4	Results of condition state prediction with SRS on complete and under- sampled test set.	174
6.5	Results of condition state prediction with SCV on complete and under- sampled test set.	174
6.6	Results of risk level prediction with SRS on the complete and under- sampled test sets.	176
6.7	Results of risk level prediction with SCV on complete and under-sampled test sets.	176
6.8	Results of maintenance advice prediction with SRS on complete and under-sampled test set.	178
6.9	Results of maintenance advice prediction with SCV on complete and under-sampled test set.	178
6.10	Results of Multi-task learning (MTL-NN) with SRS on complete test set. .	180
8.1	Overview of research activities and outcomes	206

Record of published papers

Double peer-reviewed articles included in this thesis

Chapter 2 Allah Bukhsh, Z., Stipanovic, I., Hartmaan, A., Klanker, G. (2018). Evaluation and Application of AHP, MAUT and ELECTRE for Infrastructure management. IALCEE 2018. **Published** in: *Life Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision: Proceedings of the Sixth International Symposium on Life-Cycle Civil Engineering*. Ed. Caspeepele, R. Taerwe, L., Frangopol, M. D.

Chapter 3 Allah Bukhsh, Z., Stipanovic, I., Klanker, G., O'Connor, A., & Doree, A. G. (2018). Network level bridges maintenance planning using Multi-Attribute Utility Theory. **Published** in: *Structure and infrastructure engineering*, 1-14.

Chapter 4 Allah Bukhsh, Z., Stipanovic, I., A., & Doree, A. G. (2019). Multi-year maintenance planning framework using multi-attribute utility theory and genetic algorithms. *Under review*

Chapter 5 Allah Bukhsh, Z., Saeed, A., Stipanovic, I., & Doree, A. G. (2019). Predictive maintenance using tree-based classification techniques: A case of railway switches. **Published** in *Transportation Research Part C: Emerging Technologies*, 101, 35-54.

Chapter 6 Allah Bukhsh, Z., Stipanovic, I., Saeed, A., & Doree, A. G. Predictive maintenance planning of bridges using deep neural networks. *Under review*

Appendix A Allah Bukhsh, Z., Stipanovic, I., Palic, S. S., & Klanker, G. (2018). Robustness of the Multi-Attribute Utility Model for Bridge Maintenance Planning. **Published** in: *The Baltic Journal of Road and Bridge Engineering*, 13(4), 404-415.

Scientific papers in international conferences and workshops

1. Skaric S., Stipanovic, I., **Allah Bukhsh, Z.**, Sajna., A.,(2019). Different bridge maintenance strategies and life cycle costs: Comparison of costs in Croatia, Slovenia and the Netherlands. Symposium Towards a Resilient Built Environment - Risk and Asset Management. IABSE, 2019, Guimarães, Portugal
2. Skaric S., **Allah Bukhsh, Z.**, Stipanovic, I.,(2018). Economy and societal performance indicators for bridge maintenance prioritization. In TU1406 COST ACTION Workshop, Barcelona, Spain
3. **Allah Bukhsh, Z.**, Stipanovic, I., Gavin, K., Doree, A. G. (2018). Evolution of Decision Support Systems for Railway Infrastructure Managers. Proceedings of 7th Transport Research Arena TRA 2018. Vienna, Austria
4. **Allah Bukhsh, Z.**, Saeed A., Stipanovic, I. (2018). A machine learning approach for maintenance prediction of railway assets. Proceedings of 7th Transport Research Arena TRA 2018. Vienna, Austria
5. **Allah Bukhsh, Z.**, Oslakovic, I. S., Klanker, G., Hoj, N. P., Imam, B., Xenidis, Y. (2017). Multicriteria decision making: AHP method applied for network bridge prioritization. In Joint COST TU1402–COST TU1406–IABSE WC1 Workshop: The value of Structural Health Monitoring for the reliable Bridge Management, Zagreb, Croatia
6. **Allah Bukhsh, Z.**, Hartmann, T., Stipanovic I., (2016). From Analysis of Information Needs towards an Information Model of Railway Infrastructure. In 23rd International Workshop of the European Group for Intelligent Computing in Engineering, EG-ICE 2016, Krakow, Poland

Articles not included in this PhD

1. Bukhsh F.A., **Allah Bukhsh Z.**, Daneva, M., (2019) A systematic literature review on Requirement prioritization techniques and their empirical evidences. Accepted with minor revisions in Computer Standards and Interfaces.
2. **Allah Bukhsh Z.**, van Sinderen M., Sikkell K., Quartel D. (2019) How to Manage and Model Unstructured Business Processes: A Proposed List of Representational Requirements. In: Obaidat M., Cabello E. (eds) E-Business and Telecommunications. ICETE 2017. Communications in Computer and Information Science, vol 990. Springer, Cham

3. **Allah Bukhsh, Z.**, van Sinderen, M., Sikkel K., Quartel D. (2017). Understanding Modeling Requirements of Unstructured Business Processes. In 14th International Joint Conference on e-Business and Telecommunication. Madrid, Spain

Introduction

Efficient transport services and functional infrastructure are of vital importance for Europe to economically strengthen all regions and to enable social cohesion. The condition of transport infrastructure is rapidly deteriorating due to the significant lack of investment and maintenance funding since 2008 (European Commission, 2018), aging, enhanced usage, and increased frequency of extreme weather events. Transportation agencies are facing competing demands to optimally spend the limited budget and satisfy various performance requirements related to reliability of assets, safety of users, availability of the network and impact on the environment. This thesis endeavors to improve the decision-making process of maintenance planning by developing pragmatic decision support methodologies and predictive models to facilitate transport infrastructure managers. Instead of taking an exclusive approach between experience-driven decision making and model-driven solutions, we propose to systematically integrate quantitative data and qualitative reasoning of experts for optimal maintenance planning under multiple objectives constraints. Subsequently, we investigate the effective use of the asset management data from in-use business processes of agencies for the predictive maintenance modeling of unseen (future) events. The data-driven decision methods will explicate the implicit reasoning of experts and will pave a way forwards towards evidence-based asset maintenance solutions.

1.1 An overview of transport infrastructure maintenance

The importance of transport infrastructure maintenance is increasingly recognized as we witness the aging structures and the increase in extreme weather events. The focus of transportation programs is being shifted from capital investment for construction towards the maintenance and operations of the infrastructure (Gleave, 2014). It challenges the asset owners to achieve the maximum performance from existing infrastructure while guaranteeing the safety of users and the availability of the network. This section provides an overview of transport infrastructure maintenance.

First, the practices of maintenance management of assets are introduced. Next, the factor influencing the infrastructure maintenance are highlighted in detail.

1.1.1 Practices of maintenance management

The European standard defines the maintenance as ‘the combination of all *technical, administrative and managerial actions* during the life cycle of an item intended to retain it or restore it, to a state in which it can perform the required function’ (CEN EN, 2011). According to the standard definition, the maintenance is not purely a technical process, in which the repair and replacement of a structure are sufficient. Instead, maintenance has a broader spectrum with objectives (1) to improve the reliability and availability of the equipment (public structures), (2) to optimize the social and environmental impacts, and (3) to reduce the spending of owner costs without compromising on the safety of users and workers (Van Dam, Nikolic, & Lukszo, 2012). Transport agencies employ various maintenance management policies to achieve maintenance objectives. The typical maintenance policies can be broadly categorized into the following three types (Dhillon, 2002):

1. *Corrective maintenance* is a run-to-failure approach where the unplanned maintenance is undertaken if a failure is noticed and system/asset is unable to fulfill its designated purpose.
2. *Preventive maintenance* is a use-based approach in which the maintenance of a system is performed on the predefined intervals or by specific criteria (e.g., condition thresholds).
3. *Condition-based maintenance* is a performance-oriented approach in which the physical state of the system is regularly assessed to predict and diagnose the need for maintenance accurately.

Traditionally, the planned maintenance programs and unexpected system failures trigger the maintenance actions. However, both of these maintenance triggers result in increased maintenance or replacement cost. As, planned schedules (i.e., preventive policy) lead to *the more, the better syndrome* of undue maintenance (Tsang, 1995), where the maintenance is conducted far before it is needed (Tinga, 2010). The sudden system failure (i.e., corrective policy) contributes to the higher failure cost and poses the reliability, safety and availability challenges on the overall network. Condition Based Maintenance (CBM) is one of the most preferred maintenance policy with the aim to avoid sudden system failure and loss of system life attributed to early maintenance (Khan & Haddara, 2003). CBM, also referred to predictive maintenance,

enables diagnosis of impending failures and prognosis of remaining useful life, where the decisions of maintenance actions are based on the physical condition of the system (Peng, Dong, & Zuo, 2010). A CBM policy can be broadly described in four key steps (Jardine, Lin, & Banjevic, 2006), as shown in Figure 1.1.

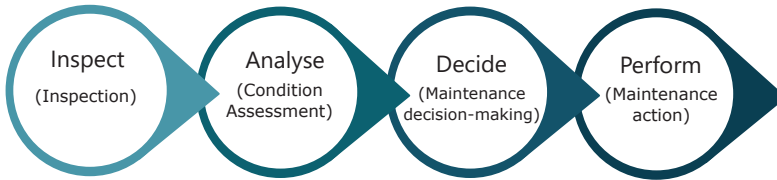


Figure 1.1: Key steps of condition-based maintenance (Jardine, Lin, & Banjevic, 2006)

The *inspect* phase covers the inspection and monitoring of a system to collect and store relevant data. The *analyse* step deals with the assessment of collected data based on absolute thresholds or probabilistic deterioration modeling. The *decide* phase involves decision-making activities which attempts to balance the costs and benefits of maintenance to develop optimal prospective plans. Finally, the *perform* phase manages the operational planning of maintenance actions. CBM policy may employ data-driven, physics-based models and knowledge-driven approaches at different phases of its implementation (Shin & Jun, 2015).

CBM policy has propelled to the forefront for crafting optimal maintenance solutions with application areas of manufacturing, transportation, and aerospace (Ahmad & Kamaruddin, 2012; Jardine et al., 2006; Peng et al., 2010; Shin & Jun, 2015). Transport infrastructure is complex systems of several assets and components having varying characteristics and lifespans. The implementation of CBM policy poses various *maintenance decision-making* challenges due to several geographically dispersed complex assets over a transportation network. For instance, imagine a scenario where the *inspect* and *analysis* steps led to the identification of many assets (e.g., bridges) that are reaching their performance limit. During the *decision* step, infrastructure managers have to make several decisions to ensure the safety and availability of the network. The example of decision questions are: *should the maintenance be performed or can it be delayed? which assets are in immediate need of maintenance? which maintenance treatment is most suitable? Is the available budget sufficient?, how does the (delayed) maintenance affect the safety and availability of network? and, what is the best time to perform the maintenance?.*

To answer these questions, the infrastructure managers analyze the condition data, estimate the future performance of assets, and decide on the maintenance actions, mostly based on their technical knowledge, past experiences, and judgments. Agencies employ different asset management systems, such as enterprise resource planning, Computerized Maintenance Management System (CMMS) and condition monitoring system to provide support in the implementation of CBM (Galar, Gustafson, Tormos Martinez, & Berges, 2012; Kans, 2013; Zhang & Karim, 2014). These systems aims to assist in the management of assets by storing data, managing work-orders, planning, and optimizing maintenance task. However, due to dispersed data across independent systems, literature reports the lack of successful implementation of these systems (Wienker, Henderson, & Volkerts, 2016) as well as their inability to support in maintenance decision-making process (Labib, 2004; Rastegari & Mobin, 2016).

Effective maintenance planning require the accurate and completeness of data to make informed cost-effective decisions. With the multiple types of transport infrastructure assets having targeted performance goals (in terms of safety, economy, society, and environment) and numerous data sources, it is intractable for infrastructure managers to systematically decide about the multiple aspects of maintenance planning based on implicit reasoning and judgments. In the following section, we briefly discuss the numerous factors influencing infrastructure maintenance and its decision-making process.

1.1.2 Factors influencing infrastructure maintenance planning

The decision-making for maintenance planning is the crucial step of infrastructure asset management. The management decisions regarding the maintenance of infrastructure can be long-term (strategic), medium-term (tactical) or short-term (operational) (Pintelon & Gelders, 1992). The *strategic planning* has to take into account long term goals which are often administrated by government regulations, and policies. *Tactical planning* addresses the problem of effective resource utilization to ensure a reliable and safe transport network. *Operational planning* manages a day-to-day scheduling, implementation, and handling of resources. The focus of this research is mainly on medium-term maintenance planning decisions. Several evitable and inevitable factors contribute to the complexity of maintenance decision-making process at all planning levels (Lidén, 2015). Figure 1.2 highlights the number of critical challenges that influence the decision-making process of maintenance planning. In the following, a brief description of each of the influencing factor is discussed.

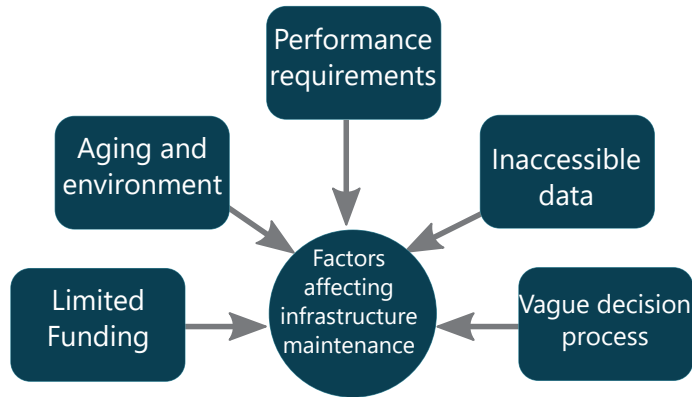


Figure 1.2: Factors affecting the infrastructure maintenance

- *Limited funding:* Many transport agencies are struggling with an insufficient budget for maintenance and operation of infrastructure since the global economic crisis in 2008. The severe budget cuts have caused irreversible damages to public assets (European Commission, 2018). The deferred maintenance has damaged the quality of roads to a level, where a maintenance cost per km of road has increased drastically. Litzka and Weninger-Vycudil (2012) noted that an increase in the annual budget of 40% is needed to maintain the current back-log level up to the year 2020.
- *Aging and environment:* Most of the transport infrastructure in Europe is constructed after the second world war (Yang & Kumaraswamy, 2011). After seventy years of service, the public structures are reaching their performance limit unless sufficient maintenance is performed. The aging and deterioration of civil structures are further escalated due to deleterious environmental events along with an increase in traffic intensity over time (Frangopol, 2011).
- *Performance requirements:* The deficiencies of transport infrastructure systems have an apparent impact on society. Thus, the government imposes Service Level Agreement (SLA) on asset owners (for an example of SLA, see Rijkswaterstaat (2013)) relating to the availability, sustainability, traffic safety, accessibility, and quality of the environment. The SLAs prompt agencies to develop maintenance plans that satisfy various performance requirements of regulators under the given budget constraints.
- *Inaccessible data:* Transport agencies use different IT-based commercial suits, e.g., Enterprise assets management, IBM MAXimo, and custom solutions, e.g.,

bridge management system, pavement management systems, and CMMS to register assets and store relevant data. These management systems, often used for specific assets types and management operations, introduce data silos with various data standards, data formats, and data granularity (Thaduri, Galar, & Kumar, 2015; Wienker et al., 2016). The data dispersed across independent systems makes it difficult to access, fuse and mine. Gualtieri (2016) reports that 70% of all collected is never used for analysis and decision-making. Resulting from qualitative interviews with asset managers, Wijnia and Herder (2010) also concluded that 'the data exist but is not accessible' for asset management tasks.

- *Vague decision-making practices*: The decision-making process that undergoes in defining when, where, what, and which asset to maintain still heavily rely on implicit reasoning based on previous experiences and expert knowledge (Ahmad & Kamaruddin, 2012; Van Noortwijk, 2009). However, due to varying personnel knowledge, cognition capacity and experiences, the intuition-based decision-making suffers from inconsistency and distortion. This results in the decisions which are difficult to follow, justify, and reproduce in the future. Additionally, without the evidence-based decision-making practices, the knowledge of experienced technicians becomes corporate assets which makes the infrastructure vulnerable in case of retiring and leaving personnel.

The task of infrastructure maintenance has become challenging due to increasing aging and deterioration, expanding performance demands, shrinking financial resources, inaccessible data and unclear decision-making approaches. The maintenance practices and highlighted factors of infrastructure maintenance have put forward the need for pragmatic and holistic decision-support methodologies that can facilitate asset managers in decision-making process of maintenance planning.

1.2 Theoretical research background

Maintenance planning of infrastructure poses several objectives related to technical, managerial and administrative aspects. Dekker (1996) defines maintenance objectives as: ensuring that the system functions (availability, efficiency and production quality); ensuring the system's life (asset management); ensuring safety; and, ensuring human well-being. The simultaneous achievement of all the maintenance objectives is not trivial job (de Almeida, Cavalcante, et al., 2015b) due to conflicting objectives, inaccessible data, and preferences uncertainties, among others. Until recently, the maintenance of transport infrastructures has been viewed solely as a

technical process, where a balance between cost and improved physical condition state is deemed sufficient (Van Dam et al., 2012). This trend is also evident in literature where numerous methods for system reliability (Jardine et al., 2006; Peng et al., 2010; Si, Wang, Hu, & Zhou, 2011; Ye & Xie, 2015) and for cost minimization and performance improvements (Frangopol, Kong, & Gharaibeh, 2001; Ghodoosi, Abu-Samra, Zeynalian, & Zayed, 2017) has been proposed.

Though valuable for detailed reliability analysis, the maintenance support methods based on single-objective criterion computes one optimal solutions which may not necessarily meet all the performance requirements of maintenance planning (Garg & Deshmukh, 2006; Liu & Frangopol, 2005; Van Dam et al., 2012). Additionally, the optimization approaches with their complicated mathematical forms do not integrate qualitative preferences of decision-makers and quantitative data systematically (Ruschel, Santos, & Loures, 2017). Multiple literature reviews have also noted a gap between theoretical progress and practical use of the maintenance optimization models (Ahmad & Kamaruddin, 2012; Ding & Kamaruddin, 2015; Ruschel et al., 2017; Sharma, Yadava, & Deshmukh, 2011). Many maintenance programs still profoundly rely on implicit reasoning of experts which is based on their judgments, past experiences and technical knowledge. Through subjective input is crucial to maintenance decision-making, they do not guarantee reliable, efficient and cost-effective solutions.

The noted challenges of infrastructure maintenance and limitations of judgment-based decision-making call for quantitative, comprehensive and transparent (followable) decision-support methods for maintenance planning tasks. Transportation agency continuously collect, and manage the large amount of operational and maintenance data. The asset management data can be used to quantify the multiple performance objectives of maintenance, where the multi-criteria methods can help to manage their conflicting nature as evident by several studies (de Almeida, Ferreira, & Cavalcante, 2015; Kabir, Sadiq, & Tesfamariam, 2014). The collected data can also be used to investigate the viability of predictive maintenance modeling using machine learning techniques. Next section provides an overview of relevant research efforts for maintenance planning based on multi-objective methods and predictive modeling. The literature overview also motivates the theoretical choices undertaken in this study.

1.2.1 Multi-objectives methods for maintenance planning

Maintenance optimization models with the principal focus on cost minimization are incapable of representing a thorough planning context and may not be suitable anymore (Ahmad & Kamaruddin, 2012; Sharma et al., 2011). A comprehensive maintenance plan must accommodate multiple performance goals (Van Dam et al., 2012). The most common types of performance goals related to maintenance planning are: improved structural performance of an asset, minimal disruptions for users, least impact on the environment, and reduced economic cost of interventions (Stipanovic et al., 2017). Due to the posed standards of environment, workers safety, and acceptable performance level, the single cost minimization objective is likely to cause a higher impact on the environment and possible risk for users and infrastructure. Various performance requirements shape maintenance planning as a complicated multi-facets problem. These objectives are often competing and conflicting in nature, where no single solution optimizes all the involved performance goals. Therefore certain trade-offs have to be made to gain the value of one performance aspect on the cost of another (Keeney & Raiffa, 1993). Based on the specific characteristics and context of decision-making, any multi-objective problem can be solved using the following approaches (de Almeida, Ferreira, & Cavalcante, 2015):

- *Multi-criteria methods* transform the multiple objectives, represented by criteria, into single representative values based on the preferences structure of decision-makers.
- *Multi-objective optimization approaches* generate numerous solutions until the Pareto optimal, i.e., non-dominant, is found. These approaches do not take into account the preferences of decision-makers.

According to recent literature reviews on maintenance and reliability research (de Almeida, Ferreira, & Cavalcante, 2015; Kabir et al., 2014; Ruschel et al., 2017), a growing trend on the application of Multi-Criteria Decision Analysis (MCDA) methods and multi-objective models for optimization of resources, strategies and intervention has been noticed. A few methods of MCDA have particularly gained attention in this regard, e.g., Pareto Front, Multi-Attribute Utility Theory (MAUT), Analytical Hierarchy Process (AHP), Multi-Attribute Value Theory (MAVT), Goal programming, ELimination and Choice Expressing REality (ELECTRE) and Technique for Order by Similarly to Ideal Solution (TOPSIS). MCDA has emerged as a decision support methods that provide an articulation procedure in order, to systematically, accommodate preferences of a decision-maker and relevant technical details. Studies from different

domains, such as environment, forestry, and water management, have also reported that the application of MCDA improves the decision-making process and are widely trusted by practitioners (Diaz-Balteiro & Romero, 2008; Hajkowicz & Collins, 2007; Huang, Keisler, & Linkov, 2011). Owing to the broad area of applications and significant variations in MCDA methodology, the selection of appropriate MCDA method pertaining to the specific needs of infrastructure management and decision maker is a problematic task (Sabaei, Erkoyuncu, & Roy, 2015). Various MCDA methods may be considered suitable for a single decision context (de Almeida, Ferreira, & Cavalcante, 2015). In the study, we begin by investigating *which methods of MCDA are most relevant to provide support in maintenance planning scenarios*.

Deterministic MCDA methods (e.g., weighted average, AHP, MAVT) are popular choices for decision-making of maintenance, in which the subjective scales are used to record the preferences of decision makers, and for each combination of alternatives and criteria, there exists a definite outcome (Huang et al., 2011). It is worth reminding that a maintenance planning problem, having multiple objectives, inherently involve uncertainty because the (set of) interventions/actions, exerted as a result of decisions, only manifest themselves in the future. The uncertainty in decision-making problem is mainly originated either due to the incomplete information or due to the fuzziness of actions and events. Few methods of MCDA such as fuzzy AHP and fuzzy TOPSIS are the extension of their primary methodologies to incorporate the fuzziness of decision context (Pedrycz, Ekel, & Parreiras, 2011). de Almeida, Cavalcante, et al. (2015a, Chapter 2) emphasized that a probabilistic approach must be used to model the uncertainty involved in maintenance planning problem. Based on expected utility theory, MAUT provides a rigorous procedure to articulate preferences and manages uncertainty by employing the notion of gambling. The literature reports many successful applications of MAUT. For instance, a model to determine inspections intervals (Ferreira, de Almeida, & Cavalcante, 2009), an additive utility model for selecting repair contracts (de Melo Brito, de Almeida Filho, & de Almeida, 2010), sustainability informed maintenance optimization of highway bridges (Sabatino, Frangopol, & Dong, 2015), allocation of tasks for reliability growth (Wilson & Quigley, 2016), and a model to determine production planning parameters (Pergher & de Almeida, 2017). The literature review on MAUT does not render any studies where a network-level multi-objective maintenance planning problem is investigated, to capture preference structure and decision uncertainty. *This study examines the application of MAUT on maintenance planning of a discrete type of assets of the network, where the multiple objectives in terms of performance, cost, user delay, and environmental impact can be optimized.*

MAUT is only practical to use when the number of alternatives is known, and criteria can take only finite possible values (Malczewski, 2006). For example, multiple assets such as bridges become alternatives and their possible condition, maintenance cost and other measures are known and finite. A maintenance plan based on MAUT can only select the assets to maintain based on the defined objective; however, it does not recommend when the maintenance should be performed. As for each alternative, the ‘when’ can take more than one value. The evolutionary algorithms such as genetic algorithms have been applied extensively in maintenance optimization problems (Bocchini & Frangopol, 2011; Denysiuk, Fernandes, Matos, Neves, & Berardinelli, 2016; Lee & Kim Sung, 2007; Morcous & Lounis, 2005; Xie, Wu, & Wang, 2018) to search for a non-dominant solution that satisfies multiple objectives. These methods search for an optimal solution within an ample solution space and provide a good quality solution within a reasonable time. However, these methods have been criticized due to their complex heuristic search methods and their inability to acknowledge preferences of decision makers (de Almeida, Ferreira, & Cavalcante, 2015). The discrete decision problem, where alternatives are finite, and the continuous decision problem, having many feasible solutions, are often categorized and handled independently. In this study, we argue that a comprehensive decision model must cover both the discrete and continuous decision context to develop optimal multi-year maintenance plan on the network-level. Since, the selection of assets for maintenance may be a discrete decision problem, whereas finding the optimal balance of cost and reliability concerning time scale is a continuous decision problem. Based on this logic, *this study investigates on how to develop an optimal multi-year network-level maintenance plan which incorporates the decision-makers’ preferences and uncertainties. The optimal strategy must find the best time for the maintenance of an asset while satisfying multiple-objectives simultaneously.*

This research endeavors of developing comprehensive maintenance planning framework to provide decision support to infrastructure managers. In addition to developing optimal maintenance plans, the suggested framework aims to equip experts to perform numerous maintenance planning scenarios under various performance requirements and budget constraints.

1.2.2 Predictive maintenance modeling

Industry 4.0 is the fourth industrial revolution attributed to current trends of the internet of things, sensor technologies, data analytics, and artificial intelligence (Wang, 2016). The focus is primarily on continuous condition monitoring of the assets for the

overall optimization of the resources and operations with minimal unplanned downtime. The idea is mainly interesting for asset-intensive industries, e.g., manufacturing units, logistics, health, and transportation.

The advancements of information and communication technologies have re-branded the CBM policy with the name of Predictive Maintenance (PdM). In its essence, PdM is similar to CBM, where the performance state of an asset drives the maintenance decisions. However, there are two critical differences between the PdM and CBM. First, instead of regular inspections, PdM proposes to perform continuous monitoring and reporting of remote assets by sensor technologies. Second, in contrast to developing physical models of structures, the (sensor) data is analyzed by artificial intelligence techniques which notify about the current state and also predicts the future condition state of assets with certain probability (Paolanti et al., 2018). By access to monitoring data, digital technologies, and artificial intelligence techniques, the ambition of PdM is to develop self-diagnostic systems which alert the assets managers in need of intervention. An example of such system is POSS condition monitoring system designed by Strukton Rail which continuously monitors the railway switches to distinguish any deviation from standard operational settings (Guzman, Hadzic, Schuil, Baars, & Groos, 2018).

Several literature studies reported the successful development of numerous predictive maintenance solutions for infrastructure maintenance. Few of the applied studies are failure prediction models using heterogeneous data (Li et al., 2014), predictive models to detect metro door failure (Manco et al., 2017), recurrent neural networks to identify and recognize the failures in railway track circuits (de Bruin, Verbert, & Babuška, 2017), remaining useful lifetime of an electrical power switch (Böhm, 2017), discovering defects of the fasteners (Chen, Liu, Wang, Núñez, & Han, 2018), an automatic prediction of intervention types for roads (Morales, Reyes, Caceres, Romero, & Benitez, 2018). Multiple PdM projects such as Infralert, Project Techlok (Kauschke, Janssen, & Schweizer, 2015), and development agenda from the European union (Frost, 2019; Niestadt, Debyser, Scordamaglia, & Pape, 2019) prove the perceived potential of PdM solutions.

Nevertheless, Tiddens, Braaksma, and Tinga (2015) noted that there is a gap between potential and realized benefits of PdM. This can be due to multiple reasons. Firstly, there is possible lack of understanding about how data analysis can help in the decision-making process. Secondly, the use of additional (sensor) monitoring devices for continuous monitoring of assets are found to be expensive and impractical for geographically distributed assets infrastructure (de Bruin et al., 2017). Thirdly, the

models developed using sophisticated algorithms of machine learning are black-boxes with little to no explanation of intrinsic prediction logic (Datta, Sen, & Zick, 2016; Doshi-Velez & Kim, 2017). Finally, there are no straightforward matrices to estimate the return on investment of PdM projects.

Transportation agencies continuously collect the data either by automated procedures, e.g., track measuring systems, or by specific business processes, such as regular visual inspection, record keeping of maintenance events, the cost spent, unplanned triggers, and so on. The asset management data can be referred as *big data* due to its large amount (volume), unstructured nature (variety), high generation rate (velocity), usefulness for business decisions (value) and its need of validation (veracity) (Lomotey & Deters, 2014). Over the time, the collected data at agencies is not only huge and dispersed across several IT systems, but it is also difficult to acquire, manage and process by traditional software and tools within a tolerable time (Galar, Kans, & Schmidt, 2016). Hence, more than 70% of collected asset management data is never used for any decision-making (Gualtieri, 2016; Wijnia & Herder, 2010). To avoid additional cost by continuous asset monitoring and to encourage the data usage for the data-driven maintenance decision-making, in this study, *we explore the viability of using available asset management data to develop predictive maintenance model by utilizing advanced machine learning techniques. Furthermore, we study how to effectively interpret the outcomes of predictive models to enable the transparency and gain the trust of infrastructure managers on the models' predictions.*

The algorithmic decision-making (i.e. decisions driven by predictive models) involves a varying level of human interventions. Therefore, it is vital to clarify that the predictive models of this study aspire to aid the experts in data-driven decision-making of maintenance planning and do not aim to develop an autonomous decision-making system.

1.3 Problem statement

The importance of transport infrastructure maintenance is increasingly being recognized due to aging infrastructure, prolonged usage, increasing deterioration of assets (Yang & Kumaraswamy, 2011), expanding performance demands, and shrinking financial resources (European Commission, 2018). Under these circumstances, the infrastructure managers are facing competing demands to optimally spend the restricted budget and satisfy various performance requirements related to the reliability of assets, the safety of users, availability of the network and the impact on the environ-

ment (Stipanovic et al., 2017). Typically, the planning of maintenance interventions and future investments mainly rely on implicit reasoning of experts which is based on their judgments, past experiences, and technical knowledge (Ahmad & Kamaruddin, 2012; Van Noortwijk, 2009). The focus on subjective analysis can be attributed to poor access and management of available data (Guler, 2012; Thaduri et al., 2015; Wijnia & Herder, 2010) and limited decision support capabilities of operating computerized systems (Rastegari & Mobin, 2016; Wienker et al., 2016). With the availability of huge amount of data dispersed across independent systems, it is beyond the cognition capacity of asset managers to retrieve appropriate data, quantify the performance requirements of a large number of assets and develop cost-effective maintenance plans.

Maintenance optimization models proposed in the literature regard the reliability of the structure and maintenance cost as an optimization criterion and do not take into account other performance requirements (Alaswad & Xiang, 2017; Bousdekis, Magoutas, Apostolou, & Mentzas, 2015; Sharma et al., 2011). Moreover, the optimization approaches having multiple mathematical formats are unable to integrate qualitative preferences of decision-makers and quantitative data systematically (Ruschel et al., 2017). Multiple literature reviews have also highlighted that the developed optimization models are unable to fulfill the functional requirements for infrastructure maintenance (Ahmad & Kamaruddin, 2012; Ding & Kamaruddin, 2015; Ruschel et al., 2017; Sharma et al., 2011). The workshops conducted during this research study with several road and railway practitioners also unveils that they lack practical decision support models for planning maintenance activities based on the existing data.

The limitations of judgment-based decision making, inaccessibility of huge amount of data, and theoretical underpinning of single-objective methods emphasizes the need to improve the decision-making process of maintenance planning. This research aspires to shift the focus of decision-making processes from implicit experience-driven approaches towards explicit data-driven methods. This study aims to develop a comprehensive computational framework that can capture the decision-makers' preferences, systematically accommodates the multiple performance objectives, and can produce an optimal multi-year maintenance plan. The framework enables experts to execute numerous planning scenarios under various performance requirements and budget limits. Furthermore, to advance the current business practices of maintenance planning with digital technologies, this thesis also investigates the viability to use a large amount of asset management data for predictive maintenance modeling. The predictive models will be able to extract useful insights from the existing data

to predict the (unseen) future events. The following section presents the guiding research questions of this study.

1.4 Research questions

The practical necessities and the theoretical underpinning has motivated the need to develop decision support methodologies that enable optimal maintenance planning by making effective use of asset management data. The overarching objective of this research is *to improve the decision-making process of maintenance planning by developing applied decision support methods and predictive models to aid transport infrastructure managers*. To achieve this goal, the following four research are formulated:

- RQ 1** Which methods of multi-criteria decision-analysis provide support in the decision-making process of the maintenance planning of transport infrastructure networks?
- RQ 2** How to prioritize the maintenance of network-wide assets by incorporating the decision-makers' preferences and satisfying multiple conflicting objectives simultaneously?
- RQ 3** How to develop multi-year maintenance plans of network-wide assets constrained by budget limitations and multiple performance requirements?
- RQ 4** How to effectively use the large amount of asset management data and improve the maintenance decision-making practices using machine learning techniques?

From the perspective of enhancing the decision-making process of maintenance planning and aiding transport infrastructure managers, this research essentially covers two themes. The first part of this study (RQs 1-3) focuses on development of methodologies that establish optimal maintenance plans for all the assets of the network and enable decision-makers to simulate numerous maintenance planning scenarios under varying performance goals. The second part of this research (RQ 4) concerns on how to perceive, acquire, manage, and process the large amount of asset management data to develop decision support models using different machine learning techniques. The predictive models aim to infer useful insights from the data and recommend appropriate maintenance actions efficiently.

1.5 Research design

The design science methodology emphasizes the link between knowledge and practice by showing that scientific knowledge can be produced by designing solutions and solving practical problems (Wieringa, 2009). Wieringa (2014) has proposed a design cycle as a research methodology to solve the *practical problems* and to answer the *knowledge questions*. This research iteratively applied the design cycle consisting of problem investigation, treatment design, and treatment validation as shown in Figure 1.3. Inspired by the design science research methodology proposed by Pefers, Tuunanen, Rothenberger, and Chatterjee (2007), Figure 1.3 also highlights the research objectives and communication as innate activities of design research.

The *problem investigation* phase concentrated on reviewing the literature of the maintenance optimization methods, the challenges of using data for decision-making, and characteristics and limitations of decision support tools utilized for maintenance planning tasks. The several interactions with road and railway infrastructure managers, in the form of project meetings, discussions, and personal interviews, also guided the investigation phase of this research. The *problem investigation* phase reveals the (1) factors influencing the infrastructure maintenance, (2) the limited decision support for multi-objective maintenance planning, (3) a missing link between the use of quantitative data and qualitative reasoning of experts and (4) inaccessibility and unusability of large amount of collected data within agencies for decision support.

The *investigation* phase establishes the *objective of this research*. The *treatment design and validation phase* examine the development and validation of proposed methods of this research. This PhD thesis approaches the problem of decision-support for maintenance planning in two related perspectives. The first three chapters (2-4) of the thesis propose the multi-objectives decision-support methodologies that generate optimal maintenance plans and enable decision-makers to simulate numerous planning scenarios under varying performance requirements. The proposed methods are validated by a real case study of road bridges. The second half of this research (chapter 5-6) concerns with the acquiring, managing and processing the large amount of asset management data to develop predictive maintenance models. These models infer useful insights from the data and recommend appropriate maintenance actions efficiently. The predictive maintenance tasks are modeled and validated on the data of railway switches and road bridges.

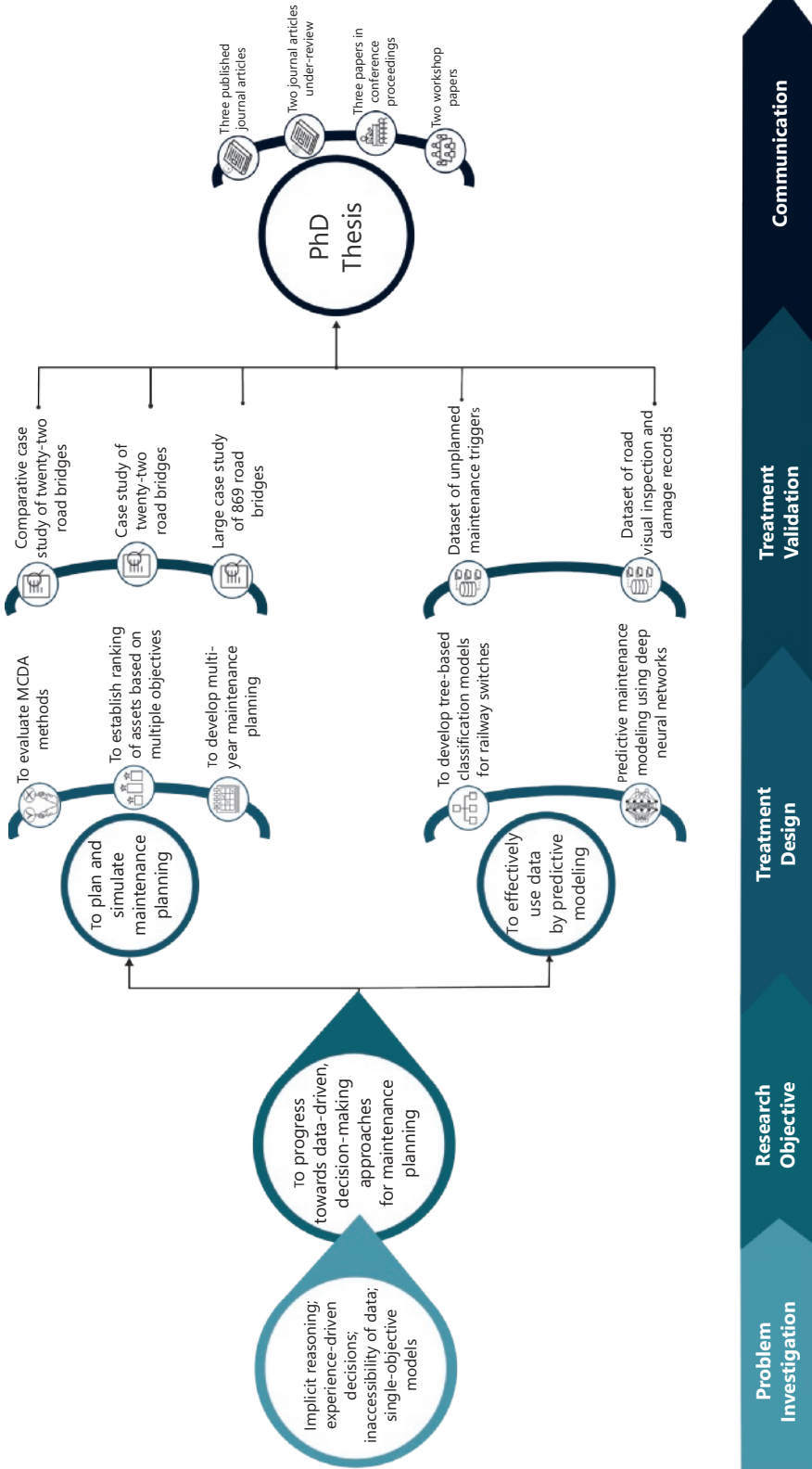


Figure 1.3: Design of this research

The predictive modeling aspires towards robust long-term maintenance planning, by predicting the maintenance needs of infrastructure ahead of failure. The data modeling approach also regulates the historical data and makes the decision-making process reliable, and consistent.

The research developments of this thesis are published in several peer-reviewed journals and conferences proceedings. The proposed data-driven decision methods were also disseminated at several workshops and project meetings to receive the feedback from the road and railway infrastructure managers for design improvements. To summarize, the research design of this study follows the design cycle methodology in order to develop several decision support approaches for maintenance planning.

The implementation of the design cycle followed an iterative approach to answer the research questions, defined in Section 1.4. The research outcomes offer several practical contributions such as support in establishing the ranking of assets based on multiple objectives; developing multi-year maintenance plans; simulating the numerous planning scenarios; estimating maintenance budget for financial planning; acquiring, managing and processing the different asset management datasets (related to (related to historical inspection, performance and maintenance) from agencies; predicting the maintenance needs of assets by using predictive models; forecasting the performance levels through deep neural networks; and finally bridging the gap between the use of data for decision-making tasks of maintenance planning. Altogether, this research provides several key insights into the decision-making process of maintenance planning, along with directions of process improvements by proposing several data-driven decision support methods.

1.6 Outline of the thesis

This thesis is composed of one conference and five journal articles that have been either published or submitted for review to peer-reviewed scientific journals. Each academic article describes the development and validation of a specific method that addresses the research questions. This thesis follows the paper-based format of thesis writing, to comply with the rules of the University of Twente. Thus, the next five chapters represent the four published and two under-review articles in their original form and can be read independently of each other.

The paper-based thesis format offers several advantages in the form of developing candidates writing, presentation, and publications skills. Moreover, the peer-review

process of journals significantly improves the quality of research and the final publication validates the credibility of work and disseminates the research outputs to a broader audience. However, there are few drawbacks of paper-based thesis writing approach. Since the academic articles have to be self-contained, the chapters of the thesis may introduce different story-lines which may not come together as a single unified narrative. Because of this, the chapters may not have a logical connection between them which can subsequently impact the readability of the overall thesis. Moreover, the individual articles presented as a chapter may contain redundancies, varying level of descriptions, and slightly different terminologies. Nevertheless, the paper-based thesis is increasingly adopted format by many universities. The overall structure of this research with respect to research questions and corresponding articles is presented in Figure 1.4.

The thesis is structured as follows. Chapter 2 evaluates the two synthesis-based methods and an outranking method methods of MCDA to investigate their support for multi-objective maintenance decision-making problem. The evaluation firstly illustrates how different MCDA methods can be applied for the prioritization of a selected number of assets and secondly provides a recommendation which method of MCDA is most useful considering the number of usability criteria.

Chapter 3 and 4 describe the method and framework for the optimal multi-year maintenance planning of several assets under the performance objectives. In particular, Chapter 3 introduces a multi-attribute utility method which seeks to optimize multiple performance objectives while accounting for involved uncertainty and risk attitudes of infrastructure managers. The model generates prioritization of alternatives which suggests that an object with higher rank contributes the most in the realization of defined performance goals. The model is further extended into a comprehensive decision-support framework. Chapter 4 introduces the multi-year maintenance planning framework to find the best time to maintain an asset under the budget and performance constrains. The introduced framework supports the infrastructure managers in the decision-making process of maintenance planning by enabling them to simulate different maintenance planning scenarios. The proposed solution approach is particularly useful for those transport agencies that mainly rely only on condition states of the assets for the maintenance decision-making.

Chapter 5 and Chapter 6 report the procedure of acquiring, managing and processing the huge amount of asset management data from agencies for development of predictive maintenance modeling using machine learning algorithms. Chapter 5 introduces the tree-based classification models for railway switches and crossings.

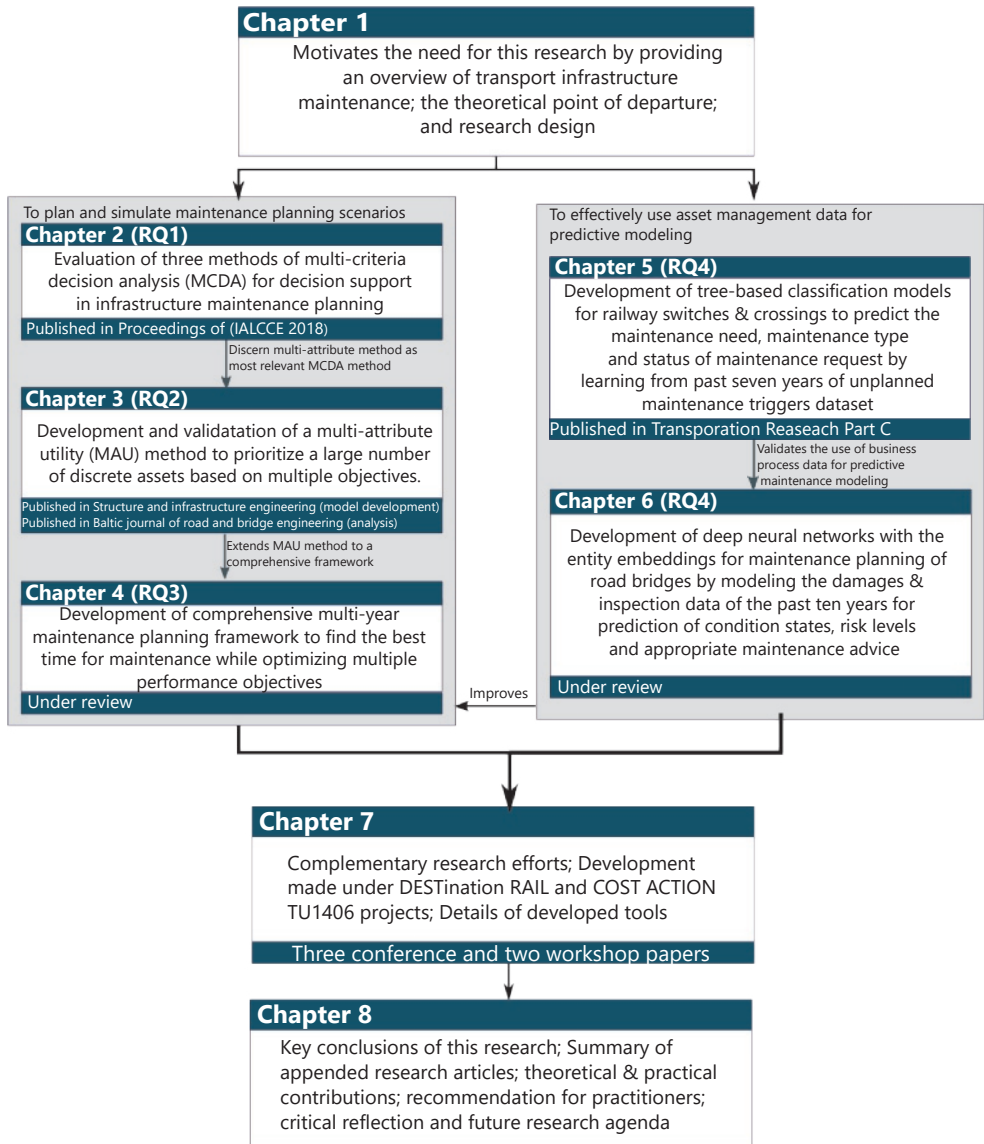


Figure 1.4: Outline of this research

The classifiers accurately identify correct maintenance notifications and treatment types by modeling the historical data of unplanned maintenance triggers. Chapter 6 further extends the application of artificial intelligence techniques to manage big data of assets. This chapter reports the development of a deep neural network with entity embeddings that predicts several related tasks for maintenance of road bridges. The multi-task neural network is also developed which jointly learns to solve multiple tasks through utilizing shared embedding and task-specific layers. The objective of employing machine learning is to effectively use data from in-use business process and develop predictive models that can be readily used as a decision aid by asset managers. With the intention to deploy the models in real decision scenarios, our work especially takes into account the interpretability of the models' outcome.

Following the five chapters that answer the research questions of this study, Chapter 7 outlines the additional research efforts that took place during this PhD trajectory. The complementary work indirectly contributed to the research objectives and provided a strong foundation for the work presented in Chapters 2-6. Finally, Chapter 8 concludes this study by reflecting on fundamental research activities and outcomes. It also provides a summary of the appended papers, along with the practical and theoretical implications of the research. The research claims are placed in perspective by outlining several recommendations for practitioners. Finally, it reflects on the limitations of this work along with setting the agenda for future research.

In the end, schematical overview of all the research activities conducted during the PhD trajectory is presented in the Figure 1.5. It provides the timeline of the PhD project containing an overview of attended conferences and project meetings along with time-frame of different articles submission and acceptance. The details of followed courses summed up to 30ECs are omitted from this overview.

1.7 References

- Ahmad, R., & Kamaruddin, S. (2012). A review of condition-based maintenance decision-making. *European Journal of Industrial Engineering*, 6(5), 519–541. PMID: 48854. eprint: <https://www.inderscienceonline.com/doi/pdf/10.1504/EJIE.2012.048854>
- Alaswad, S., & Xiang, Y. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157, 54–63.
- Bocchini, P., & Frangopol, D. M. (2011). A probabilistic computational framework for bridge network optimal maintenance scheduling. *Reliability Engineering & System Safety*, 96(2), 332–349.
- Böhm, T. (2017). Remaining useful life prediction for railway switch engines using classification techniques. *International journal of prognostics and health management*, 59.
- Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2015). A proactive decision making framework for condition-based maintenance. *Industrial Management & Data Systems*, 115(7), 1225–1250.
- CEN EN. (2011). *13306: Maintenance terminology*. Technical report, European Committee for Standardization (CEN).
- Chen, J., Liu, Z., Wang, H., Núñez, A., & Han, Z. (2018). Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Transactions on Instrumentation and Measurement*, 67(2), 257–269.
- Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *Security and privacy (sp)* (pp. 598–617). IEEE.
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., Garcez, T. V., et al. (2015a). *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis*. New York, NY, USA: Springer, 2015.
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., & Garcez, T. V. (2015b). Basic concepts on risk analysis, reliability and maintenance. In *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis* (pp. 89–160). Springer.
- de Almeida, A. T., Ferreira, R. J. P., & Cavalcante, C. A. V. (2015). A review of the use of multicriteria and multi-objective models in maintenance and reliability. *IMA Journal of Management Mathematics*, 26(3), 249–271.

- de Bruin, T., Verbert, K., & Babuška, R. (2017). Railway track circuit fault diagnosis using recurrent neural networks. *IEEE transactions on neural networks and learning systems*, 28(3), 523–533.
- de Melo Brito, A. J., de Almeida Filho, A. T., & de Almeida, A. T. (2010). Multi-criteria decision model for selecting repair contracts by applying utility theory and variable interdependent parameters. *IMA Journal of Management Mathematics*, 21(4), 349–361.
- Dekker, R. (1996). Applications of maintenance optimization models: A review and analysis. *Reliability engineering & system safety*, 51(3), 229–240.
- Denysiuk, R., Fernandes, J., Matos, J. C., Neves, L. C., & Berardinelli, U. (2016). A computational framework for infrastructure asset maintenance scheduling. *Structural Engineering International*, 26(2), 94–102.
- Dhillon, B. S. (2002). *Engineering maintenance: A modern approach*. cRc press.
- Diaz-Balteiro, L., & Romero, C. (2008). Making forestry decisions with multiple criteria: A review and an assessment. *Forest Ecology and Management*, 255(8-9), 3222–3241.
- Ding, S.-H., & Kamaruddin, S. (2015). Maintenance policy optimization—literature review and directions. *The International Journal of Advanced Manufacturing Technology*, 76(5-8), 1263–1283.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*.
- European Commission. (2018). Transport in the european union, current trends and issues. 'Mobility and Transport'. Available on: <https://bit.ly/2BuDRpy>.
- Ferreira, R., de Almeida, A. T., & Cavalcante, C. A. (2009). A multi-criteria decision model to determine inspection intervals of condition monitoring based on delay time analysis. *Reliability Engineering & System Safety*, 94(5), 905–912.
- Frangopol, D. M. (2011). Life-cycle performance, management, and optimisation of structural systems under uncertainty: Accomplishments and challenges 1. *Structure and Infrastructure Engineering*, 7(6), 389–413.
- Frangopol, D. M., Kong, J. S., & Gharaibeh, E. S. (2001). Reliability-based life-cycle management of highway bridges. *Journal of computing in civil engineering*, 15(1), 27–34.
- Frost, A. (2019). European commission to invest us475m in 69 new transportation projects. *Traffic Technology Today*.
- Galar, D., Gustafson, A., Tormos Martinez, B. V., & Berges, L. (2012). Maintenance decision making based on different types of data fusion. *Eksplatacja i Niezawodnos-Maintenance and Reliability*, 14(2), 135–144.
- Galar, D., Kans, M., & Schmidt, B. (2016). Big data in asset management: Knowledge discovery in asset data by the means of data mining. In *Proceedings of the 10th*

- world congress on engineering asset management (wceam 2015)* (pp. 161–171). Springer.
- Garg, A., & Deshmukh, S. (2006). Maintenance management: Literature review and directions. *Journal of quality in maintenance engineering*, 12(3), 205–238.
- Ghodoosi, F., Abu-Samra, S., Zeynalian, M., & Zayed, T. (2017). Maintenance cost optimization for bridge structures using system reliability analysis and genetic algorithms. *Journal of Construction Engineering and Management*, 144(2), 04017116.
- Gleave, S. D. (2014). Eu road surfaces: Economic and safety impact of the lack of regular road maintenance, study. *Policy Department Structural and Cohesion Policies*.
- Gualtieri, M. (2016). Hadoop is data's darling for a reason. Available: <https://go.forrester.com/blogs/hadoop-is-datas-darling-for-a-reason/>. [Accessed: 08-Apr-2018].
- Guler, H. (2012). Decision support system for railway track maintenance and renewal management. *Journal of Computing in Civil Engineering*, 27(3), 292–306.
- Guzman, D. N., Hadzic, E., Schuil, R., Baars, E., & Groos, J. C. (2018). Data-driven condition now-and forecasting of railway switches for. In *Phm society european conference* (Vol. 4, 1).
- Hajkowicz, S., & Collins, K. (2007). A review of multiple criteria analysis for water resource planning and management. *Water Resources Management*, 21(9), 1553–1566.
- Huang, I., Keisler, J., & Linkov, I. (2011). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *Science of the Total Environment*, 409(19), 3578–3594.
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483–1510.
- Kabir, G., Sadiq, R., & Tesfamariam, S. (2014). A review of multi-criteria decision-making methods for infrastructure management. *Structure and Infrastructure Engineering*, 10(9), 1176–1210.
- Kans, M. (2013). It practices within maintenance from a systems perspective: Study of it utilisation within firms in sweden. *Journal of manufacturing technology management*, 24(5), 768–791.
- Kauschke, S., Janssen, F., & Schweizer, I. (2015). Advances in predictive maintenance for a railway scenario-project techlok. *Knowledge Engineering Group, University of Darmstadt*.
- Keeney, & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press.

- Khan, F. I., & Haddara, M. M. (2003). Risk-based maintenance (rbm): A quantitative approach for maintenance/inspection scheduling and planning. *Journal of loss prevention in the process industries*, 16(6), 561–573.
- Labib, A. W. (2004). A decision analysis model for maintenance policy selection using a cmms. *Journal of Quality in Maintenance Engineering*, 10(3), 191–202.
- Lee, C.-K., & Kim Sung, K. (2007). Ga-based algorithm for selecting optimal repair and rehabilitation methods for reinforced concrete (rc) bridge decks. *Automation in construction*, 16(2), 153–164.
- Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., & Hampapur, A. (2014). Improving rail network velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C: Emerging Technologies*, 45, 17–26.
- Lidén, T. (2015). Railway infrastructure maintenance—a survey of planning problems and conducted research. *Transportation Research Procedia*, 10, 574–583.
- Litzka, J., & Weninger-Vycudil, A. (2012). The effect of restricted budgets for road maintenance. *Procedia-Social and Behavioral Sciences*, 48, 484–494.
- Liu, M., & Frangopol, D. M. (2005). Multiobjective maintenance planning optimization for deteriorating bridges considering condition, safety, and life-cycle cost. *Journal of Structural Engineering*, 131(5), 833–842.
- Lomotey, R. K., & Deters, R. (2014). Towards knowledge discovery in big data. In *2014 IEEE 8th international symposium on service oriented system engineering* (pp. 181–191).
- Malczewski, J. (2006). Gis-based multicriteria decision analysis: A survey of the literature. *International journal of geographical information science*, 20(7), 703–726.
- Manco, G., Ritacco, E., Rullo, P., Gallucci, L., Astill, W., Kimber, D., & Antonelli, M. (2017). Fault detection and explanation through big data analysis on sensor streams. *Expert Systems with Applications*, 87, 141–156.
- Morales, F. J., Reyes, A., Caceres, N., Romero, L., & Benitez, F. G. (2018). Automatic prediction of maintenance intervention types in roads using machine learning and historical records. *Transportation Research Record*, 0361198118790624.
- Morcous, G., & Lounis, Z. (2005). Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction*, 14(1), 129–142.
- Niestadt, M., Debyser, A., Scordamaglia, D., & Pape, M. (2019). *Artificial intelligence in transport. current and future developments, opportunities and challenges*. European Parliamentary Research Service. European Union.
- Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018). Machine learning approach for predictive maintenance in industry 4.0. In *2018 14th IEEE/ASME international conference on mechatronic and embedded systems and applications (mesa)* (pp. 1–6). IEEE.

- Pedrycz, W., Ekel, P., & Parreiras, R. (2011). *Fuzzy multicriteria decision-making: Models, methods and applications*. John Wiley & Sons.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45–77.
- Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: A review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4), 297–313.
- Pergher, I., & de Almeida, A. T. (2017). A multi-attribute decision model for setting production planning parameters. *Journal of Manufacturing Systems*, 42, 224–232.
- Pintelon, L. M., & Gelders, L. (1992). Maintenance management decision making. *European journal of operational research*, 58(3), 301–317.
- Rastegari, A., & Mobin, M. (2016). Maintenance decision making, supported by computerized maintenance management system. In *Reliability and maintainability symposium (rams), 2016 annual* (pp. 1–8). IEEE.
- Rijkswaterstaat. (2013). Summary national policy strategy for infrastructure and spatial planning-publication-government. nl.
- Ruschel, E., Santos, E., & Loures, E. (2017). Industrial maintenance decision-making: A systematic literature review. *Journal of Manufacturing Systems*, 45, 180–194.
- Sabaei, D., Erkoyuncu, J., & Roy, R. (2015). A review of multi-criteria decision making methods for enhanced maintenance delivery. *Procedia CIRP*, 37, 30–35.
- Sabatino, S., Frangopol, D. M., & Dong, Y. (2015). Sustainability-informed maintenance optimization of highway bridges considering multi-attribute utility and risk attitude. *Engineering Structures*, 102, 310–321.
- Sharma, A., Yadava, G., & Deshmukh, S. (2011). A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1), 5–25.
- Shin, J.-H., & Jun, H.-B. (2015). On condition based maintenance policy. *Journal of Computational Design and Engineering*, 2(2), 119–127.
- Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation—a review on the statistical data driven approaches. *European journal of operational research*, 213(1), 1–14.
- Stipanovic, Chatzi, Limongelli, Gavin, Bukhsh, A., & Xenidis. (2017). *WG2, Technical Report, Performance Goals for Roadway Bridges OF COST ACTION TU 1406*. Retrieved from TU1406 Website: <https://bit.ly/2Seh3NZ>.
- Thaduri, A., Galar, D., & Kumar, U. (2015). Railway assets: A potential domain for big data analytics. *Procedia Computer Science*, 53, 457–467.

- Tiddens, W. W., Braaksma, A. J. J., & Tinga, T. (2015). The adoption of prognostic technologies in maintenance decision making: A multiple case study. *Procedia CIRP*, 38, 171–176.
- Tinga, T. (2010). Application of physical failure models to enable usage and load based maintenance. *Reliability Engineering & System Safety*, 95(10), 1061–1075.
- Tsang, A. H. (1995). Condition-based maintenance: Tools and decision making. *Journal of quality in maintenance engineering*, 1(3), 3–17.
- Van Dam, K. H., Nikolic, I., & Lukszo, Z. (2012). *Agent-based modelling of socio-technical systems*. Springer Science & Business Media.
- Van Noortwijk, J. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1), 2–21.
- Wang, K. (2016). Intelligent predictive maintenance (ipdm) system–industry 4.0 scenario. *WIT Transactions on Engineering Sciences*, 113, 259–268.
- Wienker, M., Henderson, K., & Volkerts, J. (2016). The computerized maintenance management system an essential tool for world class maintenance. *Procedia Engineering*, 138, 413–420.
- Wieringa, R. (2009). Design science as nested problem solving. In *Proceedings of the 4th international conference on design science research in information systems and technology* (p. 8). ACM.
- Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Springer.
- Wijnia, Y., & Herder, P. (2010). The state of asset management in the netherlands. In *Engineering asset lifecycle management* (pp. 164–172). Springer.
- Wilson, K. J., & Quigley, J. (2016). Allocation of tasks for reliability growth using multi-attribute utility. *European Journal of Operational Research*, 255(1), 259–271.
- Xie, H. B., Wu, W.-J., & Wang, Y.-F. (2018). Life-time reliability based optimisation of bridge maintenance strategy considering lca and lcc. *Journal of Cleaner Production*, 176, 36–45.
- Yang, Y. N., & Kumaraswamy, M. M. (2011). Towards life-cycle focused infrastructure maintenance for concrete bridges. *Facilities*, 29(13-14), 577–590.
- Ye, Z.-S., & Xie, M. (2015). Stochastic modelling and analysis of degradation for highly reliable products. *Applied Stochastic Models in Business and Industry*, 31(1), 16–32.
- Zhang, L., & Karim, R. (2014). Big data mining in emaintenance: An overview. In *International workshop and congress on emaintenance: 17/06/2014-18/06/2014* (pp. 159–170). Luleå tekniska universitet.

Evaluation and Application of AHP, MAUT & ELECTRE for Infrastructure management

¹ **Abstract:** Infrastructure management renders a number of decision-making problems from assets' condition inspections to maintenance planning and resources optimization. Since management of infrastructure pertains to not only technical requirements but also societal and economic developments, these decision problems have multiple and often conflicting objectives. Various methods of Multi-Criteria Decision Analysis (MCDA) based on the decision theory and game theory are proposed to aid in decision-making problems. Owing to the wide area of applications and extensive variation in MCDA methodology, the selection of appropriate MCDA method pertaining to the specific needs of infrastructure management and decision maker is a difficult task. In this paper, two synthesis-based methods (i.e., Analytical Hierarchy Process (AHP) and Multi-Attribute Value Theory (MAUT)) and an outranking method (i.e., ELimination and Choice Expressing REality (ELECTRE III)) is applied on same maintenance decision-making problem to evaluate them for their scalability, ease of use, risk consideration, and few other aspects. The results of the evaluation suggest that a) without a computerized tool the scalability of these methods is tedious task b) only MAUT considers the risk attitude of a decision maker c) AHP and MAUT both require the data to be converted to definite scale for analysis, for instance, to Saaty scale of comparison and utility functions respectively and d) unlike other two, ELECTRE works on preference structure and yields partial pre-orders. These aforementioned results are obtained by application of AHP, MAUT, and ELECTRE III on the maintenance planning decision problem of 22 road bridges from the Netherlands road network. Despite the inherent methodology differences of these methods, the result of the case study shows a minor difference in ranking.

¹This chapter has been published as: Allah Bukhsh, Z., Stipanovic, I., Hartmaan, A., Klanker, G. (2018). Evaluation and Application of AHP, MAUT and ELECTRE for Infrastructure management. IALCEE 2018. In Life Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision: Proceedings of the Sixth International Symposium on Life-Cycle Civil Engineering. Ed. Caspeepe, R. Taerwe, L., Frangopol, M.

2.1 Introduction

The European road network consists of more than 5 million km of road, thus providing a vital link for economic competitiveness and social development (ERF, 2010). According to a report by European Union Road Foundation (ERF, 2013), roads form 71.8% of total transport modal in Europe, while rail being only 17.4%. With roads infrastructure being the backbone of nations' economy, the condition of roads infrastructure is rapidly deteriorating due to aging, increased usage, and significant lack of investment and maintenance funding since 2008 (ERF, 2013). The deteriorating infrastructure brings a higher risk of accidents, increased noise, reduced service quality, congestion, extended travel times, and increased CO_2 emission.

Due to multiple involved challenges, agencies have to deal with a number of performance requirements. This involves decision-making on maintenance scope, maintenance treatment, the future benefit of chosen maintenance treatment, acceptable service level, incurring cost, resulting user delay, and impact on the environment. Decision-making based on subjective measures (e.g., judgment, past experiences) and use of life-cycle cost analysis has yielded promising results in past where cost and condition state of an asset were the main factor for maintenance decision-making. However, these methods are usually unable to accommodate a number of performance aspects related to economy, society, environment.

Multi-Criteria Decision Analysis (MCDA) provides a systematic framework to consider and optimize a number of performance aspects based on decision-makers' preferences as well as on objective data. Among others, Patidar (2007), de Almeida, Cavalcante, et al. (2015) and Kabir, Sadiq, and Tesfamariam (2014) are excellent sources presenting the use of MCDA methods for infrastructure management and maintenance decision-making. Owing to the wide area of applications and extensive variation in MCDA methodology, the selection of appropriate MCDA method pertaining to the specific needs of infrastructure management and decision maker is a difficult task. The literature reports the use of multiple methods of MCDA for a similar decision problem. For example, de Almeida, Ferreira, and Cavalcante (2015, Table 4) presented several publications per MCDA method for the maintenance decision-making problem.

In this chapter, three methods of MCDA, namely Analytical Hierarchy Process (AHP), Multi-Attribute Utility Theory (MAUT) and ELECTRE III (Elimination and choice expressing reality) have been compared by applying them on a similar set of data for the maintenance decision-making. These methods are evaluated for their scalability,

ease of use, encoding to definite scale, risk considerations, ability to accommodate stakeholders' preferences, and understandability of results. Therefore, the objective of this work is two-fold: first to illustrate the application of various methods of MCDA for facilitating in the maintenance decision-making process and second to provide a recommendation on which method of MCDA is most useful considering the number of usability criteria.

The rest of the chapter is structured as follows: Section 2.2 provides a broader classification of MCDA methods and report few studies where MCDA methods have been applied to maintenance decision-making problem. In Section 2.3, an evaluation scale is provided to analyze the applicability of AHP, MAUT, and ELECTRE III. Section 2.4 provides the details of the case study by applying MCDA methods on real-case data. The results obtained from different MCDA methods are compared and discussed in Section 2.5, along with recommendation on methods applicability. Finally, Section 2.6 provides the conclusion of this study.

2.2 MCDA methods in decision-making of infrastructure maintenance

A literature review published by de Almeida, Ferreira, and Cavalcante (2015) shows an increasing trend on the use of MCDA methods applied on infrastructure maintenance and reliability decision problems. Few methods of MCDA have particularly gained attention in this regard e.g., Pareto Front, MAUT, AHP, Multi-Attribute Value Theory (MAVT), Goal programming, different versions of ELECTRE and Technique for Order by Similarly to Ideal Solution (TOPSIS). Considering the different solution approaches of these MCDA methods, they can be classified into three types (de Almeida, Ferreira, & Cavalcante, 2015):

- **Synthesis methods:** These are weighted aggregation methods that provide the relative ranking of all the alternatives, under considerations, based on the preference structure of the decision maker. The example of synthesis methods is AHP, MAVT, MAUT, and TOPSIS.
- **Outranking methods:** These methods seek to eliminate all the explicitly dominant alternatives. For instance, one alternative outranks another if it performs considerably well on all the attributes. The example of outranking methods are ELECTRE and Preference Ranking Organization Method for Enrichment of Evaluations (PROMETHEE).

- Interactive methods: These methods have a strong base in mathematical principles where the objective is defined in a set of targeted values. Goal programming and Pareto front are interactive methods.

As noted in Allah Bukhsh, Stipanovic, Klanker, O'Connor, and Doree (2018), interactive methods are being applied extensively in maintenance optimization problems in searching for the non-dominant solution that satisfies multiple objectives. However, interactive methods are based on complex heuristic search procedures such as particle swarm analysis, and they do not take into account the preferences of the stakeholders (de Almeida, Ferreira, & Cavalcante, 2015).

We will apply two MCDA methods belonging to synthesis class and one outranking method to the maintenance decision-making problem. The synthesis methods reduce the actual data values into certain normalization e.g., utility scores, weighted mean, to enable the comparison of heterogeneous scales of attributes such as cost in euros, delay in hours. Outranking methods take the preferences of stakeholder(s) and enable the comparison of heterogeneous scales of attributes without reducing them into certain value functions or standard scales.

2.3 Evaluation scale of multi-criteria decision analysis (MCDA) methods

In the past, several methods of MCDA have been developed for different decision-making situations. Each method of MCDA follows its procedure to assess criteria, to define the weights, to drive relative importance, and to accommodate the stakeholder(s) preferences (De Montis, De Toro, Droste-Franke, Omann, & Stagl, 2000). AHP, MAUT, and ELECTRE III are applied to the same maintenance decision-making problem. The purpose is to evaluate each of this method to determine the difference in their results as well as to examine their applicability on maintenance decision-making problems. An evaluation scale derived from (De Montis, De Toro, Droste-Franke, & Omann, 2004) and (Cinelli, Coles, & Kirwan, 2014) is provided below:

- Scalability
- Ease of Use
- Encoding to definite scale
- Uncertainty / Risk consideration

- Stakeholders' preferences
- Understandability of results

Note that, the assessment of AHP, MAUT, and ELECTRE III is performed by authors of this work by using the evaluation scale.

2.4 Case study

To illustrate how the different methods of MCDA can be applied for the maintenance decision-making, we used data of twenty-two randomly chosen bridges from the Netherlands road network. The provided data contain information of bridges' age, geometry, condition index on the element level, traffic intensity, planned maintenance activity on element level, the unit cost of chosen maintenance treatment and maintenance duration. Using this raw data, we computed condition index on overall bridge-level, owner cost incurred due to maintenance activity, user delay cost, and

Table 2.1: Data of twenty-two bridges

Bridges	Condition index	Owner cost	User delay cost	Environmental Cost
Bridge A	2.77	139.35	39.70	0.86
Bridge B	1.89	126.41	27.50	0.21
Bridge C	2.15	115.67	25.57	0.57
Bridge D	2.73	42.94	3.41	0.02
Bridge E	2.00	68.16	12.40	0.53
Bridge F	2.12	149.21	47.89	0.23
Bridge G	2.10	169.56	57.79	0.48
Bridge H	2.42	88.60	13.11	1.25
Bridge I	2.22	45.82	35.89	1.26
Bridge J	2.34	115.93	30.80	0.43
Bridge K	2.42	39.42	12.69	0.23
Bridge L	2.46	69.61	12.12	0.03
Bridge M	1.92	38.14	7.99	0.03
Bridge N	2.18	84.89	14.42	1.05
Bridge O	2.43	46.89	4.59	0.01
Bridge P	1.67	175.33	28.51	0.68
Bridge Q	2.08	161.48	55.25	0.37
Bridge R	2.30	158.89	51.04	0.22
Bridge S	2.58	65.90	8.79	0.10
Bridge T	1.96	62.22	22.83	0.42
Bridge U	2.02	84.82	25.70	0.28
Bridge V	2.34	152.60	42.91	0.27

environmental cost for each of the bridge. The computed data is provided in Table 2.1. The details of these attributes computation can be found in Section 3.4.

To compare the results of AHP, MAUT and ELECTRE, these twenty-two bridges will be ranked in a preferred order for maintenance where the objective is a) to minimize the owner cost, b) to minimize condition index (where lower value represents better condition), c) to reduce impact on road users as a result of maintenance (expressed in user delay cost) and d) to minimize environmental impact (expressed as environmental cost). It is essential to notice that these objectives are conflicting with each other e.g., to minimize the impact on users, the agency might need to use more resources, which will result in increased owner cost. Thus, a decision based on a single attribute cannot be made instead, there must be an equal representation of all the attributes in the final ranking of bridges. In the following, the brief algorithm details of each MCDA methods are provided along with their application on case study data.

2.4.1 Analytical Hierarchy Process

AHP is a group decision-making method which has been used in a wide variety of decision situations. It provides a comprehensive framework to define objectives, their quantifying criteria, and to evaluate alternative solutions. AHP performs a pairwise comparison to assign the relative importance of each criterion.

AHP Algorithm

The procedure of using AHP for decision making is outlined as follows (Saaty, 2008):

1. Identify the objectives, alternatives, and most importantly, criteria to evaluate the alternatives.
2. Perform pairwise comparison between two criteria at a time to establish the priorities among them.
3. Assign the level of importance by criteria values using the Saaty's relative scale of importance, which will convert the subjective judgments of decision makers into ratio scale.
4. Perform pairwise comparison on each criteria values.
5. Normalize the performance matrix between 0 to 1 to calculate the final weighted scale by

$$\bar{e}_{ij} = \frac{e_{ij}}{\sum_{k=1}^n e_{ik}} \quad (2.1)$$

where e_{ij} represents an element in matrix M . \bar{e}_{ij} represents an element of normalized matrix.

6. Calculate the geometric mean of the normalized matrix to compute the largest values that represent the aggregated preference.
7. Next, normalize the actual data matrix to reduce them between 0 to 1 by following Step 5.
8. Finally, multiply the geometric mean of each criterion to the normalized matrix to get the final ranking score.

AHP Application

The first step is to identify the objectives, criteria, and alternatives. In the case study, each of these aspects is established. As mentioned earlier, the objective is *to rank the alternatives in an order where owner cost, condition index, user delay cost, and environmental cost could be minimized* while the alternatives are those twenty-two randomly chosen bridges from the Netherlands road network.

In order to define, if owner cost is more important than condition index, or user delay cost is preferred over environmental cost, the pairwise comparison between criteria is performed. The pairwise comparison is based on the Saaty's fundamental scale of importance where two criteria are compared on the scale of 1 to 9, while 1 represent equal importance and 9 represents that one criterion is extremely more important than another. Table 2.2 shows the detailed steps where each criterion is compared with another to compute the geometric mean. The values of geometric mean show that owner cost is most important, followed by condition index, user delay cost, and environmental cost. Generally, the data of criteria differ in their scale and

Table 2.2: Pairwise comparison of criteria to derive their relative importance

	Relative imp. of criteria				Normalised matrix				Geometric mean
	CI	OC	UDC	EC	CI	OC	UDC	EC	total
CI	1	0.33	3	7	0.22	0.20	0.32	0.31	0.26
OC	3	1	5	9	0.67	0.60	0.54	0.40	0.54
UDC	0.33	0.20	1	5	0.07	0.12	0.10	0.22	0.12
EC	0.14	0.11	0.20	1	0.03	0.06	0.02	0.04	0.03
Sum	4.47	1.64	9.20	22	1	1	1	1	0.97

Condition Index (CI), Owner Cost (OC), User Delay Cost (UDC), Environmental Cost (EC)

magnitude. For example, owner cost could be in euros while the user delay can be computed in hours. Therefore, the normalization of data is performed to make them

comparable. In this case study, all the data value were already reduced to cost except for the condition index. After the data normalization, the final step is to perform matrix multiplication between the geometric mean of criteria and normalized data values. The final aggregated score provides the ranking of bridges, which is derived by decision makers preferences represented in the form of the geometric mean. The final result computed by AHP is provided and discussed in Section 2.5.

2.4.2 Multi Attribute Utility Theory

MAUT, proposed by Keeney and Raiffa (1993), is based on the expected utility theory, which reduces the criteria values into utility scores. In MAUT terms, the criteria are referred to as attributes. In contrast to AHP, MAUT can capture not only the preference structure of a decision maker but also the uncertainty and risk tolerance aspects. The detailed algorithm and application details on the same case study data are discussed in Chapter 3.

MAUT Algorithm

The algorithm to apply MAUT is provided as follows (Keeney & Raiffa, 1993):

1. Assuming that the objectives, criteria, and alternatives have been defined, the first step is to compute the single utility function of each attribute/criterion by the following formula

$$U_i(x_i) = A - B * e^{(\frac{-x_i}{RT})} \tag{2.2}$$

Where:

- $U_i(x_i)$ = Single utility value for attribute i of an alternative x
- A, B = Scaling constant
- e = The exponential constant i.e. 2.718
- RT is risk tolerance.

Since there exist cyclic dependency to compute A, B, and RT, the following equation can be used to compute RT value by trial and error approach

$$e^{\frac{-CE}{RT}} = 0.5 * e^{\frac{-Max(x_i)}{RT}} + 0.5 * e^{\frac{-Min(x_i)}{RT}} \tag{2.3}$$

2. Calculate the risk tolerance based on Expected Value (EV) and Certainty Equivalent (CE) where EV is median of the worst and best value of an attribute and CE is chosen by the following the principle

$$RiskAttitude = \begin{cases} Risk\ Neutral, & \text{if } EV = CE \\ Risk\ Avoiding, & \text{if } EV \geq CE \\ Risk\ Taking, & \text{if } EV < CE \end{cases}$$

3. Assign the relative importance weights k to each attribute i based on decision maker's preferences
4. Considering that the preference of decision maker for one attribute is independent to another, the aggregative utility score for each alternative is computed by using the additive form

$$U(x) = \sum_{i=1}^n k_i U_i(x_i) \quad (2.4)$$

Where:

- $U(x)$ = Multi-attribute utility of alternative x
- k = Weighting factor of each attribute i
- $U_i(x_i)$ = Single attribute utility of each attribute i for an alternative x

5. Assign the ranks based on the magnitude of the aggregated score considering the maximization of minimization function

MAUT Application

In this section, the MAUT is applied to the data of twenty-two bridges to rank them in an order where owner cost, condition index, user delay cost, and environmental cost can be kept a minimum. The first step is to compute the Single Utility Function (SUF) for each attribute. The SUF is computed based on exponential utility function (see Equation 2.2) in order to incorporate the uncertainty and risk tolerance aspects of a decision maker.

The concept of utility function and computation is inspired by lottery and gambling, where a gambler needs to take the certain risk given the equal probability to obtain the best value or worst value. Since the calculation of SUF is a computationally extensive process, the calculation details of only owner cost attribute are outlined here.

Figure 2.1 presents the lottery step to compute the CE/indifference point of owner cost. Considering the minimum value of 38.13 owner cost i across all the alternative and maximum value of 175.33, the EV (expected value) is 106 (see Table 2.1). EV value

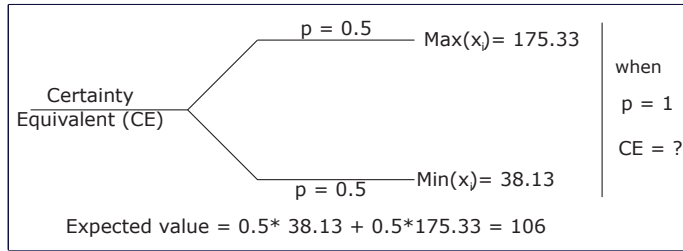


Figure 2.1: Lottery setup to discern the CE of owner cost

is used as a reference point to compute the CE, as shown in Step 2. We have assumed here that a decision maker has risk avoiding attitude, therefore the indifference point/CE value is 80. The Risk Tolerance (RT) value is 27, which is calculated by a trial-and-error approach by substituting CE, min, and max in Equation 2.3. Equation 2.2 takes the following form after computing scaling constants.

$$U_{oc}(x_{oc}) = 1.00 - 4.13 * e^{-x_{oc}/27} \quad (2.5)$$

By following the similar process, EV, CE, and RT of condition index, user delay cost, and environmental cost is calculated. Table 2.3 shows all the computed values of these attributes.

Table 2.3: Value of Single Utility Function of Attributes

Attributes	$Min(x_i)$	$Max(x_i)$	EV	CE	RT
Condition Index	1.67	2.77	2.22	1.70	0.7
Owner Cost	38.13	175	106	80	27
User Delay Cost	3.41	57.79	30.59	25	14
Environmental Cost	0.002	1.25	0.629	0.5	0.5

After the utility values of all the attributes are computed, the next step is to assign the importance weights to each attribute. The importance of these attributes is taken from pairwise comparison of AHP, as shown in Table 2.2, where owner cost is most important, followed by condition index, user delay cost, and environmental cost. Finally, the aggregated utility score for each alternative is computed by Equation 2.4. Since the objective was to minimize the values of attributes, an alternative having lower utility score achieves a higher rank.

2.4.3 ELECTRE III

ELECTRE was proposed by Roy (1968) for the decision aid. There are various versions of ELECTRE, each targeted for a different type of decision problems. ELECTRE III is used for a ranking problem where it is possible to assign relative importance weights to criteria.

ELECTRE has concepts of thresholds and outranking. The thresholds are given by decision makers who define the outranking relationship between alternatives. The outranking relationship is measured by concordance and discordance index. Concordance index measures the strength of support, which states as alternative a is at least as good as alternative b for most of the criteria. Discordance index measures the strength for those criteria which are against this hypothesis.

ELECTRE III Algorithm

A step by step procedure to apply ELECTRE III is explained as follows (Pena, Rebollo, Gibert, & Valls, 2007):

1. Like AHP and MAUT, the performance matrix outlining the criteria and alternatives is the first step of ELECTRE III.
2. Since ELECTRE is an outranking method, where an alternative outranks another alternative to establish its priority for a particular criterion, few threshold values depending on the problem statement has to be set by a stakeholder/decision maker.
 - Preference threshold [p]: A difference above which a decision maker strongly prefers an alternative a over an alternative b for criteria i .
 - Indifference threshold [q]: A difference below which a decision maker is indifferent to an alternative a over an alternative b for criteria i .
 - Veto threshold [v]: A value provided by a decision maker which blocks the outranking relationship between two alternatives for criteria i .
 - Weights [w]: The value stating the relative importance of attributes. It is a similar concept as also discussed in Section 2.4.1 and 2.4.2.

3. Calculate the concordance index per criterion. Concordance index measures the strength of support stating that alternative a is at least as good as alternative b .

$$C_i(a, b) = \begin{cases} 0, & \text{if } g_i(b) \geq g_i(a) + p_i(g_i(a)) \\ 1, & \text{if } g_i(b) \leq g_i(a) + q_i(g_i(a)) \\ \text{Otherwise,} & \frac{g_i(a) + p_i(g_i(a)) - g_i(b)}{p_i(g_i(a)) - q_i(g_i(a))} \end{cases}$$

where:

- $g_i(a)$ is a value of an alternative a for criterion i

- p is preference threshold and q is indifference threshold

4. Calculate overall concordance index as follows

$$C(a, b) = \frac{\sum w_i C_i(a, b)}{\sum w_i}$$

5. Calculate the discordance index for each criterion.

$$D_i(a, b) = \begin{cases} 0, & \text{if } g_i(b) \leq g_i(a) + p_i(g_i(a)) \\ 1, & \text{if } g_i(b) \geq g_i(a) + q_i(g_i(a)) \\ \text{Otherwise,} & \frac{g_i(b) - g_i(a) - p_i(g_i(a))}{v_i(g_i(a)) - p_i(g_i(a))} \end{cases}$$

where:

- v is veto threshold

6. Calculate credibility index as follows

$$S(a, b) = \begin{cases} C(a, b), & \text{if } D_i(a, b) \leq C(a, b) \forall i \\ \text{Otherwise} & \\ C(a, b) \prod_{D_i(a, b) \geq C(a, b)} \frac{1 - D_i(a, b)}{1 - C(a, b)} & \end{cases}$$

7. Determine the rank order by ascending or descending distillation.

ELECTRE III Application

ELECTRE III is a method which requires a decision maker to have a detailed understanding of a decision problem. This method mainly relies on the threshold values provided by a decision maker/expert. For the demonstration purpose, the authors played the role of a decision maker to state the value of preference threshold, indifferent threshold, veto threshold, and weights. The objective of bridges ranking and

the relative importance of attributes are kept the same as stated in Section 2.4.1 and 2.4.2. Table 2.4 shows the defined thresholds for each attribute.

Table 2.4: ELECTRE III thresholds values

Thresholds	Attributes			
	CI	OC	UDC	EC
Preference [p]	0.5	15	10	0.7
Indifference [q]	0.5	10	5	0.2
Veto [v]	1	1	1	1
Weights [w]	0.26	0.54	0.12	0.03

CI= Condition Index, OC = Owner Cost, UDC = User Delay Cost, EC = Environmental Cost

Once the thresholds have been defined, the next step is to generate the concordance matrix by following Step 3, where each alternative is compared with another for each criterion. Since this case has four criteria, this step yields four concordance matrices, which are aggregated based on the weight of each criterion by following Step 4. Similar to concordance, the discordance index for each alternative for each criterion is computed by following Step 5.

The last step is to compute the credibility index, which combines the concordance index and discordance matrices by checking which criteria are in favor of outranking relationship and which of them opposes it. Finally, the rank of each alternative is determined by the distillation process.

The comparison of alternative per criterion is a lengthy process as multiple concordance and discordance matrices are required to be generated. For this application, we have used an open source software J-Electre v1.0 (Pereira, 2017).

2.5 Results and Discussion

This section discusses the results of applying AHP, MAUT, and ELECTRE III on the data of twenty-two bridges. In addition to final ranking, each of the methods is also evaluated based on the scale defined in Section 2.3.

For AHP, if all the data is of quantitative nature then the main focus is to derive the relative importance of criteria/weights by pairwise comparison. In MAUT, in addition

to relative weights, the emphasis is on a selection of utility function and to reduce the data to utility scores. The application procedure of ELECTRE III is considerably different from the other two (AHP, MAUT) where the focus is on the comparison of alternatives instead of criteria/attributes. Moreover, ELECTRE III enables the comparison of heterogeneous scales of attributes e.g., the cost in euros, delay in hours, without reducing them into value functions or standard scales. Because of this, ELECTRE III does not provide a definitive ranking of alternatives as a final result (Figueira, Greco, Roy, & Slowinski, 2013).

Table 2.5: Ranking of twenty-two bridges computed by AHP, MAUT and ELECTRE III

Alternatives	AHP		MAUT		ELECTRE III
	Score	Rank	Score	Rank	Rank
Bridge A	0.061	18	0.972	22	17
Bridge B	0.048	14	0.748	13	13
Bridge C	0.048	13	0.826	15	14
Bridge D	0.026	3	0.350	3	2
Bridge E	0.032	5	0.588	7	7
Bridge F	0.060	16	0.854	16	18
Bridge G	0.068	22	0.865	18	21
Bridge H	0.043	12	0.793	14	11
Bridge I	0.036	9	0.492	5	5
Bridge J	0.050	15	0.873	19	16
Bridge K	0.026	4	0.333	2	4
Bridge L	0.033	8	0.671	9	8
Bridge M	0.022	1	0.141	1	1
Bridge N	0.040	11	0.738	12	10
Bridge O	0.026	2	0.383	4	3
Bridge P	0.061	19	0.694	10	15
Bridge Q	0.064	21	0.854	17	22
Bridge R	0.063	20	0.896	20	20
Bridge S	0.032	6	0.647	8	6
Bridge T	0.032	7	0.570	6	9
Bridge U	0.038	10	0.716	11	12
Bridge V	0.061	17	0.901	21	19

Table 2.5 provides the final ranking of twenty-two bridges where the objective was to have minimal owner cost, reduced condition index score, minimal user delay, and environmental cost. The ranking of each bridge can be compared back to its original data presented in Table 2.1. It is interesting to see that each of the MCDA methods has ranked Bridge *M* highest while this bridge does not have the lowest condition index and lowest user delay cost. This is because MCDA methods systematically account for all the attributes involved in decision-making instead of ranking on the bases of a single attribute only.

We have assumed that the ranking provided by different MCDA methods are similar if there is an only difference of one or two ranks. With this assumption notice the ranking of Bridge *B, C, D, E, F, G, H, K, L, M, N, O, Q, R, S, T,* and *U*, which have a difference of only one or two ranks for each method. This can be due to similar relative importance weights of attributes and stakeholders' preference for all the applied MCDA methods. However, there is also a notable difference in ranking for Bridge *A, I, J, P,* and *V*. For example, MAUT has ranked Bridge *A* on 22 while ELECTRE has ranked it at 17. Similarly for Bridge *V*, which is ranked at 17 by AHP and at 21 by MAUT. These difference in ranking can be referred back to the actual data values where for one method the highest maintenance cost causes a bridge to ranked lowest while for another method the lower maintenance cost and higher condition index assign the relative higher rank to a bridge.

In the following, we provide few remarks for each of the considered MCDA method based on our experience to apply them to maintenance decision-making problem.

- *Scalability*: Without the use of computerized software, it is complicated for each of these methods to include new alternatives and criteria/attribute in the assessment procedure. With a new alternative or attribute, the whole application procedure, irrespective of method type, would need to be performed again.
- *Ease of use*: AHP is easy to use as compared to MAUT and ELECTRE. However, in the case of several attributes, the application of AHP becomes a lengthy process due to the increased number of pairwise comparisons.
- *Encoding to definite scale*: AHP and MAUT both require encoding to a particular scale in order to enable different data values to be comparable. The scale conversion is performed by Saaty's scale of relative importance for AHP and by using utility function in MAUT. ELECTRE III does not require any conversion of data and work with the actual data values.

- *Uncertainty / Risk consideration:* MAUT is the only method that takes into account the uncertainty aspect of data and stakeholders' preferences. This is done by performing the trade-offs in having the best solution and worst solutions. AHP and ELECTRE III do not incorporate the concept of uncertainty.
- *Stakeholders' preferences:* The considered MCDA methods take into account the stakeholders' preferences in one manner or another. AHP requires the stakeholders' preferences to define the relative importance weights of criteria. MAUT require the stakeholders to define the indifference point for each attribute. Similarly, in ELECTRE III the threshold values are defined by stakeholders.
- *Understanding of results:* Since AHP and MAUT require the data conversion into a definite scale, the output produced by these two methods are easy to understand. Both of these methods generate an aggregated score for each alternative, while in ELECTRE III no definite aggregated scores are generated. Therefore, an additional method of distillation is applied to create alternatives ranking.

To summarize, AHP and MAUT are useful methods where a global aggregation per alternative and a definite ranking is required. If a stakeholder is uncertain of his preferences over the data values, then MAUT can be a method of preference. ELECTRE III can be used when a stakeholder has a clear understanding of his preferences over the data. At times, ELECTRE III has been referred to as a method that incorporates the fuzzy nature of decision-making in the form of thresholds. However, the experience gained with the application of ELECTRE III in this study suggests that in the case of minor differences in thresholds values, the overall ranking score is changed. This is because ELECTRE III solely relies on threshold values to compute concordance and discordance matrices.

2.6 Conclusions

Maintenance decision-making is a multi-faceted problem which demands the consideration of a number of attributes. MCDA methods provide a systematic framework where objective data along with the subjective preference of stakeholders are combined for decision making. The application of three methods of MCDA is provided to illustrate how they can facilitate in maintenance decision-making problem. We have found that, if the weighting structure of the criteria and the stakeholders' preferences are kept similar, the different methods of MCDA even having different application

procedure can provide comparable results. AHP method can be used for those maintenance decision-making problems where alternatives are not ample in number, and a definite ranking of alternatives is required. MAUT is the method of preference when there exist a large number of alternatives, and a stakeholder is uncertain of his preference choices. MAUT is one of few of MCDA methods that incorporates the concepts of uncertainty and utility theory. ELECTRE III can be a method of choice where a decision maker has a good understanding of his preferences over the data values. However, ELECTRE III requires the definition of multiple threshold values, which makes its application a tricky activity.

2.7 References

- Allah Bukhsh, Z., Stipanovic, I., Klanker, G., O'Connor, A., & Doree, A. G. (2018). Network level bridges maintenance planning using multi-attribute utility theory. *Structure and infrastructure engineering*, 1–14.
- Cinelli, M., Coles, S. R., & Kirwan, K. (2014). Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. *Ecological Indicators*, 46, 138–148.
- De Montis, A., De Toro, P., Droste-Franke, B., & Omann. (2004). Assessing the quality of different mcda methods. *Alternatives for environmental valuation*, 4, 99.
- De Montis, A., De Toro, P., Droste-Franke, B., Omann, I., & Stagl, S. (2000). Criteria for quality assessment of mcda methods. In *3rd biennial conference of the european society for ecological economics, vienna* (pp. 3–6).
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., Garcez, T. V., et al. (2015). *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis*. New York, NY, USA: Springer, 2015.
- de Almeida, A. T., Ferreira, R. J. P., & Cavalcante, C. A. V. (2015). A review of the use of multicriteria and multi-objective models in maintenance and reliability. *IMA Journal of Management Mathematics*, 26(3), 249–271.
- ERF. (2010). *European road statistics*. European Union Road Foundation. Brussels, Belgium.
- ERF. (2013). *Road asset management: An erf position paper for maintaining and improving a sustainable and efficient road network*. European Union Road Foundation. Brussels, Belgium.
- Figueira, J. R., Greco, S., Roy, B., & Slowinski, R. (2013). An overview of ELECTRE methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2), 61–85. doi: 10.1002/mcda.1482.
- Kabir, G., Sadiq, R., & Tesfamariam, S. (2014). A review of multi-criteria decision-making methods for infrastructure management. *Structure and Infrastructure Engineering*, 10(9), 1176–1210.
- Keeney, & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press.
- Patidar, V. (2007). *Multi-objective optimization for bridge management systems*. Transportation Research Board.
- Pena, R. R., Rebollo, L. P., Gibert, K., & Valls, A. (2007). *Use and evaluation of electre iii/iv*. Universitat Rovira i Virgili.

- Pereira. (2017). Mcda - multicriteria decision aid. <https://github.com/Valdecy/J-Electre>. GitHub.
- Roy, B. (1968). Classement et choix en présence de points de vue multiples. *Revue française d'informatique et de recherche opérationnelle*, 2(8), 57–75.
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1).

Network level bridges maintenance planning using Multi-Attribute Utility Theory

^{1 2} **Abstract:** Bridge infrastructure managers are facing multiple challenges to improve the availability and serviceability of aging infrastructure, while maintenance planning is constrained by budget restrictions. Many research efforts are ongoing, for the last few decades, ranging from the development of bridge management system, decision support tools, optimization models, life cycle cost analysis, among others. Since transport infrastructures are deeply embedded in society, they are not only subject to technical requirements but are required to meet the requirements of societal and economic developments. Therefore, bridge maintenance planning should accommodate multiple performance goals which need to be quantified by various performance indicators. In this paper, an application of Multi-Attribute Utility Theory (MAUT) for bridge maintenance planning is illustrated with a case study of bridges from the Netherlands road network. MAUT seeks to optimize multiple objectives by suggesting a trade-off among them and finally assigns a ranking to the considered bridges. Moreover, utility functions of MAUT appropriately account for the involved uncertainty and risk attitude of infrastructure managers. The main contribution of this study is in presenting a proof-of-concept on how MAUT provides a systematic approach to improve the decision-making of maintenance planning by making use of available data, accommodating multiple performance goals, their uncertainty, and preferences of infrastructure managers.

¹This chapter has been published as: Allah Bukhsh, Z., Stipanovic, I., Klanker, G., O'Connor, A., & Doree, A. G.(2018). Network level bridges maintenance planning using Multi-Attribute Utility Theory. *Structure and infrastructure engineering*, 1-14. doi.org/10.1080/15732479.2017.1414858.

²Extended work related to this chapter is also published (Appendix A).

3.1 Introduction

In recent years infrastructure asset management has been applied as a strategic governance approach with the aim of achieving more value from assets by making use of fewer resources. By combining engineering and economic principles with sound business practices, asset management strives for cost-effective investment decisions throughout the life-cycle of infrastructure (Tao, Zophy, & Wiegmann, 2000). Among other infrastructure objects, bridges present a vital link in any roadway network. From an economic viewpoint, it is crucial that bridges provide their designed function as part of the infrastructure network in an efficient manner. At one hand, bridges present the 30% value of the whole network while the length of bridges compared to the whole length of road networks is only approximately 2% (Chen & Miles, 2004). On the other hand, due to a longer lifespan, road bridges infrastructure is exposed to aging, adverse climate effects, and increased public demands. This puts a lot of pressure on infrastructure managers to not only improve the availability and serviceability of bridges but also to re-think the maintenance planning procedures due to the increasing budget restrictions and increased capacity demands by users. It is estimated that the ratio of expenses per route km of bridges or tunnels is 10 times the average expenses per route km of roads (Deterne, 2010).

Bridge maintenance planning is a process of deciding the scope, timing, costs, and benefits of future maintenance activities on a specific bridge while taking into account the relative importance of the bridge with respect to the overall road network. For over 20 years various Bridge Management Systems (BMS) have been used around the world to develop maintenance plans and allocate available budgets. These systems typically include single-objective optimization analyses. For example, Life Cycle Cost (LCC) analysis and reliability based concepts are well-established methodologies for single-objective analysis. Although useful for allocating budgets on the object-level, these BMS's are usually not taking into account other performance aspects related to economy, society, environment (Bush, Henning, Ingham, & Raith, 2014). LCC provides a solid base for the assessment of maintenance and budget distribution needs over a specified time period on the object-level. However, LCC does not take into account multiple objectives for the network-level bridge maintenance planning and instead yields exhaustive and detailed cost values for a number of bridges which are difficult to compare and prioritize.

Van Dam, Nikolic, and Lukszo (2012) recommended that asset management should no longer be viewed as a solely technical process, but instead should be viewed as

a socio-technical process. Since transport infrastructures are deeply embedded in society, they are not only subject to technical changes, but they also have to meet the requirements of societal and economic developments. Therefore, bridge maintenance planning should be performed at the network level where multiple performance goals can be considered. Multiple bridge performance aspects widen the scope of maintenance planning where a number of related performance goals other than minimizing owner cost must be considered. The example of such performance aspects is the structural performance of the bridge, safety, and security of users and workers, environmental impact, economic impact on the users, and impact on agency's and officials' reputation (political aspect). In addition to various aspects, these related performance goals can have a conflicting nature e.g., to minimize the impact on users, the agency might need to use more resources which will result in increased owner cost. Considering the large number of bridges on the network, it is intractable for an infrastructure manager to quantify the performance goals for each bridge and systematically perform the trade-offs among them in order to select those bridges that optimize the various performance goals. Moreover, at times, an infrastructure manager is uncertain of his preferences due to incomplete or unavailable data and due to lack of experience. So, the need to optimize multiple conflicting performance goals based on the preference uncertainty marks maintenance planning a complex decision-making problem.

For the first time, Von Neumann and Morgenstern (1945) introduced the Expected Utility Theory which deals with such decision problems where a decision maker chooses from a finite set of outcomes to balance involved risk and uncertainty by exploiting the concept of lottery. The lotteries are based on the concept of gambling where a decision maker can keep changing the values over a finite set of outcomes until an indifference point is reached. The main goal of these lotteries defined over attributes is to maximize the expected utility of the alternatives. To define a utility function over lotteries, the five axioms i.e., completeness, transitivity, continuity, monotonicity, and substitution related to preference structure must be followed. For the detailed theorem and axioms, a reader can follow (Hens & Rieger, 2010, Chapter 2). Keeney and Raiffa (1993) suggested a formal decision-making approach, namely Multi-Attribute Utility Theory (MAUT) based on Von Neumann and Morgenstern (1945) theory of Expected Utility. MAUT is also one of the fundamental methods of Multi-Criteria Decision Analysis (MCDA) due to its ability to consider probabilistic consequences.

In this paper, an application of MAUT for bridge maintenance planning is demonstrated. MAUT considers the multiple objectives (performance goals) represented

with attributes (performance indicators) and consistently captures risk attitude of decision makers as well as uncertainty in preferences (i.e., which value to choose from a finite set of outcomes) when the probability of achieving the results is not definite. The purpose of this paper is to illustrate the application of MAUT for network level maintenance planning where multiple performance goals, defined as objectives, can be optimized. We are referring to the *network level* maintenance planning because here the decision-making process is related to the selection of a single or group of bridges for maintenance that fulfills the defined objectives instead of focusing on different maintenance treatments for a single bridge only. The final results of MAUT will provide the ranking of a number of bridges based on the trade-offs of multiple performance goals.

One of the contributions of this paper is the methodological evaluation for the technical, economic, and environmental impacts assessments of bridge maintenance decisions. The main added value of the paper is that it has used available data from the road agency only and has shown how the decision-making process could be improved by implementing other aspects, other than owner cost, into the evaluation process. By giving the quantification procedure, the decision-making process can be fully followable and transparent. Nevertheless, the application of MAUT is providing the option to a decision maker to explicitly integrate risk and choice preferences. The rest of the paper is structured as follows: an overview of MCDA methods and the motivation to select MAUT is presented in Section 3.2. Section 3.3 discusses the need for a shift in bridge maintenance planning from object-level to network-level. The description of a case study, along with the quantification process of performance goals is outlined in Section 3.4. Section 3.5 illustrates the application of MAUT for bridge maintenance planning by ranking the bridges based on the trade-offs of multiple objectives. Finally, Section 3.6 provides the discussion and conclusion of this study.

3.2 Multi-Criteria Decision Analysis for Maintenance Planning

For an extensive and busy road network, maintenance planning is a complex decision-making problem. The complexity is originated mainly due to multiple objectives defined by performance goals, which are often competing and conflicting with each other. According to Keeney and Raiffa (1993), in multi-objective optimization, "It is often true that no dominant alternative will exist that is better than all other alternatives in terms of all objectives". In other words, there is never a solution that

optimizes all the involved performance goals/objectives simultaneously, for instance, they could be to maximize the reliability level of infrastructure vs. minimize the agency cost, to minimize the user delay vs. minimize the labor cost, and many others.

For the optimization of multiple objectives, a decision maker has to make certain trade-offs to gain the value from one performance aspect (e.g., reliability) on the cost of another (e.g., owner cost). The concepts of decision sciences and particularly methods of Multi-Criteria Decision Analysis (MCDA) suggest several analytical frameworks that facilitate decision makers to perform such trade-offs and rank the alternatives in an order that fulfills the defined objectives most optimally. In the following, an overview of MCDA methods used for maintenance planning is provided.

3.2.1 Methods of MCDA

According to a recent literature review of MCDA applied on maintenance and reliability research (de Almeida, Ferreira, & Cavalcante, 2015), an increasing trend on the application of MCDA methods for optimization of resources, strategies, and intervention has been noticed. A few methods of MCDA have mainly gained attention in this regard e.g., Pareto Front, MAUT, AHP (Analytical hierarchy process), MAVT, Goal programming, ELECTRE (Elimination and choice expressing reality) and TOPSIS (Technique for Order by Similarly to Ideal Solution). Traditionally, these methods of MCDA are classified into three types (de Almeida, Cavalcante, et al., 2015b):

- Synthesis method: These are weighted aggregation methods that provide the relative ranking of all the alternatives under considerations based on the preference structure of the decision maker. The example of synthesis methods is AHP, MAVT, MAUT, TOPSIS.
- Outranking method: These methods seek to eliminate all the explicitly dominant alternatives. For instance, one alternative outranks another if it performs considerably well on all the attributes. The example of outranking methods are ELECTRE and PROMETHEE (Preference Ranking Organization Method for Enrichment of Evaluations).
- Interactive method: These methods have a strong base in mathematical principles where the objective is defined in a set of targeted values. Goal programming and Pareto front are interactive methods.

With respect to the application of MCDA methods, the interactive methods are being applied extensively in maintenance optimization problems to search for the non-dominant solution that satisfies the objectives. However, interactive methods are based on complex heuristic search procedures e.g., genetic algorithms, particle swarm analysis, and they do not take into account the preferences of the decision makers (de Almeida, Ferreira, & Cavalcante, 2015). The outranking methods take the preferences of decision makers and enable the comparison of heterogeneous scales of attributes e.g., the cost in euros, delay in hours. without reducing them into value functions or standard scales (e.g., the scale of pairwise comparison). Due to this, these methods fail to enable the trade-offs among the different attributes and do not provide a definite ranking of alternatives (Figueira, Greco, Roy, & Slowinski, 2013).

The synthesis methods of MCDA are proven to be most promising, considering the multi-faceted nature of maintenance optimization problems. AHP is well known and widely applied MCDA method, but it requires pairwise comparison among alternative, which is inconceivable when alternatives are ample in number. TOPSIS works on the aggregation function by calculating the distance between a positive ideal solution and a negative ideal solution. MAVT is a variant of MAUT, which is a deterministic additive model and does not take into account the probabilistic aspects of decision-making. Since the planning of maintenance procedures are never definitive, the selected synthesis method of MCDA must take into account the preference uncertainty and risk tolerance of decision maker. Utility theory provides a systematic approach to capture the decision makers preferences under uncertainty (Ishizaka & Nemery, 2013).

As mentioned earlier, based on the quantitative axioms of Expected Utility Theory, Keeney and Raiffa (1993) introduced Multi-Attribute Utility Theory (MAUT) that reduces attributes' measure into utility values. Various functions such as linear, log, exponential, logarithmic, and quadratic can be used to determine the utility of an attribute (Meyer, 2010). The resulting utility function can only be four shaped i.e., convex, concave, S-Shaped and reverse S-shaped (Stewart, Reimers, & Harris, 2014). This paper aims to capture the preferences uncertainty and risk tolerance of a decision maker for maintenance planning attributes. Therefore we applied exponential utility functions to determine the utility values of attributes as suggested in Keeney and von Winterfeldt (2007). In the following, a few relevant studies and a procedure to apply the MAUT for maintenance planning is provided.

3.2.2 Application steps of MAUT

MAUT has been extensively used in multiple fields, e.g., for projects in investment planning (Jano-Ito & Crawford-Brown, 2017; Pujadas, Pardo-Bosch, Aguado-Renter, & Aguado, 2017), for health care (Claudio & Okudan, 2009; Sun, 2016), and performance and budget assessment of waterworks (Ismaeel, 2016). A few studies in the context of infrastructure management are also found in the literature that refers to MAUT as a base for mathematical modeling, for inspection and maintenance (de Almeida, Cavalcante, et al., 2015a; Garmabaki, Ahmadi, & Ahmadi, 2016), for bridge sustainability assessment (Dong, Frangopol, & Sabatino, 2015), for maintenance investment decision-making (Arif, Bayraktar, & Chowdhury, 2015) and for bridge network level optimization (Frangopol & Liu, 2007). However, the application of MAUT for multiple objective maintenance planning in transport infrastructure systems by incorporating the preferences of decision-makers is not fully explored. Most of the literature studies mentioned MAUT as a part of a larger framework and exploit its few relevant concepts. To explore the potential of MAUT for network level maintenance planning, this paper aims to apply MAUT for the maintenance planning decision-making where an infrastructure manager/decision maker has to deal with multiple objectives.

The steps to apply the MAUT on a maintenance planning problem are provided in Table 3.1 and discussed as follows (Keeney & Raiffa, 1993):

Table 3.1: Steps to apply MAUT on Maintenance Planning Problem

Steps	Descriptions
Step 1	Define the scope of the problem i.e., the set of objects/network for which maintenance is being planned
Step 2	Identify the performance goals or objectives of the maintenance planning
Step 3	Quantify the related performance indicators (attributes) to represent the objectives in utility functions Calculate exponential Single Utility Functions (SUF) by
Step 4	<ul style="list-style-type: none"> • Preparing the lottery question of best and worst values of attributes • Capturing the risk attitude of decision-maker with the Certainty Equivalent • Illustrating the utility values of each attribute for all alternatives
Step 5	Perform the value trade-off among attributes based on decision-makers preferences
Step 6	Considering the nature of attributes, calculate the final aggregative utility values from SUF by additive or multiplicative form
Step 7	Ranking based on value trade-offs and performance goals (objectives)
Step 8	Finally, analyze the result and perform the other related scenarios

Step 1 to Step 3 define the scope of the maintenance planning and construct the data required for the MAUT. Based on the defined performance goals, the performance indicators are quantified. The identified performance indicators are referred as attributes whereas the objects that are under consideration for maintenance are named as alternatives.

In Step 4, the utility function of each attribute's measure is calculated. As mentioned earlier, there are several forms to calculate the utility functions, and the exponential utility function is typically used to incorporate the risk and uncertainty factors (Keeney & von Winterfeldt, 2007; Keeney & Raiffa, 1993, Chapter 4 & 5). The exponential utility function presents the decision makers with the lottery question of a maximum value and a minimum value of an attribute where the indifference point has to be reached between the best and the worst possible values. Such an indifference point for a decision maker is referred to as the *Certainty Equivalent (CE)*. The probability of obtaining the best possible value or the worst possible value is referred to as *Expected Value (EV)*. The indifference point chosen by the decision maker represents his/her attitude towards risk and risk tolerance. As listed below, an indifference value where CE is equivalent to EV represents risk neutral behavior. If the value of CE is lower than EV, the decision maker has a risk-avoiding attitude, whereas the CE value higher than EV shows the risk-taking attitude. Consequently, the risk-avoiding attitude has a positive risk tolerance while a risk-taking attitude has a negative risk tolerance value.

$$\text{Risk Attitude} = \begin{cases} \text{Risk Neutral, if } EV = CE \text{ (Linear shaped)} \\ \text{Risk Avoiding, if } EV \geq CE \text{ (Concave shaped)} \\ \text{Risk Taking, if } EV < CE \text{ (Convex shaped)} \end{cases}$$

With the consideration of decision makers' uncertainty in eliciting indifference points (CE) under risk consideration, the final computed utilities of an attribute lies between $U(x_{min}) = 0, U(x_{max}) = 1$.

In Step 5, trade-offs among the attributes are made in order to find a solution that maximizes the performance goals/objectives. These trade-offs characterize the relative importance of attributes for the defined objectives.

For the final aggregation of utility functions, in Step 6, the additive or multiplicative form can be used. The additive form requires the attributes to be mutually and preferentially independent. Preferentially independent means the preferences on the level (i.e. values) of an attribute X is not dependent on the levels of an attribute

Y. For instance, let X: Maintenance treatment (coating, cleaning) and Y: Cost(100, 250). If a decision maker prefers *cleaning* treatment irrespective of its *cost* and prefer *100 cost* regardless of chosen maintenance treatment, then X and Y are mutually and preferentially independent. While, if a decision maker prefer *coating* regardless of *cost*, but the *cost* depends on the maintenance treatment chosen then X is preferentially independent of Y but not mutually preferentially independent since Y depends on X. When attributes are not mutually and preferentially independent, then the multiplicative form is used (Krishnamurty, 2006).

Finally, in *Step 7*, several alternatives having qualitative or quantitative attributes are ranked based on the Single Utility Function (SUF) and trade-off values. In *Step 8*, the ranking can be further analyzed by applying various scenarios with different trade-off values. MAUT provides a step-by-step procedure to accommodate a number of attributes related to alternatives for the realization of multiple objectives based on the preferences of a decision maker.

3.3 Object to Network-Level Maintenance Planning

In the current state of the practice, most Bridge Management Systems (BMS) are very effective at storage and retrieval of the raw data needed for maintenance planning. With these data, it is possible to analyze the change of performance over time and to identify those bridges that are in need of maintenance. Based on that, life cycle costs for different maintenance scenarios can be calculated and compared. To express and manage the spectrum of possible futures (i.e. maintenance scenarios), the concept of a 'candidate' is suggested (Patidar, 2007). A 'candidate' is defined as a life-cycle activity profile for one bridge, consisting of a sequence of agency activities, including do-nothing, in each of a sequence of future time periods. Development of other candidates (i.e., maintenance scenarios) treatments for a single bridge and a selection of the best one is an essential aspect of decision-making by the infrastructure manager. The planner then repeats this process for the other bridges in the network. By iterating between the bridge and network level, one or more optimal scenarios can be reached, resulting in the selection of a number of bridges for maintenance in a certain period. It should be noted that the focus of existing BMS's are still mainly on bridge condition level and owner's costs, rarely taking into account other impacts of the bridge, such as environmental impacts, availability, importance on the transport network and society as a whole.

In the course of developing network-level bridge maintenance plans, infrastructure managers typically face a variety of objectives and constraints. Examples of objectives are to maximize cost-effectiveness, to minimize vulnerability to damage, to maximize average condition, and to optimize a utility index that combines various objectives. Constraints include budgetary limits that cannot be exceeded or a minimum level of average bridge performance (Patidar, 2007). These multiple objectives, which are often conflicting in nature, present a complex decision-making problem for maintenance planning. Such complex decision-making problems demand clear statements of objectives and their quantification in the form of defined performance indicators. To enable the trade-offs among multiple competing objectives, it is necessary to identify a set of performance goals and a set of performance indicators to quantify them.

The project COST Action TU 1406 (2015) aims to bring together both the research and practicing community in the field of bridge assessment, which will incorporate different aspects of bridge performance goals, based on technical, environmental, economic and social factors. Within the COST Action TU 1406 bridge performance indicators have been collected through an extensive survey, with the aim to produce guideline documents linking collection and quantification of performance indicators, performance goals, standards, and practices with decision-making processes (Stipanovic, Høj, & Klanker, 2016; Strauss, Ivankovic, Matos, & Casas, 2016).

Table 3.2: Performance goals quantified by performance indicators and quantification process

Performance goals (Objectives)	Aspect	Performance Indicator	Calculation (Quantification process)
Improve reliability of assets	Reliability	i) Condition Index	Visual Inspection, Analysis of finding
		ii) Reliability Index	Probabilistic model for different limit states (Hazards)
Minimise agency cost	Economy	i) Construction Cost	Design project, quantities
		ii) Maintenance Cost	Maintenance activity, Quantities produced
		iii) End of Life Cost	Life cycle analysis output
Minimise environmental impact	Environment	i) Environmental Impact	Chosen maintenance activity, Quantities of materials produced, LCC analysis
		ii) Noise	Vehicle noise, machine noise
Minimise impact on users	Society	i) User delay cost	Extra travel time, No. of users affected, cost of an hour of a user, duration of maintenance activity
		ii) User costs	Vehicle operating costs, travel time costs, accident costs
Improve safety of network	Safety	i) Accident Rate	No. of traffic accidents, No. of fatalities, No. of injuries

This paper presents how the existing data, namely different performance indicators, at road agencies, can be used for the quantification of multiple performance goals. Table 3.2 outlines the performance goals, identified performance indicators, and their quantification process. The performance goals and relevant performance indicators used in the study are highlighted in *italic*. Performance goals highlight the objectives of the network-level maintenance planning where the performance indicators enable us to quantify and measure the defined goals and their consequences on the number of bridge alternatives. Performance indicators are also sometimes referred to as measures, attributes, or criteria. Considering the application of MAUT presented here, we will refer performance indicators as attributes (i.e., bridge condition index, owner cost, user delay cost, and environmental cost) and to the list of bridges as alternatives.

3.4 Bridges Maintenance Planning

This section presents a case study to illustrate how utility functions and MAUT can be used in decision-making processes for maintenance planning. Let us consider for the maintenance planning that, an infrastructure manager (i.e., decision maker) is presented with 100s of bridges that are in need of repair, while at the same time, he is confronted with the constraints of the limited budget, network availability, and safety, among many others. Moreover, the infrastructure manager wants to perform maintenance only where he can minimize the owner costs and maximize the object's reliability level. Such a decision-making problem requires the assessment and selection of alternatives (i.e., bridges), where a trade-off among the objectives can be made, and most of the constraints can be met.

The data of twenty-two randomly chosen road bridges from the Netherlands road agency Rijkswaterstaat is used, to demonstrate how the MAUT and utility function can facilitate in the decision-making of maintenance planning. The data provided from the agency included a description of the twenty-two bridges in terms of bridges' age, geometry, condition index on the element level, traffic intensity, planned maintenance activity on the element level, unit cost of chosen maintenance option, and maintenance duration. Generally, the condition indexes of Netherlands road bridges are found to have the optimum range (i.e., 1: Very good to 3: Fair, on the scale from 1 to 6); therefore there are not many maintenance treatments considered. The bridge data included the maintenance costs of the various maintenance treatments (e.g., replacing top-layer, coating, cleaning, etc.).

No decision regarding the selection of other optional maintenance treatments corresponding to different maintenance costs was made for this case study. Moreover, the causal factors e.g., age of the bridge, the material used in bridge, the span of a bridge are not considered independently during the decision-making procedure of this maintenance planning case as inspection and condition index assessment activities must take into account these factors (Rashidi & Gibson, 2011). Since twenty-two bridges were chosen randomly from the Netherlands road network, we did not establish any correlation concerning bridge location and its impact on surrounding objects in the road network. In other words, bridges as an individual entity are considered assuming that the maintenance of these twenty-two bridges is under consideration only.

For this illustrative application of MAUT, the twenty-two bridges are required to be ranked where the objectives are a) *to minimise condition index (where lower value represents better condition)*, b) *to minimise owner cost*, c) *to minimise the impact on users (expressed as user delay costs)* and d) *to minimize environmental impact (expressed as environmental costs)*. Considering the aforementioned objectives, the ranking of the bridges will be based on the minimization of owner cost, condition index, user delay cost, and environmental cost. However, the provided data do not have these data attributes readily available. Therefore, these attributes have to be quantified from the available raw data to be used in MAUT analysis. Following sections outline the quantification process to compute system-level condition index, owner cost, user delay cost, and environmental cost from the provided data of twenty-two bridges.

3.4.1 Condition Index

Condition indexes represent the overall ‘health’ of a bridge. The agency uses a six-level condition assessment score to assess the condition of a bridge where 1 represents very good condition and 6 represents out of service state. The objective is to minimize the condition index, which means the lower the condition index, the better will be the service level of the bridge. In the case study data, a bridge as a structure is divided into seven elements: Superstructure, Bearings, Abutments, Joints, Pavement, Railing, and Guardrail. With these sub-elements, each bridge has seven visually assessed condition indexes representing the structural performance of each bridge’s element. In order to get the overall understanding of the bridge condition, these element-level condition values must be aggregated to a single system-level condition value, namely the Bridge Condition Index (BCI).

In practice there are numerous ways, varying by country and agency, to compute the BCI, aggregated from the element-level to the system-level, using worst element score, weighted average method, ratio scale (Chase, Adu-Gyamfi, Aktan, Minaie, et al., 2016; Hooks & Frangopol, 2013; Stratt, 2010). For this case study, we have applied the weighted average method, where an expert establishes the relative weights to each element. The concepts from the Analytical Hierarchy Process is used for this purpose (Saaty, 2000). To determine the importance of each element for the overall structural integrity of a bridge, an expert must consider the failure history of an element by type, frequency of maintenance required by the elements and the failure consequence of an element on the overall bridge. The pairwise comparison of Superstructure, Bearings, Abutments, Joints, Pavement, Railing, and Guardrail, has been performed. Each of these seven elements is compared with each other based on the Saaty's fundamental scale of importance. This scale defines the five levels of importance between two elements, where 1 represents the equal importance between two elements and 5 suggests that one element is extremely important than the other. Finally, the relative importance of each element is obtained as follows:

- Formalized a matrix M for all the elements based on the gathered input
- Normalized the matrix from 0 to 1 by summing each column and dividing individual element importance values by its column's sum
- Finally, calculating the average of each row of the matrix to gain the relative weights/importance of each element

The relative importance score for each element determining their level of importance in the overall structural performance of the bridge is provided in Table 3.3. The Table shows that the superstructure is most important for the overall structural integrity of the bridge where railing and guardrails are least important. Once the relative importance of each element is obtained, the BCI is calculated by computing following equation:

$$BCI_s = \sum_{i=1}^7 (CI_i * W_i) \tag{3.1}$$

Where: BCI_s = Bridge condition index at system level s
 CI_i = Condition index of an element i
 W_i = Weighted score of an element i elicited by an expert (provided in Table 3.3)

Table 3.3: Weighted score of bridge elements

No.	Elements	Weighted score
1	Superstructure	0.3185
2	Bearings	0.2104
3	Abutment	0.1813
4	Joints	0.1288
5	Pavement	0.0618
6	Railing	0.0510
7	Guardrail	0.0478

3.4.2 Owner Cost

Owner cost is a monetary value borne by an agency as a result of construction, maintenance, and end of life cost. In this case study, only maintenance cost is considered for the owner cost, which is the money spent on the maintenance activity of a particular bridge. In order to get a unit maintenance cost per m^2 of a bridge, the maintenance cost per activity type is multiplied by the quantity of the material used, which is then divided by the size of deck area of the bridge (m^2). The maintenance cost is provided on element-level (seven elements in this case) for the chosen maintenance option, which is then summed-up to represent the cost for a bridge as a whole. The formula to calculate the maintenance cost is provided as follows:

$$OC = \frac{\sum_{i=1}^7 (UCA_i * Q_i)}{A} \quad (3.2)$$

Where: OC = Total owner (maintenance) cost per m^2 of the bridge

UCA = Unit cost of maintenance activity per element i

Q = Quantity of the area/volume per element i

A = Deck area of a bridge (m^2)

3.4.3 User delay cost

User delay cost represents the monetary value of extended travel time of road users due to the presence of work-zones on the road. The user delay cost depends on several factors as mentioned in (Tighe, Eng, & McCabe, 2005; Wang & Goodrum, 2005), a) extra travel time due to the imposed speed restriction b) the number of users affected defined by traffic intensity in terms of average traffic per hour passing over the bridge on a working day, c) the cost of an hour of the user time, and d) finally the duration of the maintenance activity. The cost of an hour of user travel time varies per country based on the income levels. For western Europe, the values mostly vary between €6

for Germany to €12.12 for Sweden, except for Switzerland, which is €31.73. For further details, please refer to (Wardman, Chintakayala, de Jong, & Ferrer, 2012).

In this study, we used €9 including VAT as the cost of an hour of an all-purpose commuting user for the Netherlands determined by Kouwenhoven et al. (2014). We used this value since the case study was considering the traffic intensity and bridge data from the Netherlands road network. The computed user delay cost is divided by the deck area of a bridge to keep the calculation procedure consistent with other attributes. In this estimation of user delay cost, no difference between passenger traffic and freight traffic is considered. Though, the user delay cost for different users type can be computed by estimating the percentage of traffic per user type in ADT_t with respect to V_{ot} .

$$UDC = \frac{ETT * ADT_t * V_{ot} * N_t}{A} \quad (3.3)$$

Where:

$$ETT = \frac{L}{V_r} - \frac{L}{V_s}$$

UDC = User delay cost

ETT = Extra travel time (h)

L = Length of the working zone (km)

V_r = Reduced velocity in working zone during maintenance (km/h)

V_s = Standard velocity during normal condition(km/h)

ADT_t = Avg. traffic per hour (number of passing vehicles per hour)

V_{ot} = Value of a time per person per hour (€)

N_t = Duration of certain maintenance activity (h)

A = Deck area of a bridge (m^2)

3.4.4 Environmental cost

Environmental cost represents the monetized value of environmental impacts of different activities during the bridge life cycle. The impact on the environment is caused mostly by the materials production and transport during the construction process, maintenance activities, and end of life of an object. According to the environmental study of Hegger and de Graaf (2013), most of the environmental impacts are caused due to the use of construction material during initial construction and the subsequent maintenance. Therefore, only the effect on the environment per kg per material produced for maintenance activity is considered in this estimation of environmental cost. Similarly to maintenance cost, the environmental cost is calculated on the

element-level, which is then summed up to the system-level to represent the overall environmental cost of a maintenance activity for a particular bridge.

The determination of environmental cost due to maintenance activity is based on three aspects. First, the environmental effect per impact category based on material type, provided by GaBi software database (Thinkstep, 2015), is considered (EE_i). Second, the material quantity per kg produced for the maintenance activity is estimated (Mq_j). Finally, to monetize environmental effect into euros, the environmental effect per impact category values are multiplied by their shadow prices (SP_i) established by Rijkswaterstaat (de Bruyn et al., 2010; TNO-MEP, 2004). Table 3.4 presents the environmental effect categories along with their shadow prices. The CO_2 emission caused by traffic during the downtime period of a bridge, such as the maintenance and repair period is not included in the calculation of environmental impact values due to their negligible impact. Equation (3.4) provides the details of environmental cost estimation.

$$EC = \sum_{e=1}^7 \left[\frac{\sum_{i=n}^m EE_i * SP_i}{A} \right] \quad (3.4)$$

$$EE_i = \sum_{j=n}^m EE_{i,j} * Mq_j$$

Where:

EC = Environmental cost

EE_i = Environmental effect per impact category i (kg of impact category equivalent (ICE) per element)

SP_i = Shadow prices per impact category i (€ per kg of ICE)

i = Environmental impact category n until m (see Table 3.4)

$EE_{i,j}$ = Environmental effect per impact category i of material j produced (kg of ICE per kg of material)

Mq_j = Material quantity per element for material j (kg of material per element)

j = Different material types

A = Deck area of bridge (m^2)

e = Element e of a bridge

Table 3.4: Environmental effect categories and shadow prices (TNO-MEP, 2004)

Environmental effect category	Shadow price (/ kg equivalent)
Abiotic depletion elements (ADP) (/Sb eq)	0.16
Abiotic depletion fossil (ADP) (/Sb eq)	0.16
Global warming potential (GWP) (/CO2 eq)	0.05
Ozone depletion potential (ODP) (/CFK-11 eq)	30
Photochemical ozone formation potential (POCP) (/C2H2 eq)	2
Acidification potential (AP) (/SO2 eq)	4
Eutrofication potential (EP) (/PO4 eq)	9
Human toxicity potential (HTP) (/1,4-DCB eq)	0.09
Freshwater aquatic ecotoxicity potential (FAETP) (/1,4-DCB eq)	0.03
Marine aquatic ecotoxicity potential (MAETP) (/1,4-DCB eq)	0.0001
Terrestrial ecotoxicity potential (TETP) (/1,4-DCB eq)	0.06

3.5 Multi-attribute Utility Analysis for maintenance planning - Case Study

With the definition of performance goals and quantification of performance attributes in Section 3.4, MAUT is applied to the case study data. The first step of MAUT is to determine Single Utility Function (SUF) for each attribute. Various functions (e.g., log, cardinal, linear, power, quadratic) can be used to determine the utility scores of attributes (Meyer, 2010). Utility functions and Expected Utility Theory have widespread use, specifically in economics and game theory, where the scientific literature suggests different procedures to compute utility scores and to capture uncertainty (Chung, 1994; Levy & Markowitz, 1979).

In this paper, we have mainly followed the computation procedures suggested by Keeney and Raiffa (1993). To determine the SUF, Keeney and von Winterfeldt (2007) recommend the use of linearity function where the value of an attribute is useful in itself and use of exponential utility functions where the utility score must exhibit the risk aversion or risk proneness. To compute the global aggregated scores for each alternative, SUF capturing the preferences uncertainty and risk attitude is determined in Section 3.5.1. The relative importance of attributes concerning decision-makers preferences is given in Section 3.5.2. The final global aggregation, based on the additive form, is provided in Section 3.5.3, and finally, Section 3.5.4 outlines the results of MAUT application. The similar MAUT application procedure

has been found in the cases of supplier selection problem (Min, 1994), product line selection (Thevenot, Steva, Okudan, & Simpson, 2006), and patients prioritization in an emergency unit (Claudio & Okudan, 2009). The equations to calculate SUF for each attribute is given by Equation (3.5) to Equation (3.9).

$$U_i(x_i) = A - B * e^{\left(\frac{-x_i}{RT}\right)} \quad (3.5)$$

Where:

$$A = \frac{e^{\left(\frac{-Min(x_i)}{RT}\right)}}{\left[e^{\left(\frac{-Min(x_i)}{RT}\right)} - e^{\left(\frac{-Max(x_i)}{RT}\right)} \right]} \quad (3.6)$$

$$B = \frac{1}{\left[e^{\left(\frac{-Min(x_i)}{RT}\right)} - e^{\left(\frac{-Max(x_i)}{RT}\right)} \right]} \quad (3.7)$$

$$RT_i = \frac{-CE_i}{\ln\left(\frac{-0.5U_i(Max(x_i)) - 0.5U_i(Min(x_i)) + A}{B}\right)} \quad (3.8)$$

Where:

$U_i(x_i)$ = Single utility value for attribute i of an alternative x

A, B = Scaling constant

e = The exponential constant i.e. 2.718

$Min(x_i)$ = Minimum value of an attribute i across all alternatives

$Max(x_i)$ = Maximum value of an attribute i across all alternatives

RT = Risk tolerance

Since there is a cyclic dependency to calculate A, B, and RT, the Equations (3.5) through (3.8) have to be solved iteratively. Theoretically, RT can be approximately equal to the maximum value x of an attribute i for which a decision maker is willing to accept the $Min(x_i)$ and $Max(x_i)$ with equal probabilities instead of obtaining 0 for certain. As reported earlier, the concepts of the utility function are inspired by gambling, where with the equal probability to obtain $Min(x_i)$ or $Max(x_i)$ values, a gambler needs to take a certain risk. To compute the exact risk tolerance value, RT must satisfy the following Equation (3.9). The trial-and-error approach can be used to satisfy the following relation by exploiting the Goal Seeker function of Microsoft Excel (Middleton, 2006).

$$e^{\frac{-CE}{RT}} = 0.5 * e^{\frac{-Max(x_i)}{RT}} + 0.5 * e^{\frac{-Min(x_i)}{RT}} \quad (3.9)$$

where CE is a certainty equivalent, for each attribute, which is calculated by presenting the decision maker with a lottery question. The indifference point between the $Min(x_i)$ and $Max(x_i)$ for a decision maker is the value of CE. The value of RT shows the willingness of a decision maker to take a risk. RT value is negative for risk-taking behavior while positive for risk-avoiding behavior. In order to avoid the complexity involved in eliciting RT, we assume that the decision maker always has risk-avoiding behavior.

3.5.1 Assessment of Single Utility Function

In this section, the utility function of each attribute is calculated. It is worth mentioning that the utility of an attribute is relative to the decision maker's choices, which can change over time. In this exercise, the authors played the role of decision-makers where the preference values mimic real decision situations. To represent the lottery question, along with the risk attitude of a decision maker, we adopted the convention from Claudio and Okudan (2009).

Single Utility Function of Condition Index

The system-level condition score of each bridge is calculated from seven sub-elements by Equation (3.1). As mentioned earlier in Section 3.4.1, a lower condition index represents a *good* bridge condition. Therefore, the objective is to improve the service level of the bridge by minimizing the condition index score. It is important to

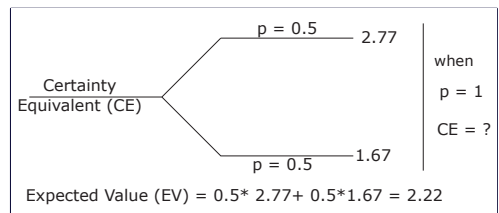


Figure 3.1: Lottery setup to discern the CE of bridge condition index

notice that the bridges chosen for the case study are generally found to have a good condition with condition scores ranging from 1.67 to 2.77 (See Table 3.6). Therefore, in the lottery question presented in Figure 3.1, the difference between the maximum condition score and minimum condition score is very small. The Expected Value (EV) of condition index is determined by considering the 50% probability of obtaining $Max(x_i)$ i.e. 2.77 and 50% probability of obtaining $Min(x_i)$ i.e. 1.67. The obtained EV is equal to 2.22. Assuming the risk-avoiding attitude of a decision maker, the chosen indifference point, also called Certainty Equivalent (CE), is 1.70, which means is an

acceptable condition score for the bridges. The risk tolerance value is calculated by satisfying the Equation (3.9) with a trial-and-error method. Due to the risk-avoiding attitude of a decision maker depicted in CE, the chosen RT value as small as 0.7. The exponential single utility function of each of twenty-two bridges is calculated by solving the Equation (3.6) through (3.8) iteratively, which finally yields following Equation (3.10). Figure 3.2 shows the utility plot of condition index values where the utility values increase steadily.

$$U_1(x_1) = 1.26 - 13.72 * e^{-x_1/0.7} \tag{3.10}$$

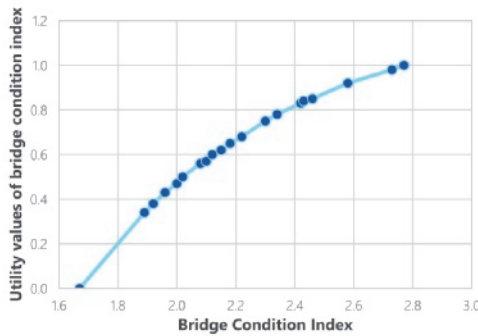


Figure 3.2: Utility plot of condition index

Single Utility Function of Owner Cost

The owner cost of all the twenty-two bridges is calculated based on Equation (3.2). The lottery question provided in Figure 3.3 can be read as 50-50% probability of having best (i.e., minimum) owner cost or worst (i.e., maximum) owner cost. The

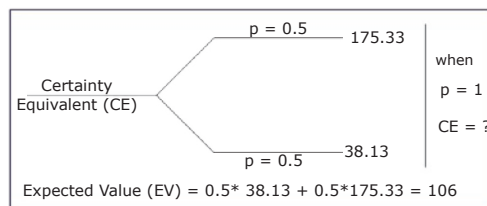


Figure 3.3: Lottery setup to discern the CE of owner cost

data of owner cost for 22 road bridges are provided in Table 3.6. Considering the maximum cost i.e. 175.33 and minimum cost i.e. 38.13, the computed EV is equal to 106. The purpose is to reduce owner cost as much as possible, therefore the

indifference point or certainty equivalent between the maximum and minimum cost is approximated to be 80. Assuming the risk avoiding behavior, the obtained RT value is 27 which is computed by substituting the values of CE, min, max and trial-and-error value of RT in Equation (3.9). By solving the Equation (3.6) through (3.8) iteratively, the exponential utility of each owner cost value is calculated as in Equation (3.11).

$$U_2(x_2) = 1.00 - 4.13 * e^{-x_2/27} \tag{3.11}$$

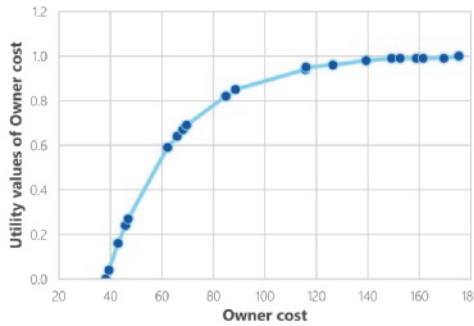


Figure 3.4: Utility plot of owner cost

Figure 3.4 provides a plot of the utility values for the owner cost, which shows the steep increase in utility values (y-axis) with respect to owner cost (x-axis). It can be said that the lower owner cost gets the small utility values in order to be ranked higher in the (aggregated) minimization function.

Single Utility Function of User delay cost

The user delay cost of each bridge is computed by the Equation (3.3). Figure 3.5 represents the lottery question presenting the minimum and maximum user delay cost of bridges (See Table 3.6 for the data), where the computed EV value is 30.59. The objective of this lottery question is to have a minimum user delay cost. The indifference point between the minimum and maximum user delay cost is reached at the value of 25.

By iteratively solving the Equation (3.6) through (3.8) and by satisfying the Equation (3.9), the value of scaling constant of A , B and RT is computed, which is provided in Equation (3.12).

$$U_3(x_3) = 1.00 - 1.34 * e^{-x_3/14} \tag{3.12}$$

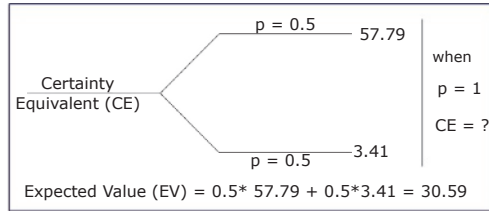


Figure 3.5: Lottery setup to discern the CE of user delay cost

Based on the values defined in Equation (3.12), the exponential utility values of user delay cost for each bridge is calculated and presented in Figure 3.6. The higher user delay costs obtain higher utility scores, which are not preferred in the minimization function.

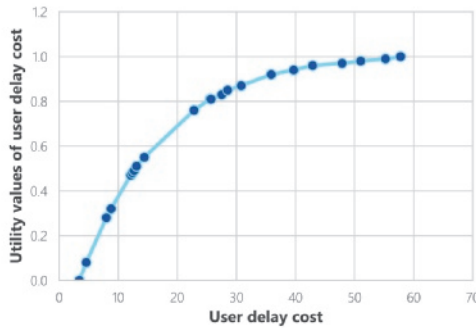


Figure 3.6: Utility plot of user delay cost

Single Utility Function of Environmental Cost

The environmental cost of the bridges is determined by Equation (3.4). Since the environmental cost is not a very dominant factor in maintenance planning and maintenance tasks, the shadow prices per environment effect category are very low, provided in Table 3.6.

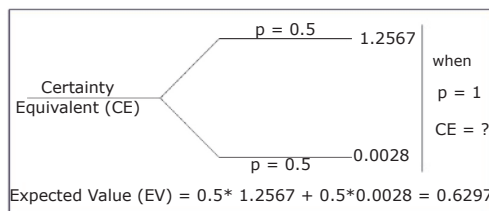


Figure 3.7: Lottery setup to discern the CE of environmental cost

The lottery question in Figure 3.7 provides the minimum and maximum amount of environmental cost of all the bridges. The EV with the 50%-50% probabilities of obtaining the minimum and maximum value of environmental cost is computed as 0.62. Assuming the risk-avoiding attitude of a decision maker regarding the environment, the CE value is chosen to be 0.5. Similarly, the RT value computed by solving Equation (3.9) is 0.50, which is not very low, considering the minimum and maximum value bounds of environmental cost. The calculation of exponential single utility function of environmental cost of each bridge is determined by Equation (3.6) through Equation (3.8), which yields Equation (3.13). The plot of environmental cost in Figure 3.8 depicts the uniform increase in the utility values associated with environmental values.

$$U_4(x_4) = 1.08 - 1.09 * e^{-x_4/0.5} \tag{3.13}$$

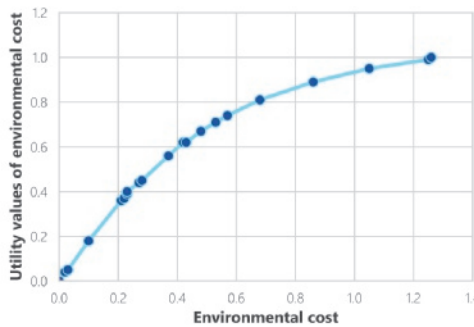


Figure 3.8: Utility plot of environmental cost

3.5.2 Attributes trade-offs

As mentioned in Section 3.4, the decision problem of maintenance planning requires the ranking of twenty-two bridges in an order where the condition of a bridge can be improved, and owner cost, user delay cost, and environmental cost can be minimized. While the minimization of one attribute might result in maximization of the other one. For instance, to minimize the user delay cost an agency might need to employ more resources, which will result in increased owner cost. Therefore, a trade-off among these attributes has to be performed. The procedure recommended by (Keeney

& Raiffa, 1993) for assigning the weighting factor to attributes is adopted for this purpose. A direct rating method is used, which is represented as follows:

$$k(x_i) = \frac{rate(x_i)}{\sum_{j=1}^n(x_j)} \quad (3.14)$$

Where:

- $k(x_i)$ = Weighting factor of each attribute i across all alternatives
- $rate(x_i)$ = Rate/weight assigned by expert for attribute i
- $\sum_{j=1}^n(x_j)$ = Total of all the weights assigned by an expert to n attributes

Based on general preferences for maintenance planning, each attribute obtains a weighting score out of 100 that depicts its importance in the maintenance planning scenario. Table 3.5 shows the obtained weights for each attribute. Since the bridges chosen for this case study have a relatively good condition, the owner cost gets the highest importance weight instead of condition index. Following owner cost, the condition index is in second place, and user delay cost is at the third place. The preferences structure shows that environmental costs have the least importance during maintenance planning.

Table 3.5: Relative rating of attributes (k)

Performance aspects	Performance Indicator	Obtained Weight*	Direct Rating
Reliability	Condition index	80	80/280 = 0.29
Economy	Owner cost	90	90/280 = 0.32
Environment	Environmental cost	40	40/280 = 0.14
Availability	User delay cost	70	70/280 = 0.25

* Each attribute obtain weight out of 100

It is worth noticing that the rating of these attributes is relative to each decision maker. The importance of these attributes in maintenance planning can largely vary from one decision maker to another. Hence, different scenarios can be visualized in the overall ranking of the bridges.

3.5.3 Aggregated Utility

The final step in the MAUT analysis is the computation of the aggregated utility of each alternative (i.e., bridge). A selection on the use of additive or multiplicative aggregation form has to be made based on the preference (in)dependency among attributes. For this case, the values maintenance cost, condition index, user delay cost, and environmental cost are dependent on each other, but this does not automatically imply that a decision maker must take into account the values of other attributes while stating preferences on one attribute. In other words, a decision maker can prefer the minimum maintenance cost, while at the same time prefer the maximum improvement in a condition index. This preference makes the maintenance cost and condition index mutually and preferentially independent of each other. Considering this, we used the additive form to compute the global aggregated score for each alternative, as shown in Equation (3.15). Since this is a minimization problem, the small aggregated score represents the most preferred alternative.

$$U(x) = \sum_{i=1}^n k_i U_i(x_i) \quad (3.15)$$

Where:

$U(x)$ = Multi-attribute utility of alternative x

k = Weighting factor of each attribute i (see Table 3.5)

$U_i(x_i)$ = Single attribute utility of each attribute i for an alternative x

3.5.4 Results

By calculating the aggregated utility from the single utility functions of attributes, the ranking of all bridges is obtained. The bridges are ranked based on the minimization of an aggregated utility function where a bridge is ranked higher, having the minimum condition index, owner cost, user delay cost, and environmental cost.

Table 3.6 outlines the actual data, single utility values, aggregated additive utility values, and a final ranking of each bridge. The final ranking takes into account the relative importance of each attribute in the form of k (see Table 3.5) and risk attitude of decision maker during the computation of a single attribute utility function. Moreover, it also incorporates the uncertainty of decision makers preference by

Table 3.6: Ranking of bridges based on utility (minimization) function

No.	Bridges Name	Attributes				Single Utility values				Additive Utility values	Rank
		BCI	OC	UDC	EC	U(BCI)	U(OC)	U(UDC)	U(EC)		
1	Bridge A	2.77	139.35	39.70	0.86	0.98	1.00	0.89	0.94	0.9621	22
2	Bridge B	1.89	126.41	27.50	0.21	0.96	0.34	0.36	0.83	0.6649	11
3	Bridge C	2.15	115.67	25.57	0.57	0.94	0.62	0.74	0.81	0.7883	16
4	Bridge D	2.73	42.94	3.41	0.02	0.16	0.98	0.04	0.00	0.3358	2
5	Bridge E	2.00	68.16	12.40	0.53	0.67	0.47	0.71	0.48	0.5711	5
6	Bridge F	2.12	149.21	47.89	0.23	0.99	0.60	0.39	0.97	0.7897	15
7	Bridge G	2.10	169.56	57.79	0.48	0.99	0.57	0.67	1.00	0.8272	18
8	Bridge H	2.42	88.60	13.11	1.25	0.85	0.83	0.99	0.51	0.7763	14
9	Bridge I	2.22	45.82	35.89	1.26	0.24	0.68	1.00	0.92	0.6454	9
10	Bridge J	2.34	115.93	30.80	0.43	0.95	0.78	0.62	0.87	0.8325	20
11	Bridge K	2.42	39.42	12.69	0.23	0.04	0.83	0.40	0.49	0.4284	4
12	Bridge L	2.46	69.61	12.12	0.03	0.69	0.85	0.05	0.47	0.5886	7
13	Bridge M	1.92	38.14	7.99	0.03	0.00	0.38	0.05	0.28	0.1836	1
14	Bridge N	2.18	84.89	14.42	1.05	0.82	0.65	0.95	0.55	0.7233	13
15	Bridge O	2.43	46.89	4.59	0.00	0.27	0.84	0.00	0.08	0.3456	3
16	Bridge P	1.67	175.33	28.51	0.68	1.00	0.00	0.81	0.85	0.6466	10
17	Bridge Q	2.08	161.48	55.25	0.37	0.99	0.56	0.56	0.99	0.8052	17
18	Bridge R	2.30	158.89	51.04	0.22	0.99	0.75	0.37	0.98	0.8301	19
19	Bridge S	2.58	65.90	8.79	0.10	0.64	0.92	0.18	0.32	0.5744	6
20	Bridge T	1.96	62.22	22.83	0.42	0.59	0.43	0.62	0.76	0.5894	8
21	Bridge U	2.02	84.82	25.70	0.28	0.82	0.50	0.45	0.81	0.6751	12
22	Bridge V	2.34	152.60	42.91	0.27	0.99	0.78	0.44	0.96	0.8411	21

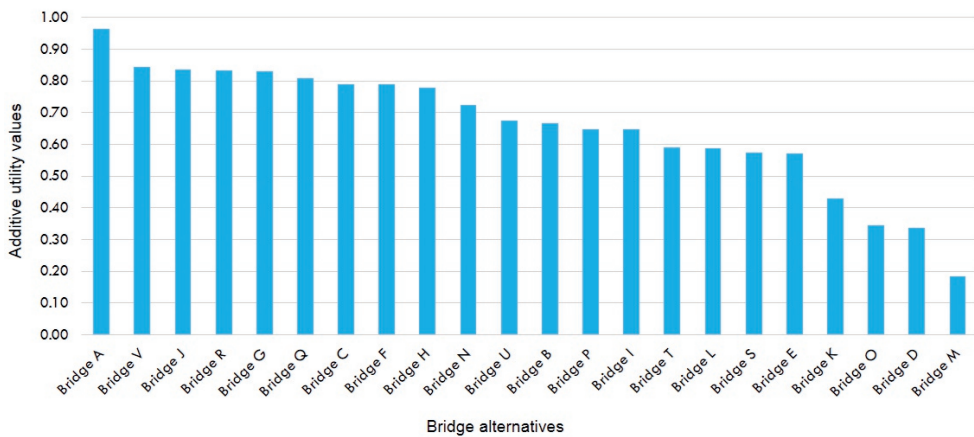


Figure 3.9: Ranking of the bridges for maintenance planning

enabling them to choose an indifference point between the best and worst possible value. It is important to mention that with the application of MAUT, the end goal is not to provide the ranking of alternatives but to transform the subjective decision-making process towards more objective ways where the utility values represent decision-makers' preferences. Notice that the higher rank have smaller additive utility scores. This is because of the minimization of objective functions, where the smaller values get the smaller utility score.

Bridge M is ranked highest based on the defined objectives, where the minimization of owner cost is most important following with minimization of condition index on the second number, user delay cost on third and finally environmental cost. It is interesting to notice the ranking of *Bridge A* and *Bridge P*. It could have been assumed that bridges having the highest owner cost will be ranked at the lowest. However, *Bridge P* having the highest owner cost is ranked at number 10, similarly, *Bridge A* is ranked at the lowest rank (i.e., 22) but does not have the highest owner cost. This is because of the relative importance of the attributes as shown in Table 3.5, where the condition index is the second most important attribute for prioritization of the bridges. It is also noted that Bridges (G, J, R, and V) are ranked at number 18-21 have a minimal difference in their aggregated additive values. The main reason is that these bridges have a condition index score in the range of 2.10-2.34. Such a small difference in values ends up providing the very close additive utility scores to the bridges alternatives. Figure 3.9 depicts the ranking of all the bridges with the obtained utility scores where a lower utility score represent the higher rank.

3.6 Discussion and Conclusions

This paper provides the proof-of-concept to apply the MAUT for optimization of network-level bridge maintenance planning. This is illustrated by considering a sample of twenty-two randomly chosen bridges from the Netherlands road network. MAUT provides a systematic procedure to transform the subjective preferences of infrastructure managers into objective values while optimizing multiple objectives during the maintenance planning process. The approach of MAUT illustrated in this paper is not aimed to replace the existing maintenance planning tools, but it can be an add-on function that is able to facilitate the further optimization of limited resources based on the goals of maintenance (e.g., minimize cost, maximize reliability).

As illustrated, MAUT yields the ranking of the bridges by computing exponential utility functions and their final aggregation by capturing the preference uncertainty and

risk tolerance of decision makers. For the case presented in this study, the ultimate ranking is based on a minimization function of condition index, owner cost, user delay cost, and environmental cost, where owner cost is rated most important. Based on the attributes scaling factors and the defined objectives, the ranking of the bridges can be different for the same values. This is because the values assigned by utility functions become higher or lower based on the weights assigned to the attributes to represent their relative importance. The final ranking obtained as a result of an aggregation of utility functions suggest that a bridge with higher rank contributes the most in the realization of defined performance goals.

Though the proof-of-concept expresses the usefulness of MAUT for the network-level bridge maintenance planning, there are also certain limitations to consider. It is noted that the range of data values or data distribution used for MAUT exercise plays a fundamental role. The attributes having a minimal difference between the maximum and minimum scores will yield final utility scores which are very similar to each other (e.g., See Table 3.6 Bridges G, J, R, V). These similar scores, though still can be ranked, make it difficult for a decision maker to justify his/her choices objectively. Besides, the procedure to calculate the utility function is complex and can be time-consuming when the number of alternatives is large in number. This can be solved by employing computation aids where a computer program can compute the utility functions. Another limitation is due to the changing values of cost, condition index with time. This means the ranking defined by MAUT is valid for a certain period of time during which the performance indicators values are considered as constant.

The future research efforts seek to mitigate these limitations by developing a computer-aided program that will minimize the tasks of manual implementation of MAUT and enable the ranking of a rather large number of alternatives based on defined objectives. Moreover, an extension of this case for through network-level maintenance planning is also part of future work, in which a number of different performance indicators e.g., bridge's age, location, the impact of surrounding objects, the possibility to cluster maintenance activities by type or locations, will be considered.

3.7 References

- Arif, F., Bayraktar, M. E., & Chowdhury, A. G. (2015). Decision support framework for infrastructure maintenance investment decision making. *Journal of Management in Engineering*, 32(1). doi: 10.1061/(ASCE)ME.1943-5479.0000372.
- Bush, S., Henning, T., Ingham, J. M., & Raith, A. (2014). An agent-based framework for improved strategic bridge asset management. In *Austrroads bridge conference, 9th, 2014, sydney, new south wales, australia* (4.4).
- Chase, S. B., Adu-Gyamfi, Y., Aktan, A., Minaie, E., et al. (2016). *Synthesis of national and international methodologies used for bridge health indices*. United States: Federal Highway Administration.
- Chen, K., & Miles, J. C. (2004). *ITS handbook 2004: Recommendations from the world road association (PIARC)*. MA Boston: Artech House 2000.
- Chung, J. W. (1994). *Utility and production functions: Theory and applications*. Blackwell Pub.
- Claudio, D., & Okudan, G. E. (2009). Utility function-based patient prioritisation in the emergency department. *European Journal of Industrial Engineering*, 4(1), 59–77.
- COST Action TU 1406. (2015). Quality specifications for roadway bridges, standardisation at a european level [project]. Retrieved from: <http://www.tu1406.eu/about-tu1406/the-project>.
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., Garcez, T. V., et al. (2015a). *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis*. New York, NY, USA: Springer, 2015.
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., & Garcez, T. V. (2015b). Multiobjective and multicriteria problems and decision models. In *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis* (pp. 1–22). Springer.
- de Almeida, A. T., Ferreira, R. J. P., & Cavalcante, C. A. V. (2015). A review of the use of multicriteria and multi-objective models in maintenance and reliability. *IMA Journal of Management Mathematics*, 26(3), 249–271.
- de Bruyn, S., Korteland, M., Markowska, A., Davidson, M., de Jong, F., Bles, M., & Sevenster, M. (2010). *Valuation and weighting of emissions and environmental impacts*. Shadow Prices Handbook.
- Deterne, J. (2010). *Report on BEXPRAC*. Conference of European Directors of Roads.

- Dong, Y., Frangopol, D. M., & Sabatino, S. (2015). Optimizing bridge network retrofit planning based on cost-benefit evaluation and multi-attribute utility associated with sustainability. *Earthquake Spectra*, 31(4), 2255–2280.
- Figueira, J. R., Greco, S., Roy, B., & Slowinski, R. (2013). An overview of ELECTRE methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2), 61–85. doi: 10.1002/mcda.1482.
- Frangopol, D. M., & Liu, M. (2007). Bridge network maintenance optimization using stochastic dynamic programming. *Journal of Structural Engineering*, 133(12), 1772–1782. doi: 10.1061/(ASCE)0733-9445.
- Garmabaki, A. H. S., Ahmadi, A., & Ahmadi, M. (2016). Maintenance Optimisation Using Multi-attribute Utility Theory. In U. Kumar, A. Ahmadi, A. K. Verma, & P. Varde (Eds.), *Current trends in reliability, availability, maintainability and safety: An industry perspective* (pp. 13–25).
- Hegger, S., & de Graaf, D. (2013). *VERGELIJKENDE LCA STUDIE BRUGGEN [LCA COMPARATIVE STUDY OF BRIDGES]*. BECO, Rotterdam, Netherlands.
- Hens, T., & Rieger, M. O. (2010). *Financial economics: A concise introduction to classical and behavioral finance*, by t. hens and mo rieger. Springer-Verlag Berlin Heidelberg.
- Hooks, J., & Frangopol, D. (2013). *LTBP Bridge Performance Primer, Report No. FHWA-HRT-13-051*. Federal Highway Administration, Washington, DC.
- Ishizaka, A., & Nemery, P. (2013). *Multi-attribute utility theory. Multi-Criteria Decision Analysis: Methods and Software*, 81–113.
- Ismaeel, M. (2016). Performance based budget allocation model for water networks.
- Jano-Ito, M. A., & Crawford-Brown, D. (2017). Investment decisions considering economic, environmental and social factors: An actors' perspective for the electricity sector of Mexico. *Energy*, 121, 92–106. doi: 10.1016/j.energy.2017.01.016.
- Keeney, R., & von Winterfeldt, D. (2007). M13 Practical Value Models. In M. R. W & von Winterfeldt D (Eds.), *Advances in decision analysis: From foundations to applications* (pp. 232–252). Cambridge University Press, New York.
- Keeney, & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press.
- Kouwenhoven, M., de Jong, G. C., Koster, P., van den Berg, V. A., Verhoef, E. T., Bates, J., & Warffemius, P. M. (2014). New values of time and reliability in passenger transport in the Netherlands. *Research in Transportation Economics*, 47, 37–49. doi: 10.1016/j.retrec.2014.09.017.
- Krishnamurthy, S. (2006). Normative decision analysis in engineering design. *Decision Making in Engineering Design*, 4(4), 21–33.
- Levy, H., & Markowitz, H. M. (1979). Approximating expected utility by a function of mean and variance. *The American Economic Review*, 69(3), 308–317.

- Meyer, J. (2010). Representing risk preferences in expected utility based decision models. *Annals of Operations Research*, 176(1), 179–190. doi: 10.1007/s10479-008-0381-7.
- Middleton, M. R. (2006). *Decision analysis using excel*. University of San Francisco, CA.
- Min, H. (1994). International supplier selection: A multi-attribute utility approach. *International Journal of Physical Distribution & Logistics Management*, 24(5), 24–33.
- Patidar, V. (2007). *Multi-objective optimization for bridge management systems*. Transportation Research Board.
- Pujadas, P., Pardo-Bosch, F., Aguado-Renter, A., & Aguado, A. (2017). MIVES multi-criteria approach for the evaluation, prioritisation, and selection of public investment projects. a case study in the city of barcelona. *Land Use Policy*, 64, 29–37. doi: 10.1016/j.landusepol.2017.02.014.
- Rashidi, M., & Gibson, P. (2011). Proposal of a methodology for bridge condition assessment. *Australasian Transport Research Forum*, 1–13.
- Saaty, T. L. (2000). *Fundamentals of decision making and priority theory with the analytic hierarchy process*. RWS publications.
- Stewart, N., Reimers, S., & Harris, A. J. (2014). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science*, 61(3), 687–705.
- Stipanovic, I., Høj, N., & Klanker, G. (2016). Multi-objective bridge performance goals. In *Life-cycle of engineering systems: Emphasis on sustainable civil infrastructure* (pp. 280–288). CRC Press.
- Stratt, R. W. (2010). Bridge Management a System Approach for Decision Making. *School of Doctoral Studies European Union Journal*, (2).
- Strauss, A., Ivankovic, A. M., Matos, J. C., & Casas, J. (2016). Wg1 technical report: Performance indicators for roadway bridges of cost action 1406.
- Sun, B. (2016). *Operational decision making for medical clinics through the use of simulation and multi-attribute utility theory*. Doctoral dissertation, University of Louisville, Louisville, Kentucky.
- Tao, Z., Zophy, F., & Wiegmann, J. (2000). Asset management model and systems integration approach. *Transportation Research Record: Journal of the Transportation Research Board*, (1719), 191–199.
- Thevenot, H. J., Steva, E. D., Okudan, G. E., & Simpson, T. W. (2006). A multi-attribute utility theory-based approach to product line consolidation and selection. *International Design Engineering Technical Conferences*, 441–450. doi: 10.1115/DETC2006-99506.

- Thinkstep. (2015). Description of the CML 2001 method. Available on: <http://www.gabi-software.com/support/gabi/gabi-lcia-documentation/cml-2001/>.
- Tighe, S., Eng, P., & McCabe, B. (2005). Incorporating User Delay Cost in Project Selection: A Canadian Case Study. *1st Annual Inter-University Symposium on Infrastructure Management*.
- TNO-MEP. (2004). *Toxiciteit heeft zn prijs [Toxicity has its price]*. Rijkswaterstaat, Ministry of Transport and Water management, Netherlands.
- Van Dam, K. H., Nikolic, I., & Lukszo, Z. (2012). *Agent-based modelling of socio-technical systems*. Springer Science & Business Media.
- Von Neumann, J., & Morgenstern, O. (1945). *Theory of games and economic behaviour*. Princeton University Press Princeton, NJ.
- Wang, Y., & Goodrum, P. M. (2005). Use of conceptual road user costs for a rapid roadway construction decision making system. In *Construction research congress 2005: Broadening perspectives* (pp. 1–11).
- Wardman, M., Chintakayala, P., de Jong, G., & Ferrer, D. (2012). European wide meta-analysis of values of travel time. *University of Leeds, Paper prepared for EIB, ITS*.

Multi-year maintenance planning framework using MAUT & genetic algorithms

¹ **Abstract:** This paper introduces a comprehensive framework for the development of optimal multi-year maintenance plans of a large number of bridges. A maintenance plan is said to be optimal when within the given budget, a maximum number of bridges is maintained in the best possible year, achieving maximum performance and minimum impact on the economy and society. The framework incorporates heuristic rules, multi-attribute utility theory, discrete Markov chain process, and genetic algorithms to find an optimal balance between limited budgets and performance requirements. The applicability of the proposed framework is illustrated on a case study of 869 highway bridges. The framework enables asset owners to execute various planning scenarios under varying budget and performance requirements where each resulting plan is optimal. The focus of this study has mainly been on the highway bridges, but the proposed framework is general and can be applied to any other infrastructure assets types.

¹This chapter is under-review as: Allah Bukhsh, Z., Stipanovic, I., A., & Doree, A. G.. Multi-year maintenance planning framework using multi-attribute utility theory and genetic algorithms.

4.1 Introduction

Bridges are one of the most fundamental structures of the transportation network. They provide the crossings at critical locations, reduce the travel times, and maintain the traffic flow. Due to their fundamental importance, the bridges must be maintained up to a specific performance level throughout their whole life cycle. Since most of the road infrastructure in Europe was built after the 1950s, the road bridges are reaching their critical age, while being exposed to adverse climate effects, and increased public demands. Under limited financial resources, agencies have to take careful investment and maintenance planning decisions to improve the serviceability and availability of the bridges, to minimize their life-cycle cost, and to maximize the return on investments.

The multitude of maintenance planning and optimization solutions have been proposed (Ding & Kamaruddin, 2015; Garg & Deshmukh, 2006; Sharma, Yadava, & Deshmukh, 2011). Most of these approaches are focused on individual bridge level, while few consider maintenance planning at the network level. Within the theme of maintenance optimisation, the literature studies can be divided into three groups: studies that estimate the remaining useful lifetime of bridges by deterioration modelling only (e.g., see review by Alaswad and Xiang, 2017), studies that compare the different maintenance alternatives to find the most optimal one (Ghodoosi, Abu-Samra, Zeynalian, and Zayed, 2017; Hu and Madanat, 2014; Kong and Frangopol, 2003; Liu and Frangopol, 2004), and studies that model probabilistic hazards function for multi-objective optimization using integer and dynamic programming techniques Barone and Frangopol, 2013, 2014; Dong and Frangopol, 2016; Kim and Frangopol, 2017). Though these advance optimization approaches are promising, yet they are computationally expensive and are complicated to comprehend and implement by assets owners. Many agencies use the Bridge Management Systems (BMS) to develop maintenance plans by allocating the budget, which is still mainly driven by subjective ranking and preferences of domain experts (Markow & Hyman, 2009; Mirzaei, Adey, Klatter, & Kong, 2012). These systems typically employ single-objective optimization analyses to allocate budget by Life Cycle Cost (LCC) analysis but do not take into account other performance aspects related to the economy, society, and environment (Bush, Henning, Ingham, & Raith, 2014).

To assist agencies with accurate yet simple maintenance planning solutions constituting multiple performance goals and resource constraints, the search-based optimization techniques like Genetic Algorithms (GAs) are popular. A brief overview

of the notable studies that employ GA for the maintenance planning is provided: Morcous and Lounis (2005a) presented an approach to determine the optimal set of maintenance alternatives using GAs. Lee and Kim Sung (2007a) proposed an algorithm to prioritize maintenance activities for bridges deck at the network level using multi-objective optimization. Bocchini and Frangopol (2011) presents a probabilistic framework to schedule preventive maintenance on the bridges with the focus on reliability assessment and Pareto optimization. A two-stage maintenance planning method for a large number of bridges is presented by Furuta, Nakatsu, Ishibashi, and Miyoshi (2014). In the first stage, the preventive maintenance plan of a single bridge is optimized, while in the second stage, the total life-cycle cost is reduced by allocating flexible intervention intervals. Denysiuk, Fernandes, Matos, Neves, and Berardinelli (2016) proposed a computational framework consisting of degradation and maintenance models along with multi-objective optimization. The purpose of the framework is to search for optimal maintenance schedules. Similarly, Xie, Wu, and Wang (2018) proposed a framework to find the optimal initial and consecutive time intervals between maintenance activities in order to minimize the life cycle environmental impact.

The literature provides numerous maintenance planning approaches which vary by the employed degradation models and optimization techniques. Some of past studies have been the source of motivation of this study (Chootinan, Chen, Horrocks, & Bolling, 2006; Frangopol, Dong, & Sabatino, 2017; Ghodoosi et al., 2017). However, it is noted that the search-based optimization solution considers the multiple objectives of maintenance planning but neglect to scrutinize the subjective preferences of the agencies/decision-makers (de Almeida, Ferreira, & Cavalcante, 2015). Moreover, most of the planning solutions are illustrated on the small set of assets (i.e., bridges), where with the best of our knowledge no maintenance planning approach provides a comprehensive methodology to manage the scheduling of hundreds of network bridges in an end-to-end manner. This study aims to mitigate the aforementioned limitations by introducing a Multi-year Maintenance Planning Framework (MMPF) for bridges. The purpose of the MMPF is to find the best time for maintenance of bridges by developing an optimal schedule over a multi-year planning period. The resulting maintenance plan by MMPF must meet the multiple objectives of performance level and cost minimisation while considering the preferences of decision makers. The main contributions of this work are the following:

- Employing the multi-attribute utility theory to incorporate multiple-objectives and decision-makers preferences for the ranking of a large number of bridges.

- Applying Markov chain process in conjunction with genetic algorithms for the performance forecasting of bridges for each year.
- Application of the MMPF and developed numerical models on the data of 869 real road bridges.
- Introducing a comprehensive two-step multi-objective optimization for the maintenance planning based on current condition state only and the future predicted condition states.
- Enabling asset owners to execute various maintenance planning scenarios under varying budget and performance constraints

The particular emphasis of this study is on road-bridges, though, the proposed MMPF is general and can be applied to other infrastructure assets. The rest of the paper is structured as follows: Section 4.2 introduces the proposed MMPF along with details of used techniques and algorithms. Section 4.3 and 4.4 illustrates the application of MMPF on the case study data, outlines the implementation details, and analyze the numerical results. The general discussion of results, potential limitations of MMPF, and concluding remarks of the study are provided Section 4.5.

4.2 Methodology of Multi-year Maintenance Planning Framework

The development of optimal multi-year maintenance plan is a complicated problem, involving several aspects relating to the budget constraints, a large number of assets, the performance of assets over time, choice of maintenance treatment, impact on availability, and preference uncertainty of stakeholders. This section introduces the methodology of Multi-year Maintenance Planning Framework (MMPF), which aims to develop optimal maintenance schedules over the multi-year planning period.

Figure 4.1 presents a flowchart of MMPF to highlight the interaction of employed techniques and algorithms. The proposed MMPF mainly constitutes of four modules: i) an impact assessment module to decide on type of maintenance intervention and its resulting impact ii) a MAUT based module for ranking of bridges on the basis of preference uncertainty and risk attitude of a decision-maker(s), iii) an Markov Chain Process (MCP) based performance prediction module to forecast the condition of a

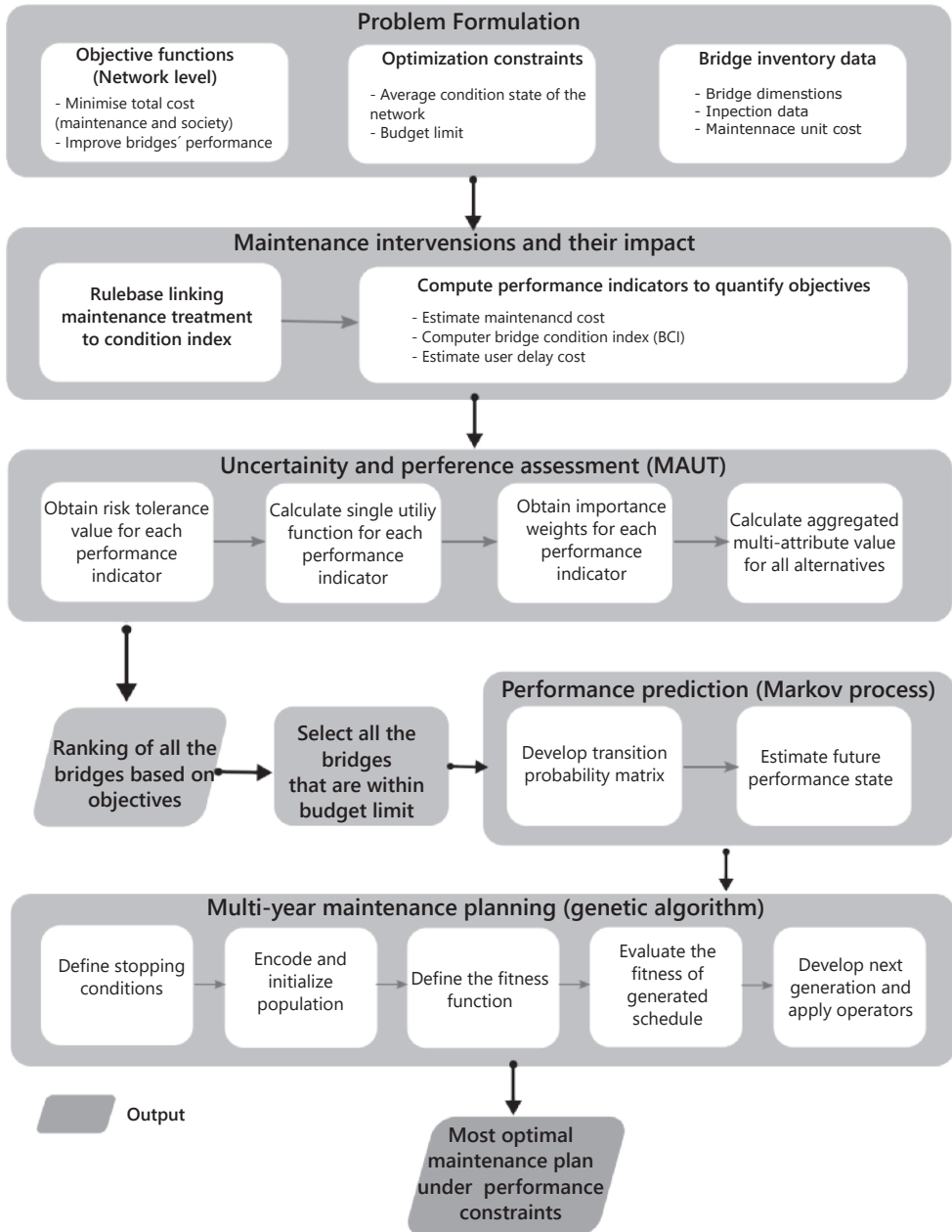


Figure 4.1: A flowchart of multi-year maintenance planning framework (MMPF).

bridge in future and finally iv) a module to apply the GAs to develop numerous multi-year maintenance plans in order to find the optimal solution. The optimal solution (maintenance plan) is defined as where bridges are allocated with maintenance treatments, and a best possible year for their execution, where no early maintenance leads to undue cost and no late maintenance activity poses safety risk. The optimal plan must also satisfy any defined objectives and constraints.

4.2.1 Problem formulation

The proposed framework considers multiple objectives of maintenance planning. The most dominant objectives are a) to maximize the performance level of bridges by minimizing their condition index and b) to minimize the maintenance cost by optimally planning the maintenance treatments. The objectives and constraints of the maintenance planning problem can be mathematically represented as:

$$\text{minimise } \sum_{t=1}^T \sum_{b=1}^B I(x_{t,b}) \quad (4.1a)$$

$$\text{minimise } \sum_{t=1}^T \sum_{b=1}^B c(x_{t,b}^m) \quad (4.1b)$$

$$\text{subject to } \sum_{t=1}^T \sum_{b=1}^B I(x_{t,b}) \leq \text{Threshold} \quad (4.1c)$$

$$\text{subject to } \sum_{t=1}^T \sum_{b=1}^B c(x_{t,b}^m) \leq \text{Budget} \quad (4.1d)$$

$$\text{where } I(x_{t,b}) = \begin{cases} I(x_{t,b}) - z, & \text{if } x_{bt} = m. \\ I(x_{t,b}) + f(d), & \text{otherwise, apply MCP.} \end{cases} \quad (4.1e)$$

The aforementioned parameters used for modeling are defined as follows:

B = Number of bridges for maintenance planning

T = Number of years in planning horizon

$I(x_{b,t})$ = Condition index of a bridge b in year t

$c(x_{b,t}^m)$ = Cost function to compute cost of maintenance activity m performed on a bridge b in year t

Threshold = Average condition index threshold of all the bridges of the network

z = Improved condition state of a bridge after maintenance/
remediation measure (based on Table 4.1)

$f(d)$ = Function based on Markov chain process (MCP) to forecast the
condition of a bridge

With this formulation, Equation 4.1a represents the first objective function, which aims to minimize the condition score I of all the bridges B over the finite planning period T . The minimization function is applied because, in a discrete condition scorecard, 1 represents the best (as new) condition state, whereas 6 represents the worst condition (loss of function). If assigned with a maintenance activity in a year, the state of the bridge will be improved (represented as z); otherwise, it is deteriorated with certain degree $f(d)$ complying to Markov chain transition matrices represented in Equation 4.1e. Equation 4.1b represents the second objective function, which is targeted for the minimization of maintenance cost resulted from m maintenance activity. Both objectives must be reached within the two main constraints. Firstly, all the bridges must be above the specified performance threshold value as represented in Equation 4.1d. This condition enforces the allocation of maintenance activity to bridges having poor condition states, irrespective of their high rehabilitation cost. Secondly, the accumulated costs of planned maintenance activities on the selected bridges must be within the specified budget limit denoted by Equation 4.1c. Currently, the impact of maintenance activities on the users is not set as an objective with the fixed value. Instead, it is used as the criteria to minimize the impact on the users by determining the user delay costs (explained in the following subsection) resulting from maintenance activities, while accounting for bridge importance in terms of the number of autos per day, the extended travel times, and duration of different maintenance options.

4.2.2 Maintenance intervention and quantification of its impact

Different maintenance interventions are performed on the bridges to preserve their state. An optimal planned maintenance strategy of a bridge can result in longer service life and reduced life-cycle costs. Since transport infrastructure is deeply embedded in society, it is not only subject to technical requirements, but the societal and economic impacts must also be contemplated. Therefore, an optimal maintenance strategy shall not only improve the performance level of a bridge but should also consider the societal and economic consequences. In this section, first, we introduce heuristic to link condition states with various maintenance treatments and secondly, we provide a

procedure to quantify the socio-economic impact of maintenance activity in terms of bridge performance state, maintenance costs and user delay costs.

Condition state and maintenance treatments

Due to the environmental impact, aging and traffic loading, bridges are deteriorating and experiencing various damages. The existence and severity of these damages define the condition state of a bridge, as well as the type of treatment. Table 4.1 provides a simplified rule-based choice of maintenance treatments related to the Bridge Condition Index (BCI). We consulted following studies (Ghodoosi et al., 2017, Section 3.5), (Morcous & Lounis, 2005b, Table 1), (Lee & Kim Sung, 2007b, Table 1 and 2) and (ERA-NET ROAD, 2012, p. D03) to define the respective maintenance activities, their impact, and associated condition ranges.

Table 4.1: Heuristics defining maintenance treatments based on condition state ranges (ERA-NET ROAD, 2012, p. D03).

Condition range (BCI)*	Treatment name	Treatment details	Impact on BCI
1 - 2	Nothing	No action needed	No change
2 - 2.7	Monitoring	Monitoring and inspection	No change
2.7 - 3.4	Minor intervention	Recoating (20%) Equipment repair (10%) Repair (10%)	Condition - 0.5
3.4 - 4.1	Medium intervention	Recoating (100%) Equipment repair (50%) Repair (20%) Renewal (10%)	Condition - 1.5
4.1 - 5	Major intervention	Recoating (100%) Equipment repair (100%) Repair (20%) Renewal (20%)	Condition - 2.5
5 - 6	Replacement	Complete replacement	1

*1 : Good as new condition state, 6: critical condition state (failure)

Performance indicators quantification

The performance indicators quantify the objectives of maintenance planning. The value of these performance indicators represents the impact of maintenance activity. For instance, the poor condition states lead to increased maintenance cost, higher user delay, and an additional CO2 due to traffic disruptions. Since bridges have multiple components with distinct damage type, many road agencies inspect the bridges on their component level (Chase, Adu-Gyamfi, Aktan, Minaie, et al., 2016). However, the maintenance decisions must be made at the system-level (as a whole bridge);

therefore, the components conditions must be translated to represent the overall health of a structure.

This section provides a procedure to quantify the impact of maintenance interventions in terms of component-level to system-level BCI, maintenance cost, and user delay cost.

- *Component-level to system-level BCI:* There are numerous ways, varying by country and agency, to compute the BCI, namely, worst element score, weighted average method, ratio scale (Chase et al., 2016; Hooks & Frangopol, 2013; Patidar, 2007; Sterritt, 2002; Stratt, 2010). This study applies a weighted-average method for calculation of BCI. In the weighted-average method, an expert establishes the importance of each component to another by considering their failure history, frequency of maintenance, and the consequences on the overall bridge in case of failure. The bridge as a structure has the following essential components, i.e., Superstructure, Bearing, Abutment, Expansion joints, Deck (pavement), and Railing. The relative importance score of each component is derived on the bases of an expert’s preference and is provided in Table 4.2. Once the relative importance of each element is obtained, the BCI is calculated by computing the following Equation:

$$BCI = \sum_{i=1}^n (CI_i * W_i) \tag{4.2}$$

where n represents number of bridge components, CI_i is condition index of component i , and W_i is weighted importance of component i elicited by an expert as shown in Table 4.2.

Table 4.2: Weighted score of bridge components (Allah Bukhsh, Stipanovic, Klanker, O’Connor, & Doree, 2018).

No.	Elements	Weighted score
1	Superstructure	0.3185
2	Bearings	0.2104
3	Abutment	0.1813
4	Joints	0.1288
5	Pavement	0.0618
6	Railing	0.0510
7	Guardrail	0.0478

- *Maintenance cost:* The maintenance cost is a monetary value borne by an agency as a result of maintenance activities. The maintenance cost of each bridge is computed by a sum of the unit costs of all the maintenance treatments (i.e., UCA), performed on a bridge in a year, multiplied by the quantity (i.e., Q) or the volume of the treated area. The formula to calculate the maintenance cost is provided as follows:

$$MC_y = \sum_{m=1}^n UCA_m * (Q) \quad (4.3)$$

where MC_y is total maintenance cost spend on a bridge in a year and n is number of bridge's component.

- *User delay cost:* The user delay cost estimates the impact of a maintenance treatment on the availability of the bridge and highlights the bridge's importance in the network. It represents the value of extended travel time of the road users due to work-zones in a monetary form. Number of factors contributes to user delay cost (Tighe, Eng, & McCabe, 2005; Wang & Goodrum, 2005) i.e., a) imposed speed restriction expressed in extra travel time (ETT) b) number of users affected defined by traffic intensity in terms of average traffic per hour passing over the bridge on a working day (AHT_h), c) the value of an hour of the user time (V_{oh}), and d) finally the duration of the maintenance activity (D_h). Mathematically, UDC is presented as:

$$UDC = ETT * AHT_h * V_{oh} * D_h \quad (4.4)$$

$$ETT = \frac{L}{V_r} - \frac{L}{V_s}$$

where L is length of working zones (km), V_s is standard velocity (km/h) and V_r is reduced velocity due to maintenance.

4.2.3 Uncertainty and preference assessment using MAUT

This module incorporates the preference uncertainty and risk attitude of experts, for maintenance planning, by employing principles of Multi-Attribute Utility Theory (MAUT). The purpose of this module in overall multi-year planning is threefold (see Figure 4.1): Firstly it considers the multiple performance indicators, ranging from societal to economic aspects instead of only the condition state of a structure. Secondly,

due to MAUT ability to capture probabilistic consequences, it enables the decision makers to explicitly state their maintenance preferences, in terms of cost, condition, user delay of a large number of bridges under uncertainty. Lastly, MAUT assigns ranks to each bridge by incorporating decision makers preferences and by performing a trade-off of multiple performance objectives. The resulting ranking facilitates to filter out the broad set of bridges from the road network that does not fulfill the objectives, therefore must not be scheduled for maintenance in the near future. The ranking can also be utilized for the next year maintenance plan as the highest ranked bridges present the urgent need of maintenance prompted by either low cost or poor condition state. In the following, a brief explanation of MAUT process is provided.

Utility theory provides a measure of preferences of a decision maker over related attributed and a group of alternatives (Ishizaka & Nemery, 2013). Keeney and Raiffa (1993) introduced MAUT based on axioms of expected utility theory. It provides a systematic approach to reduce the qualitative values of various attributes (i.e., performance indicators) into utility functions. It involves a decision maker who is willing to make certain trade-offs among the performance indicators (e.g., BCI, maintenance cost, user delay cost) while considering uncertainty and risk. The uncertainty is primarily originated from the unavailable and dynamic nature of data, and varying experience levels of decision-makers. For instance, in the bridge planning, it is difficult to estimate the number of users affected due to maintenance activity. The risk attitude is recorded by a lottery question, where a decision-maker aims to maximize his gain by taking some risk. For instance, in order to have reduced maintenance cost, a decision maker is willing to compromise on the maximum performance improvement.

The application process to implement the MAUT is briefly outlined below (Keeney & Raiffa, 1993):

1. This study applies exponential utility function to elicit the utility scores of each attribute since it captures the preference uncertainty and risk tolerance (i.e., RT) of a decision maker (Keeney & von Winterfeldt, 2007). The formula to calculate the utility scores U_a of each attribute a for the alternatives x is provided below:

$$U_a(x_a) = A - B * e^{\left(\frac{-x_a}{RT}\right)} \quad (4.5)$$

where A and B are scaling constants and e is an exponential constant. The exponential utility function requires the decision makers to respond to the

lottery question of a maximum and a minimum value of an attribute where the indifference point has to be reached between the best and the worst possible outcome. Such an indifference point for a decision maker is referred to as the *Certainty Equivalent (CE)*.

2. Calculate the risk tolerance based on *expected value (EV)* and *certainty equivalent (CE)* where *EV* is median of worst and best value of an attribute values and the value *CE* is chosen by the experts based on following principle:

$$\text{Risk Attitude} = \begin{cases} \text{Risk Neutral, if } EV = CE \text{ (Linear shaped)} \\ \text{Risk Avoiding, if } EV \geq CE \text{ (Concave shaped)} \\ \text{Risk Taking, if } EV < CE \text{ (Convex shaped)} \end{cases}$$

3. Perform trade-offs among attributes *a* by assigning the relative importance *k* based on the decision-makers preferences.

$$k_a = \frac{rate(x_a)}{\sum_{a=1}^A(x_a)} \quad (4.6)$$

where *rate(x_a)* is a weight assigned by a decision maker to attribute *a*, $\sum_{a=1}^A(x_a)$ is sum of all the weights given to attributes and *k* represents attribute's relative importance.

4. Aggregate the utility score of each attribute and their assigned weight factor to elicit the final ranking of the alternatives by computing additive aggregation as follows:

$$U(x) = \sum_{i=1}^n k_a U_a(x_a) \quad (4.7)$$

where *k_a* is the relative importance of attribute, *U_a(x_a)* is single attribute utility for each attribute *a* for an alternative *x*.

The computed MAUT score establishes the ranking of each bridge based on the objectives. In our case we have defined two objectives, namely, to minimize the total costs (maintenance and user delay costs) and to maximize the performance of the bridge (minimise condition states). Therefore, the bridge having smaller aggregative MAUT score will attain higher over ranking.

4.2.4 Performance prediction using Markov chain process

It is essential to forecast the performance of the asset in the future in order to optimally plan the maintenance activities and to estimate the cost. Based on the mathematical and statistical principles, multiple deterministic and stochastic models have been proposed for the deterioration modelling (Ahmad & Kamaruddin, 2012; Alaswad & Xiang, 2017). For this study, the discrete Markov Chain Process (MCP) are applied to statistically model the random and uncertain process of bridge deterioration as already illustrated by several studies (Cesare, Santamarina, Turkstra, & Vanmarcke, 1992; Grinstead & Snell, 2012; Robelin & Madanat, 2007).

Markov chain consists of set of states denoted as $S = s_1, s_2, \dots, s_n$, where each state represents the bridge condition state. The process starts from one state s and moves to another state s' , with a certain probability of p . The probabilities p are called transition probabilities, which quantify the probability of a bridge or its component to move from one state s to another state s' . There are multiple procedures to compute the transition probability matrices, namely expected-value method (Jiang, Saito, & Sinha, 1988), binomial regression (Butt, Shahin, Feighan, & Carpenter, 1987; Madanat & Ibrahim, 1995), and ordered probit model (Madanat, Mishalani, & Ibrahim, 1995). The simplest technique, requiring minimum data, is percentage prediction method, which forecasts the percentage of total bridges belonging to each condition state (Setunge & Hasan, 2013). The calculation procedure can be represented as:

$$p(s, s') = \frac{\sum_{n=1}^N n_{ss'}(s, s')}{\sum_{n=1}^N n_s(s)} \quad (4.8)$$

where $n_{ss'}$ is number of components transitioning from condition state s to state s' and n_s is total number of components having state s .

The computed transition probabilities are expressed in a matrix of size $w \times w$, where w is the number of discrete condition state. The matrix of transition probabilities is expressed as P . Based on the Chapman-Kolmogorov equation, the probability of a bridge moving from state s to s' after t periods can be estimated by multiplying P with itself t times (Baik, Jeong, & Abraham, 2006). Let us assume $p(0)$ represents the current state of the bridge at time 0. Then, the state vector $P(t)$ represents the probability that a bridge is in state s' after t transitions. It is expressed as:

$$P(t) = p(0) * P^t \quad (4.9)$$

The performance prediction of the bridges using the MCP is used to decide on a specific year for the maintenance execution. As the early year maintenance will lead to undue cost spending, while the late maintenance can pose the safety risk. Despite the popularity of MCP for deteriorating modeling, the computation of reliable transition probabilities matrix is a tricky task (Bu, Lee, Guan, Blumenstein, & Loo, 2012).

4.2.5 Multi-year maintenance planning using genetic algorithms

The end goal of the maintenance planning framework is to develop optimal multi-year maintenance schedules, that minimizes the life cycle costs of a bridge while fulfilling the performance requirements over a specified planning horizon. The development of maintenance plan is a multi-objective combinatorial optimization problem involving budget limits, performance requirements, planning horizon, number of alternatives (bridges) and their characteristics, different maintenance treatments and their societal and economic impact. Instead of employing traditional optimization techniques, which are complicated and time-consuming, we choose to develop Genetic Algorithms (GAs) based solution that is efficient and provides robust results. GA is a combinatorial optimization search technique motivated by Darwinian evolution theory of natural selection, genetics and survival-of-the-fittest (Holland, 1975).

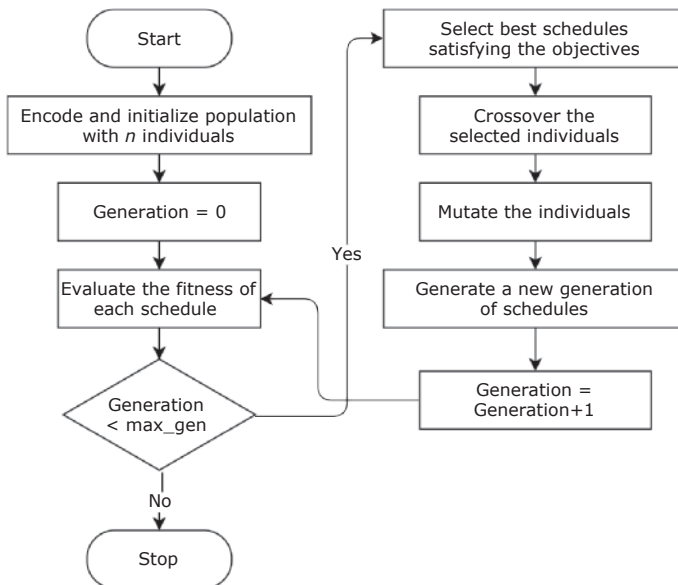


Figure 4.2: Flowchart of simple genetic algorithm.

In GA, the potential solutions are represented as *individuals*, which consist of combinations of *chromosomes*. The *chromosomes* are finite-length strings which represent the decision variable of a search problem. Unlike traditional optimization approaches, GA generates a population of potential solutions iteratively until the pre-determined population size is reached. Each individual of the population must be evaluated by *objective function* defined either based on a mathematical model or subjective rules in order to distinguish good solutions from bad solutions. Further stochastic operators i.e. *cross-over*, *mutation* are applied on selected population based on their relative fitness score assigned by *objective function* to generate next population. Figure 4.2 provides a flow chart of simple GA.

There are different encoding schemes, e.g. binary encoding, tree-based encoding, value encoding, permutation encoding to represent the optimization problem to GA search procedure (Kumar, 2013). Depending on the problem, the encoding scheme can be selected. Our optimisation problem is encoded in value scheme since the fitness of a generated schedules depends only on the discrete values i.e. cost and condition irrespective of a specific order. To evaluate each of the generated solution of GA, the objective function is defined based on the objectives and constraints introduced in subsection 4.2.1.

4.3 Case Study

The applicability of the proposed MMPF is validated with a case study of 869 concrete bridges of a road network. The data is provided by a real road agency, with a confidentiality agreement. The bridges database contains data about location, bridge structure, materials, component-level condition scores resulting from principle inspections, records of performed and planned maintenance treatment (s) and their unit costs. The agency uses condition data, damage estimation, and expert (subjective) judgment implemented into risk assessment to decide on future maintenance plans. Considering that the existing procedure still heavily relies on subjective judgments, it is not fully transparent and followable. We have developed the MMPF which primarily utilizes existing condition score data and can be applied within several agencies.

A five-year maintenance plan is programmed with the aim to keep an average network-level bridge score of at least 2.7 (see Table 4.1) with the estimated budget of €6 millions only. The objective is to improve the performance of the bridges by minimizing the impact on the society and economy within a limited budget (See Equations 4.1a and 4.1b). We have considered five treatment options, namely i) monitoring, ii)

minor intervention, iii) medium intervention, iv) major intervention, and v) complete replacement. The available treatment options are linked to the range of BCI as shown in Table 4.1.

As mentioned earlier, the bridge as a structure is divided into seven components, i.e., Abutment, Bearing, Railing, Expansion joints, Deck, Guardrail, and Superstructure. Given these components, each bridge has seven visually assessed indexes, recorded in the last 12 years, which represent the structural performance of each bridge component. To have a thorough understanding of bridge structural integrity and its deterioration over time, we will first calculate the component-level transition probabilities, which will then be aggregated by their weighted importance to obtain a system-level transition probability. The following section provides the calculation procedure to compute transition probabilities for the components of the bridge. Next, the data used to obtain the performance indicators is discussed in detail.

4.3.1 Component to system-level transition probability matrix

The transition probabilities of each component are calculated by percentage prediction method (See Equation 4.8). The percentage prediction method mainly relies on the number of changes in condition state between the inspection interval. We have considered only those components' data that shows the deteriorating trend in condition state, which means no maintenance has been performed between the two inspection activities. The transition probability is represented as $p(s, s')$, where $s \leq s'$. For the computation of transition probabilities, we accounted for principal inspection only, which is executed every five years. By using Equation 4.8, the transition probability matrix for each bridge component is computed and provided in Table 4.3.

The leftmost column represents the current condition, and the top row shows the future condition states. So, a transition probability matrix of bearings can be read as there is 0.87 probability of a bearing having condition state 1 to traverse to condition state 2 during the inspection interval of 5 years. The diagonal of the matrices represents the probability of a component of staying in its current condition state. Two key aspects can be noted from all the component-wise transition probability provided in Table 4.3.

- First, except for joints and railing, none of the components reach to poor states represented by 5 and 6 indexes. This means a repair or rehabilitation is applied in an early stage of deterioration and bridges are well-maintained.

(a) Superstructure							(b) Bearings						
	1	2	3	4	5	6		1	2	3	4	5	6
1	0.03	0.86	0.07	0.03	0.01	0	1	0.03	0.87	0.06	0.03	0	0
2	0	0.86	0.05	0.06	0.03	0	2	0	0.92	0.06	0.02	0	0
3	0	0	0.82	0.12	0.06	0	3	0	0	0.7	0.26	0.04	0
4	0	0	0	0.87	0.13	0	4	0	0	0	0.93	0.07	0
5	0	0	0	0	0	0	5	0	0	0	0	0	0
6	0	0	0	0	0	0	6	0	0	0	0	0	0

(c) Abutment							(d) Joints						
	1	2	3	4	5	6		1	2	3	4	5	6
1	0.03	0.83	0.1	0.04	0	0	1	0.04	0.59	0.08	0.29	0.01	0
2	0	0.83	0.11	0.05	0.01	0	2	0	0.68	0.09	0.22	0.02	0
3	0	0	0.84	0.14	0.02	0	3	0	0	0.65	0.35	0	0
4	0	0	0	1	0	0	4	0	0	0	0.97	0.03	0
5	0	0	0	0	0	0	5	0	0	0	0	1	0
6	0	0	0	0	0	0	6	0	0	0	0	0	1

(e) Pavements							(f) Railing						
	1	2	3	4	5	6		1	2	3	4	5	6
1	0.05	0.64	0.13	0.17	0	0	1	0.03	0.7	0.21	0.05	0	0
2	0	0.77	0.12	0.11	0	0	2	0	0.78	0.16	0.07	0	0
3	0	0	0.8	0.2	0	0	3	0	0	0.9	0.1	0.04	0
4	0	0	0	1	0	0	4	0	0	0	0.98	0.01	0
5	0	0	0	0	0	0	5	0	0	0	0	1	0
6	0	0	0	0	0	0	6	0	0	0	0	0	1

(g) Guardrail						
	1	2	3	4	5	6
1	0.01	0.77	0.17	0.05	0	0
2	0	0.83	0.13	0.03	0.01	0
3	0	0	0.83	0.17	0	0
4	0	0	0	1	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0

Table 4.3: Transition probability matrices of each bridge component elicited from inspection data of case study.

- Second, sparsely distributed probabilities are noted especially when components have good as new condition represented with 1, while they become much more stable as the age/condition progresses (Kallen, 2007). Jiang et al. (1988) noted that this early stage sparse distribution can be attributed to the human bias, as inspectors are usually reluctant to name a structure as perfect.

A system-level probability of transition is calculated for the performance prediction of the bridge as a whole. The computed relative importance of each component is used (see 4.2) to acquire the respective contribution of components in deteriorating condition state of an overall bridge. The respective transition matrix of each component is

multiplied by its weights scores, which are then all summed up and normalized to single year interval to obtain the bridge-level transition probabilities. The resulting matrix is provided in Table 4.4.

Table 4.4: System level transition probability matrix of bridges.

	1	2	3	4	5	6
1	0.0315	0.795	0.090	0.075	0.004	0
2	0	0.832	0.081	0.072	0.0144	0
3	0	0	0.779	0.188	0.031	0
4	0	0	0	0.70	0.25	0.05
5	0	0	0	0	0.179	0.821
6	0	0	0	0	0	1

As discussed earlier, only joints and railing had condition values of 5 and 6 in the database, where the rest of the components never reach to these poorer states. Due to this, most of the components have condition state 4 as their lowest (poor) condition. To eliminate this skewed data distribution, we have used historical data about bridge performance from other bridge management system to estimate probabilities of transitions from condition 4 to 5, and 5 to 6 (ERA-NET ROAD, 2012). The final computed matrix at the bridge-level is provided in Table 4.4, where the elicited probability scores are denoted in bold text.

4.3.2 Computing performance indicators

The objective set for the optimal multi-year maintenance planning is quantified by Bridge Condition Index (BCI), user delay cost, and maintenance cost. These indicators measure the impact of maintenance activity concerning the defined objectives of maximization of bridge network performance (covering structural and societal aspects), and minimization of maintenance cost. By utilizing the performance indicators quantification procedure provided in Section 4.2.2, we have calculated the BCI, maintenance cost, and user delay cost for 869 bridges of the case study.

The BCI is calculated by weighted-averaged aggregation of seven components' score. Figure 4.3a presents the range of BCI found in the case study dataset. The BCI of most of the bridges ranges in between 1.5 to 3.5, except for few bridges having greater than 3.5 condition score. This limited dispersion of BCI in the dataset shows the overall good condition state of the considered bridges. By rounding the BCI to the nearest integer, we analyzed the BCI with respect to the age of bridges using box-plot as provided in Figure 4.3b. The box with corresponding whisker and outliers represent the total spread of the condition score relative to their age. The thick line in the

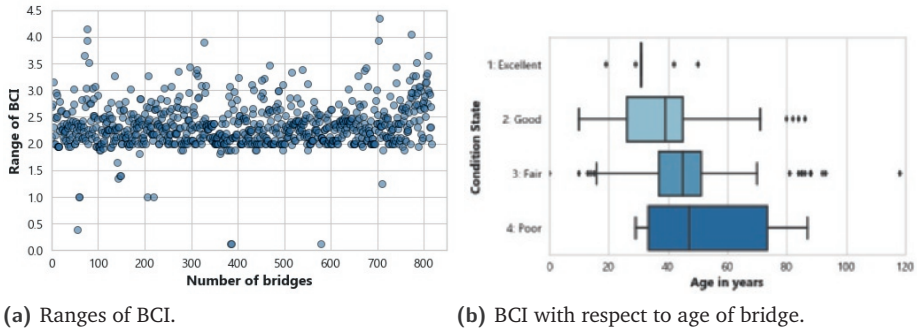


Figure 4.3: Overview of BCI in case study data.

rectangular box shows the median age with respect to each condition states. The median value of bridge age is slightly higher with each consequent condition state while it can also be noticed that the dispersion of bridge age increases as the condition gets poorer. Most of the bridges have a median range between 35 to 45 years.

We computed the approximate cost of maintenance treatment of each bridge by using the Equation 4.3 and rules defined in Table 4.1. Figure 4.4 shows the boxplot of the maintenance cost with respect to the condition state. As expected, the bridges having poor condition state require much more amount of maintenance cost as compared to bridges having relatively better condition state. In other words, a bridge is economical to maintain in its initial state of damage compared to the critical damage level.

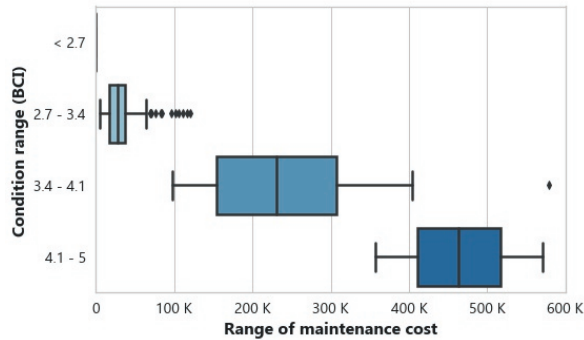


Figure 4.4: Range of computed maintenance cost.

The user delay cost is calculated by using the Equation 4.4, where the length of the work zone is estimated to be three times the total length of a bridge. The reduced speed due to maintenance is 90km/h whereas the standard speed is 130km/h. An

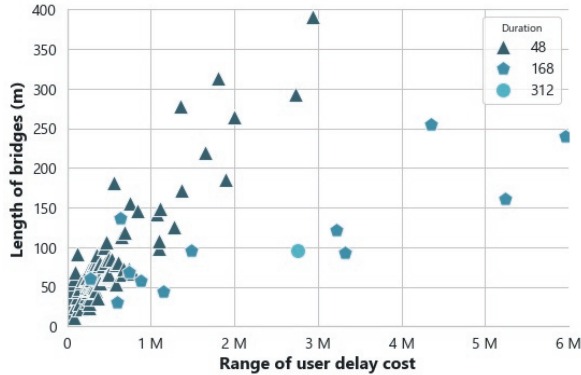


Figure 4.5: Range of computed user delay cost.

expert estimates the duration of various maintenance interventions as 48 hours for minor treatments, 168 hours for medium treatments, and 312 hours for the major treatments. Finally, the average traffic per hour over a bridge is extracted by considering the bridge location. The range of user delay cost of bridges with respect to their bridge’s length and duration of the maintenance activity is provided in Figure 4.5. The computed user delay costs are used as an indication of the impact of maintenance on the users.

4.4 Framework Implementation and Numerical Results

Based on the heuristics defined in Table 4.1, 746 out of 869 bridges have been eliminated for further consideration having BCI less than 2.7. The eliminated bridges are in good condition state and result in zero maintenance cost. In the following sections, the implementation details of MAUT, MCP and genetic algorithms applied to the case study data are provided.

4.4.1 Implementation of Multi-attribute utility theory

With the defined objectives and discerning performance indicators (attributes), MAUT method is applied on remaining 123 bridges data to rank them based on decision-makers’ risk preferences and by performing trade-offs among the multiple defined objectives. As noted earlier, the objective of maintenance planning is to improve the performance level of the bridges while resulting in minimal maintenance and user delay costs.

The first step of MAUT is to determine the single utility function (SUF) of each attribute across all the alternatives (bridges) as discussed in Section 4.2.3. The SUF is computed based on exponential utility function in order to incorporate the uncertainty as risk tolerance aspects. The concept of utility function and computation is inspired by lottery and gambling where with equal probability to obtain best value or worst value, a gambler needs to take certain risk. Since, calculation of SUF is an extensive process, the calculation details of only BCI attribute is outlined here. For this case study, the MAUT parameters, e.g., indifference points and relative weights of attributes were defined by the authors.

The condition states of remaining bridges range from 2.7 to 4.35. The expected value (EV) of BCI is then determined by considering 50% probability of obtaining $Min(x_i)$ of 2.7 and 50% probability of obtaining $Max(x_i)$ of 4.35. The computed EV is then mean of $Min(x_i)$ and $Max(x_i)$ with value 3.5. Since all the bridges having condition state up to 2 have been considered to be in an excellent condition (see Table 4.1), therefore the indifference point (CE) of BCI is set to 2. This shows that a decision-maker takes the risk-avoiding approach and prefers all the bridges to have at least BCI 2 to ensure a high-performance level. The RT value is 2.7, which is calculated by trial-and-error. By substituting the calculated values, Equation 4.5 for BCI takes the following form $U_1(x_1) = 2.18 - 5.94 * e^{-x_1/2.7}$. The similar procedure was followed to compute the utility scores of maintenance cost and user delay cost. The computation details are omitted here to conserve the space.

Next, the weight to each of the attribute is defined in order to state their importance for the objective. We have assigned the highest importance to bridge condition state with the value of 90, the maintenance cost is at a second number with 70, and user delay cost is least important with the value of 60. The relative and normalized importance weights are then calculated with these values by using Equation 4.6. From the utility function of attributes, a global aggregated score is computed for each alternative by Equation 4.7. The global aggregated score ranks all the 123 bridges in an order where the objective is to improve performance of the bridges while having minimum resulting cost.

To ensure readability, Figure 4.6 provides the MAUT scores and ranking of top 40 bridges. The bridge (B80) is 77 years old with a deck area of approximately $2000 m^2$, BCI of 4.13, and highest user delay cost is ranked as number one requiring the major intervention. The lowest ranked bridges (not included in Figure 4.6) are mostly below 30 years of age, having small deck area, small user delay cost due to less traffic flow and require minor intervention with BCI value under 3 only. Considering the budget

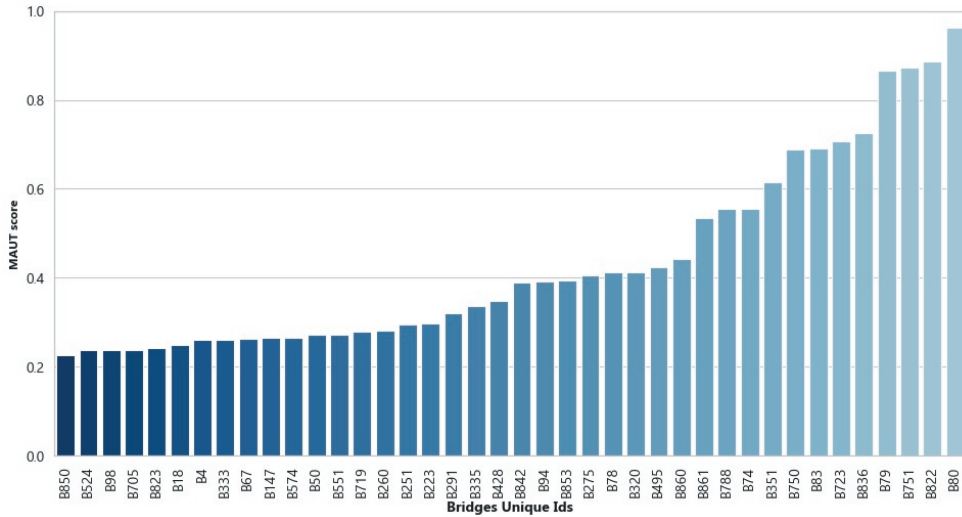


Figure 4.6: Top 40 ranked bridges by MAUT.

constraint of €6 millions, we have performed a cumulative sum of the maintenance costs of the ranked bridges and selected only those bridges that can be maintained within the available budget of 5 years. This resulted in the selection of 28 bridges having a unique identifier as B80 to B551 (See Figure 4.6).

4.4.2 Implementation of genetic algorithms and Markov chain process

The Genetic Algorithms (GAs) are applied to the 28 selected bridges to seek an optimal maintenance plan for the next five year. The purpose is to optimally plan the bridges for maintenance for a specific year, where the performance level of a maximum number of bridges can be improved in a given budget. To apply GAs on maintenance planning problem, we have utilized the evolutionary computation framework DEAP available for Python programming language (Fortin, Rainville, Gardner, Parizeau, & Gagne, 2012). A population size of 150 individuals with 10 generations is adopted for GA simulator. A partially matched crossover method with 0.2% probability is used, where the chromosomes of two individuals are randomly swapped to generate two new and unique individuals. Similarly, a mutation probability of 0.7% is applied in which, instead of mutating the chromosomes values, only the order of the chromosomes is shuffled.

Moreover, for the selection of best individuals among the number of generated individuals in each iteration, a non-dominated multi-objective optimization algorithm (NSGA-II) is applied (Deb, Agrawal, Pratap, & Meyarivan, 2000). The GA simulator is tuned multiple times with varying cross-over and mutation probability to find the settings that best converge to the given problem. The application sequences of defined stochastic operators of cross-over and mutation are standard as also introduced in Figure 4.2.

Figure 4.7 shows the number of individuals (maintenance plans) generated for a single population on the axis of condition state and budget limit having varying fitness level. A maintenance plan which compromises on the required performance threshold of 2.7 can have as low budget as €2.1 million, while comparing to those plans which achieve 2.14 of performance level but on the cost of €6.6 million. The grey (dotted) lines in Figure 4.7 show the optimization constraints for the 5 years maintenance planning where an average condition score of all the bridges of the network must have at least condition state of 2.7 enduring the €6 million budget limit. All the maintenance plans depicted with '+' marker are feasible solutions which fulfil the defined constraints. Though it is worthwhile to mention that few generated maintenance plans do not allocate all the 28 bridges for planning, therefore presents lower cost spending. The solutions represented with 'circle' are infeasible solutions and do not comply with the optimization constraints of condition states and cost.

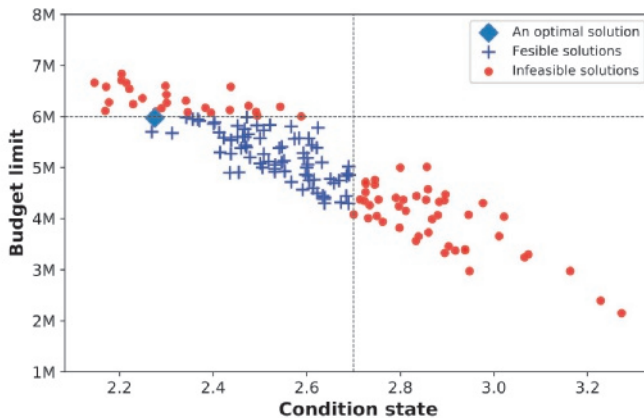


Figure 4.7: All the maintenance plans generated by GAs.

Among the 150 maintenance plans (individuals) generated in 10 iterations (generations), we choose the single most optimal maintenance plan to present and discuss here. The selected solution is also represented with a star marker on Figure 4.7.

Table 4.5: Most optimal maintenance plan of 28 bridges out of total 869 using MMPF.

No	Bridge name	Current condition	Treatment name	Maintenance cost	Improved condition
Year 1					
1	B74	3.65	Medium intervention	147,336.58	2.15
2	B719	3.36	Minor intervention	30,257.6	2.61
3	B291	3.35	Minor intervention	49,702.03	2.6
4	B822	4.04	Medium intervention	578,944.5	2.54
5	B223	3.27	Minor intervention	63,608.32	2.52
6	B842	3.07	Minor intervention	95,551.68	2.32
7	B723	3.42	Medium intervention	404,047.16	1.92
8	B428	3.24	Minor intervention	76,259.75	2.49
Yearly summary		Total yearly spending: 1,445,707.62 (24%) Remaining budget: 4,554,292.38 (76%)			
Year 2					
9	B275	3.12	Minor intervention	102,202.4	2.37
10	B853	2.93	Minor intervention	110,745.5	2.18
11	B335	3.47	Medium intervention	242,006.54	1.97
12	B79	4.08	Medium intervention	356,609.97	2.58
13	B78	3.08	Minor intervention	120,244.48	2.33
14	B750	4.08	Medium intervention	222,261.69	2.58
15	B351	4.05	Medium intervention	175,173.64	2.55
Yearly summary		Total cost spent: 1,329,244.21 (22%) Remaining budget: 3,225,048.17 (53%)			
Year 3					
16	B94	2.93	Minor intervention	116,274.03	2.18
17	B495	3.68	Medium intervention	119,176.0	2.18
18	B836	3.75	Medium intervention	238,610.66	2.25
19	B260	3.15	Minor intervention	55,593.66	2.4
20	B80	4.14	Major intervention	570,867.62	1.64
Yearly summary		Total cost spent: 1,100,521.97 (18%) Remaining budget: 2,124,526.20 (35%)			
Year 4					
21	B751	4.35	Major intervention	356,589.91	1.85
22	B251	2.98	Minor intervention	82,777.45	2.23
23	B83	3.88	Medium intervention		2.38
Yearly summary		Total cost spent: 696,271.83 (11%) Remaining budget: 1,428,254.37 (23%)			
Year 5					
24	B861	4.16	Major intervention	254,416.0	1.66
25	B860	3.71	Medium intervention	582,257.7	2.21
26	B551	3.74	Medium intervention	175,820.0	2.24
27	B788	3.89	Medium intervention	291,239.04	2.39
28	B320	3.89	Medium intervention	97,676.26	2.39
Yearly summary		Total cost spent: 1,401,409 (23%) Remaining budget: 26,845 (0.044%)			
5-Years summary		Total cost spent: 5973154.6 (99%), Remaining budget: 26845(0.044%) Number of maintained bridges: 28, Avg. condition: 2.27			

Table 4.5 presents a detailed multi-year maintenance plan of 28 bridges within defined budget and performance constraints. The plan exhibits the set of bridges allocated to a specific year along with treatment name, maintenance cost, and improved condition state. The treatment name refers to the intervention details provided in Table 4.1. As mentioned earlier, the bridge which is not maintained in a particular year (e.g., 2020)

gets its condition deteriorated until it is selected for maintenance. For instance, *B861* had a condition value of 3.65 at the time of maintenance planning (say 2018), while when scheduled to be maintained in year 5 (say 2022), the *B861* is estimated to have deteriorated condition state by 4.16. By the optimal allocation of bridge maintenance to specific years, the plan shows the cost spending of almost €5.9 million, with €0.026 million of savings. An average performance level of all the bridges is found to be 2.27.

For the sake of comparison and to establish a baseline, we also generated sequential maintenance plan without applying the GAs. The purpose of sequential plan is to mimic usual planning solution where a bridge with poor condition state are maintained first. An assumption of equal budget allocation to each year is made, where the residual budget of any year is equally distributed in all the remaining years of planning. Bridges on the basis of their condition score were sorted in an descending order and all the bridges which are within the yearly budget limit are allocated for maintenance. This sequential maintenance planning, on the basis of condition states only, is able to allocate only 18 bridges for maintenance within the given budget limit of €6 millions.

4.5 Discussion and Conclusions

This paper introduces a comprehensive framework for the development of optimal maintenance plan for a large number of road bridges over a multi-year planning period under the multiple objectives of performance requirements and budget constraints. The proposed MMPF employs the multi-attribute utility theory for the ranking of a large number of bridges by capturing the preference uncertainty and risk attitude of a decision maker. It develops several maintenance plans by genetic algorithms based heuristic search in conjunction with Markov chain processes. By forecasting the performance of assets, the objective is to assign optimal year for maintenance of bridges in order to avoid the undue maintenance cost and possible safety risk.

In addition to developing an optimal multi-year maintenance plan, the proposed framework enables the asset owners to execute various maintenance planning scenarios by changing budget limits and performance requirements from the network. Additionally, the framework can facilitate in estimating the future budget needs of agency by forecasting the performance states of assets using condition states only. The framework is validated with the large case study of 869 concrete bridges, with the objective to develop maintenance plans under €6 million of the budget limit with average condition scores of 2.7. Based on the condition states, we found that 123

bridges are below the required performance level and require some maintenance. In addition to improved performance level, the maintenance planning involved other objectives mainly the minimal impact on users (quantified as user delay cost) and reduced maintenance cost by performing maintenance at the optimal times. The proposed framework generated an optimal maintenance plan for 28 bridges within the given budget limit, whereas, the sequential maintenance plan (i.e., a bridge with the poor condition is always maintained first), provided with the same set of bridges, enables the maintenance of 18 bridges only under the same budget constraints.

The proof-of-concept on the case study data expresses the usefulness of MMPF; however, there are few limitations related to the scope and used methods. The proposed framework develops the static maintenance plan by filtering the bridges that require maintenance. Here the assumption is that filtered-out bridges are in good condition state and will not require maintenance in the next five years. However, factors like extensive usage, environmental impacts may cause extensive deterioration of assets, thus resulting in an unexpected need for maintenance. Regarding the methodology, the transition probability matrix developed in this study demands the data of at least the last two inspections and expects the ideal distribution of data over time. However, in reality, not all the bridge components can have detailed inspection records having a normal distribution, which makes the calculation of transition probability matrices a difficult activity. Similarly, despite easy implementation and reasonable running times of genetic algorithms, it is well-known that they are unable to guarantee an optimal solution due to possibly ample solution space of a combinatorial multi-objective problem. However, genetic algorithms still promise good quality solutions in a reasonable time, given the set of objectives and constraints.

From the application perspective, the MMPF incorporates heuristics, Markov chain process, and optimization frameworks, and enforces an ordered execution flow. Tool support is needed to enable the seamless execution of various maintenance planning scenarios. The future work of this study aims to improve the MMPF further. The current model mainly relies only on the pre-defined set of rules to decide on the maintenance intervention. By considering the specific structural aspects of each bridge while drawing specific maintenance treatment, we believe that a maintenance cost and performance level of the network can be further optimized. Another potential improvement of MMPF is to employ the models of machine learning for the performance prediction of each bridge on the bases of specific bridge characteristics, condition and maintenance history, and usage intensity. This will also enable the estimation of the required budget by an agency in the future.

4.6 References

- Ahmad, R., & Kamaruddin, S. (2012). A review of condition-based maintenance decision-making. *European Journal of Industrial Engineering*, 6(5), 519–541. PMID: 48854. eprint: <https://www.inderscienceonline.com/doi/pdf/10.1504/EJIE.2012.048854>
- Alaswad, S., & Xiang, Y. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157, 54–63.
- Allah Bukhsh, Z., Stipanovic, I., Klanker, G., O'Connor, A., & Doree, A. G. (2018). Network level bridges maintenance planning using multi-attribute utility theory. *Structure and infrastructure engineering*, 1–14.
- Baik, H.-S., Jeong, H. S., & Abraham, D. M. (2006). Estimating transition probabilities in markov chain-based deterioration models for management of wastewater systems. *Journal of water resources planning and management*, 132(1), 15–24.
- Barone, G., & Frangopol, D. M. (2013). Hazard-based optimum lifetime inspection and repair planning for deteriorating structures. *Journal of Structural Engineering*, 139(12), 04013017.
- Barone, G., & Frangopol, D. M. (2014). Life-cycle maintenance of deteriorating structures by multi-objective optimization involving reliability, risk, availability, hazard and cost. *Structural Safety*, 48, 40–50.
- Bocchini, P., & Frangopol, D. M. (2011). A probabilistic computational framework for bridge network optimal maintenance scheduling. *Reliability Engineering & System Safety*, 96(2), 332–349.
- Bu, G., Lee, J., Guan, H., Blumenstein, M., & Loo, Y.-C. (2012). Performance prediction of concrete elements in bridge substructures using integrated deterioration method. In *Iabse congress report* (Vol. 18, 33, pp. 108–115). International Association for Bridge and Structural Engineering.
- Bush, S., Henning, T., Ingham, J. M., & Raith, A. (2014). An agent-based framework for improved strategic bridge asset management. In *Austroroads bridge conference, 9th, 2014, sydney, new south wales, australia* (4.4).
- Butt, A. A., Shahin, M. Y., Feighan, K. J., & Carpenter, S. H. (1987). *Pavement performance prediction model using the markov process*.
- Cesare, M. A., Santamarina, C., Turkstra, C., & Vanmarcke, E. H. (1992). Modeling bridge deterioration with markov chains. *Journal of Transportation Engineering*, 118(6), 820–833.

- Chase, S. B., Adu-Gyamfi, Y., Aktan, A., Minaie, E., et al. (2016). *Synthesis of national and international methodologies used for bridge health indices*. United States. Federal Highway Administration.
- Chootinan, P., Chen, A., Horrocks, M. R., & Bolling, D. (2006). A multi-year pavement maintenance program using a stochastic simulation-based genetic algorithm approach. *Transportation Research Part A: Policy and Practice*, 40(9), 725–743.
- de Almeida, A. T., Ferreira, R. J. P., & Cavalcante, C. A. V. (2015). A review of the use of multicriteria and multi-objective models in maintenance and reliability. *IMA Journal of Management Mathematics*, 26(3), 249–271.
- Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In *International conference on parallel problem solving from nature* (pp. 849–858). Springer.
- Denysiuk, R., Fernandes, J., Matos, J. C., Neves, L. C., & Berardinelli, U. (2016). A computational framework for infrastructure asset maintenance scheduling. *Structural Engineering International*, 26(2), 94–102.
- Ding, S.-H., & Kamaruddin, S. (2015). Maintenance policy optimization—literature review and directions. *The International Journal of Advanced Manufacturing Technology*, 76(5-8), 1263–1283.
- Dong, Y., & Frangopol, D. M. (2016). Probabilistic time-dependent multihazard life-cycle assessment and resilience of bridges considering climate change. *Journal of Performance of Constructed Facilities*, 30(5), 04016034.
- ERA-NET ROAD. (2012). Asset service condition assessment methodology (ascam project), Statens väg-och transportforskningsinstitut.
- Fortin, F.-A., Rainville, F.-M. D., Gardner, M.-A., Parizeau, M., & Gagne, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13, 2171–2175.
- Frangopol, D. M., Dong, Y., & Sabatino, S. (2017). Bridge life-cycle performance and cost: Analysis, prediction, optimisation and decision-making. *Structure and Infrastructure Engineering*, 13(10), 1239–1257.
- Furuta, H., Nakatsu, K., Ishibashi, K., & Miyoshi, N. (2014). Optimal bridge maintenance of large number of bridges using robust genetic algorithm. In *Structures congress 2014* (pp. 2282–2291).
- Garg, A., & Deshmukh, S. (2006). Maintenance management: Literature review and directions. *Journal of quality in maintenance engineering*, 12(3), 205–238.
- Ghodoosi, F., Abu-Samra, S., Zeynalian, M., & Zayed, T. (2017). Maintenance cost optimization for bridge structures using system reliability analysis and genetic algorithms. *Journal of Construction Engineering and Management*, 144(2), 04017116.
- Grinstead, C. M., & Snell, J. L. (2012). *Introduction to probability*. American Mathematical Soc.

- Holland, J. H. (1975). Adaptation in natural and artificial systems. an introductory analysis with application to biology, control, and artificial intelligence. *Ann Arbor, MI: University of Michigan Press*, 439–444.
- Hooks, J., & Frangopol, D. (2013). *LTBP Bridge Performance Primer; Report No. FHWA-HRT-13-051*. Federal Highway Administration, Washington, DC.
- Hu, X., & Madanat, S. (2014). Determination of optimal mr&r policies for retaining life-cycle connectivity of bridge networks. *Journal of Infrastructure Systems*, 21(2), 04014042.
- Ishizaka, A., & Nemery, P. (2013). Multi-attribute utility theory. *Multi-Criteria Decision Analysis: Methods and Software*, 81–113.
- Jiang, Y., Saito, M., & Sinha, K. C. (1988). Bridge performance prediction model using the markov chain. *Transport Research Record*, (1180), 25–32.
- Kallen, M. J. (2007). Markov processes for maintenance optimization of civil infrastructure in the netherlands (Doctoral dissertation, Delft University of Technology, Delft, Netherlands).
- Keeney, R., & von Winterfeldt, D. (2007). M13 Practical Value Models. In M. R. W & von Winterfeldt D (Eds.), *Advances in decision analysis: From foundations to applications* (pp. 232–252). Cambridge University Press, New York.
- Keeney, & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press.
- Kim, S., & Frangopol, D. M. (2017). Efficient multi-objective optimisation of probabilistic service life management. *Structure and Infrastructure Engineering*, 13(1), 147–159.
- Kong, J. S., & Frangopol, D. M. (2003). Life-cycle reliability-based maintenance cost optimization of deteriorating structures with emphasis on bridges. *Journal of Structural Engineering*, 129(6), 818–828.
- Kumar, A. (2013). Encoding schemes in genetic algorithm. *International Journal of Advanced Research in IT and Engineering*, 2(3), 1–7.
- Lee, C.-K., & Kim Sung, K. (2007a). Ga-based algorithm for selecting optimal repair and rehabilitation methods for reinforced concrete (rc) bridge decks. *Automation in construction*, 16(2), 153–164.
- Lee, C.-K., & Kim Sung, K. (2007b). Ga-based algorithm for selecting optimal repair and rehabilitation methods for reinforced concrete (rc) bridge decks. *Automation in construction*, 16(2), 153–164.
- Liu, M., & Frangopol, D. M. (2004). Optimal bridge maintenance planning based on probabilistic performance prediction. *Engineering Structures*, 26(7), 991–1002.
- Madanat, S., & Ibrahim, W. H. W. (1995). Poisson regression models of infrastructure transition probabilities. *Journal of Transportation Engineering*, 121(3), 267–272.

- Madanat, S., Mishalani, R., & Ibrahim, W. H. W. (1995). Estimation of infrastructure transition probabilities from condition rating data. *Journal of infrastructure systems*, 1(2), 120–125.
- Markow, M. J., & Hyman, W. A. (2009). *Bridge management systems for transportation agency decision making*. Transportation Research Board.
- Mirzaei, Z., Adey, B. T., Klatter, L., & Kong, J. S. (2012). *Overview of existing bridge management systems*. Stresa, Italy. The IABMAS bridge management committee.
- Morcous, G., & Lounis, Z. (2005a). Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction*, 14(1), 129–142.
- Morcous, G., & Lounis, Z. (2005b). Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction*, 14(1), 129–142.
- Patidar, V. (2007). *Multi-objective optimization for bridge management systems*. Transportation Research Board.
- Robelin, C.-A., & Madanat, S. M. (2007). History-dependent bridge deck maintenance and replacement optimization with markov decision processes. *Journal of Infrastructure Systems*, 13(3), 195–201.
- Setunge, S., & Hasan, M. S. (2013). *Concrete bridge deterioration prediction using markov chain approach*.
- Sharma, A., Yadava, G., & Deshmukh, S. (2011). A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1), 5–25.
- Sterritt, G. (2002). *BRIDGE CONDITION INDICATORS, VOLUME 3: EVALUATION OF CONDITION INDICATORS*. Retrieved from TAP Website: <http://bit.ly/2oBt9me>. County Surveyors' Society, Lincoln, UK.
- Stratt, R. W. (2010). Bridge Management a System Approach for Decision Making. *School of Doctoral Studies European Union Journal*, (2).
- Tighe, S., Eng, P., & McCabe, B. (2005). Incorporating User Delay Cost in Project Selection: A Canadian Case Study. *1st Annual Inter-University Symposium on Infrastructure Management*.
- Wang, Y., & Goodrum, P. M. (2005). Use of conceptual road user costs for a rapid roadway construction decision making system. In *Construction research congress 2005: Broadening perspectives* (pp. 1–11).
- Xie, H. B., Wu, W.-J., & Wang, Y.-F. (2018). Life-time reliability based optimisation of bridge maintenance strategy considering lca and lcc. *Journal of Cleaner Production*, 176, 36–45.

Predictive maintenance using tree-based classification techniques: A case of railway switches

¹ **Abstract:** Growing service demands, rapid deterioration due to extensive usage, and limited maintenance due to budget cuts, the railway infrastructure is in a critical state and require continuous maintenance. The infrastructure managers have to come up with smart maintenance decisions in order to improve the assets' condition, spend an optimal cost, and keep the network available. Currently, the infrastructure managers lack the tools and decision support models that could assist them in taking (un)planned maintenance decisions effectively and efficiently. Recently, many literature studies have proposed to employ the machine learning techniques to estimate the performance state of an asset, predict the maintenance need, possible failure modes, and such similar aspects in advance. Most of these studies have utilized additional data collection measures to record the assets' behavior. Though useful for experimentation, it is expensive and impractical to mount monitoring devices on multiple assets across the network. Therefore, the objective of this study is to develop predictive models that utilize existing data from a railway agency and yield interpretable results. We propose to leverage the tree-based classification techniques of machine learning in order to predict maintenance need, activity type, and trigger's status of railway switches. Using the data from an in-use business process, predictive models based on the decision tree, random forest, and gradient boosted trees are developed. Moreover, to facilitate in models interpretability, we provided a detail explanation of models' predictions by features importance analysis and instance level details. Our solution approach of predictive models development and their results explanation have broader applicability and can be used for other asset types and different (maintenance) planning scenarios.

¹This chapter has been published as: Allah Bukhsh, Z., Saeed, A., Stipanovic, I., & Doree, A. G. (2019). Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies*, 101, 35-54.

5.1 Introduction

The railway is second rapidly growing transport modal having a steady progress growth rate of +1.5% (EuroStat, 2016). As the railway networks get busier and more developed, the demands of availability, improved service quality, and reliable infrastructure have become more critical (de Bruin, Verbert, & Babuška, 2017). With the rapid deterioration due to extensive usage, limited maintenance interventions due to budget cuts, and growing service demands, the need for infrastructure maintenance is continuously growing (ERF, 2013; European Railway Agency, 2014). Therefore, the infrastructure managers have to make maintenance decisions with the objectives to improve the assets' condition, spend optimal cost, and keep the network available.

Traditionally, the maintenance decision is derived from system failures or repetitive schedules (Marquez, 2007). The former strategy results in increased cost and longer operational delays, while the latter leads to the extra cost of undue maintenance. In condition-based maintenance, a widely adopted strategy, the decision to perform the maintenance is driven by the asset condition state (Al-Douri, Tretten, & Karim, 2016). In practice, these decisions are mainly based on expert's judgments, available budgets, and repetitive schedules. The literature has numerous Mathematical-driven Deterioration (MD) models that estimate the service lifetime of an asset. These models require multiple types of data e.g., asset's age, loading, weather, etc. for an accurate estimation of asset's service limit. Since the mentioned attributes are bound to change, MD models must be calibrated continuously. The MD models can provide sufficient decision support to infrastructure managers for planned maintenance only. But, in the case of unplanned maintenance, the decisions need to be made efficiently in order to mitigate the problem with minimal network disruptions. Currently, the infrastructure managers lack the tools and decision support models that could assist them in taking (un)planned maintenance decisions effectively and efficiently. To facilitate and accelerate the process of maintenance decision-making with less uncertainty involved, we propose to leverage machine learning techniques by utilizing a large amount of historical data of visual inspection, condition state, and maintenance records. Depending on the acquired data, the predictive models can estimate the performance state of assets, possible failure states, the best time for maintenance, and estimated operational delay minutes.

Machine Learning (ML) have brought multiple success stories from a number of industries ranging from health-care, finance, manufacturing, and marketing. ML is an umbrella term used to represent multiple computational methods for finding

hidden patterns in data and estimating predictions (Cabitza & Banfi, 2018). Few ML solution for predictive maintenance of the railway assets has been reported in the literature. For instance, Li et al. (2014) developed the failure prediction models using heterogeneous data to improve the overall railway network velocity. Similarly, Kauschke, Fürnkranz, and Janssen (2016) used the diagnostic logs for predicting the component failure of a cargo train. Another ML-based approach for detecting the metro train door failure is presented in (Manco et al., 2017). de Bruin et al. (2017) have employed the recurrent neural networks to identify and detect the failures in railway track circuits by using the signals from multiple track circuits. Based on the electrical power consumption of a switch engine, Böhm, 2017 presented an approach to predict the remaining useful lifetime of a switch. A vision-based method based on a convolutional neural network is utilized by Chen, Liu, Wang, Núñez, and Han (2018) for detecting defects of the fasteners supported on the catenary device.

The aforementioned studies highlight the increasing trends towards the data-driven predictive maintenance solutions for rolling-stocks and rail infrastructure. However, the proposed methodologies are dedicated to specific problem context and cannot be generalized to other asset types. Moreover, most of these studies have employed additional data collection measures e.g., condition monitoring devices to record the assets' behavior. Though useful for experimentation, it is expensive and impractical to mount monitoring devices on multiple assets across the network as also mentioned by (de Bruin et al., 2017). Also, the models developed using sophisticated ML techniques (e.g., deep learning models, neural networks) are accurate, but their results are difficult to interpret and communicate to infrastructure managers. Considering the infrastructure managers' needs, the objective of this study is to develop maintenance prediction models that use existing data from the railway agency and yield interpretable results.

The tree-based classification techniques for the maintenance prediction of railway switches and an approach for their interpretation is introduced in this study. The choice of tree-based classification techniques is motivated by the fact that they are comparatively easy to interpret and have empirically shown to provide optimal results for small and structured datasets (Chen & Guestrin, 2016; Hastie, Tibshirani, & Friedman, 2009). Among numerous other types of assets, we choose to illustrate the models on railway switches. Switches and crossings regulate the traffic flow and are most critical points of the network. A switch consists of multiple components, such as points, frog, joints, etc. and are vulnerable to many possible failures e.g., fatigue crack, wear failure, material deformations. An unexpected failure in switches may lead to enormous consequences. For instance, a poor condition of the switches led to

the derailment of a train in the UK on 10th May 2002, which caused seven casualties. Similarly, a major switch failure on 21st Aug. 2018 halted the traffic around Schipol, one of the busiest airport of Europe, and Amsterdam for 6 hours. Ideally, an early warning system can facilitate in timely maintenance of switches in order to avoid such incidents altogether. As also mentioned earlier, the early failure prediction models demand additional monitoring devices be mounted on the switches, which is expensive and not practical for infrastructure owners. Therefore, it is paramount to explore the possibilities to improve the current maintenance planning systems in order to bring efficiency in the maintenance decision-making process. Many railway agencies employ the SAP Enterprise Resources Planning (ERP) system to record inspections details and plan maintenance activities. We have utilized data from the in-use business process of SAP ERP for the development of maintenance prediction models. The data consist of basic properties and condition states of railway switches, accompanied by specific inspection, maintenance triggers, and historical maintenance plans. The data is provided by a railway agency though have been anonymized due to confidentiality reasons.

The primary contribution of this work is in proposing a practical solution approach for efficient maintenance planning for discrete types of infrastructure assets. The example of such assets on railway network are bridges, tunnels, switches and crossing, drainage, and slopes. In this study, we develop tree-based classification models for railway switches only. These models can predict maintenance need, activity type, and trigger's status using the existing data of a railway agency. To enable the transparency and trust of infrastructure managers on these models, a detailed explanation of models' outcomes is provided. A feature level importance analysis is performed to illustrate which features positively or negatively contributes to the model's predictive power. Moreover, we have employed Local Interpretable Model-Agnostic Explanations (LIME) framework to provide a single instance level explanations.

The rest of the paper is structured as follows: Section 5.2 gives an overview of maintenance request process and data sources. A brief introduction of tree-based classification approaches is provided in Section 5.3. Section 5.4 presents a development methodology and evaluation techniques of maintenance classification models. Section 5.5 presents the results of the models and discuss their performance differences. To interpret the models' predictions, the features importance and instance level interpretability details are discussed in Section 5.6. Finally, Section 5.7 and Section 5.8 presents discussion on the results and the conclusion of this study, respectively.

5.2 Overview of maintenance requests and data sources

SAP ERP is a software solution implemented in many railway agencies. The software manages the data and defines the key processes of a business. Among other processes, the Maintenance Request Process (MRP) is implemented by plant maintenance module of SAP ERP. An overview of MRP along with details of related data sources are explained in this section.

5.2.1 Maintenance Request Process

Railway switches are one of the most critical components on the railway network due to their intensive use, shorter lifespan, and importance for the operations. To keep the switches safe and running, the visual inspections are performed periodically. The MRP is initiated, if a switch is found to be in a poor state. Figure 5.1 shows the detailed steps of the MRP.

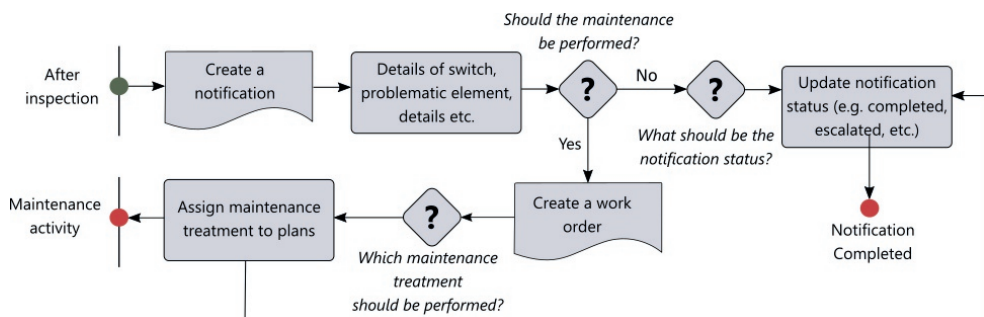


Figure 5.1: Maintenance Request Process (MRP).

As a result of an inspection, the condition state of each switch is updated in the system. If a switch is found to have a poor condition, the inspection manager generates a maintenance trigger (termed as *notification*). The created notification highlights the detected problem with a description, possible cause, and the proposed solution. Based on the reported details and respective switch properties, a maintenance engineer has to decide *if maintenance should be performed or it can be delayed*. With the decision to perform the maintenance, a work-order is generated highlighting on *which type of maintenance activity to perform* in order to mitigate the problem. Based on the maintenance need decision, the *status of the notification* (also referred to as maintenance trigger) is updated. In case maintenance was performed, then the status of notification is termed as completed. Otherwise, a notification attains the status as under observation, escalated to a manager, and technical non-compliance. The status

of notification shows the actions taken by the responsible personnel as a result of a maintenance request.

This chain of actions as a result of maintenance trigger requires decision-making on mainly three stages, as highlighted in the above paragraph with italics and with ‘?’ sign in Figure 5.1. We aim to develop classification models that could learn from historical data of notifications, work-orders, and switches properties to predict the possible answer to these decision questions.

5.2.2 Datasets

We have used four different data sources where two of them are generated through the MRP. In the following, the details of each data source are provided:

1. *Asset register* consists of basic details of 802 switches constituting features like functional location, year of installation, direction, object type, and technical details.
2. *Condition data* provides the most recent information regarding the condition state of the switches. The condition state of a switch is determined by visual inspection, where a scorecard is used to rank the asset condition from good (represented by 1) to very poor state (represented by 4).
3. *Notifications file* consists of maintenance triggers that are generated for 802 switches from 2011 to 2017. As discussed earlier, each notification is created after a visual inspection and provides the details of a detected problem, its description, and causes along with a possible mitigation measure.
4. *Work-orders file* consists of different types of maintenance works, their description, and planned execution date. The maintenance activities provided in work-orders files are either initiated by direct orders due to availability of funding, cyclic planned activities, or by unplanned triggers (i.e., notifications) reported as a result of an inspection. We have selected only those maintenance activities that have associated notification code in order to predict the specific activity types based on the reported problem.

Table 5.1 describes each of the data files in terms of data types, size, number of instances, and number of attributes (or features). Almost all the data files have mixed

Table 5.1: Description of dataset

Data files	Asset register	Condition data	Notifications	Work-orders
<i>Datatype</i>	Multivariate	Multivariate	Multivariate	Multivariate
<i>Number of instances</i>	802	4,759	13,548	12,771
<i>Number of attributes</i>	44	10	22	22
<i>Size of data (KB)</i>	349	113	1,510	1,020
<i>Prediction problem</i>	Data used for maintenance need and trigger's status prediction			
	Data used for maintenance activity type prediction			

features types constituting of a real number, integer values, text, and categorical distribution. The data sources are interconnected with a unique equipment number, notification number, and work-order number. These unique numbers have enabled us to trace a single switch across different data files. In order to get the primary characteristic of each switch, the notification file is combined with asset register based on unique equipment number. The resulting file is further combined with condition data. For the *maintenance need and trigger's status prediction*, the data from asset register, condition, and notifications files are used as also depicted in Table 5.1. For the *maintenance activity type prediction*, all the four data files are used. The notifications are combined with assets register and condition data in a similar way, as mentioned above. The resulting file is then combined with the work-order file on the basis of the order number. This results in the selection of only those notifications which has resulted in certain maintenance activity. It is important to note that for each of the prediction tasks, we have created a distinct dataset by following the MRP (See Figure 5.1) and decision aspects of the practice.

5.2.3 Exploratory Data Analysis

Before proceeding with exploratory analysis, basic data cleaning operations are performed. This includes replacing blank data fields with feature mode, renaming features' name, and conversion of data types (e.g., strings to categorical codes). In the following, few analytics and general remarks over the data are discussed.

Among the total 802 switches, most of the switches are found to lie within the age range of 10 to 35 years having a condition score of 1 or 2. The Figure 5.2 presents a box plot outlining the switches condition states concerning their age. As expected, the

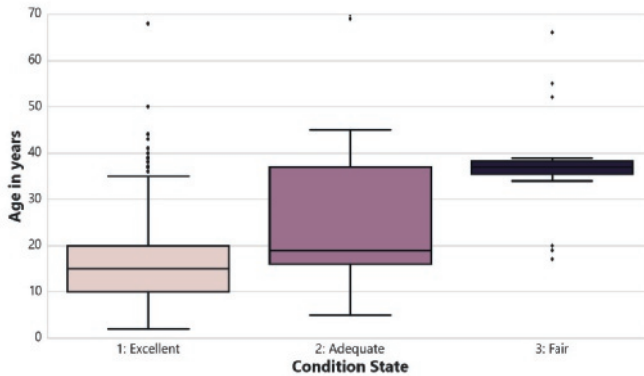


Figure 5.2: Condition states of switches with respect to age.

switches having the older age have poorer condition states as compared to the younger switches. The box with corresponding whisker and outliers represent the total spread of the condition score relative to their age. The thick line in the rectangular box shows the median age with respect to each condition states. Switches having condition state 1 have data range between 1 to 35 years with median value of 15 years. The switches with condition state 2 have a median value near 20 years of age, which means even with large spread half of the switches with this condition are younger than 20 years. Except for the outliers, at conditions score 3 the median age range of the switches is very high as compared to others because of deteriorating asset condition. Though, it is important to note that we had 25 times more data points having condition state 1 compared with condition state 3. On overall, the switches with older age shows a good condition state.

Further analysis of the number of notifications with respect to switch age group is performed. The assumption is that the older switches will generate a higher number of unplanned maintenance triggers due to their deteriorating state. To validate this, we have computed a ratio of a number of notification per number of switches in the dataset with respect to their age range. The ratios are computed because our dataset consists of a varying number of switches belonging to each age range. Figure 5.3 shows a bar plot of a number of switches and number of notifications per age group. As mentioned earlier, most of the switches belong to the age range of 11-20 years, whereas, older switches tend to have a large number of notifications generated as compared to younger switches (see for age range 'Above 40' and '31-40' years).

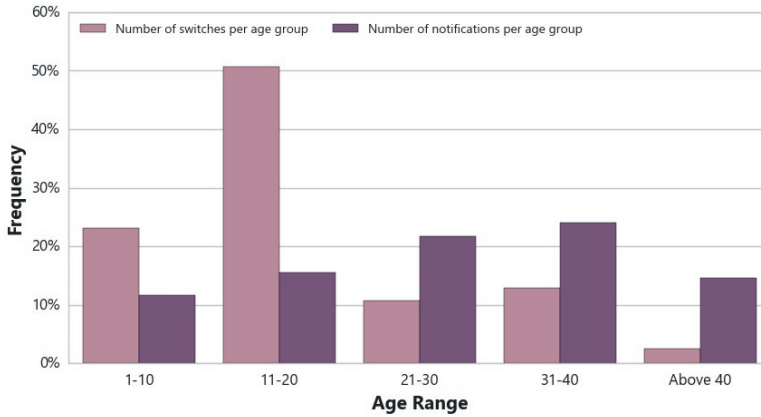


Figure 5.3: Ratio of switches and notifications generated with respect to age ranges.

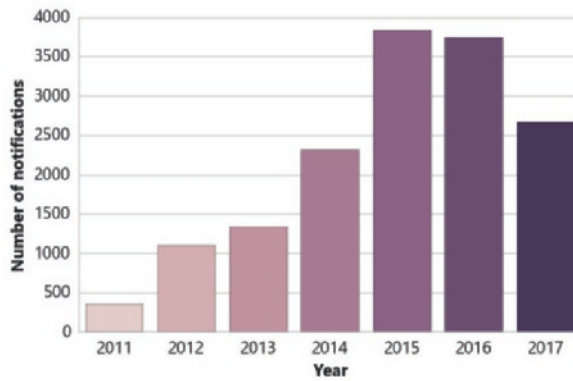


Figure 5.4: Number of notifications generated in each year.

Figure 5.4 presents the total number of unplanned maintenance triggers (notification) generated from 2011 to 2017. The relatively low number of notifications in the year 2011, 2012 and 2013 is due to the gradual adoption of the new SAP ERP system by the railway agency.

With each of the notification generated, the number of details such as detected problem, specific components, and the reason for the problem is stated. Figure 5.5 provides an overview of the identified problem during the inspection. Among others, the frequency plot outlines only the top ten most frequently reported problems.

Few interesting key aspects noted from data analysis are explained as follows:

- Among the multiple components of a switch e.g., nose, crossing panel, toes, etc., most of the problems are detected at the switch panel, while, the least problematic components are a fish plate and gauge plate.
- Lifecycle deterioration and insufficient maintenance are the top two most noted causes of problems reported.
- With the total number of notification created during 2011 and 2017, about 79% resulted in certain maintenance action.
- On average, 19 notifications are created for a single switch over the course of 7 years as a result of visual inspection.
- Among the 18 different maintenance activities noted in work-order, 90% of them belong to *M47*, which represents complete switches maintenance. We have selected the six most frequent occurring maintenance activities that have considerable data representation.

The analysis has provided an in-depth insight into the datasets. Considering the number of data attributes from multiple data sources, it is intractable for an infrastructure manager to consider the properties of previous events altogether, to decide about the maintenance need, maintenance activity type and maintenance trigger's status. Thus, the tree-based classification models developed in this paper will use the historical data and provide classification results, which can be used as the basis for the decision-making process.

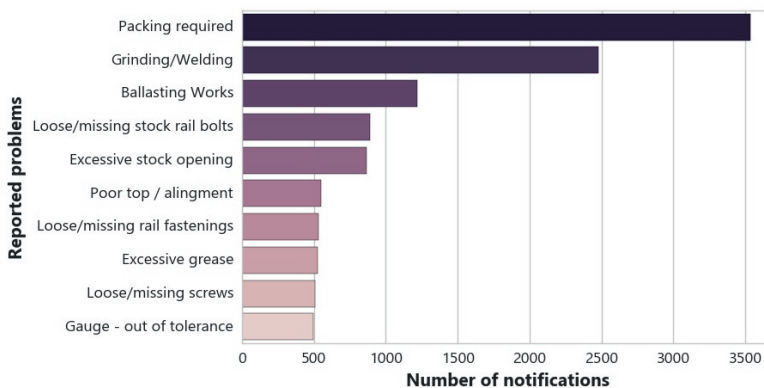


Figure 5.5: Most frequently identified problems.

5.3 Tree-based classification algorithms

Traditionally, a domain expert is required to analyze the data, establish the relationship among the features (or attributes), and derive standard rules. Since people are prone to making mistakes while analyzing a large amount of data, Machine Learning (ML) algorithms provide a more efficient alternative for capturing the knowledge in data and finding hidden patterns in order to facilitate in data-driven decision-making. The models developed using ML algorithms are successfully being used in web search, spam filtering, ad placement, credit scoring, fault detection, and many such applications. There are mainly three types of ML algorithms, supervised, unsupervised, and reinforcement learning (Raschka & Mirjalili, 2017, Chapter 1). Supervised learning algorithms require the labeled data to train the model in order to make predictions on future unseen data. The term supervised refers to the input data where the feature set and labeled-pairs are already available. In unsupervised learning, the data is without labels with unknown structure, where the algorithms try to find the clusters (or relationship) of given data points. In reinforcement learning, an agent is developed, which learns by receiving feedback from an environment in the form of some reward function. Self-Driving cars and robotics are the main application area of reinforcement learning.

In our case, the data generated from the MRP is labeled, therefore we applied the supervised learning algorithms. Irrespective of the type of learning, there are several ML algorithms to choose from for a specific prediction problem. In the context of selecting the best algorithm, the theorem of ‘no free lunch’ is prevalent in the ML community. The theorem states that there is no single algorithm that performs best for all the problems and all the datasets. In other words, an algorithm which performs best for a particular problem might perform poorly for others. The choice of an optimal algorithm depends on the size and format of the available data. The data format can be structured (tabular data) and of unstructured (e.g., audio, video, images) nature. An interested reader may refer to (Kotsiantis, Zaharakis, & Pintelas, 2007, Table 4) for the detailed explanation of supervised algorithms and their comparison for different predictive problems.

The MRP dataset, used in this study, is of structured nature where the columns represent the different features and rows represent the data instances. During the preliminary analysis, we considered several learning algorithms such as logistic regression, support vector regression, and neural net and found that the tree-based learning algorithms perform reasonably well for our dataset. These tree-based classifiers have

proven to provide optimal prediction results for the structured dataset in academic literature and many industry challenges (e.g., Kaggle competitions). For instance, Fernandez-Delgado, Cernadas, Barro, and Amorim (2014) conducted an extensive review of 179 classifiers of different types (e.g., Bayesian, neural network, support vector machine, boosting, bagging, and decision trees, etc.) on 121 structured datasets to evaluate their performance. The tree-based classifier, namely random forest, is found to be best with maximum accuracy. Similarly, in a review of supervised ML algorithms, Kotsiantis et al. (2007) concludes that rule-based system (e.g., decision tree) perform best when dealing with discrete (or categorical features), where support vector machine and neural network are better for the prediction of continuous attributes. With the reasonably good results in our preliminary analysis, and validated by the literature, we have selected the tree-based algorithms, namely decision trees, random forest, and gradient boosted trees, for the development of maintenance predictive models. These tree-based algorithms are scalable and seamlessly solve both binary and multi-class classification problems.

In addition to ease of interpretability, tree-based models offer multiple advantages. For the small data-sets, tree-based models are efficient to train and require minimal hyper-parameters tuning. These models perform internal feature selection by selecting the most prominent features as the root nodes. In contrast to various regression models, tree-based models do not require any feature normalization, thus enabling efficient implementation. A brief introduction of each of the considered tree-based model is provided below:

5.3.1 Decision Tree

The concept of Decision Tree (DT) has been implemented using different algorithms such as ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993), and CART (Classification and Regression Trees) (Breiman, Friedman, Olshen, & Stone, 1984). For DT development, CART and other algorithms work on a simple strategy of divide and conquer by employing the recursive partitioning of the training set. The key idea is to divide the training set into homogeneous subsets, where the resulting sets are pure and belong to the same target class. The algorithm quantifies the homogeneity or purity of multiple split nodes, in order to find a node split with maximum purity. The partitioning of a dataset is continued until the terminal node (i.e., target class) is reached.

Here, we have used the optimised version of CART implementation, which formulate a binary tree based on feature set as well as on thresholds. Let the data at node m be

represented by \mathcal{S} . For each candidate split $\theta = (j, t_m)$ consisting of a j^{th} feature and threshold t_m , partition the data into $\mathcal{S}_{left}(\theta)$ and $\mathcal{S}_{right}(\theta)$ subsets

$$\mathcal{S}_{left}(\theta) = (x, y) | x_j \leq t_m \quad (5.1)$$

$$\mathcal{S}_{right}(\theta) = \mathcal{S} \setminus \mathcal{S}_{left}(\theta) \quad (5.2)$$

where x is a feature value f and y represents the target class. CART generates a binary tree in which each parent node can have only two child nodes. Once a split based on θ is defined, the impurity at each node m is computed.

With the target class values as $k \in y: k = \{1, 2, \dots, K\}$ for node m , the proportion of class k belong to node m is calculated as follows:

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k) \quad (5.3)$$

where x_i represents the feature value, y_i represents the class label, N_m and R_m represent the total instances available at node m and region of m respectively. By using the p_{mk} , the Gini index, which is a measure of node impurity, is computed as:

$$H(X_m) = \sum_{k=1}^K p_{mk}(1 - p_{mk}) \quad (5.4)$$

The best node split by substituting the $H(X_m)$ takes the following form:

$$G(\mathcal{S}, \theta) = \frac{n_{left}}{N_m} H(\mathcal{S}_{left}(\theta)) + \frac{n_{right}}{N_m} H(\mathcal{S}_{right}(\theta)) \quad (5.5)$$

whereas the value of θ can be further optimized as $\text{argmin}_{\theta} G(\mathcal{S}, \theta)$ to minimize the $H(X_m)$ impurity.

Having the minimization of the Gini index as an objective function, DT applies the top-down greedy search to determine the best split of nodes. The recursive partitioning algorithm is executed until the maximum allowable length of a decision tree is reached. The mathematical formulation explained above are taken from (Pedregosa et al., 2011). In general, the DT model can be seen as a disjunction of conjunctions or as if-then rules, which are useful to aid human interpretability.

5.3.2 Random Forest

The Random Forest (RF) is a term for an ensemble approach of DT, which consists of a number of trees. Unlike classical classification techniques which builds a single tree on a complete dataset, RF randomly selects the instances and features to construct multiple trees. Each DT then casts a vote for a particular target class and a class having the majority votes is model's prediction with a certain probability.

The RF learning algorithm can be explained as follows (Breiman, 2001; Guo & Berkahn, 2016):

1. Get N bootstrap samples from the dataset.
2. For a single sample set i generate a decision tree T_i as explained in Equation 5.1 and 5.2 of Section 5.3.1.
3. Output the ensemble of T generated trees.
4. Classify a new test instance as:
 - For binary classification where $k = \{0, 1\}$, aggregate the prediction of T trees with respect to each class, where, a class having majority votes is a model's prediction
 - For multi-class classification where $k = \{1, 2, \dots, n\}$, it averages the predictions of T trees as:

$$f(x) = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (5.6)$$

In comparison with a traditional DT, RF shows good predictive performance even with considerable noise induced by the random selection of features and instances (Biau, 2012; Hastie et al., 2009). Moreover, it is able to deal with large dataset having several features with diverse data types e.g., categorical or continuous values.

5.3.3 Gradient Boosting Tree

Gradient Boosting Tree (GBT) is an alternative ensemble learning technique that consecutively produces weak tree classifiers in a stage-wise fashion as other boosting

algorithms do with a different base model (Friedman, 2001). The key idea of boosting methods is to resample strategically and incrementally build multiple models for training instances that are difficult to estimate with previous ones by minimizing some arbitrary differentiable loss function e.g., cross entropy or sum of squared errors. This is a major distinguishing factor as compared to RF, which produce each tree through random sampling with replacement. Principally, gradient boosting methods perform optimization in the functional space in an additive form by sequentially adding an estimator $h(x)$ to an overall ensemble function $F_m(x)$ for generating a new model $F_{m+1}(x)$.

Given a squared error loss function, this method employs gradient descent algorithm to select a base-learner $h(x)$ that highly correlates with negative of the gradient as:

$$\rho_m = \underset{p}{\operatorname{argmin}} \sum_{i=1}^N [\mathcal{L}(y_i, F_{m-1}(x_i) + \rho h(x_i))] \quad (5.7)$$

$$F_m(x) = F_{m-1}(x) + \rho_m h(x)$$

where N represents the total number of training instances, ρ is the gradient descent step-size, and \mathcal{L} is a loss function. More information on GBT methods can be found in (Natekin & Knoll, 2013). The GBT builds only one tree at a time, which results in longer training times as compared to RF. On the other hand, the sequential training on difficult examples provides a strong base to deal with unbalanced datasets.

5.4 Maintenance classification

To facilitate effective and efficient maintenance decision-making, we propose to employ tree-based classification methods for the development of predictive maintenance models. The development of a ML model involves few preparatory tasks such as data pre-processing, class labeling, and definition of evaluation approaches. Data pre-processing aims to eliminate the noise, and prepare the feature sets that support in the classification process. In case the data is not explicitly labeled, class labeling step is performed to analyze the data in order to define the target labels. Finally, the evaluation approaches are defined to gauge the predictive power of the developed models. The following sub-sections provide a detailed explanation of each of these tasks.

5.4.1 Feature Extraction

Feature extraction is one of the most vital steps of ML model development. It involves the domain knowledge to select the most relevant features, combine existing attributes, and often create new features from existing data. The focus is to identify the attributes that are most relevant to identify the target class, which will ultimately improve the predictive power of the model.

We initiated by eliminating any duplicating features such as, age of switch and year of construction, unique identifiers, geo-coordinates of switches, and any features that do not directly contribute to a decision-making process. This selection procedure reduced the total number of features considerably. The selected features mainly include detected problem, switch component, problem reason and cause, functional location of a switch, track type, technical details, and age, etc. We did not perform further data processing, for instance, data normalization or feature scaling, as the tree-based models are invariant to feature scales (Friedman, Hastie, & Tibshirani, 2001). Though, the textual features were processed further to eliminate the irrelevant information by performing lower casing, removing English stop words and special characters. Since ML models are unable to process the textual data directly, we calculated Term Frequency-Inverse Document Frequency (tf-idf) features from free-form text. Tf-idf is an information retrieval technique that assigns weights to each word in a documents/attribute. The TF (term frequency) calculates the number of occurrence of each term in a document, where IDF (inverse document frequency) is a measure of how important a word is to a document. The product of *tf* and *idf* form the total weight of a term. The higher is the weight of the term, the more relevant it is to a document. Further details to compute the tf-idf is provided in (Ramos et al., 2003). The weights of tf-idf serve as a feature value, where we have selected top eighty text terms with highest tf-idf score for further analysis.

5.4.2 Label Processing

The supervised ML models require the labeled data for model training and testing purposes. Often the data instances are not explicitly labeled; therefore, manual class labeling has to be performed.

The notification dataset does not contain any feature(s) that shows if the notification leads to a maintenance activity in the past. Though, an analysis of notifications showed that a work-order number is attached to few notifications, which represents

the type of maintenance performed, whereas other notifications have no associated work-order numbers, meaning no maintenance was performed. This notion leads us to label the notification dataset as follows:

$$\begin{cases} 0, \forall m \text{ if } wo \in \emptyset \\ 1, \forall m \text{ if } wo \in \mathbf{R} \end{cases}$$

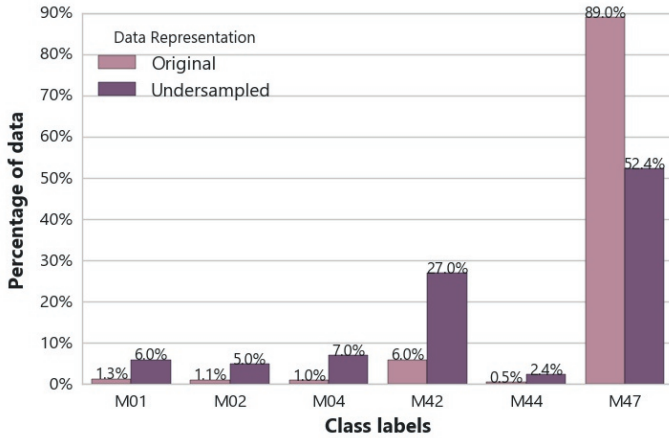
where m is a data instance and wo is a work-order number. For all the notifications m , where wo belongs to a null set, the data is marked as 0 indicating no maintenance was performed. For those notifications, where work-order wo was present and belonged to a real number, the data is marked as 1 indicating maintenance was performed. The distribution of class labels for the maintenance need prediction problem is provided in Table 5.2. Regarding the prediction of maintenance activity and trigger’s status, the data was already labeled with distinct activity types and statuses, respectively.

Table 5.2: Class representation for maintenance need prediction.

Class label	Class name	Data representation
1	Maintenance performed	79% (10,743)
0	No maintenance performed	21% (2805)

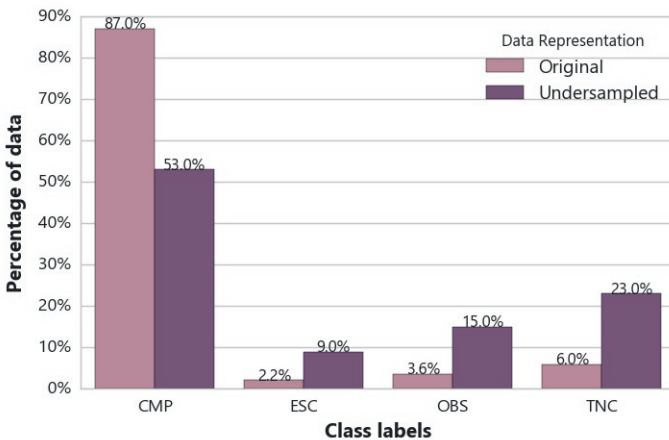
Predictive models learn well when each class has at least one-tenth representation in overall training data. Ideally, a balanced dataset has an approximately equal ratio of respective class instances for model learning. However, in case of an imbalanced dataset, the model tends to be biased towards the class having major representation, therefore performs poorly for minority class(es) (Akosa, 2017). With the class labeling, the dataset was found to be highly imbalanced, particularly for the maintenance activity and trigger’s status prediction. Our preliminary analysis also presents the high bias of model towards majority classes, which results in poor classification ability. To mitigate data imbalance and to improve the classification precision, the classes having the large data representation must be under-sampled. The simple random sampling approach is used, where each data point has an equal probability of selection on independent and identically distributed dataset (He & Garcia, 2008). By random sampling, 13-15% of the data from majority classes have been selected. This has balanced the overall representation of each class; thus reducing the considerable convergence to a single class only.

Table 5.3: Undersampling the major classes for maintenance activity prediction.



Class label	Class name	Data representation	Undersampling
M01	Hand/Cobra pack	1.33% (112)	6% (112)
M02	Replace defective material	1.11% (121)	5% (112)
M04	Oil/Grease/Shim	1% (126)	7% (126)
M42	Tamping	6% (603)	27% (603)
M44	Ballasting/Dumper	0.53% (45)	2.35% (45)
M47	Switch maintenance	89% (9,605)	52.35% (1000)

Table 5.4: Undersampling the major classes for maintenance trigger's status prediction.



Class label	Class name	Data representation	Undersampling
CMP	Completed	87% (11,795)	53% (1500)
ESC	Escalated	2.16% (244)	9% (244)
OBS	Observation	3.65% (443)	15% (443)
TNC	Technical non-compliance	6% (781)	23% (781)

Table 5.3 shows the original and undersampled representation of maintenance activity type classes. *M47* has the highest number of instance for activity type prediction. The undersampling has reduced its representation from 89% to 52%, which consequently increased the data representation of other classes. Note that, the work-order instances presented in Table 5.3 will not add up to the number of instances presented in Table 5.1. This is because, among 18 different maintenance activities, we have selected the top six most frequently occurring activities having sufficient data representation. Table 5.4 presents the original and undersampled class distribution for trigger's status prediction. *COM* has the highest share of data for prediction of maintenance triggers, which is reduced to 53% only. The undersampling of majority classes has resulted in an increased representation of other classes in the dataset. The notable difference is for classes *M44* and *M04* (see Table 5.3) and for *ESC* and *OBS* (see Table 5.4).

5.4.3 Evaluation Approaches

The performance of the predictive models for maintenance classification is assessed through the held-out test set and stratified cross-validation, which is discussed in detail in this section. Moreover, multiple performance measures to gauge the models' predictive power are introduced here.

Held-out test set

In a held-out method, the complete dataset is split into training and test sets. The training set is used to train the model, and the test set is used to evaluate the model's performance. One of the most common approaches is to randomly select the 70% of data instances for training and rest 30% for test. In our work, we have used a slightly different version of the held-out method, where the train and test split is defined based on the notification year. All the notifications generated before 2016 are used for training, while the notifications generated during and after 2016 are used for testing purposes. The split based on the year will evaluate the model performance more robustly, where the model was trained on only past data and tested-out on future unseen data. This is also in line with the purpose of the model, namely, to predict future maintenance needs. For reference about a number of notifications generated each year, see Figure 5.4.

Stratified cross validation

In Stratified Cross Validation (SCV), the whole dataset is randomly split into a number of equally sized units referred as 'folds'. Here, the term stratified to represent the equal class distribution for each fold. Having N number of the fold, the $N - 1$ are

used for the training, while the N_{th} fold is used for the model testing. This process is repeated N times until each fold had the opportunity of being used as N_{th} test and training fold (Witten, Frank, Hall, & Pal, 2016). Finally, the output is averaged across all folds to estimate the performance of the model. This method ensures that every data point is used at least once as a training example and once as a test example. The SCV is performed for the completeness of validations and the evaluation of the model's robustness.

Performance measures

Several performance measures are used to compare and evaluate the models' prediction power. In classification problems, we used confusion matrix analysis, which represents the models' predicted classes of test data for which the true values are already known. Table 5.5 shows the confusion matrix for a binary classification presenting positive and negative as class values. The True Negatives (TN) are cases belonging to *negative* class and are correctly classified as *negative* by a model. Similarly, the True Positives (TP) are a number of cases belonging to *positive* class and are correctly classified as *positive*. In contrast, False Positives (FP) are *positive* samples that are incorrectly classified as a *negative* class. While False Negative (FN) are *negative* class samples incorrectly classified as positive. A normalized confusion matrix with perfect classification has TN and TP of *one* and FP and FN of *zero*.

Table 5.5: Confusion matrix for binary classification problem.

	Classified negative	Classified positive
Actual negative	True Negatives (TN)	False Positives (FP)
Actual positive	False Negatives (FN)	True Positives (TP)

With the confusion matrix, a number of performance measures can be calculated (Flach, 2012, Chapter 2). The metrics used in this study for evaluation purposes are explained as follows:

- *Accuracy* is a measure of correct predictions of the model compared to the total data points. It shows how often the model classifies the instances correctly. It is computed as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

The accuracy is a good measure where the data is balanced for each class labels. However, in the case of imbalanced data set, the accuracy measure without other performance measures can be misleading.

- *Misclassification rate* determines how often the model has misclassified an instance compared to the total given instances. For example, a positive instance assigned to a negative class by the model. Misclassification rate is calculated as:

$$\text{Misclassification rate} = 1 - \text{Accuracy} = \frac{FP+FN}{TP+TN+FP+FN}$$

This measure is mainly used for the binary classification problems, as calculating the rate of misclassification for multi-class problems is complicated.

- *F-Score* is a combination of precision (or positive predictive value) and recall (sensitivity) measures (Sokolova & Lapalme, 2009). The precision determines the exactness of the model. It is a ratio of correctly predicted positive instances (TP) to the total positively predicted instances (TP+FP). Precision is represented as:

$$\text{Precision} = \frac{TP}{TP+FP}$$

In contrast, recall provides a measure of the model's completeness. It is a ratio of a correctly predicted positive instance to the total instance of the positive class (TP+FN) in test data. The recall is calculated as:

$$\text{Recall} = \frac{TP}{TP+FN}$$

Precision represents the model's performance with respect to false positives, whereas recall represents the performance related to false negatives. The F-score convey the balance between precision and recall by taking their weighted harmonic mean. F-score is calculated as follows:

$$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Similar to the accuracy, F-score performs well with the reasonably balanced dataset. In the case of an imbalanced dataset, the adjusted F-measure is utilized.

- *Kappa* presents an inter-rater agreement between qualitative items, which measure the relative observed agreement (p_o) with the hypothetical probability of

chance agreement (p_e) (Cohen, 1960). The kappa measure does not only calculates the percentage accuracy but also consider the possibility of an agreement between raters (qualitative items) by chance. The value of kappa is calculated as follows:

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e}$$

In case of imbalanced datasets, the kappa is a robust measure compared to F-score and accuracy. It can be said that kappa determines how well a model performed (p_o) as compared to how well it could have been performed by chance (p_e), while considering the marginal distribution of a target class.

Given the performance evaluation measures, the idea is to maximize the TP and TN and minimize the FN and FP. However, in real-world applications, there is always a trade-off between true and false positives, and negatives rate (Batista, Prati, & Monard, 2004) depending on the business needs and problem context. For instance, in the context of maintenance need prediction the false negatives (model predict no need for maintenance, where it is needed) are more critical compared to the false positives.

5.5 Results

For the development of maintenance classification models, machine learning python library scikit-learn (Pedregosa et al., 2011) is employed. Three distinct models based on DT, RF and GBT algorithms are trained with scikit-learn default hyper-parameters configurations. Only the number of trees for RF and GBT algorithm are manually tuned within a range of 100 to 2000 trees. It is found that the model performs optimally given the 1500 trees, where no improvement in performance is noted by an increasing number of trees.

It is important to note that each of the predictive models is developed independently, where every model utilizes its own set of input features. By the rigorous features selection, we have used only those input features set for modeling of prediction tasks that are available during real-time decision-making. Further, the performance of these trained models is evaluated by considering the DT as a baseline. The results explain which model performs the best for a particular classification problem. The developed predictive models will be able to predict maintenance need, maintenance activity type, and maintenance trigger's status.

5.5.1 Maintenance need prediction

Three distinct models are developed to predict if maintenance will be performed or delayed based on the respective switch properties and problems reported. Table 5.6 shows the model evaluation results tested on held-out dataset. All models show a negligible difference in performance. The GBT model obtained the f-score of 86% with 13.7% misclassification rate.

Note that, for held-out evaluation, the models were trained on a dataset of notifications generated from 2011 to 2015, while tested on data of 2016 and 2017. This warrants that the developed models will be able to perform sufficiently well in real-world settings having the notifications generated in future. Figure 5.6 shows the confusion matrix analysis to evaluate the classification accuracy. Confusion matrix analysis provides a summary of correctly and incorrectly classified instances concerning each class. The confusion matrix provided in Figure 5.6a shows that 62% of times the DT model correctly classified a ‘no’ maintenance as ‘no’ (true positives), whereas 38% of times decision tree model has incorrectly classified a ‘no’ maintenance to ‘yes’ (false positives). Similarly, 91% of times the model correctly predicts ‘yes’ (true positives), whereas 9% of times it is incorrectly classified as ‘no’ (false negative). For all three models, most of the misclassification is attributed to false positives (top right value). This means the model predicted that maintenance should be performed

Table 5.6: Results of maintenance need prediction with held-out test set.

Model	Accuracy	Misclassification	F-score	Kappa
Decision Tree (DT)	0.844	0.151	0.847	0.540
Random Forest (RF)	0.857	0.142	0.853	0.548
Gradient Boosted Tree (GBT)	0.862	0.137	0.860	0.575



Figure 5.6: Confusion matrices of maintenance need prediction on held-out test set.

while in reality, the decision was not to perform the maintenance. By using the frequency distribution per class label provided in Table 5.2, the approximate number of (in)correctly classified instances can be calculated. Though RF shows the highest number of false positives (see Figure 5.6b), DT has in total highest misclassification rate of 15%.

In the context of maintenance, the false negatives are more critical as compared to false positives since in those cases, the model suggests not to perform the maintenance; while the maintenance is needed. In addition to the trade-offs between false positives and false negatives, the misclassification by the model can be substantially reduced by gathering more training data for minority classes or by applying techniques like SMOTE (Synthetic Minority over-sampling techniques) that generates data points by interpolation from the minority class samples (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Another solution is to manually inspect false positives and false negatives by single instance-level interpretability and by relative feature importance. Further details on this are provided in Section 5.6.

Table 5.7: Results of maintenance need prediction with SCV.

Model	Accuracy	Misclassification	F-score	Kappa
DT	0.877	0.114 ± 0.006	0.886 ± 0.006	0.658 ± 0.019
RF	0.913	0.087 ± 0.008	0.911 ± 0.008	0.725 ± 0.028
GBT	0.923	0.079 ± 0.005	0.919 ± 0.006	0.752 ± 0.018

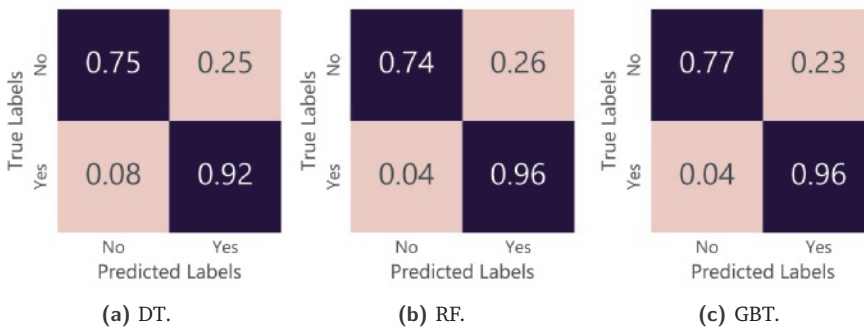


Figure 5.7: Confusion matrices of maintenance need prediction with SCV.

The trained models are also evaluated with 10-fold stratified cross-validation (SCV). Table 5.7 presents the results of SCV, which shows the averaged scores across 10-folds along with standard deviations. Based on SCV, the GBT model still performed better than others. Overall, all the models show considerable performance improvements compared with models of the held-out test set. The reason for this is due to the difference in validation approach (See Section 5.4.3 and 5.4.3). With SCV, the model

is trained and tested multiple times on randomly created data folds. The data from a single year can be used in both training and testing folds. Figure 5.7 presents the averaged confusion matrices analysis for SCV. With both validations, it can be noted that the models can learn well for both true positives and negatives, whereas false positives and negatives are critical to learning.

5.5.2 Maintenance type prediction

The purpose of these models is *to predict the specific maintenance activity type*. As mentioned earlier, we have selected only the six most occurring maintenance activities that have sufficient data representation for each class. The description of each selected maintenance type and details of their class representation is provided in Table 5.3. Based on the problem reported, any one of these six maintenance activity types can be chosen. Often the selection of maintenance activity type is based on experts' judgment. We have developed three distinct models that learn from the historical data and predict the most suitable maintenance activity type.

Table 5.8 shows the models classification results as evaluated with held-out dataset. RF model achieved the highest accuracy having the f-score and kappa of 70% and 0.42 respectively. Maintenance type prediction is a multi-class problem; therefore

Table 5.8: Result of maintenance type prediction with held-out test set.

Model	Accuracy	F-score	Kappa
Decision Tree (DT)	0.630	0.659	0.330
Random Forest (RF)	0.70	0.70	0.427
Gradient Boosted Tree (GBT)	0.678	0.695	0.409

it is not possible to calculate the straightforward misclassification rate. Though, the

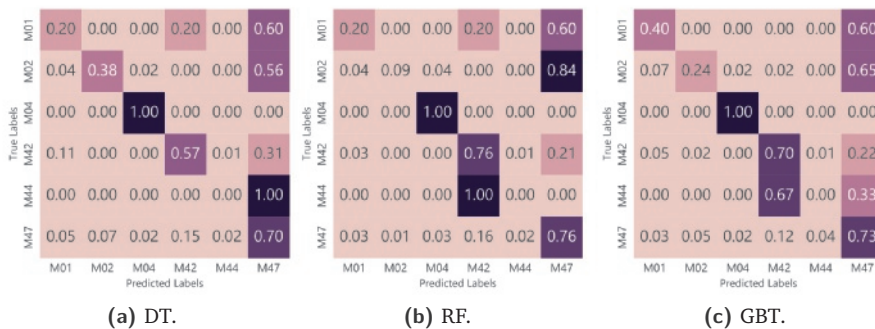


Figure 5.8: Confusion matrices of maintenance type prediction with held-out test.

misclassification of each model for a particular class can be analyzed with confusion matrices provided in Figure 5.8. For all the models, the classification of maintenance type M01, M02, M42 and M44 are confused as M47 by varying ratios. For instance, M44 is incorrectly classified by all the models. Similarly, M02 is poorly classified by all through the accuracy of DT is better than RF and GBT. The reason for M02 and M42 misclassification can be attributed due to the only 5% and 2.35% class representation, respectively (see Table 5.3). In general, the overall misclassification by the models is due to varying class distribution in train and test dataset.

These models are also evaluated with 10-folds SCV and results are represented in Table 5.9. The GBT model achieved an average f-score of 79.2% and kappa of 0.68. Likewise, the RF model has further improved the results having average f-score of 80.3% and 0.69 of kappa. Figure 5.9 shows the confusion matrices analysis of each model. Overall DT shows the highest number of misclassification rate. But, for classes M02 and M44 DT performed better than RF. Similarly, GBT have classified M01, M02 and M44 better than RF. However, RF shows the higher percentages of accurately predicted classes (see the diagonal of Figure 5.9b), which makes it the best classifier for activity type prediction.

Table 5.9: Results of maintenance type prediction with SCV.

Model	Accuracy	F-score	Kappa
Decision Tree (DT)	0.708	0.740 ± 0.018	0.602 ± 0.029
Random Forest (RF)	0.789	0.803 ± 0.02	0.699 ± 0.038
Gradient Boosted Tree (GBT)	0.768	0.792 ± 0.021	0.683 ± 0.032

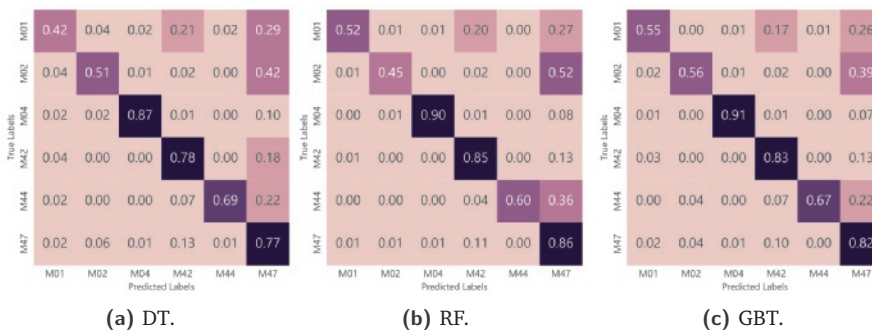


Figure 5.9: Confusion matrices of maintenance type prediction with SCV.

5.5.3 Maintenance trigger’s status prediction

The status of maintenance trigger (i.e., a notification) is decided based on the problem reported. Instead of maintenance only, the status of a maintenance trigger defines the further actions, which includes technical non-compliance (TNC), escalated (ESC), observation (OBS), and completed (COM). The details of class distribution is presented in Table 5.4. We have developed three predictive models based on DT, RF, and GBT algorithms *to predict the status of a maintenance trigger*.

Table 5.10 shows the evaluation results of models on held-out test set. The performance of the RF model with f-score of 77.7% and kappa of 0.62 is best among all the models. Figure 5.10 shows the confusion matrices, which presents relatively poor classification, where most of the classes are misclassified as *COM*. This is because of the highest class representation of this class. Comparatively, there is fewer misclassification for *TNC* because it has a sufficient number of data samples for the model to be able to learn accurately.

Table 5.10: Result of maintenance trigger’s status prediction with held-out test.

Model	Accuracy	F-score	Kappa
Decision Tree	0.707	0.714	0.494
Random Forest	0.791	0.777	0.623
Gradient Boosted Tree	0.737	0.733	0.532

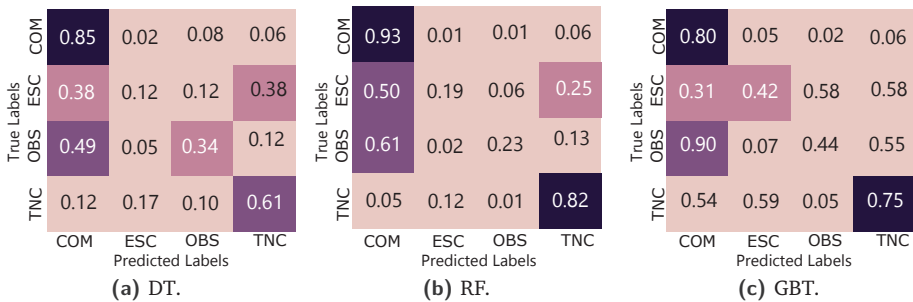


Figure 5.10: Confusion matrices of maintenance trigger’s prediction with held-out.

Table 5.11 shows the models evaluation results with SCV. The RF model performed best with 82.4% f-score and 0.73 kappa. Figure 5.11 shows the results of classifications with confusion matrices. All the models have misclassified the *OBS* with a *COM* status with a certain ratio. This is because, the postponing of a maintenance activity leads to an additional monitoring or regular visual inspection, here designated as *OBS*. Since *OBS* is just a delayed maintenance activity, these cases are most likely to share their

properties with the cases of *COM*. The similar properties of these cases may have led to the misclassification by the models.

Table 5.11: Results of maintenance trigger’s status prediction with SCV.

Model	Accuracy	F-score	Kappa
Decision Tree	0.783	0.768 ± 0.019	0.641 ± 0.03
Random Forest	0.827	0.824 ± 0.018	0.731 ± 0.03
Gradient Boosted Tree	0.834	0.823 ± 0.020	0.727 ± 0.03

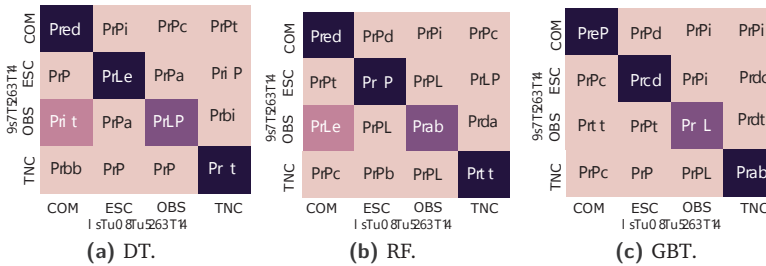


Figure 5.11: Confusion matrices of maintenance trigger’s status prediction with SCV.

5.6 Interpretability

For automated decision-aid systems, transparency and interpretability have crucial importance. The ML models are notorious for being black boxes, which provide little to no explanation of their predictions. Many recent studies have emphasized the importance of interpretable models (Doshi-Velez & Kim, 2017; Lipton, 2016). The focus is towards providing an explanation about which input features positively or negatively impact the model’s output (Ribeiro, Singh, & Guestrin, 2016) and to explain the model behavior locally for a specific instance or a case (Datta, Sen, & Zick, 2016). The model interpretability is also very important for domain experts in order to be able to trust and employ predictive models in their daily decision-making scenarios (Andreou et al., 2018). Moreover, the details on features importance and instance details also enable the further improvement of the model by reducing the critical false negatives and possible false positives. This section provides an in-depth analysis of the model behavior with respect to the features importance and local instance level explanations.

5.6.1 Features Interpretability

Among the number of features that are used for the model training, not all of them equally contribute towards learning a model. With tree-based classification models, the importance of a feature learned by a predictive model can be estimated. The feature importance score quantifies the importance of each attribute in the overall prediction of the model. The high importance score of an attribute suggests that the attribute is very important and positively contributes to the model's predictive capability. However, the negative importance score shows that the attribute is negatively influencing the performance of the model. This means by eliminating the attributes having a negative importance score, the performance of the model is likely to improve. To enable the comparison among multiple attributes, we have calculated the feature importance of only RF models. Along these lines, RF has also performed best for two of the three maintenance predictive models discussed in Section 5.5.

There are multiple methods to compute the feature importance of predictive models, namely, default mean decrease accuracy, permutation importance, and drop-feature importance (Raschka & Mirjalili, 2017, Chapter 4). Drop-feature importance is one of the most accurate methods to compute features importance but has a high computation cost. Since we have a small dataset, we choose to employ drop-feature importance method, which demands model (re)training multiple times.

The main idea of the drop-feature method is to train a RF model on all feature set, as explained in Section 5.5, and term its accuracy/error estimation as baseline represented as ps_b . Based on the n number of features, the RF model is trained n times wherein each iteration a single feature i is dropped, and accuracy of the model is recomputed. The new accuracy score is represented as ps_n . The single feature importance of i is then calculated as $\Delta = ps_b - ps_n$, which a difference in baseline and drop in overall accuracy score. This process is repeated for n number of features until the importance of each feature is computed. The higher is the difference between the resulting model's accuracy and baseline, the more important a feature is for model's prediction. Note that, these computed values provides relative importance of features that may not add to one (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008).

In order to elaborate the models' results, the feature importance level for each classification model are developed. Figure 5.12 provides the feature importance of a maintenance need prediction model. It shows that the *functional location*, *age* and *stretcher type* are main contributing factors when deciding to perform or delay the maintenance. The *problem cause*, *switch component* and *problem reason* are negatively

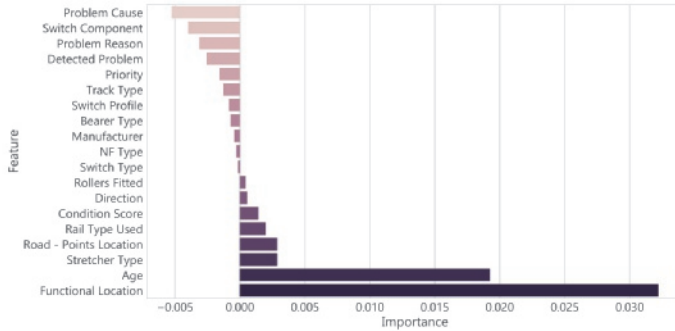


Figure 5.12: Features importance of a RF model for maintenance need prediction.

influencing the models' predictions. This is in contrary to general assumptions that the detected problem and notifications details drive the *need of maintenance* decision.

For the prediction of maintenance treatment type, the feature importance graph is presented in Figure 5.13. The *detected problem* and *functional location* are most important for maintenance type prediction whereas the *track type* and *switch component* are least important. Notice the *manufacturer* attribute has zero importance, which means if we eliminate this feature from the dataset, the performance of the model will not likely be affected. Moreover, if the features, such as *direction*, *bearer type*, having the importance level less than zero are eliminated, the overall performance scores of the model is likely to improve.

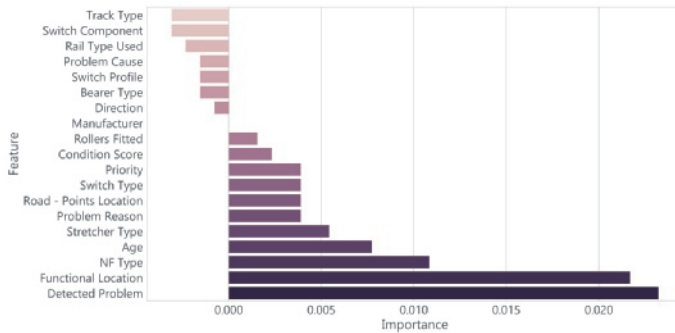


Figure 5.13: Features importance of a RF model for maintenance type prediction.

Figure 5.14 shows the feature importance of RF model for the prediction of maintenance triggers' status. As expected *detected problem* followed by *functional location* are relatively most important features to decide on maintenance triggers' status. While, *condition score* and *switch component* features can be eliminated since they have no importance and are not contributing in model's final predictions.

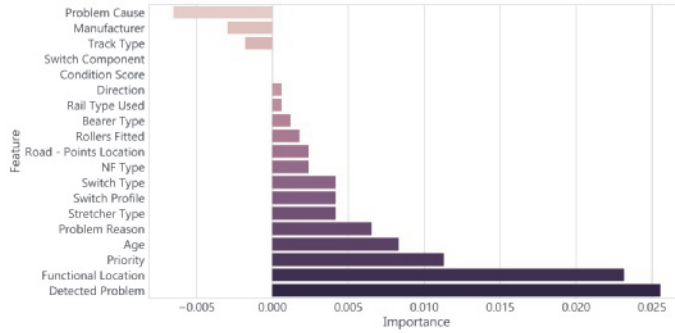


Figure 5.14: Features importance of a RF model for maintenance triggers’ status.

Comparing the feature importance graphs of Figures 5.13 and 5.14, it is interesting to note that given the similar feature set for training each model remarks the different number of features as important for the model’s performance. There are multiple benefits of performing feature importance analysis. First, it enables domain experts and decision-makers to gain an in-depth understanding model’s results by analyzing the importance of the features. Second, it illustrates how the models’ decision pattern diverts from the real decision-making practices as we noted in importance graph of maintenance need prediction (see Figure 5.12). Third, it can help in the feature selection and extraction procedure in order to train the models only on positively contributing feature set. Fourth, complying with business rules, it enables us to drop those features that are not directly relevant to the decision-making process.

5.6.2 Instance Level Interpretability

Feature importance analysis provides an understanding of the model’s prediction as a whole. However, the need to contemplate why and how a model classifies an unseen (new) instance to a specific label remains. In order to explain the model’s prediction for a single test instance, we have employed Local Interpretable Model-Agnostic Explanations (LIME) framework (Ribeiro et al., 2016). LIME explains the predictions of any model locally in an interpretable manner. This enables the domain experts and decision-makers to understand, interpret, and possibly further improve the model’s performance.

The LIME explanation of maintenance needs prediction model for a single instance is provided in Figure 5.15. The considered instance has been classified as *yes* with 98% probability. On the right side of Figure 5.15, a detailed explanation of relative features’ importance and their contribution to final classification are provided. The

explanations can be read as: A switch with *timber* bearer, *no* roller fitted, having a *age between 18 and 28*, with *ballast* track type, and with *shallow* stretcher type is noted to have *installation error*; therefore must be maintained. However, the *bi-directional* switch having *good as new condition* state with *flatbottom* rail type contributes to *no* maintenance needed. In contrast to standard performance measures to validate models, the LIME explanation has provided the feature level contribution and explicitly shows how each of them is relevant to final prediction probabilities.

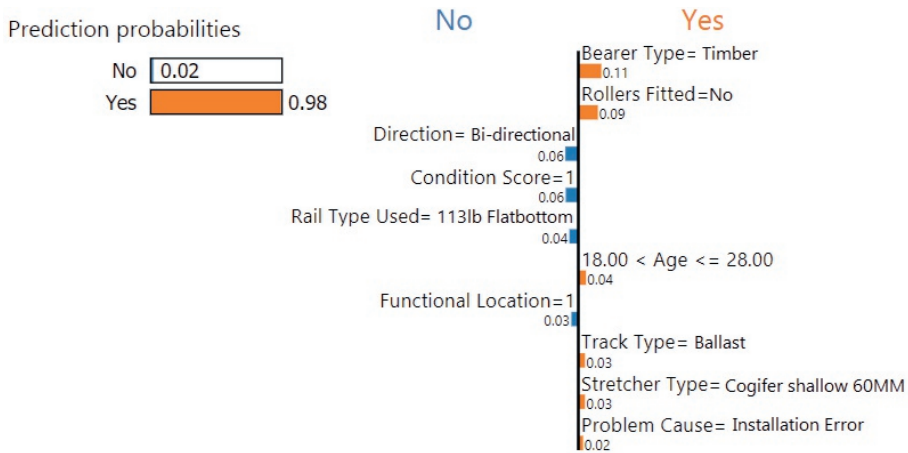


Figure 5.15: LIME explanation of maintenance need prediction model.

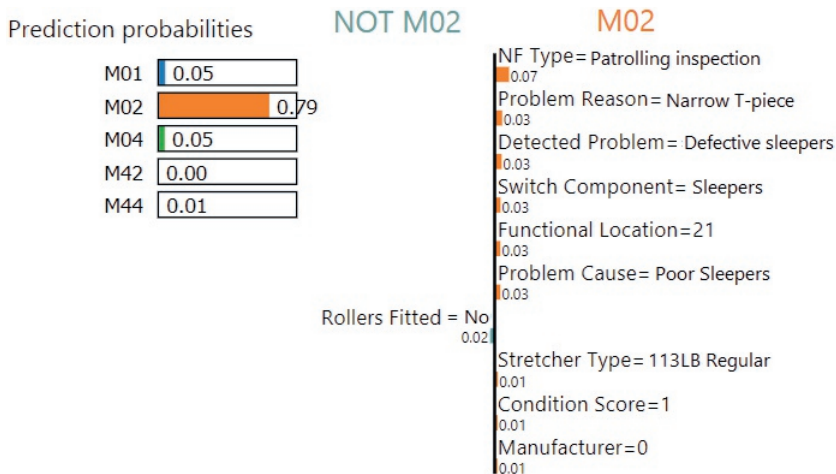


Figure 5.16: LIME explanation of maintenance treatment type prediction model.

Figure 5.16 and 5.17 provides the explanation of maintenance type prediction and maintenance trigger's status prediction models, respectively. Both of these models deal with multi-classification problem where a type of maintenance treatment or

a maintenance status trigger is predicted. By default, LIME explains the multi-classification models in a binary manner as can be noted with *M02* and *Not M02* in Figure 5.16 and as *Escalated* and *Not Escalated* in Figure 5.17. The LIME explanation of maintenance treatment type prediction model can be described as a *defective sleeper* particularly a *narrow T-piece* on *sleepers* is detected by *patrolling inspection*. Considering these features values, the models with 79% certainty predicts that *M02*: *defective material replacement* should be performed on the switch sleepers.

Similarly, for the maintenance trigger’s status prediction model the LIME explanation is outlined in Figure 5.17. The explanation shows that a switch is in *very good condition state*; however, a *free wheel passage on a switch nose* is reported due to *nose strike and wear*. This problem reported is termed as *out of tolerance* with *high priority*. Having analyzed these properties of all the features, the model with 88% certainty classifies the status of maintenance trigger as escalated to the technical manager for the final decision.

With the ability to better interpret the results of predictive models, domain experts and decision-makers can better interact with and trust the models’ prediction. This gives a reliable decision-aid to infrastructure managers to efficiently predict and plan the maintenance activities.

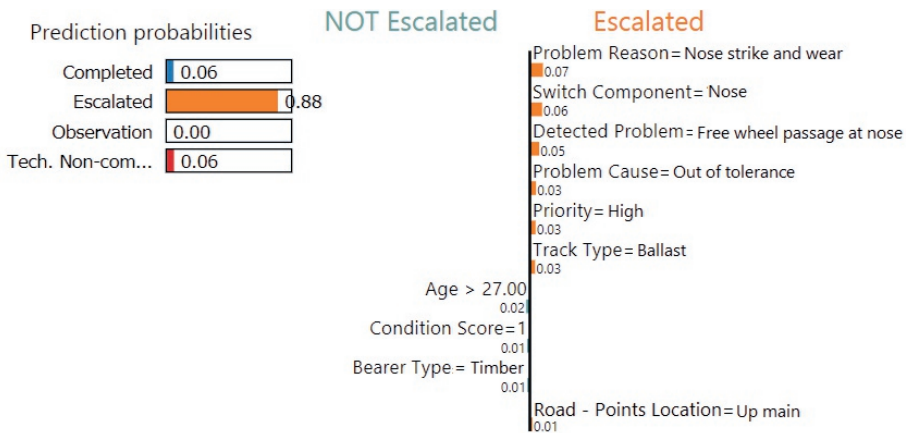


Figure 5.17: LIME explanation of maintenance triggers prediction model.

5.7 Discussion

This paper has applied the tree-based classification algorithms of machine learning for the development of maintenance predictive models. Multiple models based on decision tree (DT), random forest (RF) and gradient boosted tree (GBT) were trained and evaluated for the prediction of maintenance need, activity types and trigger's status. Unlike many studies that use additional data collection procedures such as monitoring systems and sensors, the proposed models utilize the data generated from the in-use business process of maintenance request from the SAP ERP system.

The purpose is to develop such predictive models that can be readily used as an aid in decision tool by infrastructure managers for maintenance planning. Learned from the historical data, the model analyses each maintenance trigger as an individual case and provide its prediction. Moreover, these predictive models bring efficiency, particularly for unplanned maintenance planning, as an enormous amount of notifications can be labeled with maintenance need, activity type, and trigger's status in a few seconds. Since the intention is to deploy the predictive models in real decision scenarios, our work especially takes into account the interpretability of the models' outcome. We explained the models' prediction by feature importance analysis as well as by per case details through employing LIME framework.

With the exploratory data analysis, development of maintenance classification models, and explanation of their result, few remarks can be noted:

First it is found that the data of notifications generated during 2017 was highly imbalance having only three distinct maintenance activity classes and one new trigger's status label. The data imbalance is the reason that the models show relatively weak predictive power on held-out validation.

Second for the completeness of models' evaluation, stratified cross-validation is performed where several data folds (data subsets) are used as training and test set iteratively. The models show overall good predictive performance with stratified cross-validation.

Third with feature importance analysis it is noted that functional location, age, and detected problems are the most important features. This can be very well explained with agency decision-making practices, where decision are mostly

based on age and importance of functional location and much less attention is given to other factors, namely, condition scores.

Fourth the instance level explanation using LIME shows the varying level of feature contributions for final prediction probabilities compared with feature importance analysis. This is because some feature may have a considerable impact on models' predictions for a single case, which would be visible with instance level graphs. However, the same feature may have been fragmented based on data values deep in a tree and thus attain a lower feature importance score globally.

The solution approach of predictive models' development and their results explanation have wider applicability and are generalizable to other assets types and maintenance planning scenarios. This is because we have utilized the data from a business process of plant maintenance module of ERP, which is currently being used in multiple railway agencies. Moreover, the maintenance request process, mentioned in this study, is a standard procedure of triggering unplanned maintenance irrespective of a specific asset type. Therefore, the details of datasets mentioned in Section 5.2 and structure of maintenance classification problem given in Section 5.4 can be utilized as instructions to develop predictive models for another type of assets. The sensitivity analysis of models' performance attributed to hyper-parameters tuning is deliberately skipped in this study. This is to illustrate how simple models, providing more than 70% of accuracy and trained on ERP structured data, can prove useful for maintenance prediction tasks. There are many studies in machine learning research that analyses the model's performance with respect to hyper-parameters across the different datasets. For an extensive overview, an interested reader may refer to van Rijn and Hutter, 2018.

From the practical perspective, it is worth noting that the real datasets evolve and may have new class labels in the future. Therefore, the models must be retrained periodically to accommodate the evolving nature of data while aiming for sound predictions.

5.8 Conclusions

This paper employs machine learning techniques for the development of predictive maintenance models. Multiple models based on decision tree (DT), random forest (RF) and gradient boosted tree (GBT) were trained and evaluated i) to predict if a maintenance will be performed or delayed as a result of trigger, ii) to predict

the maintenance activity type and iii) to predict the status of a maintenance trigger (notification). For the prediction of maintenance need, the GBT model performed most optimally as compared to other methods with 86% accuracy. For maintenance activity type and trigger's status prediction, the RF model attains an accuracy of 70% and 79% on the held-out test set, respectively. By collecting more data, specifically for minority classes, the predictive performance of the models can even be further improved. We show how the existing data readily available at railway agencies can be used for the development of ML models without the need for additional monitoring devices. Since the data is generated from in-use business processes, the models can be directly incorporated for maintenance decision-making. Moreover, we have introduced an approach to provide a detailed explanation of the outputs of ML models by illustrating the features' importance. The feature importance analysis shows which input features positively or negatively impact the model's output. In addition to global explanations, we have employed the local interpretable model-agnostic explanations framework to explain the model behavior locally for a specific instance. This interpretability of the models also provides an understanding of the decision-making process, and identify the most important data attributes which may help in future data collection procedures.

The future work of this study seeks to develop regression models in order to estimate the best time to perform the maintenance. This will require the data of past maintenance actions and operational details of the network. Such a regression model can be used by a railway agency for the budget planning as well as for the possession planning of the network. Another extension of this work is to implement multi-task learning using deep neural networks in order to develop a unified model for the prediction of related maintenance tasks.

5.9 References

- Akosa, J. (2017). Predictive accuracy: A misleading performance measure for highly imbalanced data. In *Proceedings of the sas global forum*.
- Andreou, A., Venkatadri, G., Goga, O., Gummadi, K. P., Loiseau, P., & Mislove, A. (2018). Investigating ad transparency mechanisms in social media: A case study of facebook's explanations. In *The network and distributed system security symposium (ndss)*.
- Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1), 20–29.
- Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, 13(Apr), 1063–1095.
- Böhm, T. (2017). Remaining useful life prediction for railway switch engines using classification techniques. *International journal of prognostics and health management*, 59.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees, wadsworth international group, belmont, california (1984). *Google Scholar*.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cabitza, F., & Banfi, G. (2018). Machine learning in laboratory medicine: Waiting for the flood? *Clinical Chemistry and Laboratory Medicine (CCLM)*, 56(4), 516–524.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357.
- Chen, J., Liu, Z., Wang, H., Núñez, A., & Han, Z. (2018). Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Transactions on Instrumentation and Measurement*, 67(2), 257–269.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). ACM.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.
- Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *Security and privacy (sp)* (pp. 598–617). IEEE.

- de Bruin, T., Verbert, K., & Babuška, R. (2017). Railway track circuit fault diagnosis using recurrent neural networks. *IEEE transactions on neural networks and learning systems*, 28(3), 523–533.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*.
- Al-Douri, Y. K., Tretten, P., & Karim, R. (2016). Improvement of railway performance: A study of swedish railway infrastructure. *Journal of Modern Transportation*, 24(1), 22–37.
- ERF. (2013). *Road asset management: An erf position paper for maintaining and improving a sustainable and efficient road network*. European Union Road Foundation. Brussels, Belgium.
- European Railway Agency. (2014). *Railway safety performance in the european union*. European Union Railway. Brussels, Belgium.
- EuroStat. (2016). *Railway passenger transport statistics - quarterly and annual data*. European Union Railway. Brussels, Belgium.
- Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, 15(1), 3133–3181.
- Flach, P. (2012). *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189–1232.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. Springer series in statistics New York.
- Guo, C., & Berkhahn, F. (2016). Entity embeddings of categorical variables. *arXiv preprint arXiv:1604.06737*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning*. Springer.
- He, H., & Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9), 1263–1284.
- Kauschke, S., Fürnkranz, J., & Janssen, F. (2016). Predicting cargo train failures: A machine learning approach for a lightweight prototype. In *International conference on discovery science* (pp. 151–166). Springer.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3–24.
- Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., & Hampapur, A. (2014). Improving rail network velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C: Emerging Technologies*, 45, 17–26.

- Lipton, Z. C. (2016). The mythos of model interpretability. *CoRR*. arXiv: 1606.03490
- Manco, G., Ritacco, E., Rullo, P., Gallucci, L., Astill, W., Kimber, D., & Antonelli, M. (2017). Fault detection and explanation through big data analysis on sensor streams. *Expert Systems with Applications*, 87, 141–156.
- Marquez, A. C. (2007). *The maintenance management framework: Models and methods for complex systems maintenance*. Springer Science & Business Media.
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7, 21.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Quinlan, J. R. (1993). C4. 5: Programming for machine learning. *Morgan Kauffmann*, 38, 48.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81–106.
- Ramos, J. et al. (2003). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133–142).
- Raschka, S., & Mirjalili, V. (2017). *Python machine learning*. Packt Publishing Ltd.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144). ACM.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, 9(1), 307.
- van Rijn, J. N., & Hutter, F. (2018). Hyperparameter importance across datasets. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 2367–2376). ACM.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Predictive maintenance planning of bridges using deep neural networks

¹ **Abstract:** Data-driven decision support can substantially aid to smart and efficient maintenance planning of road bridges. However, many infrastructure managers primarily rely on information obtained during visual inspection to subjectively decide on the follow-up maintenance actions. The subjective approach likely to lack the appropriate use of inspection data and does not promise cost-effective maintenance plans. In this paper, we show that the historical and operational data readily available at the agencies is of vital importance and can be used to recommend maintenance actions for bridges effectively. This is achieved by developing a machine learning system which is trained on the past asset management data and can provide support to the decision-makers in the condition assessment, risk analysis, and maintenance planning tasks. We have evaluated several traditional learning algorithms as well as the deep neural networks with entity embedding to find the optimal predictive models in terms of prediction accuracy. Additionally, we have explored the multi-task learning framework that has shared representation of related prediction tasks to develop a powerful unified model. The analysis of results shows that a unified multi-task learning model performed best for all the considered tasks followed by task-specific neural networks with entity embedding and class weights, which achieve close to 80% accuracy. The results of models are further evaluated by instance-level explanations, which provide insights about essential features and explain the data attributes that leads to particular predictions.

¹This chapter is under-review as: Allah Bukhsh, Z., Stipanovic, I., Saeed, A., & Doree, A. G. Predictive maintenance planning of bridges using deep neural networks.

6.1 Introduction

Functional and serviceable transport infrastructure presents one of the essential pre-dispositions for the economic growth of a country. Among other infrastructure objects, bridges represent a vital link in any roadway network. They provide the crossings at critical locations, reduce the travel times and maintain the traffic flow (Chen & Miles, 2004). Under limited financial resources (European Commission, 2018), agencies have to take prudent investment and maintenance planning decisions to improve the serviceability and availability of the bridges, to minimize their life-cycle cost, and to maximize the return on investments. To handle the amount of information required to achieve these objectives, many infrastructure owners use the computerized management systems to manage and process relevant data and to support the decision-making processes (Mirzaei, Adey, Klatter, & Kong, 2012).

Many agencies have developed Bridge Management Systems (BMS) tailored to their specific management needs. Mirzaei et al. (2012) provides an overview of BMS being used in sixteen countries. Similarly, Markow and Hyman (2009) explore how assets owners use the capabilities of BMS to get support in the decision-making of bridge programs. Many BMS primarily rely on information obtained in visual inspection to decide on the follow-up maintenance actions (Bu, Lee, Guan, Loo, & Blumenstein, 2014). These systems prompt inspectors to describe the physical state of the structure, which is quantified based on condition score card (Chase, Adu-Gyamfi, Aktan, Minaie, et al., 2016). The traditional quantification methods from visual inspection to condition rating rely on a subjective process, with a main assumption that a bridge inspector is experienced and trained personnel and has detailed knowledge of the structure (Gattulli & Chiaramonte, 2005). Since there is often no systematic procedure to record experts' preferences and their comprehension of structures and related performance objectives, the maintenance decisions become difficult to follow, justify, and replicate in the future.

Several useful reliability assessment and maintenance optimization models have been proposed in the literature (Alaswad & Xiang, 2017; Ghodoosi, Abu-Samra, Zeynalian, & Zayed, 2017; Hu & Madanat, 2014; Kong & Frangopol, 2003; Liu & Frangopol, 2004). However, frequently, the reliability assessment models are not part of computerized BMS; therefore, not every bridge gets an opportunity to have a detailed future performance profile. Likewise, the maintenance optimization models introduce complex mathematical heuristics to formulate and solve the problem. Therefore, the agencies still prefer to use traditional methods based on subjective ranking and

preferences of domain experts for the maintenance decision-making (Ahmad & Kamaruddin, 2012; Morcoux & Lounis, 2005; Sharma, Yadava, & Deshmukh, 2011). Multiple efforts have been reported in the literature to improve the functionalities of BMS for the decision-making tasks (Bush, Henning, Ingham, & Raith, 2014; Gattulli & Chiaramonte, 2005). However, the focus has mainly been on extending BMS capabilities to support in long term maintenance planning, and the whole life cycle costing of assets.

The theoretical progress and agency's practices highlighted three key challenges in the context of decision-support for maintenance planning. Firstly, a little attention is paid to investigate the solution that could improve the subjective assessment procedures from visual inspection of assets towards maintenance planning. Secondly, the historical data collected during the past visual inspections are not used for the decision-making process due to data access and analysis limitations (Wijnia & Herder, 2010). Thirdly, the condition and maintenance optimization models do not scale up to the network-level and they provide limited support in detailed condition assessment and maintenance planning. To tackle these challenges, we introduce the Machine Learning (ML) system that is trained on the historical data and can provide recommendations to the decision-makers on the condition assessment and maintenance planning tasks. By processing large and complex dataset, ML techniques can infer the patterns and rules that relate to a target class such as condition state. The automatic rules inference from the data enables the development of such systems that can either automate the decisions or provide the recommendations to the human decision-maker.

We used a large dataset of concrete bridges from the road agency to illustrate the development methodology. The dataset is collected over the years as a result of Inspection to Maintenance Advice (IMA) process which is implemented in a BMS. The IMA process collects the data of visual inspection, where, a decision-maker assess the data and decide on the condition state, risk level and recommends a maintenance advice on the bases of his/her technical knowledge and judgments. The objective of this study is to develop predictive models that can provide support in the subjective assessment procedure of bridge maintenance planning. This work mainly deals with three prediction tasks namely, assessment of *condition state*, analyses of *risk level*, and recommendation of *maintenance advice*, all by using the damage details noted during inspection activity. We applied various supervised learning algorithms from traditional machine learning to utilizing deep neural networks in order to find an optimal predictive model with the best accuracy. In addition, to discrete models for each task, we seek to develop a unified model where the shared representations can

improve the predictive power of a model by simultaneously learning to solve multiple tasks. Besides, we present the detailed interpretability of the models' predictions, which explains the most essential features and the instance-level prediction details. The interpretability of the models is important for domain experts to be able to trust and employ them in their daily decision-making scenarios (Andreou et al., 2018).

The primary contributions of this paper are summarized as:

- We develop several machine learning models and deep neural networks that can learn from the visual inspection data of the bridges. These models are introduced as a tool to support asset-owners in the subjective decision-making process by recommending appropriate condition state, risk levels, and maintenance actions.
- We present generic development methodology that utilize only the existing data generated from an in-use business process of a transport agencies. This results in predictive models that are aligned with current decision-making practices of the agency. Unlike other studies, this study does not perform additional data collection.
- This study is unique in comparing and applying logistic regression, tree-based models, neural networks with entity embedding and multi-task learning framework to find the best performing predictive model for the bridge maintenance planning.
- We provide the instance-level interpretability to explain the results of the optimal model for each task. The interpretability of the models highlights the important features and explains how a model makes certain predictions.

The rest of the paper is structured as follows: Section 6.2 presents an overview of the studies that utilize the advanced ML techniques for transport infrastructure maintenance. The problem domain and the detailed data description are discussed in Section 6.3. Section 6.4 provides an overview of the methodology by highlighting the learning algorithms, the neural network's architectural details, including hyperparameters, and the evaluation strategy. The details of the experiments and results for each prediction tasks are provided in Section 6.5. The interpretability of the models' results are explained by global and instance level interpretation in Section 6.6. The key remarks and general observation are provided in Section 6.7. Section 6.8 highlight the major outcome of this work and provide a future research agenda.

6.2 Related work

Machine learning (ML) techniques have achieved significant success in many industries ranging from health-care, finance, manufacturing, transport and marketing. Due to advancement in communication and sensor technologies, a fundamental shift in the management of transport network is noted. In the section, we discuss the studies that apply the ML techniques for the management of road and railway structures.

Masino, Thumm, Frey, and Gauterin (2017) proposed an infrastructure monitoring system based on vehicle sensors and supervised learning algorithms in order to estimate the road quality. Similarly, Souza, Giusti, and Batista (2018) introduced a low-cost system to evaluate the pavement condition by using the vibration readings from the accelerometer sensor of smartphones. Morales, Reyes, Caceres, Romero, and Benitez (2018) proposed a methodology to automate the prediction of maintenance interventions for the road pavements using the operational and historical maintenance data. Kiani, Camp, and Pezeshk (2019) applied the learning techniques to develop the fragility curves for the seismic risk assessment of the civil structures. From the railway domain, few notable studies are failure prediction models using heterogeneous data from multiple-detectors systems (Li et al., 2014), predictive models to detect metro door failure (Manco et al., 2017), recurrent neural networks to identify and capture the failures in railway track circuits (de Bruin, Verbert, & Babuška, 2017), finding and localizing damages in railway bridges (Chalouhi, Gonzalez, Gentile, & Karoumi, 2017), remaining useful lifetime of an electrical power switch (Böhm, 2017), and maintenance need and type prediction for switches and crossing (Allah Bukhsh, Saeed, Stipanovic, & Doree, 2019). An interested reader may refer to (Abduljabbar, Dia, Liyanage, & Bagloee, 2019; Spencer Jr, Hoskere, & Narazaki, 2019) for a comprehensive overview of relevant studies.

It can be noted that majority of these studies employ additional data collection means to continuously monitor the object in order to develop predictive models. Though useful for experimentation, it is expensive and impractical to mount monitoring devices on multiple assets across the network and continuously collect the data for a longer period of time (de Bruin et al., 2017). Additionally, many of the studies mentioned above focus on the development of the monitoring system which may not be aligned with transport agencies current management processes. Similarly, existing work though use sophisticated algorithms but do not address the topic of interpretability to explain the decision logic of the models. In this paper, we only utilize the historical data and propose a methodology that can be implemented

within a transport agency for the decision-making of bridge maintenance planning. Furthermore, special attention is given towards the model interpretability in order to avoid developing black-box models by elaborating on the results of the model at an instance and global level.

6.3 Problem domain and data description

A road agency has shared a large dataset of concrete bridges to analyze the applicability of machine learning approaches for providing decision-support in assessment of condition states, risk levels and maintenance actions. Here the data and name of the agency are anonymized owing to the confidentiality agreement.

The agency uses the custom BMS to store inventory, condition states, risk profiles and maintenance plans of road bridges. In total, the highway network consist of approximately 3800 bridges. To support the network-oriented asset management approach, all the civil structures (physical objects) of the road network are introduced with standard decomposition (NEN, 2010; van der Velde, Klatter, & Bakker, 2013). An example of the decomposition of road network assets is presented in Table 6.1.

The focus of this study is on *object* level, specifically on the bridges. Depending on the structural details, a bridge consists of several *elements* and *components*. It is important to note that not every element of the bridge is equally important in the overall structural integrity of a bridge. A weighted average method, conducted to elicit the relative importance of bridge components, reveals the superstructure, bearing, abutment and joints as most relevant elements for structural performance of the bridge (See Table 3 of Allah Bukhsh, Stipanovic, Klanker, O’Connor, and Doree (2018)). Therefore, the problem domain of this study covers the predictive modeling of bridges having a superstructure, bearing, abutment and joints as the main elements.

Table 6.1: Example of decomposition of road network assets (van der Velde, Klatter, & Bakker, 2013)

Level	Example
Network	Highway network
Sub network	Ring road system
Network branch	Highway between interchanges
Object	Bridge, tunnel, road section
Element	Superstructure, abutment, bearing
Component	Top layer, seal of joints

Additionally, the selection of the four elements is also motivated by the amount of data available for them.

6.3.1 The process of Inspection to Maintenance Advice (IMA)

Inspection is the integral tool for the infrastructure asset management. The inspection framework of the considered road agency consists of three types of inspection, namely routine, general and principal inspections. Routine and general inspections are aimed at the detection of unexpected failures. The principal inspection also called major maintenance inspection is targeted towards prognosis of future maintenance need of the infrastructure. The agency applies the risk-based inspection procedure, which follows the set of guidelines to foresee maintenance need and develop future maintenance strategy (Bakker & Klatter, 2012). Based on the data generation procedure of BMS, a detailed process from inspection of a physical asset to the maintenance advice is outlined in Figure 6.1.

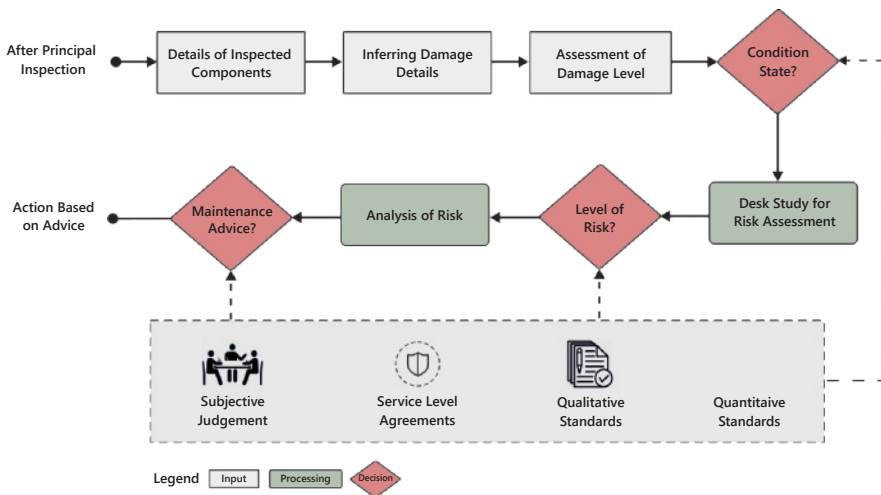


Figure 6.1: Process of Inspection to Maintenance Advice (IMA).

The IMA process primarily relies on a subjective analysis, where the desk study, interviews with inspection managers, and risk analysis are performed before and after the inspection activity to decide on the condition state, risk level, and further maintenance actions. The major inspection is executed every six years for prognosis of maintenance need and to implement control measures in time. The details of the inspection are recorded at the *element* level, whereas, any noted damages, their cause, the type of damage and its level are noted at the *component* level. Afterwards,

the *condition score of a bridge component is quantified* on a standard scorecard based on subjective analysis, quantitative standards and service level agreements. Next, the desk study is performed in which the noted damages and condition scores are regarded as risk indicators which are multiplied by their probability of occurrence to *quantify the level of risks*. The noted risk on the element of the bridge is controlled through taking certain maintenance measures. The asset owners and inspection managers issue *maintenance advice* from a standard list to trigger the maintenance actions.

The IMA process is guided by the risk-based inspection procedure where the performance of the structure is the central theme. This implies that even a component has a significant damage and poor condition state, but if the performance of the component is not likely to be affected, then the risk is regarded as negligible (Bakker & Klatter, 2012). This procedure ensures that the maintenance actions are not driven by condition score card only, instead the estimated risk profiles and the future maintenance need are considered. However, the process from inspection to maintenance advice is subjective, where due to risk consideration, a direct link between damage, condition, and risk level may not be established. The asset owners have to conduct many interviews and consult quantitative and quantitative standards along with the performance requirements to support the decision aspects of the IMA process.

In this paper, we aim to develop classification models that could learn only from historical data of the IMA process to assist in the decision-making of condition assessment, risk level, and maintenance advice. Figure 6.1 depicts the decision aspects of the IMA process with a diamond shape that will be supported by the ML classification models. Additionally, for the development of predictive models, we utilized the basic details of a bridge such as age, route, type, material and the noted damage as depicted with *rectangle shape* in Figure 6.1. Further discussion on the used data and its characteristics are presented in the following sections.

6.3.2 Data acquisition from BMS

The data generated from the IMA process is used for the development of classification models. BMS stores all the relevant data in SQL relational database system. Since, the different data are recorded based on the decomposition of the road network (as shown in Table 6.1), we have to execute several SQL queries to obtain all the required data. In the following, the brief details of the acquired data are provided.

- *Bridge inventory*: It presents the necessary details of the bridges including their description, location, construction year, route, and connection to a network branch. Besides, the bridge inventory data also provides the particulars of each bridge element along with their components (see Table 6.1).
- *Inspection data*: The principal inspection is performed every six years for each element of the bridge. The inspection data file provides the details of the principal inspections conducted from 2007 to 2017. It constitutes feature like inspection year, element id, inspection type, inspection location, and temperature on the day of inspection.
- *Damage data*: During the inspection of the elements, the damages are also inspected and recorded at the component level. The damage data file presents component id, damage types, its possible cause, a detailed description and scale of the damage. The details of the damage help in the assessment of the physical state of the element.
- *Risk data*: With the inspection and damage data, the risk over the bridge elements is assessed during a desk study. The risk data file outlines the record of all the noted risks on elements, their analysis, the risk status, and the risk type. Furthermore, an estimation of the severity of the risk is also noted. To eliminate the observed risk on bridge elements, the asset and inspector manager determine the appropriate maintenance advice.

The presented data sources are interconnected with a unique object identifier (id), element id, component id, inspection id, damage id, and risk id. Since, a SQL database consists of a collection of tables, these unique identifiers have enabled us to execute *join operations* and retrieve all the relevant data for each of the component. The obtained datasets underwent an extensive filtering and cleaning process to retrieve only those data instances and features that are relevant for the development of classification models.

6.3.3 Features engineering

Feature engineering is one of the most crucial step in the machine learning (ML) model development pipeline. The feature engineering task constitutes of preparing data for learning algorithms through extracting useful features from the given data, combining similar features, and eliminating the least relevant features. The quality

and quantity of features play an intrinsic role in the predictive ability of the model. This task mainly takes into account the domain knowledge of experts who decide the relevancy of features for the dependent variable (i.e. class label).

We performed the feature engineering task on the data collected from the IMA process during the year 2007 to 2017. Guided by multiple interactive sessions with experts, we eliminated duplicated features such as, the condition codes and their explanations and other irrelevant features, e.g., the location coordinates, the dimensional properties, and unique identifiers related to the inspection activities. We also eliminated those instances for the bridges that were constructed before the year 1900, as they follow special maintenance procedure and have a lot of missing data. Additionally, we eliminated all the data instances that do not have any noted damages, relevant risk details, or maintenance advice. Without the specific damage details, the condition assessment is purely subjective driven process with no available data for the ML model development.

The experts also directed the process of extracting relevant details of the IMA process from the BMS database. The obtained data sources are discussed in Section 6.3.2. The exhaustive feature selection procedure reduced the total number of feature from 69 to 20 features, excluding the class labels. The tasks of condition state, risk level, and maintenance advice prediction are sequential, as depicted in Figure 6.1, which means the output of one task may be used for the prediction of other tasks. However, for the sake of robust classifier and to avoid data leakage problem (Kaufman, Rosset, Perlich, & Stitelman, 2012), all prediction tasks are trained on the same set of features data (see Table 6.2), where the output of the one model is not used as a feature for learning the other task. For instance, the predicted condition state value is not used for risk level predictions.

Table 6.2 provides the selected features set along with their type. The categorical features are assigned with representative numerical codes, while considering their ordinal properties. For the continuous features, we performed feature scaling depending on the requirements of the used algorithm as discussed in Section 6.4. For instance, tree-based algorithms such as decision tree, random forest are invariant to feature scales (Friedman, 2001), whereas the neural network demands the data normalization. In order to represent the decomposition of the bridge during the training of the model, the features representing the unique identifiers of the bridge, its *elements* and relevant *components* are part of the final dataset. Moreover, the dataset also had free-text features providing the detailed description of bridges and noted damages. We initially calculated the term frequency-inverse document frequency

Table 6.2: Selected features set for developing predictive models

No.	Feature name	Type	No.	Feature name	Type
1	Bridge-id	Discrete	2	Age	Continuous
3	Bridge-nature	Categorical	4	Bridge-route	Categorical
5	Bridge-material	Categorical	6	Segment-material	Categorical
7	Element-id	Discrete	8	Element-name	Categorical
9	Element-material	Categorical	10	Component-id	Discrete
11	Component-name	Categorical	12	Component-material	Categorical
13	Temperature-insp	Continuous	14	Inspection-detail	Categorical
15	Inspection-point	Categorical	16	Damage-Cause	Categorical
17	Damage-type	Categorical	18	Damage-level	Categorical
19	Damage-category	Categorical	20	Weather-insp	Categorical

(tf-idf) from the text to use these features for training the model (Ramos et al., 2003). However, during the earlier exploration we noted that the text features do not contribute towards the models' performance. Therefore, these columns were removed for further consideration.

6.3.4 Visual analytics

This section provides some useful insights about the IMA dataset. The visual analytics helps in interpreting and analyzing the characteristics and representation of data.

The Figure 6.2 presents the age distribution of 2960 bridges that are part of the IMA dataset. Approximately, 75% of all the bridges are between the age range of 21 to 60 years old, whereas less than 10% of bridges are within 1 to 20 years old.

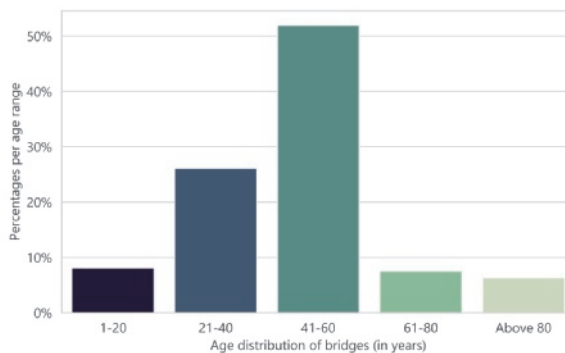


Figure 6.2: Age range of the bridges in years.

The principal inspection of the bridges leads to the identification of multiple damages. To provide an overview of the noted damages, Figures 6.3 and 6.4 presents the 15 most occurring damage types and causes, respectively. Aging is one of the most frequent causes of the damage on the bridge. It can be further verified by the age distribution of the bridges, where at least 50% of them are above 40 years of age. Similarly, in damage types (see Figure 6.4), the surface corrosion, wear, tear and end of life can be attributed to aging as well as to the traffic load.

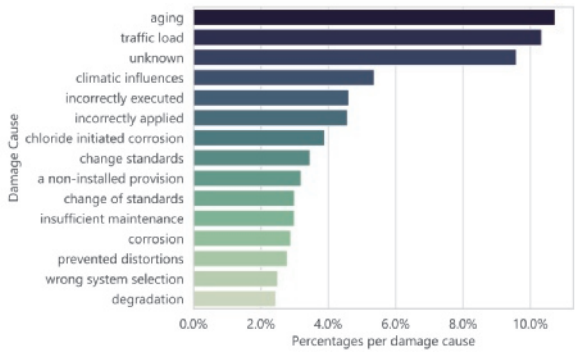


Figure 6.3: Frequently identified damage causes on elements & components of bridges.

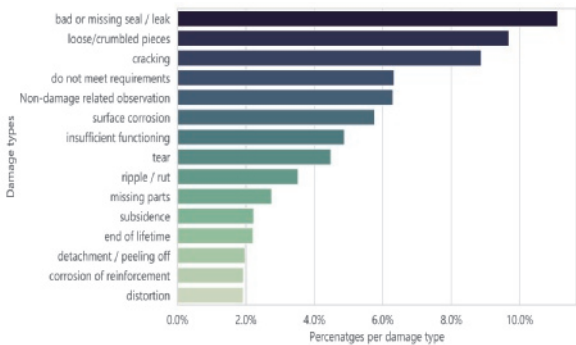
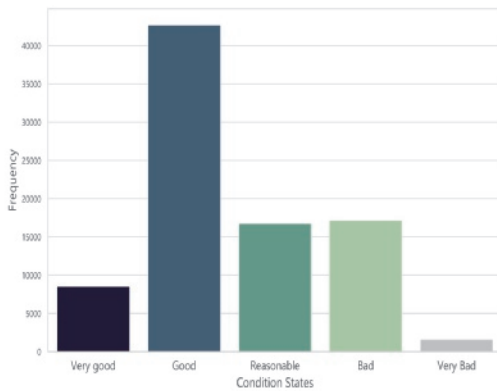
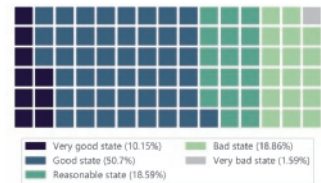


Figure 6.4: Frequently identified damages types on elements & components of bridges.

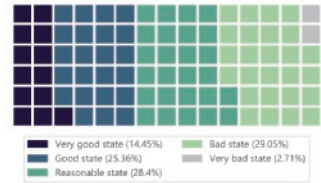
The supervised learning algorithms require the labeled data for the model training and testing purposes. In our case, the condition state, risk level, and maintenance advice are the class labels for which the distinct models are developed. The predictive models learn well (i.e. generalize to unseen data instances) when each class has at least one-tenth representation in the overall dataset (He & Garcia, 2008). A balanced dataset has an equal representation of all the classes. However, in the case of imbalanced dataset, the model tends to be biased toward the class having major representation, thus performing poorly for the minority classes. We present the frequency distribution



(a) Frequency distribution of condition state classes.

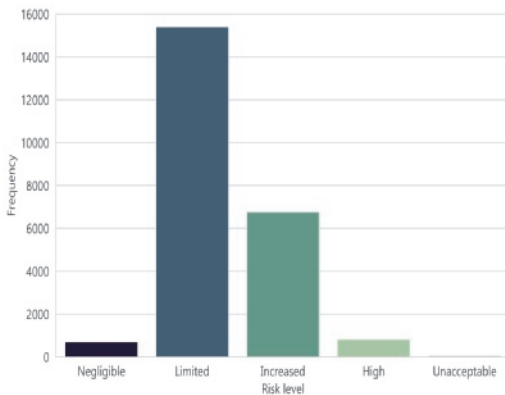


(b) Complete dataset.

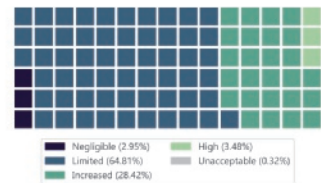


(c) Undersampled dataset.

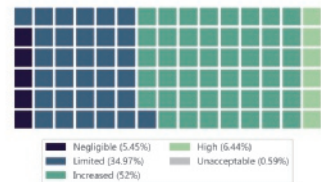
Figure 6.5: Representation of condition state classes in the overall data set.



(a) Frequency distribution of risk level classes.



(b) Complete dataset.

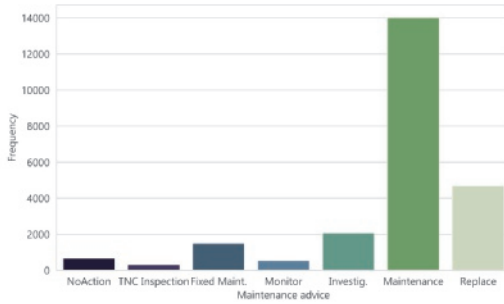


(c) Undersampled dataset.

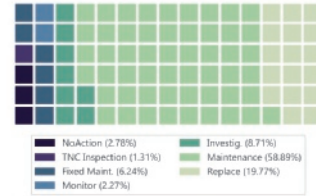
Figure 6.6: Representation of risk level classes in the overall data set.

and the percentage distribution for each class in Figures 6.5, 6.6 and 6.7 for all three classification problems. Besides, these figures also introduce the class labels of predictive tasks, where the meaning of the labels is self-explanatory.

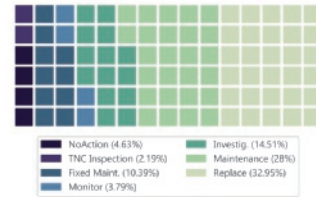
The visual analysis presented in Figures 6.5a, 6.6a, 6.7a reveals that the class distributions of condition state, risk level and maintenance advice is highly imbalanced. The *good condition state* class has more than 40k instances and represents more than



(a) Frequency distribution of maintenance advice classes.



(b) Complete dataset.



(c) Undersampled dataset.

Figure 6.7: Representation of maintenance advice classes in the overall data set.

50% of the overall dataset as shown in Figure 6.5b. The risk level classes have even higher class imbalance with the majority class i.e, *limited*, having more than 15k instances and represent 64% of the overall dataset (see Figure 6.6b). Likewise, the maintenance advice classes are also imbalanced, where *maintenance* class has 14k instances which covers approximately 59% of all the data as depicted in Figure 6.7b.

In addition to using the complete imbalanced dataset for model training and evaluation, we also performed random under-sampling of the majority classes to determine if the under-sampling approach will improve the learning ability of the predictive models. In random sampling, each data point has an equal probability of selection when the data instances are independent and identically distributed (He & Garcia, 2008). For condition state, risk-level and maintenance advice prediction, a majority class is under-sampled where only 35%-40% of its instances are randomly selected. Though, the under-sampling does not balance the dataset, it improves the representation of the minority classes to a certain extent. To put this in perspective, Figures 6.5c, 6.6c and 6.7c provides the under-sampled class distributions. As also mentioned earlier, each classification task is dealt in a discrete manner, where the same features set as presented in Table 6.2 is utilized for the condition state prediction, risk level prediction, and maintenance advice prediction. However, it is important to note that even though each element has a certain condition state, but not every element has an associated risk. Hence, the number of data instances available for developing condition state models are higher than the other classification tasks.

6.4 Methodology for prediction of maintenance related tasks

Machine Learning (ML) is a scientific study of algorithms that can extract useful patterns from the raw data in order to facilitate the data-driven decision-making. The ML techniques have enabled computers to tackle complex problems using real-world data. In supervised learning, a ML model learns from the labeled training data to find the relationship between x features and y target class. A well-trained model \hat{f} must be able to make accurate prediction \hat{y} given unseen future data instance \bar{x} . Depending on the specific learning problem, there are a multitude of algorithms to elicit \hat{f} from the dataset (Wolpert, Macready, et al., 1997). The choice of an optimal algorithm depends on the target output and the size and format of the available dataset. According to the ‘No free lunch’ theorem, no single algorithm is significantly superior to others (Wolpert, Macready, et al., 1997). In practice, we have to try a handful of a different algorithms to train, evaluate, and select the best performing model. This section presents the overall methodology to develop accurate predictive models for the IMA dataset in order to support in the subjective decision-making process of bridge maintenance planning.

The data generated from the IMA process is annotated and has a structured nature. We selected supervised algorithms from traditional machine learning and from deep learning paradigm to find the best performing model for the prediction of condition state, risk level and maintenance advice. In the following sections, we first briefly introduce the ML algorithms that are used for the development of predictive models. Next, we introduce the deep learning paradigm and also motivate our choice of utilizing neural networks for the structured dataset. The detailed explanation of each algorithm is out of the scope of this study; instead, an interested reader may refer to Trevor, Robert, and JH (2009). Finally, we discuss the various evaluation approaches and performance measures that are applied to gauge the predictive ability of the developed models.

6.4.1 Machine learning techniques

Among the several ML techniques, the utilized algorithms include logistic regression, decision tree, random forest, and gradient boosting trees. The logistic regression algorithm is selected to establish the baseline performance. The tree-based algorithms consisting of decision-tree, random forest and gradient boosting trees are chosen

because of their proven prediction performance on structured datasets in many industry challenges and academic literature (Fernandez-Delgado, Cernadas, Barro, & Amorim, 2014; Kotsiantis, Zaharakis, & Pintelas, 2007). This section briefly introduces these ML techniques that will be applied to develop predictive models for bridge condition, risk, and maintenance advice prediction. Furthermore, we explain details of the development and hyper-parameters tuning of each model.

Logistic regression is a classification algorithm that performs well for the linearly separable classes. It takes linear combination of weights and values, maps them to real-valued number, and outputs the predicted probabilities. The resulting probability presents a likelihood that a particular sample belongs to a specific class. By applying a binary threshold function, the discrete classification can also be obtained from the predicted probability.

Decision Tree works on a simple strategy of divide-and-conquer by employing the recursive partitioning of the data. The key idea is to split the dataset number of times, where the resulting sets are homogeneous and belong to the same target class. The algorithm applies the top-down greedy search to determine the best split of nodes until the maximum allowable length of a decision tree is reached, and the terminal nodes are the target classes (Breiman, 2017).

Random Forest is an ensemble approach of the decision tree. Unlike, decision trees which builds a single tree for an entire dataset, random forest randomly selects the instances and features of data to construct multiple trees in parallel fashion. The central idea of random forest is to average the result of many decision trees, which individually suffer from high variance (Breiman, 2017). This ensemble learning approach results in a robust model which is less susceptible to over-fitting.

Gradient Boosting Trees is an alternative ensemble learning technique that consecutively produces weak tree classifiers in a stage-wise fashion (Friedman, 2001). The boosting approach strategically resample and sequentially build multiple trees for instances that are difficult to estimate with previous ones by minimizing some arbitrary differentiable loss function, e.g. cross entropy or sum of squared errors. In other words, the idea is to convert the weak learners into a strong learner sequentially, where each weak learner tries to improve upon its predecessor.

Figure 6.8 shows the visual representation of tree-based models adopted from (Xristica, 2016). The decision tree develops a single tree dataset, whereas the random forest develops multiple trees at a time over the randomly selected data. The gradient

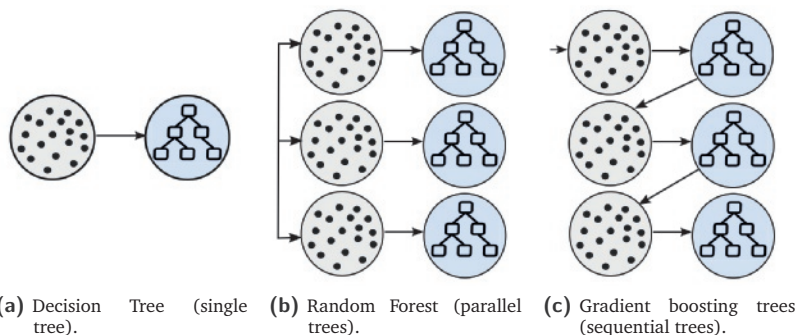


Figure 6.8: Conceptual representation of tree-based learning techniques.

boosting trees also develops several estimators but in a sequential manner. All the models are implemented on the IMA process dataset using the scikit-learn library of python (Pedregosa et al., 2011). The hyperparameters of the models are selected using the random search method (Bergstra & Bengio, 2012). The parameters are tuned by empirically optimizing the results of models over the subset of training data, also called validation set.

6.4.2 Deep learning techniques

Deep learning is a subfield class of ML whose algorithms are inspired by structure and function of brain called Neural Networks (NN). The pioneer researchers of deep learning defines it as ‘algorithms that seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels’ (Bengio et al., 2009). To put simply, the deep learning algorithm follow several layers of abstraction to learn complex function mappings, rather than direct input to output (Bengio, 2012).The NN extract the useful features from data automatically and can improve themselves without human interventions, whereas ML algorithms require the clear set of manually extracted features and may require additional data to improve its predictive performance (LeCun, Bengio, & Hinton, 2015, Chapter 1). The motivation to explore NN for the given prediction tasks is due to their state-of-the-art performance in computer vision (Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014), speech recognition (Hinton et al., 2012; Sainath, Mohamed, Kingsbury, & Ramabhadran, 2013), and natural language processing (Bengio, Ducharme, Vincent, & Jauvin, 2003; Kim, 2014).

In this study, we hypothesize that compared to traditional learning algorithms for the structured data, the NN combined with entity embeddings can result in better and robust predictive models for the condition state, risk level, and maintenance advice prediction. The ability of NN to learn complex non-linear representations from the data not only improve the prediction task at hand but can also be transferred for the related prediction tasks where the data is insufficient, outdated or unlabeled in nature (Bengio, Courville, & Vincent, 2013). In this section, we briefly introduce the basic structure of the neural network, followed by explanations of entity embedding for the categorical data. Along the lines, we present the concept of cost-sensitive learning (also called class weights) in order to manage the imbalanced nature of the IMA dataset. Next, we explain the multi-task learning framework of NN which tries to learn several tasks simultaneously instead of developing a discrete model for each task.

A **neural network** is typically represented by a network diagram consisting of several layers. The basic computation unit is a node (also called neuron) which contains an activation function such as a sigmoid function or rectified linear unit (ReLU). The supervised learning procedure within NN has a cyclic pattern, where the forward activation flow of output and backward error propagation for the weights adjustments is repeated number of times. The backpropagation is based on a learning rule such as perceptron learning, delta learning, etc., which modifies the weights of the edges based on the input pattern. To put in other words, when the data is presented to the NN for the first time, the output layer provides a mere guess of the output. This procedure is called forward activation flow. Based on the output, appropriate adjustments are made concerning the logic of learning rule and associated weights, which is referred to backpropagation.

In principle, a neural network can approximate any continuous function since the data continuity guarantees the convergence of the optimization (see Nielsen (2015) for an interactive visualization). However, structured data with their categorical features lack required continuity, which limits the application of NN. Even with coded categorical features, the NN do not work well as the numerical coding eliminates the informative relations among the features. Guo and Berkhahn (2016) proposed to use the entity embedding to learn the representation of categorical features in a multi-dimensional space. Given that IMA dataset comprises of categorical and numerical features, we implemented **Neural Networks with Entity Embeddings (NN-EE)** in this paper. The architecture of NN with entity embedding is depicted in Figure 6.9.

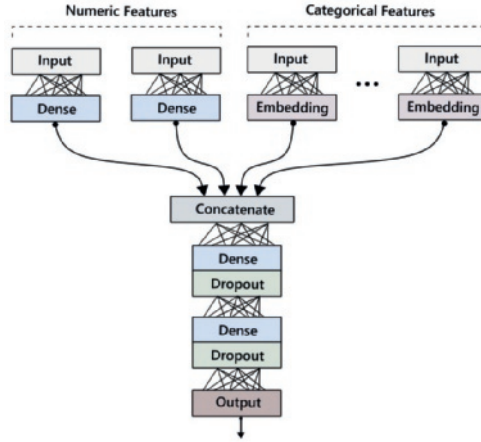


Figure 6.9: Architecture of Neural Network with Entity Embeddings (NN-EE).

All the categorical features in the dataset (see Table 6.2) are assigned with numerical codes which are mapped as vectors to develop entity embedding. The mapping is equivalent to an extra layer of neurons, which is added on the top of the input layer and is learned in an end-to-end manner. The numerical features are fed directly to the fully connected layers with 20 hidden units. The output of the embedding layers and the fully connected layer are concatenated and connected to two fully connected layers each having 128 and 64 hidden neurons. After each dense layer, we applied dropout with 0.1% probability which randomly drops the neuron from the layers to avoid overfitting and to improve the generalizability of the model. We also applied L2 regularization to the weights of dense layers. At the output layer, the softmax function is applied to obtain the normalized output probabilities, where a class having the highest probability is the predicted class. To tackle the class imbalance problem, mentioned in Section 6.3.4, we applied cost-sensitive learning through using weighted categorical cross-entropy loss function. In this case, the weights are assigned to the classes based on their distribution in the training set, where the higher weights are assigned to the minority classes and lower weights to majority class. The **NN-EE with class weights (NN-EE(cw))** handles the data imbalance problem at the algorithmic level without performing any under or oversampling of the actual training data. For all the prediction tasks, we perform experiments NN-EE with and without applying class weights to analyze the impact of class weights on the overall model prediction ability.

In the above-mentioned learning settings, there is a one task to solve by minimizing a single loss function. Though the prediction of condition state, risk level, and

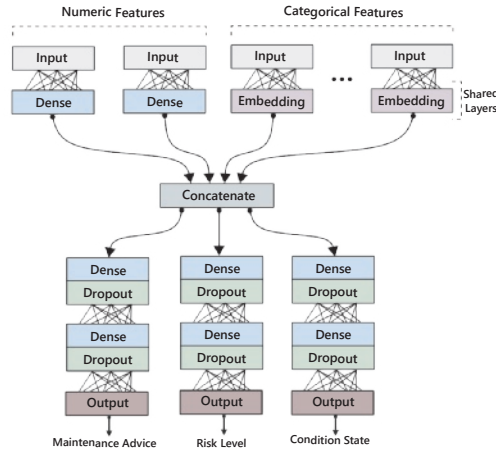


Figure 6.10: Multi-task neural networks with entity embedding and two shared layers.

maintenance advice are related, the single-task models treat them independently. The framework of multi-task learning argues that single task learning may ignore the potentially useful information that is available from the related tasks (Caruana, 1997). Multi-task learning aims to develop a unified model by using shared hidden layers which are trained in parallel on all the related tasks. Therefore, the **Multi-Task Learning Neural Networks (MTL-NN)** consists of common layers across multiple tasks as well as task-specific layers. Figure 6.10 presents the architecture of MTL-NN which seeks to develop a single unified model for the prediction of all three tasks. MTL-NN is performed through hard or soft parameter sharing (Caruana, 1997). We applied hard-parameter sharing in which initial hidden layers are shared across all the tasks whereas the final layers are problem specific. The entity embedding layers as well as dense layers for the numeric features are shared among the tasks. Likewise, the task-specific layers have the same configuration as noted in single-task architecture. We applied L2-regularization on shared and on task-specific layers to avoid over-fitting. Finally, the optimization of the loss function is done simultaneously by alternating between different tasks randomly. The categorical (weighted) cross-entropy is optimized as an objective function by ‘Adam’ optimizer. Similar to a single task NN-EE, we applied class weights to MTL-NN to tackle the class imbalance problem. The architecture of NN-EE and MTL-NN shown in Figure 6.9 and 6.10 and their parametric details are implemented with python neural network API keras (Chollet, 2015).

The objective of MTL-NN is to benefit from the shared representation learning, where the representations learned for one task may improve the learning of other tasks. The

shared representation in MTL-NN claims to improve the generalization performance of multiple tasks when they are related (Zhang & Yang, 2017). Additionally, the embedding and representation learning from multiple tasks can be used for the prediction of related tasks in the future.

6.4.3 Evaluation approaches

An optimal predictive model must be able to generalize well for new (unseen) data. We applied stratified random sampling and cross-validation to evaluate the performance of the models based on various metrics. In the following, a description of evaluation approaches and performance metrics are provided.

Stratified random sampling

In Stratified Random Sampling (SRS) approach, the entire dataset is randomly split into training and test sets. In contrast to the standard sampling, the SRS ensures that each data split has equal target class representation. The training set is used to train the model, and the test set is used to evaluate the model's performance. Typically, the 70% of the data instances are selected for training and the rest 30% are used for testing. We also further split the training set to get the validation set, which is used to tune the hyper-parameters of different models by applying a random search strategy (Bergstra & Bengio, 2012). In the final evaluation phase, the validation set was then combined, and the model is trained on the whole training set and evaluated on the held-out test set.

Stratified cross validation

In Stratified Cross Validation (SCV), the whole dataset is randomly split into a number of equally sized units referred as 'folds'. By having N number of folds, the $N - 1$ are used for the training, while the N_{th} fold is used for the model testing. This process is repeated N times until each fold had the opportunity of being used as N_{th} test and training fold. Finally, the output is averaged across all folds to estimate the performance of the model. This method ensures that every data point is used at least once as a training example and once as a test example. The SCV is performed for the completeness of validations and the evaluation of the model's robustness.

Performance metrics

There are several measures which can be used to evaluate the performance of the predictive model. For the classification tasks, the confusion matrix analysis is applied which represent the models' predicted classes on test data for which the true values

are already known. It is essential to introduce the confusion matrix first in order to explain relevant performance measures. Table 6.3 shows the confusion matrix for the binary classification problem having positive and negative as target classes.

Table 6.3: Confusion matrix for binary classification problem.

	Classified negative	Classified positive
Actual negative	True Negatives (TN)	False Positives (FP)
Actual positive	False Negatives (FN)	True Positives (TP)

The values at principal diagonal confusion matrix (i.e., TN and TP) represent the correct classification by a model. However, the secondary diagonal values (i.e., FP and FN) show the misclassifications. To elaborate further, the False Positives (FP) are *positive* samples that are incorrectly classified as a *negative* class. Similarly, False Negative (FN) are *negative* class samples incorrectly classified as *positive*. A normalised confusion matrix with perfect classification has TN and TP of value *one* and FP and FN of value *zero*. With the confusion matrix, a number of performance measures can be calculated (Flach, 2012). The metrics used in this study are explained as follows:

Accuracy is a measure of correct predictions compared to the available data instances. It shows how often the model classifies the instances correctly. The accuracy is a good measure when the data is balanced for each class. However, in the case of imbalanced dataset, this metric without other performance measures can be misleading. The accuracy is computed as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

F-score is a combination of precision (or positive predictive value) and recall (sensitivity) measures (Sokolova & Lapalme, 2009). The precision determines the exactness of the model. It is a ratio of correctly predicted positive instances (TP) to the total positively predicted instances (TP+FP). In contrast, recall provides a measure of the model's completeness. It is a ratio of correctly predicted positive instance (TP) to the total instance of the positive class (TP+FN) in test data. In other words, precision represents the model's performance with respects to false positives, whereas recall shows the performance with regards to false negatives. The F-score convey the balance between precision and recall by taking their weighted harmonic mean. F-score is calculated as follows:

$$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Similar to the accuracy, F-score performs well with the fairly balanced dataset. In the case of an imbalanced dataset, the adjusted F-measure is utilized.

Kappa presents an inter-rater agreement between qualitative items, which measure the relative observed agreement (p_o) with the hypothetical probability of chance agreement (p_e) (Cohen, 1960). The kappa measure does not only calculates the percentage accuracy but also consider the possibility of an agreement between raters (qualitative items) by chance. The value of kappa is calculated as follows:

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e}$$

In the case of imbalanced datasets, the kappa is a robust measure compared to F-score and accuracy. It can be said that kappa determines how well a model performed (p_o) as compared to how well it could have been performed by chance (p_e) while considering the marginal distribution of a target class. The value of kappa ranges between -1 to 1.

6.5 Results

For three classification tasks, several predictive models are developed by applying *learning methods* and evaluated using the *performance metrics*. The robustness of the models are validated by training and evaluating them in several folds of the dataset. The evaluations of predictive models for each tasks as well as the results of multi-task learning neural networks are reported in this section.

6.5.1 Condition state prediction

The condition state prediction is a multi-class classification problem where an instance can belong to any of the five possible condition states (see Figure 6.5). The objective of the classifier is to accurately predict the condition state of an unseen instance given the data of selected features (see Table 6.2). We developed five distinct models for this purpose where the logistic regression and decision trees are applied as baseline techniques.

Table 6.4 shows the evaluation results with a stratified random sampling (SRS) approach. In SRS, the dataset is randomly divided into train and test sets where the equal class representation is ensured. Additionally, to tackle the class imbalance

Table 6.4: Results of condition state prediction with SRS on complete and under-sampled test set.

Classifiers	Complete test set			Under-sampled majority class		
	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa
Logistic Regression (LR)	0.5082	0.3724	0.0362	0.3511	0.3126	0.1046
Decision Tree (DT)	0.698	0.7037	0.5488	0.7022	0.7032	0.6034
Random Forest (RF)	0.7196	0.7103	0.5579	0.7323	0.7317	0.6422
Gradient Boosting Trees (GBT)	0.7271	0.718	0.5674	0.7326	0.7327	0.6446
NN with Entity Embeddings (NN-EE)	0.7877	0.7914	0.6811	0.7906	0.7929	0.7224
NN-EE with class weights (NN-EE (cw))	0.7967	0.8081	0.7089	0.791	0.7946	0.7244

problem (discussed in detail in Section 6.3.4), we evaluated our model by under-sampling the majority class up to 40%. The logistic regression obtains inferior kappa value which depict the random guessing by the model. All the tree-based classification models have negligible performance differences for the complete test set. The under-sampling approach has improved the accuracy of tree-based models at least 20% with notable improvements in the kappa score. For instance, the kappa value of gradient boosting trees is improved from 56% on the test set to 64% on under-sampled set. The neural network with entity embedding (NN-EE) performed significantly good among all the models. The performance of NN-EE is further improved by assigning class weights which tackle the data imbalance problem at the algorithmic level. By the addition of class weights to NN-EE, the accuracy and f-score are approximately 80% with the kappa value of 0.70 on the test set. In other words, our NN-EE with class weighting model can predict with 80% accuracy the correct condition state of an element given its damage and their basic properties.

Table 6.5: Results of condition state prediction with SCV on complete and under-sampled test set.

Classifiers	Complete test set			Under-sampled majority class		
	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa
LR	0.5083±0.002	0.3768±0.0024	0.0439±0.0034	0.3553±0.008	0.3243±0.009	0.1113±0.0109
DT	0.7005±0.0044	0.7062±0.0046	0.5499±0.0067	0.7131±0.0068	0.714±0.0067	0.6178±0.009
RF	0.7185±0.0055	0.7143±0.0048	0.5616±0.0078	0.748±0.0038	0.7473±0.0041	0.6632±0.0052
GBT	0.7219±0.0098	0.7112±0.0117	0.5563±0.0182	0.7345±0.0146	0.7343±0.0147	0.6471±0.0194
NN-EE	0.8158±0.0062	0.8069±0.0062	0.7165±0.0074	0.8222±0.0036	0.8238±0.0037	0.7639±0.0047
NN-EE (cw)	0.8128±0.0050	0.8243±0.0048	0.7328±0.0067	0.8253±0.0029	0.8283±0.0029	0.7695±0.0038

The same set of models are further evaluated with a 10-fold stratified cross-validation (SCV) approach introduced in Section 6.4.3. Table 6.5 shows the averaged scores across the 10-folds along with standard deviations on the test set and on under-sampled majority class set. Similar to the SRS, the NN-EE with the class weights has performed best among all the models with 81% accuracy, 82% f-score and 0.73 kappa values on the complete test set. Approximately, all the models show slightly improved

performance scores compared to SRS. This is due to the difference in validation approach, wherein SRS approach the model is tested on completely unseen dataset whereas in SCV approach the model is trained and tested multiple times on randomly chosen subset of IMA dataset.

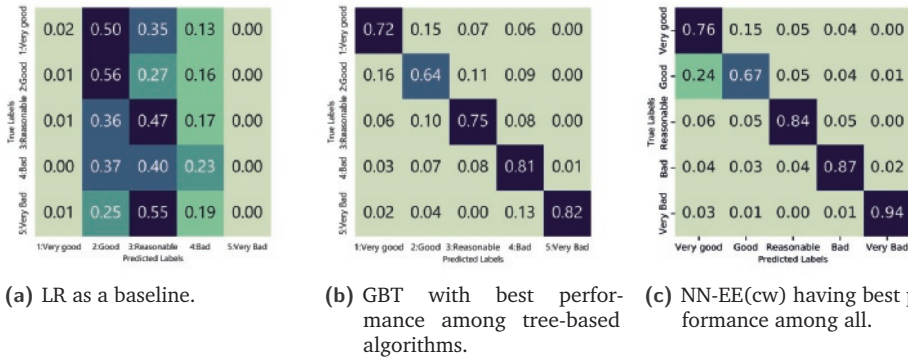


Figure 6.11: Confusion matrices of condition state prediction with SRS on under-sampled test set.

Figure 6.11 presents the confusion matrix analysis of logistic regression, gradient boosting trees and neural networks with entity embedding along with the class weights. Due to space limitation, the confusion matrices of other models are not presented. This analysis provides a summary of correctly and incorrectly classified instances for each class. Figure 6.11a presents the confusion matrix of LR results as a baseline, where the LR model shows the poor classification performance. For instance, an instance having condition state *very good* (first row of Figure 6.11a) is only 2% of times correctly classified as *very good*, whereas it is 50% times classified as *good*, 35% as *reasonable*, and 13% times as *bad*.

The results of GBT presented in Figure 6.11b are significantly better, however, the first three classes are still relatively poorly classified with below 80% accuracy. The result of NN-EE (cw) in Figure 6.11c shows further improvements in classification results in which for last three classes the model can classify an instance with a correct class atleast 85% of times. The confusion matrix also reveals that the model often (24% of times) confuses an instance of class *good* as *very good*. This can be attributed to similar damage details that leads to prediction of these classes.

6.5.2 Risk level prediction

Risk levels prediction models aim to accurately predict the risk level on a bridge (elements) given the damage and basic properties (see Table 6.2). A risk level prediction model can classify the risk to five classes namely *negligible*, *limited*, *increased*, *high*, and *unacceptable* (See Figure 6.6).

Table 6.6: Results of risk level prediction with SRS on the complete and under-sampled test sets.

Classifiers	Complete test set			Under-sampled majority class		
	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa
Logistic Regression (LR)	0.6272	0.4999	0.0236	0.5221	0.4654	0.0884
Decision Tree (DT)	0.7452	0.7433	0.4949	0.6997	0.6991	0.4908
Random Forest (RF)	0.7497	0.7417	0.4839	0.7516	0.7481	0.567
Gradient Boosting Trees (GBT)	0.8076	0.8056	0.6157	0.8037	0.8026	0.664
NN with Entity Embeddings (NN-EE)	0.842	0.8407	0.6888	0.8516	0.8487	0.7446
NN-EE with class weights (NN-EE (cw))	0.8705	0.8738	0.7626	0.8309	0.8317	0.7187

Table 6.6 provides the result of model evaluation with the SRS approach on the complete test set and under-sampling of the majority class. Comparative to condition state prediction, all the prediction models attain relatively better predictive scores. This means it is relatively easy for classifier to relate damage scores to risk level directly compared to the condition states. By applying the under-sampling approach, all the models show improved performance except for the NN-EEs (cw). It can be attributed to relatively balanced dataset resulting from under-sampling which might have eliminated the benefits of cost-sensitivity learning applied to NN-EE. On the other hand, the NN-EEs (cw) shows the best performance among all the models with an accuracy of more than 85% and kappa value of 0.76 on the test set. In other words, a GBT model can predict the risk level with 80% accuracy whereas the NN-EE (cw) model can predict the risk level with 87% accuracy. Additionally, the NN-EE (cw) obtain significantly improved kappa score of 0.76 compared to the NN-EE without class weights with only 0.68 of kappa.

Table 6.7: Results of risk level prediction with SCV on complete and under-sampled test sets.

Classifiers	Complete test set			Under-sampled majority class		
	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa
LR	0.6266±0.0028	0.5019±0.0044	0.0251±0.0076	0.5238±0.0107	0.4656±0.0105	0.0912±0.0195
DT	0.7529±0.0082	0.7512±0.008	0.5119±0.0154	0.7238±0.0117	0.7236±0.012	0.5311±0.0183
RF	0.7464±0.0108	0.74±0.011	0.4828±0.0219	0.7701±0.006	0.7677±0.0058	0.6014±0.0099
GBT	0.8041±0.0105	0.8023±0.0119	0.6089±0.0244	0.8235±0.0088	0.8232±0.0091	0.6974±0.0153
NN-EE	0.8597±0.0067	0.8594±0.0069	0.7279±0.013	0.8593±0.0087	0.8585±0.0084	0.7273±0.0169
NN-EE (cw)	0.8806±0.0066	0.8841±0.0065	0.7835±0.0117	0.8807±0.0058	0.8844±0.0056	0.784 ±0.01

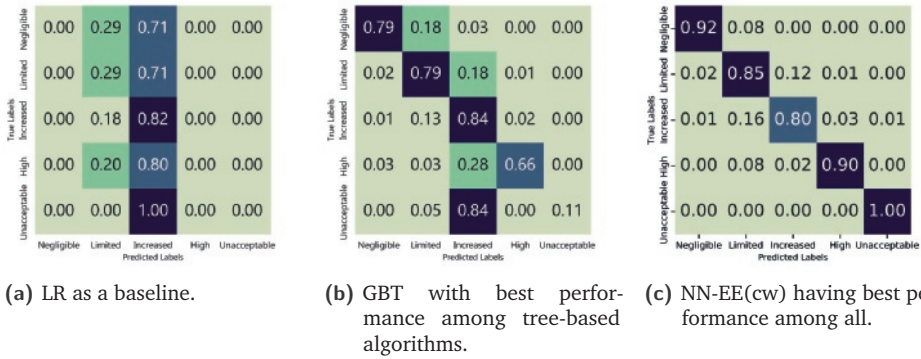


Figure 6.12: Confusion matrices of risk level prediction with SRS on under-sampled test set.

The developed models are also evaluated with SCV approach on a complete and under-sampled set. The average results of SCV across 10-folds and their standard deviation is provided in Table 6.7. Similar to the results of condition state predictions, the NN-EE (cw) performed best among all the models with 0.78 kappa, whereas the GBT has best predictive performance among tree-based models with 0.60 kappa value. The obtained results further validate the robustness of the models when trained and tested on the various folds of IMA dataset.

In addition to the numerical performance measures, we performed confusion matrices analysis to explore the classes that are difficult for the predictive models to classify. The confusion matrices of LR as a baseline, GBT as a best tree-based model and NN-EE (cw) model as best classifiers are provided in Figure 6.12. The LR model poorly classify all the risk level as *limited* and *increased*. This can be attributed to high-class imbalance which introduces a bias in favor of the majority classes. The confusion matrix of GBT shows better classification compared to LR.

However, the risk level *unacceptable* is 84% of times predicted as *increased* level. The prediction of *high* risk level is also misclassified as *increased* risk level for at least 28% of the times. The confusion matrix of NN-EE(cw) shows accurate classification of an instances to its respective risk levels. This is because NN-EE become invariant to class imbalance when aided with class weights. The confusion matrix provided in Figure 6.12 presents the results of SRS on under-sampled dataset given in Table 6.6. The NN-EE (cw) is noted to perform better on the complete test set as compared to the under-sampled set.

6.5.3 Maintenance advice prediction

As a result of damages noted during inspection, the decision-makers suggest the maintenance advice. We trained the models on the historical maintenance advice data and damages details. These classifiers can assign an instance to one of the seven categories which are *No action*, *Technical inspection*, *fixed maintenance plan*, *monitor*, *further investigation*, *maintenance*, and *replace* (see Figure 6.7).

Table 6.8: Results of maintenance advice prediction with SRS on complete and under-sampled test set.

Classifiers	Complete test set			Under-sampled majority class		
	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa
Logistic Regression (LR)	0.5804	0.4505	0.0222	0.4969	0.4389	0.2518
Decision Tree (DT)	0.797	0.8012	0.6707	0.807	0.8074	0.7404
Random Forest (RF)	0.8002	0.7997	0.6664	0.8303	0.8288	0.7704
Gradient Boosting Trees (GBT)	0.8089	0.8102	0.6842	0.8341	0.8335	0.7756
NN with Entity Embeddings (NN-EE)	0.85	0.8507	0.7522	0.871	0.872	0.8266
NN-EE with class weights (NN-EE (cw))	0.8634	0.8695	0.7909	0.883	0.8842	0.8442

Table 6.8 shows the results of various models that are trained and evaluated on complete and under-sampled test set for the accurate prediction of maintenance advice. As also noted in condition state prediction and risk level prediction, the under-sampling of the test set improves the predictive performance. The tree-based models performed significantly better with accuracy close to 80% and kappa value of 0.68 on complete test set. However as also noted with earlier tasks, the NN-EE (cw) model is the best performing model with 88% accuracy and f-score and 0.84 of a kappa on under-sampled set.

Table 6.9: Results of maintenance advice prediction with SCV on complete and under-sampled test set.

Classifiers	Complete test set			Under-sampled majority class		
	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa
LR	0.5903±0.004	0.48±0.0122	0.0716±0.0207	0.4869±0.0162	0.421±0.0167	0.2322±0.0257
DT	0.7912±0.0101	0.793±0.0093	0.6559±0.0153	0.8153±0.0121	0.8161±0.0118	0.7507±0.0161
RF	0.7927±0.0079	0.791±0.0073	0.6537±0.0121	0.8333±0.01	0.8324±0.01	0.7746±0.0131
GBT	0.8048±0.0134	0.8052±0.0116	0.674±0.0207	0.8321±0.0075	0.8311±0.0076	0.7731±0.0093
NN-EE	0.8650±0.0073	0.8663±0.0069	0.7802±0.0119	0.8685±0.0085	0.8684±0.0071	0.7828±0.0115
NN-EE (cw)	0.8690±0.0083	0.8755±0.0075	0.8009±0.0121	0.8701±0.008	0.8765±0.0071	0.8026±0.0114

The maintenance advice classifiers are further evaluated for their robustness with SCV approach on the complete test set and under-sampled set. The evaluation results using SCV are presented in Table 6.9. As noted for all the above cases, the NN-EE

(cw) performed good among all the evaluated models with accuracy and f-score of 87% and kappa close to 0.80 for both complete and under-sampled set.

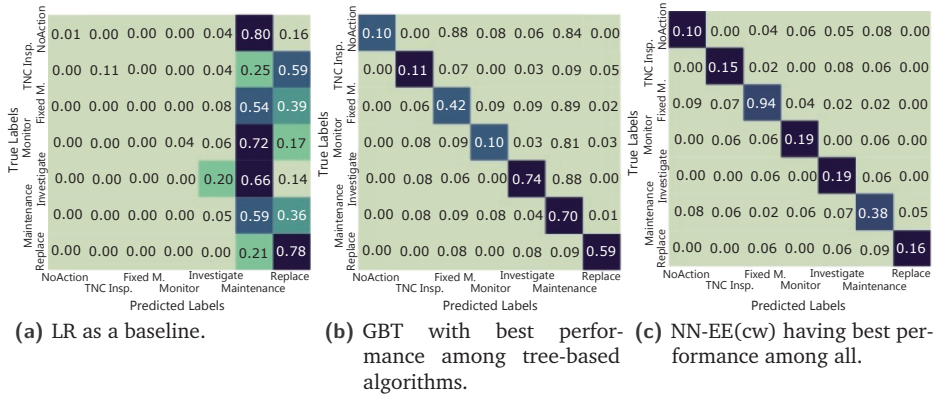


Figure 6.13: Confusion matrices of maintenance advice prediction with SRS on under-sampled test set.

For further insights into the classification capability of the models, we performed confusion matrix analysis. Figure 6.13 shows the confusion matrix of LR as a baseline, GBT as a best tree-based model, and NN-EE (cw) as the best performing model. Even after the under-sampling the majority class (i.e. *Maintenance*), the LR confusion matrix given in Figure 6.13a shows very poor performance. This is due to class imbalance as the under-sampling approach does not completely balance all the classes and a model tends to favor the majority classes over the minority classes. On the other hand, the GBT model with the same dataset shows very good classification expect for *Fixed Maintenance*, *No action*, and *Monitor* classes (see Figure 6.13b). The NN-EE (cw) model has performed best as it is also evident with the confusion matrix given in Figure 6.13c. Compared with the results of LR, the NN-EE (cw) model shows a considerable performance improvement and it can predict the maintenance advice with at least 87% accuracy.

6.5.4 Multi-task learning

The models, discussed thus far, treat each prediction problem independent of each other. In the multi-task learning framework, we developed a unified neural network that learns shared representation as well as problem-specific features to further improve the model performance (see Figure 6.10 for architectural details). This

section reports the results of multi-task learning (MTL-NN) applied for the prediction of condition state, risk level and maintenance advice prediction.

Table 6.10 presents the MTL-NN evaluation results on SRS on a complete test set. By comparing the results of NN-EE (cw) for condition state prediction on complete test set (see Table 6.4), we found that MTL-NN(cw) performed slightly better with the improvement of 0.1 of kappa value. In comparison to NN-EE without class weights, the MTL-NN shown to have improved performance. The same trends are noted for risk level prediction task, when the results of NN-EE and NN-EE(cw) given in Table 6.6 are compared with results of MTL-NN and MTL-NN (cw) presented in Table 6.10. In contrary, the MTL-NN (cw) shows a slight decline in performance accuracy when compared with NN-EE(cw) of Table 6.8.

Table 6.10: Results of Multi-task learning (MTL-NN) with SRS on complete test set.

Classifiers	Condition state			Risk level			Maintenance Advice		
	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa	Accuracy	F-Score	Kappa
MTL-NN	0.8127	0.789	0.7034	0.8587	0.8545	0.7171	0.8611	0.8615	0.7727
MTL-NN(cw)	0.798	0.8092	0.7132	0.8765	0.88	0.7753	0.8572	0.8638	0.7809

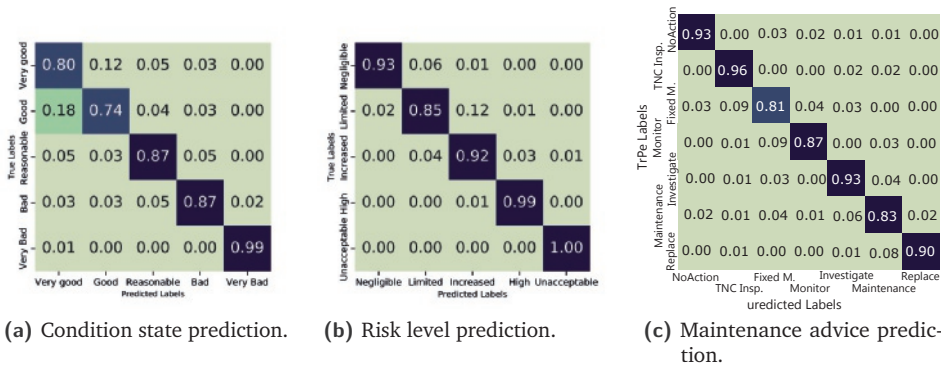


Figure 6.14: Confusion matrices of all prediction tasks with SRS on test set by MTL-NN(cw) model.

In addition to numerical performance measures, the confusion matrices analysis of MTL-NN(cw) is performed for each prediction task. The resulting confusion matrices are shown in Figure 6.14. The confusion matrix of condition prediction task given in Figure 6.14a compared to confusion metrics of NN-EE(cw) in Figure 6.11c shows significantly improved performance results specifically for the *good*, *reasonable* and *very bad* classes. For the risk level task, the MTL-NN(cw) model shows significant improvement in classification two risk classes namely *increased* and *high* risk (see Figure 6.12c for comparison). The confusion matrix of maintenance advice task in

Figure 6.14c presents slightly decline in performance compared to the NN-EE(cw) model of Figure 6.13c. The notable classes with declining in classification accuracy are *investigation*, and *monitor*. On the other hand, the MTL-NN(cw) shows improved classification for *fixed maintenance* and *technical inspection* classes.

In summary, the multi-task learning aims to improve the learning efficiency and prediction accuracy by optimizing multiple objectives from shared features representation. The goal to develop MTL-NN was not only to develop a unified models but also to improve the performance on individual tasks by learning shared representations. The MTL-NN (cw) model shows the improvement in classification results for two of prediction tasks compared to the single-task learning.

6.6 Interpretability of models

For the safety-critical domains such as health, manufacturing, and transportation, the decision-aid systems must be transparent and interpretable. The models developed using ML techniques are known for being black boxes, which provide little to no explanation of their prediction logic. The interpretability and explainable ML models are active research areas (Doshi-Velez & Kim, 2017; Lipton, 2016; Molnar et al., 2018). Miller (2018) defines the interpretability as the degree to which a human can understand the cause of the decision. In other words, a model is said to be easily interpretable if a human decision maker can comprehend and reason about the model's predictions drawn from his domain knowledge.

There are several model-agnostic techniques to interpret the performance of any ML model (see chapter 2 of Molnar et al. (2018) for a detailed overview). However, most of the techniques are mainly suitable for regression tasks. In our study, we provide the single instance level explanation of NNs-EE(cw) model by employing the Local Interpretable Model-Agnostic Explanations (LIME) framework (Ribeiro, Singh, & Guestrin, 2016). By giving a new test instance to the trained NNs-EE(cw), the LIME framework can explain the positively and negatively contributing features weights to classify an instance to respective predictive class. The instance-level explanation enables the domain experts and decision-makers to understand, interpret, and possibly further improve the model's performance. Additionally, the LIME explanation also decides if the model is trustworthy since the model may sometimes pick the spurious correlation.

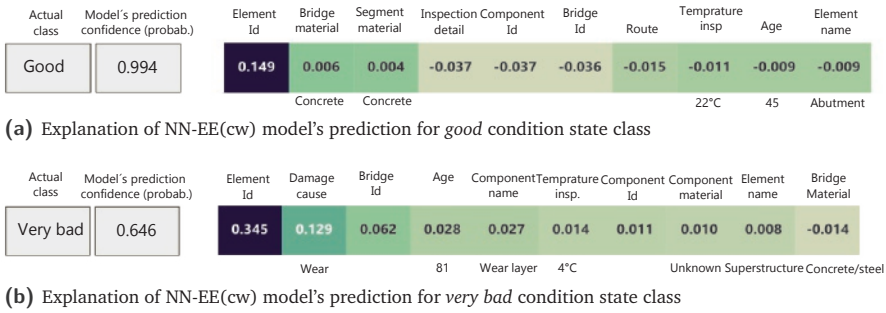
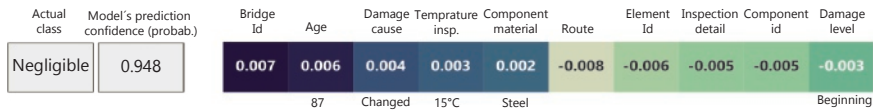


Figure 6.15: Instance-level explanation of NN-EE(cw) model for condition state prediction using the LIME framework

The instance level explanation of NNs-EE(cw) model for condition state, risk level and maintenance advice prediction are presented in Figure 6.15, 6.16, and 6.17 respectively. We provide explanation of models results for two randomly chosen instances from test set for each prediction problem. The LIME explanations shows the actual class of a test instance and the model's prediction confidence in terms of probability. The higher predicted probability shows that the model is confident in its predictions and vice versa. The LIME framework also assign the importance weights to each feature that quantifies the importance of each weight in overall prediction of the model. The positive values (depicted in darker color) represent that the feature positively contributes to the model predictions, whereas the negative weights (depicted in light color) shows that the features do not contribute towards the model's prediction. In addition to the features' weight, the instance level explanation also provide the actual data values that are used for the model prediction. Since the Id's value are difficult to contemplate, we have omitted them for the sake of brevity. It is important to note that the LIME explanation illustrate only five most and least contributing features that facilitate model in prediction, where the NN-EE (cw) was trained on twenty features set (see Table 6.2).

Figure 6.15a shows the explanation of NNs-EE(cw) model trained for the condition state prediction. The explanation of test instances shows that the element id, bridge material and segment material are most positively contributing values whereas inspection details and other ids are negatively influencing the models results. The model classified the instance as *good* with 99% confidence. When given the different test instance, the same model may find different set of features to be important as shown in Figure 6.15b. Therefore, the features that are negatively influencing single instance classification such as age, bridge Id (see Figure 6.15a) can positively contribute for



(a) Explanation of NN-EE(cw) model's prediction for *negligible* risk level class.

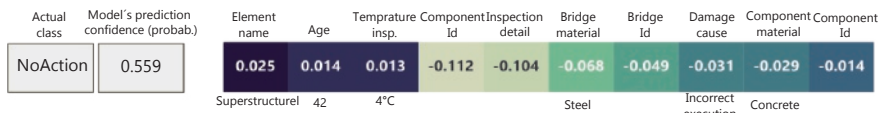


(b) Explanation of NN-EE(cw) model's prediction for *increased* risk level class.

Figure 6.16: Instance-level explanation of NN-EE(cw) model for risk level prediction using the LIME framework.

the classification of other instance (see Figure 6.15b). Thus, it is difficult to name few features as most important compared to the others.

The explanation of risk level prediction model given in Figure 6.16a and 6.16b finds the damage cause and age features quite important whereas the component id, route, and inspection details are negatively influencing the classification. The model correctly classified an instance to *negligible* class with 94% probability and the other instance is classified as *increased* with 99% probability. The classification logic of maintenance advice model is also explained and presented in Figure 6.17. For the *NoAction* class, the model shows only 55% confidence which means that the model was likely to confuse this instance with other classes.



(a) Explanation of NN-EE(cw) model's prediction for *NoAction* maintenance advice class



(b) Explanation of NN-EE(cw) model's prediction for *maintain* maintenance advice class

Figure 6.17: Instance-level explanation of NN-EE(cw) model for the maintenance advice prediction using the LIME framework

Since for each instance, different feature are found to be most and least important, it is difficult to name few features as most important compared to others. For several instances, the bridge, element and component ids are shown to have high importance weights. These ids are basically unique identifiers of bridge, element and components

(see Table 6.1 for bridge decomposition details), yet they are shown to be useful in the prediction of respective tasks. There are possibly two reason for such behavior of the model. First, in practice, the decision makers assess the condition state, risk level and maintenance advice in the IMA process based on their inherent understanding of specific bridge and their components. Second, the model find the data of similar ids from the dataset, establish inherent correlations, and learns their characteristics during training. Additionally, this instance-level analysis may also reveal the set of features that the predictive model find useful which may also diverts from the real decision-making practices.

With the ability to better interpret the results of predictive models, domain experts and decision-makers can better interact with and trust the models' prediction. This gives a reliable decision-aid in the subjective assessment of bridges maintenance planning. Furthermore, the interpretability of the models can also reveal the hidden discrepancies and can be used for features selections and models improvements activities.

6.7 Discussion

This paper has investigated several tree-based algorithms and neural networks with entity embeddings for the development of predictive models for bridge maintenance planning. Multiple learning algorithms such as gradient boosting trees, and neural networks were trained and evaluated on the principal inspection data of bridges. The objective of the predictive models is to support asset owners in subjective assessments tasks of *inspection to maintenance advice (IMA) process* for bridge maintenance decisions. Typically, the decision-maker retrieves the principal inspection data on the case-by-case basis to assess the condition state, risk level and to decide about the maintenance actions. In contrast, the predictive models introduced in this paper can aid decision-makers in these subjective procedures by predicting the condition state, risk level and maintenance advice efficiently with more than 80% accuracy.

For the development of predictive models, we paid particular attention to utilizing only those datasets that are generated from the in-use business process of BMS. The use of available data within the agencies yields practical models that are well-aligned with their business practices and can readily be employed in practice. With intention to deploy the predictive models as a support to subjective assessment procedures of BMS, our work especially takes into account the interpretability of the models' outcomes on single case-basis through employing LIME framework.

During the data analysis and development of predictive models for bridge maintenance decision-making, we noted a few key points and performed additional experiments that are discussed here.

- The overall IMA process (see Figure 6.1) substantially relies on experts' opinions and their technical knowledge to decide the condition state, risk level and maintenance advises. This makes the management of infrastructure vulnerable in case of leaving personnel or having new inspection managers.
- The IMA process records a large amount of data as a result of inspection, damage analysis, risk assessment, and risk control activities, which collectively consist of 73 features. A large number of features may give rise to missing and sometimes inconsistent data. Additionally, the manual feature analysis is time consuming and difficult for decision-makers. We performed careful feature engineering (see Section 6.3.3) on the IMA dataset and chose only 20 features for the development of machine learning models (see Table 6.2).
- The bridge as a structure is decomposed to elements and components (see Table 6.1). Assuming that each element has different characteristics, we initially developed predictive models for each element individually. Contrary to the expectations, the element-specific classifiers performed relatively poorly due to smaller subset of data. Conversely, the single model for all elements combined showed much better performance as presented in Section 6.5.
- The IMA dataset is found to be imbalanced in terms of class representation (see Section 6.3.4). In addition to applying under-sampling and cost-sensitive learning to tackle data imbalance problem, we also applied the Synthetic Minority Over-Sampling Technique (SMOTE) to generate the data instance of minority classes. However, compared to under-sampling and class weights in objective functions, the SMOTE does not improve the predictive performance of the models.
- We applied tree-based classification algorithms and neural networks for the development of predictive models for bridge maintenance decision-making. Though neural networks provide significant improvements for all prediction tasks (see Section 6.5), they are not as interpretable as the tree-based models. For the selection of an appropriate ML algorithm, there is often a trade-off between the accuracy and the interpretability of the model. Given that the maintenance decision-making problem demands both the accurate and explainable

models, we have provided feature importance analysis of the optimal performing tree-based model, as well as the instance level decision details of the neural network models.

To summarize, this paper develops several predictive models to provide support in subjective assessment procedure of bridge maintenance planning. With each new principal inspection activity that reports the damage details, our models can reliably predict the condition states, the risk levels, and the optimal maintenance advise. Since the models use the data from the standard IMA process, the proposed model development approach is generic and can also be applied to related types of structures such as culverts, tunnels, sluices which are often part of bridge management system (Mirzaei et al., 2012). From the implementation perspective of these predictive models, it is worth noting that the real data evolve over time, for instance, new condition classes may be introduced, or new features can be recorded. To accommodate the evolving nature of data, our proposed multi-task learning network can be utilized that can improve the performance on the (related) future tasks instead of training models from scratch.

6.8 Conclusions

This paper shows how the historical and operational data available at transport agencies can be used to support the owners of the assets in the subjective assessment procedure of bridge maintenance planning. We used a large *inspection to maintenance advice (IMA)* dataset of concrete highway bridges from the road agency. Based on the IMA decision procedure, we developed several machine learning models that can predict condition states, the possible risk levels, and recommends the most suitable maintenance actions by learning from the historical data collected through principal inspection within last ten years.

We explored various supervised algorithms from traditional machine learning and from deep learning paradigm to find the optimal model for the prediction tasks. The gradient boosting trees models performed best with 0.56, 0.61 and 0.68 kappa values on the complete test set for condition state, risk level, and maintenance advice prediction tasks respectively. To develop predictive models with further improved predictive ability, we explored the neural network with entity embeddings for learning from structured data. Similarly, the class weights are implemented on NN-EE to tackle the class imbalance problem. The NN-EE with class weights improved the kappa score to 0.70, 0.76 and 0.79 with accuracy close to 80% for condition state, risk

level, and maintenance advice respectively. For all the given algorithms, the discrete predictive models, i.e. task-specific classifiers are developed. Considering that the bridge maintenance planning may have numerous related tasks having shared features (input), we explored the prospect of implementing a multi-task learning framework (MTL). The MTL resulted in a unified model for all the prediction tasks where the kappa and accuracy are further improved for condition state and risk level tasks only. Also, the MTL model learns the shared representations of related tasks which are useful for learning future tasks in low-data regime.

The developed predictive models can assist the asset owners in the process of condition assessment from visual inspection by analyzing the similar cases from the past instantly. Instead of relying only on the subjective assessment, the machine learning models successfully extract the useful insights from the raw inspection data and can predict the most relevant class with a sufficient probability. The developed models can readily be applied as a part of the IMA process to support the decision-makers in maintenance planning tasks. Additionally, the predictions of the models are evaluated by instance level explanation to understand which features are important and why models predict certain classes. The future work of this study aims to employ the monitoring data and images captured during the inspection in order to accurately predict the remaining useful lifetime and to reduce the subjectively of visual inspections.

6.9 References

- Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability*, 11(1), 189.
- Ahmad, R., & Kamaruddin, S. (2012). A review of condition-based maintenance decision-making. *European Journal of Industrial Engineering*, 6(5), 519–541. PMID: 48854. eprint: <https://www.inderscienceonline.com/doi/pdf/10.1504/EJIE.2012.048854>
- Alaswad, S., & Xiang, Y. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157, 54–63.
- Allah Bukhsh, Z., Stipanovic, I., Klanker, G., O'Connor, A., & Doree, A. G. (2018). Network level bridges maintenance planning using multi-attribute utility theory. *Structure and infrastructure engineering*, 1–14.
- Allah Bukhsh, Z., Saeed, A., Stipanovic, I., & Doree, A. G. (2019). Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies*, 101, 35–54.
- Andreou, A., Venkatadri, G., Goga, O., Gummadi, K. P., Loiseau, P., & Mislove, A. (2018). Investigating ad transparency mechanisms in social media: A case study of facebook's explanations. In *The network and distributed system security symposium (ndss)*.
- Bakker, J., & Klatter, L. (2012). Risk based inspection (rbi) at rijkswaterstaat. In *Proceedings of the 6th international conference on bridge maintenance, safety and management*.
- Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. In *Proceedings of icml workshop on unsupervised and transfer learning* (pp. 17–36).
- Bengio, Y. et al. (2009). Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1), 1–127.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798–1828.
- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, 3(Feb), 1137–1155.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb), 281–305.

- Böhm, T. (2017). Remaining useful life prediction for railway switch engines using classification techniques. *International journal of prognostics and health management*, 59.
- Breiman, L. (2017). *Classification and regression trees*. Routledge.
- Bu, G., Lee, J., Guan, H., Loo, Y., & Blumenstein, M. (2014). Prediction of long-term bridge performance: Integrated deterioration approach with case studies. *Journal of Performance of Constructed Facilities*, 29(3), 04014089.
- Bush, S., Henning, T., Ingham, J. M., & Raith, A. (2014). An agent-based framework for improved strategic bridge asset management. In *Austrroads bridge conference, 9th, 2014, sydney, new south wales, australia* (4.4).
- Caruana, R. (1997). Multitask learning. *Machine learning*, 28(1), 41–75.
- Chalouhi, E. K., Gonzalez, I., Gentile, C., & Karoumi, R. (2017). Damage detection in railway bridges using machine learning: Application to a historic structure. *Procedia engineering*, 199, 1931–1936.
- Chase, S. B., Adu-Gyamfi, Y., Aktan, A., Minaie, E., et al. (2016). *Synthesis of national and international methodologies used for bridge health indices*. United States. Federal Highway Administration.
- Chen, K., & Miles, J. C. (2004). *ITS handbook 2004: Recommendations from the world road association (PIARC)*. MA Boston:Artech House 2000.
- Chollet, F. (2015). Keras. Available at: <https://github.com/fchollet/keras>. GitHub.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.
- de Bruin, T., Verbert, K., & Babuška, R. (2017). Railway track circuit fault diagnosis using recurrent neural networks. *IEEE transactions on neural networks and learning systems*, 28(3), 523–533.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*.
- European Commission. (2018). Transport in the european union, current trends and issues. 'Mobility and Transport'. Available on: <https://bit.ly/2BuDRpy>.
- Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, 15(1), 3133–3181.
- Flach, P. (2012). *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189–1232.
- Gattulli, V., & Chiaramonte, L. (2005). Condition assessment by visual inspection for a bridge management system. *Computer-Aided Civil and Infrastructure Engineering*, 20(2), 95–107.

- Ghodoosi, F., Abu-Samra, S., Zeynalian, M., & Zayed, T. (2017). Maintenance cost optimization for bridge structures using system reliability analysis and genetic algorithms. *Journal of Construction Engineering and Management*, 144(2), 04017116.
- Guo, C., & Berkahn, F. (2016). Entity embeddings of categorical variables. *arXiv preprint arXiv:1604.06737*.
- He, H., & Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9), 1263–1284.
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-r., Jaitly, N., . . . Kingsbury, B., et al. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal processing magazine*, 29.
- Hu, X., & Madanat, S. (2014). Determination of optimal mr&r policies for retaining life-cycle connectivity of bridge networks. *Journal of Infrastructure Systems*, 21(2), 04014042.
- Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(4), 15.
- Kiani, J., Camp, C., & Pezeshk, S. (2019). On the application of machine learning techniques to derive seismic fragility curves. *Computers and Structures*.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- Kong, J. S., & Frangopol, D. M. (2003). Life-cycle reliability-based maintenance cost optimization of deteriorating structures with emphasis on bridges. *Journal of Structural Engineering*, 129(6), 818–828.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3–24.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., & Hampapur, A. (2014). Improving rail network velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C: Emerging Technologies*, 45, 17–26.
- Lipton, Z. C. (2016). The mythos of model interpretability. *CoRR*. arXiv: 1606.03490
- Liu, M., & Frangopol, D. M. (2004). Optimal bridge maintenance planning based on probabilistic performance prediction. *Engineering Structures*, 26(7), 991–1002.
- Manco, G., Ritacco, E., Rullo, P., Gallucci, L., Astill, W., Kimber, D., & Antonelli, M. (2017). Fault detection and explanation through big data analysis on sensor streams. *Expert Systems with Applications*, 87, 141–156.

- Markow, M. J., & Hyman, W. A. (2009). *Bridge management systems for transportation agency decision making*. Transportation Research Board.
- Masino, J., Thumm, J., Frey, M., & Gauterin, F. (2017). Learning from the crowd: Road infrastructure monitoring system. *Journal of Traffic and Transportation Engineering (English Edition)*, 4(5), 451–463.
- Miller, T. (2018). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*.
- Mirzaei, Z., Adey, B. T., Klatter, L., & Kong, J. S. (2012). *Overview of existing bridge management systems*. Stresa, Italy. The IABMAS bridge management committee.
- Molnar, C. et al. (2018). Interpretable machine learning: A guide for making black box models explainable. *Christoph Molnar; Leanpub*.
- Morales, F. J., Reyes, A., Caceres, N., Romero, L., & Benitez, F. G. (2018). Automatic prediction of maintenance intervention types in roads using machine learning and historical records. *Transportation Research Record*, 0361198118790624.
- Morcous, G., & Lounis, Z. (2005). Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction*, 14(1), 129–142.
- NEN. (2010). *Nen 2767-4, condition assessment of infrastructure part i methodology*. Delft, The Netherlands Center for Standardization. (in Dutch).
- Nielsen, M. A. (2015). *Neural networks and deep learning*. Determination press USA.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Ramos, J. et al. (2003). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133–142).
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144). ACM.
- Sainath, T. N., Mohamed, A.-r., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for lvcsr. In *2013 ieee international conference on acoustics, speech and signal processing* (pp. 8614–8618). IEEE.
- Sharma, A., Yadava, G., & Deshmukh, S. (2011). A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1), 5–25.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.

- Souza, V. M., Giusti, R., & Batista, A. J. (2018). Asfalt: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning. *Pervasive and Mobile Computing*, 51, 121–137.
- Spencer Jr, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*.
- Trevor, H., Robert, T., & JH, F. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer.
- van der Velde, J., Klatter, L., & Bakker, J. (2013). A holistic approach to asset management in the netherlands. *Structure and Infrastructure Engineering*, 9(4), 340–348.
- Wijnia, Y., & Herder, P. (2010). The state of asset management in the netherlands. In *Engineering asset lifecycle management* (pp. 164–172). Springer.
- Wolpert, D. H., Macready, W. G. et al. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1), 67–82.
- Xristica. (2016). What is the difference between bagging and boosting? The scientific blog of ETS Asset Management Factory.
- Zhang, Y., & Yang, Q. (2017). A survey on multi-task learning. *arXiv preprint arXiv:1707.08114*.

Complementary Works

This chapter briefly outlines the work that laid the foundation of the research outcomes reported in Chapters 2-6. During the last four years, I have been actively involved in two research projects funded by the H2020 programme namely DESTination RAIL and COST ACTION TU1406. The objective of DESTination RAIL ¹ project was to develop a holistic management tool to facilitate railway infrastructure managers in maintenance decision making. The COST ACTION TU 1406 ² is focused on the roadway bridges with the aim to define the quality specifications for bridges at a European level (BridgeSpec). Both of these projects have fostered numerous academic and industrial collaborations namely with Trinity College Dublin, ETH Zurich, University of Zagreb, Irish Rail, Croatian Railways, and Rijkswaterstaat. These collaborations have proven tremendously helpful in terms of understanding the challenges faced by infrastructure managers and developing case studies on a real dataset. This chapter briefly outlines the developments made by us in DESTination RAIL and COST TU1406 project. These research developments are published in workshops, conferences and project deliverables, and are complementary to the journal articles which are part of this thesis.

7.1 Research tasks performed in DESTination RAIL project

The DESTination RAIL project aims to support railway infrastructure management by developing many novel techniques and systems for identifying, analyzing, predicting and remediating the critical assets of rail infrastructure. We were responsible for two core tasks, i.e., development of central data repository called *Information Management System (IMS)* and a holistic system named *Decision Support Tool (DST)*. The idea is to integrate all other outcomes from the Destination Rail project such as monitoring data, numerical models for structural performance, risk model, traffic flow model, and whole life cycle costing model into a DST that will allow for the holistic management of all assets of a rail network. To enable such integration, the DST must be supported

¹<http://destinationrail.eu>

²<https://www.tu1406.eu/>

by a shared data repository that collects the outputs, but also require inputs from the different identification, analysis, and remediation techniques and systems.

7.1.1 Development of Information Management System

One of the main focus in the development of the Information Management System (IMS) is to understand the end-users data needs for the decision analysis. To deal with the challenges of data silos and to comprehend the data needs of users, we choose to exploit the information modeling approach to design the information management system of railway infrastructure. Among other techniques, the information modeling structures the domain of the problem and provides its semantic interpretation. An information model provides a neutral data model, with varying level of detail, which provides support in data exchange and data maintenance. Thus, a detailed information model of railway infrastructure covering monitoring, operation, and maintenance aspects were designed. The extensive details of the development process and the final model is reported in a workshop paper³ and in a deliverable 3.1⁴. Figure 7.1 provides a high-level information model of the main components of railway infrastructure. The *RailInfrastructure* is composed of *Track*, *Platform*, *Signaling*, and *ElectrificationSystem*. Among others, *Track* is the most integral component of the rail infrastructure. The objects related to rail infrastructure are mainly associated with the physical objects (i.e. track, platform, etc) of the railway network. To understand the needs of maintenance, one or more types of *RiskAssessmentMethod* can be associated with each component.

The next step is to develop domain-specific IMS, which can store, organize and retrieve information when needed. We followed waterfall development model constituting, planning, designing, analyzing, developing and testing steps. For implementation, we opted for a system which is flexible enough to accommodate the changing data structure, inter-relationships among entities and addition of data attributes. Initially, MongoDB, a document-based database, was chosen due to its schema-less nature that does not introduce the complex joins to define the entities interrelationships. As the developed systems and techniques of projects are matured over time, the final IMS is being developed in a relational database, i.e. MySQL server – which is free and open source database management system. The final developed IMS consists of data from Irish Railway network covers the data of stations, bridges (over/under), and switches

³Allah Bukhsh, Z., Hartmann, T., & Stipanovic, I. (2016, June). From Analysis of Information Needs towards an Information Model of Railway Infrastructure. In 23rd International Workshop of the European Group for Intelligent Computing in Engineering, EG-ICE 2016.

⁴Allah Bukhsh, Z., Hartmann, T., & Stipanovic, I. (2015), Information Management System Framework, Deliverable 3.1, DESTINATION RAIL, Available at <http://www.destinationrail.eu/ajax/DownloadHandler.php?file=1850>

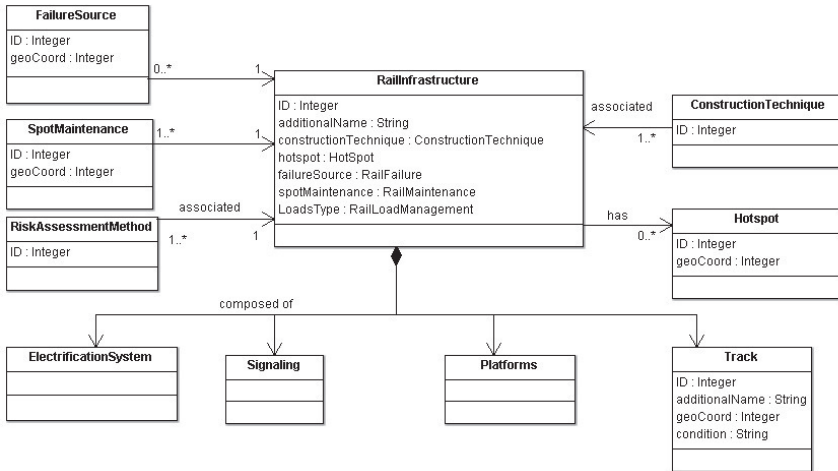


Figure 7.1: An example of an information model

and crossings. The following is the overview of the number of assets along with the number of stored properties in IMS

- Stations – Instances 87, Attributes 5
- Overbridges – Instances 833, Attributes 14
- Underbridges – Instances 1212, Attributes 15
- Switches and Crossings – Instances 951, Attributes 32

For the aforementioned assets, the data stored in IMS consist of equipment number to uniquely identify the asset, position coordinates to locate assets on the map, type, construction material, function, line status, etc. The data of each asset can then be updated with the computed values of risk, reliability levels, and life cycle costing. In contrast to other assets, the tracks data is structured in RailML format and stored in separate JSON format where a single node presents a list of coordinates, their functional location, and unique identification code. Irish Rail has 776 lines of track where 35 lines are categorized as station track, 496 lines as siding tracks, 134 as connecting and 111 as mainline tracks. The developed IMS bring the data silo into a central data repository to support all the other tasks of the project. The use of IMS is shown by the case studies where the decision support tool uses the static data of railway assets from IMS, execute the models, and store the computed results back to IMS for future reference.

7.1.2 Development of Decision Support Tool

Railway infrastructure maintenance is one of the most crucial tasks to keep the railway network functional. The activity of maintenance poses numerous decision-making challenges on the infrastructure managers, due to an extensive network of diverse railway objects (e.g., constituting tracks, bridges, switches and crossing, tunnels, electrification system), availability demands, possession time, deterioration rate, and budget constraints. Such infrastructure maintenance requirements pose decision-making dilemmas to the infrastructure managers, where maintenance planning is challenged by the number of conflicting issues (Lidén, 2015). These maintenance decision-making dilemmas are challenging to handle based only on experts' choices and on techniques that over-rely on expert judgment. Therefore, a comprehensive Decision Support Tool (DST) is needed that considers assets' assessment data, and network as a whole to provide support in a maintenance decision-making process.

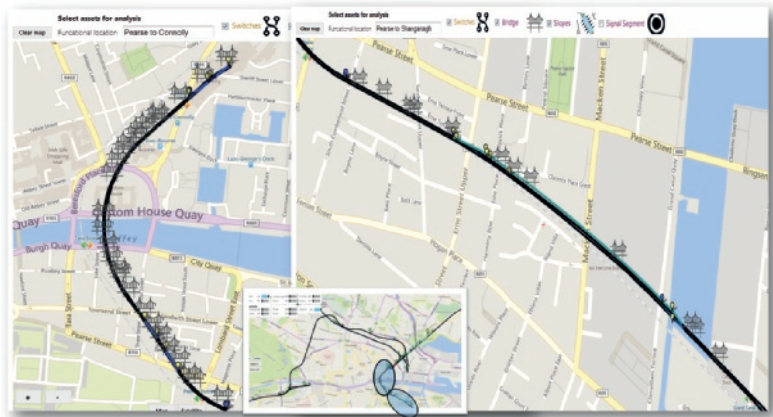
We define a few functional requirements to develop a holistic DST. The final developed version of DST should be able 1) to support basic create, read, update, and delete queries on assets static and dynamic data, 2) should locate all the assets on the network via Geographic Information System (GIS) model, 3) the models of reliability-based assessment, risk assessment, whole Life cycle cost assessment should either be implemented as a part of DST, or DST should be able to send service request(s) to these models, if deployed independently, 4) should support budget planning and maintenance planning, 5) enable the what-if scenarios and finally 6) should be able to report in variety of formats. To support the requirements mentioned above, the system architecture, and possible use cases were developed. Finally, the capabilities of developed DST are illustrated with a case study. The details are DST development and its proof of concept with a case study are explained in D. 3.3⁵ and D. 3.4⁶. The decision-driven development approach and the underlying decision support model is reported in a conference paper⁷.

Figure 7.2 provides a user interface of developed DST illustrating the proof of concept case study. Figure 7.2a shows the GIS interface of the Irish railway network. This interface enables the user to select specific assets for further analysis. Based on

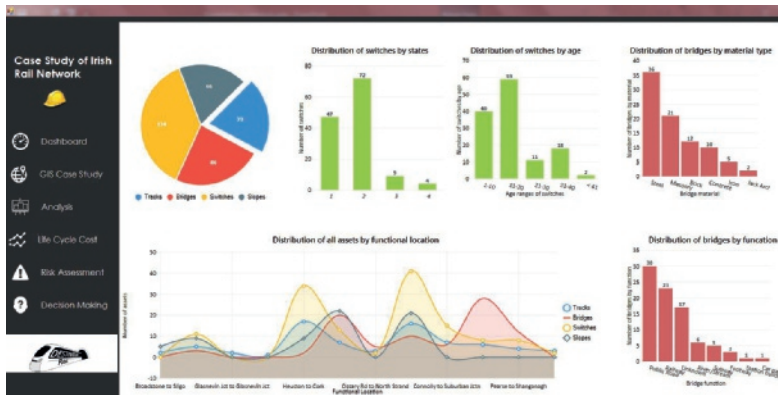
⁵Allah Bukhsh, Z., Stipanovic, I., Connolly, L., Adey, B., Papathanasiou, N., Gavin, K., Martinovic, K., Ramdas, V., Barrett, A., Schoebel, A., Report on Decision Support Tool, Deliverable 3.3, DESTINATION RAIL, Available at, <http://www.destinationrail.eu/ajax/DownloadHandler.php?file=2006>.

⁶Allah Bukhsh, Z., Stipanovic, I., Papathanasiou, N., Barrett, A., Proof of concept of Decision Support Tool, Deliverable 3.4, DESTINATION RAIL, Available at, <http://www.destinationrail.eu/ajax/DownloadHandler.php?file=2139>.

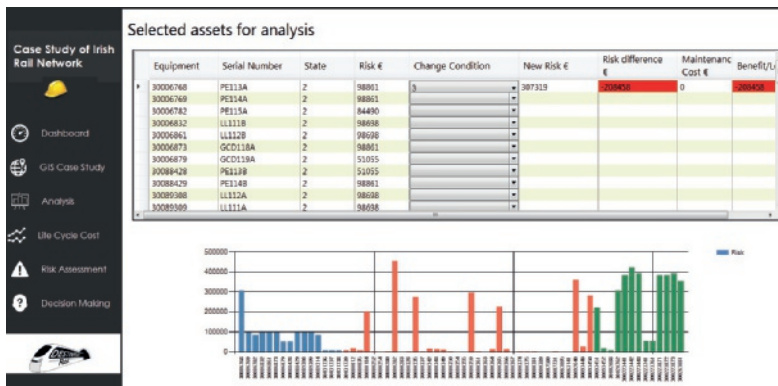
⁷Allah Bukhsh, Z., Stipanovic, I., Gavin, K., Doree, A. G. (2018) . Evolution of Decision Support Systems for Railway Infrastructure Managers. Proceedings of 7th Transport Research Arena TRA 2018



(a) Case study network selected via GIS supported DST interface



(b) DST Dashboard providing overview of selected assets



(c) DST Scenario analysis window showing how the change in condition state will increase/decrease risk, maintenance cost and affect cost/benefit ratio

Figure 7.2: Proof of concept of Decision Support Tool (DST) with a case study

the selection made a user, Figure 7.2b provides a dashboard presenting the basic information of selected assets. For instance, material distribution of the bridge, age of the switches, functional location of the track lines. Figure 7.2c shows the cost-benefit analysis window, where the current risk concerning asset's condition is quantified into monetary value. This window allows the user to modify the condition values of each asset and analyze the impact on risk and maintenance cost values. The purpose is to simulate different failure and maintenance scenario in order to assess the resulting impacts on risks and costs value.

7.2 Research tasks performed in COST ACTION TU1406

The primary objective of the Action is to develop guidelines for the establishment of quality control plans for the roadway bridges. This guideline focus on bridge maintenance and life-cycle performance at two levels: (i) performance indicators, (ii) performance goals. By integrating the most recent knowledge of quantifying and assessing bridge performance, bridge management strategies will be significantly improved, enhancing asset management of aging structures in Europe. The project is structured into five working groups, each having its specific task. I was a member of working group 2, whose goal was to define a set of performance goals based on the identified performance indicators. The research collaboration within working group 2 emphasized on the fact that often a direct link between performance goals and performance indicators cannot be established, especially when multi-objective performance-based decision making is needed and preferred.

Modern decision-making processes dealing with bridge management have to go far beyond choosing an optimal solution based just on single indicators (i.e., lowest long-term cost). Therefore, the models considering multi-performance criteria for optimal bridge management has to be developed. By utilizing the concepts of multi-criteria decision making, the multi-objective optimization models were developed. In the following, a brief description of each model is provided.

7.2.1 Multi-objective optimisation models

In the course of developing network-level bridge programs, program managers typically face a variety of objectives and constraints. Examples of such objectives are: provide a safe, responsive and sustainable network; or minimize adverse impacts on users, local communities, and the environment resulting from maintenance. These ob-

jectives are often constrained by budgetary limitations or a minimum level of average bridge performance. The overlap between objectives and constraints is a crucial issue to practical multi-objective optimization of an asset management program, as they may have conflicting nature.

Multi-Criteria Decision Analysis (MCDA) provides a systematic approach to evaluate multiple conflicting criteria in decision making. Conflicting criteria are typical in evaluating options: cost or price is usually one of the main criteria, and some measure of quality (performance level) is typically another criterion, generally in conflict with the cost. MCDA is used to identify and quantify decision-maker and stakeholder considerations about various (mostly) non-monetary factors, to compare alternative courses of action (Patidar, 2007). Alternatively, the multiple performance criteria can be combined into a so-called utility function, in which all the criteria are reduced into a single scale. Irrespective of a specific method, MCDA involves three main steps (Patidar, 2007):

1. **Weighting:** To assigns relative importance weights to the multiple criteria based on the preferences of a stakeholder.
2. **Scaling:** Because the performance criteria can be of different units, scaling provides a standard scale of measurement for comparison purposes. This may involve plain normalization or development of single-criterion utility functions.
3. **Amalgamation:** Amalgamation is combining either the normalized data or single criterion utility functions using the relative weights into one measure based on mathematical assumptions about the decision maker's preference structure.

For the COST Action project, I applied **Analytical Hierarchy Process (AHP)** on the simulated case study data of roadway bridges. AHP provides a comprehensive and rational framework for structuring a decision problem, for representing and quantifying its elements, for relating those elements to overall goals, and for evaluating alternative solutions. For the sake of demonstration and evaluation of AHP for bridge maintenance planning, the objective of the case study was defined as *prioritizing bridge for maintenance where the cost and downtime due to maintenance are aimed to be kept minimum*. Based on the objective, the selected performance indicators were reliability level, the maintenance cost of an already chosen maintenance treatment, downtime due to the maintenance works, and the importance of the asset on the network. The AHP procedure was applied to the data of bridges with the defined objective and performance indicators. The detail application procedure of AHP applied to case

study data of bridges are available in a workshop paper ⁸. By assigning different importance weights to each performance indicators, the ranking of the bridges was not based only on minimum cost or fewer downtime. Instead, the assigned weights pose trade-offs among the values of performance indicators for the final ranking. Due to rather easy application steps, AHP can be applied to the network-level bridge maintenance planning provided that the objectives and the performance indicators are clearly defined.

Even though AHP is easy to use and is one of the most widely applied MCDA methods, it is unable to capture the risk preference of a decision maker. Since an infrastructure manager is often uncertain of his choices due to incomplete or unavailable data, we opted for a method that can capture data uncertainty and risk attitude of a decision-maker. **Multi-Attribute Utility Theory (MAUT)** is an MCDA method that is capable of capturing the uncertainty of decision processes. MAUT considers the multiple objectives (performance goals) represented with attributes (performance indicators) and consistently captures risk attitude of decision makers as well as uncertainty in preferences (i.e., which value to choose from a finite set of outcomes) when the probability of achieving the results is not definite. The details of MAUT applied for bridge maintenance decision-making problem is explained in Chapter 2 and in paper ⁹. Another study was conducted to evaluate the sensitivity of the multi-attribute utility model with respect to changing weights of performance indicators and risk tolerance parameters ¹⁰. The sensitivity analysis concluded that a change in risk parameter does not influence the ranking of alternatives considerably. Instead, the magnitude of performance indicator values plays a more significant role in alternatives rankings. A one-attribute and two-attribute analysis on weights of performance indicators also reveal that change in weights of the single attribute does not affect the model's results substantially. An interested reader may refer to the online tool, explained below, for the application of multi-attribute utility model and to conduct sensitivity analysis.

⁸Allah Bukhsh, Z., Oslakovic, I. S., Klanker, G., Hoj, N. P., Imam, B., & Xenidis, Y. (2017). Multi-criteria decision making: AHP method applied for network bridge prioritization. In Joint COST TU1402–COST TU1406–IABSE WC1 Workshop: The value of Structural Health Monitoring for the reliable Bridge Management

⁹Allah Bukhsh, Z., Stipanovic, I., Klanker, G., O'Connor, A., & Doree, A. G. (2018). Network level bridges maintenance planning using Multi-Attribute Utility Theory. *Structure and infrastructure engineering*, 1-14.

¹⁰Allah Bukhsh, Z., Stipanovic, I., Palic, S. S., & Klanker, G. (2018). Robustness of the Multi-Attribute Utility Model for Bridge Maintenance Planning. *The Baltic Journal of Road and Bridge Engineering*, 13(4), 404-415.

7.2.2 Sensitivity analysis of multi-attribute utility model

In TU1406 COST Action, a web-based tool is developed to apply the multi-attribute utility model. The developed tool has implemented the concepts of multi-attribute utility theory by employing the R Utility package (Reichert, Schuwirth, & Langhans, 2013). The link to access the online tool is the following: https://maut.shinyapps.io/application_of_maut/. Currently, the tool support comma separated file with a limit of four numeric attributes only. However, there is no limit to the number of alternatives. The tool enables the user to assign weights to each attribute and modify the preference values on run time. By changing the weights and preference values, the users are able to analysis the change in utility function and resulting ranking of alternatives. The developed tool is generic and can be used with any tabular data set.

7.3 Industry Collaborations

Since the beginning of my PhD, I had the opportunity to interact with railway and road agencies during bi-annual project meetings of DESTination RAIL and COST ACTION TU1406. The project meetings and workshops have provided me guidance to understand the challenges of infrastructure maintenance faced by assets owners. This has framed my research into developing practically useful decision-support models to facilitate infrastructure managers in maintenance planning decision-making. In addition to presenting research work in several scientific workshops and conferences, I have the opportunity to do an internship at Irish Rail and Rijkswaterstaat to understand their maintenance planning processes and to avail the real data for the validation of decision models.

7.3.1 Internship at IrishRail

I had the opportunity to visit Irish Rail for three weeks in the second year of my PhD. By that time, I had developed GIS-based (see Figure 7.2a) interface of decision support tool (DST). In the monthly meeting of the board of directors, I presented the GIS-based DST tool, which mapped all the assets of IrishRail on the geographic interface, shows the necessary information of each asset on mouse hover, and present the status of trains running on the network at a particular moment. The initial ideas to develop a decision-support model based on the methods of multi-criteria decision analysis were also discussed.

To understand the challenges of infrastructure maintenance, I had interviews with a technical engineer and an infrastructure manager of the Dublin region. The interviews reveal that the switches and crossing as one of the most important yet most fragile components of the rail network. Due to the frequent failure of switches and crossing, there was a need to develop such decision support that could aid in solving the problem with minimal network disruption. The predictive models for switches and crossing data were developed in response to this, which quickly detects the maintenance need and decide on maintenance interventions type (**Chapter 5**). Though the predictive models are not evaluated at the agency, the idea of using only available data to improve the decision-making process by machine learning algorithms was valued by a technical maintenance engineer.

7.3.2 Internship at Rijkswaterstaat

The collaboration with Rijkswaterstaat (RWS) was made possible by the representative of the Netherlands in COST TU1406 project. Before starting an internship at RWS, I had worked on a small data set of bridges for the maintenance prioritization problem (**Chapter 2 and 3**). I started my six months long internship at RWS to understand their maintenance planning process for concrete bridges in further detail. The Netherlands road network consists of more than 3000 concrete bridges, where the asset owners have to plan maintenance well in advance to request the funding from central government. Though RWS has a risk-based procedure to assess and plan the maintenance of bridges, the need to plan five years motivated for a solution that can support in maintenance planning of bridges over a definite time horizon. In response, a multi-year maintenance planning framework was developed and tested on data of 869 highway bridges of Netherlands road network (**Chapter 4**).

Based on visual inspection results, the risk on the structure is assessed, and advice on maintenance action is determined. Inspired by outcomes achieved by predictive models of switches and crossing, the RWS was interested in developing predictive models that facilitate them in the assessment of condition state, risk level(s) and appropriate maintenance advice by analyzing historical data. We developed several distinct machine learning models to predict the condition states, risk levels and maintenance advice based inspection data. A single unified model using the multi-task learning framework is also developed to learn the shared representation from data and to improve the predictive capability of the tasks. **Chapter 6** reports the details of predictive maintenance modeling using machine learning and deep learning

algorithms for bridge maintenance planning. Due to time constraints, the developed models are not yet evaluated in practice.

7.4 Summary

This chapter presents the complementary work performed during the PhD trajectory. This work provided a solid foundation and motivated the research efforts presented in Chapter 2-6. A brief summary of work performed in the context of DESTination RAIL project and COST TU1406 project is the following:

- An information management system (IMS) is developed as a central database repository to support the data needs of developments made in DESTination RAIL project. The IMS contains detail data of bridges, slopes, track lines, switches and crossing and stations.
- A decision support tool (DST) based on geographic information system (GIS) is developed. A cost-benefit analysis is implemented as a part of DST, where the current risk concerning asset's condition is quantified into monetary value. The purpose of cost-benefit analyses is to simulate different failure and maintenance scenario to assess the resulting impacts on risks and costs value.
- In COST TU1406 project, the multi-objective models based on analytical hierarchy process and multi-attribute utility theory are developed to support in the maintenance planning process. These models take into account the social and environmental impact of maintenance during decision-making. A web-based tool is designed to enable the user to interact with the decision model and analyze the robustness of results by varying weights values.
- The industrial collaboration with many project partners provided the first-hand experience and knowledge of their business process of maintenance planning. The interactions with industrial partners have proven to be very useful in terms of acquiring knowledge of their maintenance planning process as well as to evaluate the developed models on real datasets.

This chapter provides an overview of complementary research activities undertaken during the PhD program. The following chapter concludes this thesis by giving the summary of each chapter and outlining the theoretical and practical contribution of this research.

References

- Lidén, T. (2015). Railway infrastructure maintenance-a survey of planning problems and conducted research. *Transportation Research Procedia*, *10*, 574–583.
- Patidar, V. (2007). *Multi-objective optimization for bridge management systems*. Transportation Research Board.
- Reichert, P., Schuwirth, N., & Langhans, S. (2013). Constructing, evaluating and visualizing value and utility functions for decision support. *Environmental Modelling and Software*, *46*, 283–291.

Conclusions

This chapter provides conclusions based on the work presented in the previous chapters. First, the main findings and a brief overview of research activities and outcomes are described. Next, an extended conclusion per chapter is given followed by addressing the theoretical and practical contributions. Subsequently, the recommendations for the practitioners, the critical reflection of this research and future research direction are listed. In brief, the key conclusions of this thesis are as follows:

- This research has made progress towards more consistent, explicit, and data-driven maintenance planning approaches, which makes the decision processes easy to implement, transparent, and reproducible.
- A developed multi-attribute method has shown to be most pertinent for establishing a prioritization of assets of the whole network based on multiple objectives.
- The comprehensive multi-year framework provides a transparent and followable procedure to the asset managers to develop optimal maintenance plans and simulate different planning scenarios under varying performance values and budget constraints.
- The machine learning models make effective use of (existing) asset management data and aid-in experience-driven decision processes by predicting maintenance need and type promptly.
- The decision logic of predictive models reveals discrepancies to the decisions taken in practice. For instance, age may be found as the most important factor to decide maintenance action in contrast to noted damage details.
- The validation of proposed models on a real dataset of highway bridges and railway switches proved their usefulness and applicability to aid the decision-making process of maintenance planning within agencies.

8.1 Overview of research activities and outcomes

This section presents the research activities and the resulting outcomes in the context of defined research questions and journal articles (see Table 8.1).

Table 8.1: Overview of research activities and outcomes

Research activities	Research outcomes
<i>Chapter: 2, RQ: 1</i>	IALCEE 2018 (Published)
<ul style="list-style-type: none"> • Determining the relevance of Multi-Criteria Decision Analysis (MCDA) methods for a maintenance decision-making problem • Analyzing the MCDA methods using an evaluation scale • Applying the three most used methods on multi-attribute data from a highway owner 	<ul style="list-style-type: none"> • Identified the characteristics of the maintenance decision-making problem • Applied analytical hierarchy process, multi-attribute utility theory and ELECTRE III on the selected number of highway bridges • Discern the multi-attribute utility theory as a most relevant MCDA method
<i>Chapter: 3 with Appendix A, RQ: 2</i>	SIE (Published) & BJRBE (Published)
<ul style="list-style-type: none"> • Determining the multiple objectives of maintenance planning • Performing maintenance planning of all the bridges of the highway network • Developing a multi-attribute method for a maintenance decision-making problem • Evaluating the method on the highway bridges to rank them on multiple objectives • Analysing the applicability and robustness of the model 	<ul style="list-style-type: none"> • Identified performance indicators to quantify the objectives • Introduced quantification procedure for different performance aspects (including socio-economic) using performance indicators • Developed a multi-attribute utility model, which accommodates data uncertainty and risk attitude of decision-makers • Validated the applicability of the model for transport infrastructure assets
<i>Chapter: 4, RQ: 3</i>	Article under review
<ul style="list-style-type: none"> • Introducing a comprehensive multi-objective optimization for the maintenance planning based on current condition state and the future predicted condition states. • Employing the previously developed multi-attribute utility model to rank bridges • Applying Markov chain process in conjunction with genetic algorithms for the performance forecasting and optimizing time for maintenance 	<ul style="list-style-type: none"> • A comprehensive multi-year maintenance planning framework • Developed transition matrices to predict the performance states in the future • Programmed a genetic algorithm with constraints of improving network performance under a limited budget • Demonstrated the framework by developing an optimal maintenance plan for a large number of bridges of the network

<p><i>Chapter: 5, RQ: 4</i></p> <ul style="list-style-type: none"> Analyzing the viability to use machine learning techniques for the effective use of available asset management data Structuring the historical data to retrieve useful insights Training the classifiers on the historical data to predict future (unseen) events Supporting the subjective analysis by predictive modeling of the data 	<p>TRC (Published)</p> <ul style="list-style-type: none"> Exploratory analysis of the data generated by the maintenance request process of railway agency Developed tree-based classifiers that predict the maintenance need and activity types of railway switches with 80% accuracy Performed feature importance analysis and instance level analysis to interpret the results of classifiers
<p><i>Chapter: 6, RQ: 4</i></p> <ul style="list-style-type: none"> Validating the claims of using available data for predictive modeling on the case study of bridges inspection data Applying the deep neural network to improve the prediction accuracy of the models Utilizing the multi-task learning framework to improve the performance of discrete models Explaining the results of the models by interpretability analysis 	<p>Article under review</p> <ul style="list-style-type: none"> Analysed the principal inspection data of road bridges from the last ten years Developed machine learning models that predict condition states, risk level and recommend maintenance advice based only on the damage details noted during the inspection Developed deep neural network models with entity embedding to improve the predictive performance of the model Developed a unified model that learns shared representations of related task Explained the models' prediction
<p><i>Chapter: 7,</i></p> <ul style="list-style-type: none"> Introducing decision-driven development approach for infrastructure maintenance Developing information models for structuring the data of infrastructure objects Evaluating the impact on maintenance cost and risk with different maintenance scenario by cost-benefit analysis method Developing an online tool to analyze the robustness of multi-attribute utility model 	<p>Several conference papers and deliverables</p> <ul style="list-style-type: none"> A GIS-based interface with the details of bridges, slopes, track lines and switches data of Irish Railways A decision-driven view on the collection of relevant data and development of concrete information models An interactive interface to execute different maintenance scenarios and to explore their impacts on cost and risk on an asset A web-based online tool to evaluate the robustness of multi-attribute model by performing sensitivity analysis

8.2 Summary of appended chapters

The implicit decision-making by assets managers, single-objective optimization methods, the distributed asset management systems, and data accessibility challenges have put forward the need for pragmatic decision-support methodologies for maintenance planning. This research has developed several practical data-driven methodologies and machine learning models to aid asset managers in the decision-making process of maintenance planning. In the following, the summary of each chapter is provided along with the key findings and conclusions.

Evaluating the multi-criteria decision making methods for Infrastructure management

Infrastructure management renders many decision-making problems from assets, condition inspections to maintenance planning and resources optimization. The methods of multi-criteria decision analysis (MCDA) have a wide area of applications with substantial variation in methodology. Therefore, the selection of an appropriate MCDA method pertaining to the specific needs of infrastructure management and decision maker is a difficult task. Prior to developing specific decision-support models for assets network, we investigated how different methods of MCDA may provide support in the ranking of a selected number of assets based on multiple performance aspects. An evaluation scale was derived from literature (Cinelli, Coles, & Kirwan, 2014; De Montis, De Toro, Droste-Franke, & Omann, 2004) to determine the difference in results as well as to examine the applicability of MCDA methods on maintenance decision-making problems.

This study has helped us to answer the first research question, which is: *Which methods of multi-criteria decision-analysis provide support in the decision-making process of maintenance planning of transport infrastructure networks?*. In following, the key findings regarding the applicability of MCDA for maintenance decision-making tasks are discussed:

- We evaluated two synthesis-based methods, namely, analytical hierarchy process (AHP), multi-attribute utility theory (MAUT) and one outranking method, namely, ELECTRE III (elimination and choice expressing reality). All three methods of MCDA are applied on a case study which included randomly selected bridges from a highway network. The objective was to analyze MCDA methods for their ability to prioritize assets of the network in an order which would

maximize their reliability and safety, while having minimum social, economic and environmental impacts.

- It is found that, if the weighting structure of the attributes and the stakeholders' preferences are kept similar, the different methods of MCDA even having different application procedure provide comparable results. Each method of MCDA has significantly unique methodology having particular strengths and weaknesses. By applying different MCDA methods on the case study of the highway bridges for prioritization, a recommendation on the use of three evaluated MCDA methods are outlined.
- The evaluation of three MCDA method also concludes that: AHP method can be used for those maintenance decision-making problems where alternatives are not large in number, and a definite ranking of alternatives is required. MAUT is a method of preference when there is a large number of alternatives, and a stakeholder is uncertain of his preference choices. MAUT is one of the few methods of MCDA that incorporates the concepts of uncertainty and (exponential) utility theory. ELECTRE III can be a method of choice where a decision maker has a good understanding of his preferences over the data values. However, ELECTRE III requires the definition of multiple threshold values, which makes its application a demanding task.

The proposed study concluded with the set of recommendations on the use of MCDA methods concerning the characteristics of a specific maintenance decision-making problem at hand.

Network level (bridges) maintenance planning using Multi-Attribute Utility Theory

Infrastructure managers have to maintain the availability and serviceability of aging infrastructure within an ever-shrinking budget. When developing maintenance plans, a decision-maker must take into account a number of related performance goals (e.g., safety and reliability of structures, environmental impact, etc.). However, incorporating multiple performance aspects for maintenance planning of assets on the whole network introduces the complexity of goals quantification, conflicting objectives, data management of several assets, data uncertainty, and unclear preferences of stakeholders. Therefore, there is a need for a method that can optimize multiple objectives simultaneously along with integrating decision-makers' preferences transparently to establish a prioritization of several assets accordingly.

The key findings to answer the second research question, that is *How to prioritize the maintenance of network-wide assets by incorporating the decision-makers' preferences and satisfying multiple conflicting objectives simultaneously?*, are presented as follows:

- It is essential to identify a set of performance goals (objectives) and quantify them by using key performance indicators. As an active participant of COST ACTION TU 1406 project, we contributed in designing an extensive survey to collect bridge performance indicators from various European road agencies. It resulted in a guideline document that linked collection and quantification of performance indicators, performance goals, standards, and practices with decision-making processes (Stipanovic, Høj, & Klanker, 2016; Stipanovic et al., 2017; Strauss, Ivankovic, Matos, & Casas, 2016).
- The multi-attribute utility theory (MAUT) method has been verified on the case study data of twenty-two highway bridges. The final prioritization obtained from MAUT model suggests that an object with higher rank contributes the most in the realization of prescribed performance goals. The MAUT model systematically transforms the different performance indicators into the utility function, where a decision-maker records his preferences in the form of risk tolerance and the importance of attributes' weights. In addition to providing decision support, the use of MAUT model for maintenance planning makes the decision-making process transparent and traceable.
- The robustness of the MAUT model is further validated by conducting sensitivity analysis, where the risk tolerance parameter and weights assigned to the attributes were dynamically changed. It is found that there are minor differences in ranking of assets with changing risk tolerance parameter. Instead, the actual magnitude of the values and weights (assigned by decision maker) play role in the final ranking of the alternatives.
- To encourage the use of the MAUT model for maintenance planning, an online and desktop tool has been developed and presented in a workshop of academics and practitioners from industry (Stipanovic et al., 2017). The proposed solution for maintenance planning by multi-objective optimization is found to be relevant, and useful. The practitioners found the support tool of the MAUT model particularly useful, as it offered them the opportunity to analyze and visualize the impact on alternatives by changing risk and weights parameters. From the researchers' perspective, the application-oriented approach of MAUT model

illustrates and guides the implementation of expected utility theory for the engineering related decision tasks.

To conclude, the proposed solution approach has used available data from the road agency only and has shown how the decision-making process could be improved by implementing other aspects into the evaluation process.

Multi-year maintenance planning framework using multi-attribute utility theory and genetic algorithms

The proposed MAUT method, introduced in the previous section, does not account for budget constraints explicitly. This limitation led us to explore the solutions that should propose an optimal time for maintenance, given the budget constraints and preferences of decision-makers. Typically, maintenance planning solutions on the 'best time to maintain' often utilize search-based optimization algorithms (Bocchini & Frangopol, 2011; Denysiuk, Fernandes, Matos, Neves, & Berardinelli, 2016; Xie, Wu, & Wang, 2018) that consider the multiple objectives (e.g., service level vs. cost spending) but neglect to scrutinize the subjective preferences of the decision-makers. The MAUT method is extended into a comprehensive framework that recommends the best time to maintain an asset under budget constraints and experts' choices.

The keys steps and findings to answer the third research question, that is *How to develop multi-year maintenance plans of network-wide assets constrained by budget limitations and multiple performance requirements?*, are outlined below:

- A Multi-year Maintenance Planning Framework (MMPF) is developed to manage the multifaceted nature of maintenance planning. The framework comprised of four different modules having the following characteristics, (1) an impact assessment module based on heuristics to decide on the type of maintenance intervention and its effect 2) a MAUT method module for ranking of assets based on preference uncertainty and risk attitude, 3) a percentage performance prediction method to forecast the performance state of an asset in the future, and 4) a genetic algorithm module to find the most optimal combination of performance goals under reliability and budget constraints.
- The proposed MMPF is validated with a case study of 869 concrete bridges of a road network. A five-year maintenance plan is programmed to achieve a reliable performance state on the network level within a given budget limit. Based on the condition states, we found that 123 bridges are below the required performance level and need some maintenance. In addition to improved performance level,

the maintenance planning involves other objectives mainly the minimal impact on users (quantified as user delay cost) and reduced maintenance cost by performing maintenance at the optimal times.

- For the case study network, the MMPF developed multi-year maintenance plans of 28 bridges that are within the given budget limit and fulfill the performance requirements. In comparison, sequential maintenance plan (i.e., a bridge with the poor condition is always maintained first) allocates the maintenance of 18 bridges given similar optimization constraints. The proposed framework is general and can be applied to other discrete infrastructure assets for multi-year planning.

The proposed solution approach is particularly useful for those road agencies that mainly rely only on condition states of the assets for the maintenance decision-making.

Predictive maintenance using tree-based classification techniques: A case of railway switches

The data at transportation agencies is huge and dispersed across several IT systems. This makes the data difficult to acquire, manage, and process by traditional software and tools within a tolerable time (Galar, Kans, & Schmidt, 2016). Hence, more than 70% of collected asset management data is never used for any decision-making (Gualtieri, 2016; Wijnia & Herder, 2010). This prompts experts to make decisions based on their implicit reasoning, where there is often not sufficient time and resources to access data and investigate past decisions of similar nature. Several studies propose to employ additional data collection procedure (Li et al., 2014; Manco et al., 2017; Masino, Thumm, Frey, & Gauterin, 2017; Souza, Giusti, & Batista, 2018) to develop self-diagnostics systems. However, the new data collection means are not only expensive but also impractical for the geographically scattered assets. This work explored the viability of using available asset management data to develop predictive maintenance model by utilizing advanced machine learning techniques.

The details of predictive maintenance modeling to answer the four research question, that is *How to effectively use the large amount of asset management data and improve the maintenance decision-making practices using machine learning techniques?* are discussed below:

- A maintenance request process of railway switches was used to investigate the viability of using available data to develop predictive models. Tree-based

classification algorithms, namely decision-tree, random forest, and gradient boosted trees, are compared and evaluated for their ability to learn from past data and predict future events. The gradient boosted trees model with more than 80% can predict the maintenance need, the treatment type, and status of maintenance request for the unplanned maintenance trigger of railway switch. It is noted that the predictive performance of the models can be further improved by generating more data for minority represented classes.

- A thorough explanation of the outputs of models is provided to avoid a black-box data-driven model. The feature importance analysis informs decision-maker about which input features positively or negatively impact the models' output. In other words, which feature plays a significant role in decision outcome. In addition to global explanations, the explanation for a single row prediction is highlighted by local interpretable model-agnostic explanations framework. This interpretability of the models capture the decision-making process of practice and identifies the most crucial data attributes which may help in future data collection procedures.
- By leveraging the potential of machine learning techniques, the developed predictive models suggest that it is possible to use the data currently available at agencies to improve the decision-making process of maintenance planning. Since the data is generated from the in-use business process, the predictive models can readily be used as an add-on tool to the existing ERP solution. Learned from the historical data, the model analyses each maintenance trigger as an individual case and provides its prediction. The proposed approach to developing predictive models along with results explanation have broad applicability and are generalizable to other asset types and maintenance planning scenarios.

The developed predictive models provide the detailed explanation of results and make the maintenance decision-process concrete, transparent, and efficient.

Predictive maintenance planning of bridges using deep neural networks

The successful development of predictive models for railway switches and crossings motivated us to extend the application of machine learning towards more sophisticated deep learning algorithms for the case of road bridges. Bridges have a longer life span where the maintenance plans are drafted much before the maintenance is performed. Many infrastructure managers primary rely on information obtained through visual inspection to decide on the follow-up maintenance actions (Bu, Lee,

Guan, Loo, & Blumenstein, 2014). In practice as well as in theory, little attention is given to investigate the solutions that could facilitate the the decision-making process for maintenance planning using historical data. Motivated by the use of data generated from in-use processes, we developed predictive models that can predict condition states, risk levels and recommend maintenance action for bridge maintenance planning.

The key steps and findings regarding the development of deep neural networks for bridges answers the four research question, that is *How to effectively use the large amount of asset management data and improve the maintenance decision-making practices using machine learning techniques?*, as:

- The road agency uses bridge management system to store inventory, condition states, risk profiles and maintenance plans of approximately 3800 road bridges. We used the data generated from Inspection to Maintenance Advice (IMA) process of BMS during the year 2006 to 2017. The IMA process primarily relies on subjective analysis, where the desk study, interviews with inspection managers, and risk analysis are performed before and after the inspection activity to decide on the *condition state*, *risk level*, and further *maintenance advice*.
- For the development of an optimal predictive model, we applied several machine learning and deep learning neural network models. Initially, the discrete models are developed for each prediction tasks, and their results are evaluated with completed and under-sampled dataset due to imbalance data nature. The neural network with entity embedding (class weights) performed best with the kappa score close to 0.80 for all the prediction tasks. We also implemented a multi-task learning framework to improve the accuracy of discrete models and to learn the shared representations from related tasks.
- Though useful for decision-support, the numeric performance measure does not explain decision logic that leads to specific prediction. For the experts to use these models and trust their predictions, it is essential to explain how models make the prediction. We provided instance level explanation that highlight the contribution score of each feature. The interpretability of the models also elaborates on how the predictive model predicts a particular class (e.g., good condition state) instead of another related class (e.g., poor condition state).

By being trained on the data of in-use IMA process, the developed models can be readily used by infrastructure managers in the decision-making process of bridge maintenance planning. In contrast to traditional machine learning models, the introduced deep models can be used for the transfer learning to gain performance improvements on tasks in a low-data regime.

8.3 Theoretical contributions

The theoretical contributions of the research are deduced according to new findings from the literature, development of novel theoretical concepts, application of existing theoretical methods; development of new models by using existing theories; implementation of existing approaches to a new dataset; and the introduction of existing methods to a new discipline.

This thesis proposes several pragmatic decision-support methodologies and predictive models to improve the maintenance planning process and to predict the maintenance needs in advance. The overarching objectives of these models are to facilitate the infrastructure managers in the decision-making process of maintenance planning. The suggested decision-support models, the comprehensive framework, and the predictive models contribute to the field of infrastructure management and maintenance. The contribution of each method are discussed as follows:

- Due to the broad area of applications and extensive variation in multi-criteria decision analysis (MCDA) methodologies, the selection of a specific MCDA method for the maintenance decision-making task remains unclear, as also noted by several literature reviews (de Almeida et al., 2015; Kabir, Sadiq, & Tesfamariam, 2014; Patidar, 2007). Unlike the studies that compare MCDA at the methodological level only (Velasquez & Hester, 2013; Zavadskas, Turskis, & Kildienė, 2014), *chapter 2* compares three commonly applied methods of MCDA, at the application level, to evaluate their applicability for the decision-making problem of infrastructure management. By highlighting the characteristics and needs of maintenance problem in the form of evaluation scale, we developed a set of recommendations on the use of MCDA methods. The recommendations can be used by the researchers and the practitioners to select an MCDA method that is most suitable for the specific maintenance decision-making problem.
- *Chapter 3* introduces a multi-attribute utility method for the maintenance planning of all the assets of the network based on multiple performance objectives.

Earlier, several studies have focused on the object-level maintenance planning by whole-life cycle costing approach (Bush, Henning, Ingham, & Raith, 2014; Frangopol & Liu, 2007a). Other studies refer to multi-attribute utility theory for optimization modeling (Arif, Bayraktar, & Chowdhury, 2015; de Almeida et al., 2015; Dong, Frangopol, & Sabatino, 2015; Frangopol & Liu, 2007b; Garmabaki, Ahmadi, & Ahmadi, 2016), but does not fully elaborate its application for network-level maintenance planning and on capturing preferences of experts for decision-making. The *chapter 3* extends the applied work of multi-criteria decision methods by quantifying the performance indicators, and by transforming the preferences uncertainty of the decision-makers into objective values for the network-level maintenance planning. The proposed model provides a followable and transparent procedure which addresses the societal and environmental impact of maintenance planning along with economic and reliability aspects.

- *Chapter 4* proposes a comprehensive multi-year planning framework that develops optimal maintenance plans for all the assets of the network under budget and performance constraints. Previous studies for the maintenance planning are concerned with deterioration modeling (e.g., see review by (Alaswad & Xiang, 2017)), and optimal maintenance planning by integer and search based heuristics (Chootinan, Chen, Horrocks, & Bolling, 2006; Denysiuk et al., 2016; Ghodoosi, Abu-Samra, Zeynalian, & Zayed, 2017; Morcoux & Lounis, 2005; Xie et al., 2018). These studies provide several promising solutions, however with the best of our knowledge, none of the literature studies scrutinize the subjective preferences of the agencies/decision-makers in the multi-year maintenance planning procedure. The proposed multi-year planning framework further extends the literature related to optimal maintenance planning by introducing an end-to-end optimization solution. The framework builds upon the multi-attribute utility method, presented in chapter 3, and implements Markov decision-process and genetic algorithms to simulate the several maintenance planning scenarios. The suggested framework is comprehensive, unique, and theoretically significant as these methods had not been scientifically applied and illustrated together for the optimal maintenance planning of an extensive case study network.
- Thereafter, the thesis continues on the subject of maintenance prediction using the asset management data from the agencies (*Chapter 5 and Chapter 6*). Compared to preventive maintenance planning, the maintenance prediction is a relatively new topic for which several studies propose to employ additional data means to develop self-diagnostic system (Kiani, Camp, & Pezeshk, 2019; Li et al.,

2014; Manco et al., 2017; Masino et al., 2017; Morales, Reyes, Caceres, Romero, & Benitez, 2018; Souza et al., 2018). However, additional data collection procedures are often not practical, and they demand substantial investments (de Bruin, Verbert, & Babuška, 2017). Additionally, most of the studies present machine learning models as black boxes and does not elaborate on how and why a model makes specific predictions. *Chapter 5* presents the methodology to develop predictive models using tree-based (machine learning) classification techniques. The classifiers accurately identify correct maintenance notifications and treatment type through modeling the historical data of unplanned maintenance triggers of railway switches. The results of predictive modeling show that, though data is difficult to access due to several independent IT systems within agencies, the available historical data is of valuable importance and can be used for the data-driven decision-making by employing machine learning techniques.

- *Chapter 6* advances the application of machine learning algorithms and further validates the claims of tree-based predictive models by developing a deep neural network with entity embeddings for the road bridges. These models have shown to learn complex non-linear features and extract insights from the damages data to predict condition states, risk levels, and maintenance advice with more than 80% accuracy. Another theoretical contribution is the implementation of a multi-task neural network which jointly learns to solve multiple tasks through utilizing shared embedding and task-specific layers. The introduced deep models can be used for the transfer learning to gain performance improvements on tasks in a low-data regime. Unlike many studies that introduce predictive models as black boxes, our work has paid particular attention to explain the output of models by highlighting the importance of each data attribute and by explaining the instance-level predictions. The research efforts presented in Chapter 6 is unique in comparing and applying tree-based models, neural networks with entity embedding, and multi-task learning framework to find the best performing predictive model for the bridge maintenance planning. The use of advanced machine learning and deep learning algorithms for the development of predictive models is novel and contributes to the industry 4.0 research trajectory (Niestadt, Debyser, Scordamaglia, & Pape, 2019), which aims to develop cyber-physical systems for smart asset management.

Pertaining to multiple challenges faced by infrastructure and an era of technological advancements, the conducted research work is well timed and provides a solid foundation towards technological advancements in infrastructure management.

8.4 Practical contributions

This thesis addresses the challenges of transport infrastructure maintenance by introducing practical decision-support approaches. The proposed solutions are relevant for infrastructure managers and decision makers for informed and data-driven maintenance decision-making. Each chapter of this thesis demonstrates the practicality of the introduced solution tailored to a specific decision-making aspect of maintenance. Each of the proposed methods is validated over the real dataset in close collaboration with practitioners. Given the successful validation of the multi-criteria methods and machine learning models in practice, Irish railways and Rijkswaterstaat have considered extending their asset management practices. In the following, we explain how the key practical contributions of this research work facilitate asset managers and improve the maintenance decision-making practices.

- *Support in tactical-level of maintenance decision-making:* The approaches (introduced in Chapter 3 and 4) based on multi-criteria decision analysis and multi-objective optimization develop multi-year maintenance plans of a large number of assets to guide the investment and planning decisions of maintenance. By following the proposed approach, the decision-makers can successfully (1) prioritize a large number of assets, (2) provide their preferences on different performance goals, (3) analyze the performance of assets over time, and (4) obtain an optimal maintenance plan in terms of budgets and performance requirements.
- *Support subjective analysis of experts in maintenance decision-making:* The predictive models of machine learning (presented in Chapter 5 and 6) can extract complicated non-linear relationship from historical data and can aid the subjective analysis of experts with data-driven inferences and insights. Since the predictive models are trained on the asset management data of the in-use business process, the outcomes are aligned with current decision-making practices of the agency. Therefore, these models can be implemented as an extension to the existing software systems.
- *Enable transparent and traceable decisions:* The solution approaches proposed in this thesis distinctly articulate 1) the objective data and subjective preferences of experts as input, 2) explain the methodological procedures of multi-objective optimization method or machine learning models and 3) yields pragmatic results in the form of multi-year planning or prediction of maintenance. The

shift from experience-driven to the evidence-driven decision-making using the asset management data and decision-support models makes the maintenance decisions transparent, traceable and replicable. The case studies using original datasets of highway bridges and railway switches also prove the utility of proposed methods in practice.

- *Enables budget estimation:* The road and railway government agencies have to estimate the financial requirements over a specified period to obtain funding from central government. One of the critical cost contributing factor of these financial plans is (un)planned maintenance of assets. The predictive models can provide support in the estimation of the future budget by frequency estimation of maintenance triggers. Moreover, the multi-year maintenance planning framework can also determine the number of assets that can be maintained within a given budget by simulating different maintenance scenarios.
- *Enables effective data usage:* The developed models of this study introduce a procedure to effectively use the available asset management data for optimal maintenance planning and predictive modeling. The proposed methods also acquire, process, consolidate, and manage the data that is dispersed across several disparate and independent computerized systems within agencies. The effective use of data enables consistent and explicit decisions by experts and ensure the cost-effective maintenance planning of infrastructure.

The practical usefulness of developed models is validated through the application on case studies from a road and railway agencies of two different countries. A web-based tool of a multi-attribute model is developed to illustrate the model's functionality and potential benefits. During the complementary work, a GIS-based decision support tool is developed to elaborate on the selection of maintenance strategies and their impact on risks and costs for all the assets of the railway network. These practical contributions achieve the overarching goal of this thesis, which was to support infrastructure managers in the decision-making process of maintenance planning by developing applicable decision-support methodologies.

8.5 Recommendations

The outcomes of this research recommend that decision-making process for maintenance planning must be firmly based on the data-driven methods rather than on intuitions of experts. In this study, we suggested several methods and support models

that can aid in the implicit reasoning of the experts. This section outlines recommendations for practitioners to eliminate the gap between the use of data and inherent decision-making practices. The recommendations are built upon the outcomes of this research; the lessons learned during several interactive sessions with road and railway practitioners; and experience of working with the real dataset of the agencies.

- Several computerized management solutions are used to manage the operations and maintenance of transport infrastructure assets. These management systems often follow their innate process logic to manage assets, which may not be aligned with the business practices of agencies. The missing alignment of the IT systems and business practices cause redundant and inconsistent functions/applications and multiple interfaces, which leads to ineffective resource management and poor collaboration among departments. Therefore, there is a need to implement a business-IT alignment procedure within transport agencies to reliably harvest the benefits of IT to achieve business objectives.
- Even though agencies are moving towards advanced technologies such as asset monitoring and data analytic solutions, limited attention is paid to quality control of the data. The attributes (i.e., clean, consistent, conformed, current and comprehensive) that determine the quality of the data are primarily affected due to increased volumes and semi-integrated systems (Tam & Kwan, 2019). Data quality is the most crucial ingredient for building decisions tools and method that support evidence-based asset management (Jardine & Tsang, 2005). Therefore, there is a need to implement data quality control pipeline within agencies to ensure that the resulting models are robust, accurate, and can be trusted for damage assessment and maintenance planning.
- The poor data quality, lack of data integrity and disparate data sources require data integration solutions to provide a unified view of data to decision-makers. Beside the semantic integration, a holistic function-driven system with various modules of budget planning, maintenance planning, traffic planning for all the assets of the network is recommended. Such a system can yield useful analytics in the form of dashboards, and can aid the experts to model prevailing performance conditions, to forecast future scenarios, and to execute the counterfactual analysis.
- The technical knowledge and intuition of experts are key aspects within agencies which drive all the decisions of infrastructure management ranging from inspection to budget estimations. This research recommends extending the decision

processes of asset management to incorporate the feedback and reflection loops. It will help to verbalize the rationale behind the decision as well as will provide an opportunity to store the corporate knowledge of experts for future use.

Most of the recommendations for the practitioners have focused on data management and accessibility challenges. This is because, even though there are several methodologies to improve the reliability of structures, the little attention is paid towards the management of data challenges, and implementation of decision-support methodologies in practices.

8.6 Critical reflection and future research agenda

This thesis introduces several decision support approaches to aid the process of maintenance decision-making. Many design decisions were made during the research process to develop practically useful methods and to manage the scope of the project. This has imposed several limitations. This section outlines the limitations of suggested methods along with the pointers for the future research.

- The multi-criteria methods introduced in Chapters 2-4 do not investigate the selection of specific maintenance treatments. This scope limitation is motivated by the fact that the choice of maintenance treatment is dependent on the reported problem and is often specific to the asset type. The proposed models utilized decision heuristics, driven from the technical knowledge of experts, which suggests the specific treatment options concerning the physical state of the asset (or the asset's component).
- The percentage prediction method, given in Chapter 4, demands data of at-least last two consecutive inspections for the performance prediction of the asset over the multi-year period. This method was adopted, as it is readily applicable in practice and has a transparent approach. In those cases, when sufficient data is not available, the percentage prediction method must be replaced with the probabilistic condition assessment method.
- The solution approach for the development of predictive models using machine learning techniques in Chapter 5 and 6 emphasizes the use of available asset management data from transport agencies. However, in the case of poor and inconsistent data, the machine learning models will give inaccurate outcomes, which can further aggravate the decision-making challenges.

- The maintenance planning and associated decision contexts are dealt from the perspective of infrastructure managers. This implies that the developed maintenance plans provide resource optimization in terms of budget and performance at the (multi) year level and do not assist in operational maintenance planning (scheduling).
- The efforts of this research work have mainly been concentrated on developing applied decision support methods for maintenance decision-making. Even though, the developed models are validated with the real case studies and results have been presented several times to academics and practitioners at workshops and conferences, the empirical evaluation of proposed solutions should be further extended to other types of assets and case studies.
- The discrete types of infrastructure assets, e.g., bridges, switches and crossings have been the primary interest of this work. Therefore, the models may not be readily usable for the continuous asset types, e.g., pavements, tracks, rails unless represented in a discrete, identifiable format.

The use of advanced autonomous monitoring systems such as sensor networks and drone imaging are expected in the near future for infrastructure monitoring and maintenance. This push of digitization in the form of data collection will make the application of machine learning techniques and self-learning algorithms more prevalent. Therefore, to progress further towards data-driven solutions and to address the noted limitation of this research, the future research agenda is established as follows.

- The research laid the foundation towards the digitization of infrastructure maintenance by utilizing artificial intelligence techniques. We demonstrated how to leverage the potential of machine learning algorithms using the available asset management data for effective and efficient maintenance decision making of railway switches and roadway bridges. The process of infrastructure maintenance can be further improved by utilizing the images data that are captured during visual inspection. With the computer vision techniques, the damage characteristics, such as type, severity, depth, length, can be extracted from the images which will facilitate in structural health assessment and maintenance planning of the structure. This will ultimately aid the inspection engineers to precisely record the damage details, whereas an asset manager can recommend the appropriate maintenance treatment by using predictive models.

- This research has also developed a web-based tool for the multi-attribute model (Chapter 2 & 3) and a decision support tool for the cost-benefit analysis (Chapter 7). Additionally, the implementation code for multi-year maintenance planning is also made publicly available (Chapter 5). The tools developed in this study support specific tasks related to maintenance, but do not provide an end-to-end decision support tool. Practitioners in several interactive sessions also highlight the need for comprehensive maintenance planning decision support tool. Hence, future research may seek to combine the capabilities of proposed methods of this research by developing a generic support tool for the discrete type of transport infrastructure assets. The objective of the tool is not only to develop concrete maintenance plans; instead, it should also enable experts to execute various maintenance planning scenarios and perform counterfactual analysis to make informed decisions based on the data.
- The proposed methods of this study utilize the results of condition assessment, collected through visual inspections, for maintenance decision-making. Further research may seek to develop a reliability assessment procedure as a part of developed models to assess the probability of failure of an asset. Additionally, since the data-driven techniques are increasingly being adopted for designing a predictive maintenance program, there is an urgency to conduct comparative studies. Both data-driven technologies and probabilistic assessment methodologies, can develop future performance profiles of structures. The comparative studies may seek to evaluate aspects like the accuracy of results, their applicability, data handling, consideration of uncertainty, among others.
- By working on the real dataset from transportation agencies, we can infer that the agencies suffer from insufficient business and IT alignment. Additionally, the data quality control processes to manage the (historical) data are limited. The future research may conduct an enterprise analysis to improve the alignment of organization's strategy and capabilities of IT. By developing the models of structure and processes of an agency, the new processes can be introduced, such as data management and governance, data quality control units. The enterprise architecture planning can eliminate the redundant processes, semi-integrated systems, and will ensure the business process compliance for the optimized use of resources.
- The knowledge of technical experts and experienced personnel is of utmost importance for the decision-making process of infrastructure management and maintenance. As also mentioned in the recommendations section, the decision-

process needs to be extended to incorporate the feedback and reflection loops to verbalize the rationale behind the decisions and to capture the experts' knowledge. Future research may seek to perform document mining from the past inspection reports in order autonomously extract useful information to develop knowledge bases.

The recommended future research activities aspire to improve the infrastructure maintenance planning activities within agencies by either extending the methods presented in this research or by endeavoring to explore related perspectives. The next section summarizes the development of this research with few closing remarks.

8.7 Closing remarks

At the start of this research, the following objective was defined: *to improve the decision-making process of maintenance planning by developing applied decision support methods and predictive models to aid transport infrastructure managers*. Given the main findings and implication of this research, the researcher is confident that this research contributes to a substantial improvement in the decision-making process of maintenance planning by introducing several practical decision support methodologies. The multi-attribute model, the multi-year maintenance planning framework and machine learning models all offer opportunities to enhance the traditional, experience-driven maintenance approaches towards the evidence-oriented data-driven approaches in orders to provide substantial decision support to infrastructure managers. Given the discussion on the adoption of sensor technologies for the predictive maintenance solutions, this research provides a timely solution on the use of available asset management data for maintenance planning while leveraging the potential of advanced machine learning techniques.

Infrastructure management and maintenance practices can be further advanced in terms of using technology and innovation for their advantage. Rather than using only visual inspection data or implementing wireless sensor networks for condition monitoring of structure, the agencies and researchers can investigate the low-cost data collection procedures. The notables examples include accelerometer data from smartphones (for related work see Mei and Gül (2018), Souza et al. (2018)) and the images from dashboard camera/smartphones (see Maeda, Sekimoto, Seto, Kashiya, and Omata (2018)) through crowdsourcing. For the real-time monitoring of critical infrastructure points, the feasibility of implementing on-device solutions that work as alert management systems can also be further explored. Lastly, the solutions developed

using artificial intelligence, and other digital technologies aim to aid experts in data-driven decision making. Therefore, it paramount to explore, and question on how and why data-driven methods make particular recommendations through interpretability analysis, which is another progressive research area of machine learning.

8.8 References

- Alaswad, S., & Xiang, Y. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157, 54–63.
- Arif, F., Bayraktar, M. E., & Chowdhury, A. G. (2015). Decision support framework for infrastructure maintenance investment decision making. *Journal of Management in Engineering*, 32(1). doi: 10.1061/(ASCE)ME.1943-5479.0000372.
- Bocchini, P., & Frangopol, D. M. (2011). A probabilistic computational framework for bridge network optimal maintenance scheduling. *Reliability Engineering & System Safety*, 96(2), 332–349.
- Bu, G., Lee, J., Guan, H., Loo, Y., & Blumenstein, M. (2014). Prediction of long-term bridge performance: Integrated deterioration approach with case studies. *Journal of Performance of Constructed Facilities*, 29(3), 04014089.
- Bush, S., Henning, T., Ingham, J. M., & Raith, A. (2014). An agent-based framework for improved strategic bridge asset management. In *Austrroads bridge conference, 9th, 2014, sydney, new south wales, australia* (4.4).
- Chootinan, P., Chen, A., Horrocks, M. R., & Bolling, D. (2006). A multi-year pavement maintenance program using a stochastic simulation-based genetic algorithm approach. *Transportation Research Part A: Policy and Practice*, 40(9), 725–743.
- Cinelli, M., Coles, S. R., & Kirwan, K. (2014). Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. *Ecological Indicators*, 46, 138–148.
- De Montis, A., De Toro, P., Droste-Franke, B., & Omann. (2004). Assessing the quality of different mcda methods. *Alternatives for environmental valuation*, 4, 99.
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., Garcez, T. V., et al. (2015). *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis*. New York, NY, USA: Springer, 2015.
- de Bruin, T., Verbert, K., & Babuška, R. (2017). Railway track circuit fault diagnosis using recurrent neural networks. *IEEE transactions on neural networks and learning systems*, 28(3), 523–533.
- Denysiuk, R., Fernandes, J., Matos, J. C., Neves, L. C., & Berardinelli, U. (2016). A computational framework for infrastructure asset maintenance scheduling. *Structural Engineering International*, 26(2), 94–102.
- Dong, Y., Frangopol, D. M., & Sabatino, S. (2015). Optimizing bridge network retrofit planning based on cost-benefit evaluation and multi-attribute utility associated with sustainability. *Earthquake Spectra*, 31(4), 2255–2280.

- Frangopol, D. M., & Liu, M. (2007a). Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost. *Structure and infrastructure engineering*, 3(1), 29–41.
- Frangopol, D. M., & Liu, M. (2007b). Bridge network maintenance optimization using stochastic dynamic programming. *Journal of Structural Engineering*, 133(12), 1772–1782. doi: 10.1061/(ASCE)0733-9445.
- Galar, D., Kans, M., & Schmidt, B. (2016). Big data in asset management: Knowledge discovery in asset data by the means of data mining. In *Proceedings of the 10th world congress on engineering asset management (wceam 2015)* (pp. 161–171). Springer.
- Garmabaki, A. H. S., Ahmadi, A., & Ahmadi, M. (2016). Maintenance Optimisation Using Multi-attribute Utility Theory. In U. Kumar, A. Ahmadi, A. K. Verma, & P. Varde (Eds.), *Current trends in reliability, availability, maintainability and safety: An industry perspective* (pp. 13–25).
- Ghodoosi, F., Abu-Samra, S., Zeynalian, M., & Zayed, T. (2017). Maintenance cost optimization for bridge structures using system reliability analysis and genetic algorithms. *Journal of Construction Engineering and Management*, 144(2), 04017116.
- Gualtieri, M. (2016). Hadoop is data's darling for a reason. Available: <https://go.forrester.com/blogs/hadoop-is-datas-darling-for-a-reason/>. [Accessed: 08-Apr-2018].
- Jardine, A. K., & Tsang, A. H. (2005). *Maintenance, replacement, and reliability: Theory and applications*. CRC press.
- Kabir, G., Sadiq, R., & Tesfamariam, S. (2014). A review of multi-criteria decision-making methods for infrastructure management. *Structure and Infrastructure Engineering*, 10(9), 1176–1210.
- Kiani, J., Camp, C., & Pezeshk, S. (2019). On the application of machine learning techniques to derive seismic fragility curves. *Computers and Structures*.
- Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., & Hampapur, A. (2014). Improving rail network velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C: Emerging Technologies*, 45, 17–26.
- Maeda, H., Sekimoto, Y., Seto, T., Kashiwama, T., & Omata, H. (2018). Road damage detection using deep neural networks with images captured through a smartphone. *arXiv preprint arXiv:1801.09454*.
- Manco, G., Ritacco, E., Rullo, P., Gallucci, L., Astill, W., Kimber, D., & Antonelli, M. (2017). Fault detection and explanation through big data analysis on sensor streams. *Expert Systems with Applications*, 87, 141–156.
- Masino, J., Thumm, J., Frey, M., & Gauterin, F. (2017). Learning from the crowd: Road infrastructure monitoring system. *Journal of Traffic and Transportation Engineering (English Edition)*, 4(5), 451–463.

- Mei, Q., & Gül, M. (2018). A crowdsourcing-based methodology using smartphones for bridge health monitoring. *Structural Health Monitoring*.
- Morales, F. J., Reyes, A., Caceres, N., Romero, L., & Benitez, F. G. (2018). Automatic prediction of maintenance intervention types in roads using machine learning and historical records. *Transportation Research Record*, 0361198118790624.
- Morcous, G., & Lounis, Z. (2005). Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction*, 14(1), 129–142.
- Niestadt, M., Debyser, A., Scordamaglia, D., & Pape, M. (2019). *Artificial intelligence in transport. current and future developments, opportunities and challenges*. European Parliamentary Research Service. European Union.
- Patidar, V. (2007). *Multi-objective optimization for bridge management systems*. Transportation Research Board.
- Souza, V. M., Giusti, R., & Batista, A. J. (2018). Asfalt: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning. *Pervasive and Mobile Computing*, 51, 121–137.
- Stipanovic, I., Høj, N., & Klanker, G. (2016). Multi-objective bridge performance goals. In *Life-cycle of engineering systems: Emphasis on sustainable civil infrastructure* (pp. 280–288). CRC Press.
- Stipanovic, Chatzi, Limongelli, Gavin, Bukhsh, A., & Xenidis. (2017). *WG2, Technical Report, Performance Goals for Roadway Bridges OF COST ACTION TU 1406*. Retrieved from TU1406 Website: <https://bit.ly/2Seh3NZ>.
- Strauss, A., Ivankovic, A. M., Matos, J. C., & Casas, J. (2016). Wg1 technical report: Performance indicators for roadway bridges of cost action 1406.
- Tam, A., & Kwan, I. (2019). Data quality in asset management—creating and maintaining a foundation for data analytics. In *Asset intelligence through integration and interoperability and contemporary vibration engineering technologies* (pp. 567–574). Springer.
- Velasquez, M., & Hester, P. T. (2013). An analysis of multi-criteria decision making methods. *International Journal of Operations Research*, 10(2), 56–66.
- Wijnia, Y., & Herder, P. (2010). The state of asset management in the netherlands. In *Engineering asset lifecycle management* (pp. 164–172). Springer.
- Xie, H. B., Wu, W.-J., & Wang, Y.-F. (2018). Life-time reliability based optimisation of bridge maintenance strategy considering lca and lcc. *Journal of Cleaner Production*, 176, 36–45.
- Zavadskas, E. K., Turskis, Z., & Kildienė, S. (2014). State of art surveys of overviews on mcdm/madm methods. *Technological and economic development of economy*, 20(1), 165–179.

Robustness of the multi-attribute utility model for bridge maintenance planning

¹ **Abstract:** Optimisation of maintenance planning is an essential part of bridge management. With the purpose to support maintenance planning, a multi-objective decision-making model is introduced in this paper. The model is based on multi-attribute utility theory, which is used for the optimisation process when multiple performance goals have to be taken into account. In the model, there are several parameters, which are freely chosen by the decision maker. The model is applied to the inventory of 22 bridges, where four Key Performance Indicators were determined for four performance aspects: reliability, availability, costs and environment. A sensitivity analysis is performed by changing risk tolerance parameter and attribute weights to determine the robustness of the model. The Multi-Attribute Utility model and sensitivity analysis presented in this paper will help decision-makers to examine the robustness of the optimal solution by dynamically changing the critical parameters.

¹This chapter has been published as: Allah Bukhsh, Z., Stipanovic, I., Palic, S. S., & Klanker, G. (2018). Robustness of the Multi-Attribute Utility Model for Bridge Maintenance Planning. *The Baltic Journal of Road and Bridge Engineering*, 13(4), 404-415.

Introduction

Functional and serviceable road infrastructure presents one of the most integral predispositions for the economic growth of countries around the world. One of the most important parts of road infrastructure are bridges which present a vital link in any roadway network. It is estimated that the ratio of expenses per route km of bridges or tunnels is 10 times the average expenses per route km of roads (Egger, 2012). Also, the length of bridges compared to the whole length of the road network is only approximately 2%, but at the same time, they present 30% value of the whole network (Das, Micic, & Chryssanthopoulos, 1999). When these statistics are taken into consideration, it is easy to understand why, an increasing number of deteriorating bridges led to the development of many Bridge Management Systems (BMS) and life cycle maintenance models (ERA-NET ROAD, 2012). Infrastructure managers are facing conflicting requirements, to improve the availability and serviceability of ageing infrastructure while budget restrictions constrain the maintenance planning. Many research efforts are ongoing ranging from development of BMS, optimisation models, life-cycle Cost Analysis, to big data analysis and implementation of artificial intelligence models into decision support tool (Allah Bukhsh, Saeed, Stipanovic, & Doree, 2019; Nunez, Hendriks, Li, De Schutter, & Dollevoet, 2014).

Since transport infrastructure is deeply embedded in society, it is not only subject to technical requirements, but it must also keep up with societal and economic developments. Therefore, bridge maintenance planning must accommodate multiple performance goals that need to be quantified by various performance indicators and Key Performance Indicators (KPIs) (Bukhsh et al., 2017; Strauss, Ivankovic, Matos, & Casas, 2016). The application of Multi-Attribute Utility Theory (MAUT) for bridge maintenance planning is presented in detail in a previous study Allah Bukhsh, Stipanovic, Klanker, O'Connor, and Doree (2018). This paper builds on the previous study with the aim to determine the robustness of the proposed model. The MAUT model was applied to the group of 22 highway bridges, for which the Condition Index (CI), chosen maintenance activity and Maintenance Costs (MC) were known. Additionally, User Delay Costs (UDC) and Environmental Costs (EC) were determined. The goal was to optimise multiple objectives by suggesting a trade-off among them and finally assign a ranking to the bridges considered. Utility functions of MAUT appropriately account for the involved uncertainty and risk attitude of infrastructure managers. Therefore, the purpose of this paper is to determine the robustness of the model through sensitivity analysis, by alternating risk attitude through the risk tolerance parameters and performance attribute weights for all

performance aspects. Multi-Attribute Utility Theory provides a systematic approach to decision making by accommodating multiple performance goals, uncertainty and preferences of infrastructure managers, thus enabling complex problems involving many parameters to be solved.

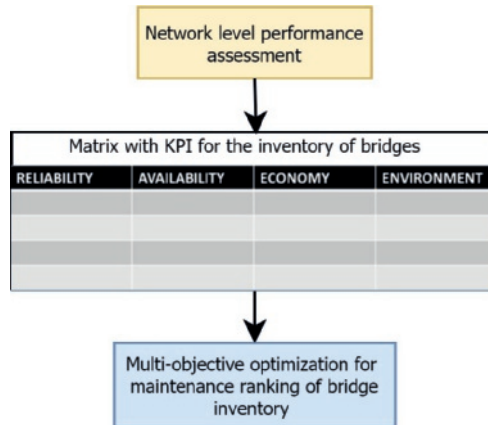


Figure A.1: Overview of the process for performance goals assessment.

Modern decision-making processes dealing with bridge management have to go far beyond choosing an optimal solution based on just single indicators (i.e. the lowest long-term cost). There is a need to have decision models that are capable of implementing multiple performance criteria as shown in Figure A.1.

Multi-attribute utility theory

Utility theory provides a measure of preferences of a decision maker over a group of alternatives (Ishizaka & Nemery, 2013). Based on the six axioms of utility theory, MAUT is introduced by (Keeney & Raiffa, 1993). Multi-Attribute Utility Theory provides a systematic approach to reduce the qualitative values of various attributes (i.e. performance indicators) into utility functions. The obtained utility scores are then aggregated based on the relative importance of attributes. The final score assigns a ranking to each alternative based on either minimisation or maximisation function. In other words, MAUT assigns the relative importance of performance indicators (e.g. condition, costs), when comparing a number of bridges. These bridges are often referred to as alternatives in MAUT.

MAUT involves the single decision maker who is willing to make certain trade-off among the performance goals while exposed to uncertainty and risk (Keeney & Raiffa,

1993). The uncertainty is usually originated because of unavailable and dynamic nature of data, and involvement of multiple stakeholders. For instance, in the bridge planning the exact estimation of a number of users affected due to maintenance activity is difficult to define. Multi-Attribute Utility Theory integrates a body of mathematical utility models and a range of decision assessment methods to assist in decision ranking problem (Thevenot, Steva, Okudan, & Simpson, 2006). The single attribute utility function is calculated for each performance aspect which reflects the risk attitude of the decision maker. The mathematical formulation of MAUT is represented as follows:

$$U(x) = k_1U(x_1) + k_2U(x_2) + \dots + k_nU(x_n) \quad (\text{A.1})$$

where $U(x)$ - multi-attribute utility value of each alternative x ; k - a scaling constant that provides the relative importance of each performance indicator (attribute i); $U_i(x_i)$ - a single attribute utility value of each performance indicator i for the alternative x .

$$U_i(x_i) = A - B * e^{\left(\frac{-x_i}{RT}\right)} \quad (\text{A.2})$$

where A and B - scaling constants; RT- risk tolerance.

The general steps to apply MAUT on decision-making problem, e.g. maintenance planning are summarised as follows:

1. Identify the decision objectives and define the attributes relevant to the problem;
2. Quantify the attributes in a form that structures and represent the defined decision objectives and goals in utility functions;
3. Calculate the single utility function for each attribute by estimating the indifference point(s) and risk attitude of a decision maker(s). These steps establish a relationship between the attributes values and their utility scores based on preferences structures of the decision maker(s);
4. Determine the relative importance of attributes build on the weighting assigned by the decision maker(s);
5. Compute the aggregative utility score for each alternative by either multiplicative form of additive form. The total aggregative score ranks the alternatives, where an alternative that is the perfect fit in a realisation of decision objective is ranked at highest.

Case study

An example case is provided to illustrate the application of MAUT for the bridge maintenance planning. The objective of this decision-making exercise is to rank the bridges alternatives regarding four KPIs reliability (KPI-Condition Index (CI)), economy (KPI-Maintenance Costs (MC)), environment (KPI-Environmental Costs (EC)) and availability (KPI-User Delay Costs (UDC)). The decision problem of maintenance planning presented in this case study requires the ranking of 22 bridges in an order where the condition level can be maximised, and at the same time MC, UDC and EC can be minimised. It is noted that the minimisation of one attribute might result in maximisation of the other one. For instance, to minimise the UDC an agency needs to employ more resources which result in increased owner costs. Therefore, a trade-off among these attributes has to be performed (Borgonovo & Cillo, 2017). With the definition of KPIs, the single utility function of each attribute is calculated. In this exercise, authors played the role of a decision maker to estimate the indifference point and the general risk attitudes. A decision maker is provided with a lottery question representing the 50-50% probability of having best (i.e. minimum MC) and worst (i.e. maximum MC) as shown in Figure A.2. The median value between the maximum and minimum MC is called the Expected Value (EV), which is 123.73. In practice, an

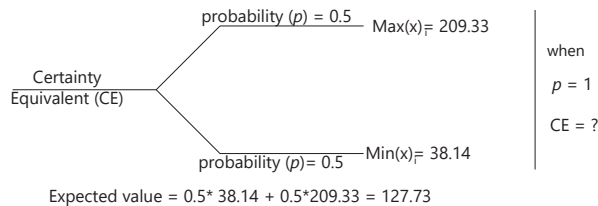


Figure A.2: Lottery question to discern Maintenance Costs.

owner is often unable to achieve minimum costs as desired. Therefore, MAUT has a concept of Certainty Equivalent (CE) which is the indifference point of a decision maker between the maximum (worst) and minimum (best) maintenance costs. In this case, the chosen CE is 90. Considering the risk tolerance value of 70, Eq. A.2 becomes:

$$U_i(x_i) = 1.09 - 1.88 * e^{\left(\frac{-x_i}{70}\right)} \quad (\text{A.3})$$

The single utility function of MC (i. e. $U(MC)$) reduces the values from 0 to 1 representing the utility values of real numbers concerning the defined objective. Figure A.3 shows the graph of MC concerning the assigned utility values.

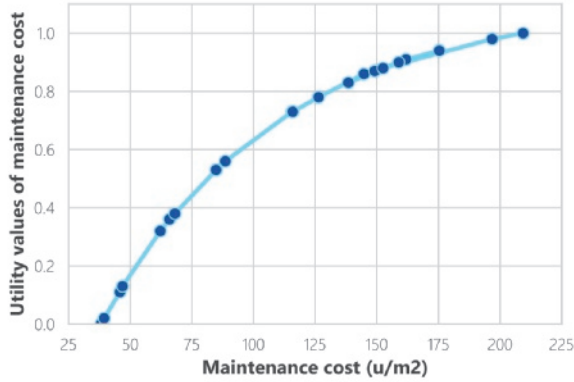


Figure A.3: Single Utility Score of Maintenance Costs.

Single utility scores of downtime (UDC), condition rating (CI) and environmental costs (EC) are computed in the same manner. Table 1 shows the actual data and the computed single utility value of each performance indicator. Finally, to obtain the total aggregative value for each bridge, where the additive multiple attribute function shown in Eq. A.1 is used. The relative importance of performance indicators is defined by k factor considering the possibility of having the multiple performance goals. A direct rating method is used as represented below:

$$k(x_i) = \frac{rate(x_i)}{\sum_{j=1}^n(x_j)} \tag{A.4}$$

where $k(x_i)$ - weighting factor of each attribute i across all alternatives; $rate(x_i)$ - rate/weight assigned by an expert for attribute i . Based on the aggregated values, the ranking of the bridges can be performed, where multiple performance goals are taken into account, i.e. the MC is kept at minimum and CI is maximized.

Table A.1: Multi-Attribute Utility model results for a group of 22 bridges

Bridge No.	CI	MC	EC	UDC	U(CI)	U(MC)	U(EC)	U(UDC)	Aggregated value
1	3.10	144.82	0.89	39.70	0.94	0.98	0.85	0.93	0.94
2	1.89	126.41	0.21	27.50	0.29	0.96	0.34	0.82	0.39
3	2.21	115.67	0.57	25.57	0.58	0.94	0.70	0.79	0.63
4	3.13	161.85	1.11	13.64	0.95	0.99	0.91	0.48	0.88
5	2.00	68.16	0.53	12.40	0.41	0.67	0.67	0.43	0.43
6	2.12	149.21	0.23	47.89	0.52	0.98	0.37	0.97	0.60
7	3.36	196.76	0.71	57.79	0.99	1.00	0.77	0.99	0.98
8	2.42	88.60	1.25	13.11	0.71	0.85	0.94	0.46	0.68
9	2.22	45.82	1.26	35.89	0.59	0.25	0.94	0.91	0.64
10	2.34	115.93	0.43	30.80	0.67	0.94	0.59	0.86	0.70
11	2.42	39.42	0.23	12.69	0.71	0.05	0.38	0.44	0.64
12	3.46	138.52	1.85	12.12	1.00	0.98	1.00	0.42	0.91
13	1.92	38.14	0.03	7.99	0.32	0.00	0.05	0.21	0.29
14	2.18	84.89	1.05	14.42	0.56	0.82	0.90	0.51	0.57
15	2.43	46.89	0.00	4.59	0.72	0.28	0.00	-0.01	0.57
16	1.67	175.33	0.68	28.51	0.00	0.99	0.76	0.83	0.18
17	3.17	209.33	0.39	55.25	0.96	1.00	0.55	0.99	0.95
18	2.30	158.89	0.22	51.04	0.64	0.99	0.36	0.98	0.70
19	2.58	65.90	0.10	8.79	0.79	0.64	0.17	0.26	0.68
20	1.96	62.22	0.42	22.83	0.37	0.59	0.58	0.74	0.44
21	2.02	84.82	0.28	25.70	0.43	0.82	0.43	0.79	0.50
22	2.34	152.60	0.27	42.91	0.67	0.99	0.42	0.95	0.71

Robustness assessment of Multi-Attribute Utility model for risk attitude

The risk attitude of the decision maker is categorized as risk-taking, risk-averse, and risk-neutral. Figure A.4 shows the resulting utility graph based on the risk attitude of the decision maker. The utility values are shown by plotting the attribute values in x-axis and utility values on y-axis ranging from 0 to 1. A risk avoiding attitude result into a concave down graph of utility values whereas risk-taking preferences show the concave up a graph of utility values. The robustness of the multi-attribute model is assessed by performing the local sensitivity analysis. The local sensitivity analysis captures the effect on the output of the model due to a small change in input parameters., There are two subjective measures in the multi-attribute model that must be defined by a decision maker. The subjective measures are risk attitude of decision

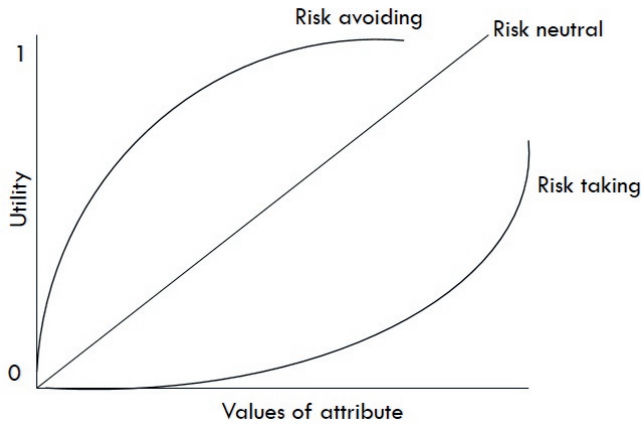


Figure A.4: Risk Attitudes of the decision maker.

makers and the weights assigned to each attribute by decision maker as shown in Section A.

The multi-attribute model is executed with the same set of 22 bridges to analyse the change in the ranking of bridges due to different risk attitudes. All the attributes were assigned equal weights of 0.5 since the idea is to access the difference in ranking with the change in risk attitude only. Figure A.5 shows the ranking of 22 bridges concerning risk attitude. The shorter bar represents the higher rank; the longer bars represent low rank. risk ranking with RT

Three main trends in the ranking of the bridges with different risk scores are noticed. First, there was no or minor difference in ranking of the bridge with different risk attitude, e.g. B7, B13, B17, B22, and B2. Second, with risk avoiding attitude the bridge ranked higher than with risk-taking. For instance, with risk avoiding, B1 is ranked as the second highest because the condition score (CI) of B1 is considerable high while the MC are low. The similar pattern is noticed with B10 and B8. The third and final pattern shows the risk-taking attitude has assigned higher scores to the bridge as compared to the risk avoiding. Take the example of B6 where risk-taking has ranked it at number 7, while with risk avoidance it is ranked at 12, the difference in rank is because of the higher magnitude values of MC and UDC of B6. To summarise, in addition to the risk attitude of a decision maker, the actual magnitude of the values plays a more significant role in ranking. It is because a decision maker states his risk attitude over actual data values, instead of computed utility scores.

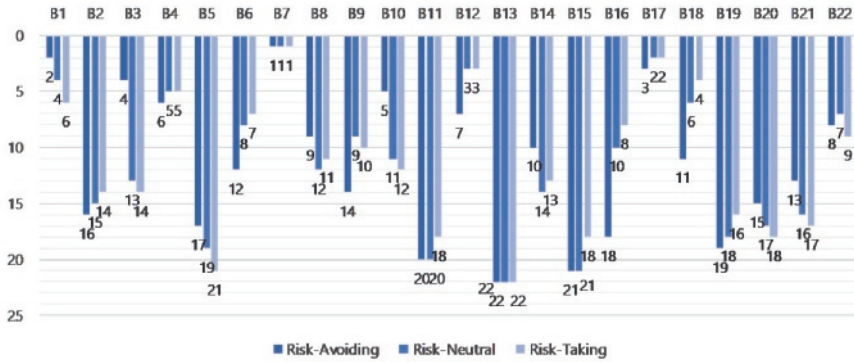


Figure A.5: Ranking of bridges with three risk attitudes.

Robustness assessment of Multi-Attribute Utility model for attributes weights

Similar to risk attribute assessment, the sensitivity analysis is performed to assess the effect on the ranking of a bridge by changing the weights assigned to each attribute. A single-attribute and two-attribute sensitivity analysis are performed to analyse the sensitivity of ranking concerning attributes weights. In the single-attribute analysis, the weights of a single attribute are changed over the range from 0.1 to 0.9 while the weights for other attributes were kept as small as 0.05. Table A.2 shows the result of one-attribute sensitivity analysis outlining the highest ranked bridge. The result of the one-attribute analysis shows that the irrespective of assigned weights the highest ranked does not change considerably. However, a substantial change in the ranking of the bridge is noted as the difference between the dynamic weight assigned to a single attribute (ranging from 0.1 to 0.9) increases compared to the constant weight of other attributes (e.g. 0.7 or 0.9). An interested reader may refer to the MAUT online tool for further analysis.

Table A.2: Single-attribute sensitivity analysis for each attribute

Attributes	Weights								
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Condition Index	B7	B12	B12	B12	B12	B12	B12	B12	B12
Maintenance Costs	B7	B7	B7	B7	B7	B7	B7	B7	B17
Environmental Costs	B12	B12	B12	B12	B12	B12	B12	B12	B12
User Delay Costs	B7	B7	B7	B7	B7	B7	B7	B7	B7

Moreover, two-attributes sensitivity analysis is also conducted to see the effect on the ranking of bridges, when the weights of two attributes are changed simultaneously. Table A.3 shows the highest ranked bridge generated by changing the weights of CI over row and MC over the column. The result of the two-attribute analysis suggests that the model is robust and the assigned weights do not influence much on the bridge ranks.

Table A.3: Condition Index and Maintenance Costs sensitivity analysis

CI vs MC	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80
0.10	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7
0.20	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7	
0.30	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7		
0.40	Bridge 7	Bridge 7	Bridge 7	Bridge 7	Bridge 7			
0.50	Bridge 7	Bridge 7	Bridge 7	Bridge 7				
0.60	Bridge 7	Bridge 7	Bridge 7					
0.70	Bridge 7	Bridge 7						
0.80	Bridge 7							

The possible reason for the static rank of the bridges while having variable weights of the attribute is because the weights are assigned to the calculated utility scores as shown in Section A. It is noted that the ranking of the bridges is sensitive to the risk attitude of a decision maker, where the preference is defined over the real values of attributes.

Conclusions

Within the Working Group 2 of COST Action TU1406 Quality Specifications for Roadway Bridges, Standardization at a European Level (BridgeSpec), a multi-objective decision-making model is developed to support bridge maintenance planning. The model was applied to the group of twenty-two bridges, where a trade-off among different performance attributes had to be performed. In the study four performance aspects, reliability, availability, cost and environment were quantified and used as an input parameter for Multi-Attribute Utility model. For each performance attribute a single utility function has been determined, and finally, the aggregative utility score for each alternative has been computed by either multiplicative form of additive form. The total aggregative score is then used for ranking the alternatives, where an alternative which is the perfect fit in a realisation of decision objective is ranked at highest. The primary purpose of this paper was to determine the robustness of

the model. The sensitivity analysis is conducted by alternating risk attitude through the risk tolerance parameter and performance attribute weights for all performance aspects. Principal conclusions from this study are the following:

1. Utility functions of Multi-Attribute Utility Theory appropriately account for the involved uncertainty and risk attitude of infrastructure managers. Multi-Attribute Utility Theory provides a systematic approach for decision making of maintenance planning by accommodating multiple performance goals, uncertainty and preferences of infrastructure managers thus enabling complex problems involving many parameters to be solved.
2. Regarding the impact of risk attitude on the final ranking, there was no or minor difference in ranking of bridges with different risk attitude. In addition to the risk attitude of a decision maker, the actual magnitude of the values plays a significant role in final ranking of the alternatives.
3. A single-attribute and two-attribute sensitivity analysis are conducted to assess the effect of the weights assigned to each performance attribute. In the single-attribute analysis, the weights of a single attribute are changed over the range from 0.1 to 0.9 while the weights for other attributes were kept as small as 0.05. The results of the one-attribute analysis show that the irrespective of assigned weights the highest ranked does not change considerably. An online tool is made available to enable the reader for further analysis. The results of the two-attribute analysis suggest that the model is robust and the assigned weights do not influence much on the bridge ranks.

Finally, we can conclude that the implementation of Multi-Attribute Utility Theory model can help decision-makers to find the optimal solution for the bridge maintenance planning while taking multiple performance goals into account.

References

- Allah Bukhsh, Z., Stipanovic, I., Klanker, G., O'Connor, A., & Doree, A. G. (2018). Network level bridges maintenance planning using multi-attribute utility theory. *Structure and infrastructure engineering*, 1–14.
- Allah Bukhsh, Z., Saeed, A., Stipanovic, I., & Doree, A. G. (2019). Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies*, 101, 35–54.
- Borgonovo, E., & Gillo, A. (2017). Deciding with thresholds: Importance measures and value of information. *Risk Analysis*, 37(10), 1828–1848.
- Bukhsh, Z. A., Oslakovic, I. S., Klanker, G., Hoj, N. P., Imam, B., & Xenidis, Y. (2017). Multi-criteria decision making: Ahp method applied for network bridge prioritization. In *Joint cost tu1402–cost tu1406–iabse wc1 workshop: The value of structural health monitoring for the reliable bridge management*.
- Das, P., Micic, T., & Chryssanthopoulos, M. (1999). *Reliability-based assessment of highway bridges*.
- Egger, M. (2012). Benchmarking of expenditures and practices of maintenance and operation (bexprac). *Procedia-Social and Behavioral Sciences*, 48, 1733–1742.
- ERA-NET ROAD. (2012). Asset service condition assessment methodology (ascam project), Statens väg-och transportforskningsinstitut.
- Ishizaka, A., & Nemery, P. (2013). *Multi-attribute utility theory. Multi-Criteria Decision Analysis: Methods and Software*, 81–113.
- Keeney, & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press.
- Nunez, A., Hendriks, J., Li, Z., De Schutter, B., & Dollevoet, R. (2014). Facilitating maintenance decisions on the dutch railways using big data: The aba case study. In *2014 ieee international conference on big data (big data)* (pp. 48–53). IEEE.
- Strauss, A., Ivankovic, A. M., Matos, J. C., & Casas, J. (2016). Wg1 technical report: Performance indicators for roadway bridges of cost action 1406.
- Thevenot, H. J., Steva, E. D., Okudan, G. E., & Simpson, T. W. (2006). A multi-attribute utility theory-based approach to product line consolidation and selection. *International Design Engineering Technical Conferences*, 441–450. doi: 10.1115/DETC2006-99506.

List of Acronyms

AHP	Analytical Hierarchy Process
BCI	Bridge Condition Index
BMS	Bridge Management Systems
CBM	Condition Based Maintenance
CE	Certainty Equivalent
CI	Condition Index
CMMS	Computerized Maintenance Management System
DST	Decision Support Tool
EC	Environmental Cost
ELECTRE	ELimination and Choice Expressing REality
ERP	Enterprise Resources Planning
EV	Expected Value
FN	False Negative
FP	False Positives
GAs	Genetic Algorithms
GBT	Gradient Boosting Tree
GIS	Geographic Information System
IMA	Inspection to Maintenance Advice
IMS	Information Management System
LCC	Life Cycle Cost
MAUT	Multi-Attribute Utility Theory
MAVT	Multi-Attribute Value Theory
MCDA	Multi-Criteria Decision Analysis
MCP	Markov Chain Process

ML Machine Learning
MMPF Multi-year Maintenance Planning Framework
MRP Maintenance Request Process
MTL-NN Multi-Task Learning Neural Networks
NN Neural Networks
NN-EE(cw) NN-EE with class weights
NN-EE Neural Networks with Entity Embeddings
OC Owner Cost
PdM Predictive Maintenance
PROMETHEE Preference Ranking Organization Method for Enrichment of Evaluations
RF Random Forest
RT Risk Tolerance
SCV Stratified Cross Validation
SLA Service Level Agreement
SMOTE Synthetic Minority Over-Sampling Technique
SRS Stratified Random Sampling
SUF Single Utility Function
tf-idf Term Frequency-Inverse Document Frequency
TN True Negatives
TOPSIS Technique for Order by Similarly to Ideal Solution
TP True Positives
UDC User Delay Cost

Acknowledgement

Getting the PhD is like having my wildest dream coming into reality. This journey was full of excitement, frustration, and uncertainty. Nevertheless, several people provided support and motivation to keep me going. Here I would like to thank them all.

I want to thank my promotor, Prof. Andre G. Doree, for giving me feedback, supporting me towards completing my PhD and assisting me with all the necessary protocols. I am indebted to Dr. Irina Stipanovic for her extensive guidance, for constructively challenging my ideas, and teaching me a great deal about scientific research and civil engineering domain. You have acted more than a supervisor to me. Without your encouragement, persistence, and guidance, this research would have never been possible. I would like to express my gratitude to the graduation committee members for their effort and time spent while reviewing my thesis.

Special thanks go to H2020 research and innovation program for funding this research by DESTination RAIL project. I would also like to acknowledge the funding opportunities by COST Action TU1406 for early-stage researchers. I am thankful to the consortium members of DESTination RAIL and COST ACTION TU1406 project for inspiring me with their work and providing me helpful comments. I want to thank Prof. Ken Gavin from TU delft for being one of the committee members of the thesis and providing support at several points during my research. I owe deepest gratitude to Irish Railways and Rijkwaterstaat for the collaboration opportunities during this research. Especially, thanks to Cathal Mangan and Alan Cummins for being exceptional hosts during my stay at Irish rail and for being generous in sharing and explaining the (much needed) data. I owe a special thanks to Giel Klanker for sharing the bridges' data and co-authoring the very first journal paper. I would also like to thank Jaap Bakker from Rijkwaterstaat for several insightful discussions and comments during the last stage of the project. I am grateful to Prof. Alan O' Conner from Trinity College Dublin for his invaluable suggestions on the very first journal paper of the PhD. In the CME department, I would like to share my appreciation for all the colleagues. The

lunch presentations followed by the discussion have been immensely helpful to set the course of my research. Thanks to the cohabiting colleagues for providing the right balance of the quiet and cheerful working environment. I am thankful to Yolanda Bosch and Jacqueline Nijhof for seamlessly managing all the administrative matters. A special mention goes to Sergei Miller for always being uplifting and motivating.

I owe thanks to Sandra and again to Irina for hosting me for a much-needed summer break. I am also thankful and happy to be part of the Pakistani student association. At times, the arrangements of the kick-in stand, the international food festival, BBQs, and bowling events were all I needed to refresh my thoughts.

Some events and people have indirectly played a vital role to set me towards the path of the PhD. I would like to show deepest appreciation to the University of Twente Scholarship program, which made the Master studies possible at Utwente. I want to express my gratitude to Marten van Sinderen for reviewing this thesis and for teaching me the basis of research. Through his supervision during the Master thesis, I felt confident to follow my vision of doing a PhD. I owe a special thanks to the teachers from Islamia university, Bahawalnagar campus for coaching me with the very basics of computer science, teaching me professionalism and for cultivating a 'can-do' attitude in me. I am in short of words to thank my parents for their generous love, support, and guidance. I cannot thank both of you enough for believing in me and giving me freedom. I am thankful to my sister Faiza for being the one that has a solution to all the problems irrespective of their academic or personal nature. You are one strong pillar that I can always lean-on. A special thanks to my brother-in-law for being constant support since the time I step in the Netherlands. I am grateful to my brother, Bilal, for being a one-stop-solution for all the things that needed to be done in Pakistan. Thanks to my sister, Javara, for being available for all sorts of gossips, and for always telling me that 'you will do it, you just complain a lot'.

Profound gratitude to Aaqib. Thanks for sharing your life with me and being my cheerleader. I am thankful to you for making me believe in myself, for our endless discussions about my research, and for guiding me in machine learning work. With you, all the potential futures are possible.

I look forward to all the opportunities to continue to carry the torch of knowledge forward.

About the Author

Zaharah Allah Bukhsh was born in 1992 in Bahawalnagar Pakistan. She completed Bachelor of Computer Science at her hometown from The Islamia University of Bahawalnagar. She secured gold-medal (equivalent to cum-laude) for her excellent academic performance. In 2013, she got admission in Master of Computer Science in University of Twente (UT) followed by UT scholarship. In 2015, she finished her MSc and started her PhD research in Department of Construction Management and Engineering at UT. During her research, she has secured two short-term scientific grants and visited Trinity College Dublin and University of Zagreb as a visiting researcher. Zaharah has published about her research in several high-impact journals and conference proceedings. She has also presented her research at several workshops, project meetings, conferences, and road and railway agencies.

